Loughborough University
Institutional Repository

Applied statistics, microcomputers and analytical chemistry

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: [https://dspace.lboro.ac.uk/2134/13781](https://dspace.lboro.ac.uk/2134/13781)

Publisher: © Gerald N. Killoran

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the
author and is made available in the Institutional Repository
(https://dspace.lboro.ac.uk/) under the following Creative Commons Licence
conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
APPLIED STATISTICS
MICROCOMPUTERS
and
ANALYTICAL CHEMISTRY

by

GERALD N. KILLORAN, B.Sc., M.Phil.

Doctoral Thesis
Submitted in partial fulfilment of the requirements
for the award of
Doctor of Philosophy
of the Loughborough University of Technology
March, 1984

Supervisor: Dr. J.F. Tyson
Department of Chemistry

(C) by Gerald N. Killoran, 1984
ACKNOWLEDGEMENTS

I would like to thank my research supervisor, Dr. J.F. Tyson, for his patience and guidance throughout the course of my studies. He has had a difficult task in attempting to guide my activities from a distance of a few thousand miles. I would also like to express a general thank-you to Professors Bentley, and Miller, and to the University for allowing me to conduct the major part of my research in Canada as an external candidate.

My studies have been aided by the use of the library, office, and computer facilities of the Northern College of Applied Arts and Technology, Ontario, Canada. Their support, and the assistance of my college colleagues, is gratefully acknowledged. I have also used the library facilities of many universities, e.g., Carleton, Toronto, British Columbia, and York; without access to their facilities this research would not have taken place.

Lastly, I wish to express my gratitude to my wife for her assistance in proofreading this thesis, and for her understanding and patience over the last several years.

Gerald N. Killoran

March, 1984
ABSTRACT

An applied statistics software package, containing a unique weighted linear regression (WLR) routine, has been developed. Its features are demonstrated using real and Monte Carlo simulated data. The WLR routine is particularly useful for absolute and comparative calibrations.

In the absolute calibration of analytical systems the statistical analysis of the linear calibration curve produces an analysis of variance (ANOVA), the curve's equation and confidence band, the regression coefficient, confidence limits for the slope and intercept, and the standard error of regression. The routine also computes an unknown's concentration and its "confidence" limits.

Both simple (SLR) and WLR can be used for absolute calibration. WLR can be used without knowing the error's standard deviation (SD); assuming the analytical error is normally distributed, the SD is a linear function of concentration or response, and the concentration range of interest is well above the detection limit. Under these conditions the computed "standard error of regression" is the "relative SD" when using WLR, or the "SD" for SLR.

Comparative calibration is used for method validation and for determining the relative economic and technical merits of analytical systems.

Ways of estimating a system's precision, as a function of concentration, are discussed. Two new, simple approaches are demonstrated. The comparison of a new analytical system to one of known accuracy, using SLR and WLR, is reviewed.

A previously reported technique for determining the merits of analytical systems, using only "raw" measurements, is reviewed and demonstrated for systems having constant SDs, RSDs, or both. The effect of transforming transmittance measurements to absorbances on the computations is examined.

The software package is also used for descriptive statistics, significance tests, and ANOVA. Many additional features, e.g., normality and outlier checks, residual analysis, and simulated data generation are demonstrated.

The role of applied statistics and the microcomputer in chemometrics and the analytical laboratory is discussed.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>STATEMENT OF RESPONSIBILITY</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
</tbody>
</table>

Chapter 1  INTRODUCTION
1.1 Background .................................. 1
1.2 Applied Statistics ........................... 2
1.3 Analytical Calibration Curves ............... 5
1.4 Comparison of Analytical Methods .......... 5
1.5 Quality Control ................................ 6
1.6 Optimisation ................................... 8
1.7 Microcomputers ................................ 10
1.8 Microcomputer Software ..................... 12
1.9 Laboratory Facilities ....................... 16
1.10 Summary ...................................... 18

Chapter 2  APPLIED STATISTICS
AND ANALYTICAL CHEMISTRY ......................... 20
2.1 Introduction .................................. 20
2.1.1 Literature Search ......................... 22
2.1.2 The Computer Program .................... 23
2.2 The Distribution of Analytical Measurements | 27 |
2.2.1 Outliers .................................. 27
2.2.2 Normal Distribution ...................... 28
2.2.3 Histograms ................................ 29
2.2.4 Availability of Routines for Outliers, Normality, and Histograms 29
2.2.5 Examples ................................. 29
2.3 Data Transformations .......................... 32
2.4 Monte Carlo Simulated Analytical Data ... 33
2.4.1 Requirements of the Simulation Model .... 34
2.4.2 The Algorithm Used ....................... 34
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>Significance Tests</td>
<td>37</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Theory</td>
<td>38</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Uses of the General &quot;t&quot; Test</td>
<td>41</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Uses of the Paired-Data &quot;t&quot; Test</td>
<td>45</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Uses of the &quot;t&quot; or &quot;U&quot; Test for the Comparison of an Experimental</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Mean with a Known Mean</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Analysis of Variance</td>
<td>52</td>
</tr>
<tr>
<td>2.6.1</td>
<td>One-Factor Analysis of Variance</td>
<td>52</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Comparison of One-, Two-, and Two-Factor with Replication, Plus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cross-Classification vs. Hierarchic Factors</td>
<td>56</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Assumptions and Limitations of ANOVA</td>
<td>58</td>
</tr>
<tr>
<td>2.6.4</td>
<td>Examples</td>
<td>58</td>
</tr>
<tr>
<td>2.7</td>
<td>Linear Regression Analysis</td>
<td>66</td>
</tr>
<tr>
<td>2.8</td>
<td>Summary</td>
<td>67</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>ANALYTICAL CALIBRATION CURVES</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>BY LEAST SQUARES ANALYSIS</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>68</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Method of Least Squares</td>
<td>69</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Weighting Factors in Regression Analysis</td>
<td>69</td>
</tr>
<tr>
<td>3.2</td>
<td>Theory</td>
<td>70</td>
</tr>
<tr>
<td>3.2.1</td>
<td>The Unweighted Calibration Curve</td>
<td>71</td>
</tr>
<tr>
<td>3.2.2</td>
<td>The Weighted Calibration Curve</td>
<td>84</td>
</tr>
<tr>
<td>3.3</td>
<td>The Microcomputer Program</td>
<td>89</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Statistical Formulae and Programming</td>
<td>91</td>
</tr>
<tr>
<td>3.3.2</td>
<td>The Formulae Used in the Program</td>
<td>91</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Output of the Program</td>
<td>95</td>
</tr>
<tr>
<td>3.4</td>
<td>The standard Addition Technique</td>
<td>102</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Atomic Absorption Spectrometry</td>
<td>102</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Differential Pulse Anodic Stripping Voltammetry</td>
<td>109</td>
</tr>
</tbody>
</table>
### Chapter 3

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>The General Comparative Analysis</td>
<td>111</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Gas Chromatography</td>
<td>111</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Atomic Absorption Spectrometry</td>
<td>115</td>
</tr>
<tr>
<td>3.6</td>
<td>Monte-Carlo Simulated Experiments</td>
<td>121</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Experimental</td>
<td>121</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Results and Discussion</td>
<td>121</td>
</tr>
<tr>
<td>3.7</td>
<td>Nonlinear Calibration Curves</td>
<td>126</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Approaches Described in the Literature</td>
<td>126</td>
</tr>
<tr>
<td>3.7.2</td>
<td>A Simplified Approach to Nonlinear Calibration Curves</td>
<td>128</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Comparison of Calculation Methods</td>
<td>132</td>
</tr>
<tr>
<td>3.8</td>
<td>Minimum Reportable Concentration</td>
<td>134</td>
</tr>
<tr>
<td>3.9</td>
<td>Practical Considerations when conducting a Statistical Analysis of Calibration Curves</td>
<td>135</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Methods of Data Analysis</td>
<td>135</td>
</tr>
<tr>
<td>3.9.2</td>
<td>Limitations of the Least Squares Method</td>
<td>136</td>
</tr>
<tr>
<td>3.9.3</td>
<td>Overview of the Calibration Procedure</td>
<td>140</td>
</tr>
<tr>
<td>3.10</td>
<td>Summary</td>
<td>142</td>
</tr>
</tbody>
</table>

### Chapter 4

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>144</td>
</tr>
<tr>
<td>4.2</td>
<td>Definition of the Problem</td>
<td>145</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Technical and Economic Merit</td>
<td>146</td>
</tr>
<tr>
<td>4.3</td>
<td>Survey of Literature Comparisons</td>
<td>149</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Number of Methods Compared</td>
<td>149</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Advantages Claimed for the New Method</td>
<td>149</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Method Validation</td>
<td>151</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Statistical Analysis Used</td>
<td>152</td>
</tr>
<tr>
<td>4.3.5</td>
<td>General Conclusions</td>
<td>153</td>
</tr>
<tr>
<td>4.4</td>
<td>Comparison of Accuracy using &quot;t&quot; tests and ANOVA</td>
<td>153</td>
</tr>
<tr>
<td>4.4.1</td>
<td>&quot;t&quot; Tests</td>
<td>154</td>
</tr>
<tr>
<td>4.4.2</td>
<td>ANOVA</td>
<td>155</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Conclusions</td>
<td>155</td>
</tr>
</tbody>
</table>
4.5 The Variation of Error Standard Deviation with Concentration .......................... 155
4.5.1 The Standard Deviation Function ... 156
4.5.2 Determining the Error Function .... 159
4.5.3 Example: "Replicate Analyses" versus "Regression of Response on Concentration" Approaches to Calculating the Precision of an Analytical Method .................. 164

4.6 Validation of Method Accuracy with Linear Regression Analysis ............... 167
4.6.1 Requirements of the Technique ...... 167
4.6.2 Usefulness of the Technique ........ 168
4.6.3 Maximum Likelihood Regression ...... 168
4.6.4 The Conditions under which SLR, WLR, and MLR can be used for the Comparative Calibration of Analytical Methods .................. 170

4.7 Quantifying the Technical and Economic Merit of Analytical Methods ...... 170
4.7.1 Mandel’s Relative Sensitivity Ratio .................. 172
4.7.2 Steps in the Application of Mandel’s Relative Sensitivity Ratio ......... 176
4.7.3 A Textbook Example of Technical Merit .................. 177

4.8 Experimental: Constant Error Standard Deviation Model .................. 183
4.8.1 Results and Discussion .................. 183

4.9 Experimental: Constant Relative Error Standard Deviation Model .................. 190
4.9.1 Results and Discussion .................. 193

4.10 Experimental: Mixed Model –
Constant Error SD and RSD .................. 200
4.10.1 Results and Discussion .................. 200

4.11 Conclusions .................. 206
4.12 Summary .................. 209
<table>
<thead>
<tr>
<th>Chapter 5</th>
<th>DISCUSSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Microcomputers in the Laboratory .................. 210</td>
</tr>
<tr>
<td>5.2</td>
<td>Applied Statistics ...................................... 212</td>
</tr>
<tr>
<td>5.3</td>
<td>Weighted Linear Regression ............................ 214</td>
</tr>
<tr>
<td>5.4</td>
<td>Monte Carlo Simulation ................................... 215</td>
</tr>
<tr>
<td>5.5</td>
<td>Absolute Calibration ................................. 215</td>
</tr>
<tr>
<td>5.6</td>
<td>Determining the Analytical Error Function .......... 217</td>
</tr>
<tr>
<td>5.7</td>
<td>Comparative Calibration ............................... 215</td>
</tr>
<tr>
<td>5.8</td>
<td>Method Comparisons based on Accuracy .............. 219</td>
</tr>
<tr>
<td>5.9</td>
<td>Comparative Calibration Method Comparison based on Precision .......... 220</td>
</tr>
<tr>
<td>5.10</td>
<td>A &quot;Vertical Program&quot; for the Analyst .............. 221</td>
</tr>
<tr>
<td>5.11</td>
<td>Summary .................................................. 223</td>
</tr>
<tr>
<td></td>
<td>Conclusions and Recommendations .................... 224</td>
</tr>
</tbody>
</table>

REFERENCES ................................................. 225

APPENDIXES

A Applied Statistics Program Manual .................. 237
B Bibliography from "Technometrics" .................. 254
C Special Symbols Used ................................. 258
D Manual Calibration Experiment ..................... 259
E A Use for Predicted Response Tolerance Limits ....... 260
F The Complete Applied Statistics Program .......... 262
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Nine Sets of Thirty-Six Transmittance Measurements</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>A Basic Program to Calculate a Series of Normally Distributed Measurements</td>
<td>35</td>
</tr>
<tr>
<td>2.3</td>
<td>Statistical Parameters Calculated for Sets of Simulated Measurements</td>
<td>36</td>
</tr>
<tr>
<td>2.4</td>
<td>Ordered Concentrations and Corresponding Differences</td>
<td>47</td>
</tr>
<tr>
<td>2.5</td>
<td>Data Layout for One-Factor ANOVA</td>
<td>55</td>
</tr>
<tr>
<td>2.6</td>
<td>One-Factor Analysis of Variance Table</td>
<td>57</td>
</tr>
<tr>
<td>2.7</td>
<td>A Comparison of Methods of ANOVA</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td>Basic Formulae Used in the Program</td>
<td>93</td>
</tr>
<tr>
<td>3.2</td>
<td>&quot;Confidence&quot; Limit Formulae for Computed Responses and Concentrations</td>
<td>94</td>
</tr>
<tr>
<td>3.3</td>
<td>Computed Transmittance Measurements</td>
<td>105</td>
</tr>
<tr>
<td>3.4</td>
<td>A Comparison of Computed and Published Results for an Analysis Using the Standard Addition Technique and Atomic Absorption Spectrometry</td>
<td>108</td>
</tr>
<tr>
<td>3.5</td>
<td>A Comparison of Computed and Published Results for an Analysis Using the Standard Addition Technique and Differential Pulse Anodic Stripping Voltammetry</td>
<td>109</td>
</tr>
<tr>
<td>3.6</td>
<td>A Comparison of Computed and Published Results for a Quantitative Analysis by Gas Chromatography</td>
<td>113</td>
</tr>
<tr>
<td>3.7</td>
<td>Descriptive Statistics for Replicate Response Measurements at Varying Levels of Analyte Concentration</td>
<td>118</td>
</tr>
<tr>
<td>3.8</td>
<td>Comparison of Determined and Actual Concentrations of Copper in Thorn Smith Copper Ore Samples</td>
<td>120</td>
</tr>
<tr>
<td>3.9</td>
<td>A Comparison of Expected and Computed Calibration Curve Parameters</td>
<td>122</td>
</tr>
</tbody>
</table>
Table 3.10 NUMBER OF SIMULTANEOUS DISCRIMINATION LIMITS
NOT BRACKETING
THE EXPECTED CONCENTRATION ...................... 124
3.11 NUMBER OF SIMULTANEOUS DISCRIMINATION LIMITS
NOT BRACKETING
THE EXPECTED CONCENTRATION ...................... 125
3.12 AN OVERVIEW OF THE CALIBRATION PROCEDURE ..... 141

4.1 SURVEY OF METHOD COMPARISONS IN THE LITERATURE. 150
4.2 RESULTS OF A % RSD DETERMINATION USING THE
"REGRESSION OF RESPONSE ON CONCENTRATION"
APPROACH ........................................... 166
4.3 CONDITIONS UNDER WHICH SIMPLE, WEIGHTED, AND
MAXIMUM LIKELIHOOD LINEAR REGRESSION CAN BE
USED FOR THE COMPARISON OF ACCURACY OF
ANALYTICAL METHODS ............................. 171
4.4 DATA FROM AN EXAMPLE USED TO ILLUSTRATE THE
CALCULATION OF TECHNICAL MERIT .............. 179
4.5 GROUPED MEASUREMENTS USED TO CALCULATE
TECHNICAL MERIT ............................... 181
4.6 SIMULATED REGRESSION DATA WITH CONSTANT ERROR
STANDARD DEVIATION ............................. 186
4.7 ESTIMATES OF THE REGRESSION PARAMETERS AND
EXPERIMENTAL ERROR STANDARD DEVIATIONS FOR
METHODS M4-1 AND M4-2 ........................... 187
4.8 DUPLICATE INTENSITY MEASUREMENTS FOR A
SET OF STANDARDS ............................... 192
4.9 THE WEIGHTED LINEAR REGRESSION OF RESULTS
FROM ANALYTICAL METHODS HAVING EXPERIMENTAL
ERRORS WITH
CONSTANT RELATIVE STANDARD DEVIATIONS ...... 194
4.10 THE SIMPLE LINEAR REGRESSION OF RESULTS
FROM ANALYTICAL METHODS HAVING EXPERIMENTAL
ERRORS WITH
CONSTANT RELATIVE STANDARD DEVIATIONS ...... 195
4.11 ESTIMATES OF THE EXPERIMENTAL ERROR'S RELATIVE
STANDARD DEVIATION FROM
DUPLICATE ANALYSES ............................. 198
Table
4.12 EXPECTED AND EXPERIMENTAL RELATIVE SENSITIVITY VALUES FOR METHODS HAVING A CONSTANT RSD ... 199
4.13 DUPLICATE TRANSMITTANCE MEASUREMENTS FOR A SET OF STANDARDS ................................. 202
4.14 ESTIMATES OF THE EXPERIMENTAL ERROR'S STANDARD DEVIATION FROM DUPLICATE ANALYSES .......... 203
4.15 DETERMINING MANDEL'S RELATIVE SENSITIVITY RATIO WHEN THE EXPERIMENTAL ERROR OF ONE SET OF RESPONSES HAS A CONSTANT STANDARD DEVIATION AND THE OTHER HAS A CONSTANT RELATIVE STANDARD DEVIATION .................. 204
4.16 A COMPARISON OF EXPECTED AND EXPERIMENTAL VALUES FOR MANDEL'S RELATIVE SENSITIVITY RATIO .......................... 205
4.17 CONDITIONS UNDER WHICH THE FOUR APPROACHES TO DETERMINING THE STANDARD DEVIATION OR RELATIVE STANDARD DEVIATION OF ANALYTICAL ERROR COULD BE USED ................ 208
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Basic statistical terms and definitions</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Menu for complete applied statistics program</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>Histogram, descriptive statistics, and an ordered set of thirty of thirty-six transmittance measurements</td>
<td>31</td>
</tr>
<tr>
<td>2.4</td>
<td>Results of a one-sided general &quot;t&quot; test</td>
<td>42</td>
</tr>
<tr>
<td>2.5</td>
<td>Results of a general &quot;t&quot; test where the two variance estimates are significantly different</td>
<td>43</td>
</tr>
<tr>
<td>2.6</td>
<td>Results of a general &quot;t&quot; test used to compare the % Recovery of two analytical procedures at many concentration levels</td>
<td>44</td>
</tr>
<tr>
<td>2.7</td>
<td>Results of two, paired-data &quot;t&quot; tests</td>
<td>46</td>
</tr>
<tr>
<td>2.8</td>
<td>Results of a paired-data &quot;t&quot; test where the data had originally been analyzed using the linear regression technique</td>
<td>48</td>
</tr>
<tr>
<td>2.9</td>
<td>Results of a paired-data &quot;t&quot; test where the data is first normalized</td>
<td>49</td>
</tr>
<tr>
<td>2.10</td>
<td>Results of a &quot;t&quot; test where an experimentally determined mean is compared to a known mean, where there is no known standard deviation</td>
<td>51</td>
</tr>
<tr>
<td>2.11</td>
<td>A two-factor, hierarchic, analysis of variance</td>
<td>60</td>
</tr>
<tr>
<td>2.12</td>
<td>A two-factor, hierarchic (model II), ANOVA with replication</td>
<td>62</td>
</tr>
<tr>
<td>2.13</td>
<td>A two-factor, cross-classification, ANOVA with replication</td>
<td>64</td>
</tr>
<tr>
<td>2.14</td>
<td>A two-factor, cross-classification, ANOVA with replication</td>
<td>65</td>
</tr>
<tr>
<td>2.15</td>
<td>Ordered data set of merged ANOVA measurements, their histogram, and descriptive statistics</td>
<td>66</td>
</tr>
<tr>
<td>3.1</td>
<td>The calibration curve with an approximate confidence band</td>
<td>73</td>
</tr>
<tr>
<td>3.2</td>
<td>Computed concentration and discrimination intervals for an unweighted calibration line</td>
<td>81</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.3</td>
<td>A weighted calibration curve and its confidence band</td>
<td>86</td>
</tr>
<tr>
<td>3.4</td>
<td>Conditions under which discrimination limits can and cannot be calculated</td>
<td>88</td>
</tr>
<tr>
<td>3.5</td>
<td>BASIC microcomputer program segment for weighted/unweighted calibration curves</td>
<td>90</td>
</tr>
<tr>
<td>3.6</td>
<td>Basic information produced by the first screen of the linear regression routine</td>
<td>97</td>
</tr>
<tr>
<td>3.7</td>
<td>The computed responses, the 95% confidence band of the computed calibration line, the weighted residuals, and the weights used in calculating the calibration line</td>
<td>97</td>
</tr>
<tr>
<td>3.8</td>
<td>Computed responses and their &quot;confidence&quot; limits for selected concentration levels</td>
<td>99</td>
</tr>
<tr>
<td>3.9</td>
<td>Computed concentrations and their &quot;confidence&quot; limits for selected responses</td>
<td>99</td>
</tr>
<tr>
<td>3.10</td>
<td>A weighted calibration curve with &quot;confidence&quot; bands around the response, the curve, and the computed concentration</td>
<td>100</td>
</tr>
<tr>
<td>3.11</td>
<td>The histogram of the weighted residuals about the calibration curve</td>
<td>101</td>
</tr>
<tr>
<td>3.12</td>
<td>Some computer results for a determination using the standard addition technique and atomic absorption spectrometry (unweighted line)</td>
<td>103</td>
</tr>
<tr>
<td>3.13</td>
<td>Some computer results for a determination using the standard addition technique and atomic absorption spectrometry (weighted line)</td>
<td>104</td>
</tr>
<tr>
<td>3.14</td>
<td>Some computer results for a determination using the standard addition technique and atomic absorption spectrometry (weights equal to transmittance^2)</td>
<td>107</td>
</tr>
<tr>
<td>3.15</td>
<td>Some computer results for a determination using the standard addition technique and differential pulse anodic stripping voltammetry</td>
<td>110</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>3.16</td>
<td>Some computer results for a quantitative analysis by gas chromatography</td>
<td>114</td>
</tr>
<tr>
<td>3.17</td>
<td>The variation of response standard deviation with response level</td>
<td>119</td>
</tr>
<tr>
<td>3.18</td>
<td>Computer results using a second order calibration curve and a weight of 1/response^2</td>
<td>130</td>
</tr>
<tr>
<td>3.19</td>
<td>A quadratic least squares BASIC program</td>
<td>131</td>
</tr>
<tr>
<td>3.20</td>
<td>Overview of the calibration procedure</td>
<td>141</td>
</tr>
<tr>
<td>4.1</td>
<td>A graphical development of Mandel's relative sensitivity (RS) ratio for Method M with respect to Method N</td>
<td>147</td>
</tr>
<tr>
<td>4.2</td>
<td>Variation of the total standard deviation and relative standard deviation with analyte concentration</td>
<td>158</td>
</tr>
<tr>
<td>4.3</td>
<td>Results of a WLR to obtain the % RSD of an analytical method</td>
<td>166</td>
</tr>
<tr>
<td>4.4</td>
<td>Bias between methods X and Y</td>
<td>169</td>
</tr>
<tr>
<td>4.5</td>
<td>Variation of the error standard deviation with stress and strain measurements</td>
<td>180</td>
</tr>
<tr>
<td>4.6</td>
<td>Variation of transformed standard deviation with log(average measurements)</td>
<td>182</td>
</tr>
<tr>
<td>4.7</td>
<td>Standard deviations estimated from triplicate responses regressed on the average response</td>
<td>184</td>
</tr>
<tr>
<td>4.8</td>
<td>Computer results for the regression of the average responses from Method M4-1 on the average responses from Method M4-2</td>
<td>188</td>
</tr>
<tr>
<td>4.9</td>
<td>Computer results for the regression of concentrations from Method M4-1 on those of Method M4-2</td>
<td>191</td>
</tr>
<tr>
<td>4.10</td>
<td>A histogram of the residuals from the regression of Method M4-1 concentrations on those of Method M4-2</td>
<td>191</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION

1.1 Background

This thesis was originally going to deal with the subject of Auger Electron Spectrometry. In 1978 I completed a Master's thesis at the Loughborough University of Technology dealing with an evaluation of the analytical potential of Auger Electron Spectrometry. At that time I received permission to register as an external candidate for a Ph.D. and my area of research was to be a continuation of the same subject.

Although I pursued the approved topic of research actively for two years it was a subject with which I was becoming more uncomfortable. Local library facilities were inadequate and there was no possibility of conducting experimental work. This latter problem was the more serious, as I consider myself to be more of a practical than a theoretical chemist. Thus, I began looking for another area of research which could be done with the available facilities and in which I had a genuine interest, or using a expression put forth by Lansbury (1), my original thesis topic and I were divorced.

While at Loughborough I had taken advantage of the opportunity to pursue one of my old interests, namely statistics as it is applied to analytical data. In addition to taking a short course in statistics I reviewed analytical journals for statistical articles. I also did some programming of my Texas Instruments TI 59 calculator for handling common statistical routines.

My second research interest is naturally programming. An IBM 360 computer (1968) was my first introduction to computing. This was followed by a Wang 700 programmable calculator (1970) and then the Digital PDP 11/70 minicomputer (1979).

When I began my Ph.D. research in Canada one of my first projects was to program a literature information retrieval program for the Digital computer. This project took place concurrently with my Auger literature search; it took about a year to produce the system.
By 1979 microcomputers were becoming very popular and by early 1980 Northern College had purchased three Heath microcomputers. In December of 1980 I purchased my own Zenith (Heath) Z89 microcomputer. At that time I stopped work on my Auger Electron Spectrometry research and turned completely to the work contained in this thesis. During the previous two years I had also been working with programming involving statistics using the Digital computer and the Heath microcomputer. I was, therefore, not starting a completely new line of research but was instead now working at it full time.

The remainder of this introduction relates specifically to the content of the thesis, i.e., applied statistics, microcomputers, microcomputer software, and the application of these in the analytical laboratory by the practising analyst. Topics which later have a chapter devoted to them are only briefly introduced.

Two additional topics, quality control and simplex optimisation, are also discussed. Originally it was thought that they would also form a major part of this research, as each is related to the thrust of the thesis. However, the amount of research conducted on analytical calibration curves and the comparison of methods was greater than originally expected; thus the scope of the thesis was reduced.

It is hoped that this thesis will in some small way help the typical analytical chemist in overcoming computerphobia (2).

1.2 Applied Statistics

My interest in statistics and its application to analytical chemistry began in 1965. At that time I was working at the Eldorado Mining and Research Laboratories with a Philips X-ray spectrometer. This technique involves the use of statistics, however, the real cause for my interest was generated by the work I did with, and the enthusiasm shown by J.M. Jardine, the laboratory's director. My introduction to the subject was expanded upon by the text "An Introduction to Scientific Research" by E. Bright
The use of statistics for the analysis of analytical data is certainly not new or novel, nor is the use of computers for doing the many calculations involved in these statistical analyses. However, the use of a computer by an average analyst, and the actual personal statistical analysis of his data, is a recent development. As well, it is an exciting new area of endeavour for the analyst.

The microcomputer, with its low cost and ever increasing computing power, has opened up a new area of research. Namely, adapting it to a small, low budget analytical laboratory.

One of the many avenues of this research is the writing of software for the analytical laboratory. An example is statistical software which will allow an analytical chemist with a modest statistical background and who has no, or at best a limited access to a statistician, to analyse his own data quickly and easily. That is, a user friendly APPLIED STATISTICS program for analytical laboratory.

Such a program has been written. The user’s manual for the original program is included as Appendix A. This manual will be rewritten and expanded at a future date to include all the additional features which were developed during the course of the research for this thesis. The use of the program with real and simulated analytical data is illustrated throughout the thesis.

In addition the thesis surveys the literature related to the use of statistics in reporting analytical data. The main chemistry journals surveyed were the Analyst, Analytical Chemistry, Analytica Chimica Acta, Talanta, Technometrics, and the Journal of Chemical Education. These journals were used to determine:
- the use of statistics in reporting analytical data
- the statistical techniques used most often
- examples of the proper use of statistical techniques
- data which could be subjected to statistical analysis
- suggested uses of statistics in chemistry
- developments in statistical methodology.
Technometrics was a journal of special importance in this research. It is a journal "for chemists, physicists, engineers and technologists interested in the application of statistics in their fields, and for statisticians with similar interests" (4). Many papers from this journal were used as references whereas others were used for general background information. In recognition of its special role, a bibliography of papers related to the topic of this thesis are included as Appendix B.


These latter reviews are a good source of introductory texts, as well as, texts on individual statistical topics. My personal recommendation of an introductory text is the book by Brookes, Betteley, and Loxton (17).

A final point is the use of the term "Chemometrics", the title of the 1980 and 1982 Analytical Chemistry review article. It is defined (16) as "the chemical discipline that uses mathematical and statistical methods, (a) to design or select optimal measurement procedures and experiments; and (b) to provide maximum chemical information by analysing chemical data. In the field of Analytical Chemistry, Chemometrics is the chemical discipline that uses mathematical and statistical methods for the obtention in the optimal way of relevant information on material systems".

This field of study is becoming increasingly important to analysts; Vandeginste (18) has discussed teaching chemometrics as a part of analytical chemistry; Belchamber, et al. (19) have discussed the application of computers in chemometrics; and Howery and Hirsch (20) have discussed the development of chemometrics and its place in the chemistry curriculum. The research conducted for this thesis began before chemometrics gained its current popularity.
1.3 Analytical Calibration Curves

The construction of calibration curves is a traditional part of analytical chemistry. Kurtz (21), in an introduction to a paper on the use of regression and statistical methods to establish calibration graphs in chromatography, has briefly described the construction of the graph. He has also mentioned the method which will be used in this research.

The thesis addresses a statement he makes about the various statistical models which have been proposed in the literature, namely "these techniques are seldom used in research and industry". It will be shown (chapter 3) that weighted linear regression (one of the recommended techniques) can be used on a routine basis in any laboratory where a microcomputer is available. It will also be shown (chapter 3 and 4) that the problem of having to know the standard deviation of the analytical error, at each level of concentration, can also be determined on a routine basis for a reasonably well behaved analytical system.

It is the intent therefore to show that these "sophisticated" techniques are of more than theoretical interest. This is done convincingly using real and synthetic data, and the computer software developed in this research. The whole of chapter 3 is devoted to this topic.

1.4 Comparison of Analytical Methods

The literature shows (sec. 4.3) that the comparison of analytical methods using statistical techniques is also a topic which apparently is mostly of academic interest to analysts. The two approaches to be explored in this research are those proposed by Mandel (22) in 1964, and by Thompson (23) in 1982. Both use statistical techniques which one would probably be correct in stating are also "seldom used in research and industry".

The two approaches are similar in that they both use regression analysis. Thompson, however, compares analytical methods on the basis of the results produced by each method, whereas, Mandel uses only the "raw" analytical data. Both approaches allow the calculation of technical and economic figures of merit.
The comparison of analytical methods, using statistical techniques, should be of more than theoretical interest. Again with the use of both real and synthetic data, and the computer software developed in this research, it will be shown that such comparisons can be made routinely in any laboratory where a microcomputer is available. The whole of chapter 4 is devoted to this topic.

1.5 Quality Control

It has long been my personal opinion that analytical chemists are in general too optimistic about the accuracy of their analyses. This has resulted in another opinion as to why analysts have tended not to use statistics more extensively; the use of statistics produces a (mathematically) best answer, but also states a range for that answer.

The chemist, and perhaps more importantly the client, prefers a single answer as this indicates that the "true result" has been found. Perhaps also the chemist has never been able to completely eradicate his university experience that the professor knew the "correct" answer. Plus, the use of statistics may give a wider range for his results then he is willing to accept. The client might possibly also think the chemist didn't know what he was doing if the mean and "too wide" a concentration interval were reported for the analysis. This is assuming the chemist had in fact done the analysis in such a manner that he could determine a concentration interval for the result.

The above is undoubtedly an exaggeration of the situation as it exists. However, the analyst does require a mechanism (24-26) whereby he can unbiassedly judge the accuracy of his results. External quality assessment (interlaboratory comparisons) are a method of judging the variation in results between laboratories. They can also do much towards dispelling an overly optimistic view about the accuracy of analytical results (27).

However, before a laboratory participates in such a test it should have a mechanism for testing it's own results. Part of that mechanism should include a method
whereby the individual analyst can also check his results.

This research unfortunately does not include the development of the software necessary for the calculations and charting required for in-house quality control. Such software would not be particularly difficult to develop; in fact much of the programming is already part of the existing program.

The expansion of the program to include quality control is part of the "vertical" software system which is discussed in chapter 5. A vertical system is one which is specific to a particular group, e.g., analysts. Ideally such a system would be capable of doing every computation required in an analytical laboratory. Some of the background required for a quality control program is discussed in the remainder of this section.

Quality Control is a term used originally in industry to describe a method of testing a product, during and after its manufacture, to determine if it conforms to previously established specifications. The associated "Control Chart" was invented in 1924 by Dr. W.A. Shewhart. It is described in his book "The Economic Control of Quality of Manufactured Product" (28). Most chemists are probably familiar with the "Average" and "Range" control charts used for process control. Juran's handbook (29) on quality control has a chapter devoted to this topic as does Brookes (18). However, a somewhat different type of quality control is required for the analytical laboratory.

In the laboratory the products are the analytical results. The specifications of the "products" are the allowable limits of error. The laboratory must control the accuracy and the precision of the analyses within the prescribed allowable limits of the method used.

Clinical laboratories (24) have done much in applying quality control techniques to their everyday work. Tonks (30), in his book "Quality Control in Clinical Laboratories" prepared for Warner-Chilcott General Diagnostics (2200 Eglinton Ave. East, Scarborough, Ontario, M1L 2N3), has discussed quality control and has outlined a plan for the establishment of quality control in the
clinical laboratory (as has Buttner et al. (31)). Tonk's book also contains an extensive bibliography of his and other papers on the subject.

Water analysis (25) has more recently brought quality assurance programs to the attention of the analyst. Many countries, including Canada, have an active water quality agency, e.g., the Inland Waters Directorate, Canada Centre for Inland Waters. This particular agency regularly publishes the results of interlaboratory quality control studies.

Taylor (26) has presented a general discussion of "Quality Assurance of Chemical Measurements" which should be of interest to any analyst thinking of this topic in regards to his own work.

Kateman and Pijpers (32) in their book "Quality Control in Analytical Chemistry", Tonks (30), and Wilson (25) have all very strongly made the point that quality control includes much more than the statistical treatment of data. Sampling and method selection for example are even more important. These other aspects of quality control are, however, outside the scope of this thesis.

1.6 Optimisation

Simplex optimisation is also a chemometric topic which could be included in a "vertical" system of software for the analyst. It is an interesting topic in that it departs from traditional statistics to offer a technique of optimising methods, procedures, strategies, and laboratories (33). The computer of course plays a large role in the application of this technique. Berridge (34) has reported using an interactive Simplex program (written in BASIC) for several years. He has also reported on the unattended optimisation in HPLC method development.

Simplex optimisation has been used in analytical chemistry for many years. The original technique was derived by Box and Wilson (35) in 1951, then developed by Spendley et al. (36) in 1962, and later applied to analytical chemistry by Long (37) in 1968. Deming and Morgan (38,39) have helped to clarify the technique for the analyst. Three
papers (40-42) in the Journal of Chemical Education indicate that the technique can be useful in undergraduate as well as graduate education.

Simplex optimisation can perhaps best be described (39) by contrasting it with the "single-factor-at-a-time" strategy. In that strategy the main factors are first identified, and then one is varied while the others are held constant at a low level; this strategy is not particularly efficient, nor is there a guarantee of locating the conditions for a maximum response. Simplex optimisation, on the other hand, allows the main factors to be altered simultaneously.

In a simple application of this technique three sets of conditions are used to obtain three responses. The poorest response is rejected and the other two are used to select another, perhaps better set of conditions. Again the conditions yielding the lowest response is rejected, etc. In this manner the strategy can "home in" on the conditions resulting in the maximum response. The original simplex has been modified; there are many examples of its use in the analytical literature (e.g., 43-49).

A program which could be used for demonstrating the technique through simulation, preferably using graphics, would be a useful addition to the scientific software base. Such a program should also be useful for actual applications of the simplex technique.

Leggett (42) has reported on the usefulness of instrument simulation using a computer to generate numerical responses for sets of "instrument" conditions. The responses are generated by students for use in an undergraduate laboratory course, where the student is required to determine the best response for the system. Leggett does not, however, have the computer perform any simplex calculations related to solving the problem for the student.
1.7 Microcomputers

The microprocessor, the integral part of the microcomputer, is having an almost "science fictional" impact upon society. The impact upon chemists range from the familiar personal items such as digital watches, to professional items such as microprocessor controlled chemical instrumentation.

The rapid development in microprocessor and microcomputer technology is probably influenced very little by their application in chemistry. These devices are produced for the mass consumer market; it is for this reason that they are relatively inexpensive. Chemists may find that they are required to select items produced for a general market, but since the market is so large and varied, this should not be too great a restriction for either the chemist or the chemical instrumentation manufacturer.

This does not mean that the scientific community is an unimportant part of the microcomputer scene; surveys conducted by computer "magazines" have shown a very large proportion of their readers are involved in science. In addition an increasing number of hardware and software products for the scientific user come onto the market each year.

Two papers related to the impact of microprocessors on instrumentation have appeared recently in the analytical literature. Betteridge and Goad (51) have covered the development of microcomputers and the application of microprocessors to analytical instrumentation. Their paper also gives a list of the microcomputers (with brief descriptions of the systems) available at that time and a glossary of associated terms. Microcomputer development has been rapid, therefore information on the systems available is often outdated within weeks of publication.

A paper by Enke et al. (52) deals with the microcircuit, their discussion of the integrated circuit (IC) is very well presented. The use of microcomputers as laboratory instruments has also been described in the popular journals (53,54). These are all important topics, however, they will not be addressed further in this thesis.
The most common microcomputers in use today are based on 8-bit microprocessors. The newer generation of microcomputers are based on 16-bit microprocessors. Although not the first such computer, the IBM-PC (55) is never-the-less a very important new product because it indicates the interest in this market by a computer giant.

The use of the 16-bit microprocessor allows an increase in the amount of read-only-memory from 64 kilobytes to 1 megabyte (16MB for the Motorola 68000), an increase in computing speed, and possibly an increase in mathematical accuracy.

The Journal of Chemical Education began a series (56) on computers in 1979. The fourth paper (57) in this "Computer Series" is an chemist's introduction to microprocessors and microcomputers. This paper is recommended to any chemist interested in a straightforward introduction to microcomputers. The authors also deal with the subject of selecting a computer system as do several papers (e.g., 58,59) in the popular journals. A selected bibliography (60) of computer programs which have appeared in the Journal of Chemical Education (1967-1974) is also available.

It has been estimated that in 1965 there were 30,000 computers in the world. Some monthly shipping rates of microcomputers (reported in October of 1983) were: Apple IIe --40,000, IBM-PC/XT--20,000+, PC-compatables--5,000 to 10,000, and CP/M systems--40,000 units. This little bit of trivia adds some dimension to the impact the microcomputer will continue to have on the analyst.

One additional item which should be mentioned before discussing software is that with the acquisition of a microcomputer, the chemist becomes the director of a small computing centre. The chemist will soon discover that a stand-alone system is more time consuming to use than a time-shared system managed by a professional computing centre director. There is, however, the advantage of having the system under the direct control of the laboratory staff.
1.8 Microcomputer Software

Although the computer hardware is important it is only part of the system. The software, i.e., the programs to be run on the computer, are probably even more important. The computer manufacturer usually provides the necessary software to get the system operating. Such software would include an operating system, which for any serious application would be a Disk Operating System. Other software such as an editor, assembler, and debugger would probably also be included with the operating system.

At least one high level language (61) would be required for user programming, and to operate some commercial programs. BASIC (62) is still the most commonly used language for microcomputers. A two-part paper edited by Dessy (63) has discussed computer languages for the laboratory.

One immediate problem the user has is the incompatibility of different computers and different versions of the same language. For example, there are many versions of BASIC. Although a program has been written in BASIC it is unlikely that it will run, without coding changes, on another manufacturer's computer. These coding changes could be substantial. There is, however, some uniformity in the versions used by different computers in that many were written by Microsoft (62), a software vendor.

A common operating system for 8-bit computers is CPM (64,65), Control Program/Microcomputers (Digital Research, Pacific Grove, California). There are many other systems available; Dessy (66) has edited a paper on the various operating systems available for the laboratory.

One might expect that a program written in Microsoft BASIC and using CPM could be run on any computer capable of using CPM and Microsoft BASIC. This is usually not the case unless the programmer had taken pains to ensure that no manufacturer specific hardware features were used in the program. Although such programs can be written, they are by design not capable of taking advantage of all the unique features of the user's computer. A user must therefore be quite selective (67) in purchasing commercial software to
ensure compatibility with their computer system, as well as, suitability for the job at hand.

Many computer users will write some of their own software. Such software can often be written relatively quickly and can be used by the programmer. However, if a program is to be useful to all laboratory staff it must be thoroughly debugged, must contain the necessary error traps, and must be well documented. Such programs take a much longer time to develop and are often a major project. Anyone writing their own programs should read the popular computing journals for hints on effective programming and documentation (68-72).

A computer owner should expect to purchase some additional software. There is a growing number of good commercial programs available. Each popular computing system is usually supported by a number of commercial software houses. The major problem for the chemist, or the scientist in general, is the lack of a wide variety of scientific software. However, some of the commercial software is useful for the laboratory.

A semiannual review (73) of software packages lists approximately 300 products under 36 different categories. These categories are listed here to give an indication of the range of software available:

- Data base management
- Data dictionary
- Query languages/report writers
- Program development/aids
- Librarians
- Project control
- Information retrieval
- Operating systems
- Utilities
- Performance measurements
- DASD/tape management
- Systems development
- Systems management
- Accounts receivable
- Accounts payable
- Payroll/personnel
- Forecasting/estimating
- Mathematical/statistical
- Word/text processing
- Graphics
- Engineering/architectural
- Manufacturing control
- Sales/marketing/distribution
- Wholesale trade/distribution
- Retail trade/distribution
- Financial industry/banking
Most laboratories would probably be interested in many of the software packages currently available.

Two categories of software packages which would be of specific use in probably every laboratory will be described briefly. These are wordprocessors and data base management systems.

Wordprocessor programs for microcomputers now have greatly improved abilities to create, edit, and print text. It is expected that they will be expanded to include the ability to store, copy, and merge documents; integrate variable and preformatted text; and perform mathematics, graphics, mailing, filing, and appointment scheduling. In fact many of these features are already available. The laboratory microcomputer can thus be used as a stand-alone wordprocessor in addition to its many other functions.

There are far too many wordprocessor packages available to even begin listing them all. Almost any popular microcomputer magazine will contain advertisements for several; such magazines are also good sources for review articles. This thesis was written using the "Magic Wand" processor. It would be difficult to prove that it speeded up the "whole" process of the research and writing of the thesis, however, it was an integral part of the entire process. It is hard to imagine what it would have been like without a wordprocessor.

The printer (an Epson MX-100) used to print this thesis was not capable of performing all operations permitted by the Magic Wand wordprocessor. In particular sub- and superscripts were unavailable; a list of special symbols used because of printer limitations are listed in appendix C. An upgrading kit, which would allow additional functions including sub- and superscripts, was ordered but wasn’t received in time to be used. Many low cost printers having
most of the functions required for wordprocessing are now available.

Two practical uses of wordprocessors are writing and submitting papers/texts for publication, and the writing of laboratory manuals and procedures. Not only does the wordprocessor allow the writer to compose at the terminal, it also allows him to submit the text electronically, either on a floppy disk for example or by telephone. In fact, most microcomputer magazine articles are now submitted in this manner. This allows them to be easily edited before publication, and typeset automatically. Scientific journals might find this to be an effective cost-cutting measure.

Analysts have a need to write and update procedures for analytical methods. Since method descriptions etc. tend to use similar components, phrases, and calculation expressions, it is possible to have a library of common "word modules". Once stored in computer memory these modules can be selected and assembled later using a wordprocessor to form the skeleton of the required document. This skeleton can then be finished by filling in the blanks etc. After a method's procedure is completed and printed it remains stored in the computer's memory; later it can be updated, and reprinted easily and quickly. Skelly (74) has reported on the implementation of such a process.

Data base management systems (DBMS) are almost as common as wordprocessor programs. A DBMS (75-77) is a program system which facilitates the entry, updating, addition, and deletion of information on computer files; it also allows the manipulation of the data for the generation of useful reports. Both wordprocessors and DBMSs are "horizontal" programs which can be used by many different types of users. These programs probably have almost as many uses as there are users.

Chemists are generally aware of the large data base systems (78) which are available for literature searching etc. There are also laboratory information management systems (LIMS) available (79), but they are expensive. These two systems will probably act as models for the development of in-house DBMSs; two factors which will promote this
development is the availability of new, powerful software, and the reduction in cost of hard-disk computer memory. Such memory is now available with storage capacities of 5 to 80 megabytes. In-house DBMSs will initially be stand-alone, but will undoubtably later be interconnected with other data base systems.

Personal computer software sales for 1983 are expected (80) to be in excess of 2.3 billion dollars. The leading wordprocessors have monthly sales of: WordStar—15,000, Applewriter—15,000, and Easywriter—7,000 units. DBMSs have monthly sales of: PFS—15,000, and DBase II—7000 units. A third type of package is the "spreadsheet" (81,82); the number one best selling application program of any type is Visicalc. Over half a million copies have been sold; current sales are 20,000 per month. Two similar programs are Lotus 1-2-3 and Multiplan. Each have sales of 15,000 units per month.

Application programs for the laboratory will not have sales of this magnitude. Therefore development of software will be slower and prices will be higher. However, some laboratory software is already being advertised in analytical journals (e.g., Analytical Chemistry).

1.9 Laboratory Facilities

The majority of the programming and computing for this thesis was done with a Zenith Data Systems Z89 microcomputer. The computing laboratory was set up in the author's home. The author's computer was occassional linked to Northern College's Digital PDP 11/70 timesharing computer by an acoustic modem and the public telephone system.

The Zenith Z89 or Heathkit H89 is an all-in-one 8-bit microcomputer which has built in floppy disk storage, a smart video terminal, two Z80 microprocessors, three serial I/O ports, and up to 64K RAM. In addition it has desirable professional qualities such as a 12 inch cathode ray tube (CRT) that permits an 80 X 24 line display plus a 25th line, an 80 key typewriter format keyboard, and it is DEC VT-52 software compatable. It is fully expandable in terms of adding peripheral devices such as printers, plotters, floppy
Several of these computers have been subjected to daily use at Northern College both as stand-alone computers and as terminals to the Digital computer. They have stood up well enough that three additional ones were later purchased to be used as word processors. One of the original computers has been used in the chemistry laboratories for over three years without any maintenance being required. Thus, this model of microcomputer has proved itself in terms of dependability. Microcomputers are however quickly outdated; the Z89 is no longer being produced—it has been replaced by a 16 bit model.

The Z89 microcomputer was also available as a kit; assembly of the kit has been discussed (83,84) and the computer has been described (85-87) in the popular journals.

The author's Z89 has 64K RAM (random access memory); one single density single-sided (48 tpi), and one single density, double sided (96 tpi) 5 1/4 inch disk drives. Total disk capacity is approximately 500 kilobytes. The computer is connected to both an Epson MX-100 dot matrix printer (89), and a Datamex acoustic modem which is interfaced with a Digital PDP 11/70 computer. (Parson (89) has described how data is sent over telephone lines.)

The computer operates on either of two disk operating systems; HDDS--the Heath Operating System, and CP/M--Control Program/Microcomputers. Both of these operating systems were available to the author. Additional software available included Benton Harbour BASIC, Microsoft BASIC, three wordprocessors (Autoscribe, Magic Wand, and WordStar), Control (a data base management system), MCS and CPS (modem programs), plus miscellaneous editors, assemblers, etc.

In addition the author's laboratory had a Texas Instrument TI59 programmable calculator and a Sinclair ZX81 microcomputer (90). The full services of Northern College's computer facility were also available.
1.10 Summary

The purpose of the introduction has been to define the scope of the research conducted. Each of the main topics of the thesis have been considered and their relationship with analytical chemistry has been established. Where necessary a brief background of the topic has been presented along with literature references for additional information. Microcomputers and computer software have received a more intensive description. In defining the scope of the research, the introduction has specifically pointed out those aspects of each topic which will not be considered further.

The thrust of the research is directed specifically at microcomputer software which can be used by the analyst. Commercial software packages were considered. The major emphasis will be on the author's software for statistical analysis of analytical data.

The literature cited has included references from the popular "home computer" magazines. Many articles in these magazines are quite technical in nature. The term "home computer" can be very misleading. It implies that these computers, and hence the supporting magazines, are for the typical non-scientific consumer. This is not the case. Usually the more expensive microcomputers are purchased by business or scientific professionals. The popular journals tend to reflect this user background. Therefore, these journals are excellent sources of information for both the novice and experienced computer owner and programmer. The term "personal computer" is now replacing "home computer" as a descriptive name as it better reflects the nature of the microcomputer.

The introduction has at times been more personal in nature than the reader might expect to see. This has been deliberate as the author felt that since the bulk of the research was done off-campus it was important to establish the "setting" in which it actually took place. In particular, the author wanted to demonstrate that a typical analyst, with an interest in computers and programming, and with a modest budget, could develop a relatively powerful
computing system for his laboratory.

The first version of this introduction was written in mid 1982, this final version was written in late 1983. Even in that short a period of time much of the information on microcomputers and the available software has become dated. The IBM-PC microcomputer has now become established as the standard, a new industry has developed to support it, other microcomputer manufacturers are dropping out of the picture, but already the replacement for the IBM-PC is expected shortly.

There has been one relatively constant factor however and that is the survival of many of the popular software packages. It is quite likely that as new microcomputers are developed the best of the existing software will be retained. It will naturally be altered to take advantage of the new computers, but it represents too large an investment to be discarded. Plus, software tends to gather a group of loyal supporters who help to maintain its popularity.

It will be interesting to watch new developments, and to see how they are applied in the analytical laboratory.
Chapter 2
BASIC APPLICATIONS OF STATISTICS

2.1 Introduction

The analyst is familiar with the terms "precision" and "accuracy" (1) and the difference between them as they are applied to analytical results. He also realizes that the accuracy of analytical results for "real" samples cannot be absolutely known.

Accurate results must usually have good precision, but precise results are not necessarily accurate. Precision, however, is a quality which can be determined reasonably easily for an analytical result, and if results are precise then there is a possibility that they are also accurate. Therefore, analysts are interested in knowing the precision of their measurements. Since statistics allows precision to be estimated therefore analysts are, or should be, interested in using statistics in their data analysis.

Statistics, for the analyst, is simply a mathematical tool for determining the precision of results. Although, it also allows many interesting tests to be performed on analytical data. Accuracy, however, remains the responsibility of the analyst. Accuracy is limited by both random error and systematic error due to bias. Statistics can measure the precision of the random uncertainty and with this knowledge the total error can be estimated for a method with the use of standard samples.

McFarren et al. (2) have discussed precision and accuracy in the determination of the so-called "total error" of analytical methods. They have shown that the precision (standard deviation of the random error) and accuracy (systematic error or bias) of analytical methods can be combined to estimate the total possible error of a method.

\[ \text{Total Error} = \text{Abs. mean error} + 2 \times \text{Std. Dev.} \times \frac{100}{\text{True value}} \] (1)
Midgley (3) has pointed out that the use of the constant "2" in the above expression overestimates the error. He suggested modifications to ensure that the calculated error does not include more than the proposed 95% of results.

It is interesting to note that methods, with "total errors" of 25% or less, are categorized as excellent by the authors; and that such methods are common. Analysts who believe that results of their routine analyses have errors of 10% or less should read this paper.

The journal of Analytical Chemistry annually publishes (fig. 2.1) a list of statistical terms to be used by authors. The fact that this is necessary implies that some statistical terms are known by more than one name, which they are, and that some authors may need to be reminded that a statistical treatment of their data is required.

This chapter will discuss and illustrate the use of microcomputer software for basic statistics in analytical chemistry. Additional uses of statistics will be discussed in chapters 3 and 4. Examples from the analytical literature have been chosen to illustrate the use of applied statistics for the analyst. Appendix A contains additional examples.
Guide for Use of Terms in Reporting Data in Analytical Chemistry

It is important to know the meaning of the terms an author uses. For publications in Analytical Chemistry, the following definitions are applicable and it is understood that they are used with a series of normally distributed replicate results with no prior information on bias of the method. They are ordered by numbers of the Analytical Society. The Guide is necessarily incomplete, and it should be used only with an understanding of its limitations; one is that a value obtained for a term is usually based on a relatively small number of observations, and it is therefore to be regarded as an estimate of the parameter. For appropriate background, the reader should consult a reputable text on the subject of data evaluation.

Set refers to a number of independent replicate measurements of some property. Authors are encouraged to report this number n. Precisions refers to the reproducibility of measurements within a set, that is, the scatter or dispersion of a set about its central value.

Accuracy normally refers to the difference (error or bias) between the mean, \( \bar{X} \), of the set of results and the true or correct value for the quantity measured. It is also used as the difference between an individual value \( X_i \) and \( \bar{X} \). The absolute accuracy of the mean is given by \( \bar{X} \pm s \), and of an individual result by \( X \pm \sigma \). The relative accuracy of the mean is given by \( \bar{X} \pm \sigma/\bar{X} \), and the precision accuracy \( 100(\bar{X} - \bar{X})/\bar{X} \).

Measures of the Central Value of a Set. Mean (or Average or Arithmetic Mean) is the sum of all \( X \), of the values of individual results divided by the number, n, of results in the set. The mean is given by:

\[
\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum X_i}{n}
\]

Median is the middle result of an odd number of results, or the average of the central pair for an even number, when they are arranged in order of magnitude. The median is less affected by extreme values than is the mean.

Measures of Precision within a Set. Standard Deviation is the square root of the quantity sum of squares of deviations of individual results from the mean, divided by one less than the number of results in the set. The standard deviation, s, is given by:

\[
s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}
\]

Standard deviation has the same units as the measurement. It becomes a more reliable expression of precision as n becomes large. When the measurements are independent and normally distributed, the most useful statistics are the mean for the central value and the standard deviation for the dispersion.

Variance, \( \sigma^2 \), is the square of the standard deviation.

Relative Standard Deviation is the standard deviation expressed as a fraction of the mean, \( s/\bar{X} \). It is sometimes multiplied by 100 and expressed as a percentage. Relative standard deviation is preferred over "coefficient of variation".

Mean (or Average) Deviation is the mean of the deviations of the individual measurements from the mean of the set without regard to sign. It is given by \( \sum |X_i - \bar{X}|/n \). The mean deviation is not recommended as a measure of precision except when the set consists of only a few measurements.

Range is the difference in magnitude between the largest and smallest result in a set. The range is not recommended as a measure of precision except when the set consists of only a few measurements. If range is used, the number of measurements in the set must be indicated.

Measure of Precision of a Mean. Confidence Limits (or Interval) are the limits around the measured mean within which the mean value for an infinite number of measurements can be expected to be found with the stated level of probability. Confidence limits for independent normally distributed measurements are given by:

\[
\text{confidence limits} = \bar{X} \pm t/s\sqrt{n}
\]

where \( t \) is the standard deviation and \( t \) is the t-table value at the stated confidence level. The use of standard error, \( s/\sqrt{n} \), to express precision of a mean is acceptable only if the authors clearly make the distinction from standard deviation.

Fig. 2.1 Basic statistical terms and definitions. Reprinted with the permission of the American Chemical Society.

2.1.1 Literature Search

The Analyst (1969 - 1983) was reviewed for papers which illustrated statistical treatment of analytical data and for data which could be used to illustrate some of the basic statistical tests. Analytical Chemistry (1969 - 1983), Talanta (1970 - 1983), and Analytica Chirina Acta (volumes 95, 103, 112, 122, and 133) were also reviewed. Most of the examples used in this chapter are from the Analyst.

In general, it seemed that the use of basic statistics could be best illustrated for the practising analyst with examples and data from the Analyst. Papers in Analytical Chemistry did receive similar or more elaborate treatment, as well, many papers contained in-depth treatment of more advanced statistical methods. This journal also has periodic reviews of statistics (chemometrics, sec. 1.2). Many of these "statistical papers" will be referred to throughout this thesis.
It is not implied, however, that the Analyst does not publish papers concerned with statistical theory (as an example see Box's paper (4) on statistical design in the study of analytical methods). The Analyst simply had a more manageable number of papers and allowed an easier overview of the use of statistics for the period reviewed.

2.1.2 The Computer Program

The program "Applied Statistics I" (Appendix A) was written to provide analysts with an easy to use "tool" for the analysis of laboratory data using basic statistical tests. "Easy to use" means a program which allows:
- it to be used with minimum reference to a manual
- data sets to be stored, recalled, edited, transformed
- user errors to be trapped with automatic recovery
- the data sets to be analyzed by many different tests
- statistics to be used without the need for statistical tables

This program will not necessarily allow the analyst to work completely independently of a statistician, but should at least allow him to assume the role of an "amateur statistician".

Program Evaluation

Programs cannot be completely tested by the programmer. The purpose of this research was to produce a program which could be used, to advantage, by students and scientists. It was decided that the best way to determine whether this purpose had been achieved was by offering the program to the market place on a commercial basis. Therefore, the original program "Applied Statistics I" was offered to a commercial software distributor (Sunflower Software, 13915 Midland Drive, Shawnee, KS 66216, U.S.A.).

Two hundred plus copies of the program have now been sold across North America. The list of purchasers includes universities, government research agencies, private industry, and the general public. Specific examples are the following:
- St. Vincent Hospital and Medical Centre - Toledo
- University of Southern California - Los Angeles
- Advanced Business Strategies - Sylvania
- Western Michigan University - Kalamazoo
- Saint Olaf College - Minnesota
- Fortrend Engineering Corp. - California
- Attorney at Law - Van Nuys
- Central Missouri State University - Warrensburg
- Environmental Research Institute of Michigan
- Gillette Co. - Cambridge, Mass.
- North Memorial Medical Centre - University of Minnesota

No serious negative comments have been received. Some favourable comments include the following:

- the program has been very complementary to my course (Quality Assurance)
- the program can do a lot of things that the statistics program on our "big computer" cannot do
- a fine program and I enjoy working with it

The experiment to determine whether this microcomputer program was useful and acceptable on both a scientific and a commercial basis has been determined to be a success. It would appear that the program is useful even in a situation where "big" computers are available.

There are two additional positive outcomes of this experiment. The program is now widely distributed and is available in many more formats (the distributor has modified it to run under four different computer operating systems). Secondly, since it was distributed on a royalty basis, an additional source of funds to expand the computing system used for this research was obtained. Since no other sources of funds were available this was an important outcome.

The Complete Program

The original program was modified and expanded for the purposes of this thesis (fig. 2.2). The actual program used ran under the CPM operating system. This system and its BASIC (MBASIC-80 by Microsoft) allowed a much larger program to be resident in memory and allowed greater mathematical accuracy.
The following modifications/additions were made to the original program "Applied Statistics I":
- greater mathematical accuracy
- improved data editing
- improved data transformations
- expansion of the ANOVA module to include hierarchic factors, and the calculation of factor standard deviations
- replacement of the simple regression module with a new and very powerful module which among other things allowed the use of weighting factors
- addition of a "normal" random number generator
- simpler output of data to a printer
- improved checks for normal distribution of data.

The program which resulted at the end of the research is at least twice as powerful as the original program.

A problem which required considerable extra time for its solution was related to the mixing of double and single precision numbers and variables. CPM MBASIC handles this problem better than HODS MBASIC, but a real problem still exists in writing statistical programs where intermediate calculations, involving small differences between the squares of large similar numbers, are required.

Single precision calculations, variables, and constants were used almost exclusively. This was done to conserve computer random access memory (RAM), disk space, and to simplify formatting of results. Some calculations, however, were done in double precision to increase the accuracy of results.

When a double precision (16 significant figures) variable is set equal to a single precision (7 significant figures) value only the first seven digits, rounded, of the converted number will be valid. Also, most arithmetic functions [e.g., SQR(X)] will return only a single precision answer, even if X is double precision. Hopefully new versions of the BASIC language and the new microcomputers will overcome this problem.
1. HISTOGRAM
2. DATA SET STATISTICS (EDIT, TRANSFORM, etc.)
3. COMPARISON OF TWO SAMPLE MEANS (t and F TEST)
4. COMPARISON OF TWO SAMPLES, DATA IN PAIRS (t TEST)
5. COMPARISON OF A SAMPLE MEAN WITH A KNOWN MEAN
   <USING THE KNOWN STANDARD DEVIATION> (U TEST)
6. COMPARISON OF A SAMPLE MEAN WITH A KNOWN MEAN
   <USING THE SAMPLE STANDARD DEVIATION> (t TEST)
7. PROBABILITY OF "F" OCCURING
8. PROBABILITY OF "U" OCCURING
9. PROBABILITY OF "t" OCCURING
10. CRITICAL VALUES OF "t" and CHI^2
11. ANALYSIS OF VARIANCE (1 and 2 FACTOR-REPPLICATE TESTS)
12. LINEAR REGRESSION ANALYSIS - CALIBRATION CURVE
13. SAVE/RECALL DATA ON/FROM DISK - END PROGRAM
14. CHANGE SET STATISTICS    15. CHANGE DISKS
16. CATALOG/KILL FILES       17. CHANGE # OF SETS IN USE
18. SIMULATED "NORMAL" MEASUREMENTS

Fig. 2.2 Menu for the complete applied statistics program.

Other Statistics Programs Available

The microcomputer user also has a choice of other
statistics programs. These are available for most commonly
available microcomputers. For the Apple computer there is
Microstat by Ecosoft (5) and Edu-Ware's Statistics 3.0 (6),
and for the Radio Shack microcomputer there is advanced
Statistics by Creative Computing Software (7). Information
on these and other programs is available in the popular
computer magazines.

The program described here was developed completely
independently of these others; it is the outcome of program
development over several years. In addition, it was
developed specifically for the practising analytical
chemist.
2.2 The Distribution of Analytical Measurements

The mean and standard deviation of any data set, whatever its distribution, can be calculated and will be useful as a means of compressing the data (8). However, the data must be normally distributed before the confidence limit of the data set and its mean can be correctly estimated, and before such data sets can be correctly used in simple (parametric) statistical tests of significance. The confidence limits for the mean have been defined in figure 2.1; the confidence limits for the data set refer to the limits around the mean within which all individual measurements are expected to be found, with the stated level of probability.

In the following discussions for estimating and testing using statistics, it will be assumed that the underlying populations (measurements) are normally distributed. These procedures can be referred to as being parametric. Nonparametric, i.e., distribution-free procedures (9), assume no such normal distribution.

This assumption of normal distribution is usually stated in textbooks but then often conveniently laid aside. How does the experimenter determine whether his data can be considered to be normally distributed? The program allows every data set to be examined in three ways to determine if it is suitable for further statistical analysis. These examinations are not exhaustive, but they should detect most problem data sets.

These three examinations are accessed using the HISTOGRAM module. This module allows a data set to be checked for outliers, tested for non-normal distribution, and arranged in a histogram for visual inspection.

2.2.1 Outliers

There are many tests for outliers, the one used in the program is "Chauvenet's Criterion" and has been discussed by Kennedy and Neville (10, p. 180). This particular test was chosen as it was convenient to program. Proschchan (11) and Dybczynski et al. (12) discuss other tests which could be used. Different tests may give different results, however,
it is the analyst who must make the final decision as to which, if any, data is to be omitted. It should be noted that many statistical tests are very sensitive to outliers in the data (13, p. 180).

2.2.2 Normal Distribution

There are also many tests for normality; Kateman and Pijpers (14) have produced a table of some of the tests available. They considered the "Kolmogorov Test" (15) in some detail and found it to be generally applicable, easy to use, and one of the dependable tests. The numerical Kolmogorov test is used in the program. Hirsch (16) has also discussed the Gaussian distribution of random errors; he also suggests the use of the Kolmogorov-Smirnov test.

For this test the data is arranged in ascending order. If the cumulative frequency lies outside the range (at a chosen confidence level) allowed for a normal distribution, then the data set is considered to be "non-normal". Again this test can also be performed without the use of tables (critical values are calculated by the program using a suitable algorithm). The program allows the user to visually observe each calculated statistic and its corresponding critical value. Thus, the user can see what points, if any, are outside the acceptable range. Only one point need be outside the range to reject the null hypothesis that the data set is normally distributed. The computer will indicate whether the data set should be considered to be non-normal (the level of significance used is $\alpha = 0.05$).

It should be noted that in this test the data set is assumed to be normal unless it can be proven otherwise. If it cannot be proven otherwise, it does not necessarily follow that the data set is normally distributed. For example, a data set with a uniform distribution (the numbers from 1 to 9) will pass the test for normality used in this program. A combined set of ten such sets also passed the test; however, a combined set of 20 such sets failed to pass. Kateman and Pijpers (14, p. 14) give another example of the application of this test to a small data set; their results were duplicated by the program.
2.2.3 Histograms
Lastly the user can plot a histogram (10, p. 10) of the data set. The upper and lower limits, and the number of intervals can be altered to produce any number of histograms on the computer screen. The histogram can also be directed to a printer if required. This histogram can be visually examined to "see" if it resembles a normal distribution. The descriptive statistics for the data set and for the mean can then be displayed on the screen or sent to the printer.

2.2.4 Availability of Routines
for Outliers, Normality and Histograms
All of the above tests, with the exception of the histogram, can also be produced by the second module in the menu. Additionally, data sets called by most of the other program modules can be tested for outliers and normal distribution, and can be transformed or edited without the need to use either of the first two modules directly.

The analyst, therefore, no longer has to rely on statements of the generality of normal distributions found in many texts (17, p. 5; 18, p. 261), e.g.,

"Many distributions which arise in science and engineering are normal or almost normal. Distributions ranging from heights or weights of animals to dimensions of articles produced by automatic processes often closely follow the law. So also, do human and other errors in physical and chemical readings. Data from such distributions can therefore be tackled using normal distribution theory."

Every data set used with this program can be tested for normality.

2.2.5 Examples
Appendix A contains an example of the histogram module. As an additional example a group of nine sets of thirty-six transmittance measurements (from sec. 3.5.2, see Table 2.1) were tested. The results in the form of the mean, variance, histogram, and results of the test for normal distribution can be found in Table 3.7. Figure 2.3 gives the detailed
results from the test on one of these data sets. These tests were performed on all data sets used in this thesis; the tests are particularly useful in checking the residuals resulting from linear regression (ch. 3).

**TABLE 2.1**

NINE SETS OF THIRTY-SIX TRANSMITTANCE MEASUREMENTS

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.986</td>
<td>.753</td>
<td>.629</td>
<td>.541</td>
<td>.448</td>
<td>.379</td>
<td>.315</td>
<td>.282</td>
<td>.231</td>
</tr>
<tr>
<td>.986</td>
<td>.753</td>
<td>.632</td>
<td>.539</td>
<td>.450</td>
<td>.378</td>
<td>.314</td>
<td>.284</td>
<td>.229</td>
</tr>
<tr>
<td>.985</td>
<td>.753</td>
<td>.632</td>
<td>.542</td>
<td>.450</td>
<td>.378</td>
<td>.314</td>
<td>.283</td>
<td>.229</td>
</tr>
<tr>
<td>.983</td>
<td>.756</td>
<td>.630</td>
<td>.539</td>
<td>.452</td>
<td>.378</td>
<td>.317</td>
<td>.283</td>
<td>.230</td>
</tr>
<tr>
<td>.983</td>
<td>.755</td>
<td>.629</td>
<td>.540</td>
<td>.452</td>
<td>.380</td>
<td>.319</td>
<td>.283</td>
<td>.230</td>
</tr>
<tr>
<td>.983</td>
<td>.754</td>
<td>.631</td>
<td>.540</td>
<td>.453</td>
<td>.379</td>
<td>.319</td>
<td>.282</td>
<td>.232</td>
</tr>
<tr>
<td>.983</td>
<td>.754</td>
<td>.629</td>
<td>.541</td>
<td>.452</td>
<td>.379</td>
<td>.318</td>
<td>.283</td>
<td>.232</td>
</tr>
<tr>
<td>.983</td>
<td>.754</td>
<td>.631</td>
<td>.540</td>
<td>.450</td>
<td>.378</td>
<td>.318</td>
<td>.284</td>
<td>.231</td>
</tr>
<tr>
<td>.983</td>
<td>.755</td>
<td>.631</td>
<td>.542</td>
<td>.449</td>
<td>.380</td>
<td>.316</td>
<td>.284</td>
<td>.230</td>
</tr>
<tr>
<td>.983</td>
<td>.755</td>
<td>.631</td>
<td>.539</td>
<td>.448</td>
<td>.381</td>
<td>.318</td>
<td>.282</td>
<td>.229</td>
</tr>
<tr>
<td>.983</td>
<td>.755</td>
<td>.631</td>
<td>.541</td>
<td>.448</td>
<td>.380</td>
<td>.317</td>
<td>.281</td>
<td>.231</td>
</tr>
<tr>
<td>.983</td>
<td>.753</td>
<td>.630</td>
<td>.539</td>
<td>.447</td>
<td>.381</td>
<td>.315</td>
<td>.283</td>
<td>.230</td>
</tr>
<tr>
<td>.983</td>
<td>.754</td>
<td>.631</td>
<td>.541</td>
<td>.447</td>
<td>.382</td>
<td>.315</td>
<td>.283</td>
<td>.232</td>
</tr>
<tr>
<td>.983</td>
<td>.755</td>
<td>.631</td>
<td>.540</td>
<td>.448</td>
<td>.382</td>
<td>.318</td>
<td>.285</td>
<td>.229</td>
</tr>
<tr>
<td>.983</td>
<td>.755</td>
<td>.631</td>
<td>.539</td>
<td>.448</td>
<td>.381</td>
<td>.318</td>
<td>.284</td>
<td>.227</td>
</tr>
<tr>
<td>.983</td>
<td>.756</td>
<td>.631</td>
<td>.540</td>
<td>.448</td>
<td>.381</td>
<td>.322</td>
<td>.283</td>
<td>.227</td>
</tr>
<tr>
<td>.983</td>
<td>.759</td>
<td>.632</td>
<td>.540</td>
<td>.451</td>
<td>.383</td>
<td>.320</td>
<td>.282</td>
<td>.227</td>
</tr>
<tr>
<td>.983</td>
<td>.757</td>
<td>.631</td>
<td>.541</td>
<td>.450</td>
<td>.380</td>
<td>.321</td>
<td>.283</td>
<td>.227</td>
</tr>
<tr>
<td>.983</td>
<td>.758</td>
<td>.630</td>
<td>.540</td>
<td>.449</td>
<td>.380</td>
<td>.320</td>
<td>.284</td>
<td>.228</td>
</tr>
<tr>
<td>.982</td>
<td>.760</td>
<td>.630</td>
<td>.541</td>
<td>.450</td>
<td>.379</td>
<td>.321</td>
<td>.280</td>
<td>.229</td>
</tr>
<tr>
<td>.983</td>
<td>.759</td>
<td>.631</td>
<td>.540</td>
<td>.448</td>
<td>.392</td>
<td>.320</td>
<td>.282</td>
<td>.228</td>
</tr>
<tr>
<td>.985</td>
<td>.757</td>
<td>.632</td>
<td>.542</td>
<td>.447</td>
<td>.382</td>
<td>.318</td>
<td>.282</td>
<td>.227</td>
</tr>
<tr>
<td>.985</td>
<td>.757</td>
<td>.633</td>
<td>.544</td>
<td>.449</td>
<td>.382</td>
<td>.317</td>
<td>.281</td>
<td>.229</td>
</tr>
<tr>
<td>.985</td>
<td>.757</td>
<td>.634</td>
<td>.544</td>
<td>.448</td>
<td>.383</td>
<td>.316</td>
<td>.262</td>
<td>.229</td>
</tr>
<tr>
<td>.985</td>
<td>.758</td>
<td>.634</td>
<td>.543</td>
<td>.447</td>
<td>.382</td>
<td>.315</td>
<td>.283</td>
<td>.228</td>
</tr>
<tr>
<td>.986</td>
<td>.760</td>
<td>.633</td>
<td>.541</td>
<td>.447</td>
<td>.381</td>
<td>.315</td>
<td>.283</td>
<td>.230</td>
</tr>
<tr>
<td>.986</td>
<td>.757</td>
<td>.634</td>
<td>.544</td>
<td>.447</td>
<td>.378</td>
<td>.316</td>
<td>.284</td>
<td>.229</td>
</tr>
<tr>
<td>.985</td>
<td>.756</td>
<td>.633</td>
<td>.540</td>
<td>.447</td>
<td>.379</td>
<td>.315</td>
<td>.283</td>
<td>.231</td>
</tr>
<tr>
<td>.985</td>
<td>.753</td>
<td>.632</td>
<td>.541</td>
<td>.447</td>
<td>.380</td>
<td>.318</td>
<td>.282</td>
<td>.231</td>
</tr>
<tr>
<td>.986</td>
<td>.756</td>
<td>.628</td>
<td>.539</td>
<td>.445</td>
<td>.380</td>
<td>.315</td>
<td>.283</td>
<td>.231</td>
</tr>
<tr>
<td>.985</td>
<td>.756</td>
<td>.629</td>
<td>.540</td>
<td>.447</td>
<td>.381</td>
<td>.314</td>
<td>.285</td>
<td>.230</td>
</tr>
<tr>
<td>.985</td>
<td>.755</td>
<td>.631</td>
<td>.539</td>
<td>.447</td>
<td>.381</td>
<td>.316</td>
<td>.283</td>
<td>.230</td>
</tr>
<tr>
<td>.986</td>
<td>.755</td>
<td>.629</td>
<td>.540</td>
<td>.445</td>
<td>.383</td>
<td>.314</td>
<td>.283</td>
<td>.228</td>
</tr>
<tr>
<td>.984</td>
<td>.754</td>
<td>.626</td>
<td>.540</td>
<td>.445</td>
<td>.382</td>
<td>.315</td>
<td>.283</td>
<td>.227</td>
</tr>
<tr>
<td>.983</td>
<td>.754</td>
<td>.627</td>
<td>.538</td>
<td>.446</td>
<td>.381</td>
<td>.315</td>
<td>.282</td>
<td>.229</td>
</tr>
</tbody>
</table>
Fig. 2.3 The histogram (A) and descriptive statistics (B) from an ordered set of thirty-six transmittance measurements (C). This data is from the third set of measurements in Table 2.1.
Nine examples from the Analyst (19-22), where the data sets were actually presented and confidence limits calculated, were checked for normality. None of the data sets were determined to be non-normal. Two cases of possible outliers were found. It is interesting to note, that although there was no indication that the distribution of the data was examined by the various authors, there was no reason to assume (at least according to the tests available in the program) that the data was not normally distributed. Evans (23) in a paper discussing the quantification of trace elements in foodstuffs did discuss the necessity for normal distributions in the application of statistical tests on experimental data. He did not, however, apparently attempt to conduct any such tests. Two papers in the educational literature (24,25), which discussed the calculation of confidence limits, also did not indicate the necessity to check for normal distribution.

Similar examples (26), but ones in which outliers were previously determined, were also reprocessed using the program. All but one of the outliers found by the authors were determined by the program; again none of the data sets could be shown to be non-normal.

Examination of figure 2.3 will show that the program gives the necessary descriptive statistics for both the entire data set and for the mean. Confidence limits, at the commonly used confidence levels of 90%, 95%, and 99%, are calculated. The previous examples cited used either the 95% or the 99% confidence levels.

2.3 Data Transformations

The analyst may be interested in making data transformations to:
- simplify (code) data
- change a non-normal data set to a normally distributed one
- change measurements to a more analytically useful form.

The program allows several types of transformations to be made.

Each measurement in a set can have a constant added to or subtracted from it, can be multiplied or divided by a
constant, can be raised to any power, or have its logarithm (common or natural) calculated. These transformations can be used to simplify or normalize (14, p. 141) a data set.

An example of a transformation for analytical convenience is the conversion of transmittance measurements to absorbance. This is accomplished (sec. 3.5.2) by first entering the transmittance measurements. These measurements are then raised to the power of -1 to obtain their reciprocals. The logarithms of the reciprocals (the absorbances) are then calculated.

Although not transformations, the data sets can also be edited. It is possible to add, delete, change, and insert measurements. As well, sets can be ordered (and a second set to the same order, e.g., for linear regression), moved, and merged. The ordering of data sets allows the median to be readily identified and gives yet another display of the data which could be useful for identifying outliers and possible non-normal distributions.

All of these possible data manipulations increase the usefulness of the program without requiring the re-entering of data; the original data set and any number of transformed sets resulting from it can be stored on the disk.

2.4 Monte Carlo Simulated Analytical Data

Davies and Goldsmith (27) have devoted a chapter of their text to simulation; the following quote is descriptive of the need for simulation in research:

"There are often difficulties in carrying out experiments on large systems without spending considerable amounts of time and money. One way of overcoming these is to build a simulation model which provides a numerical imitation of the real system."

In this particular research a model, which would allow the simulation of results from various analyses, was required. In addition to a savings of time and money, this model generates data having known statistical parameters. These parameters are a known distribution, mean, and standard deviation or relative standard deviation. Thompson (28) has made extensive use of Monte Carlo
simulation for the examination of analytical errors. He has pointed out that real data sets are in fact not appropriate for such studies as their bias parameters are unknown.

2.4.1 Requirements of the Simulation Model

The model is required to produce "measurements" which have a normally distributed error. The standard deviation of the error is to be either constant or proportional to the magnitude of the response. In addition to producing a set of measurements with a known standard deviation or relative standard deviation, a set of measurements simulating a series of responses which would be produced for a calibration curve is also required. This model is used for the studies conducted for chapters 3 and 4.

2.4.2 The Algorithm Used

Wilkins et al. (29) have given an equation for the Gaussian, or "normal" distribution of indeterminate errors which is particularly easy to work with. Their equation was rearranged to give:

\[ Y = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\text{RNDx}^2}{2}ight) \]  

(2)

In this equation \( Y \) is the probability of obtaining a value \( x \) as the result of a single measurement. \( \text{RNDx} \) is a random value of \( x \) within the range of \( 0 \) +/- infinity, although in this algorithm it will only be allowed to vary within the range of \( 0 \) +/- 6 or \( 0 \) +/- 40. It is assumed that the mean of the data set is 0 and that the standard deviation is 1.

The Monte Carlo method will be used to determine values of \( x \) which satisfy the above equation. In the Monte Carlo method (29,30) random points, i.e., pairs of \( x \) and \( y \) values, are generated and then checked to determine if they fall under the normal distribution curve (as represented by eq. 2).

The random number generator which is part of the BASIC language is used to generate a value of \( x' \) between 0 and 1; the resulting number is multiplied by 12, and then 6 is subtracted from that value; this should result in a value of
within the range of 0 +/− b. The random number generator is again used to obtain a value between 0 and 1; this value is multiplied by 1/SQR(2). If the resulting value is equal to or less than the corresponding value of Y as calculated by equation 2, then x is assumed to be a valid "measurement". This process is repeated until the required number of measurements have been obtained.

If the expected mean was not 0, and/or the expected standard deviation was not 1, then x is multiplied by the standard deviation, and the resulting value is added to the mean to obtain the final "measurement".

Table 2.2 shows a BASIC language routine to calculate simulated analytical measurements; it uses an algorithm similar to the above.

**TABLE 2.2**
A BASIC PROGRAM TO CALCULATE A SERIES OF NORMALLY DISTRIBUTED MEASUREMENTS

```basic
5 I=0: NUM=0
10 INPUT "MEAN (999 TO END); MEAN:" IF MEAN=999 THEN 100
20 INPUT "STANDARD DEVIATION"; SD:
30 INPUT "NUMBER OF MEASUREMENTS"; NUM
40 X=RND: Y=RND: X=X * 12 - 6
50 IF EXP(-X^2/2)=Y THEN X=X * SD + MEAN: I=I +1:
   PRINT I, "THE NORMAL MEASUREMENT IS"; X
60 IF I<NUM THEN 40 ELSE 5
100 END
```
The values of $x$ used for the simulations of chapters 3 and 4 were actually allowed to vary between $\pm 40$; it was therefore possible to generate "outliers" which were as much as 40 standard deviations from the mean. Table 2.3 shows the statistical parameters calculated from a series of data sets (500 measurements each) generated using this algorithm. The expected mean and standard deviation were 0 and 1.

**TABLE 2.3**

**STATISTICAL PARAMETERS CALCULATED FOR SETS OF SIMULATED MEASUREMENTS**

<table>
<thead>
<tr>
<th>Measurements could be within the range of $0 \pm 40$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MEAN</strong></td>
</tr>
<tr>
<td><strong>SD</strong></td>
</tr>
<tr>
<td><strong>RANGE</strong></td>
</tr>
<tr>
<td><strong>OUTLIER</strong></td>
</tr>
<tr>
<td><strong>DIFF</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measurements could be within the range of $0 \pm 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MEAN</strong></td>
</tr>
<tr>
<td><strong>SD</strong></td>
</tr>
<tr>
<td><strong>RANGE</strong></td>
</tr>
<tr>
<td><strong>OUTLIER</strong></td>
</tr>
<tr>
<td><strong>DIFF</strong></td>
</tr>
</tbody>
</table>

The differences between the measured means and the expected mean of zero were checked for statistical significance using a t-test (sec. 2.5.4). NS, PS, S, and HS correspond to differences which are: not significant, possibly significant, significant, and highly significant.
A visual examination of the above data suggests that the means may not be evenly distributed, i.e., there are more negative than positive means. Each mean was tested using a t-test (sec. 2.5.4) to determine if it differed significantly from zero; only one did. The mean (+0.0186341) for an additional set of measurements (-6<x<+6) was calculated using the program, and an independent routine (31). Both means were the same; therefore it was reasonable to assume that the means were correctly calculated. If the means are not normally distributed it could be as a result of the BASIC random number generator. The routine was, however, considered to be adequate for the simulations.

This algorithm was independently developed. Davies and Goldsmith (27) and Burr (32) have discussed other possible algorithms for this simulation. The one by Davies and Goldsmith is very straightforward—its suitability will be checked at a future date.

2.5 Significance Tests

The first four tests considered in this section are commonly referred to as "t tests". These tests are actually a subset (13, p. 185) of tests based on the Analysis of Variance (ANOVA) in which an "F test" is used. The "F test" will be used in this section but ANOVA is left to section 2.6. The basis for this simpler treatment is the fact that an F test value with one degree of freedom in the numerator (and any number of degrees of freedom in the denominator) can be expressed as the square of Student's "t" statistic.
2.5.1 Theory

"t" Tests

These tests (fig. 2.2, routines 3-6) will be examined and illustrated. The tests will allow:
- the means of two data sets to be compared; the two sets can have different numbers of measurements
- two data sets to be compared, the two sets have the same number of measurements and the measurements are paired
- the mean of a data set to be compared to a known mean with a known or unknown standard deviation.

In all of these tests a value of Student's "t" is calculated from the data available. This "calculated t" is then compared to the "critical t" which can be found in statistical tables. The "critical t" is the maximum value of "t" which would be expected by chance alone (at a chosen confidence level), if the two means were actually both estimates of the "true" mean. A null hypothesis, that the two means are both estimates of the same "true mean", is assumed. If the "calculated t" is equal to or less than the "critical t" then the null hypothesis has not been disproved; if it is greater than the critical "t" then the null hypothesis is rejected and the two means are assumed to be different.

The general formula for calculating "t" is as follows:

\[ t = \frac{\text{mean}_1 - \text{mean}_2}{\text{avg. SE}} \]  \hspace{1cm} (3)

(SE is the standard error)

If a specific difference is thought to exist between the means the formula is modified as follows:

\[ t = \frac{\text{mean}_1 - \text{mean}_2 - \text{test difference}}{\text{avg. SE}} \]  \hspace{1cm} (4)

This useful feature is not discussed in most textbooks, nor has it been widely used in the analytical literature. It could, for example, be used to determine if a suspected bias actually exists between two analytical methods.
The specific form of the formula for each test is as follows:

Comparison of two samples when the data do not occur in pairs.

\[ t = \frac{\text{mean}_1 - \text{mean}_2}{\text{SE of (mean 1 - mean 2)}} \]  
\[ \text{(5)} \]

Comparison of two samples when the data arise in pairs.

\[ t = \frac{\text{average difference} - \text{test difference}}{\text{SE of the differences between paired measurements}} \]  
\[ \text{(6)} \]

Comparison of a sample mean with an accurately known mean, having an unknown standard deviation.

\[ t = \frac{\text{experimental mean} - \text{known mean} - \text{test difference}}{\text{SE of the experimental mean}} \]  
\[ \text{(7)} \]

\( \text{(the estimated SD and the number of measurements are used in calculating the SE)} \)

Comparison of a sample mean with an accurately known mean, which has a known standard deviation.

\[ t = \frac{\text{experimental mean} - \text{known mean} - \text{test difference}}{\text{SE of the experimental mean}} \]  
\[ \text{(8)} \]

\( \text{(the known SD and the number of measurements are used in calculating the SE)} \)

In this last test the standard deviation of the known mean and the number of measurements in the experimental data set are used to calculate the SE. Since the number of degrees of freedom associated with the known mean is assumed to be large then the "critical t" can also be found using a normal distribution rather than a "t" distribution. This test is therefore usually referred to as a "U" or "Z" test; although only the Student's t table need actually be used.

The purpose of these tests of significance is to separate differences which could not easily have occurred by
chance from those which could have. The interpretation of the statistical significance of the test is done by the program; the analyst need only relate this interpretation to the actual problem.

The theory presented here is purposely brief as most of these tests are well known. In addition most textbooks on statistics (e.g., Brookes et al. 18, Ch. 13) cover the theory in detail. The application of these tests using the program is illustrated in Appendix A and in the following sections (2.5.2 to 2.5.4).

"F" Test

The "F" test is used to compare the variances of two data sets and thus compliments the "t" tests which are used to compare the means of two sets. For example, if two samples are suspected of being from the same source (the null hypothesis) then the variances of the results of an analysis on each sample will be estimates of the "true" variance. For very large numbers of replicate analyses of each sample the two variance estimates should be similar, if the null hypothesis is correct.

The "F" test is a comparison of the larger variance estimate to the smaller (a one sided test). When the null hypothesis is true the ratio could be as low as "1" or as high as "??" by chance alone. Statistical tables give the value of the maximum ratio "??". This critical "F" value depends on the number of measurements -1 (i.e., the degrees of freedom) for each analysis and the chosen level of confidence.

It might be suspected that since "t" can be used to estimate the confidence limits of a mean (fig. 2.1) then "F" might somehow be used to estimate the confidence limits of a variance. Chi-squared (a statistic related to F) can be used for this purpose; it is used in chapters 3 and 4. "F" can actually be used in specific cases (sec. 2.6.4).

Before applying a general "t" test to two sets of unpaired data the variances must be first examined by an "F" test to determine if they can be considered similar. This application is considered in the next section. Additional
theory on this test may be found in any statistics textbook (e.g., Brookes et al. (18), Ch. 13).

2.5.2 Uses of the General "t" Test

As a statistical tool in analytical chemistry the use of the general "t" test (eq. 4) is limited by the fact that it compares two means. If for example, the test was being used to compare the results of two analytical methods then the methods could be tested at only one analyte concentration per test, and for each test several replicates of each sample would be required. This is probably the reason why this test is not used too often for data analysis in analytical chemistry. However, example "C" illustrates how this limitation can sometimes be overcome.

The particular "t" test in this program is somewhat different from the test routinely encountered. It allows two data sets to be compared even when an "F" test indicates that the variances of the two sets cannot be considered similar. If the two estimates of variance can be considered equal then the usual "t" test (eq. 4) is conducted, otherwise the Welch test (33, p. 48; 34, p. 22-39) is conducted.

Examples

Three examples from the chemical literature will be used to demonstrate possible uses of the technique in analytical chemistry. A fourth example can be found in Appendix A; it illustrates testing for a specific difference between two means.

A: Determine whether a statistically significant difference exists between the results from two analytical procedures. In this simple example (35) the two procedures are being tested at only one concentration level. The computed results agree with the published results and both indicate that there is no significant difference between either the variance or the means of the two data sets. A single-sided test to determine if the first mean was significantly larger then the second mean failed to prove
any difference. Figure 2.4 illustrates the results for the single-sided test. Another straightforward application is given by Denney and Smith (36).

**DATA SET #1**  
**DATA SET #2**

<table>
<thead>
<tr>
<th></th>
<th>SET 1</th>
<th>SET 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>0.09896</td>
<td>0.099</td>
</tr>
<tr>
<td>VARIANCE</td>
<td>1.66667E-09</td>
<td>9.25E-09</td>
</tr>
<tr>
<td>NO.</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**CALCULATED 'F' = 5.55**  
**PROBABILITY OF 'F' OCCURRING BY CHANCE = 0.2549%**

**AVG. ESTIMATE OF POPULATION VARIANCE = 6E-09**  
**D.F. = 7**

**CALCULATED 't' = .769841**

**THIS IS A 1 SIDED TEST! THE TEST DIFFERENCE IS 0**

**THE 'NULL HYPOTHESIS (HO)' IS THAT, FIRST MEAN IS 'NOT' SIGNIFICANTLY SMALLER THAN SECOND MEAN LESS 0**

**THE PROBABILITY OF BEING WRONG IN REJECTING 'HO' IS 23.3287%**

This probability is NOT SIGNIFICANT! Therefore 'HO' is ACCEPTED.

**Fig. 2.4 Results of a one-sided general "t" test.**

B: A "t" test as above but one in which the variances of the two data sets cannot be assumed to be similar. In this example (37) replicate determinations of Si in a standard iron/steel sample are compared to the values given on the certificate of analysis. The computer results of figure 2.5 show how the number of degrees of freedom is reduced when the two variances cannot be considered similar (e.g., the number of degrees of freedom would normally be 5+7-2=10, instead there are only 6). In this case there is no estimate of an average population variance, it is reported as "0" (fig. 2.5). The test indicates no significant difference in the results between the procedure reported and those used in estimating the certificate values.
Fig. 2.5 Results of a general "t" test where the two variance estimates are significantly different.

C: Use of the general "t" test where it can actually be used to test the similarity of two analytical procedures, simultaneously, at many analyte levels. In this example (38) the actual concentration determined for each method is not compared, but rather the "% recovery" achieved by each method (Provost and Elder (39) have discussed the statistical interpretation of percent recovery data). Two possible criticisms of the actual experimental conditions under which the data was collected are:

- the concentration levels used for each method were different. Method A used a range of 2-20 whereas Method B used a range of 10-100.
- there was no indication that the standard samples were analyzed in a random order. A requirement of the "t" test is that the measurements in each set be normally distributed; a randomized order of sample analysis could assist in achieving this.

Reprocessing of the data indicated that there was a possible outlier in the first data set; neither data set could be shown to be non-normal. The computer results of the
"t" test are shown in figure 2.6; the computer calculated "t" value differed from the published result (computed, 0.1334; published, 0.148).

Many sets of literature data were reprocessed using the computer program during the duration of this research and it was not unusual to find different and occasionally, incorrectly calculated values (e.g., sec. 2.5.4). In this case a "hand calculation" confirmed both the computer and the published result. The difference arose because of the additional significant figures used in the computer calculation. The same conclusion, namely that there was no significant difference between the % recovery of the two methods, was reached using either "t" value.

<table>
<thead>
<tr>
<th>DATA SET # 1</th>
<th>DATA SET # 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>99.82</td>
</tr>
<tr>
<td>VARIANCE</td>
<td>1.904</td>
</tr>
<tr>
<td>NO.</td>
<td>10</td>
</tr>
</tbody>
</table>

CALCULATED 'F' = 1.39058
PROBABILITY OF 'F' OCCURRING BY CHANCE = 30.533 %
AVG. ESTIMATE OF POPULATION VARIANCE = 2.27983
D.F. = 18
CALCULATED 't' = .133406

THIS IS A 2 SIDED TEST! THE TEST DIFFERENCE IS 0
THE 'NULL HYPOTHESIS (HO)' IS THAT:
THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE MEANS

THE PROBABILITY OF BEING WRONG IN REJECTING 'HO' IS 89.5352 %

This probability is NOT SIGNIFICANT! Therefore 'HO' is ACCEPTED.

Fig. 2.6 Results of a general "t" test used to compare the % Recovery of two analytical procedures at many concentration levels.
2.5.3 Uses of the Paired-Data "t" Test

The paired-data "t" test finds frequent use in analytical data analysis. Its usefulness lies in the fact that it allows the comparison of two analytical methods over their useful concentration range. Another advantage of the test is that the true concentrations of the samples need not be known for the comparison. The required data consists of two sets where the individual members of one set are directly related (paired) to corresponding members of the second set. A disadvantage is that the total degrees of freedom for the test is the number of data pairs minus one.

Examples

Three examples from the chemical literature will be used to demonstrate possible uses of this test. A fourth example can be found in Appendix A. The examples used here will also illustrate that other tests are sometimes used to analyse data of this type.

As Compare the results obtained with a new method(s) to those obtained with an established method, where both methods were used to analyze the same series of samples. This example is also from the paper by Abou Ouf et al. (38); two new methods are compared with an established method. Reprocessing of the data (fig. 2.7) confirmed the original conclusions that there was no significant difference between the results of the new methods and the established method. The data was also analyzed using a two-factor ANOVA test which confirmed the results (sec. 2.6).
Fig. 2.7 Results of two, paired-data "t" tests.

A complication in the use of this test could occur if the error variance of the methods is not constant but is actually a function of the concentration. This situation is seldom mentioned or illustrated in the analytical literature. The problem is perhaps easier understood when it is realized that the paired-data "t" test is actually a one-sample "t" test (eq. 6) where the sample is the set of "differences" and the known mean is "zero". The one-sample "t" test, and hence the paired-data test, requires the data set to be normally distributed; can the "differences" be normally distributed if the error variances of the methods are not constant? Chapter 4 discusses the problem of comparing methods which have nonconstant variance.

The data in this example (40) was first ordered (the first set was ordered from lowest to highest, the second set was ordered to the first). When the data was reprocessed by the program the resulting set of differences (Table 2.4) was examined to determine if there was any correlation between concentration and magnitude of the "difference". A linear regression analysis (sec. 2.7) showed no correlation between
the difference and the concentration level. A comparison of the sign of the differences does, however, tend to indicate a possibility that the first method gives slightly higher results than the second method at low concentrations, but lower results for high concentrations.

<table>
<thead>
<tr>
<th>Concentration</th>
<th>Difference</th>
<th>Concentration</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.88</td>
<td>+0.02</td>
<td>1.27</td>
<td>+0.04</td>
</tr>
<tr>
<td>0.95</td>
<td>+0.04</td>
<td>1.32</td>
<td>+0.04</td>
</tr>
<tr>
<td>0.98</td>
<td>+0.02</td>
<td>1.32</td>
<td>0.00</td>
</tr>
<tr>
<td>1.04</td>
<td>+0.07</td>
<td>1.41</td>
<td>+0.11</td>
</tr>
<tr>
<td>1.05</td>
<td>+0.01</td>
<td>1.45</td>
<td>+0.03</td>
</tr>
<tr>
<td>1.09</td>
<td>-0.06</td>
<td>1.46</td>
<td>0.00</td>
</tr>
<tr>
<td>1.11</td>
<td>-0.10</td>
<td>1.49</td>
<td>-0.07</td>
</tr>
<tr>
<td>1.14</td>
<td>-0.07</td>
<td>1.56</td>
<td>-0.03</td>
</tr>
<tr>
<td>1.18</td>
<td>0.00</td>
<td>1.57</td>
<td>-0.04</td>
</tr>
<tr>
<td>1.20</td>
<td>+0.01</td>
<td>1.67</td>
<td>-0.14</td>
</tr>
<tr>
<td>1.21</td>
<td>+0.02</td>
<td>1.67</td>
<td>-0.05</td>
</tr>
<tr>
<td>1.22</td>
<td>+0.12</td>
<td>1.89</td>
<td>+0.17</td>
</tr>
<tr>
<td>1.26</td>
<td>-0.05</td>
<td>1.91</td>
<td>-0.04</td>
</tr>
<tr>
<td>1.27</td>
<td>+0.02</td>
<td>1.95</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

The results of the "t" test (fig. 2.8) indicated that there was no significant difference between the results of the two methods. This was the same conclusion reached by the original authors. In the previous data analysis by Michel et al. (40) a "t" test was not used, rather the data was analyzed using linear regression. This was perhaps a misuse of the linear regression technique as will be discussed in chapter 4.
Fig. 2.8 Results of a paired-data "t" test where the data had originally (40) been analyzed using the linear regression technique.

C. Hirsch (16) in a paper dealing with ANOVA has discussed the use of the nonparametric Wilcoxon Signed-Ranks test for paired comparisons. Specifically, he has considered the comparison of results from a new method with those from a reference method of analysis (see his table VI). The computer program would be useful in performing such a test; it could be altered to do the test completely.

The data used by Hirsch was reprocessed in two ways. It was first ordered and then subjected to the paired-data "t" test. The test indicated that there was no significant difference in the results from the two methods (t=1.360, probability=19.47%). However, it can be seen (fig. 2.9A) that the differences between the large concentrations must have a larger effect in determining the average difference than the differences between the smaller concentrations.

The data was then normalized (fig. 2.9B) by dividing each member of a pair by the average concentration for the pair. A paired-data "t" test was then performed on the normalized concentration data. The test again indicated (fig. 2.9C) no significant difference between the results of the two methods (t=0.0604, probability=95.37%). However, now it can be seen (fig. 2.9-A) that the differences between the large concentrations do not have a determining effect on the average difference. The program indicated that the normalized data corresponding to the concentration pair, 7.21 and 5.62, were possible outliers, as was the corresponding difference.
No data treatment similar to the above has been found in the analytical literature surveyed, although Hirsch (16) has illustrated the use of normalization for ANOVA. If such a treatment is valid then it would have an advantage over the nonparametric Wilcoxon Signed-Ranks test where information is lost and the analysis is less complete.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>DIFF</th>
<th></th>
<th></th>
<th>DIFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.13</td>
<td>.15</td>
<td>-.02</td>
<td>.928571</td>
<td>1.07143</td>
<td>-.142859</td>
</tr>
<tr>
<td>.48</td>
<td>.52</td>
<td>-.04</td>
<td>.96</td>
<td>1.04</td>
<td>-.08</td>
</tr>
<tr>
<td>1.05</td>
<td>1.28</td>
<td>-.23</td>
<td>1.00455</td>
<td>.995455</td>
<td>.009033</td>
</tr>
<tr>
<td>1.11</td>
<td>1.1</td>
<td>.01</td>
<td>1.0153</td>
<td>.984576</td>
<td>.031004</td>
</tr>
<tr>
<td>1.31</td>
<td>1.27</td>
<td>.04</td>
<td>1.0153</td>
<td>.984576</td>
<td>.031004</td>
</tr>
<tr>
<td>4.82</td>
<td>4.98</td>
<td>-.16</td>
<td>1.02397</td>
<td>.976027</td>
<td>.047943</td>
</tr>
<tr>
<td>5.98</td>
<td>5.7</td>
<td>-.28</td>
<td>1.12393</td>
<td>.876072</td>
<td>.247959</td>
</tr>
<tr>
<td>7.21</td>
<td>5.62</td>
<td>1.59</td>
<td>1.03226</td>
<td>.967742</td>
<td>.064218</td>
</tr>
<tr>
<td>8</td>
<td>7.3</td>
<td>.5</td>
<td>1.01503</td>
<td>.984786</td>
<td>.030664</td>
</tr>
<tr>
<td>9.79</td>
<td>9.3</td>
<td>.29</td>
<td>1.01503</td>
<td>.984786</td>
<td>.030664</td>
</tr>
<tr>
<td>14</td>
<td>14.4</td>
<td>-.4</td>
<td>1.01408</td>
<td>.988164</td>
<td>.028164</td>
</tr>
<tr>
<td>17.2</td>
<td>18.4</td>
<td>-.12</td>
<td>.966292</td>
<td>1.03371</td>
<td>-.067418</td>
</tr>
<tr>
<td>19.1</td>
<td>19.1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>26.4</td>
<td>26.8</td>
<td>-.4</td>
<td>.992481</td>
<td>1.00752</td>
<td>-.015039</td>
</tr>
<tr>
<td>57</td>
<td>54.1</td>
<td>2.9</td>
<td>1.0261</td>
<td>.973897</td>
<td>.052203</td>
</tr>
<tr>
<td>74.1</td>
<td>70</td>
<td>4.1</td>
<td>1.02645</td>
<td>.971848</td>
<td>.056902</td>
</tr>
</tbody>
</table>

PAIRED-DATA "t" TEST

CALCULATED 't' = .060368

THIS IS A 2 SIDED TEST! THE TEST DIFFERENCE IS 0

C

THE "NULL HYPOTHESIS (HO)" IS THAT:
THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE MEANS
THE PROBABILITY OF BEING WRONG IN REJECTING 'HO' IS 95.2637 %
This probability is NOT SIGNIFICANT! Therefore 'HO' is ACCEPTED.

Fig. 2.9 Results of a paired-data "t" test where the data is first normalized. Actual (A) and normalized (B) concentrations for the two methods, and the differences between them. "t" test results (C) for the normalized data.
2.5.4 Uses of the "t" or "U" test for the Comparison of an Experimental Mean with a Known Mean

These two tests are actually conducted using the same computer routine. The only difference in the tests lies in whether the standard deviation for the experimentally determined mean (from a limited number of measurements) or the standard deviation of the known mean (from a very large set of measurements) is used.

The test could be used to determine the absolute accuracy of an analytical procedure where an accurately known standard is available or where known additions are made to a sample. If a series of standards covering the concentration range of interest were used then the accuracy of the procedure could be evaluated for that range.

In the situation where a standard procedure has been modified, but it is expected that the analytical error will not be affected, then the previously known standard deviation of the procedure would then be used. Otherwise, the standard deviation for the new data set is used.

Examples

One example from the chemical literature will be used to demonstrate the use of this test. A second example can be found in Appendix A.

A: Compare the results (concentration = 65.92) obtained using a proposed method (41) for the analysis of iron in an iron ore sample (BCS 175/2), with the certificate value (66.10%, range 66.05-66.20). Determine if the results from the proposed method could be considered to be 0.16% lower (an arbitrarily chosen difference).

Since in this example the standard had not originally been analyzed by the proposed method it is necessary to use the standard deviation obtained from the new analysis of the standard. The computer results are shown in figure 2.10. They show (A) that there is a difference between the mean concentration estimated by the method and the certificate value and that the difference is highly significant. The second "t" test (single sided - fig. 2.10B) further showed
that the results from the proposed method were at least 0.16% lower than the certificate value. These results indicate that either the certificate value or the results from the new method are incorrect.

Many examples of the use of this test or data which could be tested by it can be found in the Analyst. Spanyar et al. (42) has compared the amount of capsaicin added to capsaicin-free paprika, to the amount recovered. Burns et al. (43) have compared the determined concentration of nitrogen, carbon, and hydrogen in acetanilide to the theoretical concentrations.

---

<table>
<thead>
<tr>
<th></th>
<th>SET 1</th>
<th>KNOWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>65.917</td>
<td>66.1</td>
</tr>
<tr>
<td>STD. Dev.</td>
<td>.0149443</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Test Variance</td>
<td>2.23333E-04</td>
<td></td>
</tr>
</tbody>
</table>

CALCULATED 't' = 38.7232
PROBABILITY OF 't' OCCURRING BY CHANCE = 0%

A

THE 'NULL HYPOTHESIS (HO)' IS THAT:
THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE MEANS

THE PROBABILITY OF BEING WRONG IN REJECTING 'HO' IS 0%

This probability is

VERY HIGHLY SIGNIFICANT(!) 'HO' can be REJECTED and it is very unlikely that this conclusion is incorrect.

---

<table>
<thead>
<tr>
<th></th>
<th>SET 1</th>
<th>KNOWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>65.917</td>
<td>66.1</td>
</tr>
<tr>
<td>STD. Dev.</td>
<td>.0149443</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Test Variance</td>
<td>2.23333E-04</td>
<td></td>
</tr>
</tbody>
</table>

CALCULATED 't' = 4.8666
PROBABILITY OF 't' OCCURRING BY CHANCE = .0443339%

B

THE 'NULL HYPOTHESIS (HO)' IS THAT:
FIRST MEAN IS 'NOT' SIGNIFICANTLY SMALLER THAN SECOND MEAN MINUS .16

THE PROBABILITY OF BEING WRONG IN REJECTING 'HO' IS .0443339%

This probability is

VERY HIGHLY SIGNIFICANT(!) 'HO' can be REJECTED and it is very unlikely that this conclusion is incorrect.

Fig. 2.10 Results of a "t" test where an experimentally determined mean is compared to a known mean, where there is no known standard deviation. A: Tested for any difference; B: tested to determine if first mean is at least 0.16% less than the second mean.
2.6 Analysis of Variance

Hirsh (16) reported in 1977 that ANOVA is a well-established technique, but it seemed to be used rarely by chemists. Lack of use was probably due in part to the amount of time required for the computations using a simple calculator. Hirsh, in the same paper, also expounded on the choice of the proper model, interpretation, and the advantages and limitations of the technique.

ANOVA is used reasonably frequently in papers in the Analyst (44-48). But in agreement with Hirsch, the applications are mostly for interlaboratory testing of methods and testing for significance in regression analysis. The use of ANOVA with linear regression is covered in chapter 3; it will not be discussed here.

In addition to the limited use of ANOVA in the chemical literature the technique is covered to varying degrees in statistics texts directed towards the engineering and scientific user. For example, Kennedy and Neville (10), in a text which is quite good overall, devotes only one very short chapter to the topic, whereas, Wine (49) covers the topic extensively.

The discussion which follows draws mainly on four sources; the papers by Hirsch (16) from Analytical Chemistry and by Evans (23) from the Analyst, and the texts by Brookes et al. (18) and by Davies and Goldsmith (27). The object of the discussion will be; to explain the workings of the simplest ANOVA method (which can then be called upon in considering more complicated cases), to produce a brief but comprehensive picture of some of the various methods available, and to discuss the appropriate application of the methods available in the program.

2.6.1 One-Factor Analysis of Variance

The total variance of a measuring system is the sum of the component variances of each of the independent sources of variation included in the overall system (10, p. 319). ANOVA is a process in which this total variance is broken down into "estimates" of the variances for each of the sources of variation. When these estimates of variance (or
combinations of variances) are obtained they can be compared using an "F" test, i.e., the relative importance of each variance can be assessed. In addition, estimates of the standard deviations for each source can be calculated, as well as, their confidence limits.

The one-factor analysis of variance has been explained quite well by Hirsch. In addition it is discussed in almost every statistics text. Table 2.5 shows the typical data layout. Table 2.6, a typical one-factor ANOVA table, will probably outline the procedure adequately for the purpose at hand.

It can be seen from Table 2.6 that the calculated mean square for the total error [MSt=SSt/(nk-1)] is actually the total variance (Vt) for the complete data set, i.e., this is the variance which would be calculated if all the data had been considered to be one set of measurements. The parameters for - Among Factor "a" - and - Residual Error - should be considered with this in mind.

The residual error mean square (SSe/k(n-1)) is an estimate of the error variance. It is assumed that the error variance is the same for each set of replicate measurements; the variance is estimated by using all of the data sets.

The variance among the factors (i.e., different methods) is actually the sum of the variance due to the factor (Va), plus the error variance of each data set (Vo/n). The calculated variance using the factor means is therefore Vo/n +Va. The null hypothesis is that Va will be zero; if this is correct then the calculated variance will be Vo/n. Hence, multiplying the calculated variance by n (assuming the null hypothesis is correct) gives Vo, therefore a second estimate of the error variance has been determined. A comparison of these two estimates of the error variance [(Vo+nVa)/Vo] using an F test (sec. 2.5.1) allows the null hypothesis to be checked.

A problem sometimes rises in the interpretation of the "Mean Squares". In the case of the Total Error (MSt) and the Residual Error [MSe=SSe/k(n-1)] they represent a variance. For Factor "a" the mean square (MSa), however, represents a combination of variances [Vo + nVa] as can be seen in the
table. These mean squares are often misinterpreted (e.g., 50, Table II) as being variances. The estimate of the variance for Factor "a" can be easily calculated as is shown at the bottom of Table 2.6.

A comparison of the mean squares of the sources of variation allows an evaluation of the significance of the variance of each of the factors. For the above example:

\[
F = \frac{\bar{V}_o + nV_a}{\bar{V}_o}
\]  

If the experimental "F" is greater than the allowable critical value for "F", then the variance estimate for the main factor "V_a" is significant, i.e., it is greater than "0".
### TABLE 2.5
DATA LAYOUT FOR ONE-FACTOR ANOVA

<table>
<thead>
<tr>
<th>REPLICATE</th>
<th>FACTOR &quot;a&quot; (j)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Y11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y1k</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Y22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>Y33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yij</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>Yn1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ynk</td>
</tr>
</tbody>
</table>

Column totals: Y1 ... Yk

### TABLE 2.6
ONE-FACTOR ANALYSIS OF VARIANCE TABLE

<table>
<thead>
<tr>
<th>SOURCE OF VARIATION</th>
<th>DEGREES OF FREEDOM</th>
<th>SUM OF SQUARES</th>
<th>MEAN SQUARE</th>
<th>F RATIO</th>
<th>EXPECTATION OF THE MEAN SQUARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among Factor &quot;a&quot;</td>
<td>k-1</td>
<td>Σ(Yj-Y)^2</td>
<td>SSa/(k-1)</td>
<td>MSa/MSe</td>
<td>Vo + nVa</td>
</tr>
<tr>
<td>Residual Error</td>
<td>k(n-1)</td>
<td>Σ(Yij-Yj)^2</td>
<td>SSa/(k(n-1))</td>
<td>Vo</td>
<td></td>
</tr>
<tr>
<td>Total Error</td>
<td>nk-1</td>
<td>Σ(Yij-Y)^2</td>
<td>SSa/(nk-1)</td>
<td>Vo</td>
<td></td>
</tr>
</tbody>
</table>

and Va = (MSa - Vo) / n
Thus, it can be seen that an ANOVA facilitates the calculation of an estimate for the identifiable sources of variation and allows their significance to be calculated. This program computes all of these estimates, plus, the standard deviations for each source of variation. It also computes "95% confidence limits" of the standard deviation for the residual error. Most applications of ANOVA stop after "F" tests on the mean squares.

It should be emphasized again that the calculated standard deviations for each source of variation are only "estimates" of the true values. The calculation of confidence limits for these estimates (using Chi-squared and "F") quite clearly shows their wide range (27,49), particularly for small degrees of freedom.

2.6.2 A Comparison of One-, Two-, and Two-Factor with Replication, Plus Cross-Classification Vs. Hierarchic Factors

The text by Brookes et al. (18) describes and gives examples of one-, two-, and two-factor ANOVA including the case of hierarchic (nested) factors. However, it does not consider the calculation of the factor variances in the same detail as Davies and Goldsmith (27). These latter authors also discuss, in detail, the calculation of confidence limits for the factor variances. Evans (23) has further illustrated these points and their value in analytical chemistry. He also lists four specific uses of ANOVA; the first three have already been discussed. The fourth deals with the significance of differences in mean values for each source of variation. In addition to the discussion by Evans, literature examples (18, p. 333; 46, Tables 5 and 6) have also been given.

A further discussion of ANOVA is not possible within the limitations of this thesis, nor is it necessary as the literature already covers it adequately. What the literature does seem to lack, however, is a simple comparison of the methods already discussed. Table 2.7 shows such a comparison. These methods will be further illustrated in section 2.6.4 with the use of examples.
<table>
<thead>
<tr>
<th>SOURCE OF VARIATION</th>
<th>DEGREES OF FREEDOM</th>
<th>MEAN SQUARE</th>
<th>( F ) RATIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{within } a )'s</td>
<td>((n-1)(k-1))</td>
<td>error ( k-1 )</td>
<td></td>
</tr>
<tr>
<td>( \text{within } b )'s</td>
<td>((k-1)(n-1))</td>
<td>error ( k-1 )</td>
<td></td>
</tr>
<tr>
<td>( \text{interaction } a \times b )'s</td>
<td>((n-1)(k-1))</td>
<td>error ( k-1 )</td>
<td></td>
</tr>
</tbody>
</table>

In a Cross-Classification ANOVA, the two main sources of experimental error variance are not real variances, but are instead variances within each possible combination of the population (finite). Mixed-model factors are also possible.

In a Hierarchical ANOVA, the two main sources of experimental error variance are not real variances, but are instead variances within each possible combination of the population (finite).

In a Random Error Factors (Model I) or Fixed Error Factors (Model II) also called One-Factor or One Source of Variation of the Mean Square, the number of replicates per cell is 1.
2.6.3 Assumptions and Limitations of ANOVA

Hirsch (16) has discussed this topic in some detail. His discussion centers around:
- random sampling
- independence of random errors
- homogeneous sample variances
- Gaussian distribution of random errors

These factors have already been discussed; they will be referred to again in later chapters. As has already been seen, statistical tests cannot be correctly used unless the data complies with the basic assumptions required by the test.

2.6.4 Examples

Many literature examples (e.g., 16,18,23,27) have been reprocessed by the program with satisfactory agreement. The examples used here have been selected to allow a comparison between computed and published results.

An experiment, for which the data will be subjected to an ANOVA, will usually be designed with the requirements of the data analysis in mind. However, it is possible that certain tests within the experiment won't work, e.g., a sample is accidently lost. This may result in an "incomplete" data set. A limitation of this program is that it will not handle this type of data (23, Table IV).

Before proceeding with the examples it is necessary to consider the usual procedure (16) involved in the use of ANOVA of experimental data. The first two points must be considered BEFORE collecting the data:
- identify the factor(s) to be studied
- design and conduct the experiment
- conduct the ANOVA
- interpret the results of the ANOVA.
A: This data was obtained from an experiment using a rapid, fully automated atomic absorption analysis system and the Delves Cup technique. The potentially high precision of the system could only be obtained if reproducible results could be obtained from one sample cup to the next.

The problem was to compare the variances of the variation, due to the use of different sample cups, with the experimental error variance (all other sources of variation pooled). In this experiment the sample cups were randomly selected from a large population of similar cups (Model II). Each cup was then used to collect nine measurements of the absorption peak for aliquots of a standard lead solution.

The data was analyzed using a two-factor hierarchic ANOVA without replication. The computer results are displayed in figure 2.11; they are similar to those by Hirsch (16), but not identical. After extensive testing it would appear that the computer results are more correct. However, as can be seen the same conclusion is reached by each analysis; the difference in cups is not significant at the 95% confidence level, although the difference is possibly significant (probability of "F" occurring is 8.4%).

In this data analysis the two factors were sampling and testing. The actual calculations performed are identical to a one-factor (Model I) ANOVA. The only difference is one of interpretation of the results. In a Model I analysis the cups would have been selected from a finite population of cups which were known to be different in some manner from each other.

An example (11) by Evan’s (23) is similar to this one, however, there is an error in his example as the mean is 12.825 and not 12.87.

This technique would be useful in evaluating a method in which linear calibration curves (chapter 3) are used. The variance of measurements due to sample preparation could be compared to the variance of testing, to determine if sample preparation resulted in a larger error than replicate readings of the measuring device.
Fig. 2.11 A two-factor, hierarchic, analysis of variance.
A: Data matrix, sum of rows and columns.
B: ANOVA table.
C: Standard deviations of each source of variation.
B: A hierarchic ANOVA with replicate measurements. In this experiment ten delivery batches of a chemical paste were selected at random. Three casks were then selected at random from each batch. Each cask was tested twice. This example could be illustrative of the answers required from a quality control laboratory for determining the consistency of a product.

The computer results, figure 2.12, agree with those of Davies and Goldsmith (27). This particular example is also considered by Wine (49). At this time the program cannot generate critical "F" values, therefore confidence limits for factors "a" and "b" are not calculated. However, the necessary expressions (27,49) are as follows (using the symbols of Table 2.5):

95% Confidence Limits for the Batches (factor a).

\[ \sqrt{\frac{\text{MS}_a}{\text{F}(k-1),k(n-1)} - \frac{\text{MS}_b}{\text{mn}}} \text{ minimum limit} \quad (10) \]

\[ \sqrt{\frac{\text{MS}_a \times \text{F}(k(n-1),(k-1)) - \text{MS}_b}{\text{mn}}} \text{ maximum limit} \quad (11) \]

95% Confidence Limits for the Casks (factor b).

\[ \sqrt{\frac{\text{MS}_b}{\text{F}(k(n-1),kn(m-1)) - \text{MSe}} / m} \text{ minimum limit} \quad (12) \]

\[ \sqrt{\frac{\text{MS}_b \times \text{F}(kn(m-1),k(n-1)) - \text{MSe}} / m} \text{ maximum limit} \quad (13) \]

if \( \alpha = 0.025 \) for the critical "F" values, then the 95% confidence limits for:

SD of factor "a" : 0.00 - 3.73

SD of factor "b" : 1.91 - 4.50
Fig. 2.12 A two-factor, hierarchic (model II), ANOVA with replication.

A: Data matrix, sum of rows and columns.  
B: ANOVA table.  
C: Standard deviation of each source of variation.
This example illustrates a two-factor cross-classification ANOVA with replicate measurements. It involves a comparison of three specific analytical methods using four different analysts. Assuming there were only three methods and four analysts available for the experiment makes it a Model I case. In this experiment it is possible to have an interaction, e.g., one analyst is particularly accurate with one method. The computer results, figure 2.13, are in agreement with those of Brookes et al. (18). The variance (not a true variance) due to the three different methods is significant, there is no interaction, and the differences between the performances of the analysts is not significant.

It can be seen that the standard deviation of the experimental error, due to errors of replication, is thought to be between 0.00804 and 0.0185 (mean standard deviation = 0.01118). Using the sums of the columns (fig. 2.13A) allows the average result for each method to be calculated. The 95% confidence intervals for each method are: (1) 0.156-0.174, (2) 0.124-0.142, (3) 0.184-0.202. They do not overlap, thus the results of the ANOVA indicating a difference between the methods is confirmed.

A similar example can be found in the Analyst (46). This data analysis results from a statistically designed collaborative test carried out in six laboratories. Two different feed samples were used. Each sample was analyzed in triplicate. The resulting data consisted of "% recovery" measurements.

The computed and the published data again agree. An added advantage of the computed results is that the standard deviation (and its confidence limits) are computed; they are 5.7222 and 4.477 to 7.978. These results are shown in figure 2.14. This example also received extensive after (ANOVA) treatment in the paper. Similar treatment was also performed using this program. Figure 2.15 shows a small part of this data treatment.
DATA MATRIX

<table>
<thead>
<tr>
<th></th>
<th>.16</th>
<th>.14</th>
<th>.19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.16</td>
<td>.18</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>.17</td>
<td>.12</td>
<td>.2</td>
</tr>
<tr>
<td></td>
<td>.17</td>
<td>.14</td>
<td>.2</td>
</tr>
<tr>
<td></td>
<td>.16</td>
<td>.11</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>.10</td>
<td>.13</td>
<td>.2</td>
</tr>
<tr>
<td></td>
<td>.16</td>
<td>.15</td>
<td>.2</td>
</tr>
<tr>
<td></td>
<td>.17</td>
<td>.12</td>
<td>.18</td>
</tr>
</tbody>
</table>

SUM OF ROWS

<table>
<thead>
<tr>
<th></th>
<th>.97</th>
<th>1</th>
<th>.97</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.97</td>
<td>1.06</td>
<td>.98</td>
</tr>
</tbody>
</table>

SUM OF COLUMNS

1.32 1.06 1.54

A

B

<table>
<thead>
<tr>
<th>SOURCE OF VARIATION</th>
<th>D. F.</th>
<th>B. B.</th>
<th>M. S.</th>
<th>F RATIO</th>
<th>PROBABILITY (X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BETWEEN METHODS</td>
<td>2</td>
<td>.0144333</td>
<td>7.21667E-03</td>
<td>57.7333</td>
<td>9.99999E-04 B##</td>
</tr>
<tr>
<td>BETWEEN ANALYSTS</td>
<td>3</td>
<td>.0001</td>
<td>3.33333E-05</td>
<td>266666</td>
<td>65.1979 N.S.</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>6</td>
<td>.0011</td>
<td>1.83333E-04</td>
<td>146666</td>
<td>28.865 N.S.</td>
</tr>
<tr>
<td>ERROR</td>
<td>12</td>
<td>.0015</td>
<td>.000125</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>23</td>
<td>.0171333</td>
<td>7.4492E-04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C

STANDARD DEVIATION OF ca. CONFIDENCE LIMITS (95%)

<table>
<thead>
<tr>
<th>METHODS</th>
<th>.0297735</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANALYSTS</td>
<td>0</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>.0054006</td>
</tr>
<tr>
<td>TESTING ERROR</td>
<td>.0111803</td>
</tr>
<tr>
<td>TOTAL ERROR</td>
<td>.0272934</td>
</tr>
</tbody>
</table>

Fig. 2.13 A two-factor, cross-classification, ANOVA with replication (18).

A: Data matrix, sum of rows and columns.
B: ANOVA table.
C: "Standard deviations" of each source of variation.
Fig. 2.14 A two-factor, cross-classification, ANOVA with replication (46).

A: Data matrix, sum of rows and columns.

B: ANOVA table.

C: "Standard deviations" of each source of variation.
2.7 Linear Regression Analysis

Linear regression is used extensively in analytical chemistry. Many uses of the technique are related to calibration curves and the comparison of two methods of analysis. These topics are covered in chapters 3 and 4 respectively.

The application of the technique is, however, often amateurish. Sometimes the regression coefficient is used rather than an ANOVA in evaluating the "goodness of fit" of the calibrated line. Another common error is fitting two sets of data, both of which are subject to error, using a
routine which assumes that one data set is known exactly. A third error is fitting a line to data, for which the error variance is proportional to the magnitude of the measurement, using a model which assumes a constant error variance.

These common misuses again emphasize the importance of considering the basic assumptions and limitations of statistical tests prior to using them for data analysis.

Chapters 3 and 4 contain additional discussion of this very important technique.

2.8 Summary

This chapter describes many of the basic statistical tests available to the practising analyst. The tests are illustrated using examples from the analytical literature, and comparisons are drawn between the computed and the reported results.

The usefulness of the computer program is well illustrated throughout the chapter. In particular, the statistics software is shown to be accurate, versatile, and capable of producing very comprehensive results. It is also sufficiently fast to allow reprocessing of the data within minutes.

The tests and the examples considered do not include all the various statistical techniques used by analysts. However, they include the simpler and more commonly used ones. The use of these tests, on a routine basis, should increase both the awareness of the analyst to the usefulness of statistics, and the accuracy of his routine analytical work.

The "blind" application of statistics can, however, result in unsound conclusions, particularly if the data used do not satisfy the assumptions of the test. Such misapplications are now much more likely to occur because of the ready availability of inexpensive computers and software. These assumptions are a major consideration of this chapter.
Chapter 3
ANALYTICAL CALIBRATION CURVES
BY LEAST SQUARES ANALYSIS

3.1 Introduction

The "Method of Least-Squares" for the calculation of linear calibration functions is now common. However, the extended use of this method for the determination of confidence limits for the calculated analyte concentrations is rare. This is probably due to the lack of an easy to understand and comprehensive description of the mathematics involved, the "odd" confidence limits presented in elementary statistics texts, and the lack of an inexpensive and readily available "tool" for doing the calculations.

This chapter deals with the results from an "extended linear least squares" analysis program, although, nonlinear analytical functions are also considered.

Why is the ruler-and-eyeball method of obtaining calibration curves being replaced by the calculation of calibration functions? It is firstly because the computing devices and programs are now available to do the calculations quickly and accurately. Secondly, it is probably because it is thought that the technique will result in greater accuracy and extra information about the calibration curve. Like any analytical technique, however, it is open to misuse, misunderstanding, and underuse. This chapter considers all these factors, however, the emphasis will be on getting more out of the technique.

Meyer (1) has recently (September, 1982) stated that "the only valid excuse a chemist can offer for not doing a completely rigorous job of curve fitting is the unavailability of a computer". This statement is easily agreed with. His comments on the simplicity of writing one's own program though, is open to discussion. The opposing statement would be that many if not most chemists would prefer to use a "canned" program. However, as Meyer implies, the use of an inappropriate "canned" program could be time consuming and could produce unreliable results.
The remainder of this chapter is devoted to what is believed to be an "appropriate" program for the practising analyst. It will be shown that the resulting weighted linear regression program is a powerful computing "tool" for use with linear calibration curves. Chapter 4 will show additional, and equally useful, applications of the program.

3.1.1 Method of Least Squares

The history of the Method of Least Squares dates from 1805; it has been reviewed by Eisenhart (2) and others (refer to the references by Draper (3, p. 11). Although the method is discussed in most elementary statistics textbooks, it is not covered to the extent required for its practical application in the analytical laboratory. This shortcoming is somewhat made up for by papers which have appeared in the chemical education (e.g., 4-14) and the analytical (e.g., 15-24) literature. Developments in linear regression methodology from 1959 to 1982 have been reviewed in Technometrics; refer to papers 11-16 in Appendix B.

A more comprehensive presentation of the technique for the analyst is still required; it is unfortunate that the most recent volume of Comprehensive Analytical Chemistry (25) did not pursue this technique more extensively in the chapter on mathematical statistics in analytical chemistry. A very useful worked example is available (26), however, it does not include weighting factors.

3.1.2 Weighting Factors in Regression Analysis

Weighting factors are the key to the successful use of linear regression analysis in analytical chemistry. Unfortunately many of the texts (e.g., 27,28) which do discuss linear regression in an easy enough manner for the nonmathematically inclined chemist to follow, don't pursue the use of these factors. Generally linear regression is explained and demonstrated with examples in which the absolute response error is assumed to be constant, i.e., the "homoscedastic" or uniform variance condition. However, real calibration curves often have a non-uniform response error,
i.e., the "heteroscedastic" or nonuniform variance condition (19). A common situation is one in which the relative standard deviations are approximately equal, i.e., the error variance of the response is approximately proportional to the square of the response.

Discussions on weighting factors can be found in many of the above references (3-7,10-12,18,20,22-24) and in the texts by Draper (3), Brownlee (29), Neter and Wasserman (30), Miller (31), and Mandel (32). These authors also develop some of the equations necessary for handling weighting factors, in addition to those required to determine concentration limits.

The concentration limits computed using weighting factors are probably more acceptable to the analyst than those produced by unweighted methods. These limits could be considered as estimates of the analytical error. The first few chapters of the text by Acton (33) may be of interest to the analyst wishing to clear up some of his thoughts on the Method of Least Squares.

3.2 Theory

Much of the theory required for the development of a computer program for simple linear regression can be found in statistics texts. Therefore, only an overview of the theory and formulae will be given. This overview will attempt to explain as simply as possible the various "estimates" of analytically useful "numbers" which can be calculated for a linear calibration curve.

The analyst should not let the "wizardry" of either the statistics or the computer overwhelm his interpretative skills. Both statistics and computers produce seemingly exact numbers; however, these numbers are at best only estimates of the "true" values sought. The analyst need only repeat the experiment and the statistics to obtain new estimates of these predictions. However, statistics and the computer can give the analyst a better idea of the reliability of his results; this can then be used in his overall assessment of the experiment.
This overview is not a repeat of textbook theory, but will instead be one analyst's interpretation of such theory. This interpretation will hopefully be useful to any chemist who is using a computer program similar to the one to be described in this chapter. What the overview lacks in statistical exactness will hopefully be made up for by its easy interpretation and general coverage of the topic.

A difficult section of the theory involves the determination of confidence limits for "new" responses, e.g., what will be the confidence limits of the average response for "one" more standard which was not used in producing the calibration curve or for "many" more standards. Related to this are perhaps the most important "estimates" which can be produced, i.e., the confidence limits of "one" or "more" concentrations calculated using the calibration curve. More than one statistical interpretation of this situation seems to be available; a "useful" formula by Miller (31) will be used.

This overview will cover both the unweighted and the more generally useful weighted linear calibration curve. Most of the theory is taken from Brownlee (29, Ch. 11) and Miller (31, Ch. 3), although Mandel (32, Ch. 12) is a recommended reference. This should make it easier for anyone wishing to check a point to do so. It should be noted that some symbols have been changed (refer to Appendix C for a list of special symbols used).

3.2.1 The Unweighted Calibration Curve

To produce a calibration curve the analyst will usually prepare a series of standards and then measure the desired response of each standard. Using the eyeball and ruler method the "best" straight line can then be drawn through the points.

Appendix D outlines a short calibration experiment conducted to compare results of the "eyeball and ruler" technique to those of the least squares technique. The experiment showed that a variety of calibration lines were obtained when the "eyeball and ruler" technique was used by a group of technical people; the calibration lines indicated that an attempt was often made to weight the calibration points.
The points \((x_i, y_i)\), where \(y_i\) is the response for concentration \(x_i\), can also be used to determine the linear equation of the calibration curve, i.e., the analytical function. A two parameter line is used since the x-intercept is not assumed to be zero; the literature (34) suggests that a background signal often exists. This line is determined using the least squares technique with which most chemists are at least somewhat familiar.

The Equation of the Calibration Curve

The equation of the two-parameter line is:

\[
Y = \bar{y} + b (x - \bar{x}) \quad \text{or the form} \quad Y = a + b x
\]

where \(a\) is the x-intercept, \(b\) is the slope, and \(Y\) is the calculated response for a given \(x\).

It is assumed that \(y\) is normally distributed about an expected value of \(\bar{y}\) with a constant standard deviation of \(\sigma\), and that all observations are independent.

The estimates for the "true" values for \(\bar{y}\) and \(b\) are:

\[
\bar{y} = \frac{\Sigma y_i}{k} \quad (k = \text{the number of data points}) \quad (2)
\]

\[
b = \frac{\Sigma (x_i - \bar{x}) y_i}{\Sigma (x_i - \bar{x})^2} \quad (3)
\]

From experience the analyst knows that the use of additional standards would change the equation of the line; the equation produced is therefore just an estimate of the true equation. In what way could the new line differ from the old? The average value of \(y\) (the average response) could be larger or smaller; this would move the line up or down (slope remains the same). Also the slope of the line could change, even though the line remained fixed at \((\bar{x}, \bar{y})\).

Most likely both the average value of \(y\) and the slope would change. Knowing this a picture of the region in which the "true" line lies could be drawn. Figure 3.1 shows the estimated line, the movement of the line due to a change in the average \(y\), the movement due to a change in slope, and the combined effect of both movements.
This combined effect produces a picture somewhat similar to the unweighted confidence limits shown in most statistics texts. The confidence limits, or bands, for the line are actually two branches of a hyperbola (eq. 14) situated symmetrically with respect to the line. The calculation of these bands are discussed later.

Fig. 3.1 Unweighted and weighted calibration curves with approximate confidence bands. The figures represent possible movements of the curve due to changes in the: A - mean response, B - slope, C - slope and mean response. The dashed lines in "A" and "B" represent extreme positions of the calibration curve, while in "C" they represent the confidence bands of the curve.
The Variance Estimates of $\bar{y}$ and $b$

Statistics allows the variance of both $\bar{y}$ and $b$ to be estimated:

\[ V[\bar{y}] = \frac{s^2}{k} \]
\[ V[b] = \frac{s^2}{\Sigma(x_i - \bar{x})^2} \]

where: $s^2$ is the variance of the $y_i$'s about the line; it can be estimated by:

\[ s^2 = \Sigma(y_i - \bar{y})^2 / (k-2) \]

This standard deviation ($s$) is an estimate of the constant standard deviation assumed for the response.

The Variance of a Response

The variance of a response "on the calibration line" for any one fixed value of $x$ can be calculated. This variance, as has been seen already from figure 3.1, will be a combination of the variance of the mean response ($\bar{y}$) and the variance of the slope ($b$). The variance of the response ($V$), due to the slope, is not constant but actually increases as the concentration ($x$) moves further away from the average $x$ ($\bar{x}$):

\[ V[y] = V[\bar{y}] + (x-\bar{x})^2 V[b] \]
\[ V[Y] = s^2 \left\{ \frac{1}{k} + (x-\bar{x})^2 / \Sigma(x_i - \bar{x})^2 \right\} \]

The above variance is different from the variance of a single "new or predicted" response at a selected $x$, i.e., the expected variance of the response of a new standard. It would be the sum of the variance expected for a response ($s^2$), plus the variance of the line (eq. 7) at that particular value of $x$:

\[ V[Y'] = s^2 + V[Y] = s^2 \left\{ 1 + \frac{1}{k} + (x-\bar{x})^2 / \Sigma(x_i - \bar{x})^2 \right\} \]

These variances have limited direct usefulness, but they will assist in following the development of confidence limits for calculated concentrations. An example of one use is illustrated in Appendix E and figure 3.3.
The Analysis of Variance Table

When ANOVA was discussed in section 2.6, it was mentioned that it was often used in testing for significance in regression analysis. The purpose of ANOVA will again be to estimate the individual sources of variance, and to test their significance.

In linear regression analysis there is the total variance of the response measurements, the variance of the residuals, and the variance explained by the regression line. The total sum of squares will equal the sum of squares of the residuals, plus the sum of squares explained by the regression line.

The total sum of squares, $\sum(y_i-\bar{y})^2$, can be obtained by summing the squares of the differences between each measured response and the average measured response. This calculation is equivalent to assuming that the responses are not dependent on the concentration.

However, if the responses are dependent on the concentration ($x$), then fitting the data to a straight line will remove much of the apparent experimental error. The appropriate sum of squares is $\sum(Y_i-\bar{y})^2$, although this quantity is usually calculated differently.

The sum of squares of the residuals is $\sum(y_i-Y_i)^2$. However, it is usually found by finding the difference between the two previous sum of squares.

This information can be used in a number of ways. A comparison of the ratio of "the mean square, explained by the line" to the "mean square of the residuals", and "the critical F ratio (using the appropriate degrees of freedom)" allows an assessment as to whether the calibration line is well represented by the estimated equation.

The coefficient of linear correlation (i.e., the correlation coefficient) can also be calculated. This coefficient is often quoted in the literature as a measure of the accuracy of the "fit" of the analytical calibration line. However, VanArendonk and Skogerboe (22) have pointed out that the coefficient is not an effective statistic for this purpose, and that the practice should be discouraged. Using this coefficient, the percentage of the overall
variance explained by the equation of the calibration line can also be calculated (35).

Section 3.3.3 and figure 3.6 illustrate the usefulness of ANOVA in linear regression analysis.

An interesting outcome of this analysis is that once the variation due to linear regression has been removed, we can consider the remaining variation to be due to the experimental error of measuring the response. Thus, the variance about the line can be used to estimate the standard deviation of any response measurement in the range being examined; this standard deviation is called the standard error of regression. It is actually the standard deviation of the "residuals", where a residual is the difference between the measured response and the corresponding calculated response.

Confidence limits for the responses at single concentrations can be calculated in the regular fashion (eq. 9), but instead of using just the replicates of that measurement to estimate the standard deviation we can use the standard error of regression. The standard error of regression will probably be a much better estimate of the "true" standard deviation of the response. It will be shown (ch. 4) that the standard error of regression is an important factor in estimating the analytical error of an analytical method.

\[
CL = Y +/− t \cdot s / (m)^{.5}
\]  

(9)

where \( m \) = number of replicates
and the number of degrees of freedom for \( t \) is the total number of response measurements used in determining the line minus 2.
Confidence Limits

The confidence limits of the slope, any point on the calibration curve, the x-intercept, and the response for any "one" future calibration standard are:

the slope:
\[ b \pm t \sqrt{\frac{V[b]}{\sum(x_i - \bar{x})^2}} \]  
\[ = b \pm t s 
\] (10)

any single point on the calibration curve:
\[ y \pm t s \left( \frac{1/k + (x - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right)^{.5} \]  
\[ = Y \pm t s \left( \frac{1/k + (x - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right)^{.5} \] (11)

the x-intercept:
\[ (y - bx) \pm t \sqrt{\frac{\sum(x_i - \bar{x})^2}{(1/k + (x - \bar{x})^2)}} \]  
\[ = (y - bx) \pm t s \left( \frac{1/k + (x - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right)^{.5} \] (12)

response for any "one" new calibration standard:
\[ = Y \pm t s \left( 1 + \frac{1/k + (x - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right)^{.5} \] (13)

Confidence Band about the Calibration Line

The calculation of an upper and lower confidence limit for a single point on the line has already been discussed. Of more practical value is the determination of upper and lower confidence limits for any number of points, i.e., banding the entire line (32). Working and Hotelling (36) solved this problem in 1929; simply replace the critical value of \( t \) in equation 11 with \( (2F)^{.5} \), where \( F \) has 2 and \( k-2 \) degrees of freedom (\( k \) is the number of data points).

The familiar hyperbolic curves above and below the calibration line are then available. Using a 95% confidence limit we would expect at least 95% of all bands calculated would contain the true calibration curve.

\[ CL(y) = Y \pm (2F)^{.5} s \left( \frac{1/k + (x - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right)^{.5} \] (14)
Simultaneous and Non-simultaneous "Confidence" Limits

A calibration curve can be used to "predict" the response which would be expected for a single standard of known concentration. This can be referred to as a non-simultaneous use of the calibration curve. The appropriate equation (eq. 13), used to determine confidence limits for this value, has already been developed. However, if an unlimited number of "predictions" were to be made using the same curve then this equation would no longer be valid.

Miller (31) has developed "confidence" limits for this latter case which he has called "simultaneous tolerance" limits; the limits for one prediction are called simply "prediction" limits.

A calibration curve could also be used to calculate the concentration of one sample of unknown concentration (the non-simultaneous case), or the concentration of an unlimited number of samples (the simultaneous case). Again the equations would differ depending upon whether one or many samples were being used with the same calibration curve.

The calculation of a concentration (x-value) from the calibration curve is referred to by statisticians as "discrimination". Discrimination is therefore the reverse of prediction. Miller (31), and Lieberman, et al. (37) refer to these confidence limits respectively as "non-simultaneous" and "unlimited simultaneous" discrimination limits.

These statistical terms will be used in the following discussion to increase its clarity.

Tolerance Intervals of Many Future Responses

Equation 13 can only be used for the calculation of the confidence or "prediction" interval of a "single" future response. There are also formulae available for the calculation of the prediction intervals of 2, 3, etc. future responses. However, a formula for the "tolerance" interval of any number of future responses is probably more useful.

The analyst may not actually wish to calculate future response intervals, but the calculation is never-the-less
necessary for the determination of concentration "discrimination" intervals. A detailed derivation of the necessary equations is given by Miller (31) and Lieberman et al. (37).

In general, the uncertainty in the calibration line at the given concentration (eq. 14) is added to the uncertainty in the sample response.

Garden et al. (21) have explained for the chemist how Miller's (31) expression for the uncertainty in the sample response, when many samples will be used, can be estimated. They have given the band on the sample response as:

\[ CL(y) = Y \pm Z \times \sqrt{k-2}/\chi^2 \times s/\sqrt{m} \]  \hspace{1cm} (15)

where

- \( m \) = the number of responses for the sample
- \( Z \) = the appropriate value of the normal distribution
- \( \chi^2 \) = the appropriate value of the chi-squares distribution.

The equation used to calculate the "tolerance limits" in the computer program is discussed in section 3.2.2 (see also Table 3.2).

Mandel (32, p. 288) suggests that the simple replacement of "t" by "\( \sqrt{2F} \)" in equation 13 allows the equation to be used to predict any number of confidence limits. His approach (which isn't used in this study) results in narrower confidence intervals than the method of Garden and Miller.

Calculating Concentrations using the Calibration Curve

As was previously stated, the use of the regression line in reverse for the determination of concentration is referred to as "discrimination". This is what the analyst in effect does, when he determines the amount of analyte in a sample by interpolation using a calibration curve. What he is probably also interested in, but usually can only guess at, is the reliability of the determination.

Statistics can improve both the estimate of the concentration and its reliability.
The concentration can be easily calculated from the equation of the regression line:

\[ X = \frac{(y - a)}{b} \]

(16)

where \( X \) is the calculated concentration and \( y \) the measured response.

Confidence or "discrimination" limits for one, several, or an infinite number of discriminations can be determined. If only one determination of concentration is to be made using the calibration curve, than its limits will be narrower than the confidence limits (actually the simultaneous discrimination limits) which would be calculated if many such discriminations were to be made. The limits of one discrimination can be estimated by using equation 13 and solving it for \( x \). The two values of \( x \) obtained would represent the upper and the lower limits of the estimated concentration.

There are different approaches to determining the "unlimited simultaneous discrimination" limits (31,37,38); Miller's solution will be used. The idea again is that a confidence band can be placed on the calibration curve (eq. 14), and confidence intervals can be placed on the expected values of future observations (eq. 15). The limits of the observed responses can then be converted into discrimination intervals on the concentration axis. This is illustrated in figure 3.2; note that the limits are not necessarily symmetrical about the calculated \( x \)-values (i.e., temperature). This figure was derived from data presented by Natrella (26), for an example in which the observed values of Young's modulus for sapphire rods were measured at various temperatures to obtain a functional relationship between the two variables.
Fig. 3.2 Computed $x$-values and discrimination intervals using an unweighted calibration line. Adapted from an example by Natrell (26, p. 5-11).
The following formula is used to calculate the discrimination limits for the estimated concentration:

Max./Min. Conc'n

= Estimated conc'n +/- Discrimination Limit  (17)

The discrimination limits are calculated by solving the following equations for the maximum or minimum concentration by an iteration process rather than by solving the quadratic equation algebraically (Table 3.2 lists the form of the equation actually used in the program).

The form of the equation for the confidence limits of one concentration prediction is: (refer to eq. 13)

\[ X' = X + \pm \frac{s}{b} \times t[(1/m + 1/k + (X' - \bar{X})^2 / \Sigma (x_i - \bar{X})^2)^{0.5} \]  (18)

where \( X' \) is the upper/lower discrimination limit.

The equation for the discrimination limits for any number of estimated concentrations is: (refer to eqs. 14 and 15)

\[ X' = X + \pm \frac{s}{b} \times \left[ Z \times \left( \frac{(k-2)/\chi^2 + 1/m}{1/m} \right)^{0.5} + \left( 2F \times \frac{1/k + (X' - \bar{X})^2 / \Sigma (x_i - \bar{X})^2)^{0.5} \right) \right] \]  (19)

where again
- \( k = \) number of standards
- \( m = \) number of replicates for the sample
- \( Z = \) appropriate value of the normal distribution
- \( \chi^2 = \) appropriate value of the chi-squared distribution.

These last two equations can be solved by iteration; Lieberman et al. (37) give equivalent equations for the direct algebraic determination of the discrimination limits (\( X' \))—for simple linear regression.

As an example assume a confidence coefficient of 0.1 (\( \alpha = 0.1 \)), a 95% (\( 1 - \alpha / 2 \)) confidence band for the
calibration line, and 95% confidence limits for the response (37). Using these limits in equation 19 results in simultaneous discrimination intervals for "all" estimated concentrations using the given calibration curve. At least 95% of the discrimination intervals will contain the true concentrations with a confidence of at least 0.9 (21,31,37).

Natrella (26, p. 2-13) has discussed the general statistical tolerance limit and has given examples which the analyst may find somewhat useful in understanding this concept.

Because the actual number of predictions made from a given calibration line is usually limited, and for statistical reasons as outlined by Miller, the confidence level is expected to be between 0.9 and 0.95.

The limits estimated by equation 18 might be assumed to be the minimum analytical error, and the limits estimated by equation 19 might be assumed to be the maximum analytical error for an analysis. This of course is dependent on having all sources of variation included in the calibration experiment.

If the range of the calibration curve to be used is narrow it is quite possible that the discrimination intervals of the concentrations will be quite similar. Thus, if the intervals were being calculated manually then only a few need be determined as the differences between them would have little practical significance.

The technique of simultaneous discrimination intervals allows the analyst to statistically estimate the accuracy of results; he can then attempt to increase the accuracy if necessary. The analyst knows that increasing the number of calibration standards, and/or increasing the number of determinations per standard, and/or increasing the number of determinations per sample can increase precision (17), and hopefully also the accuracy. With this program the analyst can determine the effect of any or all of these techniques on improving the analytical error.

Section 3.4 gives an example of this statistical technique for unweighted linear calibration curves, in addition an example can be found in Appendix A. Figure 3.2
graphically illustrates the technique.

3.2.2 The Weighted Calibration Curve

Unweighted linear regression assumes that the individual analytical responses are distributed normally about the expected value with a constant variance, regardless of the value of the response or the concentration. This is not always the observed experimental situation.

For example, the variance of the response could be approximately proportional to the response or to its square; the data must actually be examined to determine how it varies. Brownlee (29, pp. 306-313) developed equations corresponding to some of the equations previously listed for unweighted regression.

When known, the actual response variances are used to weight the residuals, i.e., \((y_i - \bar{y}_i)\) is divided by \(V[y_i]\); the weight \((w)\) is therefore \(1/V[y_i]\). However, in most analyses the error variances for each level of concentration are unknown because they are too time consuming to determine. Therefore a "proportional" weight \((w')\) can be used. The proper use of this technique requires the correct functional relationship between the error variance and the response (or concentration) level. Chapter 4 (sec. 4.5) discusses the determination of the error function.

If the error variance is assumed to be proportional to the square of the response (a common assumption) then:

\[
W_i' = \frac{1}{V[y_i]} \cdot \frac{1}{V[y_i]} = \frac{1}{V[y_i]} \cdot W_i \propto W_i \text{ (20)}
\]

since \(\text{SQR}(V[y_i]/Y_i^2)\) is the relative standard deviation of the analytical error, and it is assumed to be constant.

In the ANOVA of the unweighted regression the mean square of the residual (eq. 6) was an estimate of the constant analytical error variance. In this example of regression the mean square of the weighted residual, \(\Sigma w'(y_i - \bar{y}_i)/(k-2)\), is an estimate of the relative error variance, i.e., the square of the relative standard
deviation. This is a very important consequence of the regression analysis; it will be built upon in this chapter, as well as, chapter 4.

It should be noted that although the confidence band for the weighted calibration line appears to differ significantly from that of the unweighted line (Fig. 3.1), they are in fact both similar. The use of weights ensures that the experimental error of responses with large variances will not overshadow the effect of responses with small error variances in determining the equation of the calibration line. The net result is that the weighted averages of the responses and the concentrations (through which the line must pass) are displaced from the unweighted centre of the line. However, as before the confidence band spreads out from the calibration line as x moves, in either direction, away from $\bar{x}$.

This effect is usually not evident from drawings of confidence bands of real experimental situations. Figures 3.1 and 3.3 show the effect more clearly. Refer to Appendix E for the details of the experiment used to obtain the data for figure 3.3.
Fig. 3.3 A weighted calibration curve and its confidence band, where the response variance is proportional to the square of the response (refer to Appendix E for details of the data analysis).
"Confidence" Limits for Responses and Concentrations using Weighted Calibration Curves

It was previously stated that the limits of a future response, predicted for a given concentration, would be a combination of the variance for the response plus the variance of the line at the given concentration, as given by the confidence band about the line. Equation 13 indicates that the confidence limits for the response of "one" new calibration standard would be:

\[ Y \pm t \frac{s}{\sqrt{1/k + \frac{(x-x)^2}{\sum(x_i-x)^2}}} \]

This expression can be modified for a weighted regression (where the response variance is proportional to the square of the response, and the sum of the weights equals the sum of the responses for the standards) as follows:

response limits for "one" new calibration standard
= \[ Y \pm t \frac{s}{\sqrt{1/w + \frac{(x-xw)^2}{\sum(x_i-xw)^2}}} \]

where
s is the standard error of regression
Y is the response
\( \sum w \) is the sum of the weights and \( xw \) is the weighted mean.

Equations 18 and 19 can be modified along the same lines. Tables 3.1 and 3.2 list the general form of the equations used in the program. These equations are for weighted and unweighted calibration lines. They can readily be compared to the equations already developed for unweighted regression. Sections 3.3.3 and 3.4 give examples of weighted regression, and Appendix A shows the calibration line and its confidence band for a literature example (21).

Situations where the Discrimination Limits cannot be Calculated

In the "well behaved" case (Fig. 3.4A) it can be seen that the limits can be determined. However, if the calibration curve is too flat then one (Fig. 3.4B) or both (Fig. 3.4C) limits cannot be determined. B and C represent poor analytical situations and would probably not be observed too often.
Fig. 3.4 Conditions under which concentration discrimination limits can and cannot be calculated.

A - upper and lower limits can be determined.
B - only the lower limit can be determined.
C - no limits can be determined.
3.3 The Microcomputer Program

Two versions of this module (or program) were written. The HDOS version uses "single precision" (7 significant figure) arithmetic; it is therefore limited to using data with a maximum of six figures and it produces results which are generally accurate to at least three or four significant figures. The accuracy of the routine has been tested using textbook and literature examples (3,5,12,16,18,26,35), as well as, on quality control test data. In particular, many of the results calculated for the example used in references 18 and 26 can be reproduced to six significant figures.

The second version (the CP/M version) is also restricted to using data with a maximum of six significant figures, although it uses "double precision" arithmetic for some calculations; this can result in more accurate results in some cases. Both versions are written in Microsoft BASIC. The text on BASIC programming by Wilkins et al. (39) has some short programs which could be usefully expanded upon by the analyst for his own use. The actual coding used in this program is presented in figure 3.5. The coding is written in a compact form and it uses other routines of the main program; it therefore cannot be used "as is". However, the formulae it uses are listed and the results it produces are discussed in the following sections.

The complete program and most of the experimental data is written on a 5 1/4" diskette and can be found in Appendix F.
**Fig. 3.5** BASIC microcomputer program segment for weighted and unweighted linear analytical calibration curves.
3.3.1 Statistical Formulae and Programming

One aspect of statistical calculations which can be confusing is the number of algebraically equivalent formulae available for the same statistic (12). The choice of the appropriate formulae for a microcomputer program is critical. Huff and Carter (40), and Wanek et al. (41) have shown how the results of a standard deviation calculation by different calculators can vary; this despite the fact that calculators usually handle more significant digits than single precision BASIC. Therefore the programmer must be selective in his choice of formulae. Programs used to calculate results similar to those reported in this chapter will tend to be very involved, thus the chance of accumulating significant errors because of real, as opposed to infinitely precise calculations, is great. The program must be checked (42) to determine the conditions under which it will produce acceptable results.

3.3.2 The Formulae Used in the Program

In developing this routine it was first necessary to find the appropriate formulae to carry out the required calculations. The most difficult to find were the weighted equations used to calculate concentration limits. The equations used are listed in Tables 3.1 and 3.2; they could be used to write one's own program. A microcomputer as simple as the Sinclair could be used; in fact, the Sinclair ZX81 (43) uses 9 significant figures and thus it could produce results which are more mathematically accurate than the HDOS version of the program.

Table 3.1 lists the more basic formulae used and Table 3.2 lists the formulae for calculating "confidence" limits. Some of these formulae are described more fully later.

NOTE, in the following sections the equation numbers referred to are from Tables 3.1 and 3.2; they will be referred to by using "EQ." rather than "eq."

Confidence limits may be calculated for a single mean response on the calibration line (EQ. 14) or for the calibration line as a whole (EQ. 15), i.e., the confidence band. The prediction limits of one predicted response
(EQ. 16) or the tolerance limits of an infinite number of predicted responses (EQ. 17) may also be calculated. In addition the prediction limits of only a few predicted responses could also be calculated (30), but such limits are usually of no practical use and thus will not be considered.

Estimated concentrations from measured responses can be calculated using EQ. 21. The discrimination limits of a single concentration, as in the case of a standard addition analysis, would be calculated using EQ. 22. If many estimates of concentration are to be made from the same calibration curve then EQ. 23 is used (17,21,30).

The appropriate values of Student's "t" are calculated using EQ. 18 (44,45) and the corresponding values of "F", for 2 and k-2 degrees of freedom, are calculated using Equation 19. Chi-squared is represented by "C"; the appropriate value is calculated from a formula similar to one listed by Acton (33, p. 241). Acton's formula was empirically modified to give a formula which allows the calculation of chi-squared values for a wider range of degrees of freedom.

The value of "A" used in some of these equations depends on the weight used i.e., A = 0 when the weight is one, A = 1 when the weight is Y or 1/Y (Y is the response), and A = 2 when the weight is Y -2 or 1/Y -2. If arbitrary weights are used to determine the calibration function then the "confidence" limits for the estimated responses or concentrations cannot be determined as the appropriate weights to use are unknown, e.g., the weights are not proportional to Y etc. Equations 16, 17, 22, and 23 are solved by an iteration technique in this program.
# TABLE 3.1
BASIC FORMULAE USED IN THE PROGRAM

<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{S_{wxx}} = \Sigma w(x_i - \bar{x_w})^2$</td>
<td>Initial calculations required:</td>
</tr>
<tr>
<td>$\mathbf{S_{wyy}} = \Sigma w(y_i - \bar{y_w})^2$</td>
<td>$\bar{x_w} = \Sigma x_i / \Sigma w_i$</td>
</tr>
<tr>
<td>$\mathbf{S_{wxy}} = \Sigma w(x_i - \bar{x_w})(y_i - \bar{y_w})$</td>
<td>$\bar{y_w} = \Sigma y_i / \Sigma w_i$</td>
</tr>
</tbody>
</table>

where $w =$ appropriate weighting factor

The calibration line:

\[ Y = a + bx \] (6)
\[ b = \frac{S_{wxy}}{S_{wxx}} \] (7)
\[ a = \bar{y_w} - b \cdot \bar{x_w} \] (8)

where $b =$ slope $\quad a =$ intercept

Regression coefficient:
\[ r = \frac{(S_{wxy})^2}{S_{wxx}} \] (9)
\[ \% \text{ explained variation:} \quad = r^2 \times 100 \] (10)

The standard error of regression:
\[ s = [(S_{wyy} - (S_{wxy})^2/S_{wxx})/(k-2)]^{0.5} \] (11)

where $k =$ number of standards used

The standard deviation of the slope and intercept:
\[ \text{SD}(b) = s/(S_{wxx})^{0.5} \] (12)
\[ \text{SD}(a) = s \times (1/\Sigma w + \bar{x_w}^2/S_{wxx})^{0.5} \] (13)

Confidence intervals
for single mean $y$ values on the calibration curve:
\[ \text{CI}(y) = +/- t \times s \times [(1/\Sigma w + (x - \bar{x_w})^2 / S_{wxx})^{0.5} \] (14)

Confidence intervals for the calibration curve as a whole:
\[ \text{CI}(y) = +/- [2F]^{0.5} \times s \times [(1/\Sigma w + (x - \bar{x_w})^2 / S_{wxx})^{0.5} \] (15)
TABLE 3.2
"CONFIDENCE" LIMIT FORMULAE
FOR COMPUTED RESPONSES AND CONCENTRATIONS

The prediction limits for one predicted response (for $x$):

$$Y' = Y +/- t * s * \left[ (a+bx)^A/Q + 1/\bar{w} + (x-\bar{x})^2/S_{xx} \right]^{0.5}$$  (16)

The tolerance limits for any number of predicted responses:

$$Y' = Y +/- s * \left\{ \sqrt{2F * \left( 1/\bar{w} + (x-\bar{x})^2 / S_{xx} \right)} + Z * \sqrt{(k-2)/C * (a+bx)^A / Q} \right\}$$  (17)

where $t = 1.96 + 2.36459/0+10.4069/0^2+3.16197/0^3+31.027/0^4$  (19)

$$F = t * (1.52617+2.7903/D+0.406921/D^2)$$  (19)

$$C = \text{Chi-squared} = [0.5 * (-1.625 + 0.75/D + (2Q-5)^0.5)^2]$$  (20)

$D = \text{degrees of freedom (must be >1)}$

$Z = \text{appropriate value of the normal distribution}$

The estimated concentration ($X$) from a known response ($y$):

$$X = (y-a) / b$$  (21)

The non-simultaneous discrimination limits
for one estimated concentration:

$$X' = X +/- t*s/|b| * \left[ (a+bx)^A/Q + 1/\bar{w} + (x-\bar{x})^2/S_{xx} \right]^{0.5}$$  (22)

The unlimited simultaneous discrimination limits
for any number of estimated concentrations:

$$X' = X +/- s/|b| * \left\{ \sqrt{2F * \left( 1/\bar{w} + (x-\bar{x})^2 / S_{xx} \right)} + Z * \sqrt{(k-2)/C * (a+bx)^A / Q} \right\}$$  (23)
3.3.3 Output of the Program

Two literature examples are used to demonstrate the usefulness of a microcomputer in producing linear calibration curves. The first example does not use weighting factors; it is found in Natrella (26, pp. 5-11). Refer to the appropriate section of the manual in Appendix A; this particular example illustrates many features of the program.

Weighted linear regression is illustrated in the section which immediately follows; the example uses artificial data by Garden et al. (21, Table III).

Real analytical data is used in sections 3.4 and 3.5 to further illustrate the use of this program in the analytical laboratory. Simulated data is used in a final example (sec. 3.6) to compare the computed and the expected results.

Equation of the Weighted Calibration Line

Figure 3.6 illustrates the formatting possible for the results of a regression routine. This particular program produces the results in a format which could possibly be used directly in a report. An ANOVA analysis indicates that the data can be explained by a linear equation. Note that 97.76% of the variance (of the response measurements) about the mean response is explained by the regression equation.

In addition to the slope (0.9949) and the intercept (0.0222), their standard deviations and confidence limits are also calculated. These confidence limits could be used to determine whether the slope or intercept could be a specific value, e.g., could the intercept be zero. Simultaneous considerations of the limits for both the slope and intercept would require a consideration of joint confidence (29, Ch. 5).

The equation of the computed line is similar to that given by Garden (21). It differs slightly because the error standard deviation was assumed to be proportional to the response, rather than the concentration. The computed relative standard deviation, represented by the standard error of regression, was 0.1064. Garden et al. indicated that it was 0.1; the computed value is thus in excellent agreement with the expected value.
Calculated Mean Responses and Residuals

Figure 3.7 shows the original concentration ($X$) and response ($Y$) values, the weight ($1/Y^2$ in this case), the computed $Y$, its confidence limits, and the weighted residuals (the difference between $y$ observed and $Y$ computed). Seven weighting systems ($7,23$) are automatically available to the user: $1, Y, 1/Y, Y^2, 1/Y^2, \text{an arbitrary weight}--\text{e.g., the error variance, or a transformed weight for absorbance responses (used when the transmittance variance is assumed constant).}$

When a weight, other than $1$ or an arbitrary weight, is selected the program goes into an iteration routine until a constant slope for the calibration curve is obtained. This makes the calculations slower, but more realistic.

Some authors indicate that the error variance can be proportional to the concentration. However, Schwartz (19) offers good reasons why the response, rather than the concentration, should be used. An additional reason is readily obvious if the technique is used with the standard addition technique; in such a case the actual concentrations of the standards are unknown until after the analysis is complete.

One of two confidence limits can be selected, the confidence limits for the mean of an individual computed $Y$ or the confidence band for the complete line. Figure 3.7 shows the computed data for the confidence band of the calibration line.

The sum of the residuals acts as a check as to whether the routine is working for the particular set of data points. Ideally the sum should be zero, but because of round-off errors it usually isn't; however, it should be small. Figure 3.7 also shows how the width of the confidence band varies with increasing response.
**Fig. 3.6** Basic information produced by the first screen of the linear regression routine. Not all of the six digits in the calculated values are necessarily significant.

**Fig. 3.7** The computed responses, the 95% confidence band of the computed calibration line, the weighted residuals, and the weights used in calculating the calibration line.
"Confidence" Limits for Responses and Concentrations

Figures 3.8 and 3.9 show the results when limits for responses or concentrations are calculated; both calculation routines work similarly. The formulae used are listed in Table 3.2.

Response and concentration limits are calculated using an iteration routine; Equation 16 or 22 is used when only one prediction or discrimination is made, e.g., for a standard addition analysis. Both calculations use FNX(D) as defined in statement 13002 of the BASIC program in figure 3.5.

Equation 17 or 23 is used when many predictions or discriminations are to be made. Both calculations use FNZ(D) as defined in statement 13002 of figure 3.5.

Equation 18 can be used for calculating the appropriate value (at the 95% confidence level) of "t" for k-2 degrees of freedom, and Equation 19 will convert this value to the corresponding "F" value for 2 and k-2 degrees of freedom. The value of chi-squared for k-2 degrees of freedom (at the lower 5% confidence level) is calculated using Equation 20. The value of Z (the normal distribution) used was 1.64.

The value of Q is the number of independent response measurements actually obtained for a sample of unknown concentration or the number of response measurements which would be made for a known standard.

Although the computed information previously described is important, the real purpose of the program is to calculate a concentration and its discrimination limits. As an example, the data given by Garden et al. (21) was reprocessed. A response of 1.01717 resulted in a concentration of 1.0000 (0.4455 to 1.5061); Garden et al. obtained 1.00 (0.45 to 1.50). Figure 3.10 is derived from their data; it shows the concentration and discrimination limits for a response of 7.00.

It should not be assumed that the calculated limits are the actual analytical error; they are perhaps best interpreted as the minimum possible error that would be predicted under the conditions of the particular analysis.
ESTIMATED RESPONSES and PREDICTION/TOLERANCE LIMITS

<table>
<thead>
<tr>
<th>MEAN X</th>
<th># Y RESPONSES</th>
<th>MEAN Y ESTIMATED</th>
<th>'90 - 95%' LIMITS</th>
<th>RELATIVE LIMITS +/- %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.022435</td>
<td>-273801</td>
<td>318206</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.01717</td>
<td>310872</td>
<td>1.52348</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3.0121</td>
<td>1.16261</td>
<td>2.86139</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3.00703</td>
<td>1.71612</td>
<td>4.29795</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4.00196</td>
<td>2.242</td>
<td>5.76192</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5.99152</td>
<td>3.27327</td>
<td>6.71038</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>7.98168</td>
<td>4.29657</td>
<td>11.6668</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>9.97154</td>
<td>5.31725</td>
<td>14.6238</td>
</tr>
</tbody>
</table>

Fig. 3.8 Computed responses and their "confidence" limits for selected concentration levels. Note how the relative limits of the predicted mean response decreases with increasing projected number of responses.

ESTIMATED CONCENTRATIONS and DISCRIMINATION LIMITS (1% or MANY)

<table>
<thead>
<tr>
<th>MEAN Y</th>
<th># Y RESPONSES</th>
<th>MEAN X ESTIMATED</th>
<th>'90 - 95%' LIMITS</th>
<th>RELATIVE LIMITS +/- %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01717</td>
<td>1</td>
<td>.999997</td>
<td>.696422</td>
<td>1.29291</td>
</tr>
<tr>
<td>1.01717</td>
<td>1</td>
<td>.999997</td>
<td>.445075</td>
<td>1.50507</td>
</tr>
<tr>
<td>3.00703</td>
<td>1</td>
<td>3</td>
<td>2.21654</td>
<td>3.84719</td>
</tr>
<tr>
<td>3.00703</td>
<td>1</td>
<td>3</td>
<td>1.84743</td>
<td>4.54695</td>
</tr>
<tr>
<td>9.97154</td>
<td>1</td>
<td>10</td>
<td>7.3448</td>
<td>12.9794</td>
</tr>
<tr>
<td>9.97154</td>
<td>1</td>
<td>10</td>
<td>6.01413</td>
<td>15.671</td>
</tr>
<tr>
<td>9.97154</td>
<td>2</td>
<td>10</td>
<td>6.79732</td>
<td>14.5576</td>
</tr>
<tr>
<td>9.97154</td>
<td>3</td>
<td>10</td>
<td>7.14408</td>
<td>14.0643</td>
</tr>
<tr>
<td>9.97154</td>
<td>5</td>
<td>10</td>
<td>7.49176</td>
<td>13.5697</td>
</tr>
<tr>
<td>9.97154</td>
<td>10</td>
<td>10</td>
<td>7.84144</td>
<td>13.0719</td>
</tr>
<tr>
<td>9.97154</td>
<td>100</td>
<td>10</td>
<td>8.41905</td>
<td>12.2501</td>
</tr>
<tr>
<td>9.97154</td>
<td>9E+37</td>
<td>10</td>
<td>9.68604</td>
<td>11.8701</td>
</tr>
</tbody>
</table>

Fig. 3.9 Computed concentrations and their "confidence" limits for selected responses. Note how the relative limits of the estimated concentration decreases with increasing number of independent responses.
Fig. 3.10 A weighted calibration curve showing "confidence bands" around the response \((Y)\), the calibration curve, and the concentration \((X)\). This figure uses data by Garden et al. (21, TABLE III).
The Residuals

The program stores the weighted residuals so they can be analyzed. Figure 3.11 shows a histogram of the residuals from the previous example. An analysis of the residuals can be helpful in determining whether the data is adequately represented by a linear equation. In addition, the residuals could be checked for outliers (Beckman and Cook have reviewed the topic of outliers. Refer to papers 3-10 in Appendix B).

Neter (30, Ch. 4) and Davies (28, p. 272) both consider the informal examination of residuals. They discuss how such an examination is useful for determining:

- nonlinearity of the calibration line
- nonconstancy of error variance
- nonindependance of error terms
- presence of outliers
- nonnormality of error terms
- omission of important independent variables.

<table>
<thead>
<tr>
<th>INTERVAL</th>
<th>FREQ</th>
<th>HISTOGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.2 TO</td>
<td></td>
<td>*********</td>
</tr>
<tr>
<td>-.12</td>
<td>1</td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXXX</td>
</tr>
<tr>
<td>-.04</td>
<td>2</td>
<td>XXXXXXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXXXXXXXX</td>
</tr>
<tr>
<td>.04</td>
<td>2</td>
<td>XXXXXXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXXXXXXXX</td>
</tr>
<tr>
<td>.12</td>
<td>3</td>
<td>XXXXXXXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXXXXXXXXX</td>
</tr>
<tr>
<td>.2</td>
<td>1</td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXXX</td>
</tr>
</tbody>
</table>

Fig. 3.11 The histogram of the weighted residuals from the calibration curve. The residuals are from figure 3.7.
3.4 The Standard Addition Technique

The use of this program for a determination using the standard addition technique (46,47) is illustrated with data from Larson et al. (48) and Franke (18).

In a determination using standard addition the concentration corresponding to a response of "0" is sought. It is assumed that the response of "0" is known exactly, i.e., it has no variance associated with it. The computer is informed of this by inputing that the number of responses is infinitely large (9E37 is used here). For these two examples there is no difference in the discrimination limits calculated since if only one response is used because the variance is assumed proportional to $Y^2$ (if $Y=0$ then the variance $= 0$). However, if this is not the case (e.g., when the line is unweighted) then a large difference is possible. The results obtained agree well with the published results.

3.4.1 Atomic Absorption Spectrometry

The data for this example was originally analyzed by Larsen (48, Table I). It was assumed in that data analysis that an unweighted line could be used. The data was later reanalyzed by Franke (18) where it was assumed that the relative standard deviation was approximately constant.

For this study the data was first analyzed using the microcomputer program, and assuming an unweighted line. The results obtained (Fig. 3.12) agree with those obtained by Larsen. However, Larsen assumed only one "measurement" of the "0" response in his calculation and thus obtained too wide a concentration interval. Franke, in his analysis of the data by the same method assumed no variance for the zero response; the program reproduced the new limit determined by Franke. A comparison of the results can be found in Table 3.4.

The data was then reanalyzed. This time the variance of the absorbance measurements were assumed proportional to the square of the absorbances. Again the computer results agreed with those of Franke's. Figure 3.13 shows the detailed computer results and Table 3.4 shows a comparison of computer and literature results.
ANOVA

SOURCE
F, B, M, B. 26

EXPLAINED BY REGRESSION 1 .491053
33756.5 0.000 ***
ABOUT REGRESSION 26 .78228E-04 1.4569E-05

TOTAL 27 .491431

ER'N OF LINE
y = .179078 x + .199057

REGRESSION COEFF. = .999615
% EXPLAINED VARIATION = 99.923

STANDARD ERROR OF REGRESSION (Br) = 3.81404E-03

CONFIDENCE LIMITS (90, 95, 99%) OF INTERCEPT: .190057 +/-
1.90491E-03 1.11664E-03

CONFIDENCE LIMITS (90, 95, 99%) OF SLOPE: .199057 +/-
1.03241E-03 2.70740E-03

CALCULATED Y VALUES and CONFIDENCE BAND for the COMPLETE LINE

<table>
<thead>
<tr>
<th>X</th>
<th>WEIGHT</th>
<th>Y obs'd</th>
<th>Y comp'd</th>
<th>+/- 95% CL</th>
<th>RESIDUAL (wt'd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>.197</td>
<td>.199057</td>
<td>2.89795E-03-2.05715E-03</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>.195</td>
<td>.199057</td>
<td>2.89795E-03-4.05715E-03</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>.194</td>
<td>.199057</td>
<td>2.89795E-03-5.7146E-03</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>.197</td>
<td>.199057</td>
<td>2.89795E-03-2.05715E-03</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>.195</td>
<td>.199057</td>
<td>2.89795E-03-4.05715E-03</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1</td>
<td>.291</td>
<td>.2808596</td>
<td>2.09738E-03-2.40409E-03</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1</td>
<td>.289</td>
<td>.2808596</td>
<td>2.09738E-03-4.0409E-04</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1</td>
<td>.269</td>
<td>.2808596</td>
<td>2.09738E-03-4.0409E-04</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1</td>
<td>.267</td>
<td>.2808596</td>
<td>2.09738E-03-5.9591E-03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.381</td>
<td>.378135</td>
<td>1.89716E-03-2.86531E-03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.364</td>
<td>.378135</td>
<td>1.89716E-03-5.86531E-03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.363</td>
<td>.378135</td>
<td>1.89716E-03-6.86531E-03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.382</td>
<td>.378135</td>
<td>1.89716E-03-7.86531E-03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.381</td>
<td>.378135</td>
<td>1.89716E-03-8.86531E-03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.38</td>
<td>.378135</td>
<td>1.89716E-03-9.86531E-03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.384</td>
<td>.378135</td>
<td>1.89716E-03-1.06531E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.555</td>
<td>.557212</td>
<td>.0034055-1.21224E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.553</td>
<td>.557212</td>
<td>.0034055-4.21224E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.551</td>
<td>.557212</td>
<td>.0034055-6.21224E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.551</td>
<td>.557212</td>
<td>.0034055-8.21224E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.551</td>
<td>.557212</td>
<td>.0034055-1.01224E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.551</td>
<td>.557212</td>
<td>.0034055-1.21224E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.953</td>
<td>.957212</td>
<td>.0034055-4.21226E-03</td>
<td></td>
</tr>
</tbody>
</table>

ESTIMATED CONCENTRATIONS and DISCRIMINATION LIMITS (1% or MANY)

<table>
<thead>
<tr>
<th>MEAN Y</th>
<th># Y RESPONSES</th>
<th>MEAN X</th>
<th>'90 - 95% LIMIT'</th>
<th>RELATIVE LIMITS +/- 1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.1157</td>
<td>-1.1616</td>
<td>-1.06203</td>
<td>4.87934</td>
</tr>
<tr>
<td>2</td>
<td>-1.1157</td>
<td>-1.1554</td>
<td>-1.0801</td>
<td>2.13363</td>
</tr>
</tbody>
</table>

Fig. 3.12 Some computer results for a determination of zinc using the standard addition technique and atomic absorption spectrometry (48). The standard deviation of the absorbance was assumed to be constant.
ANOVA

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F RATIO</th>
<th>PROB (%)</th>
</tr>
</thead>
</table>
| EXPLAINED BY REGRESSION | 1 | 3.4310e6 | 3.4310e6 | 33084.2 | 0.000 #
| ABOUT REGRESSION | 26 | 2.69363E-03 | 1.03709E-04 | |
| TOTAL | 27 | 3.4375 | |

EXPLANED BY REGRESSION

\[
\text{Y} = 0.181385 \times + 0.197341
\]

% EXPLAINED VARIATION = 99.9215

STANDARD ERROR OF REGRESSION (\(\sigma_r\)) = 0.101937

CONFIDENCE LIMITS (90, 95, 99%) OF INTERCEPT; 0.197341 +/-

1.17189E-03 1.41707E-04

CONFIDENCE LIMITS (90, 95, 99%) OF SLOPE; 1.70583E-03 2.76924E-03

CALCULATED Y VALUES and CONFIDENCE BAND for the COMPLETE LINE

<table>
<thead>
<tr>
<th>X</th>
<th>WEIGHT</th>
<th>Y obs'd</th>
<th>Y comp'd</th>
<th>+/- 95% CL RESIDUAL (WT'd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25.6702</td>
<td>.197</td>
<td>.197341</td>
<td>1.78317E-03 -1.72849E-03</td>
</tr>
<tr>
<td>0</td>
<td>25.6702</td>
<td>.195</td>
<td>.197341</td>
<td>1.78317E-03 -0.01833</td>
</tr>
<tr>
<td>0</td>
<td>25.6702</td>
<td>.199</td>
<td>.197341</td>
<td>8.46277E-03</td>
</tr>
</tbody>
</table>

Fig. 3.13 Some computer results for a determination of zinc using the standard addition technique and atomic absorption spectrometry (48). The relative standard deviation of the absorbance was assumed to be constant.
A New Approach

The data was next analyzed in a completely new manner. This time the absorbance measurements were first converted to transmittance (since \( A = -\log T \), therefore \( T = 10^{-A} \)). In the following analysis it is assumed that transmittance is actually the measured response. It is transformed to absorbance because the absorbance is directly proportional to the concentration, i.e., \( A = abc \), where \( a = \) absorptivity, \( b = \) path length in absorbing medium, and \( c = \) concentration.

The data then becomes:

<table>
<thead>
<tr>
<th>Conc'n</th>
<th>Transmittance Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.6353 .6383 .6324 .6383 .6353 .6383 .6353</td>
</tr>
<tr>
<td>0.5</td>
<td>.5117 .5140 .5140 .5164 .5193 .5152 .5164</td>
</tr>
<tr>
<td>1.0</td>
<td>.4159 .4130 .4140 .4150 .4130 .4169 .4130</td>
</tr>
<tr>
<td>2.0</td>
<td>.2780 .2799 .2812 .2748 .2761 .2812 .2799</td>
</tr>
</tbody>
</table>

The variance of the transmittance for each concentration was also calculated; they were 0.0502, 0.0579, 0.0248, and 0.0631. They "looked" reasonably similar. A two-factor ANOVA of variance was conducted to determine if there was a difference between the replicates. The variation between replication was not significant at the 95% level (the probability of the actual F ratio occurring by chance was 72.46%).

Jurs (49) and Sands (10) have discussed logarithmic transformations. Even when the original data is equally weighted (as it seems to be in this case), transformed weights must be used. For this particular example the transformed weights are \( T^{-2} \). The computer program (which was altered to handle this situation) was first tested on data used by Pollnow (8); the results predicted by Sands (10)
were reproduced. The program was then used to produce the calibration function, and the concentration and discrimination limits for the unknown zinc sample. The results:

\[ 1.091 \pm (1.072 - 1.111) \] or \[ 1.091 \pm 1.80\% \]

were quite similar to those produced when the weights were \(1/A^2\). Figure 3.14 shows some computer results for this particular analysis. Table 3.4 summarizes all the results obtained for this example.

A follow-up study of error variance for atomic absorption measurements is discussed in section 3.5.2. The results of that study also indicate that transmittance error variance is approximately constant over the range studied, while the apparent absorbance error variance is approximately proportional to the square of the absorbance measurements.
**Fig. 3.14** Some computer results for a determination of zinc using the standard addition technique and atomic absorption spectrometry (48). The weights were assumed to be the transmittance squared.
TABLE 3.4
A COMPARISON OF COMPUTED AND PUBLISHED RESULTS
FOR AN ANALYSIS USING
THE STANDARD ADDITION TECHNIQUE AND
ATOMIC ABSORPTION SPECTROMETRY

<table>
<thead>
<tr>
<th>COMPUTED</th>
<th>PUBLISHED (49,18)</th>
</tr>
</thead>
</table>

UNWEIGHTED LINE
Equation:
\[ A = 0.179078 \times + 0.199057 \]
\[ A = 0.179 \times + 0.199 \]

Standard error of regression:
\[ +/- 0.00381404 \]
\[ +/- 0.00381 \]

Concentration and discrimination limits:
\[ 1.11157 +/- 0.04979 + \]
\[ (1.06203 - 1.16161) \]

\[ 1.11157 +/- 0.02372 ++ \]
\[ (1.0881 - 1.13554) \]

WEIGHTED LINE
Equation:
\[ A = 0.181385 \times + 0.197341 \]
\[ A = 0.181 \times + 0.197 \]

\[ A = 0.181136 \times + 0.197684 \]

Concentration and discrimination limits:
\[ 1.08797 +/- 0.01802 * \]
\[ (1.07015 - 1.10617) \]

\[ 1.09135 +/- 0.02614 ** \]
\[ (1.07190 - 1.11124) \]

+ one response  *
++ many responses **

* Weight = 1/A^2
** Weight = T^2
3.4.2 Differential Pulse Anodic Stripping Voltammetry

The data presented by Franke (18, Table III) was also analyzed. First, it was analyzed assuming a constant response standard deviation, and then assuming a constant relative standard deviation. Table 3.5 gives a comparison of the computer results and the published results; figure 3.15 shows some of the computer results. The data analysis is discussed further by Killoran and Tyson (50).

| TABLE 3.5 |
| A COMPARISON OF COMPUTED AND PUBLISHED RESULTS |
| FOR AN ANALYSIS USING THE STANDARD ADDITION TECHNIQUE AND DIFFERENTIAL PULSE ANODIC STRIPPING VOLTAMMETRY |

| UNWEIGHTED LINE |
| Equation: |
| Y = 14.4986 X + 2.50796 |
| Concentration and discrimination limits: |
| 0.17298 +/- 0.024466 |
| (or +/- 14.14%) |

| WEIGHTED LINE |
| Equation: |
| Y = 14.3871 X + 2.6221 |
| Concentration and discrimination limits: |
| 0.182253 +/- 0.005481 |
| (or +/- 3.01%) |
Some computer results for a determination using the standard addition technique and Differential Pulse Anodic Stripping Voltammetry (18). The relative standard deviation of the response was assumed to be constant.
3.5 The General Comparative Method

An analytical calibration curve is usually used in the situation where many estimates of concentration will be made using it. This general application of the statistical data treatment will be illustrated with two experimental examples. The data for the first example is from a paper by Bocek and Novak (51); the data was later analyzed by Schwartz (19). The second set of data is from a determination for copper in ore samples using atomic absorption spectrometry. Twelve samples were analyzed for copper and the results compared with the "true" values.

3.5.1 Gas Chromatographic Quantitative Analysis

In this example the peak height is the measured response for benzene in a toluene solution. Schwartz (19) enlarged a figure in the original paper and read off 33 visible calibrating points; these were then grouped into nine sets having the same concentration. Schwartz analyzed the data by assuming both a constant standard deviation and a constant relative standard deviation for the response measurements. Although Schwartz used the number of replicates in each of the nine calibrating data groups for weighting the calibration line, he used only seven degrees of freedom in his calculation of the confidence intervals.

The data analysis using the computer program was complicated by an apparent error in the published data. This error caused the computer calculated analytical functions to differ from the published ones. The error was tracked down by an analysis of the residuals (sec. 3.3.3).

Analyzing the data in different ways resulted in the computer indicating that the response 7.85 was probably an outlier. An examination of Schwartz's data (Table III) indicated that probably the error was in the value for the injected charge of 2.28. Eventually it was decided that the correct charge was probably meant to be 2.88 rather than 2.28; an examination of the table indicates that such an assumption was in line with the regular difference which seemed to exist between each charge. The agreement between the computed calibration equation, using this corrected
charge, and the published equation further supported this correction.

Schwartz did not calculate the concentration confidence limits in the same manner as this program, therefore a difference was expected; his limits would have been expected to be narrower. In this new analysis, the data was entered into the computer as thirty-three calibration points; in this way the resulting number of replications was introduced, but also the number of degrees of freedom was thirty-one as it should be. The computed discrimination intervals were actually narrower than Schwartz's. This was probably due in part to the resulting smaller magnitudes of the t, F, and Chi-squared values which resulted from the larger number of degrees of freedom.

Schwartz selected three of the original thirty-three data points to serve as "unknowns". The computer results are compared to the published results in Table 3.6; figure 3.16 shows some of the computer results. It can be seen that the limits calculated by the two methods are of the same order of magnitude, but because of the differences in the calculations performed they do differ.
### TABLE 3.6
A COMPARISON OF COMPUTED AND PUBLISHED RESULTS
FOR A QUANTITATIVE ANALYSIS BY GAS CHROMATOGRAPHY

<table>
<thead>
<tr>
<th>COMPUTED</th>
<th>PUBLISHED (19)</th>
</tr>
</thead>
</table>

**UNWEIGHTED LINE**

Equation:

\[ Y = 2.76007 \times - 0.0357113 \]

Concentrations and discrimination limits for the "unknowns":

- 0.532128 ± 34.5% (0.345617 - 0.712543)
- 3.36068 ± 4.43% (3.21162 - 3.50905)
- 5.98018 ± 2.91% (5.80883 - 6.15705)

**WEIGHTED LINE**

Equation:

\[ Y = 2.78215 \times - 0.099311 \]

Concentrations and discrimination limits for the "unknowns":

- 0.550765 ± 8.21% (0.505123 - 0.595597)
- 3.35687 ± 5.06% (3.18881 - 3.52879)
- 5.95558 ± 5.24% (5.64753 - 6.27211)

The expected concentrations of the "unknowns" were:

- 0.564, 3.44, 5.73
**Fig. 3.16** Some computer results for a quantitative determination by Gas Chromatography using a linear calibration curve (19,50). The relative standard deviation of the response was assumed to be constant.

<table>
<thead>
<tr>
<th>MEAN Y</th>
<th># Y RESPONSES</th>
<th>MEAN X</th>
<th>'90 - 95% LIMITS</th>
<th>RELATIVE LIMITS +/- %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.433</td>
<td>1 s</td>
<td>3.50765</td>
<td>3.2105</td>
<td>3.4986</td>
</tr>
<tr>
<td>9.24</td>
<td>1 s</td>
<td>3.35687</td>
<td>3.2105</td>
<td>3.4986</td>
</tr>
<tr>
<td>16.47</td>
<td>1 s</td>
<td>5.92558</td>
<td>5.70769</td>
<td>6.20456</td>
</tr>
<tr>
<td>1.433</td>
<td>1</td>
<td>3.50765</td>
<td>3.05123</td>
<td>3.99597</td>
</tr>
<tr>
<td>9.24</td>
<td>1</td>
<td>3.35687</td>
<td>3.18861</td>
<td>3.52879</td>
</tr>
<tr>
<td>16.47</td>
<td>1</td>
<td>5.92558</td>
<td>5.44753</td>
<td>6.27211</td>
</tr>
<tr>
<td>16.47</td>
<td>2</td>
<td>5.92558</td>
<td>5.71918</td>
<td>6.1949</td>
</tr>
<tr>
<td>16.47</td>
<td>10</td>
<td>5.92558</td>
<td>5.70769</td>
<td>6.13316</td>
</tr>
<tr>
<td>16.47</td>
<td>100</td>
<td>5.92558</td>
<td>5.86766</td>
<td>6.04589</td>
</tr>
<tr>
<td>16.47</td>
<td>9E-37</td>
<td>5.92558</td>
<td>5.89214</td>
<td>6.02076</td>
</tr>
</tbody>
</table>

| 1.433  | 1 s            | 3.50765| 3.2105            | 3.4986              | 4.1284 |
| 9.24   | 1 s            | 3.35687| 3.2105            | 3.4986              | 4.1284 |
| 16.47  | 1 s            | 5.92558| 5.70769           | 6.20456             | 4.1715 |
| 1.433  | 1              | 3.50765| 3.05123           | 3.99597             | 9.2134 |
| 9.24   | 1              | 3.35687| 3.18861           | 3.52879             | 5.08398 |
| 16.47  | 1              | 5.92558| 5.44753           | 6.27211             | 5.2434 |

**ANNOVA**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>D.F.</th>
<th>B.S.</th>
<th>M.B.</th>
<th>F RATIO</th>
<th>PROB (%)</th>
</tr>
</thead>
</table>

**EXPLAINED BY REGRESSION**

| 13.344 | 15.344 | 32906.1 | 0.000 |

**TOTAL**

| 32 | 15.3572 |

**EQN OF LINE**

| y = 2.78215 x + -0.999314 | % EXPLAINED VARIATION = 99.9059 |

**STANDARD ERROR OF REGRESSION**

| 0.0201379 |

**CONFIDENCE LIMITS**

| 90% | 95% | 99% |

**TOTAL CONFIDENCE**

| 0.012874 | 4.03552E-04 |

**ERROR LIMITS (90, 95, 99%) OF INTERCEPT, SLOPE, EDIT**

| 2.7E-03 | 3.7E-03 | 7.8E-03 |

**V VALUES**

| 3.98435E-03 | 3.98435E-03 | 4.07801E-03 |

**WEIGHTED SUM = 5.40123E-03**

**CALCULATED Y VALUES AND CONFIDENCE BAND FOR THE COMPLETE LINE**

<table>
<thead>
<tr>
<th>X</th>
<th>WEIGHT</th>
<th>Y calc'd</th>
<th>Y comp'd</th>
<th>+/- 95% CL</th>
<th>RESIDUAL (WT'd)</th>
</tr>
</thead>
</table>
3.5.2 Determination of Copper by Atomic Absorption Spectrometry

In this example, a series of preanalyzed copper ore samples were analyzed. This example has an interesting aspect which was referred to earlier (sec. 3.4.1). The response which is usually observed is the absorbance, however, it is probably the transmittance which is actually measured. The transmittance measurements being converted by the electronics of the spectrometer to absorbance \( A = -\log T \), which is then displayed.

What has actually happened then is a transformation of the data; the purpose of course was to convert a nonlinear calibration curve to a linear curve. Jurs (49) and Sands (10) have pointed out that when a transformation is made, the weights must also be transformed. In the case where the original weight is 1, a logarithmic transformation converts it to \( T^2 \).

The papers by Jurs and Sands give details on the transformation. Several references (52-55) discuss the precision and accuracy of absorption and transmittance spectrophotometric measurements, and Mandel (32, p. 72) discusses the law of propagation of errors.

Experimental

APPARATUS. All measurements were made on a Varian AA-175 atomic absorption spectrometer which was set up following the manufacture's instructions.

REAGENTS. 99.99% copper wire was used for the standards. The unknowns were preanalyzed Thorn Smith (Thorn Smith Incorporated 7755 Narrow Guage Rd., Beulah, MI. 49617) copper ore samples. The samples were from two series; the first (#52, 54, 66, 69) was dated 1954 and the second (#74, 77, 79, 82, 83, 85, 86) was dated 1966.

PROCEDURE. Nine copper standards were prepared by weighing out the appropriate weights of copper wire (0, ca. 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4g).
Two gram samples (contained in paper envelopes) of each of the twelve Thorn Smith copper ores were dried at 100 degrees celsius for one hour. Each sample and its envelope was then weighed, the sample was transferred to a beaker and the envelope was reweighted.

Each standard and sample was then dissolved in 10 ml of nitric acid to which 5 ml of HCl acid was later added. Each sample solution was then filtered; all the solutions were diluted to 100 ml. Aliquots (0.2 ml) of each solution were then diluted with 100.0 ml of water.

The samples and standards had originally been prepared for analysis by another method, thus they were more concentrated than required for this method. The last dilution probably added to the experimental error—this however, was unimportant to the data processing.

A: The standard solutions were then (pseudo)randomly ordered. A distilled water blank was aspirated and the absorbance zeroed before aspirating each standard solution. Each standard was aspirated for twenty seconds before measurements were taken.

The digital meter was covered and only uncovered to allow measurements to be made (this was done to eliminate bias in selecting a measurement). Each measurement was recorded and a new measurement taken immediately until a total of 36 absorbance measurements were obtained. The meter was then switched to the transmittance mode and 36 transmittance readings were similarly obtained.

These readings were used to determine how the response error variance changed with increasing analyte concentration.

B: The determination for copper in the unknowns was conducted by (pseudo)randomly selecting a solution and measuring its absorbance and its transmittance, once for each of the 36 times it was selected. The absorbance reading was zeroed with a water blank between solutions. Each reading was obtained as above.
Results and Discussion

A: Error Variance. Descriptive statistics for each set of absorbance, transmittance, and calculated absorbance measurements were computed. The results are summarized in Table 3.7; histograms for the distribution of the transmittance measurements are also included.

The standard deviations for the absorbances (observed and calculated) were plotted vs. the corresponding mean absorbances. A linear regression analysis indicated an approximately linear relationship (Fig. 3.17C). A similar analysis of transmittance standard deviation vs. transmittance (Fig. 3.17D) showed no correlation. SLR was used in both cases.

The results of this experiment agreed with the previous results (sec. 3.4.1) for Larsen's (48) data, and the results produced by Agterdenbos (56, Table 2), i.e.,

- the error standard deviation for absorbance measurements are approximately proportional to the absorbance
- the error standard deviation for transmittance measurements (within the range examined) are apparently independent of the transmittance.

B: Determination of Copper. The analytical data was analyzed using a linear calibration curve. A constant error variance for transmittance was assumed. The computed results were compared to the Thorn Smith values using a paired t-test; the differences between the results were significant (Table 3.8) indicating a bias in the method.
TABLE 3.7
DESCRIPTIVE STATISTICS
FOR REPLICATE RESPONSE MEASUREMENTS
AT VARYING LEVELS OF
ANALYTE CONCENTRATION

<table>
<thead>
<tr>
<th>Conc'n (ug/ml)</th>
<th>Absorbance avg.</th>
<th>Absorbance var.</th>
<th>Transmittance avg.</th>
<th>Transmittance var.</th>
<th>Transmittances</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0042</td>
<td>0.094</td>
<td>0.0070</td>
<td>0.031</td>
<td>0.9840 1.57</td>
</tr>
<tr>
<td>1.433</td>
<td>0.1186</td>
<td>1.22</td>
<td>0.1217</td>
<td>1.22</td>
<td>0.7557 3.70</td>
</tr>
<tr>
<td>2.553</td>
<td>0.1974</td>
<td>3.89</td>
<td>0.2002</td>
<td>1.86</td>
<td>0.6307 3.92</td>
</tr>
<tr>
<td>3.483</td>
<td>0.2638</td>
<td>1.93</td>
<td>0.2672</td>
<td>1.45</td>
<td>0.5405 2.26</td>
</tr>
<tr>
<td>4.202</td>
<td>0.3431</td>
<td>3.42</td>
<td>0.3483</td>
<td>3.98</td>
<td>0.4484 4.25</td>
</tr>
<tr>
<td>5.234</td>
<td>0.4138</td>
<td>6.31</td>
<td>0.4190</td>
<td>3.22</td>
<td>0.3804 2.47</td>
</tr>
<tr>
<td>6.291</td>
<td>0.4937</td>
<td>10.41</td>
<td>0.4991</td>
<td>10.69</td>
<td>0.3169 5.70</td>
</tr>
<tr>
<td>7.220</td>
<td>0.5476</td>
<td>7.10</td>
<td>0.5484</td>
<td>2.72</td>
<td>0.2829 1.15</td>
</tr>
<tr>
<td>8.303</td>
<td>0.6339</td>
<td>6.96</td>
<td>0.6394</td>
<td>9.27</td>
<td>0.2294 2.39</td>
</tr>
</tbody>
</table>

Each variance is multiplied by $10^{-6}$

* Calculated absorbance  ** Non-normal distribution
### ANOVA

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F RATIO</th>
<th>PROB (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLAINED BY REGRESSION</td>
<td>16</td>
<td>3.52024</td>
<td>.220328</td>
<td>46.9062</td>
<td>0.002***</td>
</tr>
<tr>
<td>TOTAL</td>
<td>17</td>
<td>14.3006</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EDN OF LINE**

\[ y = 3.93887 + .520376 \]

**REGRESSION COEFF.** = .668038

| STANDARD ERROR OF REGRESSION (Br) | .469391 |

**CONFIDENCE LIMITS** (90, 95, 99%) OF INTERCEPT: 3.93887 +/-

**CONFIDENCE LIMITS** (90, 95, 99%) OF SLOPE: 0.520376 +/-

### EXPLAINED BY REGRESSION

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F RATIO</th>
<th>PROB (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLAINED BY REGRESSION</td>
<td>10</td>
<td>9.020234</td>
<td>.90202345</td>
<td>1.38831</td>
<td>1.45749</td>
</tr>
<tr>
<td>TOTAL</td>
<td>17</td>
<td>14.3006</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EDN OF LINE**

\[ y = .203485 x + 1.8102 \]

**REGRESSION COEFF.** = .120727

| STANDARD ERROR OF REGRESSION (Br) | .442083 |

**CONFIDENCE LIMITS** (90, 95, 99%) OF INTERCEPT: 1.8102 +/-

**CONFIDENCE LIMITS** (90, 95, 99%) OF SLOPE: -.203485 +/-

---

**Fig. 3.17** The variation of response standard deviation with response level.

**A** - Absorbance Standard Deviation Function

**B** - Transmittance Standard Deviation Function

**C** - Absorbance Standard Deviation vs. Absorbance

**D** - Transmittance Standard Deviation vs. Transmittance
TABLE 3.8
COMPARISON OF
DETERMINED AND ACTUAL CONCENTRATIONS OF COPPER
IN THORN SMITH COPPER ORE SAMPLES
=====================================================================
<table>
<thead>
<tr>
<th>THORN_SMITH_RESULTS</th>
<th>EXPERIMENTAL_RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample</td>
<td>%</td>
</tr>
<tr>
<td>#</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>9.87</td>
</tr>
<tr>
<td>54</td>
<td>12.22</td>
</tr>
<tr>
<td>66</td>
<td>13.03</td>
</tr>
<tr>
<td>69</td>
<td>8.35</td>
</tr>
<tr>
<td>74</td>
<td>11.97</td>
</tr>
<tr>
<td>77</td>
<td>6.33</td>
</tr>
<tr>
<td>79</td>
<td>9.12</td>
</tr>
<tr>
<td>81</td>
<td>15.58</td>
</tr>
<tr>
<td>81</td>
<td>15.58</td>
</tr>
<tr>
<td>83</td>
<td>12.56</td>
</tr>
<tr>
<td>85</td>
<td>16.81</td>
</tr>
<tr>
<td>86</td>
<td>17.78</td>
</tr>
</tbody>
</table>
3.6 Monte-Carlo Simulated Experiments

Simulated analytical experiments, in which the data was generated by Monte-Carlo methods, have been used in studies of linear regression by Krutchkoff (57) and Thompson (58). The advantage of simulated data is that it can be produced with known error parameters. A pseudo-normal random number generator is required; such a generator is part of the overall program (sec. 2.4).

3.6.1 Experimental

The experiment was conducted using two sets of conditions. The first consisted of 20 standards with three independent responses per standard. The second consisted of nine standards with two independent responses per standard. The first condition perhaps corresponds to a better than average routine calibration procedure, while the second corresponds more to a routine procedure.

For both conditions the calibration equation was expected to have an intercept of zero and a slope of one. The relative standard deviation of the response was expected to be 1%, and the response error was normally distributed.

The values of the twenty standards were generated by the BASIC language's random (uniform) number generator (58) and were within the range of 0 to 10. The nine standards were selected from this group of twenty such that they had approximately equally spaced values.

3.6.2 Results and Discussion

The results of the experiments indicated that the program was able to determine the parameters of the calibration function. Table 3.9 shows that the expected values were generally within the computed 95% confidence limits of the parameters.

Simulated data was also used to determine whether the expected values of concentration fell within the computed discrimination limits. Five sets of 100 responses, corresponding to an expected concentration of 5.0 and having a normally distributed error of 1% RSD, were generated for each of the two sets of calibration curves.
### TABLE 3.9
A COMPARISON OF EXPECTED AND COMPUTED CALIBRATION CURVE PARAMETERS

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CALIBRATION DATA SET</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Slope</strong></td>
<td>1.0009</td>
<td>0.9989</td>
<td>0.9987</td>
<td>0.9988</td>
<td>0.9987</td>
</tr>
<tr>
<td>(1.0)*</td>
<td>0.9982</td>
<td>0.9994</td>
<td>0.9976</td>
<td>0.9982</td>
<td>1.0044</td>
</tr>
<tr>
<td><strong>95% CL</strong></td>
<td>0.0015</td>
<td>0.0016</td>
<td>0.0024</td>
<td>0.0019</td>
<td>0.0020</td>
</tr>
<tr>
<td>+/-</td>
<td>0.0043</td>
<td>0.0045</td>
<td>0.0051</td>
<td>0.0053</td>
<td>0.0060</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>-0.0018</td>
<td>0.0031</td>
<td>0.0013</td>
<td>-0.0021</td>
<td>-0.0006</td>
</tr>
<tr>
<td>(0.0)</td>
<td>-0.0006</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0012</td>
<td>-0.0012</td>
</tr>
<tr>
<td><strong>95% CL</strong></td>
<td>0.0021</td>
<td>0.0022</td>
<td>0.0018</td>
<td>0.0027</td>
<td>0.0028</td>
</tr>
<tr>
<td>+/-</td>
<td>0.0021</td>
<td>0.0022</td>
<td>0.0025</td>
<td>0.0026</td>
<td>0.0029</td>
</tr>
<tr>
<td><strong>RSE Regress</strong></td>
<td>0.7562</td>
<td>0.7666</td>
<td>0.8620</td>
<td>0.9476</td>
<td>1.0005</td>
</tr>
<tr>
<td>(1 %)</td>
<td>0.7788</td>
<td>0.8105</td>
<td>0.9231</td>
<td>0.9574</td>
<td>1.0663</td>
</tr>
<tr>
<td><strong>Reg Coeff</strong></td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
</tr>
<tr>
<td>(1.00)</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

* The values in brackets are the expected values. 
The first entries result from twenty standards with triplicate responses. 
The second entries result from nine standards with duplicate responses. 
The sets are arranged in order of increasing standard error of regression.
Discrimination limits for the concentrations related to each response, using each calibration curve, were then computed. These limits were examined to determine whether they bracketed the expected concentration. The number not bracketing the concentration are listed in Tables 3.10 and 3.11.

The results, as can be seen from these tables, do not appear to contradict the expectation that for at least 90% of all possible calibration curves, 95% of the simultaneous discrimination intervals will contain the expected (unknown) concentration.

A count of the number of times the non-simultaneous discrimination limits failed to bracket the expected concentration definitely indicated that the limits calculated for a single discrimination could not be used for multiple discriminations.
**TABLE 3.10**
The number of simultaneous discrimination limits not bracketing the expected concentration

<table>
<thead>
<tr>
<th>RESPONSE DATA SET</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
</tr>
<tr>
<td>SD#</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CURVE</th>
<th>NO. OUTSIDE THE DISCRIMINATION LIMITS</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3 10 10 12 15</td>
<td>8.0</td>
</tr>
<tr>
<td>2</td>
<td>1 5 4 6 6</td>
<td>4.4</td>
</tr>
<tr>
<td>3</td>
<td>1 6 5 7 6</td>
<td>5.0</td>
</tr>
<tr>
<td>4</td>
<td>1 4 3 3 5</td>
<td>3.2</td>
</tr>
<tr>
<td>5</td>
<td>1 3 3 2 4</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Calibration curves from twenty standards and triplicate responses.

* SD refers to the computed standard deviations of each set of 100 "measurements". The expected standard deviation was 0.050.
TABLE 3.11
NUMBER OF SIMULTANEOUS DISCRIMINATION LIMITS NOT BRACKETING THE EXPECTED CONCENTRATION

<table>
<thead>
<tr>
<th>RESPONSE DATA SET</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>SD*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CURVE</th>
<th>NO. OUTSIDE THE DISCRIMINATION LIMITS</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 1 5 2 4</td>
<td>2.4</td>
</tr>
<tr>
<td>2</td>
<td>0 1 2 1 3</td>
<td>1.4</td>
</tr>
<tr>
<td>3</td>
<td>0 0 1 1 3</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>0 0 1 1 3</td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>0 0 2 1 2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Calibration curves from nine standards and duplicate responses.

* SD refers to the computed standard deviation for each set of twenty "measurements". The expected standard deviation was 0.050.
3.7 Nonlinear Calibration Curves

Analytical calibration curves unfortunately are not always linear, even when theory indicates they should be.

This is often the case when the standards cover a very wide concentration range. A real calibration curve could deviate substantially from a straight line, particularly at the high concentration end. However, the analyst will normally choose analytical conditions which will result in as little deviation from a straight line as possible. In practice, when a nonlinear calibration curve is used, it will often tend to be linear at lower concentrations and will only exhibit moderate deviation from the straight line at the high concentration end.

The literature search for the data analysis of nonlinear calibration curves was restricted to the same sources searched for linear curves. Several papers (9,17,19,55,56,59-61) illustrate simple data treatment, i.e., determination of concentration only, and/or treatments where both concentrations and discrimination limits are determined. The literature (31) suggests that little work was done with regards to concentration limits of nonlinear curves prior to 1966. It also appears that to date this topic has not been covered as extensively as the linear calibration curve.

No programs duplicating the data analysis which has already been described for linear calibration have been written for this thesis. However, it is possible to do a simple analysis of nonlinear curves with the programs already described, and with a simple quadratic least squares program such as the one by Wilkins et al. (39, pp. 331-338).

3.7.1 Approaches Described in the Literature

Schwartz (19,60,61) has considered three approaches to the data treatment of the nonlinear calibration curve:

- linear segmented approximation
- three parameter function of specified form
- polynomial function of adjustable degree.
These three approaches are illustrated (61) by a simulated experiment run on a digital computer. The resulting nonlinear curve is evaluated by the three methods.

The "linear segments" method consists of selecting a portion of the calibrating curve which brackets the response of the unknown and which is approximately linear. The data treatment is then identical to that already described for a linear curve. Although on first consideration this method would seem to be a poor choice, e.g., Schwartz (19) abandoned it because of its crudity, Mitchell et al. (17) have suggested several good reasons for its serious consideration. A modified approach using this technique for the determination of concentration discrimination intervals is illustrated in section 3.7.2.

The "three parameter function" consists of fitting a two or three parameter empirical equation to the full range of the calibrating points. The most obvious form of the equation is:

\[
\text{Response} = A + B \times \text{Concentration} + C \times \text{Concentration}^2
\]

where \(A\), \(B\), and \(C\) are constants

This equation can be generalized; Schwartz provides the necessary details. It will be assumed in the approach illustrated in the next section that the second order equation will approximate most nonlinear curves that an analyst would consider using. Mitchell (17) has stated that third- and higher-order regressions, with points of inflection, are not appropriate for most analytical measurements.

The "polynomial function" method consists of constructing the curve with as many adjustable parameters as are required to fit the curvature of the calibrating data. The calculations required are fairly extensive. Again Schwartz has given details on the method.
3.7.2 A Simplified Approach to Nonlinear Calibration Curves

As was previously mentioned Schwartz (19,61) has described, illustrated, and then abandoned the linear segmented approximation. However, Mitchell et al. (17) have supported this approach. Mitchell refers to it as the "Multiple-Curve" procedure. The rigorously exact procedure would be preferred, but when it is not available this simplified approach does have merit. Mitchell attributes the following advantages to the procedure:

- discernable improvements in precision for high quality data and for analyses over narrow dynamic ranges. And

- large improvements in confidence bands (up to a factor of three)
- improvements in minimum reportable concentrations (up to a factor of seven)

for poor-quality data and for many nonlinear curves.

This approach is illustrated using the "synthetic" data produced by Schwartz (61). The data was obtained using a nonlinear calibration function, \( y = \tanh^3 x \), which was evaluated between 0.750 and 3.000. Experimental responses were computer generated assuming a standard deviation of 0.0500. Six responses for an "unknown" were generated using the same standard deviation. The "true" unknown response was 0.800 corresponding to a concentration of 1.646. Refer to figure 3.1B for the actual "experimental" data.

Schwartz determined the concentration of the unknown by interpolating between the responses of the two standards bracketing the unknown. He obtained a concentration of 1.678 and limits of 1.431 and 2.767 (a 70% confidence level had to be used to solve for the limits, refer to Fig. 3.4). Using a second order curve he obtained a concentration of 1.657 and limits of 1.543 and 1.789.
In this new analysis the second order curve was first calculated (all the calibration data, weight of 1, and the program of Fig. 3.19 was used). The concentration of the unknown was determined to be 1.65709, in agreement with Schwartz. The concentration was then similarly calculated after the responses for the last three standards were discarded; it was then 1.62204. This latter calculation was performed as it was thought that the resulting curve would be closer to the type of nonlinear curve that an analyst would actually consider using.

Since it is known that the standard deviation of the responses is constant, a weight of one would be assumed correct. However, the data is arbitrarily being fitted to a second order line. If this line is not the correct equation for the curve, which it isn’t, then one might expect the residuals to show more variability at higher response levels than at lower levels. The line was recalculated using all calibrating data, but a weight of 1/response^2 was used.

The concentration under these conditions was 1.64844. It is only coincidental, but the concentration produced by this calculation is very close to the "known" value of 1.646. It is actually closer to it than any value calculated by Schwartz using any of the three procedures. A calculation of the concentrations corresponding to the mean responses for all but the last three standards showed excellent agreement (Fig. 3.18). The concentrations of these last three standards were low indicating that a second order equation is not correct for this data.

Following Mitchell’s procedure, the data from four standards, two with a concentration greater than and two with a concentration lesser than the unknown, were used to calculate both the concentration and its limits. The results were:

1.63239 (1.26808-1.85995, +/- 18.13%)
   weight=1,    # replicates = 6, one discrimination
1.63330 (1.29037-1.85741, +/- 17.36%)
   weight=1/Y^2, # replicates = 6, one discrimination
NONLINEAR CALIBRATION CURVES

\[ Y = A_0 + A_1 \times X + A_2 \times X^2 \]

<table>
<thead>
<tr>
<th>X VALUE</th>
<th>Y VALUE</th>
<th>WEIGHT</th>
<th>CALC Y</th>
<th>WT'D RESIDUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>.75</td>
<td>.9167</td>
<td>.2164</td>
<td>.2145</td>
<td>.240819</td>
</tr>
<tr>
<td></td>
<td>.296</td>
<td>.1114</td>
<td>.240819</td>
<td>-.114901</td>
</tr>
<tr>
<td>.75</td>
<td>.223</td>
<td>.20109</td>
<td>.240819</td>
<td>-.0799043</td>
</tr>
<tr>
<td>1</td>
<td>.5103</td>
<td>.30416</td>
<td>.426501</td>
<td>.164216</td>
</tr>
<tr>
<td>1</td>
<td>.612</td>
<td>.589122</td>
<td>.426501</td>
<td>-.025176</td>
</tr>
<tr>
<td>1.25</td>
<td>.4304</td>
<td>.539828</td>
<td>.426501</td>
<td>9.05966E-03</td>
</tr>
<tr>
<td>1.25</td>
<td>.6342</td>
<td>2.45627</td>
<td>.584995</td>
<td>.0775067</td>
</tr>
<tr>
<td>1.25</td>
<td>.5483</td>
<td>3.32632</td>
<td>.584995</td>
<td>-.0646381</td>
</tr>
<tr>
<td>1.25</td>
<td>.6209</td>
<td>.259392</td>
<td>.584995</td>
<td>.0578261</td>
</tr>
<tr>
<td>1.5</td>
<td>.7932</td>
<td>1.56941</td>
<td>.7163</td>
<td>.0644922</td>
</tr>
<tr>
<td>1.5</td>
<td>.7721</td>
<td>1.67746</td>
<td>.7163</td>
<td>.0722705</td>
</tr>
<tr>
<td>1.5</td>
<td>.8686</td>
<td>2.12107</td>
<td>.7163</td>
<td>-.0434084</td>
</tr>
<tr>
<td>1.75</td>
<td>.9093</td>
<td>1.26162</td>
<td>.820417</td>
<td>.0764938</td>
</tr>
<tr>
<td>1.75</td>
<td>.7134</td>
<td>1.94647</td>
<td>.820417</td>
<td>-.15001</td>
</tr>
<tr>
<td>1.75</td>
<td>.7777</td>
<td>1.65339</td>
<td>.820417</td>
<td>-.054973</td>
</tr>
<tr>
<td>2</td>
<td>.9218</td>
<td>1.17687</td>
<td>.897346</td>
<td>.0265269</td>
</tr>
<tr>
<td>2</td>
<td>.8925</td>
<td>1.2354</td>
<td>.897346</td>
<td>-.5.42933E-03</td>
</tr>
<tr>
<td>2</td>
<td>.8945</td>
<td>1.2498</td>
<td>.897346</td>
<td>-.3.18127E-03</td>
</tr>
<tr>
<td>2.25</td>
<td>.9966</td>
<td>1.24395</td>
<td>.947006</td>
<td>-.0563063</td>
</tr>
<tr>
<td>2.25</td>
<td>.9657</td>
<td>1.0723</td>
<td>.947006</td>
<td>.0192731</td>
</tr>
<tr>
<td>2.25</td>
<td>.9339</td>
<td>1.14167</td>
<td>.947006</td>
<td>-.0119522</td>
</tr>
<tr>
<td>2.5</td>
<td>.9176</td>
<td>.118774</td>
<td>.969439</td>
<td>-.0587111</td>
</tr>
<tr>
<td>2.5</td>
<td>.9298</td>
<td>1.1567</td>
<td>.969439</td>
<td>-.0428459</td>
</tr>
<tr>
<td>2.5</td>
<td>.9743</td>
<td>1.03302</td>
<td>.969439</td>
<td>4.09811E-03</td>
</tr>
<tr>
<td>2.75</td>
<td>.9847</td>
<td>1.03312</td>
<td>.965002</td>
<td>.0200043</td>
</tr>
<tr>
<td>2.75</td>
<td>.9401</td>
<td>1.13149</td>
<td>.965002</td>
<td>-.0264888</td>
</tr>
<tr>
<td>2.75</td>
<td>.9493</td>
<td>1.10967</td>
<td>.965002</td>
<td>-.0165404</td>
</tr>
<tr>
<td>3</td>
<td>.9237</td>
<td>1.37203</td>
<td>.933177</td>
<td>-.010201</td>
</tr>
<tr>
<td>3</td>
<td>1.0205</td>
<td>1.945348</td>
<td>.933177</td>
<td>.0926813</td>
</tr>
<tr>
<td>3</td>
<td>.9661</td>
<td>1.07141</td>
<td>.933177</td>
<td>.034078</td>
</tr>
</tbody>
</table>

RESPONSE CONCENTRATION

| .240619 | .75    |
| .42501  | 1      |
| .584995 | 1.25   |
| .7163   | 1.5    |
| .820417 | 1.75   |
| .897346 | 2      |
| .947006 | 2.25   |
| .969439 | 2.5    |
| .965002 | 2.41447|
| .933177 | 3      |
| .7814   | 1.64844|

WEIGHTED SUM = 0

Fig. 3.18 Computer results using a second order calibration curve and a weight of 1/response^2 (the response relative standard deviation is assumed constant).
100 REM QUADRATIC LEAST SQUARES FIT
101 ' MODIFIED BY G.N. KILLORAN 83-01-28
103 DEFDBL A-Z: DEFINT I,J
110 DIM W(100), X(100), YC(100), A(3,3), B(3,3), C(3)
120 PRINT
140 PRINT "WEIGHTS: UNIT (1), 1/Y (2), 1/Y^2 (3), SPECIAL (4)";
160 INPUT M9 IF M9=1 OR M9=4 THEN 140
180 READ N
200 FOR I=1 TO N: READ X(I), Y(I)
220 NEXT I
230 IF N<1 OR N>4 THEN 140
250 READ N
270 KE = 0.00043+.0000177*EXP(3.4*Y(I))
290 NEXT J
300 LPRINT: LPRINT: LPRINT: LPRINT: LPRINT: LPRINT
LPRINT CHR$(14): "NONLINEAR CALIBRATION CURVES";
LPRINT "Y = A0 + A1 * X + A2 * X^2"
LPRINT: LPRINT: LPRINT: LPRINT: LPRINT
LPRINT "ZERO", A0, "STD DEY", DO
LPRINT "ONE", A1, "STD DEY", D1
LPRINT "TWO", A2, "STD DEY", D2
LPRINT "BASE", A3, "STD DEY", D3
600 LPRINT "MA
LPRINT "ZERO", A0, "STD DEY", D0
LPRINT "ONE", A1, "STD DEY", D1
LPRINT "TWO", A2, "STD DEY", D2
LPRINT "BASE", A3, "STD DEY", D3
800 FOR I=1 TO 100
810 I=999 THEN RETURN
810 INPUT "X VALUE", X(I)
820 IF I=999 THEN RETURN
830 VV=V(I)
840 NEXT I
1000:SUBROUTINE
1010 FOR I=1 TO 3: C(I,1)=0
1020 NEXT J
1030 NEXT J
1070 FOR I=1 TO N
1080 A(I,1)=A(I,1)+W(I)
1090 A(I,2)=A(I,2)+W(I)*X(I)
1100 A(I,3)=A(I,3)+W(I)*X(I)^2
1110 A(I,2,3)=A(I,2,3)+W(I)*X(I)^3
1120 A(I,3,3)=A(I,3,3)+W(I)*X(I)^4
1130 C(I,1)=C(I,1)+W(I)
1140 C(I,2)=C(I,2)+W(I)*X(I)
1150 C(I,3)=C(I,3)+W(I)*X(I)^2
1160 NEXT J
1170 C(1,2)=C(1,2)+W(I)*X(I)
1180 C(2,3)=C(2,3)+W(I)*X(I)^2
1190 C(3,3)=C(3,3)+W(I)*X(I)^3
1200 D(1,1)=D(1,1)+W(I)
1210 D(1,2)=D(1,2)+W(I)*X(I)
1220 D(1,3)=D(1,3)+W(I)*X(I)^2
1230 D(2,1)=D(2,1)+W(I)*X(I)
1240 D(2,2)=D(2,2)+W(I)*X(I)^2
1250 D(2,3)=D(2,3)+W(I)*X(I)^3
1260 D(3,1)=D(3,1)+W(I)*X(I)
1270 D(3,2)=D(3,2)+W(I)*X(I)^2
1280 D(3,3)=D(3,3)+W(I)*X(I)^3
1290 V=V(I)
1300 A(I,1)=A(I,1)+W(I)*X(I)
1310 A(I,2)=A(I,2)+W(I)*X(I)^2
1320 Q-Q+W(I)*Y(I)
1330 NEXT I
1340 Q=Q/C(N-1)
1350 RETURN
1360 END
Fig. 3.19 A quadratic least squares BASIC program (39). The program was modified to allow additional weighting factors.
Schwartz obtained a concentration of 1.678, the true concentration is 1.646. The computed concentrations are lower than the true value but actually somewhat closer to it than the value obtained by Schwartz.

3.7.3 Comparison of Calculation Methods

The concentration limits obtained by Schwartz for the "unknowns" were narrower than the limits computed in this study. His results for the four methods were:

1.657 (1.543-1.789, +/- 7.42%) * parabolic
1.635 (1.500-1.796, +/- 9.05%) * logarithmic
1.500 (1.303-1.806, +/- 16.77%) * polynomial
   (only the first of each replicate response obtained from the experiment was used)
1.678 * linear segmented

The relative 95% confidence limits for the mean of the six response measurements of the unknown are +/- 11.95%. It is thus surprising that Schwartz obtained concentration limits which were less than the relative limits of these corresponding responses. The concentration limits obtained in this study were approximately 17-18%.

In order for a calculated calibration curve to be correct, i.e., for it to yield true results, there must be a sufficiently large number of replicates of the response data (for the knowns and the unknowns) so that the means of the replicates will closely approximate the "true" population means. There must also be a sufficiently large number of calibrating points in order for the analyst to accurately compute a good estimate of the calibration function.

The "experimental" data generated by Schwartz shows that even under theoretical conditions, nonlinear experimental data is not easily fitted to an exact calibration function. This experiment should again indicate to the analyst that his results are never exactly correct.
(except by chance), and that concentration limits should always be determined.

Under everyday laboratory conditions and with limited calculating "tools" the laboratory must rely on the analyst's experience in selecting an "appropriate" nonlinear calibrating curve. However, if

- many standards are used in constructing the curve
- the concentrations of the standards are evenly distributed
- many independent replicate response measurements are taken

and if

- experimental conditions are chosen to give a curve with a minimum amount of deviation from a straight line

then the concentration and its limits can be estimated reasonably accurately with this modified "linear segmented" method.

The discrimination limits can be calculated by the usual linear segmented method. That is, prior to doing an actual determination, the limits over the range of the curve could be calculated by taking overlapping groups of at least four standards, and assuming a fixed number of response replicates for each unknown. These limits could then be tabulated. The computed concentration of the unknowns would, however, be calculated using a quadratic least squares program similar to the one already described. Weights would be used as required. The appropriate concentration limits would then be selected from the previously tabulated limits for the curve.

The use of this modified linear segmented approach produced the following results:

1.648 (1.362-1.934, +/- 17.36%) (true result = 1.646)
3.8 Minimum Reportable Concentrations

In the previous section it was noted that Mitchell et al. (17) used confidence band statistics to provide minimum reportable concentrations. Hubaux and Vos (62) have also used linear calibration curve confidence bands for the prediction of decision and detection limits. These limits will be briefly described; additional detail are available in the above literature.

Mitchell states that the minimum reportable concentration would be that concentration for which the computed concentration band (at a selected confidence level) would just include a zero concentration. This concentration could be determined by assuming lower and lower response readings until the confidence band just includes a zero concentration. The concentration corresponding to this response would be the "minimum reportable concentration".

The discussion on limits by Hubaux and Vos is more extensive and has an illustrating figure. Their paper discusses the concept of two sensitivity limits; a signal level and a content (concentration) level. Both of these limits are also the ones used by Mitchell et al.

The above determination assumes an unweighted calibration curve. However, if a weight such as 1/response^2 was used then the response standard deviation at the zero level would be zero; this would be an unlikely occurrence. Thompson and Howarth (63) suggest the concept of a limiting standard deviation at low concentrations (sec. 4.5.1). Therefore, the minimum reportable concentration might be best determined using standards of very low concentration and a weight of "1". This is probably the approach most often used.
3.9 Practical Considerations when Conducting a
Statistical Analysis of Analytical Calibration Curves

Three factors will be considered in this section, the various methods of data analysis available, the limitations of the least squares method, and a procedure which could be followed in applying the least squares method of statistical analysis to calibration curves. A fourth factor, the effect of the number of standards and replicate responses for samples and standards has already been discussed (sec. 3.2.1 - Calculating Concentrations from the Calibration Line).

3.9.1 Methods of Data Analysis

This chapter has been concerned only with least squares statistical analysis. It should be kept in mind however, that there are other ways to treat linear calibration curves. Kelly (64) has listed six methods for the data processing of linear curves. These methods, in order of increasingly "detailed picture of reality" (and increasing "difficulty of calculation"), are as follows:

- direct
- graphical
- minmax
- least squares
- maximum likelihood
- methods based on Bayes' theorem

Kelly illustrates these six methods using data from an analysis for copper in a nickel powder by the technique of standard addition with an atomic absorption spectrophotometer. The maximum likelihood method is also illustrated by Thompson (58) for data generated by Monte Carlo simulation (sec. 4.6.3).

It is obvious that the method of least squares is a compromise between numerical difficulty and the detail in which one desires to examine the data.
3.9.2 Limitations of the Least Squares Method

The classical least squares method is based on four hypotheses:

- the standards are independent of one another
- the contents of the standards are accurately known
- the variance of the response error distribution is constant
- the response error has a Gaussian distribution

It is unlikely that all, if any, of these hypotheses are exactly correct in an application of this method to an analytical calibration curve. Each of these hypotheses and the effect of departures from them will be considered.

Independent Standards

Standards should be prepared separately, i.e., they should differ as much in their preparation from one another as from the samples to be analyzed. For example, the preparation of a series of standards from a single standard solution should be avoided. This condition is sometimes difficult, and almost always time consuming, to satisfy. If this condition is not met then the accuracy of the resulting calibration curve is overestimated. Each standard should be taken through the entire analytical procedure, and should generate one average response.

Accurately Known Standards

It is very unlikely that the amount of analyte in the standards is known exactly. If this is the case then both the response and the amount ($Y_i$ and $X_i$) have an associated error variance. Thus, the simple classical form of least squares analysis could be considered to be incorrect and the more mathematically difficult calculation where both variables are assumed to be subject to error (32, p. 288; 58) should be used. However, the analyst is fortunate as this is not always the case.

Mandel (32, Ch. 12) points out that many experiments are carried out in such a way that the amount of analyte is...
a "controlled" quantity, in the sense that for each measurement of response, the corresponding value of quantity is "set" at an assigned level, or at least as close to that level as is experimentally feasible. Because of this condition (this aspect of straight-line fitting was discovered by Berkson (65)) the classical least squares method can be used. It should be noted though that the scatter about the fitted line (from which the response variance is estimated) will now include a contribution from both the error of the response and the standard. The resulting "confidence band" for the computed amount of analyte is therefore wider than it would be, had accurately known standards been used.

The study by Thompson (58) also indicates that a simple or weighted linear regression gives reasonably accurate results as long as the variance for x (the independent variable) is less than the variance of y (the dependent variable).

Another possibility is that the error associated with the measured response is less than the error associated with sample preparation. This possibility has been discussed by Smith and Mathews (5). The regression line of X upon Y would be calculated in this case. Krutchkoff (57) started a long lasting discussion when he suggested that this "inverse regression"—even when using a controlled variable X, and a random variable Y—resulted in a smaller mean squared deviation than the classical regression.

The analyst must make an effort to determine the variance due to the measurement of response and the variance due to sample preparation. A simple comparison of the variances by an F test will indicate that the variance:

- of the response measurements is greatest
- of sample preparation is greatest
- of both the response measurements and sample preparation are appreciable.

This topic is expanded upon in chapter 4 (sec. 4.6.4, Table 4.3).
Response Variance is Constant

It would be very time consuming to determine the exact variance function of the response, Schwartz (19) for example indicates that 20 or more replicate standard samples at 6 or more concentration levels would be required.

The computer program allows the calculation of a regression line assuming constant variance, variance proportional to the response or its square, or to the reciprocals of these quantities. The program could also be altered to allow for other possibilities.

An examination of the variances from a limited number of replicate measurements for the standards might allow the analyst to determine if the response variance is approximately constant. If the response variance is not constant then the literature (23) suggests that it is often proportional to the response or its square.

The regression parameters of the calibration line are not affected greatly by the weighting factor chosen. However, an inappropriate weighting factor can have a large affect on the computed concentrations for samples containing a small amount of analyte.

A special case occurs when the variance is known from theory, e.g., X-ray spectrometry (66,67) where the variance is equal to the mean response measurement. Brownlee (29, p. 306) and Mandel (32, p. 146) have discussed this case. The analyst should however, keep in mind that the experimental error of the response can be due to many factors, e.g., for x-ray spectrometry the error variance is the sum of variances due to counting statistics, generator and other equipment errors. To this could be added errors due to sample containers and positioning. It is probably safest to confirm the theoretical error variance by experiment under the actual conditions of the analysis.

Methods of estimating the analytical error are discussed in section 4.5.
Responses have a Gaussian Distribution

Youden (68, Ch. 1) has addressed some of the reasons put forward by experimenters against the use of statistical tests. One favourite argument is that the tests assume that the measurements conform to a particular mathematical law (normal or Gaussian distribution). Youden's response is fourfold:

- large sets of scientific measurements usually do show a remarkable similarity to this distribution
- moderate departures from a normal distribution have only minor influence on limits
- the only alternative to using this close approximation is to go on guessing
- if the distribution is obviously not normal then collecting the observations into groups of three or four and using the averages as individual measurements (the central limit theorem) results in an approximately normal distribution.

Kaiser (69) in a paper on "Quantitation in Elemental Analysis" has taken a different stand. He advocates the use of descriptive statistics for the reduction and compression of analytical data. However, he does not support the use of the formally calculated values for the mean and standard deviation to state statistical significance, confidence limits, etc. as they would be correct only if the distribution is normal. This is illustrated with an example showing that 100 analyses (when the distribution is unknown) are required to obtain a certain statistical reliability whereas only 7 are required if the distribution is known to be normal. Kaiser indicates that he believes the normal distribution is not the usual case in chemical analysis. This statement is followed by an interesting discussion on the special role of the normal distribution.

Thompson and Howarth (70) and Harris (71) have examined the distribution of analytical error. Harris examined errors in the manipulative and measurement steps of an analysis and concluded that non-Gaussian distribution of error are to be
expected. Thompson examined the distribution of analytical error at concentrations near the detection limit; he concluded that the error distribution approached the normal, but that small deviations, especially a positive skew, are common.

It is necessary for the distribution of the residuals about the calibration curve to be normally distributed for the concentration limits to be correct. The program allows the distribution of these residuals to be examined by means of a histogram and to be checked for non-normality (sec. 2.2.2).

3.9.3 Overview of the Calibration Procedure

The calibration procedure is illustrated by the flowchart of Table 3.12; it is based on a figure by Mitchell et al. (17). This flowchart is intended to give only an overview of a possible procedure to follow.
## TABLE 3.12
AN OVERVIEW OF THE CALIBRATION PROCEDURE

<table>
<thead>
<tr>
<th>START</th>
<th>↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Select and analyze standards</td>
<td>↓</td>
</tr>
<tr>
<td>2. Select linear or nonlinear calibrating function</td>
<td>↓</td>
</tr>
<tr>
<td>3. Select weighting factor</td>
<td>↓</td>
</tr>
<tr>
<td>4. Calculate the calibration curve etc.</td>
<td>↓</td>
</tr>
<tr>
<td>5. Examine residuals, reject outliers, check distribution</td>
<td>↓</td>
</tr>
<tr>
<td>Go to step 2, or 3 as required</td>
<td>↓</td>
</tr>
<tr>
<td>6. Analyze samples, reject outlying replicates</td>
<td>↓</td>
</tr>
<tr>
<td>7. Calculate concentrations and discrimination limits</td>
<td>↓</td>
</tr>
<tr>
<td>Select appropriate standard measurements.</td>
<td>↓</td>
</tr>
<tr>
<td>Calculate regression equations, confidence bands, and analyte concentrations.</td>
<td>↓</td>
</tr>
<tr>
<td>Select optimum data.</td>
<td>↓</td>
</tr>
<tr>
<td>8. Band includes Concentration Zero? yes -----------) is less than the minimum detectable. no ↓</td>
<td>↓</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>↓</td>
</tr>
<tr>
<td>9. Go to 7 until the concentration of the last sample has been determined.</td>
<td>↓</td>
</tr>
<tr>
<td>END</td>
<td>↓</td>
</tr>
</tbody>
</table>
3.10 Summary

The practising analyst can use personal microcomputers for the regression analysis of analytical data. The microcomputer program can do a sophisticated linear regression analysis, including the use of the appropriate weighting factors. The most important advantage of this approach is that realistic limits for the computed concentrations can be calculated.

The statistical terms—prediction and simultaneous tolerance limits for responses, and non-simultaneous and simultaneous discrimination limits for concentration—have been used in this chapter. The analyst can, however, replace them by the more familiar term "confidence limits". But, it must be remembered that the "confidence limits" for many "predictions" made from a calibration function will be wider than the limits for a single "prediction" made from the same function.

Nonlinear curves can also be evaluated using programs designed for linear regression, if the "linear segmented" approach is used. Nonlinear curves could also be subjected to a more rigorous data treatment, although they were not in this particular study.

The required programs, i.e., software, are useful for determinations using the standard addition technique, as well as, the general comparative technique using linear calibration curves. The software is "user friendly"; it will produce results quickly, reanalyze the data using different weighting factors, and produce well formatted results. The use of the program with data having the two most common analytical error functions, constant standard and relative standard deviation, have been extensively illustrated.

The mathematical accuracy of computer results are subject to round-off errors. This program has been designed to minimize these errors. The analyst, however, is responsible for interpreting the results realistically, and for testing the program whenever it is used under new conditions.
This computing toolbox, i.e., the hardware and software, will not replace the experience and judgement of the analyst. It will, however, assist him/her in assessing analytical data and will sharpen his predictive skills. Perhaps the greatest benefit it offers is in assisting the analyst in producing more realistic estimates of analytical error.
Chapter 4
COMPARISON OF ANALYTICAL METHODS

4.1 Introduction

The previous chapter dealt with the calibration of an analytical instrument or method. That topic, considered more generally, can be expanded to compare the merits of alternate methods.

Rosenblatt and Spiegelman (1), in a discussion of the statistical problems associated with the calibration of measurement instruments and procedures, pointed out that Williams (2) had stressed that there are two activities called calibration. One is the "absolute" calibration where a measurement technique is calibrated against a standard or defined measurement (having negligible experimental error). The other is a "comparative" calibration where one instrument or measurement technique is calibrated against another, with neither one being inherently a standard.

For the practising analyst, the need to compare the results of different (or modified) methods usually results from the requirement for a more economic and/or technically superior analysis. The analytical journals show that analysts continually are comparing new methods with established ones. However, this same literature also shows that many of these comparisons can, at best, be described as "elementary". Usually the conclusions reached with these comparisons--about the superiority of one of the methods--are correct. However, the data used, and more importantly the actual statistical data analysis, may not fully support such conclusions.

It will be shown in this chapter that just as there are two activities called "calibration", there are also two activities called "comparative calibration". Two methods can be compared by linear regression for either "validation" or "merit" reasons. A "validation" comparative calibration is for the purpose of comparing the accuracy of a new method to that of an established one. A "merit" comparative calibration is for the purpose of determining the relative economic and technical merits of two methods.
More complete comparisons require a larger amount of data and a more complete statistical analysis. Such data is actually often accumulated, or could be obtained, from a comparison of methods. The data analysis requires the appropriate statistical techniques and a computing device. Again the microcomputer will be shown to be capable of doing the necessary calculations. It is these calculations, the related statistical theories, and their practical application in analytical chemistry which will form the major part of this chapter.

4.2 Definition of the Problem

Two (or more) methods of measuring the same chemical property are to be compared. The comparison will consider the technical and economic merit of the methods.

It will be assumed that the methods are completely specific. However, in the application of a comparison it must be kept in mind that this is an ideal situation (3) and therefore the merits of analytical methods will depend on the materials analyzed in the study.

When simulated data is used it will also be assumed that the materials "selected" for the comparison study are representative of the samples to be subsequently analyzed by the chosen method, and that they uniformly cover the concentration range of interest.

In this problem it will be the total analytical system (4) which will be considered. This system consists of a particular set of samples, the defined analytical procedure, and the particular instruments used. Thus, the variance of an analysis must include the variance of each step and not just the variance of the final measurements. Inherent in this concept is the necessity that each measurement be from a separate sample, selected randomly, and taken through the entire analytical procedure.

Lastly, it will be assumed that measurements by all of the methods are a steadily increasing or decreasing function of the chemical property being measured. This will allow a comparison of the methods without making specific use of the chemical property.
4.2.1 Technical and Economic Merit

In choosing between analytical methods, an analyst may often select the method which is thought to give the greatest accuracy (and precision). This choice may not, however, be warranted by the requirements of the use to which the results will be subjected. It may also not be the most economical.

The analyst's desire for accuracy is understandable, however, as Mandel (3) has stated the requirement of accuracy is not pertinent to comparative calibrations where a range of concentrations are considered and where reference materials, equivalent to the samples to be analyzed, are not normally available. Accuracy is a matter best considered in the validation (5) of each individual method. However, comparing the accuracy of methods (6) is a related topic of interest and will be discussed in section 4.6.

Technical Merit

Mandel (3) has developed a "Sensitivity Ratio" as a figure of technical merit for a comparative calibration. His procedure requires only the "raw" analytical responses and their standard deviations—the corresponding analytical results are not required. This ratio will be considered in detail in a later section. The following development will suffice for now.

Figure 4.1 illustrates a calibration of method M versus method N where measurements (responses) by both methods are an increasing monotonic function of the concentration of an analyte in a series of samples. The standard deviation of measurements by each method are SDm and SDn, respectively.

To develop a figure of technical merit consider two points on the curve (A and B) which represent two very similar concentrations. It would at first appear that method M is more sensitive than method N (since dM is greater than dN). However, the scales are not the same and each measurement is subject to experimental error. Therefore, the apparent advantage of method M may be offset if SDm is larger than SDn. Mandel suggests that a valid comparison of the methods can be obtained by comparing the ratios dM/SDm.
and \( \frac{dN}{SDn} \); whichever ratio is larger corresponds to the method which best differentiates between the two samples. A comparison of these two ratios lead to his technical figure of merit (relative sensitivity of method M with respect to method N). The absolute value of \( \frac{dM}{dN} \) is used to ensure a positive value for the RS.

\[
RS(M/N) = \frac{|dM/dN|}{SDm/SDn}
\]  

(1)

It will be shown in section 4.7.1, that the same relative sensitivity is obtained when the standard deviations of the analytical results, in concentration units, are compared, i.e.,

\[
RS(M/N) = \frac{SDn'}{SDm'}
\]  

(2)

where \( SDn' \) and \( SDm' \) are the SDs of the concentrations.

Fig. 4.1 A graphical development of Mandel's relative sensitivity (RS) ratio for Method M with respect to Method N.
Economic Merit

The economic merit of a method follows from the relative sensitivity (RS) which can be used to determine the ratio of measurements required by each method to obtain equal precision, and the "cost" of measurements by each method. The cost of a measurement is a function of factors such as time, skill, equipment, reagents and sample preparation required to obtain it.

It is obvious that an analysis using a method with poorer precision will have to be repeated in order to increase the precision of its mean result. The standard deviation of the mean is \( \text{SD}/\sqrt{n} \), where \( n \) is the number of replicates. Thus, for equal precision:

\[
\text{SD}_{m'} = \frac{\text{SD}_{n'}}{\sqrt{n}}
\]

therefore:

\[
\sqrt{n} = \frac{\text{RS}(M/N)}{\text{SD}_{m'}}
\]

and hence:

\[
n = \frac{\text{RS}(M/N)^2}{\text{SD}_{m'}}
\]

Thus, if \( \text{RS}(M/N) \) is 2 to 1, then one measurement by method M is equivalent (in precision) to \( 2^2 = 4 \) measurements by method N. However, if the ratio of the cost of the measurements (by M and N respectively) is 10 to 1 then obviously method N has greater "economic merit", i.e., an equally precise analysis by method N is cheaper (4/10) than one by method M.
4.3 Survey of Literature Comparisons

The Analyst, Analytical Chemistry, and Talanta were surveyed for examples of comparisons of analytical methods. Table 4.1 summarizes the results of this review for twenty-four such references. The ones chosen are not a random selection, however, it is thought that they give a indication of the comparison methods in use, and the criteria considered important in method evaluation.

4.3.1 Number of Methods Compared

In most comparisons the "new" and the "old" method, and occasionally three or four methods, were involved. Sometimes the new method was compared, in theory, to one or more old methods.

4.3.2 Advantages Claimed for the New Method

The advantages claimed in these references were divided into twelve groups and listed in order of greatest use.

Accuracy and Precision

In most cases it was claimed that the new method was as accurate and/or precise as the old. In a few cases the new method had less accuracy/precision, but did have other desirable attributes. Occasionally the new method had better accuracy/precision. Precision and accuracy relate to technical merit.

Speed of Analysis

A reduction in the length of time required for an analysis appeared to be a very desirable characteristic of a new method. The time required for an analysis would naturally determine the number which could be done in a given time, and thus would be related to the economic merit.
<table>
<thead>
<tr>
<th>No. of Methods</th>
<th>Advantage</th>
<th>Type of Comparison</th>
<th>SD RSD</th>
<th>SLR</th>
<th>New Old</th>
<th>Other</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1 3 5 8</td>
<td>2 5</td>
<td>yes</td>
<td>yes</td>
<td>unknown</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>1 2 4</td>
<td>1-2</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>1 2 4 5</td>
<td>1 2 3</td>
<td>yes</td>
<td>yes</td>
<td>-</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4</td>
<td>1 2 3</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>1 2 4</td>
<td>2 3 4</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>1 2 4 5</td>
<td>2 3 4</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>1 2 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>-</td>
<td>-</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>-</td>
<td>30</td>
</tr>
</tbody>
</table>

Advantage
1. Accuracy/precision
2. Speed
3. Sensitivity
4. Specificity
5. Skill required
6. Versatility
7. Reagent cost
8. Sample size
9. Equipment required
10. Sample preparation
11. Response characteristics
12. Process control

Type of Validation
1. Recovery experiment
2. Independent method(s)
3. Standard reference material
4. Method(s) of known accuracy
5. Collaborative test(s)

SD
1. At one concentration level for one method
2. At one level for two or more methods
3. At two or more levels for one method
4. At two or more levels for two or more methods

BLR
Simple linear regression used?

New/Old
Are the results of the new method regressed on those of the old?
NA = Not Applicable. NS = Not Stated.
Sensitivity, Specificity, and Versatility of the Method

A new method was often claimed to have greater sensitivity, specificity, and/or versatility. These attributes would be related to technical merit.

Skill Required, Reagent Cost, Equipment Required, and Sample Preparation

The skill required by the operator, and to a lesser extent reagent cost, equipment required, and sample preparation were also factors mentioned. Again, it's obvious that these are related to economic merit.

Sample Size

A consideration of the sample size could relate to economic merit in terms of reagents required or maybe even to technical merit. However, it is probably best considered as a "special merit". Sample size is particularly important where several analyses must be run on the limited amount of sample available, or where the source of the sample would be adversely affected by the removal of a larger amount, e.g., a blood sample from a newborn infant.

Response Characteristics

The response characteristics, e.g., a straight versus a curved calibration line, could reduce the complexity of processing the data. This would be related to economic merit and possibly also to technical merit.

Process Control

Automated methods developed for quality control often need not have as great a technical merit, if they are faster than the "batch method" of analysis. This would be an example where accuracy/precision could be sacrificed for speed and convenience.

4.3.3 Method Validation

Taylor (5) lists several ways for validating an analytical method. They consist of analyzing a series of
samples and comparing the results from the "new" method with the:
- certified values of standard reference materials known to be similar to the samples to be analyzed by the new method
- results from an applicable and reliable method
- results from any independent method
- expected values of synthetic or spiked samples

The comparison of analytical results can therefore be part of the validation process for a new method. However, it can also be used to quantify the technical merit of the new method relative to the old or comparison method.

The survey indicates that the comparisons reviewed, and therefore perhaps most comparisons published, are used only for the "validation" of new methods.

4.3.4 Statistical Analysis Used

Each comparison used in the survey was reviewed to determine what statistical analysis was conducted to support the conclusions reached.

Calculation of Standard Deviations

In almost all of the comparisons surveyed, only a limited study of the standard deviations, and their variation with analyte concentration was conducted. Since many of the comparisons were based on simple linear regression analysis (which assumes a constant standard deviation for the y-variable and no error in the x-variable) this lack of interest in the precision of the methods is amazing.

Use of Simple Linear Regression (SLR)

SLR was used for half of the comparisons surveyed. It appeared that the results for the new method were often regressed on the results of the old method, without any reason for doing so being given. In some papers the authors did not indicate how the regression was conducted (e.g., 13,27). Reprocessing the data from these two papers indicated that the new methods were regressed on the old.
(Refer to Table 4.10 for the consequences of misapplications of SLR.)

Other Statistical Methods Used

Other methods which might have been used include t-tests, ANOVA, and other forms of regression analysis such as weighted or maximum likelihood linear regression. Two paired t-test comparisons were found.

4.3.5 General Conclusions

Usually the accuracy of a newly developed method was compared to that of the previously used one. The statistical testing often consisted of a simple linear regression of the results of the new on those of the old method. The standard deviations of each method and its variation with concentration (or response level) was generally not seriously considered. A paired t-test was occasionally used, and to a much lesser extent ANOVA (31).

Factors related to the economic merit of the new method were generally a prime concern. The main factors being time and skill required for the analysis, and to a lesser extent the cost of reagents and equipment.

It appeared, however, that no attempts are made to quantify the merits (technical or economic) of the new method to those of the old.

The survey supports the view that the comparison of methods for validation purposes, and for quantifying their relative technical and economic merits is a topic requiring clarification and additional study.

4.4 Comparison of Accuracy using "t" Tests and ANOVA

Both t-tests and ANOVA have been described and illustrated in chapter 2. Many of the examples used involve the comparison of accuracy of analytical methods. Therefore the use of these tests will only be outlined in this section.
4.4.1 "t" Tests

t-tests require the actual computed concentration for each sample. The tests would therefore be used for the validation of analytical procedures rather than a comparison of technical and economic merit.

- the "known" or "expected" value of a standard reference material (SRM) could be compared to the average result obtained by a method. The variance used would be that obtained in the specific test, or the variance obtained from previously conducted experiments. The disadvantage of the approach is the need to have available a number of SRMs, each with a matrix comparable to the samples. Also a number of analyses of each sample would be required, as well as, a t-test for the data from "each" sample.

- the general t-test could be used to compare the results obtained for one sample by two methods of analysis. If several samples were available then one t-test for the data, from each sample, would be required. The use of samples of different concentrations (the exact concentration would not be required) would allow a comparison of results (and perhaps the detection of a bias) across the entire range of interest. The disadvantage of this procedure is the requirement of a t-test for the data from "each" sample, and the number of replicate analyses (experimental work) required for each.

- a single, general t-test could perhaps also be used to compare the "% recovery" of spiked samples (of different concentrations) analyzed by each method. The disadvantage of this approach is that the distribution of the "% recovery" values would probably not be normal (the test assumes normality).

- a "paired" t-test would probably be preferable for the analysis of "% recovery" data. The test could also be used to compare the paired results obtained from samples covering the range of interest (the exact concentrations
of samples would not be required). This approach has the advantage of requiring only one analysis for each sample by each method, and requiring only one t-test. A disadvantage is that the degrees of freedom used in the test is the number of "pairs" minus one, therefore a large number of samples must be used to detect a small difference between the methods. In addition, it is likely that the technique would be inappropriate whenever the range of the concentrations is wide and the precision of the methods is a function of concentration.

4.4.2 ANOVA

The method of ANOVA also requires the actual computed concentration for each sample used in the study. One advantage is that several methods can be compared; the test will determine whether or not all the methods can be considered equivalent. A disadvantage is the need for the variance of each method to be equivalent and to be independent of analyte concentration. Determining whether these conditions are met requires additional work; it is probable that the conditions will "not" be satisfied.

4.4.3 Conclusions

These tests have limited use in the comparison of analytical methods. They would be useful only for the validation of accuracy. It will be shown that linear regression analysis is a preferrable statistical technique because it allows the calculation of relative technical and economic method, and because it more easily takes into account the variation of method precision with concentration.

4.5 The Variation of Error

Standard Deviation with Concentration

It has been shown (sec. 4.2.1) that knowledge of the standard deviation of analytical methods is required to determine their relative merits. As well, it has been noted (sec. 4.4) that t-tests and ANOVA generally assume a constant standard deviation for the data, i.e., the error
variance of the analysis must be independent of concentration. It has also been shown (Table 4.1) that in many published comparisons the variance of the methods is not fully determined. Finally, it will be seen that the proper use of linear regression analysis demands fairly extensive knowledge of a method's error variance and its variation with concentration or response level.

4.5.1 The Standard Deviation Function

The point has already been made (ch. 3) that the standard deviation of an analysis tends to vary with the concentration or response level. Thompson and Howarth (4) have also stated this tendency, and have developed an expression for this variation. An important concept which they stress is a fixed standard deviation at zero concentration. They begin by expressing the variation as a function of concentration:

\[ SDc = f(c) \]  \hspace{1cm} (6)

where SDc is the standard deviation at concentration c. The expression is then expanded:

\[ SDc = k_0 + k_1c + k_2c^2 + ... + k_n c^n \]  \hspace{1cm} (7)

where \( k_0, k_1, k_2, \) and \( k_n \) are constants. The first term \( k_0 \) can be replaced by SD(0), the standard deviation at zero concentration. Ignoring all but the first order concentration term results in the following expression:

\[ SDc = SD(0) + k_1c \]  \hspace{1cm} (8)

Figure 4.2 shows the variation of the total standard deviation (SDc) and the RSD with concentration which would be predicted by this expression. Sthapit et al. (13) have produced a very similar figure from experimental data; thus indicating that this type of variation can be obtained in practice.
If SDo were small, and if the analytical system was being used at concentrations much greater than the detection limit, then the expression could be further simplified to:

\[ S_{Dc} = k1\times c \]  \hspace{1cm} (9)

The RSD would therefore approach:

\[ \text{RSD} \rightarrow \frac{S_{Dc}}{c} = \frac{(k1\times c)}{c} = k1 \]  \hspace{1cm} (10)
Fig. 4.2 Variation of the total standard deviation (SDc) and relative standard deviation (RSD) with analyte concentration. The analytical system has a small limiting SD at "zero" concentration, and an essentially constant RSD at high concentrations. (DL = detection limit concentration)
4.5.2 Determining the Error Function

The lack of a convenient and easy to apply technique for the determination of analytical error functions is evident from a study of the literature. In this section four approaches will be considered. The first two have already been described in the literature. The third, is based on the assumption of either a constant analytical error standard deviation or relative standard deviation, and standard reference materials of known concentration. The fourth, is based on the assumption of either a constant error SD or RSD, and duplicate (or multiple) analyses; the concentrations of the samples are not required. These last two approaches do not seem to have been described previously in the analytical literature.

I Replicate Analyses

The most straightforward approach to obtaining the error variance function would seem to be the analysis of replicate samples at each of a number of concentration levels which are uniformly distributed across the range of interest. Schwartz (32), as was previously noted, suggests that twenty or more replicate samples, at each of six or more concentration levels, should be analyzed in order to determine the variance function.

Thompson and Howarth (4) have commented on this procedure. The following are several important points which should be considered whenever this approach is adopted:
- the samples to be analyzed for the study should be similar to those which will be analyzed later by the method.
- the samples must be taken through the complete analytical system.
- it may be difficult to obtain a large enough uniform sample from which to obtain the required number of replicates.
- if the samples are analyzed contiguously, and particularly if they are recognized as such by the analyst, then the variance estimate will probably be low.
- a large number of replicates are required to obtain a good estimate of the variance.
- obtaining the necessary number of replicate results is probably an inefficient use of analytical resources.

II Duplicate Analyses

Thompson and Howarth (4) have suggested that a better approach is the use of duplicate analyses where each sample is analysed twice by each method. The advantages claimed are:
- results truly represent the samples.
- the preparation of standard materials is not required.
- no data are wasted, averaged duplicates improve precision of results.
- conclusive information on error variance is obtained at various analyte concentrations.

To obtain proper error variance estimates each duplicate must be taken through the complete analytical system as if it were a separate sample, and the sequence of analysis must be random.

Assuming

\[ \text{RSD} = \frac{\text{SD}_c}{c} \]  

(11)

then combining this expression with equation (8) gives:

\[ \text{RSD} = \frac{\text{SD}_0}{c + k_1} \]  

(12)

It can be seen from this expression (and Fig. 4.2) that the RSD rises rapidly as the concentration approaches zero, and that it approaches k1 at high concentrations. Thompson and Howarth (4) have shown that for values of k1 ranging from 0.15 (15% RSD) to 0.0005 (0.05% RSD) an analyte concentration of approximately 30 to 10,000 times the detection limit is required before the RSD approaches to within 10% of k1. Supporting evidence can be obtained from the figure by Sthapit et al. (13) referred to earlier; it shows a constant value of RSD at a concentration of 300 ug/l, which is 50 times the detection limit of approximately 6 ug/l.
The detection limit referred to is DL, where DL is a concentration of 2\(\times\)SDo (the corresponding RSD is 0.5, see Fig. 4.2). In general it is suggested that the RSD approaches to within 10\% of its limiting value (k1) when the concentration is approximately 5/k1 times the detection limit.

The above authors have shown how both SDo and k1 can be estimated using duplicate analyses obtained from one analytical system. They state that a minimum of fifty pairs of duplicates are necessary. [Although not directly related to this application, the use of duplicates to estimate standard deviations has also been described by Kaiser (33).]

The concentration range is usually divided into 11 narrower ranges. For each narrow range the median, of the absolute differences between members of each pair within that range, is determined. This median is multiplied by 1.0483 to produce an estimate of the standard deviation (y-variable) for the arithmetic mean (x-variable) of all concentrations within that narrow range.

The relationship between the estimated standard deviations and the mean concentrations is then determined by linear regression analysis. The x-intercept of the resulting line is SDo, the standard deviation at zero concentration, and the slope is the value of k1 (see eq. 8). SLR is suggested as being adequate for this analysis although WLR would probably produce a better estimate of the intercept, SDo. Howarth and Thompson (34) have used Monte Carlo simulation to test the robustness of the method. The method has also been applied to data generated for section 4.9 and results of its application are listed in Table 4.11.

III Regression of Response on Concentration: Assumption of Constant Error SD or RSD, and Known Concentrations

It has already been shown (ch. 3) that weighted linear regression can be conducted without knowing the error variances, if the RSD is constant. When the concentration range of interest is well above the detection limit, a few replicate analyses (or prior knowledge) may be sufficient to justify an assumption of an approximately constant error
The standard error of regression (as computed by the WLR routine) would then approximate the relative standard deviation of the error for that method.

Supporting evidence for this concept can be found in section 3.6 and Table 3.9. For that particular analysis, data was collected from two sets of simulated calibration experiments. In each set five calibrations were conducted for which the expected RSD was 1%. In one set of experiments twenty standards were used and triplicate analyses were conducted; the values of the computed RSDs were 0.76, 0.77, 0.86, 0.95, and 1.00%. In the other set nine standards were used and duplicate analyses were conducted; the values of the computed RSDs were 0.78, 0.81, 0.92, 0.96, and 1.07%. These values correspond to the constant \( k_1 \) referred to by Thompson and Howarth (4) (see Fig. 4.2).

This technique has many of the advantages of the "duplicate analyses" approach while being faster and more convenient to use. It is probably also able to produce an estimate of the limiting precision (\( \%\text{RSD} \)) from a smaller amount of data. As shown above, as few as nine standards and duplicate analyses produced a reasonable estimate of the RSD. It could not, however, produce an estimate of the standard deviation for a "zero concentration".

A similar approach could also be used with SLR to determine the standard deviation of a system which is assumed to have a constant analytical error. An additional use might be to estimate the standard deviation (SDo) at the "zero concentration" of a system in which the total standard deviation is actually a function of concentration. Assuming the data used were from SRMs with concentrations close to the detection limit, then it could be expected that the standard deviation would be approximately SDo as \( k_1c_1 \) would be approximately zero (see eq. 8).

A further use of this approach would be to estimate the constant SD of a nonlinear response. In fact all four approaches might be useful in this regard.
IV Self Regression: Assumption of Constant Error SD or RSD, Replicate Analyses, and Unknown Concentrations

The experimental work required by this approach is identical to that of the "Duplicate Analyses" technique except that triplicate or other multiple analyses could also be used. In the data analysis the values are not grouped, but rather the data set is regressed on itself. One would expect to obtain a slope of 1.0 and an x-intercept of 0.0 from this regression, if the data were not "rearranged".

The "theory" of this approach is simply an attempt at explaining a calculation which seems to give good estimates of error SD and RSD for simulated experimental data.

In the usual SLR or WLR the experimental error is assumed to be in only the y-variable. Under these conditions it has been shown that this error (SD or RSD) is estimated by the standard error of the appropriate regression. When both variables have an error variance, and the variances are equal, SLR and WLR are still valid techniques. In this situation it will now be assumed that the standard error of regression (SE regression) will be "made up of" both error variances. The mathematical approach for the SLR case is as follows:

\[
\text{Variance (total)} = \text{Variance (x)} + \text{Variance (y)} \tag{13}
\]

but \[
\text{Variance (y)} = \text{Variance (x)} \tag{14}
\]

thus \[
\text{Variance (total)} = 2 \times \text{Variance (y)} \tag{15}
\]

if \[
(\text{SE regression})^2 = \text{Variance (total)} \tag{16}
\]

then \[
\text{Variance (y)} = (\text{SE regression})^2 / 2 \tag{18}
\]

and \[
\text{SDy} = \text{SE regression} / \sqrt{2} \tag{19}
\]
Mandel (3) in a discussion on the "law of propagation of errors" states that for any function of a purely multiplicative form, that the total relative standard deviation is the sum of the individual relative standard deviations. Assuming therefore that the square of the RSDs can be substituted in equations 13 to 19, leads to the following equation for WLR:

\[ \text{RSD}_y = \frac{\text{SE regression}}{\sqrt{2}} \]  

(20)

The usefulness of this approach should be based on the results it produces and not necessarily the above reasoning. As an example of its usefulness, the data previously used with Approach III was reprocessed; "Self Regression" produced the following results:

- 0.848 0.812 0.956 0.835 0.926 (20 standards, 3 replicates)
- 0.706 0.963 0.436 0.536 0.470 (9 standards, 2 replicates)

where the expected %RSD was 1%.

Section 4.9 and 4.10 will also show that it can produce "good" estimates.

4.5.3 Example: "Replicate Analyses" versus "Regression of Response on Concentration" Approaches to Calculating the Precision of an Analytical Method

Sthapat et al. (13) appear to have used the "Replicate Analyses" approach to obtain the standard deviation function. The data (their Fig. 5) indicates a limiting RSD of 1% for their procedure B. An independent estimate of the limiting % RSD for the same procedure using data from their Table V (procedure 2) was conducted using the "Assumption of Constant Error RSD, and Known Concentration" approach.

Regressing the results of procedure 2 on the means for the quality control samples resulted in a RSD of 11.36% for the method. One value appeared to be an outlier (631 for a sample expected to be 455); this value was deleted and the data was reprocessed to obtain a limiting RSD of 7.32%. Table 4.2 lists the relevant data and Fig. 4.3 shows the results of the WLR analysis.
A possible cause for the difference in the % RSDs (the published versus this new one) could be due to the concentrations of the quality control samples not being known exactly; WLR requires that they be known without error. This would increase the estimated RSD, but probably not sufficiently to explain the wide difference noted. Perhaps the procedure used by Sthapit et al. for obtaining the standard deviations resulted in them being underestimated. One would also intuitively expect that determining the % RSD for the entire concentration range simultaneously, as opposed to individual estimates at single concentrations, would result in a higher and probably more realistic estimate of % RSD.
TABLE 4.2
RESULTS OF A % RSD DETERMINATION USING THE "REGRESSION OF RESPONSE ON CONCENTRATION" APPROACH

<table>
<thead>
<tr>
<th>QC Sample</th>
<th>Result by Procedure 2</th>
<th>Difference</th>
<th>+ 1 SD</th>
<th>- 1 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conc'n</td>
<td>Conc'n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>240</td>
<td>256</td>
<td>16</td>
<td>237</td>
<td>275</td>
</tr>
<tr>
<td>248</td>
<td>243</td>
<td>-5</td>
<td>228</td>
<td>261</td>
</tr>
<tr>
<td>302</td>
<td>289</td>
<td>-13</td>
<td>266</td>
<td>310</td>
</tr>
<tr>
<td>306</td>
<td>292</td>
<td>-14</td>
<td>271</td>
<td>313</td>
</tr>
<tr>
<td>382</td>
<td>418</td>
<td>36</td>
<td>387</td>
<td>449</td>
</tr>
<tr>
<td>385</td>
<td>398</td>
<td>13</td>
<td>369</td>
<td>427</td>
</tr>
<tr>
<td>440</td>
<td>472</td>
<td>32</td>
<td>437</td>
<td>507</td>
</tr>
<tr>
<td>441</td>
<td>411</td>
<td>-30</td>
<td>381</td>
<td>441</td>
</tr>
<tr>
<td>472</td>
<td>407</td>
<td>-65</td>
<td>377</td>
<td>437</td>
</tr>
<tr>
<td>495</td>
<td>514</td>
<td>19</td>
<td>476</td>
<td>552</td>
</tr>
<tr>
<td>600</td>
<td>605</td>
<td>5</td>
<td>561</td>
<td>649</td>
</tr>
<tr>
<td>640</td>
<td>564</td>
<td>-76</td>
<td>523</td>
<td>605</td>
</tr>
<tr>
<td>646</td>
<td>633</td>
<td>-13</td>
<td>587</td>
<td>679</td>
</tr>
<tr>
<td>691</td>
<td>718</td>
<td>27</td>
<td>665</td>
<td>771</td>
</tr>
<tr>
<td>772</td>
<td>862</td>
<td>90</td>
<td>799</td>
<td>925</td>
</tr>
<tr>
<td>816</td>
<td>854</td>
<td>38</td>
<td>791</td>
<td>916</td>
</tr>
<tr>
<td>820</td>
<td>852</td>
<td>32</td>
<td>790</td>
<td>914</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F RATIO</th>
<th>Prob (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explained by regression</td>
<td>1</td>
<td>2.44294</td>
<td>2.44294</td>
<td>456.175</td>
<td>0.000 #1</td>
</tr>
<tr>
<td>About regression</td>
<td>15</td>
<td>0.0803289</td>
<td>5.33326E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>2.52327</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EDN OF LINE Regression Coeff. = .963954
y = 1.01926 x + -5.84122
% Explained Variation = 96.8165

Standard Error of Regression (Sy) = .0731797

Confidence Limits (90, 95, 99%) of Intercept: -5.84122 +/- .0856372 .101712 .140355
Confidence Limits (90, 95, 99%) of Slope: 1.01926 +/- .0477222 .101712 .140355

Fig. 4.3 Results of a WLR to obtain the % RSD of an analytical method using the data of Table 4.2.
4.6 Validation of Method Accuracy with Linear Regression Analysis

Thompson (6), through the use of Monte Carlo simulation, has produced a very practical study of the robustness of SLR, WLR, and MLR analysis for the comparison of accuracy using the "comparative calibration" technique. In this context "robustness" refers to the ability of regression analysis to give approximately correct predictions despite deviations from the theoretical assumptions of the technique, e.g., normal error distribution for the y-data set (sec. 4.1). The results of his study are important as they provide guidance in the use of linear regression techniques where the results of both methods (the x- and y-variables) are each subject to experimental error.

As was already shown (Table 4.1) linear regression, and particularly SLR, has been the most common statistical technique used to compare the accuracy of two methods. The theoretical requirements of SLR are that the x-values be known without error, and that the error variance of the y-values is homoscedastic. Thompson's conclusion that SLR may give accurate estimates of bias between methods, if the results from the method with the smaller variance are used as the independent variable (x-values), will be useful to those analysts who must continue to use SLR. However, analysts should have been aware of this requirement previously as the same fact was presented by Mandel (3) in 1964. Thompson states that two additional requirements for the valid use of SLR are that ten or more samples must be used, and that their concentrations must cover the concentration range uniformly from zero upwards.

4.6.1 Requirements of the Technique

Since the technique is used for the comparison of accuracy the responses obtained for each method must be converted to concentration, and the error variance of each method (or the variance ratios) must be known at least approximately.
4.6.2 Usefulness of the Technique

Comparative calibrations using linear regression analysis allows a test for:
- significant intercept, resulting from "translational" bias
- significant deviations in the slope from unity, resulting from "rotational" bias
- significant deviations from linearity, perhaps resulting from unexpected "curvature" in one of the calibration curves.

Figure 4.4, an adaptation of figures by Thompson (6), illustrate the above conditions of bias.

It has been shown in chapter 3 that the computer program is capable of determining all of the above. This will be further illustrated in the experimental section of this chapter.

4.6.3 Maximum Likelihood Regression

Maximum Likelihood Regression (MLR) can be considered as the third method of linear regression. It follows WLR in terms of increased usefulness; SLR is the least useful of the three techniques. The main advantage of MLR is that, unlike SLR and WLR, the x-values are not assumed to have zero variance. It would therefore seem to be the ideal technique to use in comparing the accuracy of methods. However, having to know the exact error variance of each method, for each sample, limits its usefulness in routine situations. One might suspect that knowing the proportional error function (as was the case with WLR) of each method (variable) could allow a simpler application of the technique. This idea was not pursued in this research.

Mandel (3) and Irvin and Quickenden (35) have also discussed linear regression analysis where there are errors in both the x- and y-variable. They each give equations for the general least squares treatment.
Fig. 4.4 Bias between methods X and Y.

A: Constant or "translational" bias caused by an uncorrectable systematic error in one method.

B: Variable or "rotational" bias caused by an incorrect slope of one of the calibration curves.

C: A "combination" of translational and rotational bias.

D: A "nonlinear" bias perhaps caused by an unknown curvature in one of the calibration curves.

The broken lines indicate a comparative calibration curve with a slope of one and an intercept of zero. This curve would be expected when both methods give accurate results.
4.6.4 The Conditions under which SLR, WLR, and MLR can be used for the Comparative Calibration of Analytical Methods

Thompson's (6) tests of the robustness of the three regression techniques, for the comparison of accuracy of analytical methods, leads to the conclusions presented in Table 4.3. A reasonable assumption to be drawn from these conclusions is that the method of WLR, as developed in this thesis and using the appropriate proportional error weighting function, is a more appropriate technique than SLR and is a simpler approach than MLR for the comparative calibration of method accuracy.

4.7 Quantifying the Technical and Economic Merit of Analytical Methods

The technical and economic merit of analytical methods was discussed briefly in section 4.2.1. The concept will now be more fully discussed. In particular, it will be shown that Mandel's technique for determining the technical merit of analytical methods (systems) has the following distinct advantages:

- the "raw" response data can be used directly
- no recourse to the calibration curves, in terms of the analyte, is necessary
- transformation of the response data does not change the value of the technical merit (relative sensitivity ratio)
- the "sensitivity ratio" can reflect the variation of technical merit with concentration
- the relative economic merit of an analytical method can be determined from a knowledge of the technical merit.

In addition the technique would be useful in the statistical comparison of several methods (all for measuring the same analyte in the same samples) when no referee method is available. Lawton, Sylvestre, and Young-Ferraro (36) have discussed this particular situation in regards to clinical chemistry. The discussions of this paper in the same issue of "Technometrics", particularly the discussion by Fung and Hunter (37), elaborate on the topic. They have compared this problem to "that of selecting the best archer from among a
group of archers, on the basis of data indicating where their shots had landed while shooting at a target, when the position of the target is unknown.

**TABLE 4.3**

<table>
<thead>
<tr>
<th>CONDITIONS UNDER WHICH SIMPLE, WEIGHTED, AND MAXIMUM LIKELIHOOD LINEAR REGRESSION CAN BE USED FOR THE COMPARISON OF ACCURACY OF ANALYTICAL METHODS</th>
</tr>
</thead>
</table>

SLR gives accurate estimates of bias between methods when:
- ten or more samples are used
- sample concentrations range uniformly from zero upwards, and
- the method with the smaller error variance is used as the independent (x) variable.

If the analytical error variances are homoscedastic then the estimated standard errors of the intercept and slope are similar to those obtained by WLR and MLR. If the error variances are heteroscedastic then the estimates by SLR are high!

WLR can accurately detect a smaller bias than SLR when:
- the constraints are the same as for SLR
- the error variances are heteroscedastic, and
- the analytical error variance is known for the dependent (y) variable.

MLR generally gives results similar to WLR and:
- it is free of the constraints of SLR and WLR, however
- the analytical error variances must be known for both variables (x and y)

It is therefore the safest, but most demanding method for general use.
It was shown (sec. 4.2.1) that Mandel's figure of technical merit, the relative sensitivity ratio of method M with respect to method N, could be expressed as:

\[ RS(M/N) = \frac{\log_{10} dM/dN}{SD_m/SD_n} \]  

Thus, in the validation of a new analytical method against a reference method (using a series of samples covering the concentration range of interest) the relative technical merit could probably be determined from the information which would be available. This would be a straightforward application as \( dM/dN \) would normally be "1". But it should be noted that in this situation the raw response data would have been already converted to concentration data. This conversion, for each method, would involve considerable effort.

The general application of Mandel's technical merit determination would be useful when it is required to select between two (or more) analytical methods which are already known to be suitable in terms of accuracy or when the only choice is to compare the new method to an independent method (sec. 4.3.3). The analyst would be attempting to determine which system had a higher technical and/or economic merit for the particular samples, instruments, and other laboratory conditions at hand. Mandel's technique, using the raw response data, would allow a faster evaluation of the systems.

4.7.1 Mandel's Relative Sensitivity Ratio

As was already indicated, if two methods are to be compared then the response data for each could be converted into analyte values. The method yielding the lower standard deviation, in analyte concentration terms, would be the more precise. Mandel (3) has shown, however, that the raw response data does not have to be converted to concentration. The following is an abbreviated derivation of his technical merit "relative sensitivity ratio".
If $M$ is the response measurement and $Q$ the concentration of the analyte then:

$$M = f(Q)$$  \hspace{1cm} (21)$$

This equation represents the calibration curve of the method. To determine $Q$ from a measured response $(M)$, the equation is inverted, i.e., $Q''$ is a function of $M$.

$$Q'' = g(M)$$  \hspace{1cm} (22)$$

But $M$ is a measured value (with an error of $m$) therefore, from the law of propagation of errors the standard deviation of $Q''$ is: (this law was also used in sec. 3.4.1 for transformed weights)

$$SD(Q'') = |dg/dM| \times SDm$$  \hspace{1cm} (23)$$

Since $g$ is the inverse function of $f$, therefore:

$$SD(Q'') = SDm/|df/dQ|$$ \hspace{1cm} (24)$$

The standard deviation of the analyte concentration $(Q'')$ can be found from the standard deviation of the response and the tangent to the calibration curve (the derivative of the function $f$).

If two analytical systems are used for the determination of the same concentration $Q$, there will be two calibration functions:

$$M = f(Q)$$  \hspace{1cm} (25)$$

$$N = h(Q)$$  \hspace{1cm} (26)$$

where $M$ and $N$ represent response measurements from each system.
If
- m and n represent the errors of measurement for M and N
- \( Q^M \) and \( Q^N \) represent the concentration estimates by each method, and
- \( fQ' \) and \( hQ' \) represent the derivatives of function Q
then:

\[
\begin{align*}
SD(Q^M) &= \frac{SDm}{|fQ'|} \quad (27) \\
SD(Q^N) &= \frac{SDn}{|hQ'|} \quad (28)
\end{align*}
\]

The method with the smaller standard deviation for the concentration will be the method with the greater technical merit. Comparing the ratios of the concentration standard deviations results in:

\[
\begin{align*}
SD(Q^M) &= \frac{SDm}{SDn} \quad (29) \\
SD(Q^N) &= \frac{|fQ'|}{|hQ'|}
\end{align*}
\]

The ratio of the derivatives (\( fQ'/hQ' \)) taken with respect to the same quantity Q (the concentration of the analyte) has a simple interpretation. Since the responses M and N are both related to the concentration Q, they are functionally related to each other. Thus, the derivative \( dM/dN \), i.e., \( fQ'/hQ' \), can be obtained from a plot of M versus N. Therefore equation 29 becomes:

\[
\begin{align*}
SD(Q^M) &= \frac{SDm}{SDn} \quad (30) \\
SD(Q^N) &= \frac{|dM/dN|}{|fQ'|/|hQ'|}
\end{align*}
\]

This equation now shows that the ratio of the standard deviations of two analytical systems, expressed in the same concentration units, can be determined without using calibration curves to first convert the response data to concentration units.

From an experimental point of view a series of samples covering the concentration range of interest are measured by each method. Sufficient replicate measurements are obtained to estimate the standard deviation with concentration function, or one of the other approaches described in
section 4.5.2 could be used. A plot of the M versus N averages yields the ratio dM/dN. With this data the ratio of the standard deviations of the concentrations can be estimated for any concentration level.

The Sensitivity Ratio

It can be seen from eq. 30 that as the ratio SD(Q"M)/SD(Q"N) increases, the technical merit of system M decreases with respect to that of N. The reciprocal of this ratio therefore expresses the "relative sensitivity of method M with respect to method N".

Relative Sensitivity \( \frac{M}{N} \) = \frac{\text{ldM/dN}}{\text{SDm/SDn}} \quad (31)

When the relative sensitivity (RS) ratio exceeds "1", method M is technically superior to method N, i.e., it has a greater ability to detect a real difference in concentration.

If more than two analytical systems are being considered then the value "one" can be assigned to the absolute sensitivity of any one of the systems, and the RSs of the other systems can be determined with respect to the selected one. The resulting RSs can be used to rank the systems with respect to their technical merit.

The Relative Sensitivity Curve

It has already been established that the standard deviation of the analytical error is not always constant; it often varies with concentration. Therefore, it could also be expected that the RS of a method may also not be constant as it depends on the standard deviation of the analytical error for both methods, as well as, the ratio dM/dN. A plot of RS versus the concentration (or M or N) results in a curve which shows the variation of RS with concentration.

Transformation of Scale

An interesting and important feature of the RS ratio is that it is not affected by a transformation of scale of
either or both responses. For example, the comparison of two methods where the responses are transmittance and emission level. In this case the average transmittances and their standard deviations can be transformed and then used directly in the calculation of the RS.

If the transformation is given by:

\[ M^* = f(M), \quad \text{then} \]

\[ \frac{\text{RS}(M^*/N)}{\text{SD}(M^*/N)} = \frac{\text{RS}(M/N)}{\text{SD}(M/N)} = \frac{|dM^*/dN|}{\text{SD}(M/N)} = \frac{|dM/dN|}{\text{SD}(M/N)} \]  

The sensitivity ratio, unlike the standard deviation, relative standard deviation, variance, range, and the confidence limits is the only measure of precision which is invariant with respect to any transformation of scale.

4.7.2 Steps in the Application of Mandel's Relative Sensitivity Ratio

The following are the steps in determining Mandel's relative sensitivity ratio of two analytical systems:

1. Select two appropriate analytical methods (M and N).
2. Verify that a functional relationship exists between the quantities representing the two responses.
3. Study the experimental errors affecting both systems.
4. Transform the data as required to obtain a linear relationship between the paired responses. The standard deviations must also be transformed (section 3.4.1)!
5. Plot the above data and determine dM/dN (i.e., the slope of the above line).
6. Plot the ratio, SDm/SDn, versus one set of the (transformed) response measurements, and determine the relationship between the standard deviation ratio and response level.
7. Calculate the relative sensitivity ratio for the response (concentration) level of interest.
4.7.3 A Textbook Example of Technical Merit

A textbook example by Mandel (3, p. 375) was reworked and is presented here as a bridge to the simulated experiments and data processing of the next section. The measurements are from an actual experiment; some details of the problem have been left out as they are not necessary for the purpose at hand.

In this problem, two testing methods for the characterization of rubber are available [Step 1]; one is a strain measurement and the other a stress measurement. Eighteen samples of rubber were measured four times by each test method. The average of each set of four measurements, and their standard deviations were then calculated (Table 4.4). When the averages of the strain test were plotted against the averages of the stress test, a hyperbolic type of relation was observed. Thus, the two tests were functionally related to each other [Step 2].

Table 4.4 lists the data in increasing order of strain; this aids in examining the error standard deviation and the percent relative standard deviation [Step 3]. The data was also plotted (Fig. 4.5). It can be seen that the error SD of the strain measurements is a function of their intensity. Although the range of the strain measurements is limited (in particular they do not include zero), the variation of the SD is somewhat in accord with the model suggested by Thompson and Howarth (4) in which there is a limiting error SD (eq. 8). Both Table 4.4 and Fig. 4.5 indicate that the error standard deviation of the stress measurements can be assumed to be approximately constant.

Thompson (6), in his study of linear regression, considered only two possibilities for the variation of standard deviation. They were (i) both methods had a constant SD, and (ii) both methods had a constant relative SD. This example indicates that mixed models are also possible.

A plot of the log(strain) versus log(stress) resulted in a straight line relationship and allowed the calculation of \( d\log(\text{strain})/d\log(\text{stress}) \) which was \(-1.838\) [Steps 4 and 5]. The "transformed" standard deviations were then
calculated:

\[ \text{SD}[\log(\text{strain})] = \text{SD}(\text{strain})/\left[\ln(10) \times \text{strain}\right] \] (34)

\[ \text{SD}[\log(\text{stress})] = \text{SD}(\text{stress})/\left[\ln(10) \times \text{stress}\right] \] (35)

The ratio of the transformed standard deviations were then plotted against \( \log(\text{strain}) \). The equation of the resulting line was [Step 6]:

\[ \frac{\text{SD}[\log(\text{strain})]}{\text{SD}[\log(\text{stress})]} = -3.595 + 2.335 \times \log(\text{strain}) \]

The sensitivity ratio of the strain test with respect to the stress test was therefore [Step 7]:

\[ \text{RS}[\log(\text{strain})/\log(\text{stress})] = \frac{1.838}{-3.595 + 2.335 \times \log(\text{strain})} \]

But since the relative sensitivity is invariant with respect to transformations, therefore:

\[ \text{RS}[\text{strain/stress}] \] also equals \( \frac{1.838}{-3.595 + 2.335 \times \log(\text{strain})} \)

As can be seen the RS is expressed as a function of the strain measurement. The strain measurements varied from 56.8 to 284.4; the corresponding values of RS are 3.67 and 0.86. Thus, it can be seen that the precision of the strain test is superior over most of the range of interest.

SLR was used to calculate all of the linear relationships, with stress being the x-variable. This was somewhat reasonable as the error standard deviations of the log(measurements), both stress and strain, were similar (Table 4.4). However, it would have been more reasonable to use the strain and its log as the x-variable; the RS ratio indicates that it is the more precise measurement.
TABLE 4.4
DATA FROM AN EXAMPLE
USED TO ILLUSTRATE
THE CALCULATION OF TECHNICAL MERIT

<table>
<thead>
<tr>
<th>Strain</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>56.8</td>
<td>0.24</td>
</tr>
<tr>
<td>63.1</td>
<td>0.24</td>
</tr>
<tr>
<td>65.5</td>
<td>0.49</td>
</tr>
<tr>
<td>75.5</td>
<td>0.73</td>
</tr>
<tr>
<td>82.5</td>
<td>0.73</td>
</tr>
<tr>
<td>93.0</td>
<td>0.73</td>
</tr>
<tr>
<td>106.4</td>
<td>1.21</td>
</tr>
<tr>
<td>109.1</td>
<td>1.46</td>
</tr>
<tr>
<td>111.1</td>
<td>2.43</td>
</tr>
<tr>
<td>117.1</td>
<td>2.19</td>
</tr>
<tr>
<td>126.1</td>
<td>3.64</td>
</tr>
<tr>
<td>140.0</td>
<td>0.97</td>
</tr>
<tr>
<td>147.1</td>
<td>6.31</td>
</tr>
<tr>
<td>173.0</td>
<td>7.28</td>
</tr>
<tr>
<td>199.5</td>
<td>4.13</td>
</tr>
<tr>
<td>246.8</td>
<td>8.50</td>
</tr>
<tr>
<td>284.4</td>
<td>9.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>log(Strain)</th>
<th>log(Stress)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>1.754</td>
<td>0.0017</td>
</tr>
<tr>
<td>1.800</td>
<td>0.0018</td>
</tr>
<tr>
<td>1.803</td>
<td>0.0033</td>
</tr>
<tr>
<td>1.878</td>
<td>0.0042</td>
</tr>
<tr>
<td>1.908</td>
<td>0.0026</td>
</tr>
<tr>
<td>1.916</td>
<td>0.0038</td>
</tr>
<tr>
<td>1.968</td>
<td>0.0034</td>
</tr>
<tr>
<td>2.027</td>
<td>0.0049</td>
</tr>
<tr>
<td>2.038</td>
<td>0.0058</td>
</tr>
<tr>
<td>2.046</td>
<td>0.0095</td>
</tr>
<tr>
<td>2.069</td>
<td>0.0080</td>
</tr>
<tr>
<td>2.101</td>
<td>0.0125</td>
</tr>
<tr>
<td>2.146</td>
<td>0.0030</td>
</tr>
<tr>
<td>2.168</td>
<td>0.0188</td>
</tr>
<tr>
<td>2.238</td>
<td>0.0183</td>
</tr>
<tr>
<td>2.300</td>
<td>0.0090</td>
</tr>
<tr>
<td>2.392</td>
<td>0.0152</td>
</tr>
<tr>
<td>2.454</td>
<td>0.0144</td>
</tr>
</tbody>
</table>

Data from Mandel (3), but rearranged to better show the variation of the error standard deviation.
Fig. 4.5 Variation of the error standard deviation with stress and strain measurements. The small dots represent the standard deviations of groups of three replicate measurements. The large dots represent the standard deviations of groups of twelve measurements. The dashed lines were visually fitted.

Reprocessed Data after Grouping of Measurements

Table 4.4 shows that the estimated standard deviations have in some cases a fairly large amount of scatter. In an attempt to reduce this scatter the data was grouped in sets of twelve measurements. New means were calculated, as well
as, new standard deviations (33); the standard deviations versus the means are plotted in Fig. 4.5. The data was then transformed (Table 4.5) as before and replotted (Fig. 4.6).

Reprocessing the data resulted in the following expression for the relative sensitivity of the strain method to the stress method:

\[
\text{RS[strain/stress]} = \frac{1.973}{-5.075 + 3.058 \times \log(\text{strain})}
\]

The new values of \( RS \) were 6.46 (strain = 56.8) and 0.77 (strain = 284.4). Again the precision of the strain test is shown to be superior over most of the range of interest.

<table>
<thead>
<tr>
<th>( \log(\text{Strain}) )</th>
<th>( \log(\text{Stress}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td><strong>Standard Deviation</strong></td>
</tr>
<tr>
<td>1.786</td>
<td>0.0024</td>
</tr>
<tr>
<td>1.901</td>
<td>0.0036</td>
</tr>
<tr>
<td>2.012</td>
<td>0.0050</td>
</tr>
<tr>
<td>2.072</td>
<td>0.0104</td>
</tr>
<tr>
<td>2.186</td>
<td>0.0158</td>
</tr>
<tr>
<td>2.387</td>
<td>0.0138</td>
</tr>
</tbody>
</table>
Fig. 4.6 Variation of transformed standard deviation with log(average measurements). The points represent the standard deviations and averages of groups of twelve transformed measurements.
4.8 Experimental:

Constant Error Standard Deviation Model

Three sets of simulated experiments were conducted to investigate Thompson's and Mandel's techniques for the comparison of analytical methods, and the procedures for estimating the experimental error variance. These experiments assumed: (i) a constant error standard deviation for each method, (ii) a constant relative error standard deviation for each method, and (iii) a constant error standard deviation for one, and a constant relative standard deviation for the second method. The responses were generated using the normal number generator discussed in section 2.4 and used in section 3.6.

For this first experiment, using a constant error standard deviation model, the same series of twenty concentrations generated in section 3.6.1 were used. Responses for Method M4-1 were in the range of 0.0 to 0.6 units and had a constant standard deviation of 0.006. Responses for Method M4-2 were in the range of 0.0 to 10.0 units and had a constant standard deviation of 0.015.

Three normally distributed responses were generated for each concentration by each method; the responses were rounded off to three decimal places.

4.8.1 Results and Discussion

The responses, their averages, and standard deviations are listed in Table 4.6.

Mandel's Technique

It was assumed that the two analytical methods were appropriate for the analysis, that a functional relationship existed between them, and that no transformations were required (sec. 4.7.2).

Three of the approaches discussed previously (sec. 4.5.2) were used for estimating the error's SD. The first was the "Replicate Analyses" approach. Fig. 4.7 shows the standard deviations plotted versus the response averages; the large scatter of the SD estimates is obvious. A SLR analysis (SD versus response averages) yielded estimates of
0.005683 and 0.013773 for methods M4-1 and M4-2 respectively. Estimates were then obtained using ANOVA (one factor, concentration level); the SDs in this case were 0.005433 and 0.01407. Finally, the means of the SDs for each method were 0.005048 and 0.01233; they obviously result in low estimates.

![Graphs showing standard deviations as a function of average response for methods M4-1 and M4-2.](image)

Fig. 4.7 Standard deviations estimated from triplicate responses regressed on the average response. SLR was used to obtain the regression line.
The responses for each method were then plotted against the known concentrations (Approach III); SLR was used since the SD is known to be constant. The estimates of the SD were 0.005205 and 0.01403 respectively.

Lastly, the responses were plotted against themselves (Approach IV), after rotating the positions of the triplicates for the x-variable set. The estimates were 0.005525 and 0.01433. All estimates are listed in Table 4.7 for easier comparison.

The responses of one method were then plotted against the other to determine the dM/dN ratio. However, the question as to which set of responses should be used as the independent (x) variable had to be answered.

Mandel (3) has explained that the calculation of the relative sensitivity will in fact answer this question. From his explanation it is possible to state that the x-variable should be the responses from the method having the smaller relative magnitude of experimental error, with respect to the total response range. This results in a smaller error for the calculated slope than the opposite choice.

In the experiment, Method M4-1 has a relative error of 0.01 \( 0.006/(0.6-0.0) \) and Method M4-2 has a relative error of 0.0015 \( 0.015/(10.0-0.0) \). The responses of Method M4-2 were therefore used as the x-variable. Regressing the average responses of Method M4-1 on the average responses of M4-2 resulted in the following equation (Fig. 4.8):

\[
y = 0.06026 x - 0.0008137
\]

Using a regression of M4-2 on M4-1, and inverting the equation resulted in:

\[
y = 0.06027 x - 0.0008789
\]

Thus, it is evident that for this particular example there is little difference between the two regressions.
TABLE 4.6
SIMULATED REGRESSION DATA WITH
CONSTANT ERROR STANDARD DEVIATION

<table>
<thead>
<tr>
<th>Standards</th>
<th>Responses</th>
<th>M4-1</th>
<th>M4-2</th>
<th>Avg</th>
<th>SD</th>
<th>Avg</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conc'n</td>
<td>M4-1</td>
<td>M4-2</td>
<td>Avg</td>
<td>SD</td>
<td>Avg</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>.165</td>
<td>.0085</td>
<td>.199</td>
<td>.0112</td>
<td>.003245</td>
<td>.168</td>
<td>.01308</td>
<td></td>
</tr>
<tr>
<td>.165</td>
<td>.0148</td>
<td>.183</td>
<td>.0915</td>
<td>.006077</td>
<td>1.477</td>
<td>.00529</td>
<td></td>
</tr>
<tr>
<td>.165</td>
<td>.0103</td>
<td>.162</td>
<td>.1297</td>
<td>.004996</td>
<td>2.157</td>
<td>.01405</td>
<td></td>
</tr>
<tr>
<td>1.466</td>
<td>.0851</td>
<td>.147</td>
<td>.136</td>
<td>.005268</td>
<td>2.319</td>
<td>.01249</td>
<td></td>
</tr>
<tr>
<td>1.466</td>
<td>.0987</td>
<td>.1481</td>
<td>.1598</td>
<td>.004386</td>
<td>2.677</td>
<td>.02322</td>
<td></td>
</tr>
<tr>
<td>2.158</td>
<td>.1229</td>
<td>.2158</td>
<td>.1762</td>
<td>.009777</td>
<td>2.911</td>
<td>.01625</td>
<td></td>
</tr>
<tr>
<td>2.158</td>
<td>.1314</td>
<td>.2142</td>
<td>.2106</td>
<td>.003586</td>
<td>3.315</td>
<td>.02003</td>
<td></td>
</tr>
<tr>
<td>2.158</td>
<td>.1317</td>
<td>.217</td>
<td>.2185</td>
<td>.006222</td>
<td>3.625</td>
<td>.02003</td>
<td></td>
</tr>
<tr>
<td>2.316</td>
<td>.1299</td>
<td>.2305</td>
<td>.2268</td>
<td>.003531</td>
<td>3.824</td>
<td>.01305</td>
<td></td>
</tr>
<tr>
<td>2.316</td>
<td>.1296</td>
<td>.2332</td>
<td>.2406</td>
<td>.002508</td>
<td>3.927</td>
<td>.00089</td>
<td></td>
</tr>
<tr>
<td>2.316</td>
<td>.1304</td>
<td>.2329</td>
<td>.2751</td>
<td>.004299</td>
<td>4.505</td>
<td>.01007</td>
<td></td>
</tr>
<tr>
<td>2.688</td>
<td>.1548</td>
<td>.2686</td>
<td>.3913</td>
<td>.0048</td>
<td>6.535</td>
<td>.00051</td>
<td></td>
</tr>
<tr>
<td>2.688</td>
<td>.1616</td>
<td>.2696</td>
<td>.4272</td>
<td>.008251</td>
<td>7.08</td>
<td>.00282</td>
<td></td>
</tr>
<tr>
<td>2.688</td>
<td>.1637</td>
<td>.2448</td>
<td>.4223</td>
<td>.003142</td>
<td>7.685</td>
<td>.01793</td>
<td></td>
</tr>
<tr>
<td>2.688</td>
<td>.1675</td>
<td>.284</td>
<td>.4520</td>
<td>.003351</td>
<td>7.57</td>
<td>.00889</td>
<td></td>
</tr>
<tr>
<td>2.684</td>
<td>.1718</td>
<td>.2851</td>
<td>.5066</td>
<td>.003045</td>
<td>8.368</td>
<td>.0205</td>
<td></td>
</tr>
<tr>
<td>2.684</td>
<td>.161</td>
<td>.2849</td>
<td>.5203</td>
<td>.004216</td>
<td>8.629</td>
<td>.01609</td>
<td></td>
</tr>
<tr>
<td>2.927</td>
<td>.187</td>
<td>.2931</td>
<td>.5953</td>
<td>.003842</td>
<td>9.932</td>
<td>.00252</td>
<td></td>
</tr>
<tr>
<td>2.927</td>
<td>.1673</td>
<td>.2898</td>
<td>.6052</td>
<td>.002715</td>
<td>9.997</td>
<td>.01457</td>
<td></td>
</tr>
<tr>
<td>2.927</td>
<td>.1744</td>
<td>.2901</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.519</td>
<td>.2099</td>
<td>.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.519</td>
<td>.2106</td>
<td>.333</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.519</td>
<td>.211</td>
<td>.3508</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.623</td>
<td>.2124</td>
<td>.364</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.623</td>
<td>.2244</td>
<td>.3626</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.623</td>
<td>.2191</td>
<td>.3644</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.813</td>
<td>.2267</td>
<td>.3634</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.813</td>
<td>.2215</td>
<td>.3609</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.813</td>
<td>.2322</td>
<td>.3828</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.927</td>
<td>.2379</td>
<td>.3926</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.927</td>
<td>.2349</td>
<td>.3927</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.927</td>
<td>.2469</td>
<td>.3927</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.574</td>
<td>.2609</td>
<td>.4594</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.574</td>
<td>.2774</td>
<td>.4586</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.574</td>
<td>.279</td>
<td>.4574</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.525</td>
<td>.3913</td>
<td>.6539</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.525</td>
<td>.3865</td>
<td>.6538</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.525</td>
<td>.3961</td>
<td>.6529</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.096</td>
<td>.4108</td>
<td>.7002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.096</td>
<td>.4357</td>
<td>.7077</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.096</td>
<td>.4274</td>
<td>.708</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.096</td>
<td>.4191</td>
<td>.7074</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.096</td>
<td>.4282</td>
<td>.7106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.096</td>
<td>.4195</td>
<td>.7076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.542</td>
<td>.4529</td>
<td>.7567</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.562</td>
<td>.4494</td>
<td>.758</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.562</td>
<td>.4561</td>
<td>.7563</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.345</td>
<td>.5099</td>
<td>8.392</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.345</td>
<td>.5039</td>
<td>8.357</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.345</td>
<td>.506</td>
<td>8.356</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.345</td>
<td>.5273</td>
<td>8.614</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.345</td>
<td>.5164</td>
<td>8.627</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.345</td>
<td>.5171</td>
<td>8.646</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.944</td>
<td>.5995</td>
<td>9.935</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.944</td>
<td>.5943</td>
<td>9.932</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.944</td>
<td>.592</td>
<td>9.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.6049</td>
<td>10.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.6027</td>
<td>9.982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.6001</td>
<td>9.999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 4.7
ESTIMATES OF THE REGRESSION PARAMETERS AND EXPERIMENTAL ERROR
STANDARD DEVIATIONS FOR METHODS M4-1 AND M4-2

<table>
<thead>
<tr>
<th>Method</th>
<th>Slope</th>
<th>Intercept</th>
<th>SE of Reg</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method M4-1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.06023</td>
<td>-0.00068</td>
<td>0.00521</td>
</tr>
<tr>
<td>Expected</td>
<td>0.06000</td>
<td>0.00000</td>
<td>0.00600</td>
</tr>
<tr>
<td><strong>Method M4-2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.99958</td>
<td>0.00230</td>
<td>0.01403</td>
</tr>
<tr>
<td>Expected</td>
<td>1.00000</td>
<td>0.00000</td>
<td>0.01500</td>
</tr>
</tbody>
</table>

Estimates of the Error Standard Deviation

<table>
<thead>
<tr>
<th>Approach</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLR</td>
<td></td>
<td>Mean</td>
<td>ANOVA</td>
<td>SLR</td>
</tr>
<tr>
<td>Method M4-1</td>
<td>0.00568</td>
<td>0.00505</td>
<td>0.00543</td>
<td>0.00520</td>
</tr>
<tr>
<td>Method M4-2</td>
<td>0.01377</td>
<td>0.01233</td>
<td>0.01407</td>
<td>0.01403</td>
</tr>
<tr>
<td>SD Ratio</td>
<td>0.4125</td>
<td>0.4096</td>
<td>0.3859</td>
<td>0.3706</td>
</tr>
</tbody>
</table>

(Expected: SD(M4-1)=0.006, SD(M4-2)=0.015, SD Ratio=0.4)

I - Replicate Analyses
III - Regression of Responses on Concentration
IV - Self Regression

* The expected slopes and intercepts were within the 95% confidence limits of the corresponding experimental values. The observed standard errors of regression were slightly underestimated (as they were in section 3.6).
### A

**ANOVA**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>D.F.</th>
<th>B.B.</th>
<th>M.S.</th>
<th>F RATIO</th>
<th>PROB (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLAINED BY REGRESSION</td>
<td>1</td>
<td>.59417</td>
<td>.59417</td>
<td>82452.1</td>
<td>0.000 $##$</td>
</tr>
<tr>
<td>ABOUT REGRESSION</td>
<td>18</td>
<td>1.29712E-04</td>
<td>2.70625E-06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>19</td>
<td>.5943</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EDN OF LINE**  

REGRESSION COEFF. = .9999891  

% EXPLAINED VARIATION = 99.9782

**STANDARD ERROR OF REGRESSION (Gr) = 2.68445E-03**

**CONFIDENCE LIMITS (90, 95, 99%) OF INTERCEPT:**  

- .000402545 ± .013658E-04

- .00020839 2.25469E-03 3.45722E-03  
  SD = 1.20177E-03

**CONFIDENCE LIMITS (90, 95, 99%) OF SLOPE:**  

- .0402565 ± .06024505  
  -1.38.3879E-04 4.4084E-04 6.0368E-04  
  SD = 2.09847E-04

**TOTAL**  

19  

163.645

**EDN OF LINE**  

REGRESSION COEFF. = .9999891  

% EXPLAINED VARIATION = 99.9782

**STANDARD ERROR OF REGRESSION (Gr) = .0445419**

**CONFIDENCE LIMITS (90, 95, 99%) OF INTERCEPT:**  

- .0145829  
  SD = .0145829

**CONFIDENCE LIMITS (90, 95, 99%) OF SLOPE:**  

- .0145829  
  SD = .0145829

- .100189 .121381 .166215  
  SD = .0577784

---

**Fig. 4.8** Computer results for the; A: regression of the average responses from Method M4-1 on the average responses of Method M4-2, and B: regression of Method M4-2 average responses on the average responses of M4-1.

The experimental relative sensitivity of Method M4-1 to M4-2 (using the SD ratio from Approach IV) was:

\[
RS \left( \frac{M4-1}{M4-2} \right) = \frac{d(M4-1)}{d(M4-2)} = \frac{.06026}{.01564} = 0.1564
\]

Method M4-1 is therefore less precise than Method M4-2. It would require (see sec. 4.2.1) approximately 41 \[1/(0.1564)^2\] measurements by Method M4-1 to equal the precision of one measurement by M4-2. The expected number was 45.
An interesting feature of Mandel's technique is that either or both methods could have a non-zero blank measurement without it changing the relative sensitivity. Such blank measurements would, however, change the relative standard deviation of each method and hence the ratio of the relative standard deviations.

Thompson's Technique

In this technique the responses are first converted to concentrations. The responses from Method M4-1 were each divided by 0.06 to convert them to concentration (range of 0 to 10, and a standard deviation of 0.1 (0.006/0.06)). The responses for Method M4-2 were assumed to already correspond to concentration (range of 0 to 10, and a standard deviation of 0.015). The average concentration estimates of Method M4-1 were then regressed on those of Method M4-2. The resulting computer analysis is shown in Fig. 4.9.

According to Thompson (6) tests for bias would involve testing for: (i) a significant intercept, (ii) a significant deviation in the slope of the line from zero, and (iii) significant deviations from linearity.

In this experiment the intercept and its 95% confidence limits are -0.01356 +/- 0.04207; the expected intercept of "zero" is within this range. The slope and its 95% confidence limits are 1.0043 +/- 0.0073; the expected slope of "one" is within this range. The alternate regression of M4-2 on M4-1 produced an intercept of 0.01458 +/- 0.04180 and a slope of 0.9955 +/- 0.0072. Both regressions appeared to produce similar results.

An examination of the residuals resulted in the histogram shown in Fig. 4.10; a visual "signs test" of the residuals showed no pattern indicating nonlinearity. The ANOVA table and regression coefficient of Fig. 4.9 also indicate a good linear "fit" of the data.
The expected ratio of the standard deviation of concentration was:

$$\frac{SD(M4-1)}{SD(M4-2)} = \frac{0.1}{0.015} = 6.67$$

The relative sensitivity is the inverse of the above ratio, i.e., 0.15. This is the same answer that was expected by Mandel's technique.

4.9 Experimental:

Constant Relative Error Standard Deviation Model

The concentrations (range of 0 - 100) for fifty standards were produced using a random number generator. Two "measurements" of response were produced for each standard using the "normal" number generator. These measurements had relative standard deviations of 0.01, 0.02, 0.04, 0.08, and 0.16. This resulted in analytical errors of approximately 2, 4, 8, 16, and 32% (at the 95% confidence level). This data is shown in Table 4.8.

The following data processing was performed:
- each set of responses was regressed against each other to obtain an estimate of $dM/dN$
- and to obtain estimates of experimental error RSD:
- each set of responses was regressed against the concentration (Approach III)
- each set of responses was regressed against itself (Approach IV)
- standard deviations and means were calculated for each set of measurements using Thompson's (4) duplicate analyses method. The standard deviations were then regressed against the means (Approach II).
**Fig. 4.10** A histogram of the residuals from the regression of Method M4-1 concentrations on those of Method M4-2.
<table>
<thead>
<tr>
<th>Standards Concentration</th>
<th>Percent Relative Standard Deviation %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>7%</td>
<td>8%</td>
</tr>
<tr>
<td>8%</td>
<td>9%</td>
</tr>
<tr>
<td>9%</td>
<td>10%</td>
</tr>
<tr>
<td>10%</td>
<td>11%</td>
</tr>
<tr>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>12%</td>
<td>13%</td>
</tr>
<tr>
<td>13%</td>
<td>14%</td>
</tr>
<tr>
<td>14%</td>
<td>15%</td>
</tr>
<tr>
<td>15%</td>
<td>16%</td>
</tr>
<tr>
<td>16%</td>
<td>17%</td>
</tr>
<tr>
<td>17%</td>
<td>18%</td>
</tr>
<tr>
<td>18%</td>
<td>19%</td>
</tr>
<tr>
<td>19%</td>
<td>20%</td>
</tr>
<tr>
<td>20%</td>
<td>21%</td>
</tr>
<tr>
<td>21%</td>
<td>22%</td>
</tr>
<tr>
<td>22%</td>
<td>23%</td>
</tr>
<tr>
<td>23%</td>
<td>24%</td>
</tr>
<tr>
<td>24%</td>
<td>25%</td>
</tr>
<tr>
<td>25%</td>
<td>26%</td>
</tr>
<tr>
<td>26%</td>
<td>27%</td>
</tr>
<tr>
<td>27%</td>
<td>28%</td>
</tr>
<tr>
<td>28%</td>
<td>29%</td>
</tr>
<tr>
<td>29%</td>
<td>30%</td>
</tr>
<tr>
<td>30%</td>
<td>31%</td>
</tr>
<tr>
<td>31%</td>
<td>32%</td>
</tr>
<tr>
<td>32%</td>
<td>33%</td>
</tr>
<tr>
<td>33%</td>
<td>34%</td>
</tr>
<tr>
<td>34%</td>
<td>35%</td>
</tr>
<tr>
<td>35%</td>
<td>36%</td>
</tr>
<tr>
<td>36%</td>
<td>37%</td>
</tr>
<tr>
<td>37%</td>
<td>38%</td>
</tr>
<tr>
<td>38%</td>
<td>39%</td>
</tr>
<tr>
<td>39%</td>
<td>40%</td>
</tr>
<tr>
<td>40%</td>
<td>41%</td>
</tr>
<tr>
<td>41%</td>
<td>42%</td>
</tr>
<tr>
<td>42%</td>
<td>43%</td>
</tr>
<tr>
<td>43%</td>
<td>44%</td>
</tr>
<tr>
<td>44%</td>
<td>45%</td>
</tr>
<tr>
<td>45%</td>
<td>46%</td>
</tr>
<tr>
<td>46%</td>
<td>47%</td>
</tr>
<tr>
<td>47%</td>
<td>48%</td>
</tr>
<tr>
<td>48%</td>
<td>49%</td>
</tr>
<tr>
<td>49%</td>
<td>50%</td>
</tr>
<tr>
<td>50%</td>
<td>51%</td>
</tr>
<tr>
<td>51%</td>
<td>52%</td>
</tr>
<tr>
<td>52%</td>
<td>53%</td>
</tr>
<tr>
<td>53%</td>
<td>54%</td>
</tr>
<tr>
<td>54%</td>
<td>55%</td>
</tr>
<tr>
<td>55%</td>
<td>56%</td>
</tr>
<tr>
<td>56%</td>
<td>57%</td>
</tr>
<tr>
<td>57%</td>
<td>58%</td>
</tr>
<tr>
<td>58%</td>
<td>59%</td>
</tr>
<tr>
<td>59%</td>
<td>60%</td>
</tr>
<tr>
<td>60%</td>
<td>61%</td>
</tr>
<tr>
<td>61%</td>
<td>62%</td>
</tr>
<tr>
<td>62%</td>
<td>63%</td>
</tr>
<tr>
<td>63%</td>
<td>64%</td>
</tr>
<tr>
<td>64%</td>
<td>65%</td>
</tr>
<tr>
<td>65%</td>
<td>66%</td>
</tr>
<tr>
<td>66%</td>
<td>67%</td>
</tr>
<tr>
<td>67%</td>
<td>68%</td>
</tr>
<tr>
<td>68%</td>
<td>69%</td>
</tr>
<tr>
<td>69%</td>
<td>70%</td>
</tr>
<tr>
<td>70%</td>
<td>71%</td>
</tr>
<tr>
<td>71%</td>
<td>72%</td>
</tr>
<tr>
<td>72%</td>
<td>73%</td>
</tr>
<tr>
<td>73%</td>
<td>74%</td>
</tr>
<tr>
<td>74%</td>
<td>75%</td>
</tr>
<tr>
<td>75%</td>
<td>76%</td>
</tr>
<tr>
<td>76%</td>
<td>77%</td>
</tr>
<tr>
<td>77%</td>
<td>78%</td>
</tr>
<tr>
<td>78%</td>
<td>79%</td>
</tr>
<tr>
<td>79%</td>
<td>80%</td>
</tr>
<tr>
<td>80%</td>
<td>81%</td>
</tr>
<tr>
<td>81%</td>
<td>82%</td>
</tr>
<tr>
<td>82%</td>
<td>83%</td>
</tr>
<tr>
<td>83%</td>
<td>84%</td>
</tr>
<tr>
<td>84%</td>
<td>85%</td>
</tr>
<tr>
<td>85%</td>
<td>86%</td>
</tr>
<tr>
<td>86%</td>
<td>87%</td>
</tr>
<tr>
<td>87%</td>
<td>88%</td>
</tr>
<tr>
<td>88%</td>
<td>89%</td>
</tr>
<tr>
<td>89%</td>
<td>90%</td>
</tr>
<tr>
<td>90%</td>
<td>91%</td>
</tr>
<tr>
<td>91%</td>
<td>92%</td>
</tr>
<tr>
<td>92%</td>
<td>93%</td>
</tr>
<tr>
<td>93%</td>
<td>94%</td>
</tr>
<tr>
<td>94%</td>
<td>95%</td>
</tr>
<tr>
<td>95%</td>
<td>96%</td>
</tr>
<tr>
<td>96%</td>
<td>97%</td>
</tr>
<tr>
<td>97%</td>
<td>98%</td>
</tr>
<tr>
<td>98%</td>
<td>99%</td>
</tr>
<tr>
<td>99%</td>
<td>100%</td>
</tr>
</tbody>
</table>

TABLE 4.8 DUPLICATE INTENSITY MEASUREMENTS FOR A SET OF STANDARDS
4.9.1 Results and Discussion

The results of this experiment were used:
- to verify some of Thompson's conclusions in regards to the use of SLR and WLR for comparing two analytical procedures
- to verify the use of WLR in determining the experimental error's relative standard deviation when the concentration of the standards are "known"
- to verify the use of WLR in determining the experimental error's relative standard deviation when the concentration of the standards are "unknown"
- to determine the error standard deviation function using Thompson's (4) duplicate analyses technique.

Effectiveness of Weighted Linear Regression

Table 4.9 shows the slope and intercept of each WLR and indicates whether they are within the computed 95% confidence limits. Each pair of measurement sets were regressed against each other twice using WLR; each was used as the x-, and then the y-variable.

It can be seen from Table 4.9 that similar slopes and intercepts for the regression lines were obtained by regressing x on y and y on x. Thompson's (6) conclusions imply that the variable with the smaller RSD should be used as the independent (x) variable. The results of this experiment do not indicate this requirement to be critical. The computed regression coefficients and confidence limits for the slopes and intercepts only slightly favoured the regression having the variable with the smaller RSD as the x-variable.

Effectiveness of Simple Linear Regression

Each pair was also regressed using SLR; again each was used as the x-, and then the y-variable. The results in Table 4.10 clearly show that a SLR using the measurements with the higher RSD as the x-variable is not correct. Such a procedure leads to very high intercepts and slopes. The slopes diverge increasingly from the expected values as the differences between the RSDs increase.
### TABLE 4.9

**THE WEIGHTED LINEAR REGRESSION OF RESULTS FROM ANALYTICAL METHODS HAVING EXPERIMENTAL ERRORS WITH CONSTANT RELATIVE STANDARD DEVIATIONS**

<table>
<thead>
<tr>
<th>%RSD</th>
<th>Slope</th>
<th>Intercept</th>
<th>%RSD</th>
<th>Slope</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.0026</td>
<td>-0.02907</td>
<td>2</td>
<td>0.9977</td>
<td>0.03213#</td>
</tr>
<tr>
<td>4</td>
<td>0.9920</td>
<td>0.06917#</td>
<td>4</td>
<td>1.0096</td>
<td>-0.05947</td>
</tr>
<tr>
<td>8</td>
<td>0.9891</td>
<td>0.06863</td>
<td>8</td>
<td>1.0203</td>
<td>-0.06994</td>
</tr>
<tr>
<td>16</td>
<td>0.9961</td>
<td>0.07725</td>
<td>16</td>
<td>1.0285</td>
<td>-0.04887</td>
</tr>
<tr>
<td>16</td>
<td>1.0248</td>
<td>0.04569</td>
<td>16</td>
<td>0.9976</td>
<td>0.04045</td>
</tr>
<tr>
<td>4</td>
<td>0.9898</td>
<td>0.09800</td>
<td>4</td>
<td>1.0123#</td>
<td>-0.09250#</td>
</tr>
<tr>
<td>8</td>
<td>0.9865</td>
<td>0.10091</td>
<td>8</td>
<td>1.0226#</td>
<td>-0.09923</td>
</tr>
<tr>
<td>16</td>
<td>0.9942</td>
<td>0.10617</td>
<td>16</td>
<td>1.0316</td>
<td>-0.10116</td>
</tr>
<tr>
<td>16</td>
<td>1.0463#</td>
<td>-0.02469</td>
<td>16</td>
<td>0.9871</td>
<td>0.10643</td>
</tr>
<tr>
<td>8</td>
<td>0.9982</td>
<td>0.01305</td>
<td>8</td>
<td>1.0116</td>
<td>0.00376</td>
</tr>
<tr>
<td>16</td>
<td>1.0050</td>
<td>0.02288</td>
<td>16</td>
<td>1.0178</td>
<td>0.00424</td>
</tr>
<tr>
<td>16</td>
<td>1.0328</td>
<td>0.00554</td>
<td>16</td>
<td>0.9886</td>
<td>0.11287</td>
</tr>
<tr>
<td>16</td>
<td>1.0165</td>
<td>0.01573</td>
<td>16</td>
<td>1.0177</td>
<td>0.00424</td>
</tr>
<tr>
<td>16</td>
<td>1.0463#</td>
<td>0.02469</td>
<td>16</td>
<td>0.9871</td>
<td>0.10643</td>
</tr>
</tbody>
</table>

The expected slope and intercept were 1.0000 and 0.00000 respectively.

- The confidence limits increased with increasing difference between the %RSDs.
- The regression coefficients decreased with increasing differences between the %RSDs (0.9997 - 0.9775)

# the value falls outside its 95% confidence limits
### TABLE 4.10

THE SIMPLE LINEAR REGRESSION OF RESULTS
FROM ANALYTICAL METHODS HAVING EXPERIMENTAL
ERRORS WITH CONSTANT RELATIVE STANDARD DEVIATIONS

<table>
<thead>
<tr>
<th>%RSD</th>
<th>Slope</th>
<th>Intercept</th>
<th>%RSD</th>
<th>Slope</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x-variable = 1%</td>
<td></td>
<td></td>
<td>x-variable = 1%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.0008</td>
<td>0.00360</td>
<td>2</td>
<td>0.9973</td>
<td>0.09599</td>
</tr>
<tr>
<td>4</td>
<td>0.9815</td>
<td>0.52645</td>
<td>4</td>
<td>1.0096</td>
<td>-0.06566</td>
</tr>
<tr>
<td>8</td>
<td>0.9988</td>
<td>-0.24036</td>
<td>8</td>
<td>0.9631</td>
<td>2.20300</td>
</tr>
<tr>
<td>16</td>
<td>1.0036</td>
<td>-0.25729</td>
<td>16</td>
<td>0.9053#</td>
<td>4.96034#</td>
</tr>
<tr>
<td></td>
<td>x-variable = 2%</td>
<td></td>
<td></td>
<td>x-variable = 2%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.9790#</td>
<td>0.60990</td>
<td>4</td>
<td>1.0107</td>
<td>-0.07313</td>
</tr>
<tr>
<td>8</td>
<td>0.9970</td>
<td>-0.19146</td>
<td>8</td>
<td>0.9647</td>
<td>-2.16323</td>
</tr>
<tr>
<td>16</td>
<td>1.0021</td>
<td>-0.22282</td>
<td>16</td>
<td>0.9071#</td>
<td>4.91211#</td>
</tr>
<tr>
<td></td>
<td>x-variable = 4%</td>
<td></td>
<td></td>
<td>x-variable = 4%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.0113</td>
<td>-0.45309</td>
<td>8</td>
<td>0.9479#</td>
<td>2.55046#</td>
</tr>
<tr>
<td>16</td>
<td>1.0156</td>
<td>-0.44114</td>
<td>16</td>
<td>0.9032#</td>
<td>4.76389#</td>
</tr>
<tr>
<td></td>
<td>x-variable = 6%</td>
<td></td>
<td></td>
<td>x-variable = 6%</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.96554</td>
<td>2.00693</td>
<td>16</td>
<td>0.9032#</td>
<td>4.76389#</td>
</tr>
</tbody>
</table>

- The expected slope and intercept were 1.0000 and 0.00000 respectively.
- The confidence limits increased with increasing difference between the %RSDs.
- The regression coefficients decreased with increasing differences between the %RSDs (0.9990 - 0.9339)

# the value falls outside its 95% confidence limits
Using the measurements with the smaller RSD as the x-variable improves the estimates of both the slope and intercept. However, it can be seen by comparing the results of Tables 4.9 and 4.10 that WLR produces much better results.

Estimating Relative Standard Deviations

It was previously mentioned (sec. 4.5.2) that there are at least four methods of estimating the experimental error function. The simplest, the "Replicate Analyses" method is not used in this experiment as only duplicate measurements were obtained; twenty replicates are usually preferred.

The "Duplicate Analyses" (II), the "Regression of Responses on Concentration" (III), and the "Self Regression" (IV) approaches were used to generate estimates of the RSDs.

The data provided an additional opportunity to test the reasoning behind these latter two methods. In this experiment the RSDs of both methods are known. It is therefore possible to calculate an estimate of the RSD of one set by using the computed standard error of regression (SE reg) and the known RSD of the other set. For example, if the measurement set having a RSD of 2% is regressed on the set with a RSD of 1%, then using the computed SE of regression and the known RSD of both sets will allow an estimate of the "unknown" RSD to be calculated. The formula used to calculate RSD\{1\%\} would be:

\[
\text{RSD}\{1\%\} = \sqrt{(\text{SE reg})^2 + (\text{RSD}\{1\%\})^2 + (\text{RSD}\{2\%\})^2} / \text{RSD}\{1\%\}
\]

These calculated RSDs are listed in Table 4.11 "C". The results on average tended to be about 10% higher than the expected values. It was therefore again assumed that the method had merit. In particular, it could be used as a fast method of roughly estimating the analytical error, whenever it can be assumed that the error has an approximately constant RSD.

The results ("A") from the "Regression of Responses on Concentration" approach are also listed in Table 4.11. They
agree reasonably well with the expected values except for the 8% RSD estimate which is high! This estimate is probably a result of the particular data set, as all approaches gave high values for the set.

The estimates ("B") from the "Self Regression" approach are in good agreement (except for the 8% RSD estimate) with the above results ("A"). This method is particularly useful and simple to use. The original data set is used as one variable, and then reused as the second variable--after each duplicate measurement is swapped with the other.

Lastly, the %RSDs ("D") were estimated using Thompson’s (4) "Duplicate Analyses" approach. A short routine in BASIC was written and temporarily inserted into the program to process the raw data already stored on the computer disk. The data was analysed using both SLR and WLR (the weight used was proportional to the square of the SD estimate). The results obtained agreed reasonably well with the expected values; the WLR results appeared to be "better".

It can be seen from Table 4.11 that all three approaches produced reasonable estimates of the %RSD, with Thompson’s approach perhaps giving the poorest results. Assuming that the prerequisites of all are met and that the appropriate computer software is available, than all three approaches can be used. The time actually required for the data processing was short. Approaches II and IV have the advantage of not requiring the actual concentrations of the standards used.
TABLE 4.11
ESTIMATES OF THE EXPERIMENTAL ERROR'S
RELATIVE STANDARD DEVIATION FROM DUPLICATE ANALYSES

<table>
<thead>
<tr>
<th>% Relative Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected Value 1%  2%  4%  8%  16%  16%</td>
</tr>
</tbody>
</table>

"A" Regression of Response on Concentration - Approach III:
(Constant RSD and Known Concentration)
0.98  2.19  4.20  9.33  16.0  15.7

"B" Self Regression - Approach IV: *
(Constant RSD and Unknown Concentration)
0.84  2.27  4.16  10.58  15.7  15.9

"C" Constant and Known RSDs, Unknown Concentration: **
1.06  2.12  4.46  9.31  15.8  ----
1.11  2.18  4.36  9.01  19.3  ----
1.16  2.25  4.61  9.22  15.7  ----
0.99  2.41  3.93  8.69  17.4  ----

"D" Thompson's Duplicate Analyses - Approach II:

<table>
<thead>
<tr>
<th></th>
<th>SLR</th>
<th>WLR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.70</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>1.81</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>5.36</td>
<td>4.49</td>
</tr>
<tr>
<td></td>
<td>13.15</td>
<td>11.54</td>
</tr>
<tr>
<td></td>
<td>21.65</td>
<td>14.22</td>
</tr>
</tbody>
</table>

* \[ \text{RSD} = \text{SE regression} / \sqrt{2} \]

** \[ \text{RSD}_x = \sqrt{\left(\text{SE regression}\right)^2 + \left(\text{RSD}_x\right)^2} \]

or \[ \text{RSD}_y = \sqrt{\left(\text{SE regression}\right)^2 + \left(\text{RSD}_y\right)^2} \]
Relative Sensitivity of the Methods

If in this experiment it is assumed that the responses correspond exactly with concentration then Mandel's Relative Sensitivity need not be applied. The determination of technical merit becomes simply a matter of calculating the ratio of the analytical error of each method as the

\[ RS(M/N) = SD_n / SD_m \]

Table 4.12 lists the expected and calculated values.

**Table 4.12**

EXPECTED AND EXPERIMENTAL RELATIVE SENSITIVITY
VALUES FOR METHODS HAVING A CONSTANT RSD

%RSDs of the Variables

<table>
<thead>
<tr>
<th>y-variable (M)</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>8</th>
<th>16</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-variable (N)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

Relative Sensitivity

Expected (N/M) | 2 | 4 | 8 | 16 | 2 | 4 | 8 | 2 | 4 | 2
Experimental Approach:

III (N/M) | 2.2 | 4.3 | 9.5 | 16.3 | 1.9 | 4.3 | 7.3 | 2.2 | 3.8 | 1.7
IV (N/M) | 2.7 | 5.0 | 12.3 | 18.7 | 1.8 | 4.7 | 6.9 | 2.5 | 3.8 | 1.5
II (N/M SLR) | 2.6 | 7.7 | 18.8 | 30.9 | 3.0 | 7.3 | 12.0 | 2.5 | 4.0 | 1.6
II (N/M WLR) | 2.4 | 6.0 | 15.5 | 19.0 | 2.5 | 6.5 | 8.0 | 2.6 | 3.2 | 1.2

Relative Standard Deviations were taken from Table 4.11
4.10 Experimental:

Mixed Model - Constant Error BD and RSD

The concentrations from the previous experimental section (4.9) were used to generate fifty corresponding absorbances. The absorbances were in the range of 0.0 to 0.5 (corresponding to concentrations of 0.0 to 100.0). A "blank measurement" of 0.2 was added to each absorbance and they were then converted to transmittance. Each was then used to generate two normally distributed "measurements". Five sets of these "measurements" were generated; they had constant standard deviations of 0.005, 0.01, 0.02, 0.04, and 0.08. The concentrations and the five sets of transmittance measurements are listed in Table 4.13.

The data was processed to obtain estimates of the experimental error and to determine Mandel's (3) relative sensitivity ratio using the theoretical and the experimental parameters.

4.10.1 Results and Discussion

The results of this experiment were used:
- to verify the use of SLR in determining the experimental error's standard deviation when the concentrations are "unknown" (Approach IV)
- to examine the use of Approach II (Thompson's) in determining a constant experimental error.
- to illustrate the use of Mandel's relative sensitivity technique in a case consisting of a constant standard deviation, a constant relative standard deviation, and a single transformation.

Estimating Constant Standard Deviations

Only Approaches II and IV could be used to estimate the experimental error; there was insufficient replication for Approach I, and Approach III requires the response to be proportional to concentration (which it isn't). Again, as in the WLR/Constant RSD case, regressing (using SLR) the duplicates on themselves produced good estimates of the experimental error.
<table>
<thead>
<tr>
<th>Standards</th>
<th>Concentration</th>
<th>0.005</th>
<th>0.01</th>
<th>0.02</th>
<th>0.04</th>
<th>0.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0844</td>
<td>.61763</td>
<td>.62720</td>
<td>.62720</td>
<td>.62827</td>
<td>.70482</td>
<td>.62083</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62995</td>
<td>.62043</td>
<td>.62043</td>
<td>.62117</td>
<td>.62739</td>
<td>.62544</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
</tr>
<tr>
<td>1.0844</td>
<td>.61763</td>
<td>.62720</td>
<td>.62720</td>
<td>.62827</td>
<td>.70482</td>
<td>.62083</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62995</td>
<td>.62043</td>
<td>.62043</td>
<td>.62117</td>
<td>.62739</td>
<td>.62544</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
</tr>
<tr>
<td>1.0844</td>
<td>.61763</td>
<td>.62720</td>
<td>.62720</td>
<td>.62827</td>
<td>.70482</td>
<td>.62083</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62995</td>
<td>.62043</td>
<td>.62043</td>
<td>.62117</td>
<td>.62739</td>
<td>.62544</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
</tr>
<tr>
<td>1.0844</td>
<td>.61763</td>
<td>.62720</td>
<td>.62720</td>
<td>.62827</td>
<td>.70482</td>
<td>.62083</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62995</td>
<td>.62043</td>
<td>.62043</td>
<td>.62117</td>
<td>.62739</td>
<td>.62544</td>
</tr>
<tr>
<td>1.0844</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
<td>.62005</td>
</tr>
</tbody>
</table>

**TABLE 4.13 DUPLICATE TRANSMITTANCE MEASUREMENTS FOR A SET OF STANDARDS**

<table>
<thead>
<tr>
<th>Standards</th>
<th>Concentration</th>
<th>0.000</th>
<th>0.01</th>
<th>0.02</th>
<th>0.04</th>
<th>0.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>54.8797</td>
<td>.337356</td>
<td>.338356</td>
<td>.348356</td>
<td>.358356</td>
<td>.368356</td>
<td>.378356</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
<tr>
<td>55.2377</td>
<td>.332377</td>
<td>.333377</td>
<td>.334377</td>
<td>.335377</td>
<td>.336377</td>
<td>.337377</td>
</tr>
</tbody>
</table>

*Responses*
Thompson’s approach, however, produced both good and bad estimates; SLR and WLR were each used although only SLR was thought to be appropriate. The results are tabulated in Table 4.14.

The standard deviations used in calculating Mandel’s Relative Sensitivity ratio were those determined by the regression of the response data set on itself.

Mandel’s Relative Sensitivity Ratio

The development of three expressions for the determination of the RS ratio are outlined in Table 4.15. The specific data used is for a set of intensity measurements having a RSD of 4% (Table 4.8) and a set of transmittance measurements having a SD of 0.02 (Table 4.13).

The first expression (RS = 0.023103*T*I) is used to determine the RS when the known parameters, by which the data sets were generated, are used. This will also be the expression used to determine the RS from the comparison of the concentration standard deviations; both dA/dI and SD(A) are each multiplied by the same constant when the conversion from response to concentration is made.

The second expression (RS = 0.02682*T*I) is used to determine the RS when the experimentally determined parameters are used, and where the transformation (sec. 3.4.1) of the transmittance standard deviation is accounted for (3). The third expression (RS = 0.003854*I/A) uses the experimentally determined RSDs of both the intensity and the absorbance data sets, i.e., the fact that the transmittance has a constant SD is ignored. This last expression for calculating the RS is incorrect for this situation. However, it is likely that this approach might actually be used in practice as the analyst may not have prior knowledge that the transmittances have a constant error standard deviation. The real data used in sec. 3.4.1 have already indicated that using this approach produces results similar to using the transformed standard deviations.

The RSs at six response levels were calculated using the above three expressions. Table 4.16 allows a comparison
of the results; it can be seen that the RSs calculated from the experimental data are somewhat similar to the expected values. The RSs indicate that the intensity method yields more precise measurements. The table also lists the number of absorbance measurements which would be necessary to equal the precision of one intensity measurement. If, as is very likely, the analyst simply wants a reasonable indication of the relative technical merit then ignoring the transformation of the transmittance SD probably won't have a significant impact on the result.

The RS calculated at an absorbance of 0.2 and an intensity of zero was "0". However, the value is meaningless because at this level the intensity would be expected to have a zero standard deviation (RSD=4%) which experimentally would not be possible.

TABLE 4.14
ESTIMATES OF THE EXPERIMENTAL ERROR’S STANDARD DEVIATION FROM DUPLICATE ANALYSES

<table>
<thead>
<tr>
<th>Expected Value</th>
<th>0.005</th>
<th>0.010</th>
<th>0.020</th>
<th>0.040</th>
<th>0.080</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Regression - Approach IV: *</td>
<td>0.0048</td>
<td>0.0098</td>
<td>0.018</td>
<td>0.046</td>
<td>0.085</td>
</tr>
<tr>
<td>Thompson’s Duplicate Analyses Method - Approach II:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLR</td>
<td>0.0015</td>
<td>0.0084</td>
<td>-0.0051</td>
<td>0.038</td>
<td>0.109</td>
</tr>
<tr>
<td>WLR</td>
<td>0.0009</td>
<td>0.0083</td>
<td>-0.0047</td>
<td>0.038</td>
<td>0.111</td>
</tr>
</tbody>
</table>

* SD = SE regression / SQRT(2)
TABLE 4.15
DETERMINING MANDEL'S RELATIVE SENSITIVITY RATIO
WHEN THE EXPERIMENTAL ERROR OF ONE SET OF RESPONSES
HAS A CONSTANT STANDARD DEVIATION AND THE OTHER
HAS A CONSTANT RELATIVE STANDARD DEVIATION

*** Expected Relative Sensitivity Ratio ***
\[
RS(T/I) = \frac{dI/dT}{dI/dI} = RS(A/I) = \frac{dA/dI}{dA/dI} = RS(CA/CI)
\]
\[\frac{dA/dI}{dA/dI} = (0.7 - 0.2)/(100.0 - 0.0) = 0.005\]
\[SD(A) = 1/2.303 \times SD(T)/T = 0.02/(2.303 \times T)\]
\[SD(I) = RSD(I) \times I = 0.04 \times I\]

\[RS(A/I) = \frac{0.005 \times 2.303 \times T \times 0.04 \times I}{0.02} = 0.02303 \times T \times I\]

*** Experimental Relative Sensitivity Ratio ***
Assuming SD of transmittance is used:

\[dA/dI = 0.005089\]
\[SD(T) = 0.01818\]
\[SD(I) = 0.04160 \times I\]

\[RS(A/I) = \frac{0.005089 \times 2.303 \times T \times 0.04160 \times I}{0.01818} = 0.02682 \times T \times I\]

Assuming constant SD of transmittance is ignored:

\[dA/dI = 0.005089\]
\[SD(A) = 0.05493 \times A\]
\[SD(I) = 0.04160 \times I\]

\[RS(A/I) = \frac{0.005089 \times 0.04160 \times I}{0.05493 \times A} = 0.003854 \times I/A\]

\[RS(CA/CI) = RS\] using standard deviations of concentration
T = transmittance, I = intensity, A = absorbance
Data sets used: transmittance, SD=0.02; intensity, RSD=4%
**Table 4.16**

A Comparison of Expected and Experimental Values for Mandel’s Relative Sensitivity Ratio

<table>
<thead>
<tr>
<th>Levels at which the RS is calculated</th>
<th>Absorbance</th>
<th>Transmittance</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2 0.3 0.4 0.5 0.6 0.7</td>
<td>0.6310 0.5012 0.3981 0.3162 0.2512 0.1995</td>
<td>0 20 40 60 80 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Calculated Relative Sensitivities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
</tr>
<tr>
<td>0 0.2309 0.3667 0.4369 0.4628 0.4594</td>
</tr>
<tr>
<td>*19 8 6 5 5</td>
</tr>
<tr>
<td>Experimental*</td>
</tr>
<tr>
<td>0 0.2688 0.4271 0.5088 0.5390 0.5350</td>
</tr>
<tr>
<td>*14 6 4 4 4</td>
</tr>
<tr>
<td>Experimental**</td>
</tr>
<tr>
<td>0 0.2569 0.3854 0.4625 0.5139 0.5506</td>
</tr>
<tr>
<td>*16 7 5 4 4</td>
</tr>
</tbody>
</table>

* computed SDs for transmittance and intensity were used.
** computed SD for transmittance, and RSD for absorbance were used.

* the number of absorbance measurements required to equal the precision of one intensity measurement.

The expressions used to calculate the relative sensitivities are those developed in Table 4.15.
4.11 Conclusions

Monte Carlo simulation has been used to produce analytical "measurements" with varying standard deviations and relative standard deviations. These measurements have been used to test and illustrate four approaches of determining a method's analytical error across a wide range of concentration. These experimental error estimates have in turn been used to determine Mandel's relative sensitivity ratio for three sets of conditions:

(i) both methods have a constant standard deviation of error
(ii) both methods have a constant relative standard deviation of error
(iii) one method has a constant RSD, the other has a constant SD and its response must be transformed to make it directly proportional to concentration.

Table 4.17 summarizes information on the four approaches to determining analytical error.

The application of Mandel's technique to determine the relative sensitivity ratio, a measure of the relative technical merit of the methods, is straightforward for cases (i) and (ii).

In case (i) simple linear regression (SLR) is used. The response measurements with the smaller relative SD are used as the x-variable. Any of the four approaches to determining the SDs of the methods could be used. Approach IV (regressing the responses on themselves) appeared to be less dependent on having the concentrations range from zero upwards than did Thompson's approach (II).

Case (ii) proceeded similarly to case (i) except that weighted linear regression (WLR) was used. Again any of the four approaches could be used to determine the experimental error. When the concentrations were unknown, but the RSD was assumed to be constant, Approach IV appeared to be somewhat better than Thompson's approach to determining the RSD.

Case (iii) calculations were more involved due to the transformation of the transmittances and their standard deviation. The decision as to which response set should be used as the x-variable was also complicated. This decision
would affect the choice of SLR or WLR for the determination of the relationship between the two sets of responses, one of which is transformed. The particular data used, intensity measurements with a constant RSD and transmittance measurements with a constant SD, indicated that the transformation of the transmittance SD could be ignored without grossly affecting the value for the relative sensitivity. There could be many different examples of this case; each would require individual consideration to successfully determine the relative sensitivity.
TABLE 4.17
CONDITIONS UNDER WHICH THE FOUR APPROACHES TO DETERMINING
THE STANDARD DEVIATION OR RELATIVE STANDARD DEVIATION OF
ANALYTICAL ERROR COULD BE USED

<table>
<thead>
<tr>
<th>REQUIREMENT</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples of Known Concentration</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Duplicate Analyses</td>
<td>no</td>
<td>yes</td>
<td>no-c</td>
<td>yes</td>
</tr>
<tr>
<td>Replicate Analyses</td>
<td>yes</td>
<td>no</td>
<td>no-c</td>
<td>no-c</td>
</tr>
<tr>
<td>Normally Distributed Responses at Each Level</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Constant Error SD or RSD</td>
<td>no</td>
<td>either</td>
<td>either</td>
<td>either</td>
</tr>
<tr>
<td>WLR when RSD is Constant</td>
<td>u</td>
<td>u</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>100 plus analyses</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

no-c not required but could be used
u useful, probably would give a better result

APPROACH
I Replicate Analyses; calculated SDs regressed on known concentrations
II Duplicate Analyses; constant SD or RSD, estimated SDs regressed on estimated mean responses/concentrations
III Regression of Response on Concentration; assumption of constant SD or RSD and standards of known response/concentration
IV Self Regression; duplicate or replicate analyses, and assumption of constant SD or RSD.
4.12 Summary

The comparison of analytical methods for the purposes of validation or the assessment of technical and economic merit is of great practical importance. Linear regression techniques are essential in these comparisons. Weighted linear regression, in which the "proportional" rather than "actual" weights are used, is particularly useful—if not essential—for practical applications of method comparison.

Simple linear regression allows the constant error standard deviation of a method to be estimated. Weighted linear regression allows the constant relative standard deviation of a method be to estimated. These two techniques are uniquely useful as many analytical methods have errors which are approximately constant with, or proportional to concentration. The comparison of a method with a constant SD to one with a constant RSD is more involved—transformation of data increases the complexity of the problem.

Simulated experiments have shown the usefulness of linear regression for comparing the precision of analytical methods using either concentrations or responses. Comparisons require good estimates of the error standard deviation for each method.

Two additional approaches to estimating a method's error have been added to the two described in the analytical literature. Accurate estimates are dependent upon satisfying the assumptions of the chosen approach. "Duplicate Analyses", "Regression of Response on Concentration", and "Self Regression" are more useful than the "Replicate Analyses" approach. Often several approaches are applicable.

The "Relative Sensitivity" technique has been shown to be useful in determining the technical and economic merit of methods having constant SDs, constant RSDs, and combinations of the two with transformation of response data. Hopefully this "old" technique will finally be recognized and used.

The microcomputer, and the program for basic statistical analyses and linear regression have proved to be useful in the practical application of method comparisons. With these "tools" the routine statistical comparison of methods by the analyst is possible.
5.1 Microcomputers in the Laboratory

The potential impact of the microcomputer on the analytical laboratory is just beginning to be realized. However, because of the rapid development of computer hardware and software, the impact is still very speculative.

The manufacture of computers did not arise from a perception of an established need (1); the need was created. Perhaps the widespread use of laboratory microcomputers must also wait for the need to be more fully created.

A capsule view of current uses of microcomputers in the laboratory (2) is contained in eight papers presented at the Annual (1982) Chemical Congress of the Royal Society of Chemistry (3-10). One important point made is the low cost of microcomputers. The low cost allows it to be a "personal" computing device of one analyst or of a small group of analysts. It can exist even where a mainframe computer is readily available.

The microcomputer has hopefully now moved out of the phase in which it was used to automate laboratory calculations or operations which were previously done manually. A problem of this phase was that the microcomputer was being used to perform operations which had been deliberately simplified in order to allow a manual operation. The next phase is using the microcomputer to do these operations as the analyst would have preferred them done originally. The third phase is to use the microcomputer for operations which had not been previously considered. This phase will hopefully burst upon us soon.

An interesting aspect of microcomputers is the lack of a major commercial impact by the scientific and engineering community on its development. The microcomputer will certainly play an increasingly important role in science and engineering, but its commercial impact will be minimal compared to the business community. This is not to say that hardware and software won't be developed specifically for the laboratory. However, such developments will probably be
less extensive, more expensive, and slower than those directed towards the business market. The fall-out from the business market, in terms of software and hardware, is and should continue to be, beneficial to the scientific user.

A second interesting aspect of microcomputer development is the impact of the "home" market. Interest in the home computer has spurred the development of the microcomputer, and has lead to the creation of a multitude of computer magazines. These magazines contain advertisements and articles about software and hardware; many of these articles are very technical in nature. It is probably correct to state that most analysts are not computer experts. These magazines are therefore very important sources of information for analysts considering the use of microcomputers, as well as, analysts using microcomputers.

An article occasionally found in computer magazines looks into the microcomputer's future. One such article (11) has looked at the cycle of 8-bit microcomputer development. Four stages were evident: revolutionary hardware, but little software; stabilized hardware, powerful software; integration of hardware components and software; and lastly a stage in which the hardware was stretched to its limits.

These four stages occurred between 1973 and 1981. It is suggested that 16-bit microcomputers may follow the same cycle: 1979-81, hardware development; 1981-83, software development; 1983-85 mature hardware and software; and 1985-87 the development of powerful microcomputer "systems". The predictions of authors close to the microcomputer scene will be useful to analysts for predicting developments within their own laboratories.

The microcomputer has become a chemical instrument as much as any other "tool" found in the laboratory. However, it is unusual in that it is something that the individual analyst could afford to purchase, could use at home, and which could be considered an asset by his family. The microcomputer offers the analyst a new, personal and professional challenge.
5.2 Applied Statistics

The calculation of the mean and standard deviation of a set of data, and the standard error of the mean is a basic application of statistics. The calculation of confidence limits for the set and the mean is often the next application.

It is at this step that the analyst must begin to make some basic assumptions about the data. The correctness of these decisions will determine the validity of the conclusions he will reach about the data.

The statistics used for the program developed in this research assumes that the error of the analytical data has a normal (Gaussian) distribution, that the measurements are independent, and that the data set represents the population of measurements from which the set was drawn.

The program can check the data set for a non-normal distribution and for outliers. It can also plot the data in the form of a histogram. The independence of the measurements, and their ability to represent the complete population of such measurements, is dependent upon the design of the experiment conducted.

The analyst can never be sure that these important assumptions upon which the "statistics" are based are in fact satisfied. This is particularly true in the case of small data sets. The answers that result from the statistical treatment of the data cannot therefore be considered as absolutely correct. They range from representing the truth, to grossly misrepresenting it.

A second problem faced by the analyst using statistics is the tedium of carrying out the necessary calculations and of using statistical tables to determine the appropriate inferences. This is a chore that is efficiently handled by a combination of computer software and hardware. The microcomputer can thus aid, perhaps even encourage an increased use of statistics in the laboratory.

Much of the statistical software described and used in this research is based on common techniques (although these techniques are not necessarily commonly used). In particular, tests of significance, analysis of variance,
linear regression, and descriptive analysis are included. To maximize the usefulness of the program in the laboratory each of these functions have been customized for the analyst. In addition, any or all of the functions can be utilized with the same data set; the program is completely integrated.

A noteworthy feature of the significance tests (t-tests, sec. 2.5.1) are their ability to check for a bias between two sets of measurements.

Both cross-classification and hierarchic (nested) ANOVA (sec. 2.6) can be conducted. A table (Table 2.7) was produced to summarize the operations of each type of ANOVA. In addition the standard deviation for each tested source of variance in an experiment can be estimated. The confidence limits for the standard deviation of the residual (experimental) error can also be determined. However, the confidence limits for the standard deviations of the other sources of variance cannot presently be computed using the program; it would require an algorithm for generating critical "F" values, and one hasn’t been located yet.

These simpler applications of statistics were used to reprocess analytical data from the literature; the program worked well.

However, what became apparent in this research is the limited usefulness of most of these functions to the analyst. They were perhaps more important when the application of statistics to analytical data had to be conducted without a computer.

For the comparison of analytical methods in particular, these simple statistical tests looked very feeble compared to the power of linear regression. Their main shortcoming is the assumption of a constant analytical error variance which is independent of concentration. Since most analytical data do not have a constant variance then this problem must be considered before an expanded use of statistics in the laboratory can take place.
5.3 Weighted Linear Regression

The next thrust of this research was to develop a weighted linear regression program which could be used for processing calibration data. Existing programs seemed to depend on knowing the response variance of each standard used. This is an experimental drawback as the calculation of a good estimate of variance may require as many as twenty replicate measurements.

On reflecting, one realizes that simple linear regression by the least squares technique does not require that the error variance be known in advance. All that is required is knowledge that the variance is constant across the concentration range of interest. The variance, in the form of the standard error of regression, is actually calculated in the regression analysis.

The program allows the assumption of constant, as well as, several other assumptions about the error variance to be used. The most useful assumption is that the relative standard deviation of the error is constant,

\[ \frac{SD(\text{response error})}{\text{response}} = \text{RSD} = \text{a constant} \]

The computed standard error of regression is actually the analytical error's standard deviation.

If the assumption of constant relative standard deviation, i.e., constant precision, is correct, then the equation of the calibration function can be calculated without knowing the actual response variances. In addition the relative standard deviation of the response is determined by the program. This proves to be an extremely powerful statistical technique in the comparison of analytical methods.

The other assumptions which can be made about the error variance will also allow the equation of the calibration function to be computed. However, none of them appear to have a useful interpretation for the resulting standard error of regression.
5.4 Monte Carlo Simulation

A problem faced in this research was in testing the statistical tests as formulated in the program. In some cases previously published data was reprocessed. However, published data was limited in terms of both the amount, and its applicability. In general it was of limited use.

The generation of simulated data using the Monte Carlo technique solved this problem. Large data sets with known, and specifically selected, statistical parameters could be readily computed. A program to generate analytical data is described in section 2.4.

A small negative bias in the mean calculated for several data sets (Table 2.3) generated using this program was suggested by a cursory review of the results. This could not be confirmed by a series of t-tests. If the bias does exist it could be as a result of the random number generator built into the version of BASIC used; it is known that the quality of random number generators vary. Additional testing of this program for generating "Normal" analytical measurements may be warranted.

Simulated data was used in testing the weighted linear regression routine (ch. 3). Such data was also used for determining error functions and for the comparison of analytical methods (ch. 4). Thompson and Howarth (12-14) have made extensive use of Monte Carlo simulation in their study of the robustness of statistical methods.

5.5 Absolute Calibration

If it can be assumed that an analytical method has constant precision, then the problem of determining a good estimate of the linear calibration function is solved. Thompson and Howarth (12) have suggested that at concentrations well above the method’s detection limit that this assumption may be approximately correct.

Equations for estimating the concentration of "unknowns", and their confidence limits is then required. Scheffe' (15) and others (refer to ch. 3) have considered this problem. Miller’s (16) approach was followed in this research.
Miller considered two situations: the calibration curve is used only once or it is used an indefinite number of times. The first case would correspond to an analysis by the standard addition technique. The second to the more common application of a calibration curve for estimating the concentration of many "unknowns". The details of these calculations are discussed in chapter 3.

The end result of such calculations is that the confidence interval for "one" estimate of concentration using a calibration curve, is narrower than the confidence interval for the same unknown when the curve is used for "many" estimates. Lieberman, Miller, and Hamilton (17) refer to the confidence interval in the case of many estimates as the "unlimited simultaneous discrimination interval". The confidence interval for a concentration determined by a single use of a calibration curve is referred to by the same authors as a "non-simultaneous discrimination interval".

These calculations could be done manually—but it is doubtful that a manual method would be used on a routine basis. One disadvantage of manual calculations are the errors which routinely occur. Babbage’s (1) desire to produce a computing machine has been linked to this problem.

A relevant example are the unlimited simultaneous discrimination intervals produced by Lieberman et al. (17); they are incorrect due to an error in reading the statistical tables. A computer program, once "debugged", will allow the repeated calculations of these intervals, with the only chance of error resulting from the input of incorrect data.

The second advantage of a computer program is the speed of the calculations and the lower degree of operator skill required. The program allows the confidence limits to be calculated as quickly as the concentrations. Such limits can thus be calculated on a routine basis.

The weighted linear regression routine and Miller’s approach to determining confidence intervals was tested using Monte Carlo generated data (Tables 3.9, 3.10, 3.11). The regression routine produced "good" estimates of the expected equation parameters. In addition the number of
estimated concentrations falling outside Miller’s discrimination intervals were in reasonable agreement with the number expected.

An additional weighting scheme available allows the absorbance measurements to be weighted using transmittances (sec. 3.4.1 and 4.10) rather than the absorbances. Both weighting factors produced similar results; the absorbance factor is more easily applied, but it is less accurate.

5.6 Determining the Analytical Error Function

It is evident from the chemical literature (Table 4.1) that the analytical error of a method is usually assumed to be constant, and that its magnitude is very seldom evaluated thoroughly. This is understandable as a thorough evaluation of the error over the concentration range of interest could be very time consuming, and expensive. However, if the standard deviation of the error is constant or proportional to the response, then the experimental effort required to determine its magnitude could be substantially reduced.

Two approaches of evaluating the error function (sec. 4.5.2) have been described in the literature. Schwartz (18) has discussed the traditional method of replicate measurements at each of many concentrations across the range of interest. Thompson and Howarth (12,13) have used a method based on duplicate analysis of samples which cover the concentration range of interest.

According to Schwartz, the traditional approach (Replicate Analyses) to determining the error variance function requires 20 or more replicate measurements at 6 or more concentration levels. Each measurement must be from a separate sample taken through the complete analytical procedure. Thompson and Howarth (12) have pointed out the disadvantages of this method.

In Thompson’s approach (Duplicate Analyses) the analytical system is assumed to have a limiting error at zero concentration. At concentrations very much greater than the detection limit the system is assumed to have constant precision (fig. 4.2). A minimum of 100 independent responses are required, e.g., duplicate measurements of 50 samples.
In theory, Thompson's approach allows both the limiting constant standard deviation, and the constant relative standard deviation components of the error variance to be determined simultaneously. However, the determination of the error variance at zero concentration might be better estimated by using samples having concentrations close to the detection limit, and by using simple linear regression.

Two additional approaches (sec. 4.5.2) have been examined in this research. In the first, the response and concentration is required for each reference material. Simple or weighted linear regression is then used to calculate the standard error of regression.

In the second, the exact concentrations of the reference materials are not required. Duplicate, independent measurements of the response for each reference are obtained; these measurements are regressed against themselves using either simple or weighted linear regression.

The standard error of regression is either the standard or relative standard deviation of analytical error depending upon whether simple or weighted linear regression was appropriate.

The "Regression of Response on Concentration" approach (sec. 4.5.2) was tested using Monte Carlo simulation (secs. 3.6, 4.8.1, 4.9.1, 4.10.1; Tables 3.9, 4.7, 4.11, 4.14). The results produced agree closely with the expected standard or relative standard deviations. The estimates were as good as, or better, than those produced using either the Replicate or Duplicate Analyses approach. Although no direct reference to this approach was found in the literature it is a rather straightforward use of the weighted linear regression routine.

The "Self Regression" approach is unconventional; no literature reference to a similar calculation was found. A possible explanation as to why it works was formulated (sec. 4.5.2 eqs. 13-20). In this approach the original measurements (y-data set) are regressed against themselves, after the order of the replicate measurements for each reference sample is altered (x-data set).
Testing of this approach using Monte Carlo simulation produced good estimates of the expected analytical error (secs. 4.8.1, 4.9.1, 4.10.1; Tables 4.7, 4.11, and 4.14). The estimates were as good as, or better, than those produced by the Replicate or Duplicate Analyses approaches. They were also comparable to those produced by the "Regression of Response on Concentration" approach.

These latter two techniques produced satisfactory results with fewer measurements than both the Replicate and Duplicate Analyses approaches. An advantage of the Self Regression over the Regression of Response on Concentration approach is that the exact concentration of the reference samples is not required for the calculation.

The conditions under which these four approaches could be used are tabulated in Table 4.17. It is possible that in many experimental situations two or more approaches could be used to produce error estimates. Similar estimates would tend to substantiate one another; it should not, however, be expected that they would be identical.

Two important reasons for determining the analytical error function are related to comparative calibrations, and the determination of the relative technical merits of analytical methods.

5.7 Comparative Calibration:

Method Comparisons based on Accuracy

Rosenblatt and Spiegelman (19), and Williams (20) have discussed absolute versus comparative calibration. An absolute calibration is basically the preparation of the usual calibration curve, where the responses for an analytical method are regressed on the concentration of the standards. A comparative calibration is one in which the results of one analytical method are calibrated against the results of another.

The straightforward approach (sec. 4.6) has been discussed by Thompson (14). In this approach the concentrations produced by each method are regressed against one another. This allows a comparison of accuracy (sec. 4.6.2) in terms of possible bias. Thompson evaluated
the use of simple, weighted, and maximum likelihood regression methods for this evaluation (sec. 4.6.4). He concluded that the maximum likelihood method was the best.

Thompson also determined, however, that WLR tended to produce results very similar to those of MLR, although there are constraints to its application (Table 4.3). Studies conducted in this research confirmed the usefulness of WLR.

An important advantage of the WLR technique is that the error variances for each method do not have to be known. Although, some information on the nature of the error functions is required, hence one of the reasons for the necessity of being able to determine analytical error functions (sec. 5.6). WLR is therefore a more practical technique than MLR for comparative calibrations.

5.8 Comparative Calibration:

Method Comparisons based on Precision

The comparative calibration technique suggested by Mandel (21) does not require the conversion of "raw" response measurements to concentration. This technique results in a reasonably easy method for the evaluation of the relative "technical" and "economic" merits of competing analytical methods.

The comparison of analytical methods tends to be a favourite past-time of analysts, with the purpose usually being to find a cheaper method which produces comparable or better results (Table 4.1). Although the terms are not usually used, the experiments are evaluations of technical and economic merit.

Mandel has shown how his "relative sensitivity ratio" is a measure of technical merit, and how it can be determined (sec. 4.7.2) without having to convert response data to concentration. His work used only SLR.

In this research the relative sensitivity ratios were determined for methods simulated by the Monte Carlo technique. These included pairs of methods with analytical errors having: (i) constant standard deviations, (ii) constant relative standard deviations, and (iii) constant SD and RSD. The results produced were
comparable to the expected ratios (sec. 4.8.1, 4.9.1, 4.10.1; Tables 4.12, 4.16).

The relative sensitivity ratio also allows a calculation of the number of measurements required from the less precise method, to equal the precision of one measurement from the more precise method. The calculated results were again comparable to the expected values (e.g., Table 4.16). This calculation allows a determination of the economic merits of the methods once an evaluation of the cost for a measurement by each method has been made.

Although the comparison of analytical methods is common, it is uncommon to find a quantitative evaluation of the economic and technical merits of the methods. This is very surprising considering the frequency and importance of these endeavours.

5.9 A "Vertical Program" for the Analyst

The program developed in this research should be helpful in encouraging the routine use of statistics in the analytical laboratory. However, it represents only a very small step in the development of a "vertical" program for the analyst.

Its limitations are partly related to the capacity of ordinary 8-bit microcomputers, i.e., limited random access memory, mathematical accuracy, and computing speed. The advent of 16-bit microcomputers reduces these limitations.

The real limitation of this, and probably most analytical programs developed to date, is the wide scope of the potential analytical applications. The magnitude of the endeavour will require teams of analysts, programmers, and chemometricians to produce the substantial programs which will be required.

It will probably also require commissions, such as the IUPAC commissions, to study, evaluate, and recommend acceptable algorithms and procedures. Such commissions, however, will have to work in such a way that they will not hinder the software development which should arise from each new hardware development.
Analysts will not be able to rely on commercial software producers to the extent that the business community does. Business users represents a market which allows the development costs of successful software to be easily recouped. In addition, the marketplace will automatically select, by means of sales, the best of the software produced.

The development of vertical software packages, i.e., software which will be used only in an analytical setting and which will attempt to do an ever increasing number of functions, will be frustrated by developments in hardware, operating systems, and computer languages. Developments in all of these areas will continue; it appears that it will require much more time for obvious winners to be chosen. An example of the many possible choices can be seen in a recent issue of Microprocessors and Microsystems (22) which devoted a special issue to 16-bit operating systems.

If a choice of a system had to be decided upon today, the following might be a reasonable decision:

- the upcoming replacement (or look-alike) for the IBM PC. Hopefully this microcomputer will have a better microprocessor, i.e., the Intel 8086 or 80286, an arithmetic data processing unit such as the Intel 8087, better graphics resolution, one megabyte of memory or virtual memory, and a ten-plus megabyte hard-disk.

- an operating system capable of concurrent operation. The choice, would probably be between MS-DOS (using the MS-Windows (23,24) device driver) and CP/M-86. MS-DOS might be chosen because of the large amount of software currently running under it. The entry of AT and T into the microcomputer market will increase the importance of the UNIX operating system.

- the BASIC language. BASIC is known by most computer users, and in its interpretive form it allows the easy alteration and debugging of programs. In its compiled form (e.g., CBASIC) it has many of the advantages of newer languages, e.g., conserves memory, eliminates line numbers, more accurate mathematics, advanced handling of files, speed,
and structured programming (25). There are many other languages (26, 27); however, it might be difficult to get a consensus of opinion in selecting any one of them.

The one constant in the use of microcomputers is their constant development. This will remain an advantage and a disadvantage for years.

Immediate extensions to this applied statistics program could include:

- the capacity to handle larger data sets
- the capacity to handle more data sets simultaneously
- increased ability for manipulating and transforming data
- quality control routines
- the addition of routines to allow a similar treatment of linear and non-linear calibration data.

An additional development in the program would be the interfacing of analytical instrumentation (28) to the computer system allowing the direct reading of analytical data and its real-time processing. Inexpensive interfacing devices are available for this task.

5.10 Summary

Chapters 2 to 4 each discuss one of the major aspects of the research conducted. The purpose of this concluding chapter was to summarize these various discussions and to give an overall view of the main research results.
5.11 Conclusions and Recommendations

It was concluded that the microcomputer is a useful tool for the laboratory and that with the software produced in this research it will allow the routine determination of the following:

- useful analytical linear calibration functions
- confidence limits for estimated concentrations
- error functions of analytical methods
- quantitative measures for the relative technical and economic merits of analytical methods.

It will also allow the routine use of descriptive statistics, ANOVA, and significance tests by the analyst.

It is recommended that analytical societies work toward setting up committees to establish guidelines for the development of fully integrated software systems for the analytical laboratory, and that they more actively encourage the statistical analysis of analytical data.
REFERENCES - Chapter 1


56. Computer Series, J. Chem. Educ., 1979, 56, 79. (the Computer Series began with the above, there are now over 23 papers in the series)
58. Pace, B.T., Microcomputing, 1981, Mar., 120.
61. Price, D., Microcomputing, 1980, June, 82.
62. Lindsay, J., Microcomputing, 1980, May, 72.
64. Hogan, T., Microcomputing, 1980, Feb., 183.
76. Akers, R.L., Microcomputing, 1982, April, 46.

REFERENCES - Chapter 2

32. Burr, I.W., Technometrics, 1967, 9, 647.

REFERENCES - Chapter 3


43. Tebbutt, D., Creative Computing, 1981, 7(11), 27.
54. Erickson, J.O., and Surles, T., Am. Lab., 1976, 8(6), 41.
REFERENCES

17. Farnell, P.J., Analyst, 1975, 100, 661.

REFERENCES - Chapter 5

APPLIED STATISTICS I

Appendix A

User's Manual

For

Applied Statistics I

A Computer Program

By

G.N. Killoran

Lab Data Systems
This user's manual is written to describe the use of the computer program "APPLIED STATISTICS I." This program has been designed for the HEATH/ZENITH 89 computer. It is written for students and professionals studying or using statistics for the evaluation of experimentally obtained data. Although the examples used to illustrate the usefulness of the program are mostly chemical in nature, both the manual and the program are designed to be used in most scientific fields. Specifically they should be of use to students, engineers, scientists, and technologists of all subjects associated with chemistry. The program should also have some business applications.

It is expected that this program will be followed by a program dealing more specifically with statistical quality control (quality control charts, etc.). Users of APPLIED STATISTICS I are requested to forward their comments on this program and their suggestions for a STATISTICAL QUALITY CONTROL program.

The author wishes to thank the authors of the many statistics and quality control books and papers he has read and used over the years. Specific thanks are extended to John Wiley & Sons Ltd. for permission to use examples from "MATHEMATICS AND STATISTICS" for students of chemistry, chemical engineering, chemical technology and allied subjects" by C.J. Brooks, I.G. Bettsly, and B.M. Low (all such examples are indicated by an "*). Users of this manual may also be interested in referring to the "QUALITY CONTROL HANDBOOK" (J.M. Juran, Editor-in-Chief, published by McGraw-Hill, Inc.)

It should be pointed out that the author is an analytical chemist and teacher interested in statistical quality control but that he is not a statistician.

This program is the outgrowth of many programs the author has written and used over the years. These previous programs were written for the Wang 700 programable calculator, the Texas Instrument TI-59, and the Digital PDP 11/70 computer. This program, "APPLIED STATISTICS I", combined with the Zenith/Heath 89 computer has resulted in what the author believes to be an easy to use and comprehensive tool for basic applied statistics.

G.N. Killoran

TABLE OF CONTENTS

Disclaimer of All Warranties and Liability 11
Preface iii
Introduction 1
Preparing the Diskettes 3
For the First Time User 4
Histograms and Data Set Statistics 11
Comparing Two Sets of Data, t and F Tests 15
Statistical Tables 20
Analysis of Variance 21
Regression Analysis 25
Accuracy of Calculated Values 26
Program Support 28
Appendices 29

TABLE OF SCREENS

1 Program Title Page 5
2 Program Menu 5
3 Statistics for Data Set 7
4 Histogram 12
5 Statistics for Data Set 13
6 Results of General t Test 16
7 Comparison of Two Samples, Data in Pairs 17
8 Comparison of a Sample Mean with a Known Mean 19
9 Analysis of Variance Table - Two Factor 22
10 Two Factor ANOVA Table - Multiple Readings per Cell 23
11 Regression Analysis 26
12 Predicted Y Values and Confidence Limits 27
INTRODUCTION

Applied Statistics I is an integrated package of programs designed to carry out basic statistical analysis of experimentally obtained numerical data. It can be used by both students and professionals in an educational, industrial, or research setting.

This program was written to make available within a single package, a set of the more commonly used statistical routines. The uniqueness of this program is that it makes available to the user, all of the routines that it contains for the analysis of data, without the need to re-enter data or to load additional programs. Thus, although individual programs can be found to do parts of what this program will do, few microcomputer programs are available that can do what it does in such a simple manner. The user may actually find that this program is capable of doing more than what he is expecting of it.

As an educational tool, it can be used in introductory/intermediate level courses in statistics and statistical quality control at the community college level. In addition it can also be used as a computational tool for the analysis of data from undergraduate and graduate level experiments or research at the university level.

In an industrial/research setting it can be used for the more routine analysis of data. It has been designed to be used without the need of statistical tables; it will in fact, generate many of the tables normally used. The program is conversational in nature. It contains numerous error traps and prompts to ensure ease of use.

A minimal statistical/computer background is required to use the program. The user should either have taken or be taking a course in statistics. It is worth noting that the program will determine, automatically, the significance of the test results obtained, if the user has operated a Zenith/Heath 89 computer before then he should have little difficulty in using the program after a couple of practice sessions. One chapter of the manual is designed specifically for the "first time user"; it is tutorial in nature and will lead the user through the program's more basic features.

The program has been designed to take advantage of many of the unique features of the Zenith/Heath 89 computer. It is written as one complete program which, once loaded, can be used to do any of the available statistical tests on data in the computer memory. The data is initially entered via the keyboard. It can be viewed, edited, and stored/recalled from a diskette. The program also contains many features, for example automatic checking for outliers and data transformations which are not indicated in the menu. After the program is loaded the (HDOS, M Basic, Applied Statistics 1) disk can be removed and replaced by a data diskette for the storage of data sets. Over 200 sets of data with individual descriptions can be stored on each diskette; if the data sets contain more than 49 numbers, then fewer sets can be stored. These data diskettes can be replaced in the disk drive as required and thus an unlimited amount of data can be stored and later recalled. A multi-disk system is also supported by the program.

THE PROGRAM MENU

1. HISTOGRAM
2. DATA SET STATISTICS
3. COMPARISON OF TWO SAMPLE MEANS (t & F TEST)
4. COMPARISON OF TWO SAMPLES, DATA IN PAIRS (t TEST)
5. COMPARISON OF A SAMPLE MEAN WITH A KNOWN MEAN (USING THE KNOWN STANDARD DEVIATION) (T TEST)
6. COMPARISON OF A SAMPLE MEAN WITH A KNOWN MEAN (USING THE SAMPLE STANDARD DEVIATION) (T TEST)
7. PROBABILITY OF 'F' OCCURRING
8. PROBABILITY OF 'U' OCCURRING
9. PROBABILITY OF 't' OCCURRING
10. CRITICAL VALUES OF 't'
11. ANALYSIS OF VARIANCE (1 & 2 FACTOR - REPLICATE TESTS)
12. REGRESSION ANALYSIS
13. SAVE/RECALL DATA SETS ON/FROM DISK - END PROGRAM
14. CHANGE SET STATISTICS
15. CHANGE DISKS
16. DRIVE SELECT, CATALOG/KILL FILES
17. RECONFIGURE MEMORY

The Package includes the program, screen printer, and test data already recorded on a 5 1/4" diskette and this user's manual with its numerous examples. It requires Microsoft Basic (Rev. 4.7 or 4.82) and HDOS (1.6 or 2.0).

Hardware requirements are a Heathkit HBP or a Zenith 899 computer with a single disk drive and 48 K memory (64 K memory if a printer is used).
PREPARING THE DISKETTES

The programs for APPLIED STATISTICS I and the test data is shipped to you on a non-erasable disk. This means that the diskette does not contain
the Heath operating system (HDOS) or the BASIC language, both
of which are necessary to run this program. You are required to
prepare a disk containing HDOS and Microsoft BASIC to which you must
add this program and the test data.

First you should make a back-up copy of the distribution diskette.
Store the original distribution diskette in a safe place and use the
duplicate copy in the following procedures. (Two copies would be even
better, and please add copyright notices to each duplicate.)

To make a working diskette for APPLIED STATISTICS I follow these
directions and refer to your HDOS manual as necessary:

BOOT - Load a copy of HDOS (from your System Distribution disk).
INIT - Initialize a new diskette.
TEST - Test the media of the disk for bad sectors.
INIT - Re-initialize the diskette.
SYSGEN - copy HDOS to the new diskette (SYSGEN/WH with ver. 2.0)
ONECOPY - copy MBASIC to the new diskette.
ONECOPY - copy APPLIED STATISTICS I (APPLSTAT.BAS) and the test
data etc. to the new disk. Use "R:" to copy everything.

If you are using HDOS 2.0, MBASIC Rev. 4.02 and only 48 K memory you
must delete one routine to have sufficient free memory. Refer to
Appendix A for information on configuring the program to fix your
computer's memory size. If you are going to use a printer refer to
Appendix B for information on memory requirements and instructions on
loading the device drivers.

If you have not previously prepared a diskette it would be advisable
to seek the assistance of someone who has. This job need only be done
once or twice. While you are doing it you should also prepare a
couple of additional blank diskettes which will be used for data
storage. These need only be initialized, tested, and then
re-initialized.

The experienced H/Z 89 user may choose to prepare the working disk in
some other equally useful manner.

FOR THE FIRST TIME USER

This section of the manual will lead you through the basic steps which
will allow you to use APPLIED STATISTICS I. When you are finished with
this section you will probably be able to use the other routines
available in this program. The following sections of the manual will
explain these other routines in more detail, as well as, pointing out
additional features of the program.

In order to carry out the following instructions you should have a
working copy of the program prepared as outlined in the previous
section of this manual. If you have not already prepared such a disk
then do so now.

LOADING THE PROGRAM

Turn on your computer.

Ensure that the Upper Case lock is Down.

Insert the disk (your working copy) into the disk drive.

Boot-up the system.

Type:
MBASIC/F1

LOAD "APPLSTAT.BAS"

RUN

"DATA SET STATISTICS" ROUTINE #2

The title page will now be on the screen (SCREEN 1). Press the RETURN
key to continue. The menu is now on the screen (SCREEN 2). Press the
"2" key, then the RETURN key to select the "DATA SET STATISTICS"
routine. When asked what "DATA SOURCE" reply "0". This indicates
that the computer will be entering data from the disk, i.e., a set of
test data.

SCREEN TITLES

You should have noticed that the name of each of the routines that you
have been using are appearing as black type on a white background at
the bottom of the screen. Check SCREEN 1 and 2, as well as, some of
the others to see this. The program does this continually to tell you
what is happening as you use it. In addition the computer "beeps" each
time the screen title changes. This "beep" is an indication that
you should check the screen.
RECALLING DATA FROM THE DISKETTE

You are now being asked “EXAMINE FILE NAMES”? If you are using MBASIC, Rev. 4.7, reply “N” (this feature is not available in this version of MBASIC) otherwise reply “Y”. That response should have given you an error message. Always reply “Y” or “N” to a question of this nature. Reply with a “Y”. The computer will respond with a listing of the files contained on the disk and then asks “NAME?” answer “FIRST”. To the next question, “SET #?” answer “1”; the disk drive will again become active. The data in the file “FIRST.STA” is being loaded into the computer memory as set #1. (NOTE! You do NOT add the extension “.STA”, this is done automatically.) If for some reason the file could not be found then an error message would be given and you would then have to indicate whether you were saving or recalling data, your response would be “R” for recalling data in this case.

EDITING DATA

At this point the computer is asking you “EDIT DATA?”. This is your opportunity to see the data and to correct or add to it; answer “Y”. The data set will be printed on the screen; this is a short data set. Up to 330 numbers can be contained in a data set but only 100 can be shown on the screen at any one time. To the question “ALL CORRECT?”, respond “Y”. Had the data been incorrect and you had responded with “N” then a correction, additions, or a deletion could have been made. The “STATISTICS FOR DATA SET #1” should now be on the screen (SCREEN X). An explanation of this display is in the next section of the manual.

ACCESSING DATA IN THE COMPUTER MEMORY

Data already in the computer memory can also be accessed. Press “RETURN” to return to the menu. Again select routine #2. (NOTE! If you ask for a non-existent routine you will receive an error message.) This time select the memory (M) as the DATA SOURCE. A response of anything but “1” to the request for a set # will result in a message of “NO DATA” as only set #1 presently contains data. Try a few different replies, then reply with “1”. View the data, reply “N” to the question “ALL CORRECT”. Then try asking some corrections and deletions by responding with “C” or “D”. You will notice that you can make only one correction or deletion at a time, although you can add any number of additions “UP TO” the maximum of the set size selected (more on that later).

ENTERING DATA FROM THE KEYBOARD

When you have finished trying the editing routine have the statistics printed and then return to the menu. Again select routine #2; this time, however, select the keyboard (K) as the data source. Enter the following data as set #1:

SCREEN 1. Program Title Page

SCREEN 2. Program Menu
When the program asks for the sixth number enter "E", to indicate the end of data entry. The mean should be 61159.4 and the standard deviation is 1.14018.

RECONFIGURING THE COMPUTER MEMORY

Routine #17 "RECONFIGURE MEMORY", allows you to re-allocate the computer memory available for data set storage. 1 set of 330, 2 sets of 165, or up to 11 sets of 30 values each, may be selected using this routine. The first set (#0) is usually reserved as a work file unless it is the only file selected; therefore, up to 10 data sets (of 30 values) may be in the memory at one time. The default condition of 11 sets of 30 is automatically selected on program start-up. (NOTE! Be sure to check the menu label to determine the number of data sets stored in memory before entering new data sets from either the keyboard or the disk. Error checks will usually correct for a size mismatch but a simple manual check will be less time consuming.)

SAVING DATA ON THE DISK

Select routine #17 from the menu. Before you reconfigure the memory you must save (put) any data already in the memory (and which you want to save) onto the disk, otherwise it will be lost. To the question "SAVE1NG DATA?" respond "Y". Now practice using the save routine (#17 in the menu) by saving data set #2, call it "MYDATA15" and describe it as "TEST DATA". (NOTE! Use of an existing data file name will destroy that file. Also, the name must begin with a letter of the alphabet, cannot contain special symbols, and must not be more than eight characters long.) Reply "N" to the question "SAVING ANOTHER?" and reconfigure the memory to 1 set of 330 using routine #17. Return to the menu, examine the screen title. If you have more than one disk drive refer to the next section where the Drive Select routine (R16) is described.

USING LARGE DATA SETS - SIZE OF NUMBERS ALLOWED

Select routine #2 again. Try entering 300 numbers using the keyboard, use the repeat key to speed up the entry of numbers. You will notice that it is somewhat slower editing a set of this size. Only single precision numbers may be entered, i.e. numbers containing 6 or fewer significant figures. Numbers can be entered in scientific notation, e.g. 1.23456E+25. If the numbers to be entered contain more than 8 significant figures you can probably subtract a constant from each of them to reduce the number of significant figures to 6 or fewer. This will affect only the calculated mean; later add the constant to the calculated mean. File "TEST330" contains 330 numbers as an example.

Reconfigure the memory to 11 sets of 30 after completing this routine and returning to the menu.

ERROR HANDLING

You have already seen that some errors are handled in a rather fancy manner. Other errors, which you may encounter later, may be handled
more simply. If the program encounters an error which it cannot handle, then the error # and the line # in which the error occurred is printed. Check your Microsoft manual to determine what the error # refers to. An error message you may sometimes see is "REDO FROM START"; it just means to answer the question again. You will usually be able to continue the program without difficulty after typing CONTINUE. However, save your data on the disk occasionally just in case the program should happen to "bomb out".

If you do give an incorrect response and it is accepted by the program you may do one of three things:

- continue through with the routine if possible and then repeat it correctly.
- stop the program and return to the menu, then continue on if possible. Do this by pressing "CTRL C", then type "GOTO900".
- restart the program after attempting to save on the disk any important data already keyed in (use "CTRL C", GOTO900 as above). Then press "CTRL C" and type "RUN".

CHANGING DISKETTES

You have now used most of the standard features utilized in many of the menu routines; you could now possibly use the other routines selectable from the menu. If you do, it may be advisable to first "CHANGE DISK" (*15). Replace the working disk with a blank data disk (see the previous section for instructions on preparing it). The working disk has only limited space available for extra data storage. If you don't want to experiment yet on your own, then turn to the following sections for additional instructions and examples.

ENDING THE PROGRAM

The last routine to be discussed in this section is how to end the program. Select ROUTINE #13, reply "E" for "END"; you must confirm this decision. Should you want to restart the program with the same data in the memory, then immediately type "GOTO900". You will then return to the menu. Typing "RUN" will clear the memory of any existing data sets.

REVIEW

1) There are three sources of data; the keyboard, the computer memory, and the data diskette.
2) Data can be stored (with a label) on a data diskette for later use.
3) Data in the memory can be viewed, edited, and transformed.
4) The computer memory can contain 330 data set numbers as 1 set (#0), or up to 11 sets (#0 to #10) of 30 values each. The default setting is 11 sets of 30.
5) Data set numbers can contain no more than 6 significant digits; scientific notation is allowed.
6) All errors are trapped and most are handled automatically.
7) After ending the program normally or with "CTRL C", it may be restarted without losing memory data, by typing "GOTO900".
The HISTOGRAM routine (§1) calls on the DATA SET STATISTICS routine (§2). Both routines will be discussed in this section. As well, the CHANGE SET STATISTICS routine (§14) and the DRIVE SELECT, CATALOG/KILL FILES routine (§16) will be explained.

HISTOGRAM

The histogram routine could be used in a problem such as the following:

Construct a histogram for the following list of 80 laboratory determinations of the specific gravity of sulphuric acid.

To obtain the data for this problem select routine §1 from the menu. Recall from the disk, the file "HISTOGRAM" and store it as set §1. You have probably noticed that these directions have not worked. If the memory was set for 11 sets of 30 (the default condition), then because this set contains 80 numbers the program will automatically transfer to routine §17. That is, you are asked if you want to save any important data already in memory and then you reconfigure the memory. Select 3 sets of 110 numbers, and then follow the above directions to recall the data. Select the EDIT option to view the data; reply "y" to "ALL CORRECT?".

The program now returns with the range of the data set. You must enter the lower limit and the upper limit for your histogram. Any set number equal to or greater than the lower limit and LESS than the upper limit will be used in plotting the histogram; the (>) is a reminder to normally choose an upper limit greater than the upper limit of the data set. Of course, data can be deliberately excluded by selecting appropriate upper and lower limits. A consideration in choosing limits is to try and get easy to work with group intervals, e.g., an interval of .02 rather than .017. Next you must enter the number of intervals you want to have; any number from 1 to 20 is allowed. Use the following data:

LOWER LIMIT 1.675
UPPER LIMIT 1.915
NUMBER OF INTERVALS 12

The program prints the group width and then determines the frequency of each interval. The length of time this takes depends on the size of the data set and the number of intervals chosen. After pressing the return key upon request, the histogram (SCREEN 4) will be plotted. The histogram is scaled automatically to ensure that it will fit on the screen. If you require a copy of the histogram then record the group limits and the group frequencies. The histogram can then be easily drawn in a format which would be most appropriate for your particular application.

You may decide that the histogram is not as you want it; if so then repplot it using different upper and/or lower limits, and/or a different number of group intervals. Once you are satisfied with the histogram record the information you want and then view the statistics for the set.

SCREEN 4. Histogram
round-offs. Refer to the section on Accuracy of Calculated Values.)

CHANGE SET STATISTICS

If you are only given the mean, variance, and number of measurements in a data set (a common textbook situation) then this routine (814) can be used for entering them. (NOTE! The data set does not actually exist in the computer memory so DON'T try to edit it.) This routine can also be used for changing one of these statistics for an actual data set whenever a better estimate is obtained by some other means. For example if a better estimate of the variance is obtained in a "t" test by the combination of another data set with the first. An example of this is given in the next section. This is a simple routine: you must re-enter the mean, variance, and the set size. Enter the data as one entry, with each number being separated by a comma. Try altering the values for set #1. If you press RETURN only, the existing data will not be altered.

DRIVE SELECT, CATALOG/KILL FILES

This routine (816) allows you to specify one of three drives (0, 1, 2) which will be used for all disk operations until another drive is selected with this routine. The original default condition is drive 0. The routine then shows all the files on the disk currently in the selected drive. It allows you to KILL (i.e. delete any "STA" file). Use it to check the data files on your diskette and to delete any that you no longer require. Try deleting your data file, "MYDATA", when you decide not to delete any files (for any more files) reply "N" and return to the menu (remember HRASIC, Rev. 4.7 does not allow you to examine the file names).

REVIEW

1) This section discusses the HISTOGRAM, DATA SET STATISTICS, CHANGE SET STATISTICS, and the DRIVE SELECT, CATALOG/KILL FILES routines.

2) Histograms can be replotted until a suitable format is achieved. They are automatically scaled for the screen.

3) The descriptive statistics can be obtained for any set in memory.

4) The confidence limits of the set and of the mean, at the 90, 95, and 99% levels, are automatically calculated.

5) The mean, variance, and size of any data set (in memory) can be changed. As well these values can be entered for any set where the actual set numbers are not available.

6) One of three disk drives can be selected for all disk operations.

7) The files on the disk being used can be catalogued and any data files can be deleted.
This section of the manual describes ROUTINES #3 to #6. These routines (significance tests) are used when:

- the means of two data sets are compared
- paired data sets are compared
- a data set mean is compared to a known mean

These methods use hypothesis testing; i.e. the two means (or data sets) are considered to be similar unless it can be shown that the chance of this being so is very small. If you are unfamiliar with these techniques then you should refer to a statistics textbook such as Brooks’.

These tests are interpreted COMPLETELY by the program. An example of each of the tests is used to explain the operation of the routines.

COMPARISON OF TWO SAMPLE MEANS (t & F TEST)

This is a general test used for comparing the means of two data sets. It can be used for problems of the following type:

The following data (files T5A and T5B on the disk) give yields from two chemical processes. Process A is both more lengthy and more costly than Process B. It is considered worthwhile using Process A only if the average yield is more than 4 lb/batch higher than would be obtained using Process B. The probability of wrongly introducing Process A is to be no more than 0.05.

Should Process A be recommended?

This rather complicated problem will allow the use of many of the features of the program. Call ROUTINE #5, recall file T5A as set #1 and T5B as set #2; view the data as you recall it. You will then be asked about the comparison you want to make between the means; whether you are checking for any difference or whether the first is larger or smaller than the second and by how much (often the difference is just zero). In this example you are testing to determine if Process A gives a yield which is at least 4 lb/batch higher (larger) than Process B.

Before the means can be compared the variances of the two sets must be compared. Then, depending on whether they can be considered similar or not, the program selects one of two tests to perform on the means. If the variances are comparable then a value of "t" is calculated using the method outlined in most statistics textbooks. If the variances are not similar then the method (Welch test) outlined in Juran’s handbook (page 22-39) is used. This is a less frequently used test, often no test is done under these conditions.

The results of the test are then given (SCREEN #6). The mean and variance of each set is given, the calculated "F" ratio and the probability of it occurring by chance, the average estimate of the population variance as calculated from the two sets, the total degrees of freedom, and the calculated value of "t". Whether it was a one-sided or two sided test is then indicated, as well as, the test difference. The appropriate "null" hypothesis which is being tested is then stated and the probability of the null hypothesis being correct is indicated, PLUS its significance. That is, you are told whether the program considers the two means to be similar or not.

In this example it is determined that Process A does give a yield which is at least 4 lb/batch greater than Process B. This is assumed since the test indicates that the chance of the null hypothesis being correct is only 4.3%.

If the variances of the two groups cannot be considered to be equivalent then the Welch test is performed. (NOTE! In a printout similar to SCREEN #6 the population variance will be shown as "D" - no value is actually calculated - and the degrees of freedom is not (n1 4 n2 4 2) as is usual, but is actually something less.) As an example use routine #14 to enter the variance, mean, and set size of two sets e.g. 0.00003, 5.269, 13 and 0.00001, 5.261, 60. Then use routine #3 to test the similarity of the means. (NOTE! This test uses the "t table" of critical values which results in an approximate answer.) The results indicate that there is a very definite difference between the means.
COMPARISON OF TWO SAMPLES, DATA IN PAIRS (t TEST)

This test (routine #4) compares data which occurs naturally in pairs. The following question is an example of the situation in which such data is collected. (NOTE: In this example the previous test is not used as there is no reason to believe that the fibre strength is consistent amongst all samples.)

Six lengths of a certain type of synthetic fibre are halved, and one half of each pair (selected at random) dyed. The following data (files T4A and T4B) gives the coded values of a strength test.

Determine whether or not the strength is affected by dying if the level of significance is to be taken at the 0.05 probability level. Put 95% confidence limits on the expected change in strength due to dying.

Using routine #4 recall data sets T4A and T4B, view the data as you recall it. In this question you are testing for "any difference". The program will calculate the statistics for data set #0, i.e. the difference data set. The answer for the second question is given here as $0.95 \pm 0.91$. Next the results of the "t test" are given (Screen 7). In this case the null hypothesis is rejected and a real difference is asserted.

COMPARISON OF A SAMPLE MEAN WITH A KNOWN MEAN USING EITHER A KNOWN OR THE SAMPLE STANDARD DEVIATION

An example of a problem where the first routine is used is as follows:

The average daily scrap from a certain manufacturing process is $25.5 \pm$ (with standard deviation of $1.6 \pm$). A modification is suggested in an attempt to reduce the scrap.

It can be assumed from the nature of the modification that daily variability in the amount of scrap will not be affected.

The following data (file T5A) was obtained from a fortnight's (two week) trial of the modification.

Is the modification effective?

Put 95% confidence limits on the expected average daily scrap, using the modification.

Use routine #5, recall file T5A and enter the known mean as 25.5 and the known standard deviation as $1.6$. (NOTE: The first data set referred to is the new set of values and the second is the one for which you know only the mean and perhaps the standard deviation.) In this example the null hypothesis is rejected, and the alternate hypothesis that the modification reduces scrap is accepted (SCREEN 8).

In order to calculate the 95% confidence change the calculated variance of the data set to 1.6, i.e. $1.6 \pm$ 1.6, the known variance; use routine #14 to do this. Now use routine #2, the correct answer should be 23.90 $\pm$ 0.46. Unfortunately the routine does not give this answer; this is because it uses the value of "t" for 11 degrees of freedom rather than an infinite degree of freedom. Therefore, use routine #10 and a large degree of freedom e.g. 1000, to determine the correct value of "t". Then calculate the confidence limits manually, i.e. $\pm 1.96 \times 0.462 = 0.91$ (0.462 = S.E.).

Redo this problem (use routine #6), but this time use the calculated standard deviation of the data set; enter "0" to indicate this.(NOTE: You require the actual set variance. If you changed it as was suggested previously then recall the set from the disk and not from the memory.) Again the program asserts a real decrease in scrap. In this example the mean and its 95% confidence limit can be obtained directly using routine #2; they are $23.90 \pm 0.35$. 

SCREEN 7. Comparison of Two Samples, Data in Pairs
This section discusses significance tests. Two sets of data are compared.

2) Routine #14 can be used to change data set statistics to "better" values. Then routine #2 can be used to recalculate the data set statistics.

3) Five tests are illustrated.

4) Although it was not illustrated, a histogram of any data set in memory can be plotted.

5) Any of the data sets in memory can be compared; they need not be #1 and #2, or even consecutive data sets.
This routine (#1) will do a ONE or TWO factor analysis of variance. As well it handles replicate measurements for the TWO factor ANOVA. From 2 to 10 sets of data can be involved; the sets must be consecutive from #1 up to #10. The routine indicates the sum of the rows and the columns, produces the Analysis of Variance Table, calculates the "F" ratios, and indicates the significance of each ratio. In addition, in the case of a TWO factor ANOVA with replicate measurements, it will recalculate the "F" ratios if the interaction is significant and redetermine the significance of each new ratio.

The following example illustrates a one factor ANOVA:

Three methods of analysis are used to determine in parts per million, the amount of a certain constituent in a given sample. Six analyses are made by each method.

Determine whether or not the methods can be assumed equivalent and put 95% confidence limits on the mean values for each method. 

Call up routine #11; recall files ANOVA11, ANOVA12, and ANOVA13 from the diskette, view them, and place them in data sets #1, 2, and 3. You will note that you could also have entered new data via the keyboard. For this routine you would normally indicate the memory as the source only if you already had all your data in memory and you were simply repeating the ANOVA routine, just indicate a valid data set and then reply "N" to "ANOTHER SET?". Next indicate whether there are 1 or 2 factors (1 in this example). Then enter the names and size of the first (column) factor (METHODS, 3 in this example). This is followed by the name and size of the second (row) factor if there is one, else enter the number of measurements in each data set (4 in this example). If it is a 2 factor ANOVA then enter the number of replicate measurements per cell. In all of the above the program checks for any obvious input errors.

Next the program will print the data matrix table; for this example answer "Y" to the question "ALL CORRECT?". If there was an error use routine #2 to correct it and then recall the ANOVA routine. This is followed by the sum of the rows, and then the sum of the columns. The formatting is not elaborate, particularly if you are using 10 factors. This is then followed by the Analysis of Variance Table. Screen 9 is an example of this table when the six analyses by each method were made, one by each of six different analysts; i.e. a two factor ANOVA.

SCREEN 9. Analysis of Variance Table - Two Factor

After you have produced the ANOVA table for the first problem, calculate the confidence limits as previously illustrated. The mean and confidence limits of the first set are 0.163 +/- 0.013.

Next try this problem as a two factor ANOVA (methods and analysts), refer to Screen 9 for the table (remember, state the memory as the data source but just recall one data set).

A brief explanation of the symbols for the significance of the "F" ratios are as follows:

N.S. NOT SIGNIFICANT, probability greater than 10%.
P.S. POSSIBLY SIGNIFICANT, probability between 10 and 5%.
S SIGNIFICANT, probability between 5 and 1%.
S** HIGHLY SIGNIFICANT, probability between 1 and 0.1%.
S*** VERY HIGHLY SIGNIFICANT, probability less than 0.1%.

Refer to a statistics text for a more detailed explanation of these terms and the appropriate conclusions to be drawn. Detailed explanations of these significance tests are used in the comparison of means routines (t and F tests) illustrated previously.
As a final example consider the following problem:

Inhibitors are added to water treatment solutions to minimize their corrosive side effects. The effect of four different inhibitors were tested by adding them to three different water treatment solutions. The corrosive effect of each combination of solution and inhibitor was measured in duplicate. The coded information is stored in files ANOVA21, ANOVA22, ANOVA23, and ANOVA24.

Determine if the type of solution, type of inhibitor, or interaction (solution/inhibitor) is significant.

Screen 10 illustrates the ANOVA table for this problem. In this problem only the interaction is significant.

The user could carry on with the results produced by this routine to carry out analyses which are not done directly by the routine. For example, if interaction can have no physical meaning then combine the appropriate sources of variation by combining the calculated degrees of freedom and the sum of squares contained in the ANOVA table and then manually complete the problem.

The user can also use any of the previous routines with the ANOVA data. The data can be edited, histograms can be plotted, the statistics for individual groups can be obtained, better estimates of variance could be entered and the statistics recalculated, and "t" tests could be performed on any two sets of data. Finally of course any or all of the data sets could be stored on the diskette for future use.

**REVIEW**

1) One factor, two factor, and two factor ANOVA with multiple readings per cell can be performed.
2) Sums of rows and columns are printed.
3) ANOVA tables are printed.
4) F ratios and their significance are included in the tables.
5) F ratios are recalculated when interaction is significant.
REGRESSION ANALYSIS

This routine is used in the situation where it is believed that a possible linear relationship exists between two variables, x and y. The data available consists of two data sets, both of which are subject to experimental error, and which are related to each other by an exact physical or chemical law or by an approximate relationship.

The routine calculates the:

- linear regression equation
- coefficient of linear regression
- % variation in "y" explained by regression
- ANOVA of linear regression
- confidence limits of the slope
- confidence limits of mean "y"

In addition, the routine produces a table of the input data (x and y data sets) and the calculated y values with their 90% confidence limits plus their residuals. Lastly, the routine calculates the predicted "y" value and its 90% confidence limits for any value of "x". Thus it allows the calculation of the confidence limits of the "x" intercept (let x=0).

The following question illustrates the use of this routine.

A biochemical technologist conducts an experiment to examine how the growth rate of a certain bacteria increases with the passage of time. She sets up six cultures of the bacteria and studies each one's growth for a different period of time. The time periods, hours, are stored on the disk as file REGJ1 and the growth rate, units per hour, as file REG12.

At the 5% significance level, can the growth rate be considered to be linearly related to the length of the time period?

What is the equation of the regression line that will predict growth rates for the various time periods?

What proportion of the variation in growth rates among the bacteria can be attributed to the linear relationship between time and growth rate?

What are the 90% confidence limits on the slope and the intercept of the regression equation?

What are the predicted growth rates (and their 90% confidence limits) when the time of growth is 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20 hours?
### ACCURACY OF CALCULATED VALUES

The accuracy of the values calculated by this program is limited by both the BASIC language used and by the algorithms chosen by the author. The accuracy of the probabilities and the critical "t" values calculated has already been discussed. In general the author believes that the accuracy of the values generated by this program are sufficient for their use in Applied Statistics. However, the calculated values will in some cases be different from those which may be obtained by long hand calculations or those produced by other computer systems. (NOTE! Not all digits in calculated values are significant. The user must exercise his own judgement in rounding-off answers.)

### PROGRAM SUPPORT

LAB DATA SYSTEMS will support all program problems to the best of its ability. Depending on the nature of the particular problem, we may issue an updated disk, a written notice for software modification, incorporate the change in a future release, or follow any other course of action which we believe to be appropriate.

Users who would like the program customized for their own use are invited to contact the author with their specific requirements. Any program customizing will be done on a fee basis.

---

**ANOVA**

<table>
<thead>
<tr>
<th>Source</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F Ratio</th>
<th>Prob (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explained by regression</td>
<td>1</td>
<td>28.9</td>
<td>28.9</td>
<td>47.507</td>
<td>0.357</td>
</tr>
<tr>
<td>About regression</td>
<td>4</td>
<td>2.43333</td>
<td>0.608</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>5</td>
<td>31.333</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Regression**

- **REGRESSION COEFF.** = 0.960386
- **% EXPLAINED VARIATION** = 92.2341
- **CONFIDENCE LIMITS (90, 95, 99%) OF Y avg.**
  - +/- 0.67896, 0.884053, 1.46574
- **CONFIDENCE LIMITS (90, 95, 99%) OF SLOPE**
  - +/- 0.26289, 0.342392, 0.567478
APPENDIX A. CONFIGURING THE PROGRAM FOR YOUR COMPUTER

If your Z/H 89 has 64 K memory than the program as supplied will run on your computer without any alteration, this includes the two device drivers used for printing. If, however, your computer has only 48 K of memory than you will probably have to delete some lines of the program in order for it to have sufficient working memory. The following are some suggestions; you may wish to configure the program otherwise for your specific situation.

- a 48 K computer, HDOS 1.6, and either MBASIC Rev. 4.7 or 4.82 will run the program as supplied but without hard copy.
- a 48 K computer, HDOS 2.0, and MBASIC Rev. 4.7 may run the program as supplied (but without hard copy).
- a 48 K computer, HDOS 2.0, and MBASIC Rev. 4.82 will require that either routine # 11 or # 12 be deleted to run the program as supplied (but without hard copy). Load the program as per the instructions then type:
  
  DELETE 12000-12340 (this deletes routine # 11) or
  DELETE 12300-12320 (this deletes routine # 12)

(Replace the deleted line 12000(12300) with "12000 RETURN")

After the program and the drivers are loaded type:

PRINT FRE(0)FRE(R)

If the number returned is at least 2000 then the program should execute without running out of memory.

APPENDIX B. USING A PRINTER FOR HARDCOPY

The program is designed such that it can be used without a printer. The results of the different routines are presented as "screens". Each screen presents specific data which may be of interest to the user. This allows the program to be used with a "minimum" system; it also means that there is no paper cost. However, some users will have a printer and will want hardcopy results. A screen printer "device driver" is supplied with the program; the user must supply his own printer "device driver". These two drivers must be present on the disk that will be used for booting up the system. They must be loaded before loading "MBASIC".

Using the screen printer the user can now obtain a hardcopy of any screen by typing: CTRL T. Therefore, only the desired screens are printed.
APPENDIX B
BIBLIOGRAPHY - TECHNOMETRICS

7. Prescott, P., "Discussion of (3)," 1983, 25, 156.
40. Fung, C.A., and Hunter, W.B., "Discussion of (33)," 1979, 21, 411; Other discussions are available.
APPENDIX C
SOME SYMBOLS USED

Mathematical

+ addition - subtraction
* multiplication / division

X^Y X raised to the power of Y
SQR(X) Square root of X
EXP(X) "e" raised to the power of X
RND(X) a random number between 0 and 1
LOG(X) the natural logarithm of X
|X| the absolute value of X
xi x subscript i
9E37 9 x 10^-37

Abbreviations

s, SD, S.D. standard deviation
SE, S.E. standard error
V[x], s^2 variance of x, variance
CL, CI confidence limits, interval
D.F., DF degrees of freedom

RS relative sensitivity ratio
SLR simple linear regression
WLR weighted linear regression
MLR maximum likelihood regression
Ten different engineers, mathematicians, and technologists were asked to draw "the best straight line" between nine points. The data represented three replicate measurements at three different concentrations. It can be seen from the following figure that the majority of the lines drawn were between the computed simple and the weighted linear regression lines. There was an obvious attempt on the part of some participants to weight the line. The scatter of the lines, however, points out the lack of precision of the "eyeball and ruler" technique of line fitting.

APPENDIX E

A USE FOR PREDICTED RESPONSE TOLERANCE LIMITS

Figure 3.3, "A weighted calibration curve and its confidence band", was produced from Monte Carlo generated data. The confidence band was obtained by plotting computed response tolerance limits.

Figure A shows the equation of the line (expected: \( y = 1.0x + 1.0 \)) and its relative standard deviation (expected: 0.20).

Figure B shows the simulated data and the confidence band for the line. Note that only confidence band values for the used concentrations are given.

Figure C shows the confidence band values for concentrations from -1 to +4. They were determined by inputing the required value of concentration, and by inputing that the average of 9E37 responses was required.

\[ y = 0.924308x + 1.15766 \]

**ANOVA**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F RATIO</th>
<th>PROB (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLAINED BY REGRESSION</td>
<td>1</td>
<td>4.30221</td>
<td>4.30221</td>
<td>90.6892</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>ABOUT REGRESSION</td>
<td>30</td>
<td>1.42317</td>
<td>.0474391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>31</td>
<td>5.72529</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EQ'N OF LINE**

\[ y = 0.924308x + 1.15766 \]

**REGRSSION COEFF.** = 0.866849

**% EXPLAINED VARIATION** = 75.1428

**STANDARD ERROR OF REGRESSION (br) =** 0.217805

**CONFIDENCE LIMITS (90, 95, 99%) OF INTERCEPT**

\[ 1.15766 +/- .193627 , .252978 , .313553 \]

**ED = .114084**

**CONFIDENCE LIMITS (90, 95, 99%) OF SLOPE**

\[ .924308 +/- .165425 , .199284 , .268191 \]

**ED = .0975848**
### CALCULATED VALUES and CONFIDENCE Band for the COMPLETE LINE

<table>
<thead>
<tr>
<th>X</th>
<th>WEIGHT</th>
<th>Y obs'd</th>
<th>Y comp'd</th>
<th>+/- 95% CL</th>
<th>RESIDUAL INT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.746171</td>
<td>1.30754</td>
<td>1.27755</td>
<td>0.29715</td>
<td>0.129468</td>
</tr>
<tr>
<td>0</td>
<td>0.746171</td>
<td>1.30754</td>
<td>1.27755</td>
<td>0.29715</td>
<td>0.129468</td>
</tr>
<tr>
<td>0</td>
<td>0.746171</td>
<td>1.30754</td>
<td>1.27755</td>
<td>0.29715</td>
<td>0.129468</td>
</tr>
<tr>
<td>0</td>
<td>0.746171</td>
<td>1.30754</td>
<td>1.27755</td>
<td>0.29715</td>
<td>0.129468</td>
</tr>
</tbody>
</table>

### WEIGHTED SUM = -1.11014E-06

### PREDICTED RESPONSES and CONFIDENCE/TOLERANCE LIMITS

<table>
<thead>
<tr>
<th>MEAN X</th>
<th># Y RESPONSES</th>
<th>MEAN Y PREDICTED</th>
<th>90 - 95% LIMITS</th>
<th>RELATIVE LIMITS +/- %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9E+37</td>
<td>2.06897</td>
<td>1.87687</td>
<td>2.29706</td>
</tr>
<tr>
<td>1</td>
<td>9E+37</td>
<td>2.06897</td>
<td>1.87687</td>
<td>2.29706</td>
</tr>
</tbody>
</table>

### TOLERANCE LIMITS

**WEIGHTED SUM = -1.11014E-06**
The complete program with much of the experimental data is contained on a 5 1/4" single sided, hard-sectored floppy disk, which is located in the back cover of the thesis. A Zenith/Heath Z/H89 microcomputer with the CPM operating system and MBASIC-80 is required to read and use the program. An IBM-PC version of the program may be available at a later date.