Dichotomous transformations for statistical inference about odds ratios

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


Additional Information:

- This article was published in the Journal of Nonparametric Statistics [© Taylor & Francis]. The definitive version is available from: http://www.tandfonline.com/doi/abs/10.1080/10485250802471551

Metadata Record: https://dspace.lboro.ac.uk/2134/9165

Version: Accepted for publication

Publisher: © Taylor & Francis

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available 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/
Dichotomous transformations for statistical inference about odds ratios

Xiangning Huang\textsuperscript{a} and Baibing Li\textsuperscript{b}

\textsuperscript{a}Biostatistics, AstraZeneca R&D Charnwood, Loughborough LE11 5RH, United Kingdom

\textsuperscript{b}Loughborough University, Loughborough LE11 3TU, United Kingdom

Abstract

Dichotomous transformations for continuous outcomes are commonly used in practice. In this paper, we investigate dichotomisations for statistical inference about odds ratios in a situation where two underlying distributions from which independent samples are drawn are skewed and unknown. Under some mild conditions it is shown that a suitable choice of the cutpoint of a dichotomous transformation must lie within the range bounded by the two medians of the two underlying distributions, within which there exists a unique optimal cutpoint in terms of the asymptotic efficiency of point estimation and hypothesis testing. The issue of selecting a cutpoint is also linked to the choice amongst some existing nonparametric tests.

Keywords: Clinical trial; Cutpoint; Dichotomisation; Median test; Odds ratio

* Corresponding author. Email: b.li2@lboro.ac.uk; Tel. +44 1509 228841
1. INTRODUCTION

In clinical trials, when observed outcomes are continuous, two methods can be used to compare the treatment efficacy of a test drug against a placebo. One is to directly use the original continuous outcomes to draw statistical inference for the effect size defined as the standardised mean difference in a pre-specified efficacy measure. It is not unusual in clinical trials, however, particularly when the normality assumption of underlying distributions is questionable, for researchers to incorporate an alternative, where continuous outcomes are split into two categories and the treatment effect is characterised using an odds ratio. This approach is related to responder analysis in regulatory practice (see, e.g. [10] and [13]). Both approaches are quite common in practice. For example, Gotsche et al. [8] collects twelve studies on house mite control measures in the management of asthma for a systematic review, where seven studies report an effect size but the other five report an odds ratio. Abramson et al. [1] also find a mixture of approaches reported for immunotherapy in asthma.

Clearly the latter approach, dichotomisation, has an advantage of clinical relevance in many applications where a clinical diagnosis of a disease by experiencing a level beyond a specific cutpoint is of most interest. Examples where a continuous variable is dichotomised because a range of the variable has distinct clinical significance include blood pressure, birth weight, respiratory function, blood glucose, olfactory function, and depression ([17]). As Ragland [17] has pointed out, it is common practice in clinical and epidemiologic studies to dichotomise continuous outcome variables, in part because therapeutic and policy decisions are usually based on thresholds and cutpoints rather than on continuous values.

Further, from a statistical perspective, using the method of dichotomisations may also be beneficial in many circumstances. It is robust against outliers and the required assumptions about the underlying distributions are kept to a minimum [12]. One common criticism to dichotomisations is that it reduces statistical power. This is true under the normality assumption but becomes irrelevant in the general situation of non-Gaussian distributions.
When continuous outcomes are split into two categories, the resulting quantity of interest, such as the odds ratio for a treatment group compared to a control group and the difference in the two population proportions, depends on the chosen cutpoint. In practice, the cutpoint of a dichotomisation is usually selected on the basis of background knowledge, past experience, published studies or sometimes even in a relatively arbitrary manner. The selection of cutpoints in the absence of statistical guidance can, however, be potentially problematic. Friedman *et al.* [7, pp. 308] gives an example to demonstrate it can be misleading to apply a dichotomous transformation without a suitable cutpoint.

In the Friedman example above, two therapies were administered to two groups of patients. The change in heart rate in beats per minute was measured for each patient before and after the treatments. Each of three cutpoints was used to split the data into two categories and a chi-square test and Fisher’s exact test were performed for each of the scenarios. It was found the significance levels change from significant (or not significant) to not significant (or significant) when different cutpoints were used. This example suggests that by manipulating the cutpoint, one can observe a significance level less than 0.05 when there does not really seem to be a difference.

Since the choice of the cutpoint of a dichotomous transformation is important, considerable attention has been paid to this issue in both statistical and medical literatures. Ragland [17] and Deyi *et al.* [5] have investigated the issue of power reduction in dichotomisations when the outcomes of normal continuous variables are dichotomised. Not surprisingly, the power is shown to be lost to some extents when a dichotomous transformation of a normal variable is carried out. Selvin [18] and Kieser *et al.* [11] have further considered the relationships between dichotomisations and sample size determinations. Selvin [18] has calculated the ratio of required sample sizes with and without using a dichotomous transformation under the normality assumption. Kieser *et al.* [11] have investigated power and sample size determination when assessing the clinical relevance of trial results by responder analyses. The analysis in [11] is mainly on the basis of normality assumption but when the assumption of normal distributions is relaxed, a bootstrap procedure has been investigated.
In this paper we consider dichotomising continuous outcome variables under a general situation where the normality assumption is no longer valid. Rather, the underlying distributions are assumed to be skewed and unknown. We investigate dichotomous transformations of continuous outcomes for statistical inference about odds ratios and explore the links with some nonparametric methods, including Mood’s median test [16] and Mathisen’s test [15].

An issue related to the problem investigated in this paper is to dichotomise a predictor variable when outcomes are binary (see, e.g. [6] and [14]). This paper, however, does not pursue this topic. We also note that there are extensive studies regarding dichotomisation of variables in the epidemiologic field (see e.g. [17] and [18]). In this paper we focus on the context of clinical trials although the results also apply to epidemiologic studies.

The following Section is devoted to the main results. Practical examples are examined in Section 3 and concluding remarks are given in Section 4. Proofs for the theorems are given in the Appendix.

2. MAIN RESULTS

Consider a two-arm parallel group, placebo controlled clinical trial to investigate the efficacy of a test drug in the treatment of a disease. Eligible subjects are randomly allocated into test drug or placebo groups. Let \( x_i \) \( (i = 1, \ldots, m) \) be continuous efficacy outcomes observed in the test drug group having a distribution of \( F(x - \mu_x) \), and let \( y_j \) \( (j = 1, \ldots, n) \) be continuous efficacy outcomes observed in the placebo group having a distribution of \( G(y - \mu_y) \), where \( \mu_x \) and \( \mu_y \) are the medians of the two distributions so that \( F(0) = G(0) = 1/2 \).

Suppose \( F(x) \) is known \textit{a priori} to differ from \( G(x) \) in the shape of distributions so that testing the difference between \( F(x - \mu_x) \) and \( G(y - \mu_y) \) is not of interest. The purpose of the analysis is rather to investigate whether the two populations have the same median, i.e. the statistical inference problem that we consider is a test of

\[
H_0: \mu_x = \mu_y \quad \text{versus} \quad H_{11}: \mu_x \neq \mu_y,
\]
with no specific shape assumptions on $F$ and $G$.

It is assumed the two unknown distributions $F(x)$ and $G(x)$ are absolutely continuous and possibly skewed. Let $f(x)$ and $g(x)$ denote the corresponding densities. To uniquely define a median, we suppose $f(0)>0$ and $g(0)>0$. However, it is not assumed the distributions have any specified functional forms. Clearly neither the t-test nor the Wilcoxon-Mann-Whitney test is suitable in this particular circumstance, for the former requires a normality assumption and the latter assumes a location shift is the sole distributional difference (see e.g. [20]).

Now consider an alternative approach where binary coding is applied to split the continuous outcomes. Define a dichotomous transformation of a continuous variate $z$ to be

$$T(z \mid \tau) = \begin{cases} 0 & \text{if } z \leq \tau \\ 1 & \text{if } z > \tau \end{cases},$$

where $\tau$ is the cutpoint of the dichotomous transformation $T(z \mid \tau)$ such that the continuous outcomes in each treatment group are split into two categories: favourable response and non-favourable response. Without loss of generality, suppose the category where observations are less than, or equal to, the cutpoint $\tau$, represents the favourable response.

The dichotomous transformation $T(z \mid \tau)$ is applied to the observations, $x_i$ ($i=1,...,m$) and $y_j$ ($j=1,...,n$). Denote $a$ as the number of observations in the test drug group with a favourable response and $b = m - a$. Likewise, define $c$ and $d$ for the placebo group, as displayed in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>$\leq \tau$ (favourable)</th>
<th>$&gt; \tau$ (unfavourable)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test drug</td>
<td>$a$</td>
<td>$b$</td>
<td>$m$</td>
</tr>
<tr>
<td>Placebo</td>
<td>$c$</td>
<td>$d$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

In practice, the logarithm of an odds ratio is usually estimated by $\ln(ad/bc)$, and the corresponding variance is estimated by the Woolf’s method: $\hat{V} = 1/a + 1/b + 1/c + 1/d$ (see, e.g. [9]).
The null hypothesis for the detection of whether there is a treatment difference between the test drug and placebo is:

\[ H_{02} : \ln(OR) = 0 \text{ versus } H_{12} : \ln(OR) \neq 0, \]

where \( \ln(OR) \) is the logarithm of the odds ratio when a dichotomous transformation \( T(z \mid \tau) \) is applied to split the outcomes:

\[ \ln(OR) = \ln \left( \frac{F(\tau - \mu_x) \{1 - G(\tau - \mu_y)\}}{[1 - F(\tau - \mu_x)] G(\tau - \mu_y)} \right). \]

When splitting data into two categories, it is important to raise the question: is testing \( H_{02} \) equivalent to testing \( H_{01} \)?

**THEOREM 1.** Suppose two distributions \( F(x - \mu_x) \) and \( G(y - \mu_y) \) are absolutely continuous with \( F(0) = G(0) = 1/2 \). Their densities \( f(x - \mu_x) \) and \( g(y - \mu_y) \) satisfy \( f(0) > 0 \) and \( g(0) > 0 \). If the cutpoint \( \tau \) of a dichotomous transformation \( T(z \mid \tau) \) is taken anywhere between \( \mu_x \) and \( \mu_y \), then \( \mu_x = \mu_y \) if and only if \( \ln(OR) = 0 \).

Theorem 1 shows that if the cutpoint of a dichotomous transformation is taken somewhere between the two medians of the underlying distributions, then it is guaranteed that testing the null hypothesis \( H_{01} \) is equivalent to testing \( H_{02} \).

Next, we show that the choice of a cutpoint relates to the asymptotic efficiency of point estimation and hypothesis testing. We first note that when the sample sizes become large, counts \( a \) and \( c \) approach to their expected frequencies, \( mF(\tau - \mu_x) \) and \( nG(\tau - \mu_y) \) respectively. Consequently, the estimated variance approaches

\[ V = \frac{r F(\tau - \mu_x) + r [1 - F(\tau - \mu_x)] + 1 / G(\tau - \mu_y) + 1 / [1 - G(\tau - \mu_y)]}{n}, \]

where it is assumed \( n/m \to r \), where \( 0 < r < +\infty \), as sample sizes approach infinity.
THEOREM 2. Suppose densities \( f(x - \mu_x) \) and \( g(y - \mu_y) \) are log-concave and continuously differentiable. Then there exists a cutpoint \( \tau^* \) lying between \( \mu_x \) and \( \mu_y \) such that the asymptotic variance of \( \ln(ad/bc) \) attains the global minimum.

Furthermore, it is immediate from Theorem 2 that we have

COROLLARY 1. Under the conditions of Theorem 2, if the cutpoint \( \tau \) of a dichotomous transformation \( T(z|\tau) \) lies between \( \mu_x \) and \( \mu_y \), then the asymptotic variance of \( \ln(ad/bc) \) attains the global minimum under \( H_{02} \).

Theorem 2 ensures the existence of the optimal cutpoint that leads to the asymptotically most efficient estimate of \( \ln(OR) \) among all dichotomous transformations \( T(z|\tau) \). It, however, does not indicate, except for a range, where exactly this optimal cutpoint lies. From the proof of Theorem 2 the location of the optimal cutpoint relies on the specific forms of the underlying distributions. In practice, two reasonable choices for cutpoints can be: (a) the median of the combined sample; and (b) a weighted midway point of the two medians, \( \tau_0 = (\omega_y \mu_x + \omega_x \mu_y)/(\omega_y + \omega_x) \), where \( \omega_x \) and \( \omega_y \) are the corresponding inter-quartile ranges of the two underlying distributions. For the special case investigated in [11] where \( F(x - \mu_x) \) and \( G(y - \mu_y) \) are two normal distributions having equal variance with equal sample sizes, it can be seen that \( \tau_0 \) satisfies the first-order condition in the proof of Theorem 2 in the Appendix so that \( \tau_0 \) is an optimal cutpoint for that particular circumstance.

In practice, the z-test is frequently used to test the null hypothesis \( H_{02} \) when sample sizes are not small [9]. Not surprisingly, the power of the z-test relies on the choice of the cutpoint \( \tau \). In fact, it is easy to find that the power of the z-test is a strictly decreasing function of \( \hat{V} \). Hence, asymptotically the power of the z-test attains maximum, i.e. the type II error is kept to a minimum, if the optimal cutpoint \( \tau^* \) indicated in Theorem 2 is used in a dichotomous transformation.
COROLLARY 2. Under the conditions of Theorem 2, applying the dichotomous transformation with the optimal cutpoint $\tau^*$ leads to a maximum power of the $z$-test for testing $H_{02}$ versus $H_{12}$ among all dichotomisations $T(z|\tau)$.

Remark: The conclusions drawn in Theorem 2 and Corollaries 1 and 2 do not rely on any specific forms of underlying distributions. Rather, they can be anyone in the log-concavity (also called strongly unimodal) distribution family. The assumption of log-concavity is widely used in statistical literature. Mathematically, this assumption is critical to guarantee a global minimum of the asymptotic variance. From a practical perspective, the log-concavity distribution family includes many commonly used unimodal distributions, such as normal, uniform, logistic, exponential, extreme values, gamma with shape parameter greater than or equal to 1, Weibull with the shape parameter greater than or equal to 1 (see [2] and [3] for details).

Before concluding this section, we note that selecting a cutpoint of a dichotomous transformation for testing $H_{02}$ is linked to the choice amongst nonparametric tests, including Mood’s median test [16] and Mathisen’s test [15], where the former method incorporates a cutpoint of the median of the combined sample, and the latter method has a cutpoint of $\mu_x$. For either method, the favourable and non-favourable responses are counted for each treatment group, as displayed in Table 1 (see, e.g. [20, pp. 86]). The null hypothesis $H_{01}$ is then tested using a hypergeometric distribution under $H_{01}$, or based on a normal approximation when sample sizes are not small. It is worth noting that the median $\mu_0$ of the distribution, $H(x) = \{1/(1+r)\}F(x-\mu_x) + \{r/(1+r)\}G(x-\mu_y)$, from which the combined sample is drawn lies between $\mu_x$ and $\mu_y$, although, in general, it is not the optimal cutpoint. Theoretically selecting any other percentile of $H(x)$ lying between $\mu_x$ and $\mu_y$ leads to a non-parametric test. However, without knowledge of the shapes of the two underlying distributions $F(x)$ and $G(x)$, it is hard to justify that such a choice is better than the Mood’s median test.
3. PRACTICAL EXAMPLES

To illustrate the results obtained in the previous section, we consider an example investigated in [4], where the data comprise the average capillary densities for both feet for 19 healthy subjects and 23 ulcer patients. To test if the population of ulcer patients has the same mean value as the healthy population, Bland [4] performs the t-test, resulting in a calculated t-value of 5.08. The null hypothesis was thus rejected at the significance level of 1 %.

We applied dichotomous transformations to the capillary density data to investigate the impact of the choice of a cutpoint. Totally 100 cutpoints were selected between the median of the ulcerated patient group, 22.5, and the median of the healthy group, 34.5, with a constant increment step of 0.12. The weighted midway point of the two sample medians is $\hat{\tau}_0 = 28.1$ and the median of the combined sample is $\hat{\mu}_0 = 27.25$.

Figure 1 gives the plot of the estimated standard deviation versus cutpoint. It is noted that an appropriate cutpoint should be somewhere between 27 and 28.5, where the corresponding estimated standard deviation is much smaller than that in the remaining area. In particular, a cutpoint less than 27 or greater than 29 results in a much larger value of the estimated standard deviation, demonstrating a very poor quality of the estimated log odds ratio.

Fig.1: Estimated standard deviation versus cutpoint
(Figure 1 is about here)

Figure 2 displays a plot of p-value versus cutpoint when the z-test was performed to test the null hypothesis $H_0$. Choosing a cutpoint in the interval [27, 29] has a maximum chance to give a result being consistent with that obtained by the t-test for the continuous outcomes. In this example both $\hat{\tau}_0 = 28.1$ and $\hat{\mu}_0 = 27.25$ are suitable choices.
Returning to the heart rate data mentioned in Section 1, the two sample medians coincide so the only reasonable choice for the cutpoint is the median \textit{per se}. This leads to the optimal estimate of OR which is equal to 1, and a result for hypothesis testing for dichotomised observations which is consistent with a t-test for the raw measures carried out in [7].

4. CONCLUDING REMARKS

This paper investigates dichotomisations for continuous outcomes. It is motivated by the example in [7] where the choice of a cutpoint could be misleading by manipulating the cutpoint.

Overall, dichotomisations are very useful in clinical trials due to its clinical relevance [12] but it is controversial from a statistical perspective. As demonstrated in previous studies such as [5], [11], [17] and [19], dichotomisations reduce statistical power because some of the information contained in the underlying distribution is discarded. We share the concerns raised in these studies and believe that, from a statistical perspective, data should not be dichotomised when they are approximately normally distributed.

In practice, however, it is not unusual that true distributions are unknown and robustness is a big issue. In these circumstances dichotomisations can play an important role where most of the criticisms on dichotomisations become irrelevant because the efficiencies in these criticisms were usually calculated under the normality (or symmetry) assumption in the first place.

In this paper we have investigated the general situation of non-Gaussian distributions. Our study shows that a suitable cutpoint of a dichotomous transformation must lie within a certain range bounded by the two medians of the two independent samples. Further, within this range, there exists a unique optimal cutpoint in terms of point estimation and hypothesis testing. We have also linked the issue of selecting a cutpoint to the choice of a nonparametric test. Consequently we have provided
statistical guidance for the choice of cutpoints of dichotomisations from the perspectives of estimation
and hypothesis testing of odds ratios. We hope this is helpful for researchers in defining
dichotomisations which are both statistically justified and clinically relevant.

APPENDIX

Proof of Theorem 1. Without loss of generality, we suppose \( \mu_x \leq \mu_y \). Since the cutpoint \( \tau \) lies
between the two medians, we have

\[
F(\tau - \mu_x) \geq 1/2 \quad \text{and} \quad G(\tau - \mu_y) \leq 1/2. \tag{A1}
\]

Now if \( \ln(OR) = 0 \), then \( F(\tau - \mu_x) = G(\tau - \mu_y) \). From (A1) we obtain \( F(\tau - \mu_x) = G(\tau - \mu_y) = 1/2 \).

Since \( f(0) > 0 \) and \( g(0) > 0 \), we have \( \mu_x = \mu_y \). On the other hand, if \( \mu_x = \mu_y \), then \( \tau = \mu_x \) and \( \tau = \mu_y \) since \( \tau \) lies between the two medians. Hence, we have \( \ln(OR) = 0 \). This completes the proof.

LEMMA. Suppose that \( u(x) \) is a log-concave and differentiable density. Let \( U(x) \) be the
corresponding cumulative distribution. Then \( J(x) = 1/U(x) + 1/[1 − U(x)] \) is convex.

Proof. By some algebra we obtain

\[
d^2 J / dx^2 = u^2(x)[1/U^3(x) + 1/[1 − U(x)]^3]
+ [u^2(x) − u'(x)U(x)]/[U(x)]^3 + [u^2(x) + u'(x)\{1 − U(x)\}]/[1 − U(x)]^3,
\]

where \( u'(x) = du(x)/dx \). Since \( u(x) \) is log-concave, from Lemma 2 in [2] we have

\[
u^2(x) − u'(x)U(x) \geq 0 \quad \text{and} \quad u^2(x) + u'(x)\{1 − U(x)\} \geq 0,
\]

which lead to \( d^2 J / dx^2 \geq 0 \). This completes the proof.

Proof of Theorem 2. Without loss of generality, we suppose \( \mu_x \leq \mu_y \). It is easy to find that

\[
dV / d\tau = [rf(\tau - \mu_x)(1 − F(\tau - \mu_x))^2 − (F(\tau - \mu_x))^2]
+ g(\tau - \mu_y)(1 − G(\tau - \mu_y))^2 − (G(\tau - \mu_y))^2]/n.
\]
Let $\delta = \mu_y - \mu_x \geq 0$ and denote $dV/d\tau$ as $h(\tau)$. Then we have

$$h(\mu_x) = g(-\delta) \left[\{1 - G(-\delta)\}^{-2} - \{G(-\delta)\}^{-2}\right] / n.$$  

Since $G(-\delta) \leq 1/2$, we obtain $h(\mu_x) \leq 0$. Similarly we can show $h(\mu_y) \geq 0$. Hence, there exists $\tau^* \in [\mu_x, \mu_y]$ such that $h(\tau^*) = 0$. From the Lemma above, $V$ attains the global minimum at $\tau^*$. This completes the proof.

ACKNOWLEDGEMENT

The authors would like to thank the referee for his/her valuable comments on an earlier version of this paper which have improved the quality of this paper.

REFERENCES


Fig. 1: Estimated standard deviation *versus* cutpoint
Fig. 2: The p-value versus cutpoint