TWO-SAMPLE TESTS WITH UNORDERED PAIRS

BY

D. V. HINKLEY

TECHNICAL REPORT NO. 28
AUGUST 10, 1971

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-14

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
TWO-SAMPLE TESTS WITH UNORDERED PAIRS

by

D. V. HINKLEY

TECHNICAL REPORT NO. 28
August 10, 1971

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-14

This report was supported in part by
National Science Foundation Grant GP-24918

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SUMMARY

Pairs of observations, \( x \) and \( y \), are taken on \( n \) individuals such that all observations are independent normal with a common variance, and \( x \) and \( y \) have possibly different means. The ordering of each pair \((x, y)\) is unknown, so that only \( \min(x, y) \) and \( \max(x, y) \) are available. Some simple tests of equality of means are discussed for known and unknown variance.
1. INTRODUCTION

The problem we are concerned with is the following. Pairs of observations \( x \) and \( y \) are taken on \( n \) individuals, where we may assume for simplicity that all observations are independent, \( x \) is \( N(\mu_1, \sigma^2) \), and \( y \) is \( N(\mu_2, \sigma^2) \). We wish to test \( H_0: \mu_1 = \mu_2 \) and possibly estimate \( \mu_1 - \mu_2 \), but the catch is that the ordering of each pair \( (x, y) \) is not known. Effectively we observe \( u = \min(x, y) \) and \( v = \max(x, y) \) on each individual, and we are forced to work with \( |\mu_1 - \mu_2| \geq 0 \) instead of \( \mu_1 - \mu_2 \).

It may be that this situation is uncommon, but it certainly does arise. For example, in genetics some studies are concerned with the question: are two chromosomes different in some way, and can we establish the existence of two "populations" by measuring lengths of pairs in several cells? A specific example arises in a recent unpublished paper by Efron, Miller and Brown (1971) at Stanford University.

In Section 2 we look at the problem under the assumption of known variance. Three statistics are considered for testing \( H_0: \mu_1 = \mu_2 \), and the three are shown to differ very little with respect to power.

The more likely situation, when the variance is unknown, is discussed in Section 3, where we also consider the possibility that between-pairs effects are present. The discussion is at a formal rather than rigorous level since the problem is quite regular. However, we might note that the problem provides a simple example where standard asymptotic properties of likelihood inference fail to hold in the null case \( \mu_1 = \mu_2 \).
2. TWO-SAMPLE TESTS WITH KNOWN VARIANCE

Suppose that we observe a random sample

\[ u_i = \min(x_i, y_i) \quad \text{and} \quad v_i = \max(x_i, y_i) \quad (i=1, \ldots, n), \]

where \( x_i \) and \( y_i \) are independent \( N(\mu_1, \sigma^2) \) and \( N(\mu_2, \sigma^2) \) respectively with \( \sigma^2 \) known. We wish to test \( H_0: \mu_1 - \mu_2 = 0 \) and possibly estimate \( |\mu_1 - \mu_2| \). Only two-sided alternatives to \( H_0 \) can be considered, and the differences \( d_j = v_j - u_j \) are minimal sufficient for \( \theta = |\mu_1 - \mu_2|/\sigma \).

We look at three statistics for testing \( H_0 \).

2.1 The sample average \( \overline{d} \)

First let us consider inference about \( \theta \) using the naive statistic \( \overline{d} = \sum d_j / n \). Simple calculation shows that

\[(2.1) \quad E(d) = \sigma \{ \theta - 2A(\theta) \} = \sigma a(\theta) \]

and

\[ \text{var}(d) = \sigma^2 \{ 2 + 4A(\theta) - 4A^2(\theta) \} = \sigma^2 b^2(\theta), \]

say, where

\[ A(\theta) = \theta \Phi(-\theta/\sqrt{2}) - 2 \Phi(\theta/\sqrt{2}). \]

Note that

\[ A(0) = -\pi^{-1/2}, \quad A'(0) = 0 \quad \text{and} \quad \lim_{\theta \to \infty} A(\theta) = 0. \]

Thus our first test of \( H_0 \) has rejection region

\[(2.2) \quad D_n = \frac{\sqrt{n} \left\{ \overline{d} + 2\sigma A(0) \right\}}{\sigma \sqrt{2 - 4A^2(0)}} \geq c_n, \]
where \( c_n \) determines test size. The asymptotic normality of \( D_n \) implies that
\[
c_n = c(\alpha) = \Phi^{-1}(1-\alpha) \quad \text{for test size } \alpha.
\]
Further, the large sample power function is
\[
P(D_n, \alpha; \theta) = 1 - \Phi \left\{ \frac{c(\alpha)b(0)+\sqrt{n}[a(0)-a(\theta)]}{b(\theta)} \right\}.
\]
(2.3)

It is easy to see that the \( D_n \) test is unbiased and that
\[
P''(D_n, \alpha; 0) = \{ I_2 c(\alpha) + \sqrt{\frac{n}{2\pi-4}} \} \Phi[c(\alpha)].
\]
(2.4)

This contrasts with the usual two-sample statistic (for ordered pairs)
\[
Z_n = \sqrt{n} \frac{(\bar{x} - \bar{y})}{(\sigma \sqrt{2})},
\]
whose power function has second derivative of order \( n \) at \( \theta = 0 \).

Consequently for fixed \( \alpha \) and power \( \beta \) at \( \theta = \theta^{(n)}_a \), \( \theta^{(n)}_a \propto n^{-1/2} \) for the \( D_n \) test whereas \( \theta^{(n)}_a \propto n^{-1/4} \) for the \( Z_n \) test. Table 2.1 gives a few values of \( P(D_n, \cdot .05; \theta) \) and the corresponding power \( P(Z_n, \cdot .05; \theta) \) for \( n = 20 \) and \( n = 40 \). The accurate normal approximation (2.3) is used for \( D_n \).

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( P(D_{20}, .05; \theta) )</th>
<th>( P(Z_{20}, .05; \theta) )</th>
<th>( P(D_{40}, .05; \theta) )</th>
<th>( P(Z_{40}, .05; \theta) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.063</td>
<td>0.124</td>
<td>0.068</td>
<td>0.201</td>
</tr>
<tr>
<td>0.50</td>
<td>0.114</td>
<td>0.353</td>
<td>0.144</td>
<td>0.609</td>
</tr>
<tr>
<td>1.00</td>
<td>0.426</td>
<td>0.885</td>
<td>0.620</td>
<td>0.994</td>
</tr>
<tr>
<td>1.50</td>
<td>0.851</td>
<td>0.997</td>
<td>0.976</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 2.1. Powers of \( D_n \) and \( Z_n \) with \( \alpha = 0.05 \)
We can estimate \( \theta \) consistently by the method of moments using \( \overline{d} \).

If this estimate is denoted by \( \tilde{\theta} \), then from (2.1) we have

\[
\overline{d} = \sigma(\tilde{\theta} - 2A(\tilde{\theta})) > \tilde{\theta}\sigma \quad (\overline{d} > \sigma a(0))
\]
\[
\tilde{\theta} = 0 \quad (\overline{d} < \sigma a(0))
\]

Since \( a(\theta) \geq \theta \) and is convex, the iterative solution

\[
\tilde{\theta}_{r+1} = \tilde{\theta}_r + \frac{\tilde{\theta}_r - a(\tilde{\theta}_r)}{\sigma a'(\tilde{\theta}_r)}, \quad \tilde{\theta}_1 = \overline{d}/\sigma
\]

converges down to \( \tilde{\theta} \) monotonically. That \( \tilde{\theta} \) is consistent follows from the convergence of \( \overline{d} \) and continuity of \( a(\theta) \). Simple calculation also shows that for \( \theta > 0 \),

\[
\chi^2[\sqrt{n} (\tilde{\theta} - \theta)] \to N \left( 0, \frac{b^2(\theta)}{[a'(\theta)]^2} \right) \quad (n \to \infty).
\]

When \( \theta = 0 \), \( \tilde{\theta} \) is not asymptotically normal because by (2.5)

\[
\text{pr}(\tilde{\theta} = 0) = \text{pr}(\overline{d} < \sigma a(0)) \approx \frac{1}{2}.
\]

The asymptotic distribution of \( \tilde{\theta} \) at \( \theta = 0 \) can be derived by expanding the right-hand side of (2.5) in Taylor series to get

\[
\tilde{\theta}^2 = 2 \sqrt{n}(\overline{d}/\sigma - a(0)) + o_p(n^{-1/2}) \quad (\overline{d} > \sigma a(0)).
\]

Hence

\[
\text{pr}(\tilde{\theta} \leq t | \theta = 0) = \Phi\left( t^2 \sqrt{\frac{n}{2\pi - 16}} \right) \quad (t \geq 0)
\]

2.2 The sum of squares \( \sum d_i^2 \)

A second statistic for testing \( H_0 \) is

\[
F_n = \sum \frac{d_j^2}{(2\sigma^2)}
\]

whose distribution is that of a \( \chi^2_n \) variable with non-centrality parameter \( n\sigma^2/2 \). We shall see later that the maximum likelihood estimate of \( \theta \) is
zero if \( F_n \leq n \), which suggests \( F_n \) to be superior to \( D_n \) close to \( \theta = 0 \) for large \( n \). The exact rejection region for the size-\( \alpha \) \( F_n \) test is \( F_n \geq \chi^2_n(1-\alpha) \), and the power function is

\[
P(F_n, \alpha; \theta) = \phi \left\{ \frac{\frac{\theta^2}{\sqrt{n}} - 2\sqrt{2} c(\alpha)}{2 \sqrt{(2+2\theta^2)}} \right\}
\]

where \( c(\alpha) = \Phi^{-1}(1-\alpha) \). Some exact values of \( P(F_n, \alpha; \theta) \) are given by Fix (1949) and Patnaik (1949). It is easy to see from (2.4) and the second derivative of (2.8) that the asymptotic efficiency of \( D_n \) relative to \( F_n \) is \( (\pi-2)^{-1} = 0.93 \). However, for moderate values of \( \theta \) (2.3) and (2.8) indicate that \( D_n \) is more powerful than \( F_n \). Some values of \( P(D_n, \alpha; \theta) \) and \( P(F_n, \alpha; \theta) \) are given in Table 2.3, together with the corresponding values for the likelihood ratio test.

2.3 The likelihood

Both \( D_n \) and \( F_n \) are simpler to use than the likelihood ratio statistic which we now consider. Since \( \sigma^2 \) is known, we may assume \( \sigma^2 = 1 \). Then the likelihood of the observed \( (u_i, v_i) \) is

\[
L_n(\mu_1, \mu_2) = \prod_{j=1}^{n} \{ \phi(u_j-\mu_1) \phi(v_j-\mu_2) + \phi(u_j-\mu_2) \phi(v_j-\mu_1) \}.
\]

Re-parametrizing to \( \mu_1 + \mu_2 = 2\gamma \) and \( |\mu_1 - \mu_2| = \theta \), we see that the likelihood is maximized at \( \hat{\gamma} = (u+v)/2 = (x+y)/2 \) for any fixed \( \theta \). It is then convenient to work with the ratio

\[
R_n(\theta) = \frac{L_n(\hat{\gamma}, \theta)}{L_n(\hat{\gamma}, 0)} = \exp(-n\theta^2/4) \prod_{j=1}^{n} \cosh(d_j \theta/2).
\]
The maximum likelihood estimate $\hat{\theta}$ is thus a solution of

$$\sum d_j \tanh(\theta d_j / 2).$$

The solution $\theta = 0$ maximizes the likelihood if and only if $\sum d_j^2 \leq n$, otherwise $\hat{\theta}$ is the unique positive solution of (2.10). Note that $\hat{\theta} \leq \bar{d}$ since $\tanh(x) \leq 1$, also that if $\hat{\theta} = 0$ then $\tilde{\theta} = 0$, One might calculate $\hat{\theta}$ approximately by evaluating both sides of (2.10) for some values of $\theta$ less than $\bar{d}$. More accurate is the iterative solution

$$\hat{\theta}_{r+1} = \hat{\theta}_r - \frac{2n\hat{\theta}_r - 2 \sum d_j \tanh(d_j \hat{\theta}_r / 2)}{2n - \sum d_j^2 \text{sech}^2(d_j \hat{\theta}_r / 2)}, \quad \hat{\theta}_1 = \bar{d},$$

for which $\bar{d} > \hat{\theta}_2 > \ldots > \hat{\theta}_r > \hat{\theta}$ when $\hat{\theta} > 0$.

The likelihood ratio statistic for testing $H_0$ is

$$G_n = R_n(\hat{\theta}).$$

Large-sample theory for $G_n$ and $\hat{\theta}$ is complicated by the discrete component at the origin in the distribution of $\hat{\theta}$. If $\theta > 0$, then $p_r(\hat{\theta} = 0) \to 0$ as $n \to \infty$, and standard asymptotic theory implies

$$\chi^2 \{\sqrt{n}(\hat{\theta}-\theta)\} \to N(0, I(\theta)^{-1}) \quad (\theta > 0)$$

where

$$I(\theta) = \frac{1}{2} - \exp(-\theta^2/4) \int_0^\infty u^2 \phi(u) \text{sech}(\theta u / \sqrt{2}) du.$$

Table 2.2 gives some values of $I(\theta)$ and the corresponding variance reciprocal $\{a'(\theta)\}/b^2(\theta)$ for $\sqrt{n}(\hat{\theta}-\theta)$; the entries at $\theta = \infty$ are equal to the ordered-sample information. Apparently for large $n$ there is little to choose between $\hat{\theta}$ and $\tilde{\theta}$. 

6
Table 2.2. Reciprocals of limiting variances for $\sqrt{n}(\hat{\theta} - \theta)$ and $\sqrt{n}(\hat{\theta} - \theta)$

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0</th>
<th>0.25</th>
<th>0.50</th>
<th>1.00</th>
<th>1.50</th>
<th>2.00</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I(\theta)$</td>
<td>0</td>
<td>0.029</td>
<td>0.101</td>
<td>0.266</td>
<td>0.383</td>
<td>0.448</td>
<td>0.500</td>
</tr>
<tr>
<td>${a'(\theta)}^2/b^2(\theta)$</td>
<td>0</td>
<td>0.026</td>
<td>0.094</td>
<td>0.260</td>
<td>0.381</td>
<td>0.447</td>
<td>0.500</td>
</tr>
</tbody>
</table>

Complications occur when $\theta = 0$, since this is a boundary point of the parameter space. Properties of likelihood inference in such cases have been treated by Chernoff (1956) and others. Here the first derivative of the likelihood at $\theta = 0$ vanishes, so more terms are required in the expansion of the likelihood. Suppose $\hat{\theta} > 0$, and expand (2.10) in Taylor series to the $\hat{\theta}^3$ term. Then we get

$$n\hat{\theta} \approx \frac{1}{2} \hat{\theta} \sum d_j^2 - \frac{1}{24} \hat{\theta}^3 \sum d_j^4,$$

whose positive solution is

$$\hat{\theta} = \left(\frac{\frac{1}{2n} \sum d_j^2 - 1}{\frac{1}{24n} \sum d_j^4}\right)^{1/2} \approx \left(\frac{\frac{1}{n} \sum d_j^2 - 2}{\frac{1}{2n}\sum d_j^4}\right)^{1/2}.$$

Notice that $\hat{\theta} \sim \mathcal{O}(n^{-k})$. Now expand $\log R_n(\hat{\theta})$ in Taylor series (up to the $\hat{\theta}^4$ term) and substitute from (2.11). The result is

$$\log C_n = \log R_n(\hat{\theta}) \approx \frac{n}{16} \hat{\theta}^4 \approx \frac{n}{16} \left(\frac{1}{n} \sum d_j^2 - 2\right)^2 \quad (\hat{\theta} > 0, \theta = 0)$$

We can apply the central limit theorem to $\sum d_j^2$ to deduce from (2.11), (2.12) and the fact that $\sum d_j^2 < 2n$ implies $\hat{\theta} = 0$, that
\begin{align*}
(2.13) \quad \text{pr}(2 \log G_n \geq c | \theta = 0) & = \frac{1}{2} \text{pr}(\chi^2_1 \geq c) \quad , \quad (c > 0) \\
\text{pr}(\log G_n = 0 | \theta = 0) & = \text{pr}(\hat{\theta} = 0 | \theta = 0) \approx \frac{1}{2} \\
(2.14) \quad \text{pr}(\hat{\theta} < t | \theta = 0) & \approx \Phi \left( t \sqrt{\frac{n}{\theta}} \right) \quad (t > 0).
\end{align*}

Comparison of (2.7) and (2.14) shows \( \hat{\theta} \) to be slightly superior to \( \hat{\theta} \) when \( \theta = 0 \), as is the case when \( \theta > 0 \).

Notice from (2.12) that the rejection regions for \( F_n \) and \( G_n \) are asymptotically the same when \( \theta = 0 \). For large \( \theta \) one can also deduce from (2.10) that \( \hat{\theta} \approx \bar{d} \), and consequently for \( \theta > 0 \) both \( G_n \) and \( D_n \) have approximately the same large-sample power.

Thus far we have examined the large-sample behavior of the test statistics \( D_n, F_n \) and \( G_n \). We might conclude that \( D_n \) and \( F_n \) differ very little, while the more complicated \( G_n \) effectively combines \( D_n \) and \( F_n \). Are these conclusions valid for moderate values of \( n \) and \( \theta \)? We have calculated some approximate values of power for the three statistics, and these are given in Table 2.3. For \( F_n \) we have interpolated the exact power from Fix's (1949) table, and for \( D_n \) we have used the accurate normal approximation (2.3). For each \( n \) we generated 1000 samples on a digital computer and determined the empirical 0.95 quantiles for \( G_n, D_n \) and \( F_n \) when \( \theta = 0 \). Then with \( \theta > 0 \) in the same 1000 samples we calculated the empirical powers using the empirical critical values. These power values are entered in Table 2.3 with an asterisk.

Evidently there is little to choose between \( D_n, F_n \) and \( G_n \) on the basis of power. From a practical viewpoint \( G_n \) is the worst to compute, while \( F_n \) is the easiest and its exact distribution is known.
<table>
<thead>
<tr>
<th>n</th>
<th>θ</th>
<th>$D_n$</th>
<th>$F_n$</th>
<th>$G_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.25</td>
<td>.060</td>
<td>.060</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.059*</td>
<td>.062*</td>
<td>.063*</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>.095</td>
<td>.095</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.091*</td>
<td>.099*</td>
<td>.101*</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>.297</td>
<td>.270</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.273*</td>
<td>.293*</td>
<td>.292*</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>.647</td>
<td>.608</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.639*</td>
<td>.638*</td>
<td>.648*</td>
</tr>
<tr>
<td>20</td>
<td>0.25</td>
<td>.063</td>
<td>.064</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.064*</td>
<td>.065*</td>
<td>.069*</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>.114</td>
<td>.114</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.117*</td>
<td>.121*</td>
<td>.130*</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>.426</td>
<td>.404</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.388*</td>
<td>.414*</td>
<td>.419*</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>.851</td>
<td>.834</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.854*</td>
<td>.851*</td>
<td>.866*</td>
</tr>
<tr>
<td>40</td>
<td>0.25</td>
<td>.068</td>
<td>.068</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.061*</td>
<td>.070*</td>
<td>.071*</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>.144</td>
<td>.142</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.124*</td>
<td>.152*</td>
<td>.158*</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>.620</td>
<td>.615</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.615*</td>
<td>.631*</td>
<td>.644*</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>.976</td>
<td>.975</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.978*</td>
<td>.982*</td>
<td>.983*</td>
</tr>
</tbody>
</table>

(*) empirical
A large-sample interval estimate for $\theta$ can be computed from the likelihood in the usual way, using (2.13) and the classical result for $\theta > 0$,

$$\text{Pr}(2 \log R_n(\hat{\theta}) - 2 \log R_n(\theta) > c | \theta) \approx \text{Pr}(\chi^2_1 > c).$$

3. TWO-SAMPLE TESTS WITH UNKNOWN VARIANCE

If an independent variance estimate $\bar{\sigma}^2$ is available, independent of $\theta$ and $(d_1, \ldots, d_n)$, then the tests described in Section 2 are appropriate with $\bar{\sigma}$ replacing $\sigma$. Large-sample normal and chi-square distributions are replaced by $t$- and $F$-distributions in the usual way. Such a variance estimate is defined by the $u_i$ and $v_i$ only when no between-pairs effects are present.

3.1 Tests with no between-pairs effects

If the $(u_i, v_i)$ are identically distributed pairs (as assumed in Section 2), then a suitable independent variance estimate is

$$\bar{\sigma}^2 = \frac{\sum (u_j + v_j - \bar{u} - \bar{v})^2}{2(n-1)};$$

$2(n-1)\bar{\sigma}^2$ is distributed as $\sigma^2 \chi^2_{n-1}$ independently of $(d_1, \ldots, d_n)$, $\gamma$ and $\theta$.

Use of additional degrees of freedom from $(d_1, \ldots, d_n)$ must give an estimate whose distribution depends on $\theta$. A "natural" estimate to use in testing $H_0: \theta = 0$ would be

$$\bar{\sigma}^2 = \frac{\sum (u_j - \bar{u})^2 + \sum (v_j - \bar{v})^2}{2(n-1)(1-\pi^{-1})},$$
which is unbiased under $H_0$, but which is not independent of $\bar{d}$ or $\sum d_j^2$ or $\theta$. Consequently the properties of $D_n$ with $\bar{\sigma}$ replacing $\sigma$ are quite complicated for small $n$, when the additional degrees of freedom in $\bar{\sigma}^2$ are likely to be effective; use of $F_n$ with $\bar{\sigma}^2$ is inappropriate since they have $\sum d_j^2$ in common. Surprisingly enough the null distribution of the t-statistic $D_n(\bar{\sigma})$ agreed well with the $t_{2(n-1)}$ distribution in 1000 artificially generated samples. For example, for $n = 10, 20$ and $40$ the empirical probabilities beyond the 0.95 percentiles of $t_{2(n-1)}$ were 0.044, 0.048 and 0.047 respectively. To see if $D_n(\bar{\sigma})$ performs better than $D_n(\bar{\sigma})$ we computed some power values empirically from 1000 samples, and Table 3.1 gives the results.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>1.00</th>
<th>1.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{10}(\bar{\sigma})$</td>
<td>.051</td>
<td>.058</td>
<td>.075</td>
<td>.149</td>
<td>.298</td>
</tr>
<tr>
<td>$D_{10}(\bar{\sigma})$</td>
<td>.052</td>
<td>.053</td>
<td>.073</td>
<td>.134</td>
<td>.290</td>
</tr>
<tr>
<td>$D_{20}(\bar{\sigma})$</td>
<td>.052</td>
<td>.061</td>
<td>.082</td>
<td>.213</td>
<td>.546</td>
</tr>
<tr>
<td>$D_{20}(\bar{\sigma})$</td>
<td>.049</td>
<td>.060</td>
<td>.081</td>
<td>.210</td>
<td>.553</td>
</tr>
</tbody>
</table>

Apparently $D_n(\bar{\sigma})$ has no advantage over $D_n(\bar{\sigma})$ in terms of power for such cases.

11
3.2. Tests with between-pairs effects present

Now suppose that between-pairs effects are present, so that the
appropriate model for the (unobserved) \( x_j \) and \( y_j \) is

\[
\begin{pmatrix}
  x_j \\
  y_j
\end{pmatrix} = \begin{pmatrix}
  \gamma_{-1} + \theta \sigma \\
  \gamma_{+1} + \theta \sigma
\end{pmatrix} + \begin{pmatrix}
  a_j \\
  1
\end{pmatrix} + \begin{pmatrix}
  \epsilon_j \\
  \eta_j
\end{pmatrix} \quad (j=1, \ldots, n),
\]

where \( \epsilon_j \) and \( \eta_j \) are iid \( N(0, \sigma^2) \) and \( a_1, \ldots, a_n \) are fixed or random effects. Without concomitant information about the \( a_j \)'s, only \( d_1, \ldots, d_n \) are available for inference about \( \theta \) and \( \sigma^2 \). The naive \( t \)-statistic is

\[
D_n^* = \sqrt{n} \left\{ \frac{\overline{d}}{\hat{\sigma}(d)} - \frac{a(0)}{b(0)} \right\}
\]

where \( \hat{\sigma}^2(d) = \sum (d_j - \overline{d})^2 / (n-1) \). (Note that

\[
D_n^* = \sqrt{n} \left\{ \frac{\overline{d}}{\hat{\sigma}(d)} - \frac{E(d|\theta=0)}{SD(d|\theta=0)} \right\} \neq \sqrt{n} \left\{ \frac{\overline{d} - E(d|\theta=0)}{\hat{\sigma}(d)} \right\},
\]

since the latter involves the unknown \( \sigma \).)

That \( \hat{\sigma}^2(d) \) is the component added to \( \hat{\sigma}^2 \) to get \( \hat{\sigma}^2 \) in (3.2), together with our empirical study of \( D_n(\hat{\sigma}) \) and \( D_n(\hat{\sigma}) \), should warn of weak power for \( D_n^* \). Also the half-normal distribution of \( d_j \) under \( H_0 \) suggests that the null distribution of \( D_n^* \) will not be close to Student's \( t \)-distribution; the effects of asymmetry on the one-tailed \( t \)-test are discussed by Scheffe (1957, Chapter 10). This turns out to be the case, as the empirical (1000 samples) critical values in Table 3.2 show; for \( n > 50 \) the normal approximation is adequate.
The lack of power is illustrated in Table 3.3 by some values of corresponding empirical power given for \( n = 10, 20 \) and \( 40 \). These should be compared to the values in Table 3.1.

| Table 3.2. Empirical critical values for \( D^*_n \) from 1000 samples |
|-------------------|---|---|---|---|---|---|---|
| size \( n \)     | 5 | 10 | 15 | 20 | 30 | 40 | 50 |
| \( .100 \)        | 2.52 | 1.74 | 1.57 | 1.54 | 1.48 | 1.38 | 1.37 |
| \( .050 \)        | 3.56 | 2.31 | 2.14 | 2.05 | 1.94 | 1.85 | 1.67 |
| \( .025 \)        | 4.77 | 2.74 | 2.59 | 2.50 | 2.34 | 2.14 | 2.10 |
| \( .010 \)        | 7.66 | 3.57 | 3.22 | 3.13 | 2.82 | 2.69 | 2.65 |

| Table 3.3. Empirical power of \( D^*_n \) from 1000 samples (\( \alpha = .05 \)) |
|-------------------|---|---|---|---|---|---|
| \( n \) \( \theta \)  | 0.10 | 0.25 | 0.50 | 1.00 | 1.50 |
| 10                | \( .050 \) | \( .046 \) | \( .057 \) | \( .074 \) | \( .098 \) |
| 20                | \( .045 \) | \( .041 \) | \( .043 \) | \( .060 \) | \( .125 \) |
| 40                | \( .051 \) | \( .052 \) | \( .052 \) | \( .099 \) | \( .261 \) |

Estimation of \( \theta \) (and \( \sigma \)) by moments now involves solution of

\[
\bar{\sigma} = \sigma a(\theta) \\
\tilde{\sigma}(\sigma) = \sigma b(\theta)
\]  

(3.4)

similar to (2.5). We know from (2.1) that \( a(\theta) \) and \( b(\theta) \) are monotonically decreasing and increasing respectively, so it follows from (3.4) that the moment estimator \( \tilde{\theta} \) is given by
\[ \bar{d} = \tilde{\sigma}(d) \frac{a(\tilde{\theta})}{b(\tilde{\theta})} \quad (\bar{d} > \tilde{\sigma}(d) \frac{a(0)}{b(0)}) \]

\[ \tilde{\theta} = 0 \quad (\bar{d} \leq \tilde{\sigma}(d) \frac{a(0)}{b(0)}) . \]

Limiting distributions for \( \tilde{\theta} \) similar to (2.6) and (2.7) can be derived by tedious calculation, but we have not done this. The results for \( \text{var}(\tilde{\theta}) \) in Section 2.1 and Table 2.2 will be lower bounds in the present situation, and the strong dependence on \( \theta \) will remain.

The likelihood approach is a little more complicated with \( \sigma \) as a nuisance parameter, but the same type of results as in Section 2.3 obtain. We give an outline here. First let \( \alpha = \theta / \sqrt{2} \) and \( \beta = \sigma / \sqrt{2} \).

Then the log likelihood of \( d_1, \ldots, d_n \) may be written

\[ \ell_n(\alpha, \beta) = c - n \log \beta - \frac{1}{2} \beta^{-2} \Sigma d_j^2 - \frac{1}{2} n\alpha^2 + \Sigma \log \cosh(\alpha d_j / \beta) . \]

Setting first derivatives of (3.5) equal to zero gives the estimating equations

\[ n\alpha = \beta^{-1} \Sigma d_j \tanh(\alpha d_j / \beta) , \]

(3.6)

\[ n\beta^2 = \Sigma d_j^2 / (1 + \alpha^2) . \]

(3.7)

The solution \( \alpha = 0 \) to (3.6) is the maximum likelihood estimate if \( 3(\Sigma d_j^2)^2 \leq n \Sigma d_j^4 \), and from (3.6) and (3.7) we note that

\[ \hat{\alpha} \leq \sqrt{n} \bar{d} / \sqrt{\Sigma (d_j - \bar{d})^2} . \]

If the conditional solution \( \hat{\beta}(\alpha) \) to (3.7) is substituted into (3.6), an iterative solution for \( \hat{\alpha} \) is easily constructed, similar to that for \( \hat{\theta} \) in Section 2.3.
The log likelihood ratio statistic for testing $H_0$ may be written, in terms of the conditional estimate $\hat{\beta}(\alpha)$, as

$$\log G^*_n = \ell_n(\hat{\Theta}, \hat{\beta}) - \ell_n(0, \hat{\beta}(0))$$

where $\hat{\beta} = \hat{\beta}(\hat{\Theta})$. Expansion of $\log G^*_n$ under $H_0$ leads to the asymptotic distribution (2.13). An asymptotically equivalent test statistic exists, analogous to $F_n$ in Section 2.3, but now involves the first three sample moments of $d^2$.

Empirical comparisons of the power functions of $D^*_n$ and $G^*_n$ indicate very little advantage to $G^*_n$, similar to the comparison of $D_n$ and $G_n$ in Table 2.3.

When $\theta$ is non-zero, standard asymptotic theory applies to the likelihood function and $\hat{\Theta}$ in particular. We have not computed numerical values of the relevant information matrix $I(\Theta, \sigma)$, but it is quite clear that the variance of $\sqrt{n}(\hat{\Theta} - \Theta)$ with $\sigma$ known (Table 2.2) is a lower bound when $\sigma$ is unknown and the strong dependence on $\Theta$ remains. Thus the asymptotic normal distribution of $\hat{\Theta}$ is very difficult to use in constructing large-sample confidence intervals for $\Theta$. However, the likelihood may be used directly, making use of the large-sample results

$$\Pr[\ell_n(\hat{\Theta}, \hat{\beta}) - \ell_n(\alpha_0, \hat{\beta}(\alpha_0)) > c|\alpha = \alpha_0] \approx \Pr(\chi^2_1 < 2c) \quad (\alpha_0 > 0)$$

$$\Pr[\ell_n(\hat{\Theta}, \hat{\beta}) - \ell_n(0, \hat{\beta}(0)) > c|\alpha = 0] \approx \frac{1}{2} \Pr(\chi^2_1 < 2c),$$

where $\alpha = \theta/\sqrt{2}$. 

15
4. DISCUSSION

If the variance \( \sigma^2 \) can be estimated independently of the differences 
\( d = |x - y| \), the \( F_n \) statistic discussed in Section 2.2 should be used 
together with that variance estimate to test \( H_0: \theta = 0 \). The information 
about \( \sigma^2 \) in the \( d \)'s is apparently negligible compared to that in the 
\( x+y \)'s. There seems to be little to choose between \( \bar{\theta} \) and \( \hat{\theta} \) as point 
estimators, but their distributional properties depend heavily on \( \theta \). 
For that reason, interval estimates for \( \theta \) should be based on the likeli-
hood, taking account of the special case \( \theta = 0 \).

If at all possible one should avoid the situation where \( \sigma^2 \) must 
be estimated from the \( d \)'s, since the results in Section 3.2 show that 
tests of \( H_0 \) have little chance of detecting moderate values of \( \theta \) even 
in medium-sized samples.

A generalization of the problem we have considered is that where 
unordered \( k \)-tuples (\( k > 2 \)) are sampled, and equality of means is to 
be tested. Presumably calculations similar to those in the present paper 
would not be too difficult. It would be interesting to see how efficient 
are the ranges of the \( k \)-tuples.

A problem related to the unordered-pairs testing and estimation is 
that of population identification. Given that we decide \( \mu_1 \neq \mu_2 \), we 
might then want to decide which individuals in the sample came from 
each of the two populations with a view to examining other characteristics 
of the populations. A pragmatic compromise would be to decide that all 
\( v_i \) measurements with rank \( d_i > m \) belong to one population, while
the remainder are left undecided. Here \( m/n \) would depend on \( \hat{\theta} \). Of course if one wished to test for difference in a characteristic \( Z \) between the populations, emphasis on the pair \( Z_{11}, Z_{12} \) for the \( i \)-th individual should increase with rank \( (d_i) \).

I am grateful to Louise Knight, Stanford Medical School, for bringing this problem to my attention.
REFERENCES

