Chapter 6: Distribution free methods

χ² goodness of fit test

Completely specified hypothesis

data:

<table>
<thead>
<tr>
<th>class</th>
<th>C₁</th>
<th>C₂</th>
<th>...</th>
<th>Cₖ</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>f₁</td>
<td>f₂</td>
<td>...</td>
<td>fₖ</td>
</tr>
</tbody>
</table>

null hypothesis, \( H_0 \): \( \text{Pr}(C_j) = p_j, \quad j = 1, 2, \ldots, k \)

<table>
<thead>
<tr>
<th>class</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>43</td>
<td>39</td>
<td>18</td>
</tr>
</tbody>
</table>

\( H_0 \): \( \text{Pr}(A) = 0.5, \text{Pr}(B) = 0.3, \text{Pr}(C) = 0.2 \).

Are the data and the hypothesis compatible?

What do we “expect” to get if \( H_0 \) is true?

<table>
<thead>
<tr>
<th>class</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>50</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

If \( H_0 \) is true then \( F_i \overset{d}{=} \text{Bi}(n, p_i) \), so that \( E(F_i) = np_i \) — these are called the “expected frequencies”. But note that they are actually expectations and so need not be integers.

Our test is based on the deviation of the observed frequencies from the expected frequencies:

Test statistic:

\[
U^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} = \sum \frac{(o - e)^2}{e}
\]

If \( H_0 \) is true then \( U^2 \overset{d}{=} \chi^2_{k-p-1} \),

where \( k \) is the number of classes
and \( p \) the number of parameters estimated;

If \( H_0 \) is not true then \( U^2 \) is inclined to be large, so:

reject \( H_0 \) if \( U^2 > c_{1-\alpha}(\chi^2_{k-p-1}) \).

Note:

1. In the case of a completely specified hypothesis, \( p = 0 \).

2. In using the \( \chi^2 \) distribution we are approximating binomial by normal. This is reasonable provided \( np_i > 5 \).

3. The larger the value of \( k \) (number of classes), the more powerful the test; but we must have \( np_i > 5 \) which limits the number of classes.

4. \( U^2 \) too small \( \leftrightarrow \) fit “too good” — rigging

example (die rolling) Test \( H_0 \): the die is fair.

<table>
<thead>
<tr>
<th>( C_i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_i )</td>
<td>6</td>
<td>10</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>( e_i )</td>
<td>( \frac{8}{3} )</td>
<td>( \frac{8}{3} )</td>
<td>( \frac{8}{3} )</td>
<td>( \frac{8}{3} )</td>
<td>( \frac{8}{3} )</td>
<td>( \frac{8}{3} )</td>
</tr>
</tbody>
</table>

\[
u^2 = \left( \frac{2}{\frac{8}{3}} \right)^2 + \left( \frac{1}{\frac{8}{3}} \right)^2 + \cdots + \left( \frac{1}{\frac{8}{3}} \right)^2 = 1.92
\]

page 6.1
Since \( c_{0.95}(\chi^2) = 11.07 \), we accept that the die is fair.

example

<table>
<thead>
<tr>
<th>class</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>obs freq</td>
<td>36</td>
<td>49</td>
<td>15</td>
</tr>
</tbody>
</table>

Test the hypothesis that these occur in the ratio 1:2:1.

\( H_0: \Pr(AA) = 0.25, \Pr(Aa) = 0.50, \Pr(aa) = 0.25 \)

<table>
<thead>
<tr>
<th>class</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>exp freq</td>
<td>25</td>
<td>50</td>
<td>25</td>
</tr>
</tbody>
</table>

\[
\chi^2 = \frac{(36 - 25)^2}{25} + \frac{(49 - 50)^2}{50} + \frac{(15 - 25)^2}{25} = 8.86.
\]

\( P = \Pr(U^2 > 8.86 | H_0) = \Pr(\chi^2 > 8.86) \approx 0.02 \)

Hence we reject \( H_0 \).

An alternative approach is to use the asymptotic likelihood ratio test. Here we have

\[
L = \theta_1 f_1 \theta_2 f_2 \cdots \theta_k f_k
\]

where \( \theta_j = \Pr(X \in C_j) \).

\[
\ln L = \sum f_j \ln \theta_j.
\]

It follows that

\[
\begin{align*}
\ln L_1 &= \ln L(\hat{\theta}) = \sum f_j \ln \frac{\hat{f}_j}{n} \quad (\hat{\theta}_j = \frac{\hat{f}_j}{n}); \\
\ln L_0 &= \ln L(\theta_0) = \sum f_j \ln p_j \quad (H_0 \Rightarrow \theta_j = p_j).
\end{align*}
\]

Therefore

\[
\ln L_1 - \ln L_0 = \sum f_j \ln \frac{f_j}{np_j}.
\]

And, if \( H_0 \) is true, we have \( \ln L_1 - \ln L_0 \approx \frac{1}{2} \chi^2_{k-1} \) for large \( n \).

So we define

\[
G^2 = \sum_{i=1}^k 2f_i \ln \frac{f_i}{e_i} \approx \chi^2_{k-p-1} \quad \text{under the null hypothesis.}
\]

\( G^2 \) has the advantage that the effect of small \( e_i \) is less serious.

\( U^2 \) and \( G^2 \) are not equal. They just happen to have the same distribution under the null hypothesis. However, in most cases \( U^2 \) and \( G^2 \) will be comparable: \( U^2 \) and \( G^2 \) are asymptotically equivalent as \( n \to \infty \) (see Notes 202=p105).

For the above example,

\[
g^2 = 2 \left( 36 \ln \frac{36}{25} + 49 \ln \frac{49}{50} + 15 \ln \frac{15}{25} \right) = 8.95
\]

example  Test the goodness of fit of the Poisson distribution to the following data:

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>freq(( x ))</td>
<td>15</td>
<td>19</td>
<td>13</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

\( n = 60; \ e_i = np_i \), where \( p_i = \frac{e^{-\lambda} \lambda^i}{i!}; \ \hat{\lambda} = \bar{x} = 1.5 \)

<table>
<thead>
<tr>
<th>class</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>( \geq 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>obs freq</td>
<td>15</td>
<td>19</td>
<td>13</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>exp freq</td>
<td>13.39</td>
<td>20.08</td>
<td>15.06</td>
<td>7.53</td>
<td>2.83</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Since \( e < 5 \), we must combine classes; it is permitted to have one in five is between 1 and 5.

<table>
<thead>
<tr>
<th>class</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>( \geq 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>obs freq</td>
<td>15</td>
<td>19</td>
<td>13</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>exp freq</td>
<td>13.39</td>
<td>20.08</td>
<td>15.06</td>
<td>7.53</td>
<td>3.94</td>
</tr>
</tbody>
</table>
Then we obtain:
\[ u^2 = \frac{(15 - 13.39)^2}{13.39} + \cdots + \frac{(5 - 3.94)^2}{3.94} = 0.85 \]

\[ df = 5 - 1 - 1 = 3; \text{ so reject } H_0 \text{ if } u^2 > 7.815. \]
Thus the Poisson distribution is a good fit (in fact an excellent fit) to the data.

Note the alternative test of “Poissonicity” based on the result that if the Poisson model is correct, then \( S^2/\bar{X} \approx \chi^2_n - 1. \)
Here \( s^2/\bar{x} = 1.282^2/1.5 = 1.096; \text{ and } c_{0.025}(\chi^2_{59})/59 = 0.672, \]
\( c_{0.975}(\chi^2_{59}) = 1.392; \text{ so this also indicates acceptance of the Poisson hypothesis.} \)

**Example:** (2×2 contingency table)

**Table:**

<table>
<thead>
<tr>
<th>obs.freq</th>
<th>B</th>
<th>B-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>A-</td>
<td>14</td>
<td>24</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>36</td>
<td>70</td>
</tr>
</tbody>
</table>

\( H_0: A \text{ and } B \text{ are indept} \rightarrow \text{Pr}(A \cap B) = \text{Pr}(A) \text{Pr}(B) \)

<table>
<thead>
<tr>
<th>exp.freq</th>
<th>B</th>
<th>B-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>( n\alpha\beta )</td>
<td>( n\alpha(1 - \beta) )</td>
<td>( n\alpha )</td>
</tr>
<tr>
<td>A-</td>
<td>( n(1 - \alpha)\beta )</td>
<td>( n(1 - \alpha)(1 - \beta) )</td>
<td>( n(1 - \alpha) )</td>
</tr>
</tbody>
</table>

in which \( \alpha = \text{Pr}(A) \) and \( \beta = \text{Pr}(B) \) are unknown, so we must estimate them:
\[ \hat{\alpha} = \frac{32}{70}, \quad \hat{\beta} = \frac{34}{70}; \]
so that:
\[ \text{exp.freq}(A \cap B) = n\hat{\alpha}\hat{\beta} = 70 \times \frac{32}{70} \times \frac{34}{70} = \frac{32 \times 34}{70} \]

**Rule:**
\[ \text{exp.freq} = \frac{\text{row total} \times \text{column total}}{\text{grand total}} \]

<table>
<thead>
<tr>
<th>exp.freq</th>
<th>B</th>
<th>B-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15.54</td>
<td>16.46</td>
<td>32</td>
</tr>
<tr>
<td>A-</td>
<td>18.46</td>
<td>19.54</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>36</td>
<td>70</td>
</tr>
</tbody>
</table>

\[ u^2 = \frac{4.46^2}{15.54} + \frac{4.46^2}{16.46} + \frac{4.46^2}{18.46} + \frac{4.46^2}{19.54} = 4.58 \]

And under \( H_0, U^2 \rightarrow \chi^2_1, \text{ since } df = k - p - 1 = 4 - 2 - 1 = 1. \)
But there is a catch! The statistic \( U^2 \) is actually discrete valued; and so we must make a correction for continuity: we define \( o_e \) to be 0.5 closer to \( e \) than \( o \) is; and then calculate \( u^2_e \) using \( o_e \) instead of \( o \). Or equivalently, reduce each \(|o - e|\) by 0.5. This gives:
\[ u^2_e = \frac{3.96^2}{15.54} + \frac{3.96^2}{16.46} + \frac{3.96^2}{18.46} + \frac{3.96^2}{19.54} = 3.61. \]

Then:
\[ P = \Pr(U^2 \geq u^2) \approx P_e = \Pr(\chi^2_1 > u^2_e) = \Pr(\chi^2_1 > 3.61) = 0.057; \]
and we accept \( H_0 \) since \( P > 0.05. \)
In this case the exact $P$-value can be calculated (see later) and it is found that $P = 0.057$, so the (continuity-corrected) $\chi^2$ approximation is accurate to three decimal places.

If no correction for continuity is used then we would get

$$P \approx P_u = \Pr(\chi^2 > u^2) = \Pr(\chi^2 > 4.58) = 0.032,$$

which is a pretty bad approximation — and also gives the wrong decision!

Using the critical value approach, we reject $H_0$ if $u^2 > c_{0.95}(\chi^2)$;
and since $c_{0.95}(\chi^2) = 3.841$, we accept $H_0$ . . . and conclude that there is no significant association between $A$ and $B$.

If we are using $G^2$, then we need to make a similar correction for continuity: again we define $o_c$ to be 0.5 closer to $e$ than $o$ is, and calculate $g^2_c$ using $o_c$ instead of $o$. Here

$$g^2_c = 2 \times 19.5 \ln \frac{19.5}{15.54} + 2 \times 12.5 \ln \frac{12.5}{16.36}$$
$$+ 2 \times 18.5 \ln \frac{18.5}{18.46} + 2 \times 23.5 \ln \frac{23.5}{19.54} = 3.64.$$

which gives $P_c(G^2) = 0.056$; as compared with $P_u(G^2) = 0.031$.

Note:

1. to test against one-sided alternative, we use $U = \pm \sqrt{U^2}$
   (where the sign depends on the direction of the deviation from expectation), so that $U \approx N(0, 1)$ if $H_0$ is true.

   For the example above, to test $H_0$: $A$ & $B$ independent vs $H_1$: $A$ & $B$ positively related, $u_c = +\sqrt{3.61} = 1.90$. So we would reject $H_0$ as $u_c > 1.6449$; $P \approx \Pr(N > 1.90) = 0.029$.

2. testing other null hypotheses — use the same procedure, after finding expected frequencies: for example to test the null hypothesis $\Pr(A) = \Pr(B)$ [McNemar’s test]:

$$\begin{array}{ccc}
\text{obs.freq} & 10 & 40 & 50 \\
20 & 30 & 50 \\
30 & 70 & 100 \\
\end{array}
\quad
\begin{array}{ccc}
\text{exp.freq} & 10 & 30 & 40 \\
30 & 30 & 60 \\
40 & 60 & 100 \\
\end{array}$$

$u^2 = \frac{0^2}{10} + \frac{10^2}{40} + \frac{10^2}{20} + \frac{0^2}{50} = 5$; $u_c^2 = \frac{9.5^2}{40} + \frac{9.5^2}{20} = 4.76$.

Since $u_c^2 > 3.84$, we would reject $H_0$ against a two-sided alternative.

2×2 contingency table — small sample (exact) test

$$\begin{array}{c|cc}
& B & B \\
\hline
A & X & n \\
\bar{A} & \bar{X} & \bar{n} \\
\hline
R & N \\
\end{array}$$

If $H_0$ is true then $X \overset{d}{=} Hg(n, R, N)$.

This follows since, if $A$ and $B$ are independent, choosing which of the $B$-elements are also in $A$ is like choosing which of the $(R)$ marked items in a population (of $N$) should be in a random sample (of $n$).
example

\[
\begin{array}{ccc}
A & B & \text{Total} \\
6 & 7 & 13 \\
14 & X = 2 & R = 16 \\
20 & n = 9 & N = 29
\end{array}
\]

\[H_0 \Rightarrow X \sim \text{Hg}(9, 16, 29)\]

\[
\begin{array}{ccccccc}
x & 0 & 1 & 2 & 3 & 4 & 5 & \ldots \\
p(x) & 0.000 & 0.002 & 0.021 & 0.096 & 0.234 & 0.312 & \ldots
\end{array}
\]

\[P = 2 \Pr(X \leq 2) = 2 \times 0.023 = 0.046,\]

so we reject \(H_0\).

This is the basis of the calculation of the exact \(P\)-value mentioned above. The approximation involved is to assume that the Hypergeometric distribution is approximately Normal; and approximating a discrete distribution by a continuous distribution requires a correction for continuity.

\(r \times c\) contingency table \quad (p.111)

For testing independence, the procedure is exactly the same as for the \(2 \times 2\) case (large sample test):

\[
\hat{e}_{ij} = n \hat{\alpha}_i \hat{\beta}_j = \frac{f_i \cdot f_j}{f_{\cdot\cdot}}
\]

\[u^2 = \sum \frac{(o - e)^2}{e} \approx \chi^2_\nu\] if \(H_0\) true (provided \(e_{ij} \gtrsim 5\)).

\[\nu = rc - [(r - 1) + (c - 1)] - 1 = (r - 1)(c - 1)\]

example comparing cough mixtures

\begin{tabular}{ccc}
A & B & C \\
“little or no relief” & 11 & 13 & 9 \\
“moderate relief” & 32 & 28 & 27 \\
“total relief” & 7 & 9 & 14 \\
\end{tabular}

In Minitab the procedure for an \(r \times c\) contingency table is no different. With the tabled data in columns c1, c2 and c3, the command chis c1-c3 produces the following output.

Chi-Square Test

Expected counts are printed below observed counts

\[
\begin{array}{ccc}
A & B & C \\
1 & 11 & 13 & 9 & 33 \\
2 & 29.00 & 29.00 & 29.00 & 87 \\
3 & 7 & 9 & 14 & 30 \\
Total & 50 & 50 & 50 & 150
\end{array}
\]

\[\text{ChiSq} = 0.000 + 0.364 + 0.364 + 0.310 + 0.034 + 0.138 + 0.900 + 0.100 + 1.600 = 3.810\]

\[df = 4, p = 0.433\]

Thus, Minitab \(\Rightarrow U = 3.81\) \((df = 4)\)

\[P \approx 0.43\] and so we accept \(H_0\): there is no significant difference in the perceived effects of the cough mixtures.

Here \(g^2 = \sum \sum 2o \ln \frac{o}{e} = 3.74\); and, as before, \(H_0 \Rightarrow G^2 \approx \chi^2_1\).

Note: for tables bigger than \(2 \times 2\) a correction for continuity is not appropriate: we don’t have a direction to make it in.
Inference on the median

Inference on $m$ is based on $Z = \text{freq}(X \leq m)$ (i.e. the number of observations in the sample less than or equal to $m$)

We know that $\text{freq}(A) \overset{d}{=} \text{Bi}(n, \Pr(A))$.

So the distribution of $Z$ is known and simple:

$$Z = \text{freq}(X \leq m) \overset{d}{=} \text{Bi}(n, \frac{1}{2})$$

and if $n$ is large, then $Z \overset{d}{=} N(\frac{1}{2}n, \frac{1}{4}n)$.

example Use the following sample to test the hypothesis that $m = 10$ against a two-sided alternative (i.e. $m \neq 10$)

$$4.6 \ 7.3 \ 4.7 \ 6.1 \ 11.2 \ 9.8 \ 2.4 \ 6.7$$
$$23.1 \ 10.2 \ 8.7 \ 6.2 \ 7.3 \ 2.9 \ 9.0$$

$n = 15 \quad z = \text{freq}(X \leq 10) = 12 \quad H_0 \Rightarrow Z \overset{d}{=} \text{Bi}(15, \frac{1}{2})$

Thus, $P = 2 \Pr(Z \geq 12) = 2 \times 0.0176 = 0.0352$

Hence we reject $H_0$: $m = 10$.

The sample suggests that $m < 10$ ($\hat{m} = 7.3$)

example new type of light globe: is $m > 1000$?

$n = 400 \quad z = \text{freq}(T \leq 1000) = 175$

$H_0: m = 1000 \quad \text{vs} \quad H_1: m > 1000$.

$H_0 \Rightarrow Z \overset{d}{=} N(200, 100)$

\[\therefore \quad P = \Pr(Z \leq 175) \approx \Pr(Z^* < 175.5)\]
\[\approx \Pr(Z^*_* < -2.45) \approx 0.0071\]

Thus we reject $H_0$ and conclude that the median is significantly greater than 1000 hours.

Confidence interval for $m$ based on $Z = \text{freq}(Z \leq m)$

[A confidence interval for $\theta$ based on $T$ is obtained by “inverting” a probability statement in $T$]

$$a \leq \text{freq}(X \leq m) \leq b \quad \text{Pr}(a \leq \text{freq}(X \leq m) \leq b) = p$$

$$\Downarrow$$

$$X_{(a)} < m < X_{(b+1)} \quad \text{Pr}(X_{(a)} < m < X_{(b+1)}) = p$$

If there are at least $a$ observations less than $m$, then $X_{(a)}$ must be less than $m$; i.e. $Z \geq a \Rightarrow X_{(a)} < m$; and

if $X_{(a)}$ is less than $m$, then there must be at least $a$ observations less than $m$; i.e. $X_{(a)} < m \Rightarrow Z \geq a$.

The argument at the other end is similar, remembering that the complement of $(Z \leq b)$ is $(Z \geq b + 1)$.
example \( n = 10 \) \( Z \overset{d}{=} \text{Bi}(10, \frac{1}{2}) \)

Binomial tables \( \Rightarrow \Pr(2 \leq Z \leq 8) = 0.9784; \)
thus a 97.84% CI for \( m \) is \( x(2) < m < x(9) \).

\[
2 \leq \text{freq}(X \leq m) \leq 8
\]
\[
\uparrow
\]
\[
X(2) < m < X(9)
\]

example \( n = 100 \) \( Z \overset{d}{=} \text{N}(50, 25) \)

\[
a \approx 50 - 1.96 \times 5 = 40.2 \rightarrow a = 40
\]
\[
b \approx 50 + 1.96 \times 5 = 59.8 \rightarrow b = 60
\]

\[
\Pr(40 \leq Z \leq 60) \approx \Pr(-2.10 < Z^* < 2.10) = 0.9642
\]
\[
\Pr(41 \leq Z \leq 59) \approx \Pr(-1.90 < Z^* < 1.91) = 0.9426
\]

Thus a 96.42% CI for \( m \) is \((x(40), x(61))\).

MTB > random 100 c1;
SUBC> normal 40 10.
MTB > sint c1

Sign confidence interval for median

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{ACHIEVED} & \text{MEDIAN} & \text{CONF IDENTITY} & \text{CONF INTERVAL} & \text{POSN} \\
\hline
C1 & 100 & 37.64 & 0.9426 & (36.58, 40.44) & 41 \\
 & & 0.9500 & (36.58, 40.45) & \text{NLI} \\
 & & 0.9643 & (36.58, 40.48) & 40 \\
\hline
\end{array}
\]

MTB > stest 40 c1
Sign test of median = 40.00 versus N.E. 40.00

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{N} & \text{BELOW} & \text{EQUAL} & \text{ABOVE} & \text{P-VALUE} & \text{MEDIAN} \\
\hline
C1 & 100 & 58 & 0 & 42 & 0.1336 & 37.64 \\
\hline
\end{array}
\]

Note: We have \( \hat{m} \overset{d}{=} \text{N}(m, \frac{1}{4n f(m)}) \), so why not use this?

The problem is that this result is not much use without a reasonable estimate of \( f(m) \), which is difficult to obtain without making assumptions about \( f \) (unless the sample is very large).

**Test of \( m = m_0 \) assuming symmetry**

[mainly useful for considering differences \( X - Y \), where \( X \overset{d}{=} Y \): then the distribution of \( X - Y \) is symmetric about zero, since \( Y - X \overset{d}{=} X - Y \).]

The procedure is best explained by example:

example Test \( H_0: m = 4 \) vs \( H_1: m \neq 4 \)

for the following sample:

\[
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
x & 1.7 & 2.9 & 3.4 & 4.2 & 5.7 & 5.9 & 6.4 & 7.2 \\
\hline
|x - 4| & 2.3 & 1.1 & 0.6 & 0.2 & 1.7 & 1.8 & 1.9 & 2.4 & 3.2 \\
\hline
\text{rank}(|x - 4|) & 7 & 3 & 2 & 1 & 4 & 5 & 6 & 8 & 9 \\
\hline
w_- = 12 & w_+ = 33 \\
\hline
\end{array}
\]

\([\text{Note: } w_- + w_+ = 1 + 2 + \cdots + 9 = 45]\)

Test: reject \( H_0 \) unless \( a \leq W_- \leq b \)  \( \text{(or } a \leq W_+ \leq b) \)

Thus we need to know the distribution of \( W_- \) (or \( W_+ \)) under \( H_0 \).

If \( H_0 \) is true then \( W_- \overset{d}{=} W_+ \) by symmetry — and the common distribution can be derived by enumeration of the \( 2^n \) possibilities: each rank \( (1, 2, \ldots, n) \) may be above or below \( m_0 \).
example

\[ \begin{array}{c|ccc|}
   & w_- & w_+ \\
\hline
123 & 6 & 0 \\
23 & 1 & 5 & 1 \\
13 & 2 & 4 & 2 \\
12 & 3 & 3 & 3 \\
3 & 12 & 3 & 3 \\
2 & 13 & 2 & 4 \\
1 & 23 & 1 & 5 \\
123 & 0 & 6 \\
\end{array} \]

For \( n \) anything but small, this would be tedious. However, it is something easily handled by a computer.

The tables (Table 13: 202=p251) give critical values.

\[ n = 9: \quad Pr(W \leq 5) \leq 0.025, \quad Pr(W \geq 40) \leq 0.025 \]

So, for a test of size \( \alpha \leq 0.05 \): reject \( H_0 \) unless \( 6 \leq W \leq 39 \).

Thus in the above (where \( w_- = 12, \ w_+ = 33 \) we accept \( H_0 \).

Minitab simplifies matters even further:

```
MTB > print c1
1.7 2.9 3.4 4.2 5.7 5.8 5.9 6.4 7.2
MTB > wtest 4 c1
Wilcoxon Signed Rank Test
TEST OF MEDIAN = 4.000 VERSUS MEDIAN N.E. 4.000
N FOR WILCOXON  ESTIMATED
N  TEST  STATISTIC  P-VALUE  MEDIAN
C1  9  9  33.0  0.236  4.900
```

```
MTB > ttest 4 c1
Test of mu = 4.000 vs mu not = 4.000
N  Mean  StDev  SE Mean  T  P-Value
C1  9  4.80  1.833  0.611  1.31  0.23
```

```
MTB > wint c1
Wilcoxon Signed Rank Confidence Interval
ESTD ACHIEVED
N  MEDIAN  CONFIDENCE  CONF INTERVAL
C1  9  4.90  95.6  ( 3.15, 6.40)
```

The confidence interval can be thought of as the set of values \( m_0 \) for which the null hypothesis \( H_0: m = m_0 \) is accepted (against a two-sided alternative).

**Rank-based tests for comparison of populations**

We wish to test

\( H_0: X_1, X_2, \ldots, X_k \) have the same distribution, \ i.e. \( X_i \overset{d}{=} X_0 \)

against

\( H_1: \) the distributions differ in location, \ i.e. \( X_i \overset{d}{=} X_0 + a_i \)

(These assumptions are analogous to the equal variance assumptions made for the \( t \)-test and analysis of variance.)

**Procedure:** rank all observations from smallest to largest and replace the observations by their ranks; then carry out tests analogous to \( t \)-tests and analysis of variance — these tests are simplified since it is known that the ranks are \( 1, 2, \ldots, N \) (in some order), so that \( \sigma_R^2 \) is effectively known (\( \sigma_R^2 = \frac{1}{12}N(N+1) \)) and doesn’t need to be estimated. Consequently the tests are the counterparts of \( z \)-tests and \( \chi^2 \)-tests (rather than \( t \)-tests and \( F \)-tests).
Comparison of two populations

Independent samples

Note: a simple test for equality of medians using a 2×2 contingency table is given in the Notes 202=pp114. It is not very powerful. However, it is simple and is readily extended to a k×2 table to compare k populations.

Suppose we wish to test $H_0$: $X_1 \distr X_2$ against $H_1$: $X_1 \distr X_2 + a$.
We assume that $n_1 \leq n_2$.

Test statistic: \[
T = \frac{\bar{R}_1 - \bar{R}_2}{\sigma_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{W_1 - \frac{1}{2} n_1(n_1 + n_2 + 1)}{\frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)}
\]

Thus, an equivalent statistic is $W_1 = \text{sum of the ranks in the } X_1 \text{ sample}$. This is the Wilcoxon rank sum statistic.

To specify the test, we need the distribution of $W_1$ under $H_0$.
If $H_0$ is true then all \( \binom{N}{n_1} \) possible combinations of ranks are equally likely. Thus all we need to do is to list them and count the number for which $W_1 = k$. This will specify the distribution of $W_1$.

Fortunately, we can persuade a computer to do this and the results (in the form of critical values) are given in the tables (Table 12: 202=p250).

For example: If $n_1 = 8$, $n_2 = 9$, the table give:

<table>
<thead>
<tr>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.005</td>
</tr>
<tr>
<td>0.01</td>
<td>0.025</td>
</tr>
<tr>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>0.9</td>
<td>0.95</td>
</tr>
<tr>
<td>0.975</td>
<td>0.99</td>
</tr>
<tr>
<td>0.99</td>
<td>0.995</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>51</td>
<td>54</td>
</tr>
<tr>
<td>58</td>
<td>86</td>
</tr>
<tr>
<td>90</td>
<td>93</td>
</tr>
<tr>
<td>97</td>
<td>99</td>
</tr>
</tbody>
</table>

thus for a two-sided test of size $\leq 0.05$, we would reject $H_0$ if $W_1 \leq 51$ or $W_1 \geq 93$, i.e. unless $52 \leq W_1 \leq 92$.

The one-sided tests of size ($\leq 0.05$) are to reject $H_0$ if $W_1 \leq 54$, and reject $H_0$ if $W_1 \geq 90$ respectively.

Example: Random samples on $X_1$ and $X_2$ yield the following results:

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>37</th>
<th>27</th>
<th>38</th>
<th>41</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_2$</td>
<td>39</td>
<td>29</td>
<td>51</td>
<td>43</td>
<td>52</td>
</tr>
</tbody>
</table>

To test $H_0$: $X_1 \distr X_2$ vs $H_1$: $X_1 \distr X_2 + a$ (where $a > 0$), using a test of size $\leq 0.05$,
we would reject $H_0$ if $W_1 \leq 23$ [tables: $m = 5$, $n = 8$, $p = 0.05$].

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>5</th>
<th>2</th>
<th>6</th>
<th>8</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_2$</td>
<td>7</td>
<td>3</td>
<td>12</td>
<td>9</td>
<td>13</td>
</tr>
</tbody>
</table>

Observations give $w_1 = 22$, so we reject $H_0$. $[0.025 < P < 0.05]$.

Note: A two-sided test of size $\leq 0.05$ would be to reject $H_0$ if $W_1 \leq 21$ or $W_1 \geq 48$ (i.e. unless $22 \leq W_1 \leq 47$).

Large sample approximation (for large $n_1$, $n_2$):

\[
T = \frac{\bar{R}_1 - \bar{R}_2}{\sigma_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \approx N(0, 1)
\]
example (as above — $n_1 = 5$, $n_2 = 8$)

\[ P = \Pr(W_1 \leq 22) \approx \Pr(W_1^* < 22.5) = \Pr(W_1^* < -1.830) \]

\[ \therefore P \approx 0.034 \] (which agrees with the above result)

This approximation is what MINITAB uses for all sample sizes!

MTB > mann c1 c2;
SUBC> alte -1.
C1 N = 5 Median = 37.00
C2 N = 8 Median = 45.50
Point estimate for ETA1-ETA2 is -10.50
95.2 Percent C.I. for ETA1-ETA2 is (-22.00, 2.00)
W = 22.0
Test of ETA1 = ETA2 vs. ETA1 < ETA2: P = 0.0336

Compare this with the result of the $t$-test:

MTB > twos c1 c2;
SUBC> alte -1.

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>5</td>
<td>33.80</td>
<td>6.83</td>
</tr>
<tr>
<td>C2</td>
<td>8</td>
<td>43.13</td>
<td>8.44</td>
</tr>
</tbody>
</table>

95% C.I. for $\mu$ C1 - $\mu$ C2: (-18.8, 0.2)

T-Test $\mu$ C1 = $\mu$ C2 (vs <): $T$=-2.18 $P$=0.027 DF=10

Paired samples

As for normally distributed data, we consider the sample of differences:

\[ D_i = X_{1i} - X_{2i} \quad (i = 1, 2, \ldots, n) \]

Under $H_0$: $X_1 \overset{d}{=} X_2$, the distribution of $D$ is symmetric about zero.

Thus we simply apply the test for the median of a symmetric population to the sample of differences (i.e. the signed rank sum test).

example (text example 202=p117)

MTB > let c3=c1-c2
Row  x1  x2  d
1  6.3  5.3  1.0
2  6.5  6.7 -0.2
3  6.9  6.3  0.6
4  7.2  7.6 -0.4
5  6.8  6.5  0.3
6  9.0  8.2  0.8
7  8.4  8.3  0.1
8  7.7  7.4  0.3

MTB > utest d
TEST OF MEDIAN = 0 VERSUS MEDIAN N.E. 0

<table>
<thead>
<tr>
<th>N</th>
<th>TEST STATISTIC</th>
<th>P-VALUE</th>
<th>MEDIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>29.0</td>
<td>0.141</td>
<td>0.3000</td>
</tr>
</tbody>
</table>

MTB > ttest d
Test of $\mu$ = 0.000 vs $\mu$ not = 0.000

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.312</td>
<td>0.479</td>
<td>0.169</td>
<td>1.84</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Comparison of $k$ populations

Independent samples

Replace observations by ranks and compute sums of squares (cf. one-way anova).

between groups SS:

$$B_R = \sum_{i=1}^{k} \frac{R_i^2}{n_i} - \frac{R_{••}^2}{N}.$$  

If $H_0$ is true then $B_R \approx \sigma_R^2 \chi^2_{k-1}$.

Define

$$H = \frac{B_R \sigma_R^2}{\sigma_R^2} = \frac{12}{N(N+1)} \left\{ \sum_{i=1}^{k} \frac{R_i^2}{n_i} - \frac{[\frac{1}{2}N(N+1)]^2}{N} \right\} = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

$H$ is the Kruskal-Wallis statistic

If $H_0$ is true, then $H \approx \chi^2_{k-1}$;

and if $H_0$ is not true, then $H$ tends to be large.

Thus, we reject $H_0$ if $H > c$.

Note: since $\sigma_R^2$ is effectively known, there is no need to estimate it (by $s_R^2 =$ within groups MS) and use an $F$-test.

The $\chi^2$ approximation is reasonable, except:

- $k = 3$ $n_1, n_2, n_3 \leq 5$: (Table 14, 202=p252)
- $k = 2$ $n_1, n_2 \leq 10$: (Table 12, 202=p250)

The Kruskal-Wallis test is readily implemented on MINITAB:

example (data of Problem 5.14)

MTB > table c2;
SUBC> data c1.
1 510.0 520.0 466.0 510.0 530.0 508.0
2 540.0 528.0 540.0 518.0
3 468.0 484.0 456.0 516.0 454.0
4 544.0 532.0 520.0 518.0
5 460.0 452.0 474.0 468.0 410.0 462.0

MTB > krus c1 c2
LEVEL NOBS MEDIAN AVE. RANK Z VALUE
1 6 510.0 14.3 0.48
2 4 534.0 20.9 2.33
3 5 468.0 8.3 -1.60
4 4 526.0 20.5 2.22
5 6 461.0 5.4 -2.90
H = 17.32 d.f. = 4 p = 0.002

MTB > oneway c1 c2
Source DF SS MS F p
litter 4 22528 5632 13.38 0.000
Error 20 8418 421
Total 24 30946