Normal approximations with Wilcoxon tests

Starter

1. (Review of last lesson) Eleven randomly selected primary school children are weighed. The results, in kg, are shown in the table below.

Girls	43.7	34.4	53.1	38.3	48.5	48.9
Boys	41.4	36.6	34.6	32.5	32.4	

Test at the $1\,\%$ level, whether the data supports the view that, on average, primary school girls have greater weights than the boys.

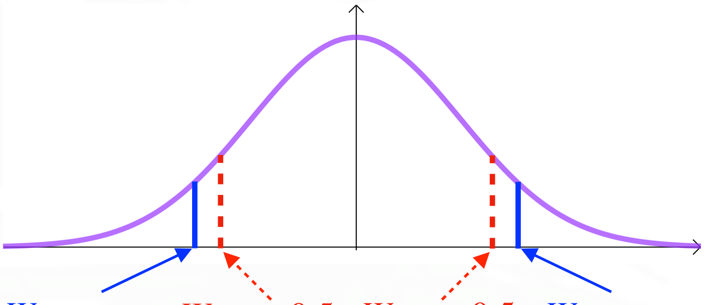
Notes

The tables for Wilcoxon's tests (i.e. the signed-rank and rank-sum tests) only deal with quite small samples sizes i.e. up to 20. With higher sample sizes a normal approximation is used, since under H_0 the test statistic is normally distributed.

Continuity correction

Since $W_{+/-}$ is an integer and the normal distribution is continuous a ***continuity correction*** needs to be made.

In effect $W_{+/-}$ covers the interval between $W_{+/-} - 0.5$ and $W_{+/-} + 0.5$.
When calculating p, choose the end that ***makes the tail as large as possible***.
If $W_{+/-} <$ mean, let $x = W_{+/-} + 0.5$.
If $W_{+/-} >$ mean, let $x = W_{+/-} - 0.5$.



Wilcoxon signed-rank test for large samples

The null and alternative hypotheses are still:

H_0 : the median is the stated value
H_1 : the median is not/greater than/less than the stated value

Success criteria — Wilcoxon signed-rank test for large samples

1. Calculate the mean using $\dfrac{1}{4}n(n+1)$.
2. Calculate the variance using $\dfrac{1}{24}n(n+1)(2n+1)$ — remember to square root it to find σ.
3. If test statistic, $W_{+/-} <$ mean, use $W_{+/-} + 0.5$.
4. Use your calculator to find $p \approx P(T \le W_{+/-} + 0.5)$.

...or...

3. If test statistic, $W_{+/-} >$ mean, use $W_{+/-} - 0.5$.
4. Use your calculator to find $p \approx P(T \ge W_{+/-} - 0.5)$.
5. Accept H_0 if $p \nleq \alpha$ (one-tailed tests) or $p \nleq \dfrac{\alpha}{2}$ (two-tailed tests)

   Reject H_0 if $p \le \alpha$ (one-tailed tests) or $p \le \dfrac{\alpha}{2}$ (two-tailed tests)

In reality, W_+ and W_- would need to be calculated the usual way but in an exam it would take too long to calculate them with a large sample.

E.g. 1 When a Wilcoxon signed-rank is carried out on a sample size of 30, the value obtained for T is 154. Test whether this result is significant at the $5\ \%$ level for a one-tail test.

Working:

Mean $= \dfrac{1}{4}n(n+1) = \dfrac{1}{4} \times 30 \times 31 = 232.5$

Variance $= \dfrac{1}{24}n(n+1)(2n+1) = \dfrac{1}{24} \times 30 \times 31 \times 61 = 2363.75$

$\sigma = \sqrt{2363.75}$

Continuity correction: since $154 < 232.5$, use $154 + 0.5 = 154.5$

Using a calculator:

Normal CD

- Lower: -99999
- Upper: 154.5
- σ: $\sqrt{2363.75}(\approx 48.618)$
- μ: 232.5

$p = P(T \leq 154) \approx P(T' \leq 154.5) \approx 0.0543$

Since $p \approx 0.0543 \nleq 0.05$, we do not reject H_0.

The result is not significant at the $5\ \%$ level.

E.g. 2 From a sample size of 52, the values of W_+ and W_- are 420 and 958 respectively. Test at the $2\ \%$ level for a two-tail test whether this result is significant.

E.g. 3 Determine the upper and lower $2\ \%$ one-tailed critical values for a Wilcoxon signed-rank test based on a sample size of 76.

Wilcoxon rank-sum test for large samples

When one of the samples has more than 10 values a normal approximation can be used,

The method for the rank-sum test is similar to that above with $W_{+/-}$ replaced by R_m and different calculations for the mean and variance.

The null and alternative hypotheses are still:

H_0 : The medians of the distributions are the same.
H_1 : The medians of the distributions are different.

Success criteria — Wilcoxon rank-sum test for large samples

1. Calculate the mean using $\dfrac{1}{2}m(m+n+1)$.
2. Calculate the variance using $\dfrac{1}{12}mn(m+n+1)$ — remember to square root it to find σ.
3. If test statistic, $R_m <$ mean, use $R_m + 0.5$.
4. Use your calculator to find $p \approx P(T \leq R_m + 0.5)$.

...or...

3. If test statistic, $R_m >$ mean, use $R_m - 0.5$.
4. Use your calculator to find $p \approx P(T \geq R_m - 0.5)$.
5. Accept H_0 if $p \nleq \alpha$ (one-tailed tests) or $p \nleq \dfrac{\alpha}{2}$ (two-tailed tests)

   Reject H_0 if $p \leq \alpha$ (one-tailed tests) or $p \leq \dfrac{\alpha}{2}$ (two-tailed tests)

E.g. 4 A Wilcoxon rank-sum test is carried out on samples of size 12 and 17. The R_m value is 210. State the value of the test statistic, W, you are using and test at the $5\,\%$ level for a one-tail test whether this result is significant.

Working: $m = 12, n = 17$: $\quad m(m+n+1) - R_m = 12(12+17+1) - 210 = 150$

$\therefore W = 150$

Mean $= \dfrac{1}{2}m(m+n+1) = \dfrac{1}{2} \times 12 \times (12+17+1) = 180$

Variance $= \dfrac{1}{12}mn(m+n+1) = \dfrac{1}{12} \times 13 \times 14 \times (13+14+1) = 510$

$\sigma = \sqrt{510}$

Continuity correction: since $150 < 180$, test statistic is $150 + 0.5 = 150.5$

Using a calculator:

Normal CD

Lower: -99999

Upper: 150.5

σ: $\sqrt{510}(\approx 22.583)$

μ: 180

$p = P(W \le 150) \approx P(W' \le 150.5) \approx 0.0957$

Since $p \approx 0.0957 \not\le 0.005$, we do not reject H_0.

The result is not significant at the $5\,\%$ level.

Alternatively:

Since $210 > 180$, the test statistic is $210 - 0.5 = 209.5$

Using a calculator:

Normal CD

Lower: 209.5

Upper: 99999

σ: $\sqrt{510}(\approx 22.583)$

μ: 180

$p = P(W \ge 210) \approx P(W' \ge 209.5) \approx 0.0957$

N.B. When using a normal approximation, there is no need to choose the smallest of R_m and $m(m+n+1) - R_m$ as long as it is known which area is being calculated.

E.g. 5 When a Wilcoxon rank-sum test was carried out the value obtained was $W = 126$ with $m = 13$ and $n = 14$. Is this result significant at the $1\,\%$ level for a two-tail test?

E.g. 6 Determine the upper and lower $10\,\%$ one-tailed critical values for a Wilcoxon rank-sum test based on a sample sizes of $m = 10$ and $n = 14$.

What happens when the sum of the ranks of the larger sample is given?

When the sum of the ranks of the larger sample is given, use sum of ranks $= \dfrac{1}{2}N(N+1)$ where $N = m + n$ to find the sum of the ranks of the smaller sample.

E.g. 7 Two cold remedies, A and B, are compared by a consumer group. Two random samples of people with colds were given either remedy A or remedy B. The size of the sample for A was 32 and sample size of B was 27. The sum of the ranks of A was 1086. Test at the $5\ \%$ significance level whether there is difference in the ranks of the two remedies.

Working: $m = 27, n = 32$

The sum of the ranks of the smaller sample is needed.

Sum of ranks $= \frac{1}{2}n(n+1) = \frac{1}{2} \times (27+32) \times (27+32+1) = 1770$

So the sum of the ranks of sample $B = 1770 - 1084 = 686$

Mean $= \frac{1}{2}m(m+n+1) = \frac{1}{2} \times 27 \times (27+32+1) = 810$

Variance $= \frac{1}{12}mn(m+n+1) = \frac{1}{12} \times 27 \times 32 \times (27+32+1) = 4320$

$H_0: \ m_A = m_B$

$H_1: \ m_A \neq m_B$ where m_A and m_B are the medians of the rankings given to remedy A and B respectively.

This is a two-tailed test.

$R_B \sim \text{N}(810, 4320)$

Continuity correction: as $R_B = 686 < 810 =$ mean, add 0.5.

$p = P(R_B \leq 686) \approx P(T' \leq 686.5) \approx 0.0301$

Since $p \approx 3.01\ \% > 2.5\ \%$ (two-tailed test), we do not reject H_0.

There is evidence to suggest that the medians of the cold remedies are not different i.e. there is insufficient evidence that there is a difference in the cold remedies effectiveness.

E.g. 8 A school wants to test whether online learning is as effective as normal class teaching. Two groups of randomly chosen students were taught the same course and then took an exam to test their understanding. Group X, of 18 students, was taught by a classroom teacher while Group Y, of 24 students, followed a course online. The sum of the ranks of Group Y in the exam was 444. Test at the $5\ \%$ significance level whether normal teaching methods are better than online learning.

Video (password): Single sample tests: normal approximation for large samples

Video (password): Paired sample tests: normal approximation for large samples

Video: Normal approximations with Wilcoxon tests

Solutions to Starter and E.g.s

Exercise

p60 4E Qu 1-4

Summary

Success criteria — Wilcoxon signed-rank test for large samples:

1. Calculate the mean using $\frac{1}{4}n(n+1)$.
2. Calculate the variance using $\frac{1}{24}n(n+1)(2n+1)$ — remember to square root it to find σ.
3. If test statistic, $W_{+/-} <$ mean, use $W_{+/-} + 0.5$.
4. Use your calculator to find $p \approx P(T \leq W_{+/-} + 0.5)$.

...or...

3. If test statistic, $W_{+/-} >$ mean, use $W_{+/-} - 0.5$.
4. Use your calculator to find $p \approx P(T \geq W_{+/-} - 0.5)$.
5. Accept H_0 if $p \nleq \alpha$ (one-tailed tests) or $p \nleq \frac{\alpha}{2}$ (two-tailed tests)

   Reject H_0 if $p \leq \alpha$ (one-tailed tests) or $p \leq \frac{\alpha}{2}$ (two-tailed tests)

Success criteria — Wilcoxon rank-sum test for large samples:

1. Calculate the mean using $\frac{1}{2}m(m+n+1)$.
2. Calculate the variance using $\frac{1}{12}mn(m+n+1)$ — remember to square root it to find σ.
3. If test statistic, $R_m <$ mean, use $R_m + 0.5$.
4. Use your calculator to find $p \approx P(T \leq R_m + 0.5)$.

...or...

3. If test statistic, $R_m >$ mean, use $R_m - 0.5$.
4. Use your calculator to find $p \approx P(T \geq R_m - 0.5)$.
5. Accept H_0 if $p \nleq \alpha$ (one-tailed tests) or $p \nleq \frac{\alpha}{2}$ (two-tailed tests)

   Reject H_0 if $p \leq \alpha$ (one-tailed tests) or $p \leq \frac{\alpha}{2}$ (two-tailed tests)