

Statistical Methods

Nels Grevstad

Metropolitan State University of Denver
ngrevsta@msudenver.edu

December 3, 2019

Nels Grevstad

Topics

- 1 Wilcoxon Rank Sum Test for Two Population Means μ_1 and μ_2
- 2 Large Sample Version of the Wilcoxon Rank Sum Test

Nels Grevstad

Objectives

Objectives:

- Carry out a Wilcoxon rank sum test for two population means.
- Carry out the large sample version of the Wilcoxon rank sum test for two population means.

Nels Grevstad

Wilcoxon Rank Sum Test for Two Population Means μ_1 and μ_2

- The **Wilcoxon rank sum test** is a **nonparametric** alternative to the **two-sample t test**.

(It's equivalent to a test called the *Mann-Whitney test*, which uses a different test statistic but always produces the same p-value.)

Nels Grevstad

Notes

Notes

Notes

Notes

- We only assume X_1, X_2, \dots, X_m and Y_1, Y_2, \dots, Y_n are two *independent* random samples from continuous distributions that have the **same shape** but possibly different means μ_1 and μ_2 .

Nels Grevstad

- The **null hypothesis** is that there's no difference between μ_1 and μ_2 .

Null Hypothesis:

$$H_0 : \mu_1 - \mu_2 = 0$$

Nels Grevstad

- The **alternative hypothesis** will depend on what we're trying to "prove":

Alternative Hypothesis: The alternative hypothesis will be one of

1. $H_a : \mu_1 - \mu_2 > 0$ **(one-sided, upper-tailed)**
2. $H_a : \mu_1 - \mu_2 < 0$ **(one-sided, lower-tailed)**
3. $H_a : \mu_1 - \mu_2 \neq 0$ **(two-sided, two-tailed)**

depending on what we're trying to verify using the data.

Nels Grevstad

Wilcoxon Rank Sum Test Statistic for $\mu_1 - \mu_2$:

1. If the sample sizes *aren't* the same, denote the **smaller** sample by X_1, X_2, \dots, X_m and the **larger** one by Y_1, Y_2, \dots, Y_n .

If they *are* the same, **either** sample can be denoted X_1, X_2, \dots, X_m and the other Y_1, Y_2, \dots, Y_n .

2. **Combine** the two samples and **rank** the observations from smallest (rank = 1) to largest (rank = $m + n$), keeping track of which are X_i 's and which are Y_i 's. For **ties**, use the **average** of the **ranks** that would've been assigned if there weren't any ties.

Nels Grevstad

Notes

Notes

Notes

Notes

3. The **test statistic**, denoted W , is

$$W = \text{Sum of ranks of } X_i\text{'s.}$$

Example

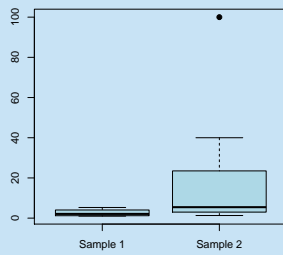
Consider the following (fabricated) samples from populations whose (unknown) means are μ_1 and μ_2 :

Sample 1: 1.0, 1.4, 2.8, 5.3

Sample 2: 1.3, 2.5, 3.5, 5.4, 5.5, 7.0, 40.0, 100.0

Boxplots of the two samples are on the next slide.

Boxplots of Samples 1 and 2



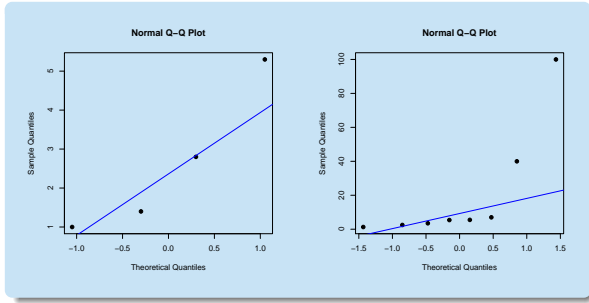
Normal probability plots (next slide) indicate that **both** samples are from **right skewed** populations, so the two-sample *t* test *isn't* appropriate.

Notes

Notes

Notes

Notes



Nels Grevstad

Here are the samples **combined, sorted, and ranked**.

Observation	1.0	1.3	1.4	2.5	2.8	3.5	5.3	5.4	5.5	7.0	40.0	100.0
Sample	X	Y	X	Y	X	Y	X	Y	Y	Y	Y	Y
Rank	1	2	3	4	5	6	7	8	9	10	11	12

Note that the **smaller** sample is denoted X , and its size is $m = 4$. The **larger** sample is denoted Y , and its size is $n = 8$.

The **test statistic** for a Wilcoxon rank sum test to decide if μ_1 is less than μ_2 is

$$\begin{aligned}
 W &= \text{Sum of ranks of } X_i\text{'s} \\
 &= 1 + 3 + 5 + 7 \\
 &= 16.
 \end{aligned}$$

Nels Grevstad

- W will be large when the observations in the X sample tend to be *larger* than those in the Y sample, as would be the case if μ_1 was greater than μ_2 . W will be small when the observations in the X sample tend to be *smaller* than those in the Y sample, as would be the case if μ_1 was less than μ_2 . Therefore:

Nels Grevstad

1. Large values of W provide **evidence against H_0 in favor of $H_a : \mu_1 - \mu_2 > 0$.**
2. Small values of W provide **evidence against H_0 in favor of $H_a : \mu_1 - \mu_2 < 0$.**
3. Large *and* small values of W provide **evidence against H_0 in favor of $H_a : \mu_1 - \mu_2 \neq 0$.**

Nels Grevstad

Notes

Notes

Notes

Notes

- **Comment:** Sometimes we want to test

$$H_0 : \mu_1 - \mu_2 = \Delta_0,$$

for some (non-zero) value Δ_0 . In this case, simply compute W as described above, but replacing each observation X_i in the X sample by $X_i - \Delta_0$.

Nels Grevstad

Sampling Distribution of the Test Statistic Under H_0 :

If W is the Wilcoxon rank sum test statistic, then when

$$H_0 : \mu_1 - \mu_2 = 0$$

is true, W follows a so-called **Wilcoxon rank sum distribution**, which has two parameters m and n (the **sample sizes**), i.e.

$$W \sim \text{Wilcoxon}_{RS}(m, n).$$

Nels Grevstad

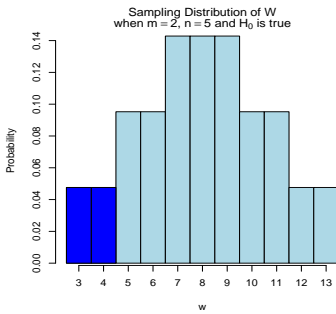


Figure: $\text{Wilcoxon}_{RS}(m, n)$ when $m = 2, n = 5$. The shaded area is the p-value for a lower-tailed test when $W = 4$.

Nels Grevstad

• **Properties of $\text{Wilcoxon}_{SR}(n)$ distributions**

- They're **symmetric, discrete** distributions.
- The probability lies between $m(m + 1)/2$ ($= 1 + 2 + \dots + m$) and $mn + m(m + 1)/2$ ($= (n + 1) + (n + 2) + (n + 3) + \dots + (n + m)$).
- They're **centered** on $m(m + n + 1)/2$ (which is the *mean* of the distribution).
- As m and n both increase, the $\text{Wilcoxon}_{RS}(m, n)$ distributions approach a **normal** distribution.

Nels Grevstad

Notes

Notes

Notes

Notes

Recall that the **Wilcoxon rank sum test statistic** for deciding if μ_1 is less than μ_2 from the previous example is $W = 16$.

By symmetry of the Wilcoxon_{RS}(m, n) distribution, when $m = 4$ and $n = 8$,

$$P(W \leq 16) = P(W \geq 4(4 + 8 + 1) - 16) = P(W \geq 36).$$

From **Table A14**, the **p-value** for the lower-tailed test is **0.055**.

There's *no statistically significant* evidence that the true mean μ_1 is less than μ_2 .

Nels Grevstad

Exercise

In an experiment to compare plastics produced by two different processes, six specimens of each plastic were tested for strength (breaking loads in 1000s of lbs per square inch). Here are the data.

Plastic 1	Plastic 2
15.3	12.1
18.7	22.4
22.3	18.3
17.6	19.3
19.1	17.1
14.8	27.7

Nels Grevstad

Here are the samples combined and sorted, with their ranks.

Observation	12.1	14.8	15.3	17.1	17.6	18.3	18.7	19.1	19.3	22.3	22.4	27.7
Sample	Y	X	X	Y	X	Y	X	X	Y	X	Y	Y
Rank	1	2	3	4	5	6	7	8	9	10	11	12

Note that the **sample sizes** are the **same**, $m = n = 6$, and that **Plastic 1** is denoted **X** and **Plastic 2** denoted **Y**:

Nels Grevstad

Without assuming normality of the data, carry out a **Wilcoxon rank sum test** to decide if there's statistically significant evidence for **any difference** between the true mean breaking strengths for the **two plastics**, μ_1 and μ_2 .

Nels Grevstad

Notes

Notes

Notes

Notes

Large Sample Version of the Wilcoxon Rank Sum Test

- When m and n are both large, the Central Limit Theorem (which applies not just to means, but also to *sums* of random variables) says that W follows approximately a normal distribution.

Nels Grevstad

Proposition

- When $H_0 : \mu_1 - \mu_2 = 0$ is true, the mean and standard error of the distribution of W , denoted μ_w and σ_w , are

$$\begin{aligned} \mu_w &= E(W) = \frac{m(m+n+1)}{2} \\ \sigma_w &= \sqrt{V(W)} = \sqrt{\frac{mn(m+n+1)}{12}} \end{aligned} \quad (1)$$

- When $H_0 : \mu_1 - \mu_2 = 0$ is true and m and n are both large ($m \geq 10$ and $n \geq 10$),

$$W \sim N(\mu_w, \sigma_w)$$

(approximately).

Nels Grevstad

Large Sample Wilcoxon Rank Sum Test Statistic for $\mu_1 - \mu_2$:

$$Z = \frac{W - \mu_w}{\sigma_w} = \frac{W - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}}$$

Nels Grevstad

- Now suppose the sample sizes m and n are *both large*. In this case, the sampling distribution of the test statistic is as follows.

Sampling Distribution of the Test Statistic Under H_0 :

If Z is the large sample Wilcoxon rank sum test statistic, then when

$$H_0 : \mu_1 - \mu_2 = 0$$

is true,

$$Z \sim N(0, 1).$$

Notes

Notes

Notes

Notes

Nels Grevstad

- The $N(0, 1)$ curve gives us:
 - The **rejection region** as the **extreme $100\alpha\%$ of Z values** (in the direction(s) specified by H_a).
 - The **p -value** as the **tail area(s) beyond the observed Z value** (in the direction(s) specified by H_a).

Nels Grevstad

- **Comment:** Most statistical software uses a slightly more accurate **continuity corrected** version of the test statistic Z .

The correction adjusts for the fact that a **continuous** distribution (the $N(0, 1)$ distribution) is being used to approximate a discrete one (the true $\text{Wilcoxon}_{RS}(n)$ distribution).

Nels Grevstad

Notes

Notes

Notes

Notes
