Statistical Methods

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

December 3, 2019

Topics

Objectives

Objectives:

- Carry out a Wilcoxon signed ranks test for a population mean.
- Carry out the paired samples version of a Wilcoxon signed ranks test for two population means.
- Carry out the large sample version of the Wilcoxon signed ranks test for a population mean.

Wilcoxon Signed Ranks Test for a Population Mean μ

Parametric and Nonparametric Tests

- A ***parametric test*** is one that requires an **assumption** that the data are a sample from a some **specific probability distribution** (e.g. **normal**).

Wilcoxon Signed Ranks Test for a Population Mean μ

Parametric and Nonparametric Tests

- A ***parametric test*** is one that requires an **assumption** that the data are a sample from a some **specific probability distribution** (e.g. **normal**).

  A ***nonparametric test*** is one that requires no such assumption.

Wilcoxon Signed Ranks Test for a Population Mean μ

Parametric and Nonparametric Tests

- A ***parametric test*** is one that requires an **assumption** that the data are a sample from a some **specific probability distribution** (e.g. **normal**).

  A ***nonparametric test*** is one that requires no such assumption.

  The t **tests** and **ANOVA** F **tests** are **parametric** tests.

- The ***Wilcoxon signed ranks test*** is a **nonparametric** alternative to the **one-sample t test.**

- The ***Wilcoxon signed ranks test*** is a **nonparametric** alternative to the **one-sample t test.**
- We only assume only that $X_1, X_2, \ldots, X_n$ are a random sample from ***some*** **continuous**, **symmetric distribution** whose mean is μ.

- The **null hypothesis** is that μ is equal to a claimed value μ_0.

Null Hypothesis:

$$H_0: \ \mu \ = \ \mu_0$$

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

Alternative Hypothesis: The alternative hypothesis will be one of

1. $H_a: \mu > \mu_0$ (**one-sided, upper-tailed**)
2. $H_a: \mu < \mu_0$ (**one-sided, lower-tailed**)
3. $H_a: \mu \neq \mu_0$ (**two-sided, two-tailed**)

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

- **Comment**: Because the population distribution is *symmetric*, its mean, μ, is also its median (50th percentile), $\tilde{\mu}$.

- **Comment**: Because the population distribution is *symmetric*, its mean, $\boldsymbol{\mu}$, is also its median (50th percentile), $\tilde{\boldsymbol{\mu}}$.

  Thus we could state H_0 and H_a in terms of the **population median**, e.g.

$$H_0 : \tilde{\mu} \ = \ \mu_0$$

Wilcoxon Signed Ranks Test Statistic for μ:

1. Discard any observations X_i that equal μ_0, and diminish the sample size by the number of discarded X_i's before proceeding with Steps 2 - 3.
2. **Rank** the absolute deviations $|\,X_i - \mu_0\,|$ from smallest (rank $= 1$) to largest (rank $= n$), keeping track of each deviation's original sign. For **ties**, use the **average** of the **ranks** that would've been assigned if there weren't any ties.
3. The **test statistic**, denoted S_+, is

$$S_+ \;=\; \text{Sum of ranks of } |\,X_i - \mu_0\,| \text{ 's that were positive.}$$

Example

The **price-to-earnings** (**P/E**) **ratio** of a stock is an important tool in finance. A low P/E ratio indicates a ”value” or ”bargain” stock.

Example

The **price-to-earnings** (**P/E**) **ratio** of a stock is an important tool in finance. A low P/E ratio indicates a ”value” or ”bargain” stock.

A recent issue of the *Wall Street Journal* indicated that the P/E ratio of the entire S&P 500 stock index is **19.0**.

Example

The **price-to-earnings** (**P/E**) **ratio** of a stock is an important tool in finance. A low P/E ratio indicates a ”value” or ”bargain” stock.

A recent issue of the *Wall Street Journal* indicated that the P/E ratio of the entire S&P 500 stock index is **19.0**.

The next slide shows the **P/E ratios** in a sample of $n = 14$ large U.S. banks, as reported in *Forbes* magazine. Also shown are the **deviations** away from **19.0**.

P/E Ratio (X_i)	Deviation ($X_i - 19.0$)
24.3	5.3
15.8	-3.2
22.1	3.1
14.4	-4.6
11.7	-7.3
13.2	-5.8
17.0	-2.0
22.1	3.1
15.4	-3.6
19.0	0.0
23.0	4.0
13.2	-5.8
10.9	-8.1
18.2	-0.8

We want to decide if the true **mean P/E ratio** for large U.S. banks, μ, is **different** from **19.0**:

$$\begin{aligned} H_0 : \mu &= 19.0 \\ H_a : \mu &\neq 19.0 \end{aligned}$$

We want to decide if the true **mean P/E ratio** for large U.S. banks, μ, is **different** from **19.0**:

$$\begin{aligned} H_0 : \mu &= 19.0 \\ H_a : \mu &\neq 19.0 \end{aligned}$$

Ratios of random variables, like the **P/E ratio**, often follow a *heavy tailed*, *non-normal* (but symmetric) distribution called the *Cauchy* distribution.

We want to decide if the true **mean P/E ratio** for large U.S. banks, μ, is **different** from **19.0**:

$$\begin{aligned} H_0 : \mu &= 19.0 \\ H_a : \mu &\neq 19.0 \end{aligned}$$

Ratios of random variables, like the **P/E ratio**, often follow a *heavy tailed*, *non-normal* (but symmetric) distribution called the *Cauchy* distribution.

(We could also check the normality assumption with a normal probability plot of the **P/E ratios**.)

We want to decide if the true **mean P/E ratio** for large U.S. banks, μ, is **different** from **19.0**:

$$\begin{aligned} H_0 : \mu &= 19.0 \\ H_a : \mu &\neq 19.0 \end{aligned}$$

Ratios of random variables, like the **P/E ratio**, often follow a *heavy tailed*, *non-normal* (but symmetric) distribution called the *Cauchy* distribution.

(We could also check the normality assumption with a normal probability plot of the **P/E ratios**.)

Thus a one-sample t test *isn't* appropriate.

We'll carry out a **Wilcoxon signed ranks test.** Here are the **sorted absolute values** of the **deviations** along with their original **signs** and their **ranks**:

We'll carry out a **Wilcoxon signed ranks test.** Here are the **sorted absolute values** of the **deviations** along with their original **signs** and their **ranks**:

$\mid X_i - 19 \mid$	0.0	0.8	2.0	3.1	3.1	3.2	3.6	4.0	4.6	5.3	5.8	5.8	7.3	8.1
Sign	NA	-	-	+	+	-	-	+	-	+	-	-	-	-
Rank	NA	1	2	3.5	3.5	5	6	7	8	9	10.5	10.5	12	13

We'll carry out a **Wilcoxon signed ranks test.** Here are the **sorted absolute values** of the **deviations** along with their original **signs** and their **ranks**:

$\mid X_i - 19 \mid$	0.0	0.8	2.0	3.1	3.1	3.2	3.6	4.0	4.6	5.3	5.8	5.8	7.3	8.1
Sign	NA	-	-	+	+	-	-	+	-	+	-	-	-	-
Rank	NA	1	2	3.5	3.5	5	6	7	8	9	10.5	10.5	12	13

Note that the **zero** deviation is discarded, and so now $n = 13$, and the **ties** are assigned the **average rank.**

The **test statistic** is

$$\begin{aligned} S_+ &= \text{Sum of ranks of } |\, X_i - 19 \,| \text{'s that were positive} \\ &= 3.5 + 3.5 + 7 + 9 \\ &= \mathbf{23}. \end{aligned}$$

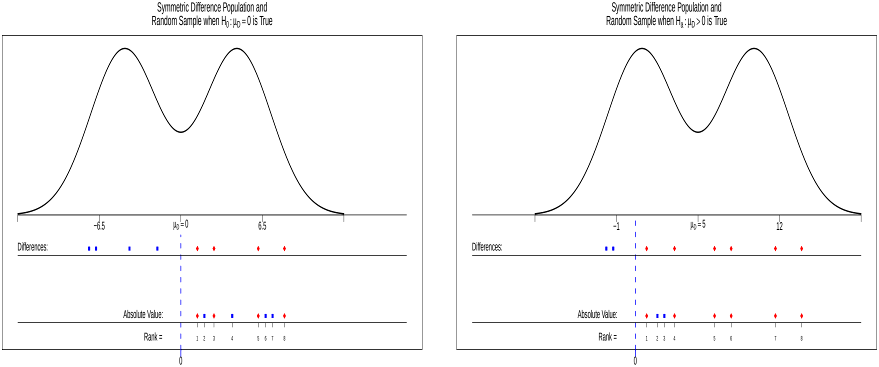

Figure: Symmetric populations and samples from them. Absolute values of positive (blue squares) and negative (red diamonds) at the bottom. Left plot, $H_0 : \mu = 0$ is true. Right plot, $H_a : \mu > 0$ is true.

- S_+ will be **large** when the deviations $X_i - \mu_0$ are *larger* in the *positive* direction than in the *negative* direction, as would be the case if $\mu > \mu_0$.

- S_+ will be **large** when the deviations $X_i - \mu_0$ are *larger* in the *positive* direction than in the *negative* direction, as would be the case if $\mu > \mu_0$.
- S_+ will be **small** when the deviations in the *positive* direction are *smaller* than those in the *negative* direction, as would be the case if $\mu < \mu_0$.

1. **Large** values of S_+ provide **evidence against H_0 in favor of $H_a : \mu > \mu_0$.**
2. **Small** values of S_+ provide **evidence against H_0 in favor of $H_a : \mu < \mu_0$.**
3. **Large *and* small** values of S_+ provide **evidence against H_0 in favor of $H_a : \mu \neq \mu_0$.**

Sampling Distribution of the Test Statistic Under H_0:
If S_+ is the Wilcoxon signed ranks test statistic, then when

$$H_0 : \mu = \mu_0$$

is true, S_+ follows a so-called ***Wilcoxon signed ranks distribution***, which has one parameter n (the **sample size**), i.e.

$$S_+ \sim \text{Wilcoxon}_{\text{SR}}(n).$$

- **Properties of Wilcoxon$_{\text{SR}}(n)$ distributions**
  - They're **symmetric**, **discrete** distributions.

- **Properties of Wilcoxon$_{\text{SR}}(n)$ distributions**
  - They're **symmetric**, **discrete** distributions.
  - The probability lies between **0** and $\boldsymbol{n(n+1)/2}$ $(= 1 + 2 + \ldots + n)$.

- **Properties of Wilcoxon$_{\text{SR}}(n)$ distributions**
  - They're **symmetric**, **discrete** distributions.
  - The probability lies between **0** and $\boldsymbol{n(n+1)/2}$ $(= 1 + 2 + \ldots + n)$.
  - They're **centered** on $\boldsymbol{n(n+1)/4}$ (which is the *mean* of the distribution).

- **Properties of Wilcoxon$_{\text{SR}}(n)$ distributions**
  - They're **symmetric**, **discrete** distributions.
  - The probability lies between **0** and $n(n+1)/2$ $(= 1 + 2 + \ldots + n)$.
  - They're **centered** on $n(n+1)/4$ (which is the *mean* of the distribution).
  - As n increases, the Wilcoxon$_{\text{SR}}(n)$ distributions approach a **normal** distribution.

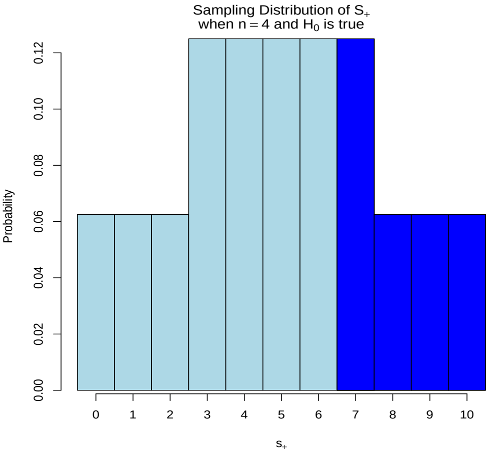

Figure: Wilcoxon$_{\text{SR}}(n)$ distribution when $n = 4$. The shaded area is the upper-tailed p-value when $S_+ = 7$.

Example

Recall that for a test of

$$H_0: \mu = 19.0$$
$$H_a: \mu \neq 19.0$$

where μ is the true **mean P/E ratio** for large U.S. banks, a sample of $n = 14$ (diminished to $n = 13$) banks produced a **Wilcoxon signed ranks test statistic**

$$S_+ = \mathbf{23}.$$

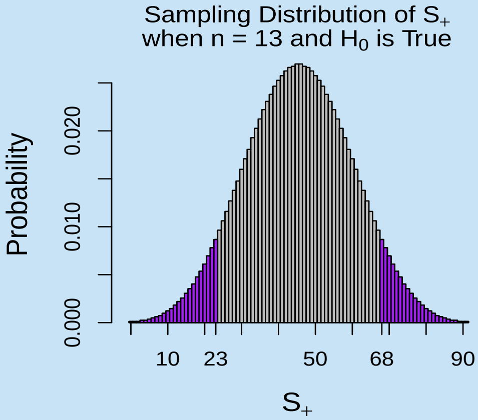

Figure: Wilcoxon$_{\text{SR}}(n)$ distribution when $n = 13$. The shaded area is the two-tailed p-value when $S_+ = 23$.

By symmetry of the Wilcoxon$_{\text{SR}}(n)$ distribution, when $\boldsymbol{n = 13}$,

$$P(S_+ \leq 23) \; = \; P(S_+ \geq 13(14)/2 - 23) \; = \; P(S_+ \geq 68)\,.$$

By symmetry of the Wilcoxon$_{\text{SR}}(n)$ distribution, when $\boldsymbol{n = 13}$,

$$P(S_+ \leq 23) \;=\; P(S_+ \geq 13(14)/2 - 23) \;=\; P(S_+ \geq 68)\,.$$

From **Table A13**, the **p-value** for the two-tailed test is **between 2(0.055)** and **2(0.095)**, i.e. **between 0.110** and **0.190**.

By symmetry of the Wilcoxon$_{\text{SR}}(n)$ distribution, when $\boldsymbol{n = 13}$,

$$P(S_+ \leq 23) \;=\; P(S_+ \geq 13(14)/2 - 23) \;=\; P(S_+ \geq 68)\,.$$

From **Table A13**, the **p-value** for the two-tailed test is **between 2(0.055)** and **2(0.095)**, i.e. **between 0.110** and **0.190**.

There's *no statistically significant* evidence that the true mean **P/E ratio** for U.S. banks differs from **19.0**.

Exercise

The data on the next slide are estimated values of the true (unknown) **ratio** μ of the mass of the earth to that of the moon obtained from seven different Mariner and Pioneer spacecraft in the 1960's.

Spacecraft	Estimated Ratio X_i	Deviation $X_i - 81.3035$
Mariner 2 (Venus)	81.3001	-0.0034
Mariner 4 (Mars)	81.3015	-0.0020
Mariner 5 (Venus)	81.3006	-0.0029
Mariner 6 (Mars)	81.3011	-0.0024
Mariner 7 (Mars)	81.2997	-0.0038
Pioneer 6	81.3005	-0.0030
Pioneer 7	81.3021	-0.0014

Prior to obtaining these estimates, scientists had considered μ to be **81.3035**. We'll use the data to test

$$\begin{aligned} H_0 : \mu &= 81.3035 \\ H_a : \mu &\neq 81.3035 \end{aligned}$$

Because the data are *ratios*, they (likely) follow the (non-normal) *Cauchy distribution*, so a one-sample t test isn't appropriate.

Carry out a **Wilcoxon signed ranks test** using $\alpha = 0.05$. Here are the **sorted absolute values** of the **deviations** along with their original **signs** and their **ranks**:

$\lvert X_i - 81.3035 \rvert$	0.0014	0.0020	0.0024	0.0029	0.0030	0.0034	0.0038
Sign	-	-	-	-	-	-	-
Rank	1	2	3	4	5	6	7

Carry out a **Wilcoxon signed ranks test** using $\alpha = 0.05$. Here are the **sorted absolute values** of the **deviations** along with their original **signs** and their **ranks**:

$\lvert X_i - 81.3035 \rvert$	0.0014	0.0020	0.0024	0.0029	0.0030	0.0034	0.0038
Sign	-	-	-	-	-	-	-
Rank	1	2	3	4	5	6	7

Hint: You should get $S_+ = 0$ and a **p-value** of **2(0.008) = 0.016**.

Lack of Power of Nonparametric Tests

- Notice (Table A13) that when $n = 4$, the **largest** possible value, $S_+ = 10$, gives an upper-tailed p-value of **0.0625.**

Lack of Power of Nonparametric Tests

- Notice (Table A13) that when $n = 4$, the **largest** possible value, $S_+ = 10$, gives an upper-tailed p-value of **0.0625.**

  Thus for this sample size, it's **not possible** to **reject** $\boldsymbol{H_0}$ (using $\alpha = 0.05$).

Lack of Power of Nonparametric Tests

- Notice (Table A13) that when $n = 4$, the **largest** possible value, $S_+ = 10$, gives an upper-tailed p-value of **0.0625.**

  Thus for this sample size, it's **not possible** to **reject H_0** (using $\alpha = 0.05$).

  In general, when n is **small**, **nonparametric** tests ***lack power*** for rejecting H_0.

Lack of Power of Nonparametric Tests

- Notice (Table A13) that when $n = 4$, the **largest** possible value, $S_+ = 10$, gives an upper-tailed p-value of **0.0625.**

  Thus for this sample size, it's **not possible** to **reject H_0** (using $\alpha = 0.05$).

  In general, when n is **small**, **nonparametric** tests ***lack power*** for rejecting H_0.

  Intuitively, it's because some information is discarded when raw data are converted to ranks.

- Key takeaway: **Parametric** tests are ***more powerful*** than **nonparametric** ones.

- Key takeaway: **Parametric** tests are ***more powerful*** than **nonparametric** ones.

  Thus, for example, when the **normality** assumption is **met**, the t **test should be used** instead of the Wilcoxon signed ranks test.

Paired Samples Version of the Wilcoxon Signed Ranks Test

- The **Wilcoxon signed ranks test** can serve as a **nonparametric** alternative to the **paired t test.**

Paired Samples Version of the Wilcoxon Signed Ranks Test

- The **Wilcoxon signed ranks test** can serve as a **nonparametric** alternative to the **paired** t **test.**
- We'll assume only that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are **paired samples** from continuous distributions whose means are $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$, and that the **differences** $D_1, D_2, \ldots, D_n$, where

$$D_i = X_i - Y_i,$$

follow a **symmetric** distribution.

- The next fact says that as long as the X and Y distributions have the **same shape**, the **differences** will follow a **symmetric** distribution.

- The next fact says that as long as the X and Y distributions have the **same shape**, the **differences** will follow a **symmetric** distribution.

  In particular, the X and Y distributions *don't* themselves have to be symmetric.

Proposition

Suppose X and Y are random observations from *any* two continuous distributions that have the **same shape** (but possibly different means μ_1 and μ_2). Let

$$D = X - Y.$$

Then the distribution of D is continuous and **symmetric** about the value

$$\mu_d = \mu_1 - \mu_2.$$

- The **null hypothesis** is that μ_d is equal zero.

Null Hypothesis:

$$H_0: \ \mu_d \ = \ 0$$

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

Alternative Hypothesis: The alternative hypothesis will be one of

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

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

- The test is conducted exactly as a ***one-sample*** **Wilcoxon signed ranks test**, but using the sample of **differences** $D_1, D_2, \ldots, D_n$.

Large Sample Version of the Wilcoxon Signed Ranks Test

- When n is **large**, the **Central Limit Theorem** (which applies not just to *means*, but also to ***sums*** of random variables) says that S_+ (which is a ***sum***) follows a **normal** distribution (approximately).

Proposition

1. The mean and standard error (standard deviation) of the Wilcoxon$_{\text{SR}}(n)$ distribution, denoted $\boldsymbol{\mu_{s_+}}$ and $\boldsymbol{\sigma_{s_+}}$, are

$$\begin{aligned} \mu_{s_+} &= E(S_+) = \frac{n(n+1)}{4} \\ \sigma_{s_+} &= SD(S_+) = \sqrt{\frac{n(n+1)(2n+1)}{24}} \end{aligned} \tag{1}$$

2. When n **is large ($n > 20$)**, the Wilcoxon$_{\text{SR}}(n)$ distribution is (approximately) **normal**, i.e.

$$S_+ \sim \text{N}(\mu_{s_+},\ \sigma_{s_+})$$

(approximately).

Large Sample Wilcoxon Signed Ranks Test Statistic for μ:

$$Z = \frac{S_+ - \mu_{s_+}}{\sigma_{s_+}} = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}.$$

- Now suppose the sample size n is *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 signed ranks test statistic, then when

$$H_0 : \mu = \mu_0$$

is true,

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

- The $N(0, 1)$ curve gives us:

- The $N(0, 1)$ curve gives us:
    - The ***rejection region*** as the **extreme 100α% of Z values** (in the direction(s) specified by H_a).

- The $N(0,1)$ curve gives us:
  - The ***rejection region*** as the **extreme 100α% 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).

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

- **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 Wilcoxon$_{\text{SR}}(n)$ distribution).