9 Paired Samples Hypothesis Tests (Cont'd)

MTH 3240 Environmental Statistics

Spring 2020

Notes

Objectives

Objectives:

- Carry out a signed rank test for the difference between two population means.
- Carry out a paired sign test for a population median difference.
- Decide which test (the paired t test, signed rank test test, or paired sign test) is most appropriate for a given set of data.

Notes

Dealing With Non-Normal Data

- The paired t procedures require that the sample of *differences* is from a **normal** population (or that n is **large**).

  If this **normality** assumption isn't met (and n isn't large), there are two possible remedies:

  1. **Transform** the data to normality before carrying out the hypothesis test, or
  2. Carry out a **nonparametric** test (i.e. one that doesn't require normality).

  We'll look at these two approaches one at a time.

Notes

Transforming Data To Normality

- The first approach to testing hypotheses with *paired samples* whose **differences** are **non-normal** is to ***transform*** the data (**both** samples) to normality first.

  (It can be shown that if the X and Y samples are both from **normal** populations, the **differences** will also be **normal**.)

Notes

Carrying Out a Nonparametric Test

- The second approach to testing hypotheses with *paired samples* whose **differences** are **non-normal** is to use a ***nonparametric*** test procedure, i.e. one that **doesn't** rely on a normality assumption.

  The ***signed rank test*** and ***paired sign test*** described ahead are both **nonparametric** alternatives to the *paired t test*.

Notes

The Signed Rank Test

- The ***signed rank test*** is a **paired samples nonparametric** test for the difference between two population **means** $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$.

Notes

- The **null hypothesis** is that there's **no difference** between $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$.

Null Hypothesis:

$$H_0 : \mu_x - \mu_y \ = \ 0.$$

(Same hypothesis as for the paired and two-sample t tests.)

Notes

- The **alternative hypothesis** is one of the following.

Alternative Hypothesis:

1. $H_a : \mu_x - \mu_y > 0$ (**upper-tailed test**)
2. $H_a : \mu_x - \mu_y < 0$ (**lower-tailed test**)
3. $H_a : \mu_x - \mu_y \neq 0$ (**two-tailed test**)

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

(Same hypothesis as for the paired and two-sample t tests.)

Notes

- As for the *paired t test*, we act as though the **differences** $\boldsymbol{D_1, D_2, \ldots, D_n}$ ...

  ... are a random sample from a **population** of **differences** whose mean is $\boldsymbol{\mu_d}$.

Notes

- The hypotheses can be reformulated in terms of μ_d as:

	Hypothesis About $\boldsymbol{\mu_x - \mu_y}$	**Equivalent Hypothesis About $\boldsymbol{\mu_d}$**
Null	$H_0 : \mu_x - \mu_y = 0$	$H_0 : \mu_d = 0$
	$H_a : \mu_x - \mu_y > 0$	$H_a : \mu_d > 0$
Alternatives	$H_a : \mu_x - \mu_y < 0$	$H_a : \mu_d < 0$
	$H_a : \mu_x - \mu_y \neq 0$	$H_a : \mu_d \neq 0$

(Same hypotheses as for the paired t test.)

Notes

- The ***test statistic***, denoted $\boldsymbol{W^+}$, is obtained by taking the **absolute values** of the **differences**, ***ranking*** them (smallest to largest), and then **summing** the **ranks** of the differences that were **originally positive**.

Signed Rank Test Statistic:

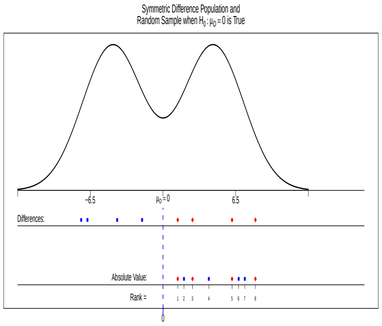
1. If any of the differences $D_1, D_2, \ldots, D_n$ are **zero**, **discard** them and diminish n by the number of discarded D_i's.
2. Take **absolute values** of the remaining **differences**, keeping track of which ones were originally **positive**.

Notes

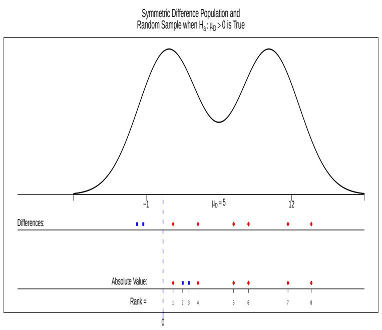
3. **Sort** the **absolute differences** and ***rank*** them from smallest to largest. If two or more are **tied**, assign to each of them the **average** of the **ranks** they would've been assigned if they hadn't been tied.
4. **Sum** the **ranks** of the absolute differences that were originally **positive**. This gives the **test statistic**:

$W^+ \;=\;$ Sum of the ranks of the $|\,D_i\,|$'s for which D_i is positive.

Notes



Notes



Notes

- W^+ reflects whether the **positive** and **negative differences** are **evenly intermingled** or **segregated** (after taking absolute values and sorting).
  - **If H_0 was true**, ...

    ... we'd expect the positive and negative differences to be **intermingled**.
  - But **if H_a was true**, ...

    ... we'd expect the positive and negative differences to be **segregated**, and the $|D_i|$'s for which D_i is positive to mostly lie near the end in the direction specified by H_a.

Notes

- It can be shown that ...
  1. W^+ will be approximately **equal to** $n(n+1)/4$ (most likely), if H_0 is true.
  2. W^+ will **differ from** $n(n+1)/4$ (most likely) in the direction specified by H_a if H_a is true.

Notes

1. *Large* values of W^+ (larger than $n(n+1)/4$) provide evidence in favor of $H_a : \mu_x - \mu_y > 0$ (or $H_a : \mu_d > 0$).
2. *Small* values of W^+ (smaller than $n(n+1)/4$) provide evidence in favor of $H_a : \mu_x - \mu_y < 0$ (or $H_a : \mu_d < 0$).
3. Both *large and small* values of W^+ (larger or smaller than $n(n+1)/4$) provide evidence in favor of $H_a : \mu_x - \mu_y \neq 0$ (or $H_a : \mu_d \neq 0$).

Notes

- Now suppose the sample of **differences** is from **any** (continuous) **population** that has a (roughly) **symmetric** shape.

  In this case, the **null distribution** is as follows.

Notes

Sampling Distribution of W^+ Under H_0: If W^+ is the signed rank test statistic, then when

$$H_0 : \mu_x - \mu_y \;=\; 0 \qquad \text{(or equivalently } H_0 : \mu_d \;=\; 0\text{)}$$

is true, W^+ follows a distribution called the ***Wilcoxon signed rank distribution***, which will depend on n. We write this as

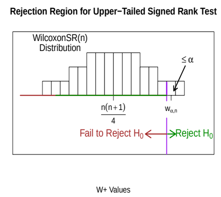
$$W^+ \sim \text{WilcoxonSR}(n).$$

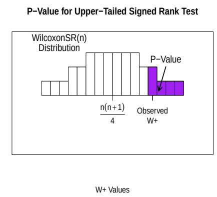
Notes

- **P-values** and **rejection regions** are obtained from the appropriate tail(s) of the **WilcoxonSR(n) distribution**, as shown on the next slides.

Notes

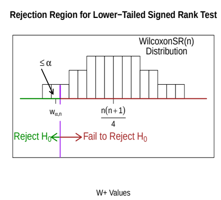
1. $H_a : \mu_x - \mu_y > 0$ (Upper-Tailed Test)

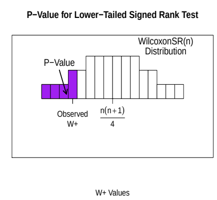




Notes

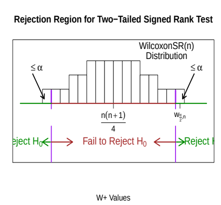
2. $H_a : \mu_x - \mu_y < 0$ (Lower-Tailed Test)

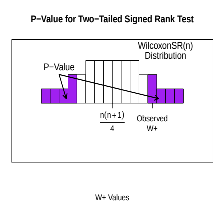




Notes

3. $H_a : \mu_x - \mu_y \neq 0$ (Two-Tailed Test)





Notes

Paired Samples Signed Rank Test for μ_d

Assumptions: $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are two random samples that are paired and the differences $d_1, d_2, \ldots, d_n$ form a single sample from a *continuous* population whose distribution is *symmetric*.

Null hypothesis: $H_0 : \mu_d = 0$.

Test statistic value: $w^+ =$ sum of the ranks of $|\, d_i \,|$'s for which $d_i > 0$.

Decision rule: Reject H_0 if p-value $< \alpha$ or w^+ is in rejection region.

Notes

Null hypothesis: $H_0 : \mu_d = 0$.

Test statistic value: $w^+ =$ sum of the ranks of $|\,d_i\,|$'s for which $d_i > 0$.

Decision rule: Reject H_0 if p-value $< \alpha$ or w^+ is in rejection region.

Alternative hypothesis	P-value = tail probability of the W^+ distribution under H_0: *	Rejection region = w^+ values such that: **
$H_a : \mu_d > 0$	to the right of (and including) w^+	$w^+ \geq w_{\alpha,n}$
$H_a : \mu_d < 0$	to the left of (and including) w^+	$w^+ \leq w^*_{\alpha,n}$
$H_a : \mu_d \neq 0$	$2 \cdot$ (the smaller of the tail probabilities to the right of (and including) w^+ and to the left of (and including) w^+)	$w^+ \leq w^*_{\alpha/2,n}$ or $w^+ \geq w_{\alpha/2,n}$

Notes

* For a given sample size (after deleting the zero-valued d_i's) n, in Table B6, the p-value can be taken to be less than the smallest α for which H_0 would be rejected using the rejection region approach.

** For a given level of significance α and sample size (after deleting the zero-valued d_i's) n, in Table B6 the upper tail critical value $w_{\alpha,n}$ is the *large* W entry associated with row n, column α. The lower tail critical value $w^*_{\alpha,n}$ is the *small* W entry

Notes

Exercise

A method for measuring ground-level atmospheric mercury (Hg) requires holding air samples for up to 120 h (five days) before analyzing them at a laboratory.

A quality assurance study was carried out to ensure that the long holding time wouldn't affect the measurement results.

Notes

Air sampling devices were **placed in pairs** in the field.

For each pair of sampled air specimens, one was held for **4 hours** and the other for **120 hours** before being analyzed in the lab.

The table on the next slide shows particulate-bound **Hg measurements** (pg/m^3) for each of the **10 pairs** along with their **differences**.

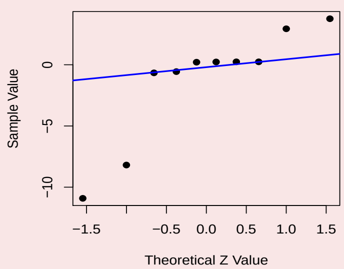
A **normal probability plot** of the **differences** is two slides ahead.

Notes

	Particulate-Bound Hg		
Air Sample	Long	Short	
Pair	Holding Time	Holding Time	Difference
1	4.27	1.33	2.94
2	2.77	3.43	-0.66
3	1.50	1.29	0.21
4	5.70	6.26	-0.56
5	3.80	11.99	-8.19
6	5.64	1.88	3.76
7	3.67	14.58	-10.91
8	0.78	0.54	0.24
9	3.92	3.69	0.23
10	1.85	1.61	0.24

Notes

Normal Probability Plot of Differences



Notes

The ”backward S” shape of the normal probability plot suggests that the normality assumption required for the *paired t test* **isn’t** met.

Instead, we’ll carry out a **signed rank test**.

Notes

Carry out the **signed rank test** to decide if the long holding period has **any effect** on Hg measurements. Use $\alpha = 0.05$.

Hint: The **combined**, **sorted**, **absolute values** of the **differences** are below.

Differences that were **positive** before taking absolute values are denoted by $+$ and ones that were **negative** by $-$.

Sign	$+$	$+$	$+$	$+$	$-$	$-$	$+$	$+$	$-$	$-$
Obs.	0.21	0.23	0.24	0.24	0.56	0.66	2.94	3.76	8.19	10.91
Rank										

You should get $\boldsymbol{W^+ = 25}$ and **p-value** $= 2(0.423) = \mathbf{0.846}$.

Notes

Paired Samples Sign Test

- The ***paired samples sign test***, like the *signed rank test*, is a **paired samples nonparametric** test for the difference between two population **centers**.

  We'll act as though the **differences** $D_1, D_2, \ldots, D_n$ are a random sample from a **population** of **differences** whose ***median*** is $\tilde{\mu}_d$.

- The ***paired samples sign test*** is just a **one-sample sign test** for $\tilde{\mu}_d$ based on the **differences**.

Notes

- The **null hypothesis** is that $\tilde{\mu}_d$ is **zero**.

Null Hypothesis:

$$H_0 : \tilde{\mu}_d \ = \ 0.$$

(This says a typical difference is zero.)

Notes

- The **alternative hypothesis** is one of the following.

Alternative Hypothesis:

1. $\tilde{\mu}_d > 0$ (**upper-tailed test**)
2. $\tilde{\mu}_d < 0$ (**lower-tailed test**)
3. $\tilde{\mu}_d \neq 0$ (**two-tailed test**)

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

Notes

Paired Samples Sign Test Statistic:

$$S^+ \ = \ \text{Number of } D_i\text{'s that are greater than 0.}$$

(If any D_i's equal 0, they're discarded, and n is diminished by the number of discarded D_i's).

This is just the **one-sample sign test statistic** using the sample of **differences** and a null-hypothesized value **zero**.

Notes

1. *Large* values of S^+ (larger than $n/2$) provide evidence in favor of $H_a : \tilde{\mu}_d > 0$.
2. *Small* values of S^+ (smaller than $n/2$) provide evidence in favor of $H_a : \tilde{\mu}_d < 0$.
3. *Both large and small* values of S^+ (larger or smaller than $n/2$) provide evidence in favor of $H_a : \tilde{\mu}_d \neq 0$.

Notes

Paired Samples Sign Test for $\tilde{\mu}_d$

Assumptions: $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are two random samples that are paired, and the differences $d_1, d_2, \ldots, d_n$ form a single sample from *any continuous* population.

Null hypothesis: $H_0 : \tilde{\mu}_d = 0$.

Test statistic value: $s^+ =$ number of positive d_i's.

Decision rule: Reject H_0 if p-value $< \alpha$ or s^+ is in rejection region.

Notes

Paired Samples Sign Test for $\tilde{\mu}_d$

Null hypothesis: $H_0 : \tilde{\mu}_d = 0$.

Test statistic value: $s^+ =$ number of positive d_i's.

Decision rule: Reject H_0 if p-value $< \alpha$ or s^+ is in rejection region.

Alternative hypothesis	P-value = tail probability of the binomial$(n, 0.5)$ distribution: *	Rejection region = s^+ values such that: **
$H_a : \tilde{\mu}_d > 0$	to the right of (and including) s^+	$s^+ \geq s_{\alpha,n}$
$H_a : \tilde{\mu}_d < 0$	to the left of (and including) s^+	$s^+ \leq s^*_{\alpha,n}$
$H_a : \tilde{\mu}_d \neq 0$	$2\cdot$(the smaller of the tail probabilities to the right of (and including) s^+ and to the left of (and including) s^+)	$s^+ \leq s^*_{\alpha/2,n}$ or $s^+ \geq s_{\alpha/2,n}$

Notes

Paired Samples Sign Test for $\tilde{\mu}_d$

* For a given sample size (after deleting the zero-valued d_i's) n, the p-value for a one-tailed test is obtained from a binomial$(n, 0.5)$ distribution table by locating the upper or lower tail probability (depending on the direction of H_a) associated with the observed S^+ value. For a two-tailed test, locate both the upper and lower tail probabilities and multiply the smaller of these by two.

** For a given sample size (after deleting zero-valued d_i's) n and level of significance α, $s_{\alpha,n}$ is obtained from a binomial$(n, 0.5)$ distribution table by locating the smallest s for which the upper tail probability is less than α. $s^*_{\alpha,n}$ is obtained by locating the largest s for which the lower tail probability is less than α. For the two-tailed test, $s_{\alpha/2,n}$ and $s^*_{\alpha/2,n}$ are defined analogously but with $\alpha/2$ used in place of α. In practice, due to the discreteness of the distribution, it's not always possible obtain a rejection region having exact probability α.

Notes

Exercise

Consider again the quality assurance study to ensure that a long holding time wouldn't affect mercury (Hg) measurements in air samples.

The table below shows the data again.

Notes

Air Sample Pair	Particulate-Bound Hg: Long Holding Time	Particulate-Bound Hg: Short Holding Time	Difference
1	4.27	1.33	2.94
2	2.77	3.43	-0.66
3	1.50	1.29	0.21
4	5.70	6.26	-0.56
5	3.80	11.99	-8.19
6	5.64	1.88	3.76
7	3.67	14.58	-10.91
8	0.78	0.54	0.24
9	3.92	3.69	0.23
10	1.85	1.61	0.24

Notes

Recall that the normality assumption for these data was questionable, so a *paired t test* wasn't appropriate.

Carry out a **sign test for paired samples** to decide if the long holding period has **any effect** on Hg measurements. Use $\alpha = 0.05$.

Hints: You should get $\boldsymbol{S^+ = 6}$ and **p-value** $= 2(0.3770) = \mathbf{0.7540}$.

Notes

Notes