8 Two-Sample Hypothesis Tests (Cont'd)

MTH 3240 Environmental Statistics

Spring 2020

Notes

Objectives

Objectives:

- Carry out a rank sum test for the difference between two population means.
- Decide which test (the two-sample t test or the rank sum test) is more appropriate for a given set of data.

Notes

Dealing With Non-Normal Data

- The two-sample t procedures require that the samples were drawn from **normal** populations (or that $\boldsymbol{n_x}$ and $\boldsymbol{n_y}$ are **large**).

  If this **normality** assumption isn't met (and n_x and n_y aren'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 hypothesis with **non-normal** data is to ***transform*** the data (**both** samples) to normality first.

Notes

Carrying Out a Nonparametric Test

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

  The ***rank sum test*** (described next) is a **nonparametric** alternative to the *two-sample t test*.

Notes

Rank Sum Test

- The ***rank sum test*** is a **nonparametric** test for the difference between two population **means** μ_x and μ_y.
- The **null hypothesis** is that there's **no difference** between μ_x and μ_y.

Null Hypothesis:

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

(Same hypothesis as for the two-sample t test.)

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 hypotheses as for the two-sample t test.)

Notes

- The ***test statistic***, denoted W_{rs}, is obtained by combining the two samples, ***ranking*** the observations (smallest to largest), and then **summing** the **ranks** of the *smaller* sample.

Rank Sum Test Statistic:

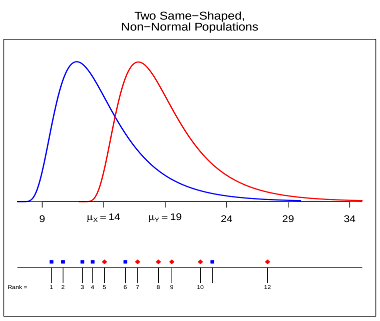
1. Label the observations in the sample whose size is **smaller** by $X_1, X_2, \ldots, X_{n_x}$ and those in the sample whose size is **larger** by $Y_1, Y_2, \ldots, Y_{n_y}$ (i.e $n_x \leq n_y$). If the sample sizes are the same, either sample may be labeled as X_i's and the other as Y_i's.

Notes

2. **Combine** the two samples, **sort** the observations, 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.
3. **Sum** the **ranks** of the X_i's. This gives the **test statistic**:

$$W_{rs} \;=\; \text{Sum of ranks of } X_i\text{'s.}$$

Notes



Notes

- $\boldsymbol{W_{rs}}$ reflects whether the two samples are **evenly intermingled** or **segregated** (when combined and sorted).
  - **If $\boldsymbol{H_0}$ was true**, ...

    ... we'd expect the two samples to be **intermingled**.
  - But **if $\boldsymbol{H_a}$ was true**, ...

    ... we'd expect the samples to be **segregated**, and the X_i's to mostly lie near the end in the direction specified by H_a.

Notes

- It can be shown that ...
  1. $\boldsymbol{W_{rs}}$ will be approximately **equal to $\boldsymbol{n_x(N+1)/2}$** (most likely), where $\boldsymbol{N} = n_x + n_y$ denotes the overall (combined) sample size, if H_0 is true.
  2. $\boldsymbol{W_{rs}}$ will **differ from $\boldsymbol{n_x(N+1)/2}$** (most likely) in the direction specified by H_a if H_a is true.

Notes

1. *Large* values of W_{rs} (larger than $n_x(N+1)/2$) provide evidence in favor of $H_a : \mu_x - \mu_y > 0$.
2. *Small* values of W_{rs} (smaller than $n_x(N+1)/2$) provide evidence in favor of $H_a : \mu_x - \mu_y < 0$.
3. Both *large and small* values of W_{rs} (larger or smaller than $n_x(N+1)/2$) provide evidence in favor of $H_a : \mu_x - \mu_y \neq 0$.

Notes

- Now suppose the samples are from **any** two (continuous) populations whose means are μ_x and μ_y.

  If the populations have (roughly) the same shape, the **null distribution** is as follows.

Notes

Sampling Distribution of W_{rs} Under H_0: If W_{rs} is the rank sum test statistic, then when

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

is true, W_{rs} follows a distribution called the ***Wilcoxon rank sum distribution***, which will depend on n_x and n_y. We write this as

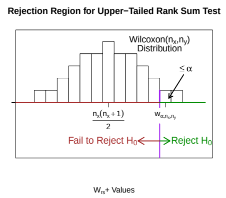
$$W_{rs} \sim \text{Wilcoxon}(n_x, n_y).$$

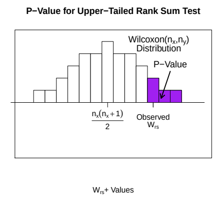
Notes

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

Notes

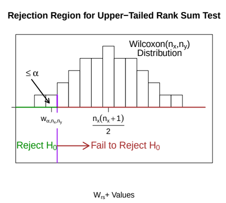
1. $H_a : \mu_x - \mu_y > 0$ (Upper-Tailed Test)

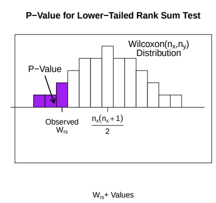




Notes

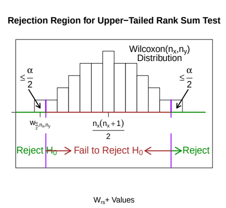
2. $H_a : \mu_x - \mu_y < 0$ (Lower-Tailed Test)

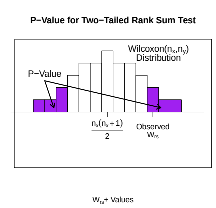




Notes

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





Notes

Rank Sum Test for μ_x and μ_y

Assumptions: The data $x_1, x_2, \ldots, x_{n_x}$ and $y_1, y_2, \ldots, y_{n_y}$, with $n_x \leq n_y$, are independent random samples from two continuous populations whose distributions differ, if at all, by their means μ_x and μ_y but not their shapes.

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

Test statistic value: w_{rs} = Sum of the ranks of x's in the combined, sorted set of x's and y's.

Decision rule: Reject H_0 if p-value $< \alpha$ or w_{rs} is in rejection region.

Notes

Rank Sum Test for μ_x and μ_y

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

* Tail probabilities of the W_{rs} distribution under H_0, for given sample sizes n_x and n_y, are found in the Wilcoxon rank sum distribution table under the column p for different observed values of W_{rs} (denoted w in the right tail and w' in the left tail).

Notes

Rank Sum Test for μ_x and μ_y

** For a given level of significance α and sample sizes n_x and n_y, the upper tail critical value w_{α,n_x,n_y} is obtained from the Wilcoxon rank sum distribution table by locating the smallest w for which the tail area p to its right is less than α. Likewise, w'_{α,n_x,n_y} is obtained by locating the largest w' for which the tail area p to its left less than α.

Notes

Exercise

In a study of the impact of an inactive gold mine on water quality in the North Fork Humboldt River, several variables were measured **upstream** and **downstream** of the former mine.

The table below shows the **arsenic** (As) concentrations (μg/L) made upstream on $n_x = 6$ days and downstream on $n_y = 7$ days.

Notes

As in Water

Upstream	Downstream
5	10
4	9
6	8
10	7
12	10
9	16
	10

Carry out a **rank sum test** to decide if the true mean As concentration is **higher downstream than upstream**. Use $\alpha = 0.05$.

Notes

Hints: The smaller (**upstream**) sample is denoted X and the larger (**downstream**) sample Y.

The **combined**, **sorted** data are below.

Sample	X	X	X	Y	Y	X	Y	X	Y	Y	Y	X	Y
Obs.	4	5	6	7	8	9	9	10	10	10	10	12	16
Rank													

You should get $W_{rs} = 34$ and **p-value** > 0.117.

Notes

Exercise

Another variable measured in the study of the impact of the inactive gold mine on water quality was sodium (Na).

The table below shows the Na concentrations (in mg/L) made **upstream** on $n_x = 6$ days and **downstream** on $n_y = 7$ days.

Notes

Na in Water

Upstream	Downstream
2.17	4.23
2.22	4.43
0.85	3.80
2.03	3.80
1.54	3.66
2.05	2.58
	3.48

Carry out a **rank sum test** to decide if the true mean Na concentration is **higher downstream than upstream**. Use $\alpha = 0.05$.

Notes

Hints: The smaller (**upstream**) sample is denoted X and the larger (**downstream**) sample Y.

The **combined**, **sorted** data are below.

Sample	X	X	X	X	X	X	Y	Y	Y	Y	Y	Y	Y
Obs.	0.85	1.54	2.03	2.05	2.17	2.22	2.58	3.48	3.66	3.80	3.80	4.23	4.43
Rank													

You should get $W_{rs} = 21$ and **p-value** < 0.007.

Notes