Week 12: Nonparametric testing

Phase IV

Key ideas

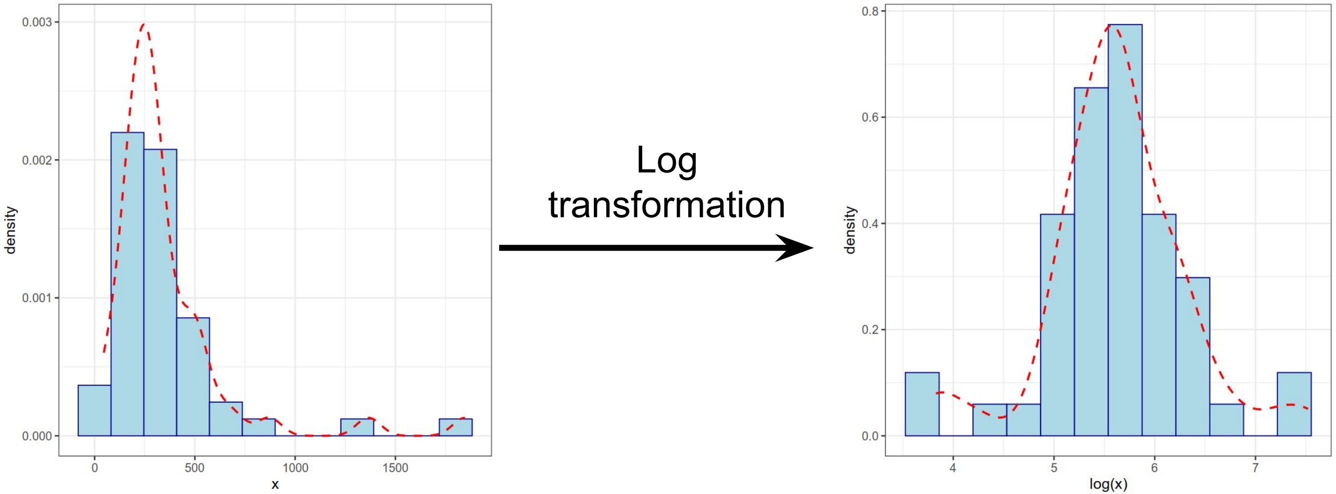
- Statistical tests studied so far are called **parametric** because they rely on an underlying probability distribution.
- These tests also **depend** on certain **assumptions** (e.g. Normality, lack of outliers, etc).
- **Nonparametric** tests come to our rescue when **assumptions** are **not met**, since they are **distribution-free** methods.
- Most of them are based on **ranks** instead of the actual observed data.

What to do when normality assumption fails?

- If lack of Normality is due to **outliers**, it may be legitimate to **remove** the outliers.

What to do when normality assumption fails?

- ➢ If lack of Normality is due to **outliers**, it may be legitimate to **remove** the outliers.
- ➢ Sometimes we can **transform** our **data** so that their distribution is more nearly normal (not covered in this course).



What to do when normality assumption fails?

- If lack of Normality is due to **outliers**, it may be legitimate to **remove** the outliers.
- Sometimes we can **transform** our **data** so that their distribution is more nearly normal (not covered in this course).
- Maybe consider a **different distribution** (e.g.g. binomial, Poisson, etc).

What to do when normality assumption fails?

- If lack of Normality is due to **outliers**, it may be legitimate to **remove** the outliers.
- Sometimes we can **transform** our **data** so that their distribution is more nearly normal (not covered in this course).
- Maybe consider a **different distribution** (e.g. binomial, Poisson, etc).
- **Resampling methods** like bootstrapping or permutation tests (Not covered in this course).

What to do when normality assumption fails?

- If lack of Normality is due to **outliers**, it may be legitimate to **remove** the outliers.
- Sometimes we can **transform** our **data** so that their distribution is more nearly normal (not covered in this course).
- Maybe consider a **different distribution** (e.g. binomial, Poisson, etc).
- **Resampling methods** like bootstrapping or permutation tests (Not covered in this course).
- **Nonparametric tests** that do not require any specific form for the distribution of the population.

When to use nonparametric tests?

- ➢ You have very small sample size.
- ➢ Your area of study is better represented by the median.
- ➢ You have ordinal data, ranked data, or outliers that you can’t remove.

Ranks

➢ The rank of an observation in the column of dataset is its position when all observations are sorted, starting with rank 1 for the smallest observation.

Example:

(21, 18, 24, 18, 19, 7, 11, 13)

↓ Sort values

group a
group b

(7, 11, 13, 18, 19, 20, 21, 24)

↓ Ranking

(1, 2, 3, 4, 5, 6, 7, 8)

Ranks

- ➢ The rank of an observation in the column of dataset is its position when all observations are sorted, starting with rank 1 for the smallest observation.
- ➢ What if there are **ties**? →The average of the ranks they occupy.

(21, 18, 24, 19, 19, 7, 11, 13)

↓ Sort values

group a
group b

(7, 11, 13, 19, 19, 20, 21, 24)

↓ Ranking

(1, 2, 3, 4.5, 4.5, 6, 7, 8)

The Wilcoxon rank-sum test

- It is used when dealing with one **quantitative** variable (i.e. one sample), or one **quantitative** and one **binary** variable (i.e two samples).
- It is usually the **nonparametric** alternative to the **two sample t-test**.
- It **tests** whether two populations have **the same distribution** or not.
- Under special cases (both populations having the same distribution shape), it can be reformulated as a test on the **medians**.
- In the case of **two samples**, it can also be called the **Mann–Whitney U test** in literature.

The Wilcoxon rank-sum test

- Null Hypothesis $\mathbf{H_0}$: the two populations have the same distribution.

- Test statistic: $W = R_1 - \frac{N_1 \cdot (N_1 + 1)}{2}$

- Alternative Hypothesis $\mathbf{H_A}$:

  One distribution has systematically smaller or larger values (one-sided)

  One distribution has systematically different values (two-sided)

The Wilcoxon rank-sum test (same shape assumption)

- Null Hypothesis $\mathbf{H_0}$: the two populations have the same **medians**.

- Test statistic: $W = R_1 - \frac{N_1 \cdot (N_1 + 1)}{2}$

In this case it can be expressed as a test on the medians!

- Alternative Hypothesis $\mathbf{H_A}$:

  One distribution has a systematically smaller or larger **median** (one-sided)

  One distribution has systematically different **medians** (two-sided)

The Wilcoxon rank-sum test

- Null Hypothesis $\mathbf{H_0}$: the two populations have the same distribution.

- Test statistic: $W = R_1 - \frac{N_1 \cdot (N_1 + 1)}{2}$

wilcox.test(X_1, X_2 mu = Δ_0, alternative=”smaller”)

- Alternative Hypothesis $\mathbf{H_A}$:

  One distribution has systematically smaller or larger values (one-sided)

  One distribution has systematically different values (two-sided)

The Wilcoxon rank-sum test

- Null Hypothesis $\mathbf{H_0}$: the two populations have the same distribution.

- Test statistic: $W = R_1 - \frac{N_1 \cdot (N_1 + 1)}{2}$

wilcox.test(X_1, X_2 mu = Δ_0, alternative="greater")

- Alternative Hypothesis $\mathbf{H_A}$:

  One distribution has systematically smaller or larger values (one-sided)

  One distribution has systematically different values (two-sided)

The Wilcoxon rank-sum test

- Null Hypothesis $\mathbf{H_0}$: the two populations have the same distribution.

- Test statistic: $W = R_1 - \frac{N_1 \cdot (N_1 + 1)}{2}$

- Alternative Hypothesis $\mathbf{H_A}$:

  wilcox.test(X_1, X_2 mu = Δ_0, alternative="two.sided")

  One distribution has systematically smaller or larger values (one-sided)

  One distribution has systematically different values (two-sided)

Example

In one study, researchers are interested in comparing the levels of anxiety between two groups of participants: Group A, which consists of 5 individuals diagnosed with an anxiety disorder, and Group B, which consists of 5 control individuals. Anxiety scores are measured on a Likert-type scale ranging from 1 to 10, where higher scores indicate higher levels of anxiety.

<u>Research question:</u> do anxiety scores differ between both groups?

Group A (Anxiety Disorder)	Group B (Control)
7	2
9	3
4	1
8	4
5	2

Example

Research question: do anxiety scores differ between both groups?

Data: (7, 9, 4, 8, 5, 2, 3, 1, 4, 2)

Sorted data: (1, 2, 2, 3, 4, 4, 5, 7, 8, 9)

Ranks : (1, 2.5, 2.5, 4, 5.5, 5.5, 7, 8, 9, 10)

$\sum \text{Rank}_i = 5.5, 7, 8, 9, 10 = 39.5$

$N_1 * (N_1 + 1)/2 = 5 \times 6/2 = 15$

$W = 39.5 - 15 = 24.5 \rightarrow p = 0.016$

Group A (Anxiety Disorder)	Group B (Control)
7	2
9	3
4	1
8	4
5	2

At α = 0.05 we should **reject** the null that both distributions are equal.

Paired data: The Wilcoxon signed-rank test

- It is usually the **nonparametric** alternative to the **paired sample t-test**.
- It is a test about the symmetry of the data: it tests the **median** against some null value (it can be one-sided or two-sided differences).
- It is based on the sum of the ranks of the **positive** (or **negative**) differences when we rank the **absolute** values of the differences.

The Wilcoxon signed-rank test

- Null Hypothesis $\mathbf{H_0}$: the population's median is equal to some value.

- Test statistic: $V = R_+$

- Alternative Hypothesis $\mathbf{H_A}$:

  The population's median is smaller or larger values than a given value (one-sided).

  The population's median is different from a given one (two-sided).

The Wilcoxon signed-rank test

- Null Hypothesis $\mathbf{H_0}$: the population's median is equal to some value.

- Test statistic: $V = R_+$

wilcox.test(X_1, X_2 mu = Δ_0,
alternative="smaller", **paired=TRUE**)

- Alternative Hypothesis $\mathbf{H_A}$:

  The population's median is smaller or larger values than a given value (one-sided).

  The population's median is different from a given one (two-sided).

The Wilcoxon signed-rank test

- Null Hypothesis $\mathbf{H_0}$: the population's median is equal to some value.

- Test statistic: $V = R_+$

wilcox.test(X_1, X_2 mu = Δ_0,
alternative="greater", **paired=TRUE**)

- Alternative Hypothesis $\mathbf{H_A}$:

  The population's median is smaller or larger values than a given value (one-sided).

  The population's median is different from a given one (two-sided).

The Wilcoxon signed-rank test

- Null Hypothesis $\mathbf{H_0}$: the population's median is equal to some value.

- Test statistic: $V = R_+$

wilcox.test(X_1, X_2 mu = Δ_0, alternative="two.sided", **paired=TRUE**)

- Alternative Hypothesis $\mathbf{H_A}$:

  The population's median is smaller or larger values than a given value (one-sided).

  The population's median is different from a given one (two-sided).

Example

In one study, researchers are interested in comparing the levels of anxiety in a group of 5 participants before therapy (Group A) and after therapy (Group B). Anxiety scores are measured on a Likert-type scale ranging from 1 to 10, where higher scores indicate higher levels of anxiety.

<u>Research question:</u> Are anxiety scores before therapy greater than scores after therapy?

Group A (before therapy)	Group B (after therapy)
7	2
9	3
4	1
8	4
5	2

Example

Are anxiety scores before therapy greater than scores after therapy?

Data: (7, 9, 4, 8, 5, 2, 3, 1, 4, 2)

Absolute differences: (5, 6, 3, 4, 3)

Sorted data: (3, 3, 4, 5, 6)

Ranks : (1.5, 1.5, 3, 4, 5)

Signs: (+, + ,+, + , +)

$V = \sum Rank_+ = 15 \rightarrow p = 0.029$

Group A (before therapy)	Group B (after therapy)
7	2
9	3
4	1
8	4
5	2

At α = 0.05 **reject** the null: therapy does not reduce anxiety levels.

The Kruskal-Wallis test

- It is used when dealing with **one quantitative** and **one categorical** variable of **two or more categories**.

- It is usually the **nonparametric** alternative to **ANOVA**.

- <u>Intuition</u>: The Kruskal-Wallis test ranks all the observed values from all groups together and compare the sum of the ranks for each group. Under the null hypothesis, we would expect these sum of ranks to be more or less equal.

The Kruskal-Wallis test

- Null Hypothesis $\mathbf{H_0}$: We have the same distribution in all populations.

- Test statistic (no ties): $H = \frac{12}{N \cdot (N+1)} \sum_{i=1}^{k} \frac{R_i^2}{N_i} - 3 \cdot (N+1)$

- Alternative Hypothesis $\mathbf{H_A}$:

  At least one distribution has systematically different values than others (two-sided)

The Kruskal-Wallis test

- Null Hypothesis $\mathbf{H_0}$: We have the same distribution in all populations.
- Test statistic (no ties): $H = \frac{12}{N \cdot (N+1)} \sum_{i=1}^{k} \frac{R_i^2}{N_i} - 3 \cdot (N+1)$ ~ χ^2(df=k-1)
- Alternative Hypothesis $\mathbf{H_A}$:

  At least one distribution has systematically different values than others (two-sided)

The Kruskal-Wallis test

- Null Hypothesis $\mathbf{H_0}$: We have the same distribution in all populations.

- Test statistic (no ties): $H = \frac{12}{N \cdot (N+1)} \sum_{i=1}^{k} \frac{R_i^2}{N_i} - 3 \cdot (N+1)$

kruskal.test(formula, data)

- Alternative Hypothesis $\mathbf{H_A}$:

  At least one distribution has systematically different values than others (two-sided)

Kruskal-Wallis rejects the null: Post-hoc testing

One very simple recipe:

First, **pairwise** Wilcoxon rank-sum test; then, **Bonferroni** procedure that keeps Type I error under control (We'll come back to this in the future)**.**

```
pairwise.wilcox.test(x=Our continuous variable,
                     g=Our categorical variable,
                     p.adjust.method="bonf")
```

Kruskal-Wallis rejects the null: Post-hoc testing

<u>One very simple recipe:</u>

First, **pairwise** Wilcoxon rank-sum test; then, **Bonferroni** procedure that keeps Type I error under control (We'll come back to this in the future)**.**

Important: Some people claim against the Wilcoxon rank-sum test for post-hoc analysis and recommend a **Dunn's test** instead. This test is not found natively in R, so it is beyond the scope of this course. But in the future you may want to take this into account if using the Kruskal-Wallis test…

Example

A group of psychologists is interested in the effect of different types of therapy on reducing symptoms of anxiety. 18 participants were recruited and randomly assigned to one of three groups: Cognitive Behavioral Therapy (CBT), Mindfulness-Based Stress Reduction (MBSR), and Supportive Therapy (ST). Each participant completed a standardized anxiety scale before and after therapy, and the difference scores were calculated (post-test score minus pre-test score) to measure the change in anxiety symptoms.

CBT	MBSR	ST
6	5	4
5	8	2
3	9	5
7	7	6
2	9	1
4	6	4

<u>Research question</u>: do distribution in anxiety levels differ across groups?

Example

Research question: do distribution in anxiety levels differ across groups?

Data = (6, 5, 3, 7, 2, 4, 5, 8, 9, 7, 9, 6, 4, 2, 5, 6, 1, 3)

Sorted = (1, 2, 2, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9, 9)

Ranks = (1, 2.5, 2.5, 4.5, 4.5, 6.5, 6.5, 9, 9, 9, 12, 12, 12, 14.5, 14.5, 16, 17.5, 17.5)

$R_1 = 49$, $R_2 = 86.5$, $R_3 = 35.5$ → $H \cong 8.27$ (after correction) → $p \cong 0.016$

CBT	MBSR	ST
6	5	4
5	8	2
3	9	5
7	7	6
2	9	1
4	6	4

At $\alpha = 0.05$ we should **reject** the null: distribution in anxiety differ across groups.

The Spearman's correlation test

- It is used when testing the **monotonic** relationship (i.e. a trend not necessarily at a constant rate) between **two quantitative** variables.
- It is one of the **nonparametric** alternatives to the **Pearson's correlation**.
- It is based on computing the **Pearson's correlation on the ranks** of both X and Y, $r_{rankx,\ ranky}$, instead of using the observed values.
- It is appropriate when **relationship** are **nonlinear**, when **data** are **ordinal**, or in the presence of **outliers**.
- <u>Statistical inference</u>: test for the Pearson's correlation with $r_{rankx,\ ranky}$.

Example

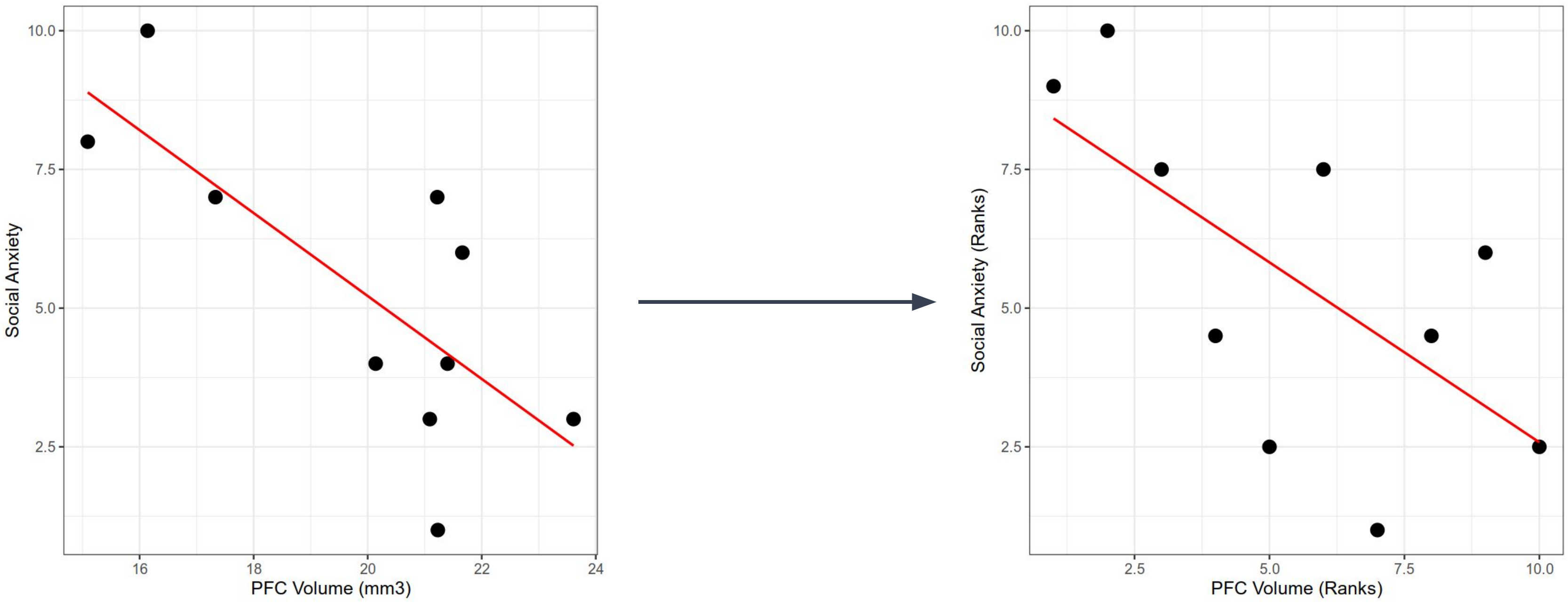
In order to understand the relationship between the volume of grey matter in the prefrontal cortex (PFC) of the brain and social anxiety, a study recollected neuroimaging data from 10 subjects, who were also administered with a likert test that measures social anxiety. The test asked encoded in a scale from 1 to 10 statements such as "I feel anxious in social situations".

Research question: Is social anxiety associated with grey matter volume?

PFC Volume	Social Anxiety test
16.14	10
21.22	7
21.09	3
21.23	1
23.61	3
21.4	4
15.09	8
17.33	7
21.66	6
20.04	4

Example

A correlation between ranks:



Example

<u>Research question:</u> Is social anxiety associated with grey matter volume?

$X = (16.14, 21.22, 21.09, 21.23, 23.61, 21.4, 15.09, 17.33, 21.66, 20.14)$

$rank_X = (2, 6, 5, 7, 10, 8, 1, 3, 9, 4)$

$Y = (10, 7, 3, 1, 3, 4, 8, 7, 6, 4)$

$rank_Y = (10, 7.5, 2.5, 1, 2.5, 4.5, 9, 7.5, 6, 4.5)$

$r_{rankx,\ ranky} = -0.654 \rightarrow p = ??$

PFC Volume	Social Anxiety test
16.14	10
21.22	7
21.09	3
21.23	1
23.61	3
21.4	4
15.09	8
17.33	7
21.66	6
20.04	4

Recap

- **Nonparametric** testing comes to our rescue when **assumptions** are **not met**, since they are **distribution-free** methods.
- Most of these tests are based on **ranks** instead of the actual observed data.

Parametric test		Nonparametric alternative
Two sample t-test	⟷	Wilcoxon rank-sum test
Paired t-test	⟷	Wilcoxon rank-signed test
ANOVA test	⟷	Kruskal-Wallis test
Pearson's correlation test	⟷	Spearman's correlation test