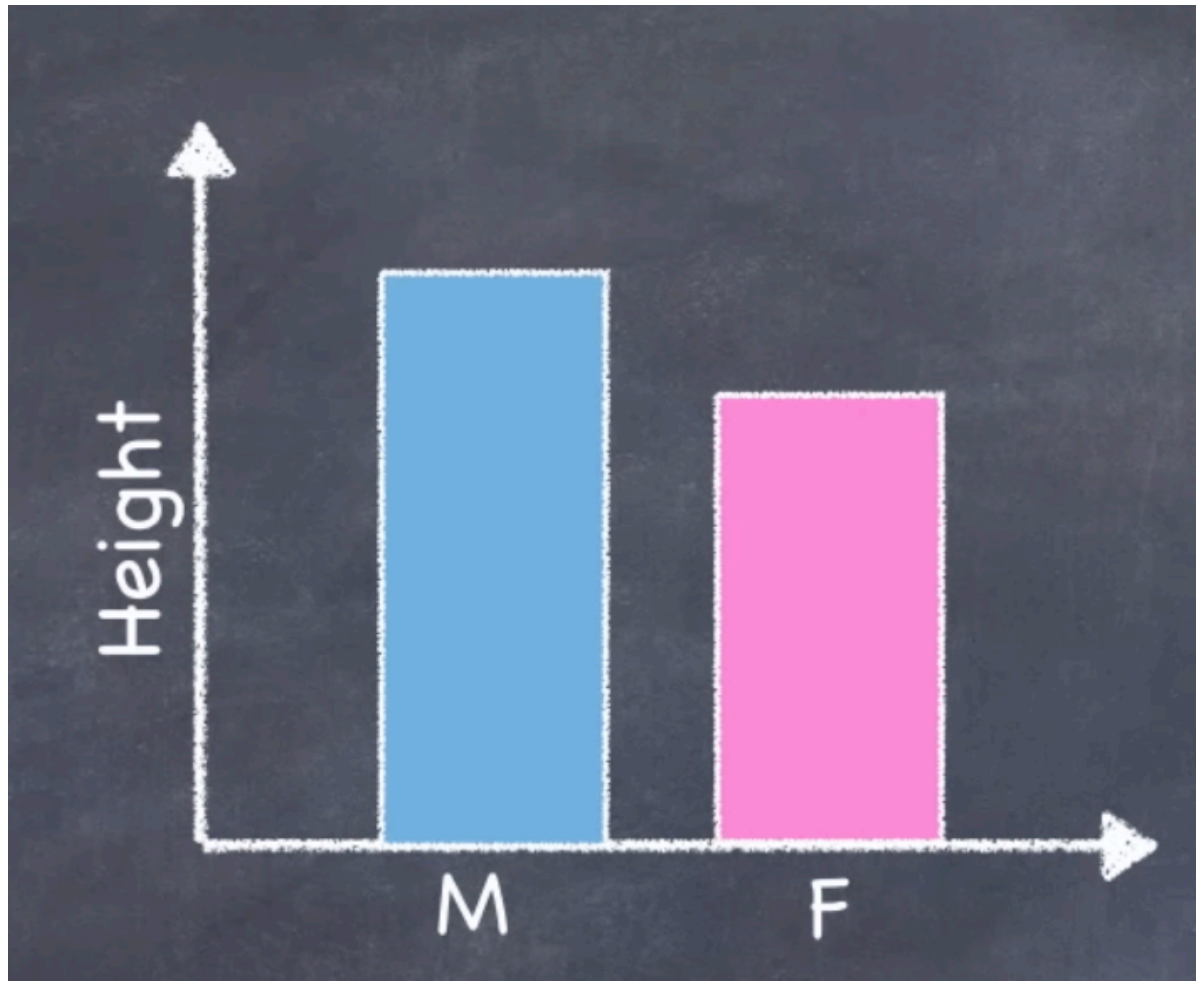
Statistical tests II

Rachel Warnock

05.04.2023

Yesterday's adventures



compare the means of 2 groups → z-test

→ t-test

compare 2+ groups → ANOVA

Today's adventures

post hoc ANOVA test → Tukey-Kramer

Steel Dwass

equivalent to a one-sample t-test, non-parametric → Wilcoxon test

equivalent to ANOVA, non-parametric → Kruskal-Wallis

Dependent vs. Independent samples

Say I want to know, does the Hypothesis Testing course impact stress levels in students? 😵‍💫

I have a questionnaire to measure stress levels **before** and **during** the course.

Two options:

1. Give the questionnaire to some students before the course & to some other students after the course.
2. Give the questionnaire to the **same students** before & after the course.

Dependent vs. Independent samples

Two options:

1. Give the questionnaire to some students before the course & to some other students after the course. This is an **independent sample**.

2. Give the questionnaire to the **same students** before & after the course. This is a **dependent sample** (measures are available in pairs, better option).

Parametric vs. Non-parametric

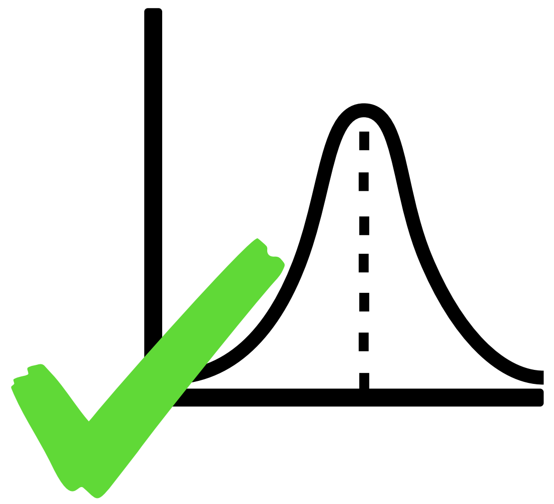
For any hypothesis, you must check the assumptions of the test.

A very common assumption is that the data must have a certain underlying distribution, usually the normal distribution.

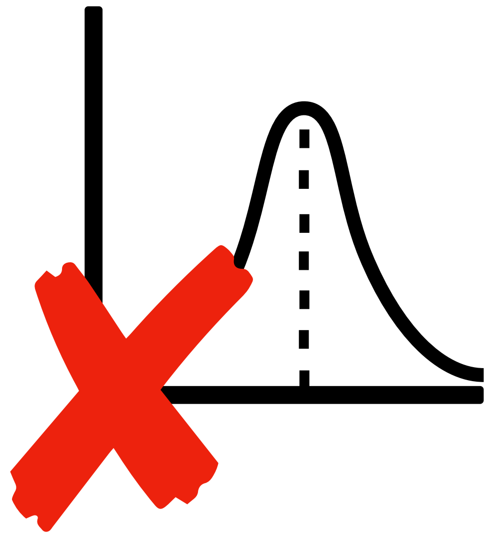
Parametric statistics are based on assumptions about the distribution of population.

In **non-parametric statistics** do not make this assumption — the data can be collected from a sample that does not follow a specific distribution.

Parametric vs. Non-parametric



If your data are normally distributed you can use a **parametric test**, e.g., t-test, ANOVA.



If your data are not normally distributed you use a **non-parametric test**, e.g., Wilcoxon test, Kruskal-Wallis.

Parametric vs. Non-parametric

We have to check the other assumptions, but in general there are less assumptions associated with non-parametric tests.

But parametric tests are usually more powerful — a smaller difference in values, or a smaller difference in sample sizes is required to reject the null hypothesis.

If possible - use parametric tests!

For all parametric tests there is (usually) a parametric counterpart

	Parametric Tests	Nonparametric Tests
One Sample	Simple t-Test	Wilcoxon test for one sample
Two dependent samples	Paired Sample t-Test	Wilcoxon Test
Two inpendent samples	Unpired Sample t-Test	Mann-Whitney U Test
More than two independent samples	One factorial ANOVA	Kruskal-Wallis Test
More than two dependent samples	Repeated Measures ANOVA	Friedman Test
Correlation between two variables	Pearson-Korrelation	Spearman-Korrelation

DATAtab

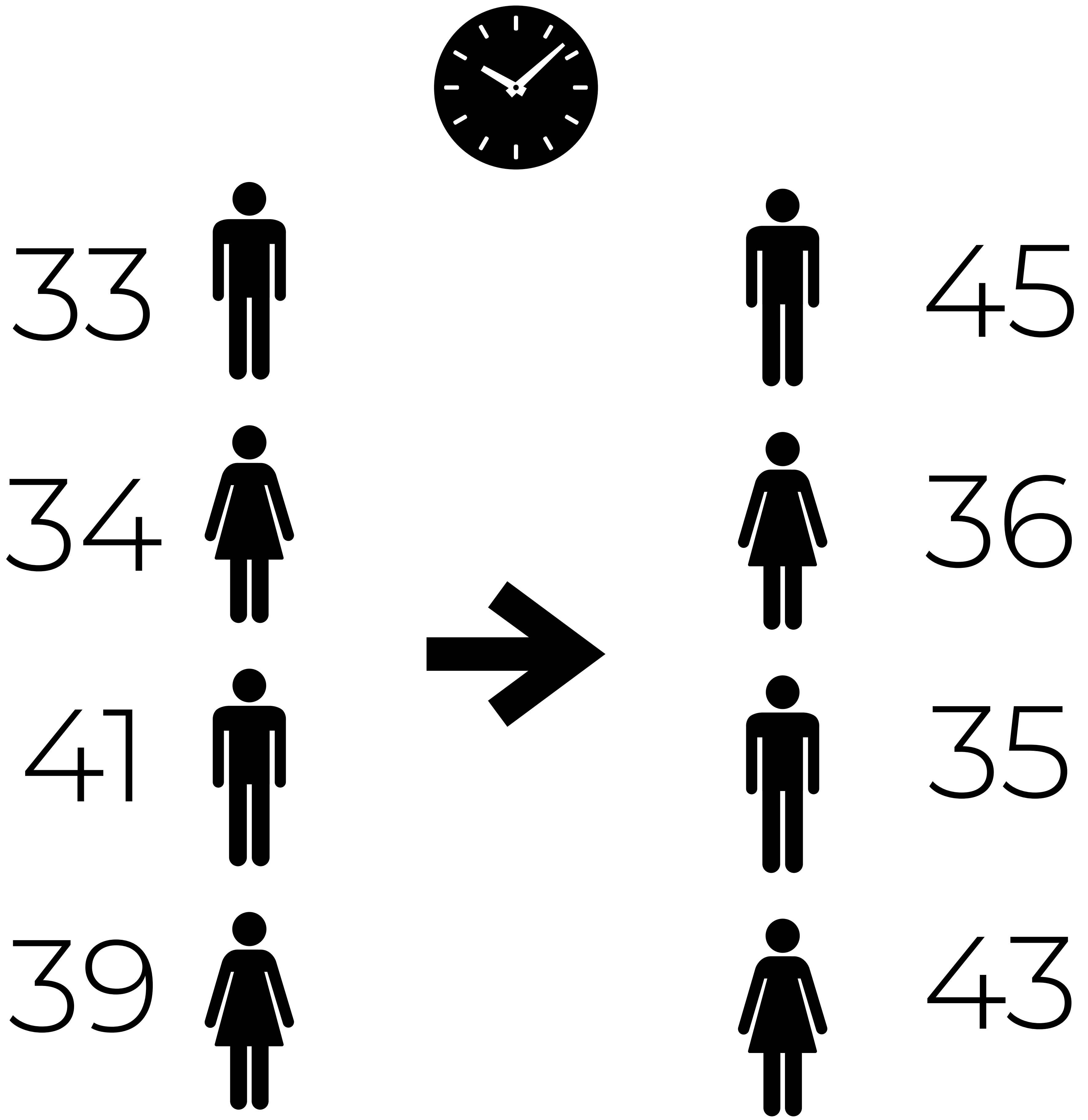
The Wilcoxon test

The non-parametric version of the t-test.

Asks, is there a difference between two **dependent** samples?

Data is **not normally distributed** and the data is **available in pairs**.

The Wilcoxon test

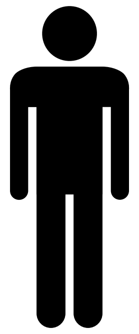
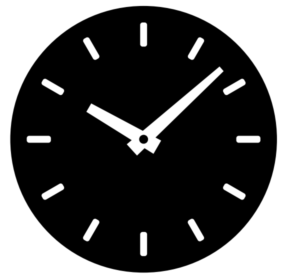


t-test: is there a diff. in the **mean**? i.e., significantly diff. from 0?

Wilcoxon test: is there a diff. in the **ranks**?

= 12	4	(+)
= 2	1	(+)
= -6	3	(-)
= 4	2	(+)

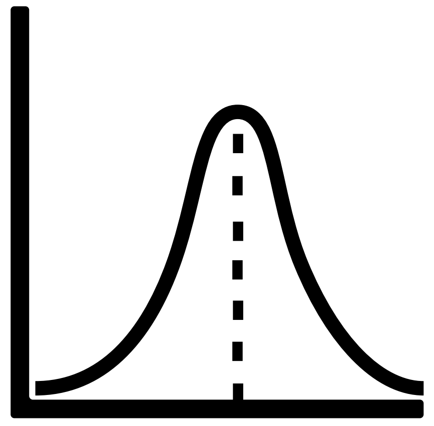
The Wilcoxon test

Student	morning	evening	diff (AM - PM)	ranks (from diff)	rank sums
1	34	45	12	7 (+)	
2	33	36	3	2 (+)	+ve total = 7 + 2 + 3 + 4 + 6
3	41	35	-6	5 (-)	= 22
4	39	43	4	3 (+)	
5	44	42	-2	1 (-)	-ve total = 5 + 1
6	37	42	5	4 (+)	= 6
7	39	46	7	6 (+)	

If there is not difference between AM and PM values the difference between +ve and -ve ranks should be approximately equal.
Null hypothesis: both +ve and -ve rank sums follow the same distribution.

The Wilcoxon test



Null hypothesis:
there *is no difference* in the **central tendencies** between dependent samples (i.e., follows a symmetric distribution around zero).

Alternative hypothesis:
there *is a difference* in the **central tendencies** between dependent samples.

The Wilcoxon test

rank sums

+ve total = 7 + 2 + 3 + 4 + 6 = 22

-ve total = 5 + 1 = 6

Test statistic *W*

W = min (+ve total, -ve total)
W = min (22, 6) = 6

expected value (if there was no diff. between AM & PM)

u_W = n (n + 1) / 4
u_W = 7 (7 + 1) / 4 = 14 (for +ve & -ve)

standard deviation

$$\sigma_W = \sqrt{\frac{n(n+1)(2n+1) - \sum \frac{t_i^3 - t_i}{2}}{24}}$$

z-score

$$z = \frac{W - \mu_W}{\sigma_W}$$

(if > 25 samples, else use table)

The Wilcoxon test

In our example:
the morning group had lower values than the evening group.

p = 0.321

This diff. is not (?) significant.

expected value (if there was no diff.)

$mu_W = n\ (n + 1) / 4$
$mu_W = 7\ \ (7 + 1) / 4 = 14$ (for +ve & -ve)

standard deviation

$$\sigma_W = \sqrt{\frac{n(n+1)(2n+1) - \sum \frac{t_i^3 - t_i}{2}}{24}}$$

z-score

$$z = \frac{W - \mu_W}{\sigma_W}$$

↑

(if > 25 samples, else use table)

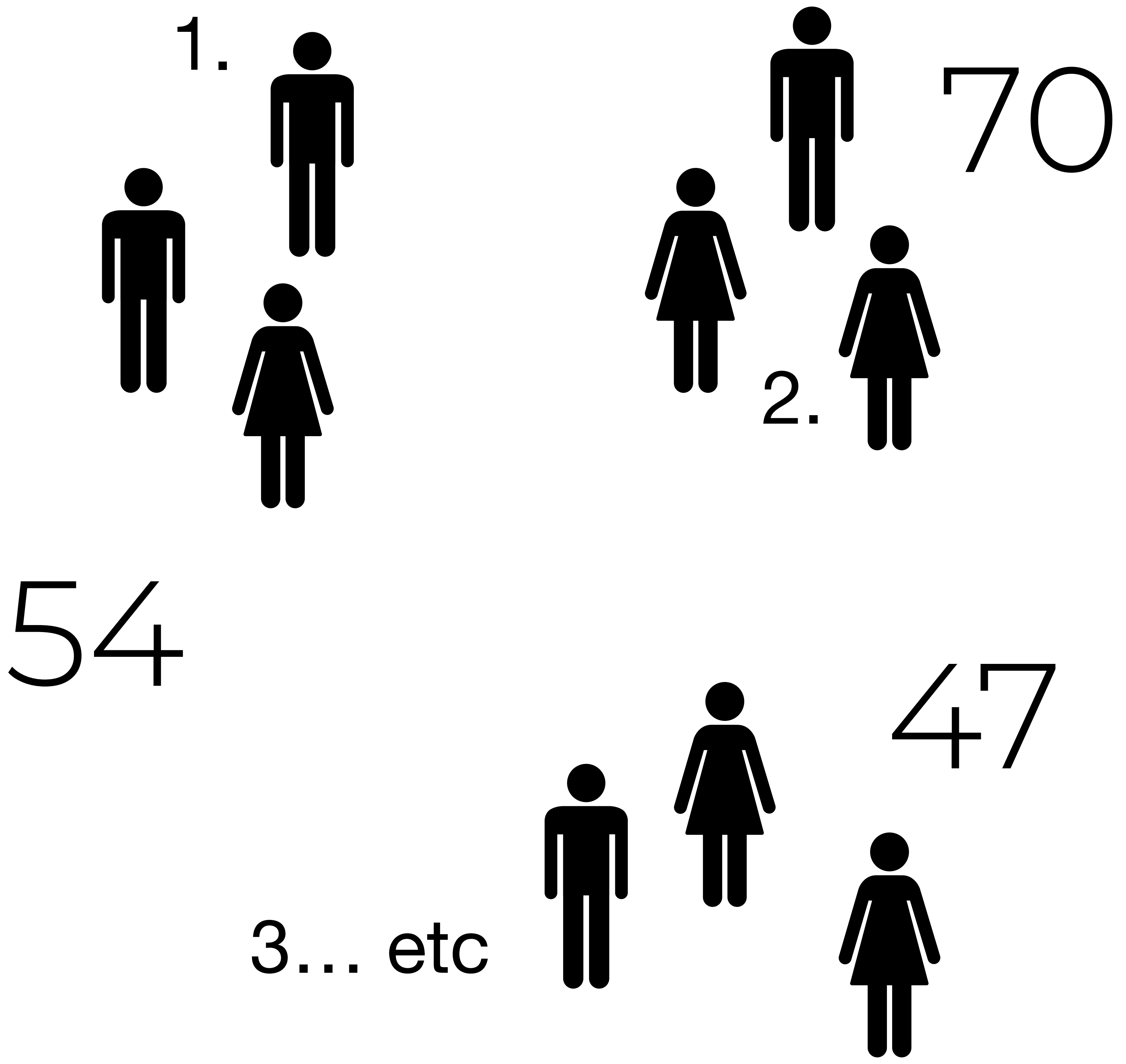
Kruskal-Wallis

The non-parametric version of ANOVA.

Asks, is there a difference between several independent samples?

Data is *not* normally distributed and the data is *not* available in pairs.

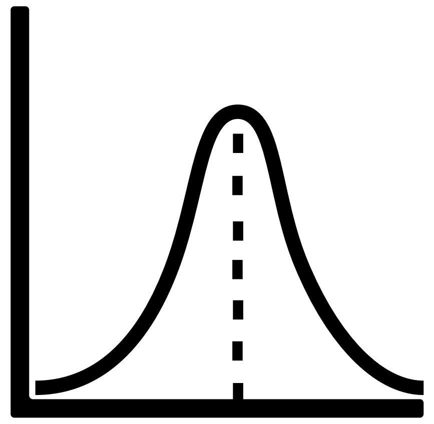
Kruskal-Wallis



ANOVA: is there a diff. in the **mean**?

Kruskal Wallis test: is there a diff. in the **ranks**?

Kruskal-Wallis



Null hypothesis:
The independent samples all have the same central tendency.

Alternative hypothesis:
At least one of the independent samples does not have the same central tendency.

Kruskal-Wallis

Is there a difference in reaction time?

Group	Response time	Rank
A	34	2
A	36	4
A	41	7
A	43	9
B	44	10
B	37	5
B	45	11
B	33	1
C	35	3
C	39	6
C	42	8
C	46	12

If there is no diff. ranks should be distributed randomly.

Kruskal-Wallis

Group	Response time	Rank
A	34	**2**
A	36	**4**
A	41	**7**
A	43	**9**
B	44	**10**
B	37	**5**
B	45	**11**
B	33	**1**
C	35	**3**
C	39	**6**
C	42	**8**
C	46	**12**

Rank sums:

$R_A = 2 + 4 + 7 + 9 = 22$

$R_B = 10 + 5 + 11 + 1 = 27$

$R_C = 3 + 6 + 8 + 12 = 29$

Mean Rank Sum:

$\bar{R}_A = 22 / 4 = 5.5$

$\bar{R}_B = 27 / 4 = 6.75$

$\bar{R}_C = 29 / 4 = 7.25$

$$E_R = \frac{n+1}{2} = \frac{12+1}{2} = 6.5$$ (expected value)

These next ones a little bit hairy 😬
just remember we're doing the same things as before

Kruskal-Wallis

Group	Response time	Rank
A	34	**2**
A	36	**4**
A	41	**7**
A	43	**9**
B	44	**10**
B	37	**5**
B	45	**11**
B	33	**1**
C	35	**3**
C	39	**6**
C	42	**8**
C	46	**12**

Number of cases

$n = 12$

Expected value of the rankings

$E_R = 6.5$

Mean Rank Totals:

$\bar{R}_A = 22 / 4 = 5.5$

$\bar{R}_B = 27 / 4 = 6.75$

$\bar{R}_C = 29 / 4 = 7.25$

Degrees of freedom

$df = 2$

Rank variance

$$\sigma_R^2 = \frac{n^2 - 1}{12} = \frac{12^2 - 1}{12} = 11.92$$

Number of cases

$n = 12$

Expected value of the rankings

$E_R = 6.5$

Mean Rank Totals:

$\bar{R}_A = 22 / 4 = 5.5$

$\bar{R}_B = 27 / 4 = 6.75$

$\bar{R}_C = 29 / 4 = 7.25$

Degrees of freedom:

$df = 2$

Rank variance

$$\sigma_R^2 = \frac{n^2 - 1}{12} = \frac{12^2 - 1}{12} = 11.92$$

Test value H $\rightarrow$ equivalent to χ^2

$$H = \frac{n - 1}{12} \cdot \sum_{i=1}^{k} \frac{n_i\,(\bar{R}_i - E_R)^2}{\sigma_R^2}$$

$$H = \frac{12 - 1}{12} \cdot 4\,\frac{(5.5 - 6.5)^2 + (6.75 - 6.5)^2 + (7.25 - 6.5)^2}{11.92}$$

$= 0.5$

Table of chi-squared distribution

Significance level Alpha	0.995	0.975	0.2	0.1	0.05	0.025	0.02	0.01
Degrees of freedom								
1	0	0.001	1.642	2.706	3.841	5.024	5.412	6.635
2	0.01	0.051	3.219	4.605	5.991	7.378	7.824	9.21
3	0.072	0.216	4.642	6.251	7.815	9.348	9.837	11.345
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277
5	0.412	0.831	7.289	9.236	11.07	12.833	13.388	15.086

Quiz

Under what circumstances do we apply the following tests?

z-test

t-test

ANOVA

Wilcoxon test

Kruskal-Wallis

r-bloggers.com/2021/08/how-to-perform-tukey-hsd-test-in-r/