Non-parametric tests
The Kruskal-Wallis test and Friedman's ANOVA

PSYC234: Statistics: from association to modelling causality

Dr Amy Atkinson
Lecturer in Developmental Psychology
amy.atkinson@lancaster.ac.uk

The plan

My aim: to add a few final statistical tests to your toolbox for when the statistical test you've learned about might not be appropriate

Lecture 7 → Non-parametric tests → A situation where the One factor between-participant ANOVA and the one factor within-participant ANOVA is not appropriate

Learning objectives

- To understand how to assess normality when you have three or more independent groups or three or more repeated measures
- To understand the theory behind the Kruskal-Wallis test and Friedman's ANOVA
- To understand how to conduct the Kruskal-Wallis test and Friedman's ANOVA in R
- To understand how to interpret the Kruskal-Wallis test and Friedman's ANOVA

Part 1

Assessing normality with more than two independent groups

The Kruskal-Wallis test

Assessing the assumption of normality

- Very similar to with only two independent groups - Q-Q plots and Shapiro-Wilk test used

Data in **each** group should follow a normal distribution

Run these steps separately for each group

Research question

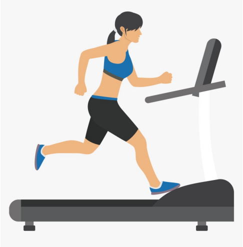
You are a researcher interested in whether drinking protein shakes increases the amount of minutes spent at the gym.

You assign participants to one of three groups: 0 protein shakes, 1 protein shakes, or 2 protein shakes. Once they've drunk the protein shakes, they are given access to gym equipment. You time how long they spend exercising.

0 protein shakes

1 protein shake

2 protein shakes

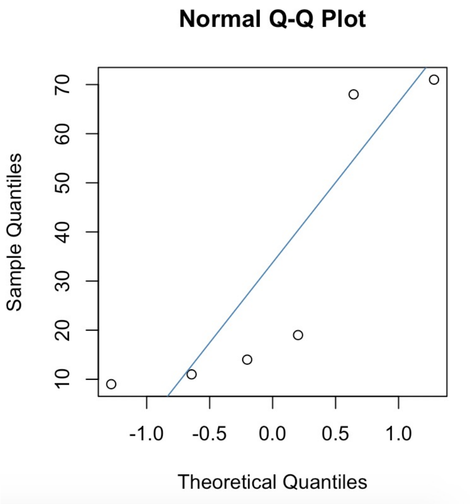


Data

Group	Minutes
No shakes	11
No shakes	9
No shakes	19
No shakes	68
No shakes	71
No shakes	14
One shake	10
One shake	89
One shake	101
One shake	108
One shake	15
One shake	82
Two shakes	87
Two shakes	17
Two shakes	91
Two shakes	103
Two shakes	134
Two shakes	153

Assessing the normality assumption
Q-Q plots and Shapiro-Wilk: per group

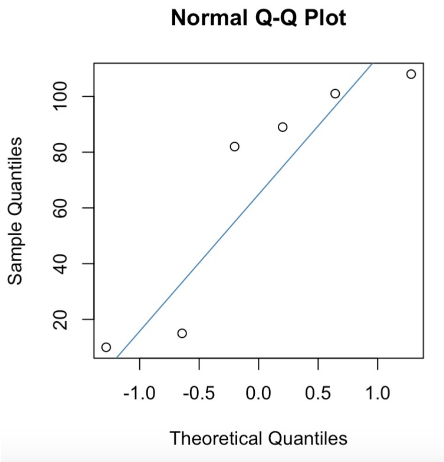
No shakes



Shapiro-Wilk normality test

data: no_shakes$Minutes
W = 0.74074, p-value = 0.01614

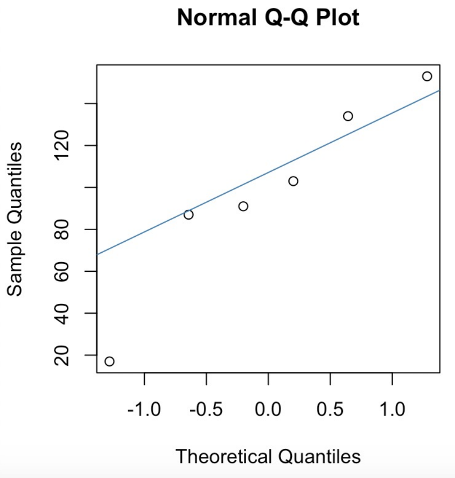
One shake



Shapiro-Wilk normality test

data: one_shake$Minutes
W = 0.81112, p-value = 0.07382

Two shakes



Shapiro-Wilk normality test

data: two_shakes$Minutes
W = 0.93379, p-value = 0.6097

Kruskal-Wallis test

- Alternative to the One-Factor Between-Participants ANOVA
- Appropriate if you have a design with three or more independent groups (and no repeated measures)

The theory behind the Kruskal-Wallis test

Kruskal-Wallis test

Step 1: Order the dependent variable from smallest to largest

Group	Minutes
No shakes	11
No shakes	9
No shakes	19
No shakes	68
No shakes	71
No shakes	14
One shake	10
One shake	89
One shake	101
One shake	108
One shake	15
One shake	82
Two shakes	87
Two shakes	17
Two shakes	91
Two shakes	103
Two shakes	134
Two shakes	153

Group	Minutes
No shakes	9
One shake	10
No shakes	11
No shakes	14
One shake	15
Two shakes	17
No shakes	19
No shakes	68
No shakes	71
One shake	82
Two shakes	87
One shake	89
Two shakes	91
One shake	101
Two shakes	103
One shake	108
Two shakes	134
Two shakes	153

Kruskal-Wallis test

Step 2: Rank the data (smallest =1, second smallest = 2, etc)

Group	Minutes
No shakes	11
No shakes	9
No shakes	19
No shakes	68
No shakes	71
No shakes	14
One shake	10
One shake	89
One shake	101
One shake	108
One shake	15
One shake	82
Two shakes	87
Two shakes	17
Two shakes	91
Two shakes	103
Two shakes	134
Two shakes	153

Group	Minutes	Rank
No shakes	9	1
One shake	10	2
No shakes	11	3
No shakes	14	4
One shake	15	5
Two shakes	17	6
No shakes	19	7
No shakes	68	8
No shakes	71	9
One shake	82	10
Two shakes	87	11
One shake	89	12
Two shakes	91	13
One shake	101	14
Two shakes	103	15
One shake	108	16
Two shakes	134	17
Two shakes	153	18

Same ranking rules as the tests covered last week

Kruskal-Wallis test

Step 3: Sort back into the original groups (e.g. no shakes, one shake, two shakes)

Group	Minutes	Rank
No shakes	9	1
One shake	10	2
No shakes	11	3
No shakes	14	4
One shake	15	5
Two shakes	17	6
No shakes	19	7
No shakes	68	8
No shakes	71	9
One shake	82	10
Two shakes	87	11
One shake	89	12
Two shakes	91	13
One shake	101	14
Two shakes	103	15
One shake	108	16
Two shakes	134	17
Two shakes	153	18

Group	Mins	Rank
No shakes	11	3
No shakes	9	1
No shakes	19	7
No shakes	68	8
No shakes	71	9
No shakes	14	4

Group	Mins	Rank
One shake	10	2
One shake	89	12
One shake	101	14
One shake	108	16
One shake	15	5
One shake	82	10

Group	Mins	Rank
Two shakes	87	11
Two shakes	17	6
Two shakes	91	13
Two shakes	103	15
Two shakes	134	17
Two shakes	153	18

Kruskal-Wallis test

Step 4: Add up the ranks for each group

Group	Mins	Rank
No shakes	11	3
No shakes	9	1
No shakes	19	7
No shakes	68	8
No shakes	71	9
No shakes	14	4
SUM OF RANKS		**32**

3+1+7+8+9+4 = 32

Group	Mins	Rank
One shake	10	2
One shake	89	12
One shake	101	14
One shake	108	16
One shake	15	5
One shake	82	10
SUM OF RANKS		**59**

2+12+14+16+5+10 = 59

Group	Mins	Rank
Two shakes	87	11
Two shakes	17	6
Two shakes	91	13
Two shakes	103	15
Two shakes	134	17
Two shakes	153	18
SUM OF RANKS		**80**

2+12+14+16+5+10 = 80

Kruskal-Wallis test

Step 5: Use these values to calculate the test statistic (*H*)

What is the *H* statistic and how do I calculate it?

Don't worry – not as confusing as it seems!

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ \ldots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

- N = total sample size
- R_1 = sum of ranks for group 1, R_2 = sum of ranks for group 2, R_k = simply tells you to repeat this for each group
- n_1 = sample size for group 1, n_2 = sample size for group 2, n_k = simply tells you to repeat this for each group

Kruskal-Wallis test

Step 5: Use these values to calculate the *H* statistic

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ \dots + \frac{R_k^2}{n_k}\right) - 3(N + 1)$$

Replace the statistical letters with numbers

$$H = \frac{12}{18(19)}\left(\frac{32^2}{6} + \frac{59^2}{6} + \frac{80^2}{6}\right) - 3(19)$$

Kruskal-Wallis test

Step 5: Use these values to calculate the *H* statistic

$$H = \frac{12}{18(19)}\left(\frac{32^2}{6} + \frac{59^2}{6} + \frac{80^2}{6}\right) - 3(19)$$

18*19 = 342

(32*32)/6 = 170.67

(59*59)/6 = 580.17

(80*80)/6 = 1066.67

(3*19) = 57

$$H = \frac{12}{342}(170.67 + 580.17 + 1066.67) - 57$$

Kruskal-Wallis test

Step 5: Use these values to calculate the *H* statistic

$$H = \frac{12}{342}(170.67 + 580.17 + 1066.67) - 57$$

Added up the values inside the brackets

$$H = \frac{12}{342}(1817.51) - 57$$

Kruskal-Wallis test

Step 5: Use these values to calculate the *H* statistic

$$H = \frac{12}{342}(1817.51) - 57$$

$$H = 63.77 - 57$$

(12/342)*1817.51

$$H = 6.77$$

Kruskal-Wallis test

Step 6: Calculate the degrees of freedom

How are the degrees of freedom calculated?

Degrees of freedom (df) = Number of groups - 1

In our gym example, there are 3 groups:

Df = 3 -1

Df = 2

Running the analysis in R

Basic code to run the Kruskal-Wallis test

Keep output

Dependent variable

Tells R which dataframe the data is stored in

```
model <- kruskal.test(Minutes ~ Group, data = gym_data)
model
```

Function to run test

Independent variable

Displays the model output

Output

The degrees of freedom

```
        Kruskal-Wallis rank sum test

data:  Minutes by Group
Kruskal-Wallis chi-squared = 6.7719, df = 2, p-value = 0.03384
```

The H statistic

The p-value: A significant effect of group on the number of minutes spent exercising

Where do the differences lie?

Like with parametric tests, we can perform post-hoc tests

Need to load in the 'FSA' library'

Independent variable

Which method should R use to correct for multiple comparisons?

```
library(FSA)

dunnTest(Minutes ~ Group, data=gym_data, method = "holm")
```

Function to conduct the post-hoc tests

Dependent variable

Tells R which dataframe the data is stored in

Where do the differences lie? Output

Tells you that the p-values were adjusted using the Bonferroni-Holm method

```
> dunnTest(Minutes ~ Group, data=gym_data, method = "holm")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Holm method.

          Comparison         Z      P.unadj      P.adj
1   0 Shakes - 1 Shake -1.459993 0.144292055 0.2885841
2 0 Shakes - 2 Shakes -2.595543 0.009444166 0.0283325
3   1 Shake - 2 Shakes -1.135550 0.256144967 0.2561450
```

Adjusted p-value

P value before adjustment

Where do the differences lie? Output

```
> dunnTest(Minutes ~ Group, data=gym_data, method = "holm")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Holm method.

          Comparison         Z      P.unadj      P.adj
1  0 Shakes - 1 Shake -1.459993 0.144292055 0.2885841
2 0 Shakes - 2 Shakes -2.595543 0.009444165 0.0283325
3  1 Shake - 2 Shakes -1.135550 0.256144967 0.2561450
```

A significant difference between 0 shakes and 2 shakes

In which direction?

```
gym_data %>%
  group_by(Group) %>%
  summarise(med = median(Minutes))
```

```
# A tibble: 3 × 2
  Group       med
  <fct>     <dbl>
1 0 Shakes   16.5
2 1 Shake    85.5
3 2 Shakes   97
```

Participants in the 2 Shakes group exercised for longer than participants in the 0 shakes group

Effect size

- No easy way to calculate an effect size for the Kruskal-Wallis test

Reporting in APA format

```
        Kruskal-Wallis rank sum test

data:  Minutes by Group
Kruskal-Wallis chi-squared = 6.7719, df = 2, p-value = 0.03384
```

	Comparison	Z	P.unadj	P.adj
1	0 Shakes - 1 Shake	-1.459993	0.144292055	0.2885841
2	0 Shakes - 2 Shakes	-2.595543	0.009444166	0.0283325
3	1 Shake - 2 Shakes	-1.135550	0.256144967	0.2561450

There was a significant effect of protein shakes on the number of minutes spent exercising, $H(2) = 6.77$, $p = .034$. Post-hoc comparisons were conducted using Dunn's test. P-values were corrected using Bonferroni-Holm. There was a significant difference between the 0 shakes (median = 16.5, range = 9-71) and 2 shakes groups (median = 97, range = 17-153; $p = .028$), with participants in the 2 shakes group exercising for significantly longer. No significant difference emerged between the 0 shake and 1 shake groups (median = 85.5, range = 10-108; $p = 0.289$), or the 1 shake and 2 shakes group ($p = 0.256$).

Reporting in APA format

There was a significant effect of protein shakes on the number of minutes spent exercising, $H(2) = 6.77$, $p = .034$. Post-hoc comparisons were conducted using Dunn's test. P-values were corrected using Bonferroni-Holm. There was a significant difference between the 0 shakes (median = 16.5, range = 9-71) and 2 shakes groups (median = 97, range = 17-153; $p = .028$), with participants in the 2 shakes group exercising for significantly longer. No significant difference emerged between the 0 shake and 1 shake groups (median = 85.5, range = 10-108; $p = 0.289$), or the 1 shake and 2 shakes group ($p = 0.256$).