

# **Introduction to Statistics with GraphPad Prism: Answers**

## Contents

<b>Introduction to Statistics with GraphPad Prism: Answers</b> .....	<b>1</b>
<i>Version 2024-05</i> .....	1
<b>Exercise 1: Power</b> .....	<b>3</b>
Activated T cells .....	3
Mice weight .....	4
Arachnophobes .....	4
<b>Exercise 2: t-test</b> .....	<b>7</b>
Working memory .....	7
Coffee .....	9
<b>Exercise 3: ANOVA</b> .....	<b>11</b>
Crop yield .....	11
Neutrophils .....	13
<b>Exercise 4: Correlation</b> .....	<b>15</b>
Roe deer .....	15
Exam anxiety .....	16
<b>Exercise 5: Non-parametric</b> .....	<b>20</b>
Oxbridge rivalry: smelly T-shirts [Mann-Whitney] .....	20
Botulinum [Wilcoxon paired] .....	20
Creatine [Kruskal-Wallis] .....	21
Violin [Friedman] .....	22
Dominance [Spearman Rank] .....	23
<b>Exercise 6: Categorical/qualitative</b> .....	<b>25</b>
Cats & dogs .....	25
Cane toads .....	25
<b>Exercise 7: Mixed</b> .....	<b>28</b>
Iris flowers .....	28
Recycling .....	31
Income data .....	32
Coffee – part 2 .....	33
<i>With outliers</i> .....	34
<i>Without outliers</i> .....	35

## Exercise 1: Power Activated T cells

- *Providing the observed difference between WT and KO cells is of scientific interest, what sample size is needed to achieve 80% power?*

	Polarised	Not Polarised	Proportions
WT	10	31	10/41=0.24
KO	14	21	14/35=0.40

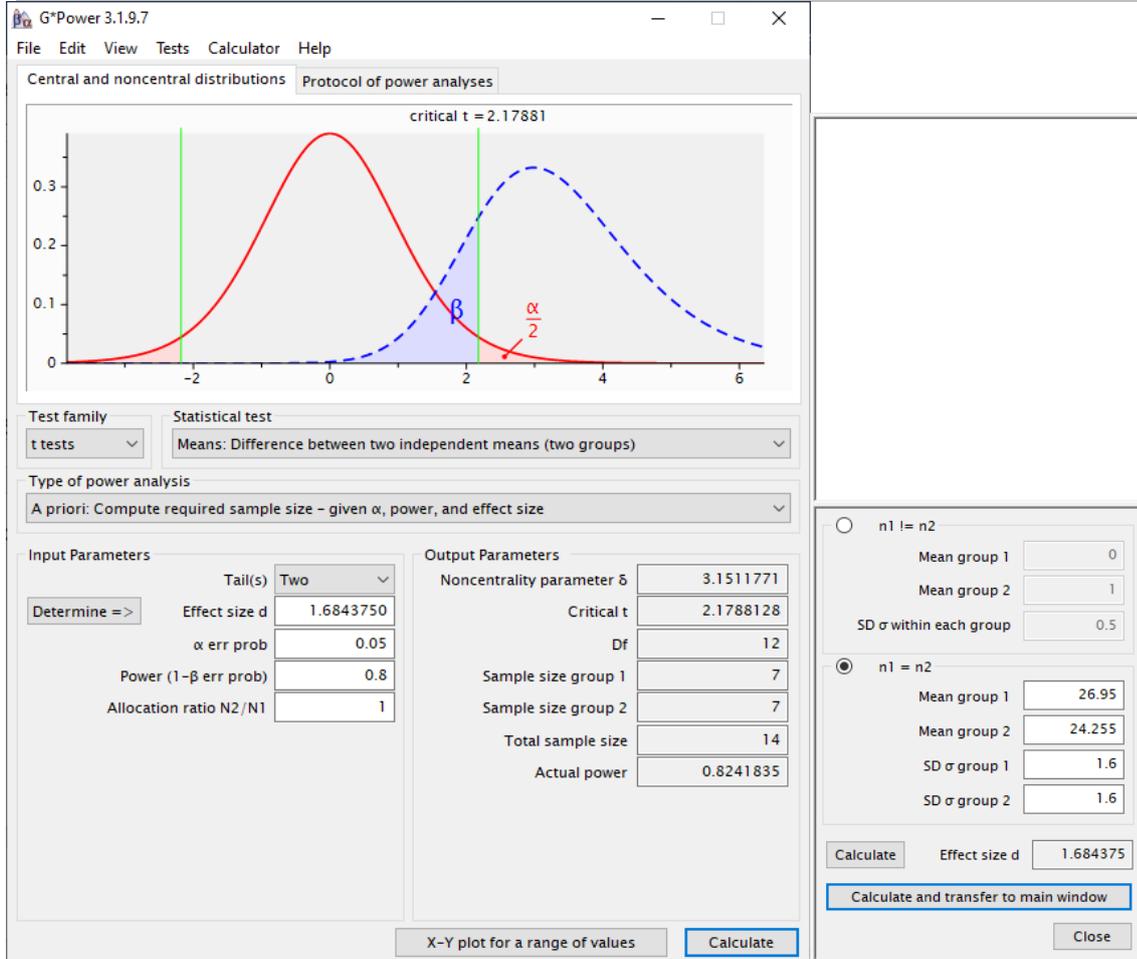
The screenshot shows the G\*Power 3.1.3 interface. The 'Exact' test family is selected, with the statistical test set to 'Proportions: Inequality, two independent groups (Fisher's exact test)'. The type of power analysis is 'A priori: Compute required sample size - given alpha, power, and effect size'. The input parameters are: Tail(s) = Two, Proportion p1 = 0.24, Proportion p2 = 0.40, alpha err prob = 0.05, Power (1-beta err prob) = 0.80, and Allocation ratio N2/N1 = 1. The output parameters are: Sample size group 1 = 144, Sample size group 2 = 144, Total sample size = 288, Actual power = 0.8032472, and Actual alpha = 0.0365897. The values 144 and 288 are circled in red.

- To achieve a power of 80%, you will need a total sample of **288 cells**.

## Mice weight

- What sample size is needed to be able to spot a 10% difference with 80% power?

					Mean	SD	10% difference
<b>Weight</b>	27.2	25.5	26	29.1	26.95	1.6	26.95*0.9 = 24.255



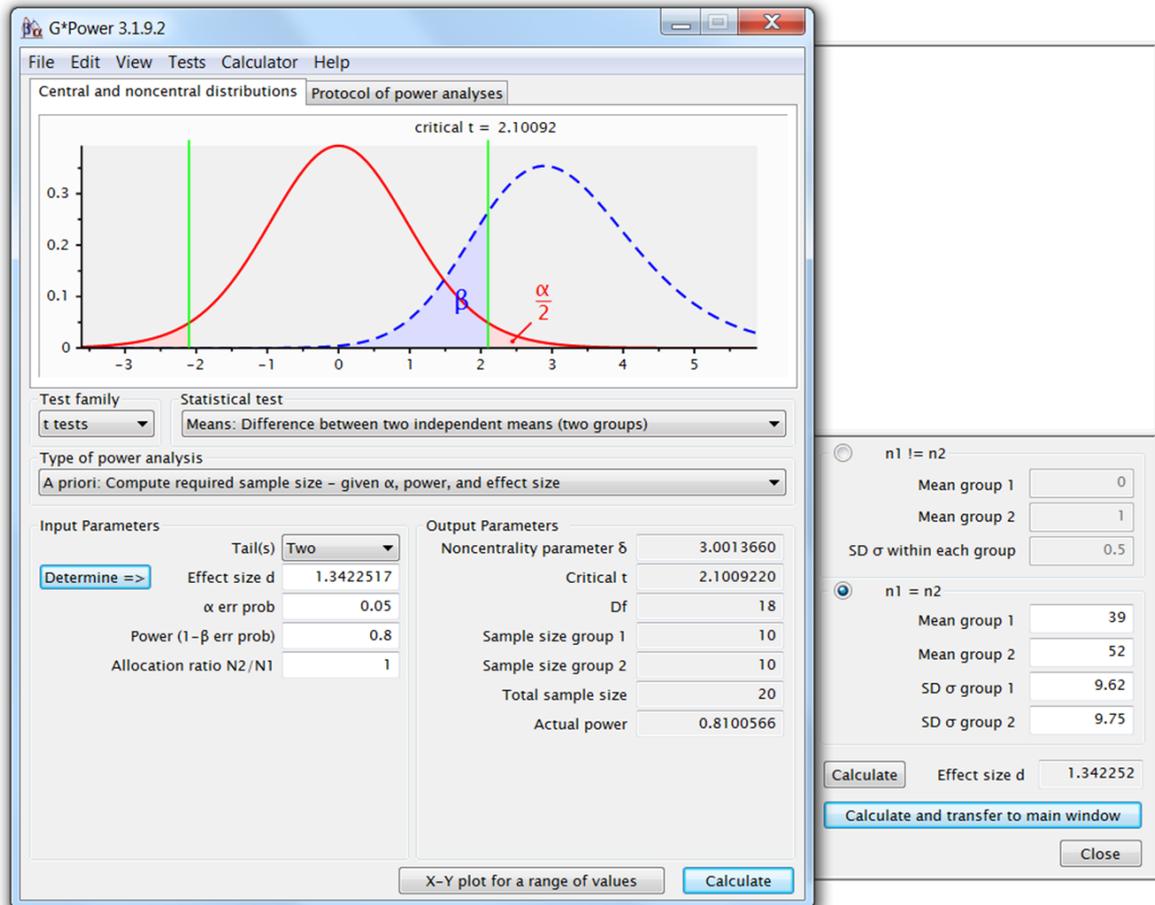
- To be able to spot a 10% difference in the mice weight, at 80% power, assuming the KO group have the same variance as the WT, you will need **7 mice per group**.

## Arachnophobes

- Use the data to calculate the values for a power calculation

	Picture	Real Spider
	25	45
	35	40
	45	55
	40	55
	50	65
<b>Mean</b>	39	52
<b>SD</b>	9.6177	9.7468

o Run a power calculation



- Providing the preliminary results are to be trusted, to be able to pick up such a difference between the 2 groups, with a power of 80% and a significance level of 5%, we will need at least **10 arachnophobes in each group**.
- How many arachnophobes would you need to achieve a power of 80%, if based on a paired design? Answer: **4 arachnophobes**

Picture	Real Spider	Difference
25	45	-20
35	40	-5
45	55	-10
40	55	-15
50	65	-15
	<b>Mean</b>	-13
	<b>SD</b>	5.700877

G\*Power 3.1.9.7

File Edit View Tests Calculator Help

Central and noncentral distributions Protocol of power analyses

critical t = 3.18245

Test family: t tests

Statistical test: Means: Difference between two dependent means (matched pairs)

Type of power analysis: A priori: Compute required sample size - given alpha, power, and effect size

Input Parameters

Tail(s): Two

Determine => Effect size dz: 2.2807018

alpha err prob: 0.05

Power (1-beta err prob): 0.8

Output Parameters

Noncentrality parameter delta: 4.5614036

Critical t: 3.1824463

Df: 3

Total sample size: 4

Actual power: 0.8462264

From differences

Mean of difference: 13

SD of difference: 5.7

From group parameters

Mean group 1: 1

Mean group 2: 1

SD group 1: 1

SD group 2: 1

Correlation between groups: 0.5

Calculate Effect size dz: 2.280702

Calculate and transfer to main window

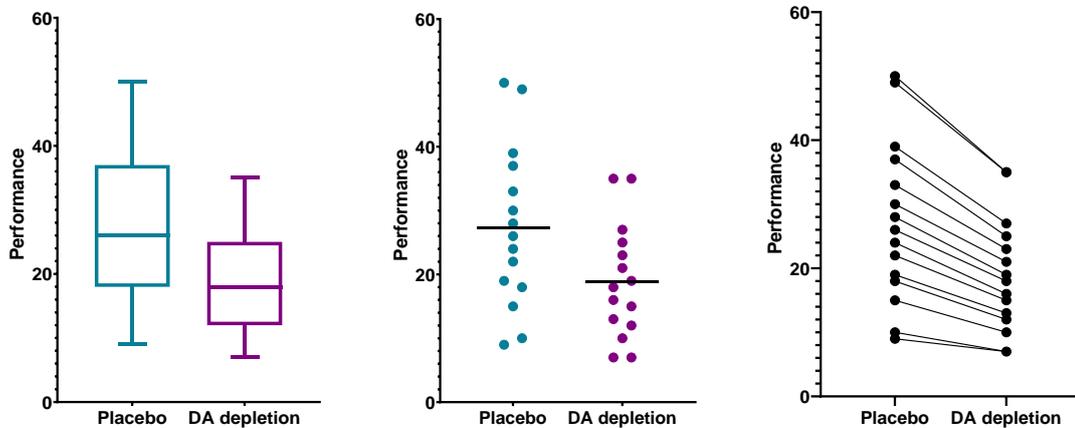
Close

X-Y plot for a range of values Calculate

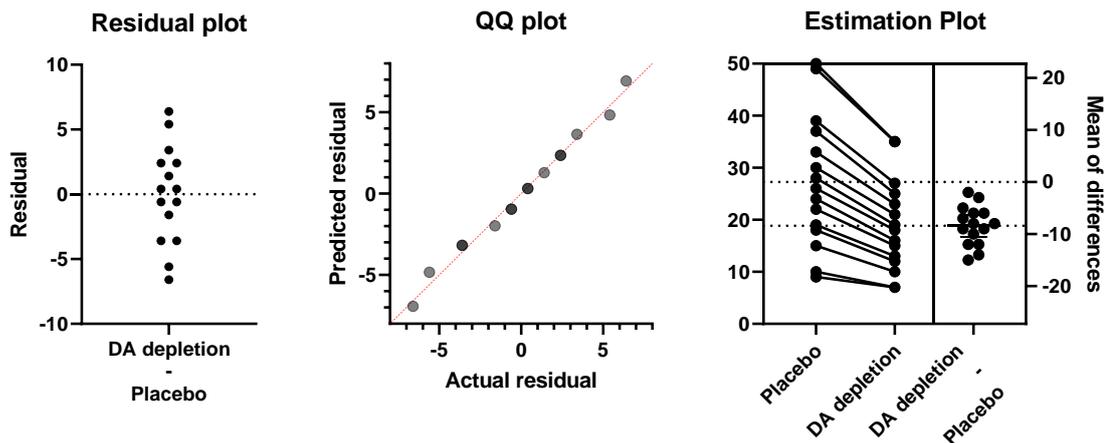
## Exercise 2: t-test

### Working memory

- *Is there an effect of treatment on the monkeys' performance?*
- First, plot the data – it looks normal and there aren't any outliers on the boxplot. It looks like there is likely to be an effect.

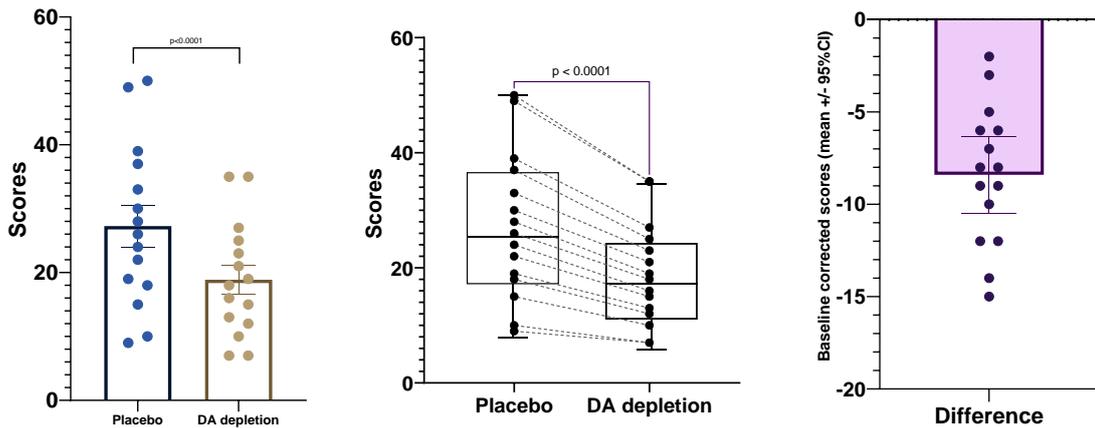


- Then, run the paired t test in Prism, with residual and QQ plots. Again, it looks like it meets the assumptions for parametric testing, and it passes all the tests for normality of the residuals (see the results page below). The estimation plot shows the paired points and mean differences, so can be a useful summary.



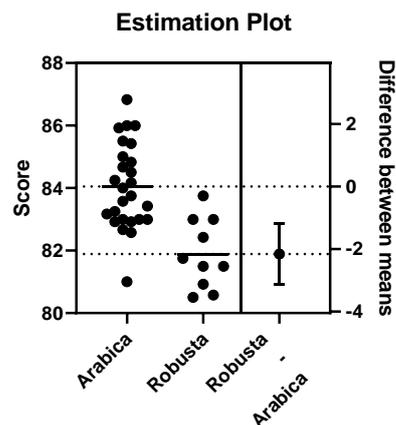
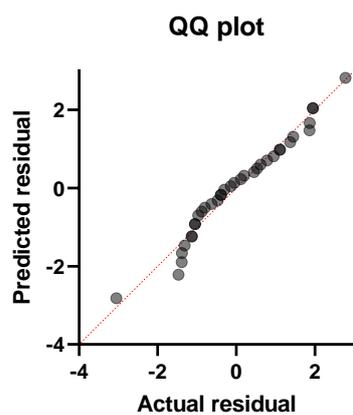
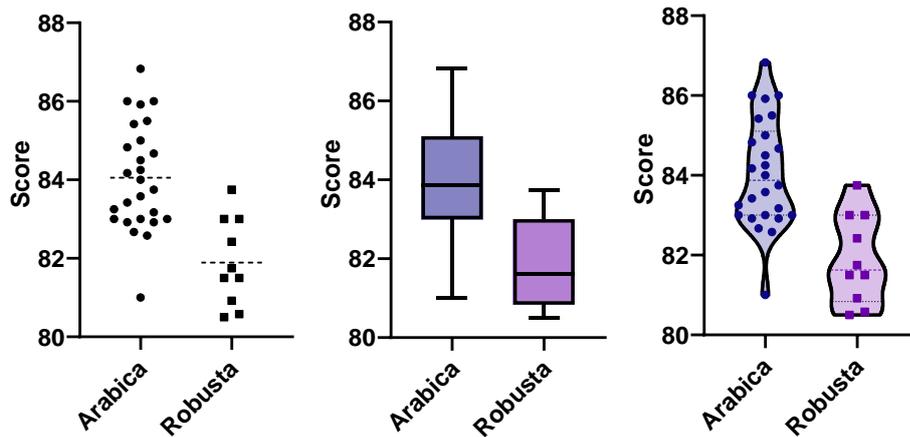
Paired t test				
Tabular results				
1	Table Analyzed	Working memory		
2				
3	Column B	DA depletion		
4	vs.	vs.		
5	Column A	Placebo		
6				
7	<b>Paired t test</b>			
8	P value	<0.0001		
9	P value summary	****		
10	Significantly different (P < 0.05)?	Yes		
11	One- or two-tailed P value?	Two-tailed		
12	t, df	t=8.616, df=14		
13	Number of pairs	15		
14				
15	<b>How big is the difference?</b>			
16	Mean of differences (B - A)	-8.400		
17	SD of differences	3.776		
18	SEM of differences	0.9749		
19	95% confidence interval	-10.49 to -6.309		
20	R squared (partial eta squared)	0.8413		
21				
22	<b>How effective was the pairing?</b>			
23	Correlation coefficient (r)	0.9986		
24	P value (one tailed)	<0.0001		
25	P value summary	****		
26	Was the pairing significantly effective?	Yes		
27				
28	<b>Normality of Residuals</b>			
29	<b>Test name</b>	<b>Statistics</b>	<b>P value</b>	<b>Passed normality test (alpha=0.05)?</b>
30	Anderson-Darling (A2*)	0.1569	0.9395	Yes
31	D'Agostino-Pearson omnibus (K2)	0.1823	0.9129	Yes
32	Shapiro-Wilk (W)	0.9773	0.9474	Yes
33	Kolmogorov-Smirnov (distance)	0.1035	0.1000	Yes
34				

- The results page for the paired t test gives a statistically significant result, with a p value <0.0001. The 95% confidence interval does not include 0 and the pairing was highly effective.
- **Answer: There is a significant effect of dopamine depletion on working memory in rhesus monkeys (p<0.0001).**
- There are a number of ways this result can be presented graphically. It is misleading to show a p-value associated with a paired t-test without showing the pairing on the graph (as in the first plot). It is preferable to show what has actually been tested, by either showing the pairing on the plot or plotting the differences between the groups, as in the second and third plots.
- Just showing the difference can be a useful, especially if there are a lot of points – to do this:
  - Analyse > Remove baseline and column math
  - Definition of baseline: Selected column(s) = Every other dataset (column): 1<sup>st</sup>, 3<sup>rd</sup>, 5<sup>th</sup>...
  - Calculation: Difference = Value – Baseline.
  - Create a graph from the resulting data.



## Coffee

- You are tired and want a decent coffee but are not sure which type to have – the coffee shop has the option of Robusta or Arabica beans. You find some data where they have been tasted and rated for their overall score and decide to use this to base your decision on. Which should you choose for your coffee?
- Start by plotting the data and checking for normality – the QQ plot looks slightly wobbly but still mostly follows the line, and our data are passing normality tests, so we are going to use a t-test.



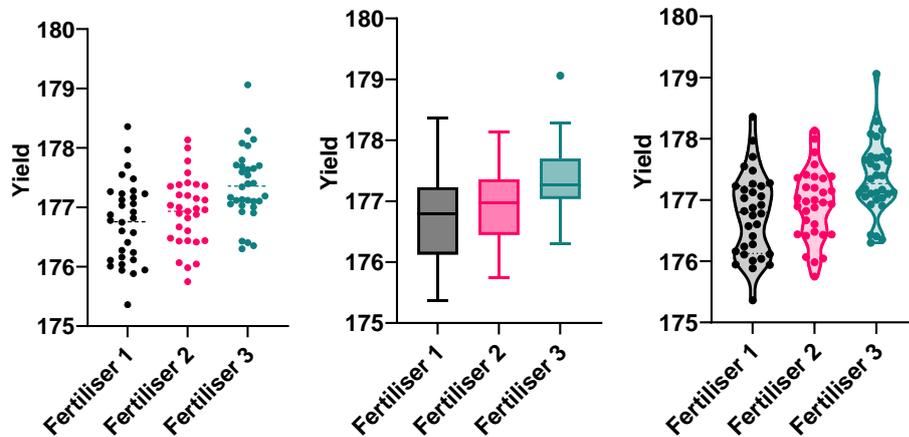
<b>Unpaired t test</b>				
P value	<0.0001			
P value summary	****			
Significantly different (P < 0.05)?	Yes			
One- or two-tailed P value?	Two-tailed			
t, df	t=4.466, df=34			
<b>How big is the difference?</b>				
Mean of column A	84.05			
Mean of column B	81.89			
Difference between means (B - A) ± SEM	-2.160 ± 0.4836			
95% confidence interval	-3.143 to -1.177			
R squared (eta squared)	0.3697			
<b>F test to compare variances</b>				
F, DFn, Dfd	1.500, 25, 9			
P value	0.5390			
P value summary	ns			
Significantly different (P < 0.05)?	No			
<b>Normality of Residuals</b>				
<b>Test name</b>	<b>Statistics</b>	<b>P value</b>	<b>Passed normality test (alpha=0.05)?</b>	<b>P value summary</b>
Anderson-Darling (A2*)	0.5179	0.1763	Yes	ns
D'Agostino-Pearson omnibus (K2)	0.2451	0.8846	Yes	ns
Shapiro-Wilk (W)	0.9655	0.3163	Yes	ns
Kolmogorov-Smirnov (distance)	0.09749	0.1000	Yes	ns
<b>Data analyzed</b>				
Sample size, column A	26			
Sample size, column B	10			

- **Answer: There is a significant difference in taste score between the two coffee beans ( $p < 0.0001$ ). You should choose the Arabica beans as these get a higher score.**

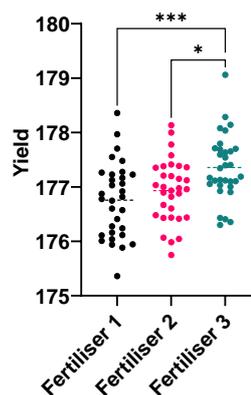
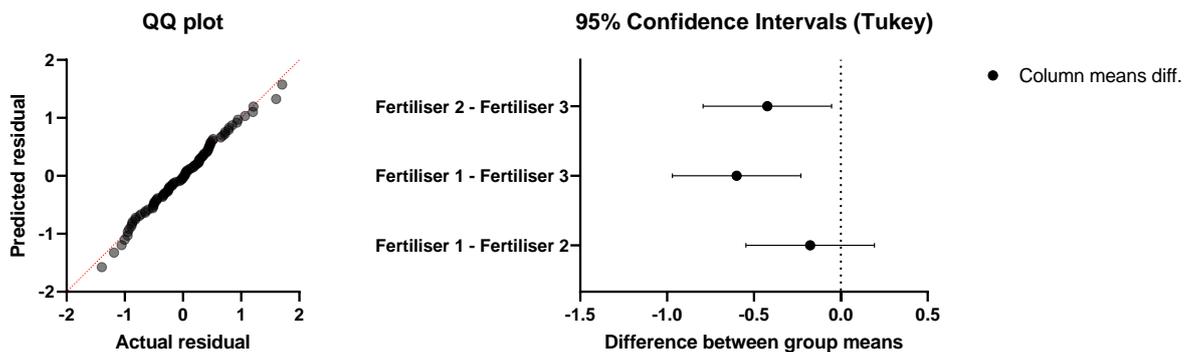
## Exercise 3: ANOVA

### Crop yield

- As a crop researcher, you want to test the effect of three different fertiliser mixtures on crop yield. You carry out an experiment using three fertiliser types (1, 2, and 3) and measure crop yield. Do the fertilisers have an impact on yield and, if so, which fertiliser gives the highest yield?



- After plotting the data and looking at the QQ plot and normality tests, a parametric approach looks reasonable. The box and whiskers plot highlights a potential outlier in fertiliser 3 but it doesn't look implausible and, given no other reason to exclude it, we will leave it in. If we wanted to, we could run additional outlier identification tests (which also come back as negative).



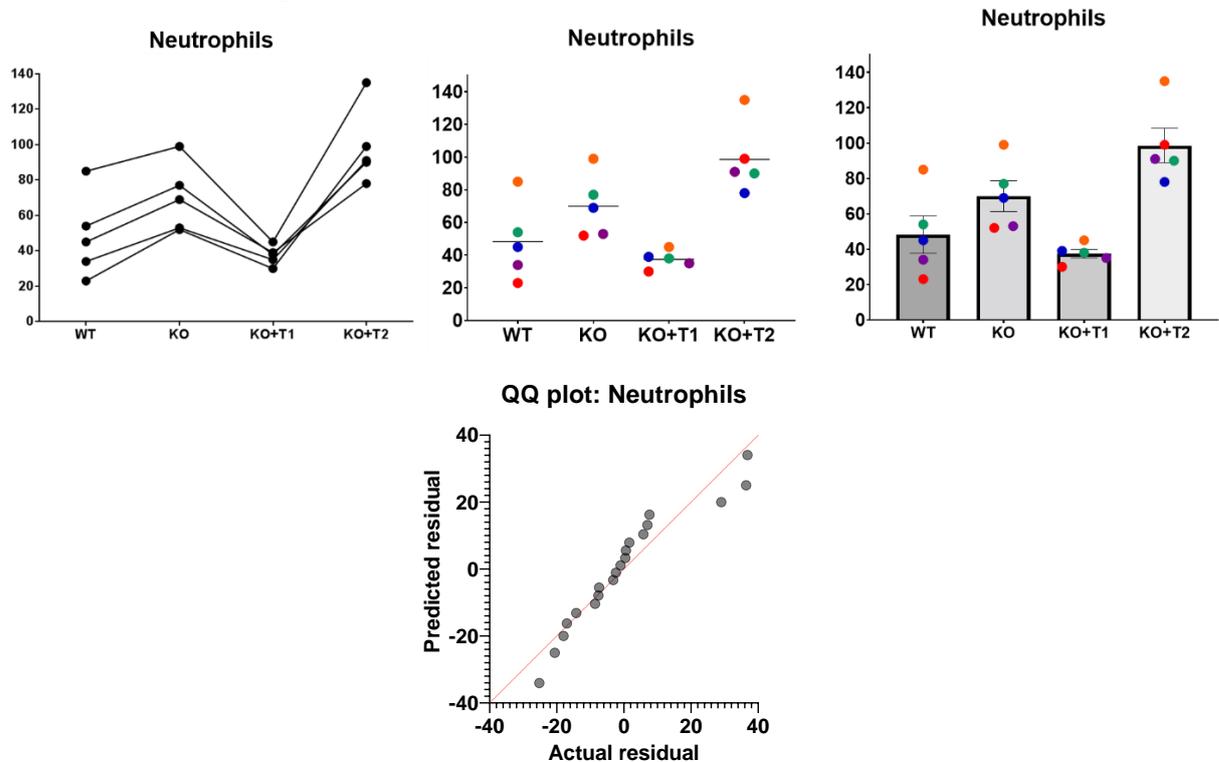
Ordinary one-way ANOVA ANOVA results						
4	<b>ANOVA summary</b>					
5	F	7.863				
6	P value	0.0007				
7	P value summary	***				
8	Significant diff. among means (P < 0.05):	Yes				
9	R squared	0.1446				
10						
11	<b>Brown-Forsythe test</b>					
12	F (DFn, DFd)	0.8470 (2, 93)				
13	P value	0.4320				
14	P value summary	ns				
15	Are SDs significantly different (P < 0.05):	No				
16						
17	<b>Bartlett's test</b>					
18	Bartlett's statistic (corrected)	1.062				
19	P value	0.5880				
20	P value summary	ns				
21	Are SDs significantly different (P < 0.05):	No				
22						
23	<b>ANOVA table</b>	<b>SS</b>	<b>DF</b>	<b>MS</b>	<b>F (DFn, DFd)</b>	<b>P value</b>
24	Treatment (between columns)	6.068	2	3.034	F (2, 93) = 7.863	P=0.0007
25	Residual (within columns)	35.89	93	0.3859		
26	Total	41.95	95			
27						
28	<b>Normality of Residuals</b>					
29	<b>Test name</b>	<b>Statistics</b>	<b>P value</b>	<b>Passed normality test (alpha=0.05)?</b>	<b>P value summary</b>	
30	D'Agostino-Pearson omnibus (K2)	0.8769	0.6451	Yes	ns	
31	Anderson-Darling (A2*)	0.2067	0.8651	Yes	ns	
32	Shapiro-Wilk (W)	0.9909	0.7595	Yes	ns	
33	Kolmogorov-Smirnov (distance)	0.05294	0.1000	Yes	ns	
34						
35	<b>Data summary</b>					
36	Number of treatments (columns)	3				
37	Number of values (total)	96				

Ordinary one-way ANOVA Multiple comparisons									
1	Number of families	1							
2	Number of comparisons per family	3							
3	Alpha	0.05							
4									
5	<b>Tukey's multiple comparisons test</b>	<b>Mean Diff.</b>	<b>95.00% CI of diff.</b>	<b>Below threshold?</b>	<b>Summary</b>	<b>Adjusted P Value</b>			
6	Fertiliser 1 vs. Fertiliser 2	-0.1762	-0.5461 to 0.1937	No	ns	0.4954			
7	Fertiliser 1 vs. Fertiliser 3	-0.5991	-0.9690 to -0.2292	Yes	***	0.0006			
8	Fertiliser 2 vs. Fertiliser 3	-0.4230	-0.7928 to -0.05306	Yes	*	0.0209			
9									
10	<b>Test details</b>	<b>Mean 1</b>	<b>Mean 2</b>	<b>Mean Diff.</b>	<b>SE of diff.</b>	<b>n1</b>	<b>n2</b>	<b>q</b>	<b>DF</b>
11	Fertiliser 1 vs. Fertiliser 2	176.8	176.9	-0.1762	0.1553	32	32	1.604	93
12	Fertiliser 1 vs. Fertiliser 3	176.8	177.4	-0.5991	0.1553	32	32	5.456	93
13	Fertiliser 2 vs. Fertiliser 3	176.9	177.4	-0.4230	0.1553	32	32	3.852	93

- After running the ANOVA, we can see there is a significant difference between the three fertilisers from the omnibus test ( $p=0.0007$ ), so we can reject the null hypothesis that there is no difference. From the multiple comparisons test, we can see that fertiliser 3 is significantly different from both 1 and 2 ( $p=0.0006$  and  $p=0.0209$ , respectively), but there is no significant difference between fertilisers 1 and 2 ( $p=0.4954$ ). **Therefore, we would want to use fertiliser 3 to maximise yield.**

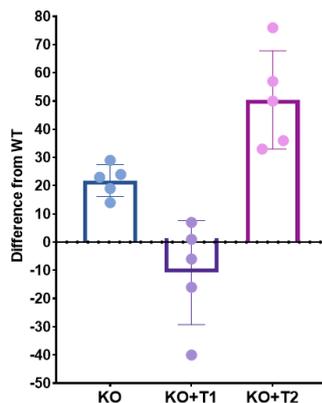
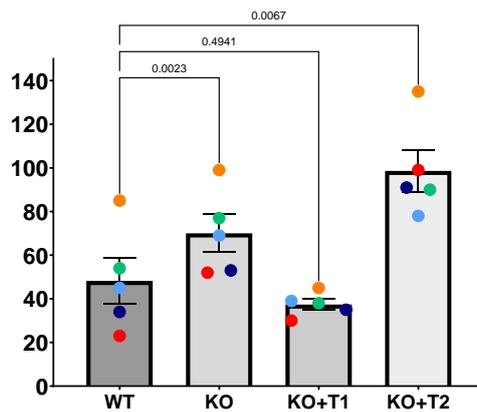
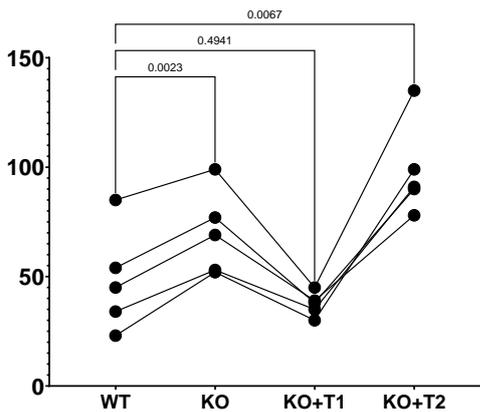
## Neutrophils

- A researcher is looking at the difference between 4 cell groups. They have run the experiment 5 times. Within each experiment, they have neutrophils from a WT (control), a KO, a KO+Treatment 1 and a KO+Treatment2. Is there a difference between KO with/without treatment and WT?
- Again, we want to start with plotting our data. As the experiment has been run independently 5 times, this makes it a repeated measures design, and this should be reflected in our graphs.



- *Hint: to colour the points by experiment you highlight the data on the data table, right click and select Format Points > Symbol Color*
- Although the QQ plot is not perfect, it still mostly follows the line and our data look okay otherwise, so we will use a repeated-measures one-way ANOVA. From this, **there is a significant difference between our groups ( $p=0.0002$ ). KO and KO+T2 are significantly different from WT ( $p=0.0023$  and  $p=0.0067$ , respectively) but there is no significant difference between KO+T1 and WT ( $p=0.4941$ ).**

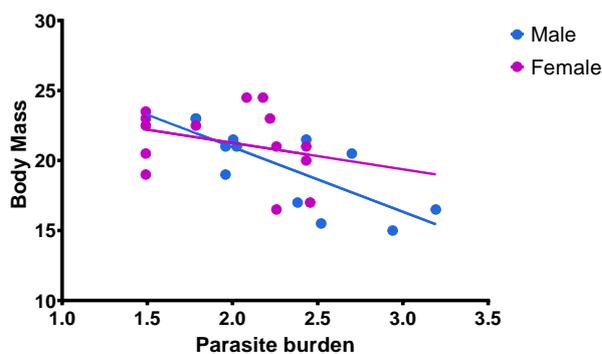
RM one-way ANOVA						
ANOVA results						
Table Analyzed	Neutrophils					
<b>Repeated measures ANOVA summary</b>						
Assume sphericity?	No					
F	28.57					
P value	0.0002					
P value summary	***					
Statistically significant (P < 0.05)?	Yes					
Geisser-Greenhouse's epsilon	0.6916					
R squared	0.8772					
<b>Was the matching effective?</b>						
F	8.239					
P value	0.0020					
P value summary	**					
Is there significant matching (P < 0.05)?	Yes					
R squared	0.2522					
<b>ANOVA table</b>						
	<b>SS</b>	<b>DF</b>	<b>MS</b>	<b>F (DFn, DFd)</b>	<b>P value</b>	
Treatment (between columns)	10948	3	3649	F (2.075, 8.299) = 28.57	P=0.0002	
Individual (between rows)	4209	4	1052	F (4, 12) = 8.239	P=0.0020	
Residual (random)	1533	12	127.7			
Total	16689	19				
<b>Dunnett's multiple comparisons test</b>						
	<b>Mean Diff.</b>	<b>95.00% CI of diff.</b>	<b>Below threshold?</b>	<b>Summary</b>	<b>Adjusted P Value</b>	<b>A-?</b>
WT vs. KO	-21.80	-30.91 to -12.69	Yes	**	0.0023	B KO
WT vs. KO+T1	10.80	-19.02 to 40.62	No	ns	0.4941	C KO+T1
WT vs. KO+T2	-50.40	-78.53 to -22.27	Yes	**	0.0067	D KO+T2



## Exercise 4: Correlation

### Roe deer

- *Is there a relationship between parasite burden and body mass in roe deer?*
- For this we can look at correlation or linear regression – although we want to do a correlation as our test, linear regression can give us some more options to explore our data, and gives us a line of best fit. Linear regression gives us  $R^2$  and a p-value for whether the slope is non-zero (the same p as from the correlation), although it does not directly give us Pearson r.



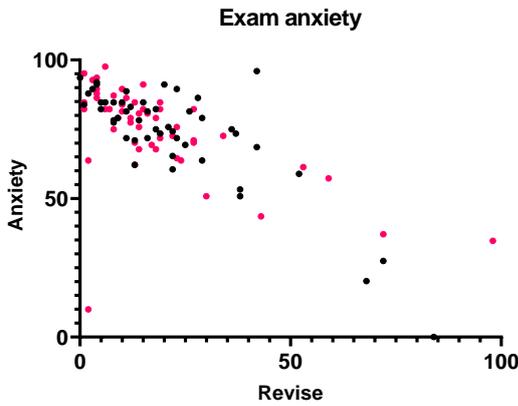
Correlation		A	B
		PL vs. Male	PL vs. Female
1	<b>Pearson r</b>		
2	r	-0.7504	-0.3020
3	95% confidence interval	-0.9256 to -0.3099	-0.7176 to 0.2722
4	R squared	0.5630	0.09119
5			
6	<b>P value</b>		
7	P (two-tailed)	0.0049	0.2940
8	P value summary	**	ns
9	Significant? (alpha = 0.05)	Yes	No
10			
11	Number of XY Pairs	12	14

Linear reg. Tabular results		A	B
		Male	Female
1	<b>Best-fit values</b>		
2	Slope	-4.621	-1.888
3	Y-intercept	30.20	25.04
4	X-intercept	6.536	13.26
5	1/slope	-0.2164	-0.5297
6			
7	<b>Std. Error</b>		
8	Slope	1.287	1.721
9	Y-intercept	3.025	3.453
10			
11	<b>95% Confidence Intervals</b>		
12	Slope	-7.490 to -1.753	-5.637 to 1.861
13	Y-intercept	23.46 to 36.94	17.51 to 32.56
14	X-intercept	4.902 to 13.47	5.738 to +infinity
15			
16	<b>Goodness of Fit</b>		
17	R square	0.5630	0.09119
18	Sy.x	1.980	2.512
19			
20	<b>Is slope significantly non-zero?</b>		
21	F	12.89	1.204
22	DFn, DFd	1, 10	1, 12
23	P value	0.0049	0.2940
24	Deviation from zero?	Significant	Not Significant
25			
26	<b>Equation</b>	Y = -4.621*X + 30.20	Y = -1.888*X + 25.04
27			
28	<b>Data</b>		
29	Number of X values	12	26
30	Maximum number of Y replicates	1	1
31	Total number of values	12	14
32	Number of missing values	0	12

- **Answer:** There is a negative correlation between parasite load and fitness but this relationship is only significant for the males ( $r = -0.7504$ ,  $p = 0.0049$  vs. females:  $r = -0.3020$ ,  $p = 0.2940$ ).

### Exam anxiety

- *Is there a relationship between time spent revising and exam anxiety? And, if yes, are males and females different? How good is the model?*
- From plotting our data, we can see there are potentially a couple of outliers. We could run the correlation straight away, however this gives us very limited options to learn more about our data.



- Anxiety F
- Anxiety M

Correlation	Revise vs. Anxiety F	Revise vs. Anxiety M
	Y	Y
Pearson r		
r	-0.8214	-0.5974
95% confidence interval	-0.8945 to -0.7055	-0.7483 to -0.3877
R squared	0.6746	0.3568
P value		
P (two-tailed)	<0.0001	<0.0001
P value summary	****	****
Significant? (alpha = 0.05)	Yes	Yes

- If we run a non-linear regression, but choose a linear model this gives us many more options, including normality testing and outlier identification.

Parameters: Nonlinear Regression

Parameters: Nonlinear Regression

Parameters: Nonlinear Regression

Model Method Compare Constrain Initial values Range Output Confidence Diagnostics Flag

**How to quantify goodness-of-fit?**

R squared  Sy.x  Sum-of-Squares  
 Adjusted R squared  RMSE  AICC

**Are residuals Gaussian (normal)?**

D'Agostino-Pearson omnibus normality test  
 Anderson-Darling test  
 Shapiro-Wilk normality test  
 Kolmogorov-Smirnov normality test with Dallal-Wilkinson-Lilliefor P value

**Are residuals clustered or heteroscedastic?**

Runs test  Replicates test  Test for appropriate weighting (homoscedasticity)

**Which residual graphs to create?**

Residual vs X plot  
 Residual vs Y plot  
 Homoscedasticity plot  
 QQ plot  
 Actual vs Predicted plot

**Are the parameters intertwined, redundant or skewed?**

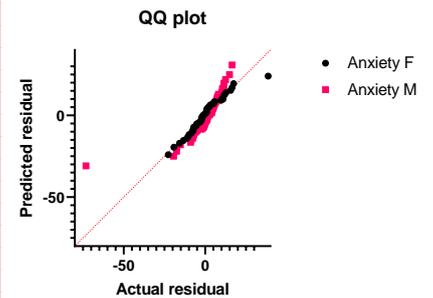
Covariance of parameters  
 Dependency  
 Hougaard's measure of skewness

Make these diagnostics choices the default for future fits.

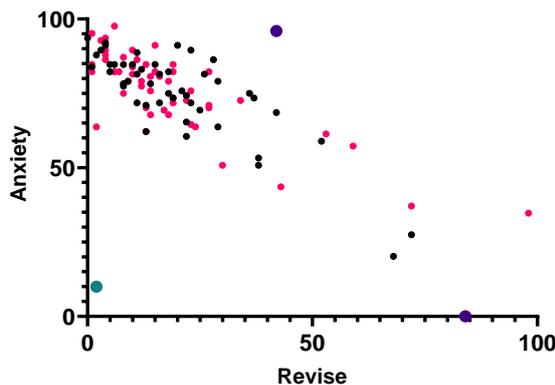
Learn Cancel OK

Nonlin fit Table of results	A	B
	Anxiety F	Anxiety M
<b>Normality of Residuals</b>		
<b>D'Agostino-Pearson omnibus (K2)</b>	14.43	68.42
P value	0.0007	<0.0001
Passed normality test (alpha=0.05)?	No	No
P value summary	***	****
<b>Anderson-Darling (A2')</b>	0.7493	3.041
P value	0.0478	<0.0001
Passed normality test (alpha=0.05)?	No	No
P value summary	*	****
<b>Shapiro-Wilk (W)</b>	0.9429	0.6997
P value	0.0161	<0.0001
Passed normality test (alpha=0.05)?	No	No
P value summary	*	****
<b>Kolmogorov-Smirnov (distance)</b>	0.1243	0.1887
P value	0.0475	<0.0001
Passed normality test (alpha=0.05)?	No	No
P value summary	*	****
<b>Number of points</b>		
# of X values	51	103
# Y values analyzed	51	52
Outliers (not excluded, Q=1%)	2	1

Nonlin fit Table of results	A	B	C
	Anxiety F	Anxiety M	Global (shared)
<b>Comparison of Fits</b>			
1 Null hypothesis			Slope same for all data sets
2 Alternative hypothesis			Slope different for at least one data set
3 P value			0.0299
4 Conclusion (alpha = 0.05)			Reject null hypothesis
5 Preferred model			Slope different for at least one data set
6 F (DFn, DFd)			4.852 (1, 99)
<b>Slope different for at least one data set</b>			
<b>Best-fit values</b>			
11 YIntercept	91.94	84.19	
12 Slope	-0.8238	-0.5353	
<b>95% CI (profile likelihood)</b>			
14 YIntercept	87.36 to 96.52	78.93 to 89.46	
15 Slope	-0.9880 to -0.6596	-0.7394 to -0.3312	
<b>Goodness of Fit</b>			
17 Degrees of Freedom	49	50	
18 R squared	0.6746	0.3568	
19 Sum of Squares	5322	8845	
20 Sy.x	10.42	13.30	



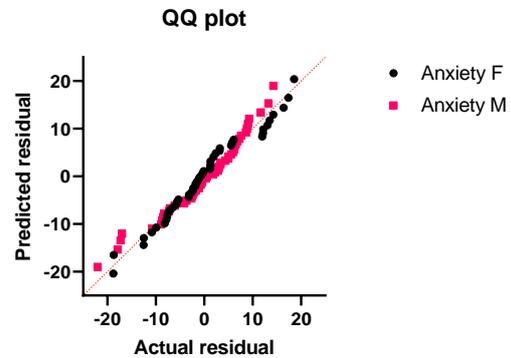
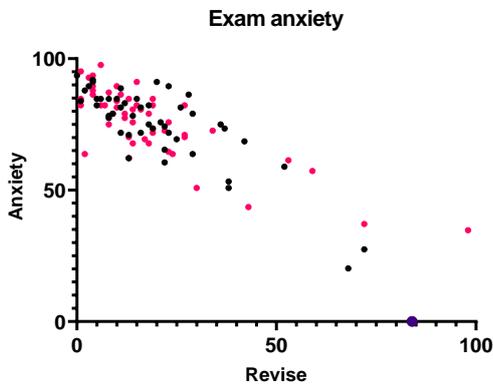
Exam anxiety



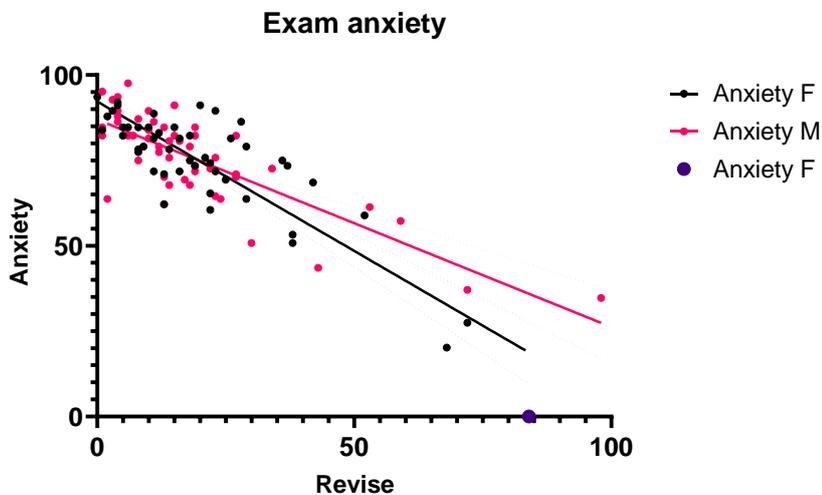
- Anxiety F
- Anxiety M
- Anxiety M
- Anxiety F

Revise	Anxiety F	Anxiety M
84.000	0.056	
42.000	95.970	
2.000		10.000

- This shows that our data are not normal and there are clearly outliers. The test found three outliers, however from looking at our data, one of them does not look particularly far removed. If we instead just remove the two worst looking outliers, we can see that our data are now passing normality tests and follow the QQ plot line more closely.



Normality of Residuals			Correlation	A	B
				Revise vs. Anxiety F	Revise vs. Anxiety M
D'Agostino-Pearson omnibus (K2)	0.5158	5.132			
P value	0.7727	0.0768			
Passed normality test (alpha=0.05)?	Yes	Yes			
P value summary	ns	ns			
Anderson-Darling (A2*)	0.6528	0.6557	Pearson r		
P value	0.0834	0.0821	r	-0.8737	-0.8081
Passed normality test (alpha=0.05)?	Yes	Yes	95% confidence interval	-0.9267 to -0.7866	-0.8863 to -0.6851
P value summary	ns	ns	R squared	0.7633	0.6530
Shapiro-Wilk (W)	0.9673	0.9566			
P value	0.1787	0.0597	P value		
Passed normality test (alpha=0.05)?	Yes	Yes	P (two-tailed)	<0.0001	<0.0001
P value summary	ns	ns	P value summary	****	****
Kolmogorov-Smirnov (distance)	0.1167	0.09776	Significant? (alpha = 0.05)	Yes	Yes
P value	0.0863	>0.1000			
Passed normality test (alpha=0.05)?	Yes	Yes	Number of XY Pairs	50	51
P value summary	ns	ns			



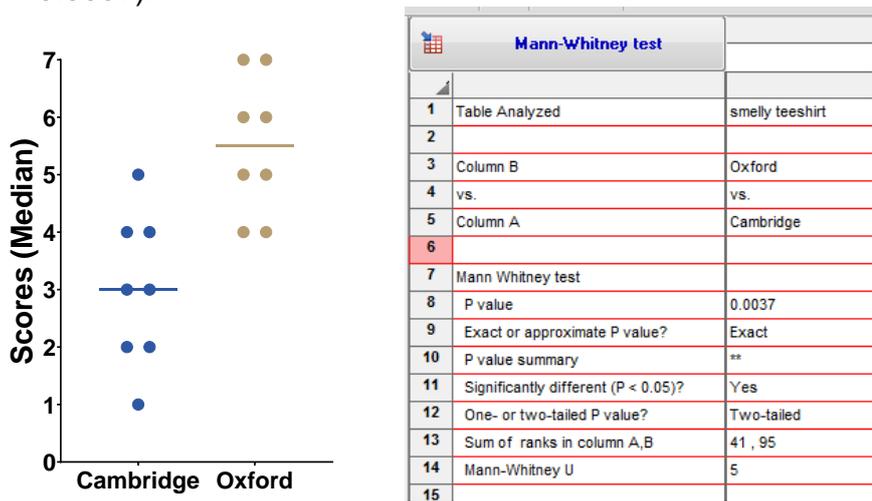
Nonlin fit Table of results	A	B	C
	Anxiety F	Anxiety M	Global (shared)
<b>Comparison of Fits</b>			
Null hypothesis			Slope same for all data sets
Alternative hypothesis			Slope different for at least one data set
P value			0.0056
Conclusion (alpha = 0.05)			Reject null hypothesis
Preferred model			Slope different for at least one data set
F (DFn, DFd)			8.022 (1, 97)
<b>Slope different for at least one data set</b>			
<b>Best-fit values</b>			
YIntercept	92.25	86.97	
Slope	-0.8750	-0.6075	
<b>95% CI (profile likelihood)</b>			
YIntercept	88.35 to 96.14	83.66 to 90.29	
Slope	-1.016 to -0.7336	-0.7347 to -0.4804	
<b>Goodness of Fit</b>			
Degrees of Freedom	48	49	
R squared	0.7633	0.6530	
Sum of Squares	3759	3306	
Sy.x	8.849	8.213	

- In this case, the outliers are meaning our data are not normal and they are far removed from the rest of the data, so I opted to remove. Removing them also greatly increases the goodness of fit, particularly in males where the  $R^2$  increases from 0.3568 to 0.6530.
- However, how to deal with outliers can be tricky to decide, especially when it is not your own data and you do not have the context behind it. You should state how you will deal with outliers before seeing your data, and be consistent across experiments (i.e. **do not delete them when you get a significant result without them and leave them in when you get a significant result with them**).
- **The correlation is highly significant for both males and females ( $p < 0.0001$ ) so we can conclude that there is a negative correlation between anxiety and revision (Pearson  $r = -0.8737$  and  $-0.8081$  for females and males, respectively).**
- However, the slope for the line of best fit is significantly different between males and females, indicating that for each point increase in revision, females tend to reduce anxiety by more than males.

## Exercise 5: Non-parametric

### Oxbridge rivalry: smelly T-shirts [Mann-Whitney]

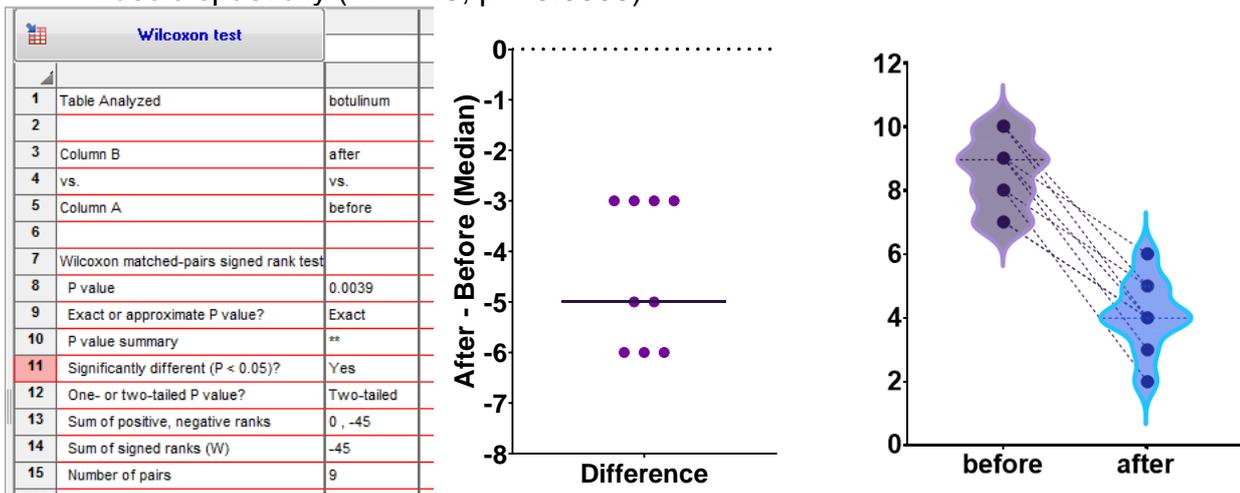
- *Study:* Two groups of Cambridge University students are presented with one of two worn T-shirts with university logos. Disgust score: 7 = most disgusting. Can Cambridge students tell the difference between T-shirts from Oxford or Cambridge?
- As this is a non-parametric test, need to select 't tests (and nonparametric tests)' in analysis menu
- **Answer:** Cambridge students can tell the difference between Oxford and Cambridge ( $U = 5$ ,  $p = 0.0037$ ).



- What do you think about the design?
  - The biggest flaw is that the T-shirts had logos on, meaning the study is not blinded, and subject to bias. It seems likely that the students rated the Oxford t-shirts worse because they had the logo on, rather than anything to do with the smell. A paired design would also have been better.

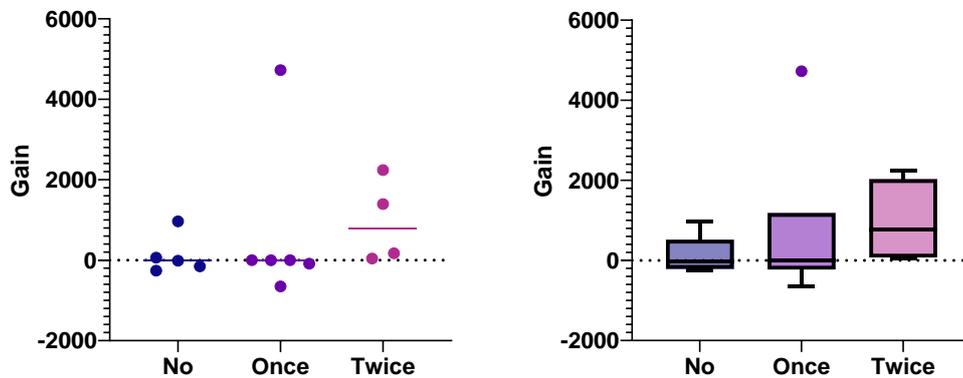
### Botulinum [Wilcoxon paired]

- *Question:* do botulinum toxin injections reduce muscle spasticity levels?
- **Answer:** There was a significant difference pre- and post- treatment in ratings of muscle spasticity ( $W = -45$ ,  $p = 0.0039$ ).



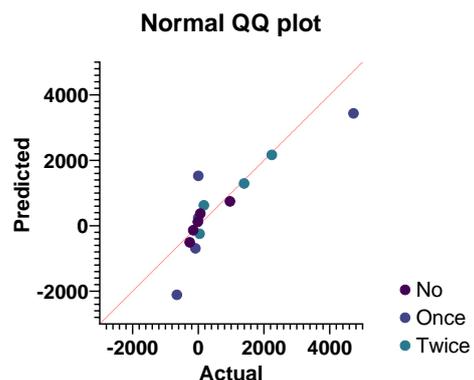
## Creatine [Kruskal-Wallis]

- Question: does the average weight gain depend on the creatine group to which people were assigned? Explore the data and answer the question with a non-parametric approach.

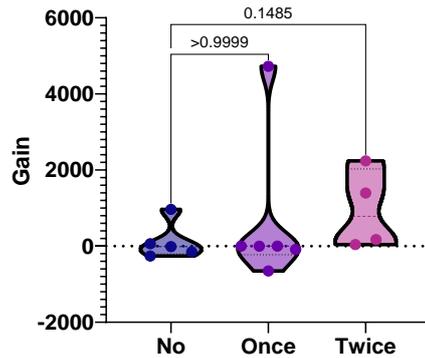


- It looks like the outlier in the 'once' group might be too high to be plausible, so should probably be removed but below are the results both including and excluding the outlier. If these were your own data, you would know if it was plausible and could check whether there were any errors in measurement, etc. **The outlier should not be excluded just because there is a significant result after it is removed.**
- Results (with outlier):

Kruskal-Wallis test ANOVA results		
1	Table Analyzed	Creatine
2		
3	<b>Kruskal-Wallis test</b>	
4	P value	0.1458
5	Exact or approximate P value?	Exact
6	P value summary	ns
7	Do the medians vary signif. (P < 0.05)?	No
8	Number of groups	3
9	Kruskal-Wallis statistic	3.868
10		
11	<b>Data summary</b>	
12	Number of treatments (columns)	3
13	Number of values (total)	15
14		
15		

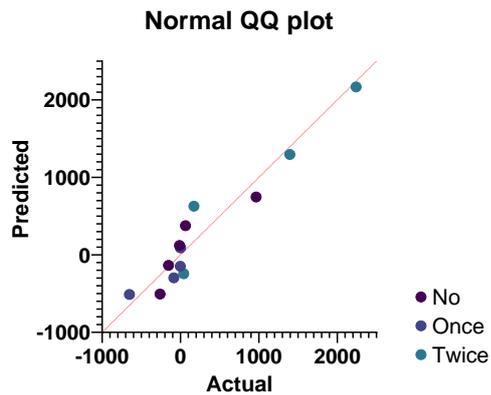


Dunn's multiple comparisons test	Mean rank diff.	Significant?	Summary	Adjusted P Value
No vs. Once	-0.4333	No	ns	>0.9999
No vs. Twice	-5.350	No	ns	0.1485

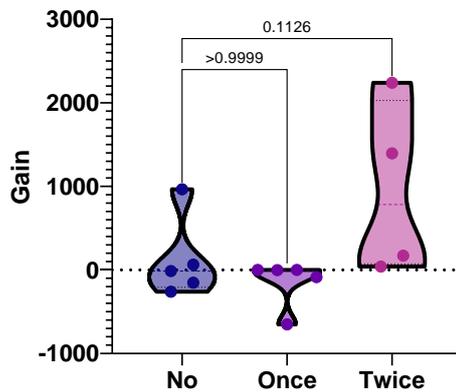


- Results (without outlier):

Kruskal-Wallis test		
ANOVA results		
1	Table Analyzed	Creatine
2		
3	<b>Kruskal-Wallis test</b>	
4	P value	0.0397
5	Exact or approximate P value?	Exact
6	P value summary	*
7	Do the medians vary signif. (P < 0.05)?	Yes
8	Number of groups	3
9	Kruskal-Wallis statistic	5.999
10		
11	<b>Data summary</b>	
12	Number of treatments (columns)	3
13	Number of values (total)	14
14		



Dunn's multiple comparisons test	Mean rank diff.	Significant?	Summary	Adjusted P Value
No vs. Once	1.200	No	ns	>0.9999
No vs. Twice	-5.350	No	ns	0.1126



## Violin [Friedman]

- An auction house is putting three violins, A, B, and C, up for bidding. Ten violinists are blindfolded and asked to rate the instruments and each player plays the violins in a randomly determined sequence (BCA, ACB, etc.).
- After each violin is played (Instructor), the violinist (Rater) rates the instrument on a 10-point scale of overall excellence (Likert: 1=lowest, 10=highest).

- Question: which violin is the best according to the 10 violinists?

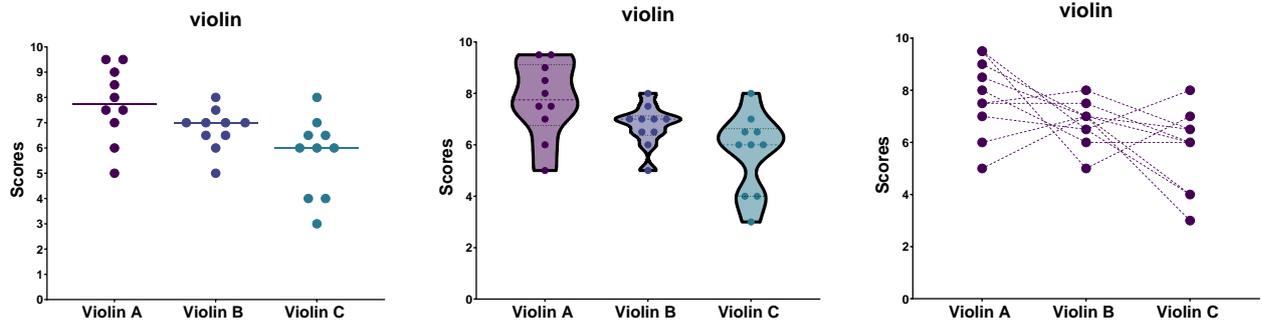
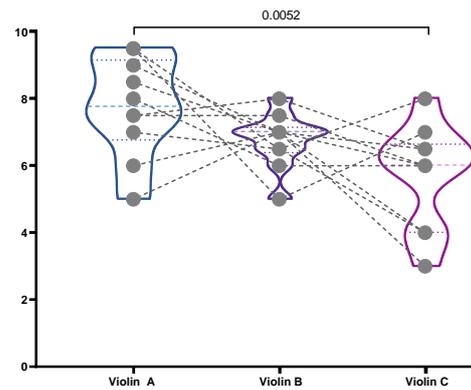


Table Analyzed	violin
<b>Friedman test</b>	
P value	0.0033
Exact or approximate P value?	Exact
P value summary	**
Are means signif. different? (P < 0.05)	Yes
Number of groups	3
Friedman statistic	10.47
<b>Data summary</b>	
Number of treatments (columns)	3
Number of subjects (rows)	10

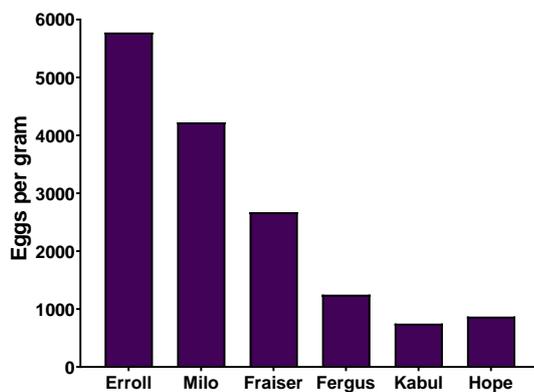


Dunn's multiple comparisons test	Rank sum diff.	Significant?	Summary	Adjusted P Value
Violin A vs. Violin B	5.500	No	ns	0.6563
Violin A vs. Violin C	14.00	Yes	**	0.0052
Violin B vs. Violin C	8.500	No	ns	0.1720

- There is a significant difference between the three violins ( $p=0.0033$ ), with violin A being ranked higher than violin C ( $p=0.0052$ ).

### Dominance [Spearman Rank]

- Six male colobus monkeys ranked for dominance and the eggs of *Trichirus nematode* per gram of monkey faeces measured. Is social dominance associated with parasitism?



Correlation		Dominance vs. Eggs per gram
1	Spearman r	
2	r	-0.9429
3	95% confidence interval	
4		
5	P value	
6	P (two-tailed)	0.0167
7	P value summary	*
8	Exact or approximate P value?	Exact
9	Significant? (alpha = 0.05)	Yes
10		
11	Number of XY Pairs	6
12		

- **Answer:** the relationship between dominance and parasitism is significant ( $\rho = -0.94$ ,  $p = 0.017$ ) with high ranking males harbouring a heavier burden.

## Exercise 6: Categorical/qualitative

### Cats & dogs

- Run a test for dogs & compare to cats, graphically represent results for both

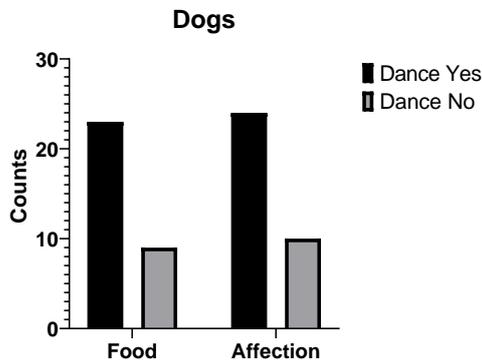
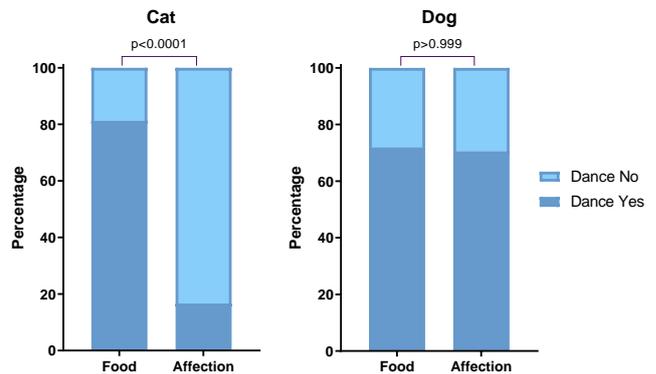


Table Analyzed	Dog	
<b>P value and statistical significance</b>		
Test	Fisher's exact test	
P value	>0.9999	
P value summary	ns	
One- or two-sided	Two-sided	
Statistically significant (P < 0.05)?	No	
<b>Effect size</b>		
	<b>Value</b>	<b>95% CI</b>
Odds ratio	1.065	0.3716 to 3.164
Reciprocal of odds ratio	0.9391	0.3160 to 2.691

Table Analyzed	Dog	
<b>P value and statistical significance</b>		
Test	Chi-square	
Chi-square, df	0.01331, 1	
z	0.1154	
P value	0.9081	
P value summary	ns	
One- or two-sided	Two-sided	
Statistically significant (P < 0.05)?	No	
<b>Effect size</b>		
	<b>Value</b>	<b>95% CI</b>
Odds ratio	1.065	0.3716 to 3.164
Reciprocal of odds ratio	0.9391	0.3160 to 2.691



- Answer:** Odds Ratio = 1.065. If you are a dancing dog, you are pretty much as likely to have received food than affection as a reward ( $p > 0.99$ ). This is a contrast to cats, who are more likely to dance having received food.

### Cane toads

- Is the proportion of cane toads infected by intestinal parasites the same in 3 different areas of Queensland? Produce graphical and statistical answers.

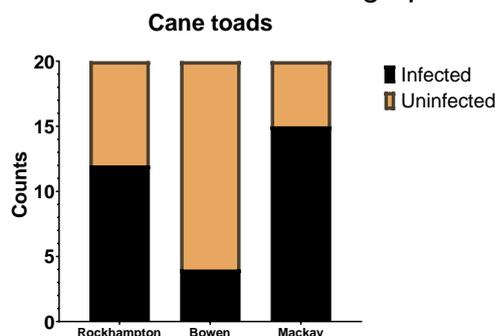


Table Analyzed	Cane toad	
<b>Chi-square</b>		
Chi-square, df	12.95, 2	
P value	0.0015	
P value summary	**	
One- or two-tailed	NA	
Statistically significant? (alpha<0.05)	Yes	
<b>Data analyzed</b>		
Number of rows	3	
Number of columns	2	

- Answer 1:** The proportion of cane toads infected by intestinal parasites varies significantly between the 3 different areas of Queensland ( $p = 0.0015$ ), the

animals being more likely to be parasitized in Rockhampton and Mackay than in Bowen.

- *Is the proportion of infected cane toads different in Bowen than in the other 2 areas?*
- For this, we need to run 3 separate tests to get results for the individual comparisons (i.e. Rockhampton vs Mackay, Rockhampton vs Bowen, and Mackay vs Bowen), then add the resulting p-values to a column table and select “Analyze a stack of P values” to get multiple comparison corrections:

	Group A	Group B	Group C	Group D	Group E	Group F	Group G	Group H	Group I
	P value	Title							
1	0.0225								
2	0.5006								
3	0.0012								
4									

The input to this analysis should be a stack of P values (computed elsewhere)

**How to decide which P values are small enough to investigate further**

False Discovery Rate (FDR) approach  
 Method: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommended)  
 Desired FDR (Q) = 1 %

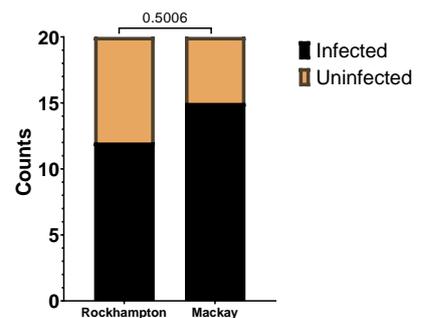
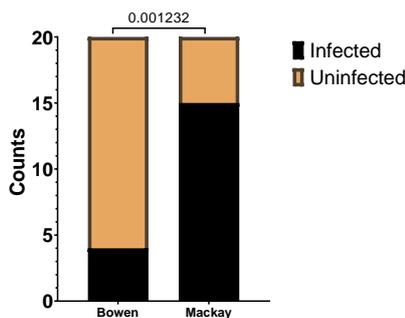
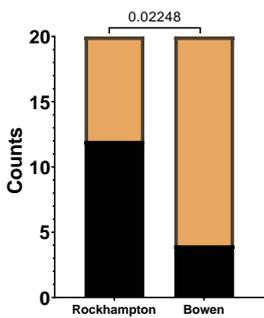
Statistical significance  
 Method: Holm-Šidák (more power)  
 Alpha: 0.05 (Definition of 'statistically significant')

**Graphing**

Graph ranked P values

Make these choices the default for future analyses

	Below threshold?	P value	Adjusted P Value
1	Yes	0.0225	0.0445
2	Yes	0.0012	0.0036
3	No	0.5006	0.5006

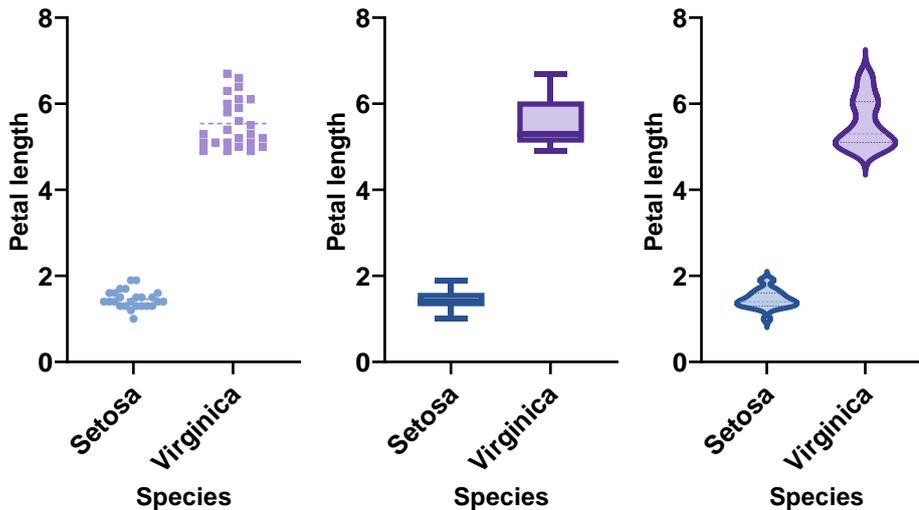


- **Answer 2:** The proportion of cane toads infected by intestinal parasites is significantly lower in Bowden than Mackay and Rockhampton (adjusted  $p=0.0035$  and  $0.0445$ , respectively).

## Exercise 7: Mixed

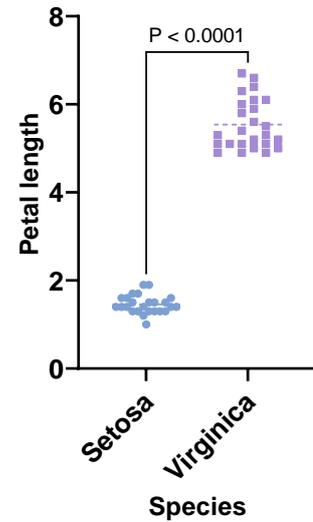
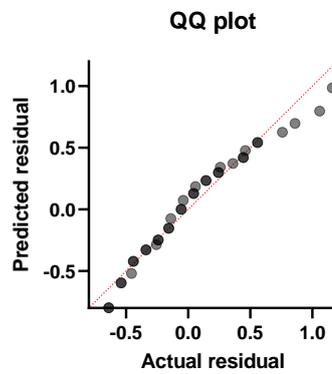
### Iris flowers

- You want to know whether the mean petal length of iris flowers differs according to their species. You find two different species of irises growing in a garden and measure 25 petals of each species.
- This looks very clear cut, so you hardly need to do stats at all! However, to follow the process, the data look like they might be deviating a little from normality, with the Virginia species being fairly skewed and them failing three of the normality tests. The QQ plot doesn't look too bad but I have opted for non-parametric to be safe – in this instance either is going to give a highly significant result anyway.
- Answer:** the Virginia species of iris has significantly longer petal length than the Setosa species ( $P < 0.0001$ )

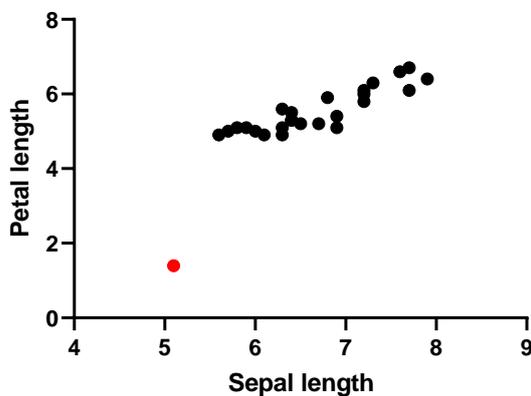


F test to compare variances				
F, DFn, Dfd	7.616, 24, 24			
P value	<0.0001			
P value summary	****			
Significantly different ( $P < 0.05$ )?	Yes			
Normality of Residuals				
Test name	Statistics	P value	Passed normality test (alpha=0.05)?	P value summary
Anderson-Darling (A2*)	0.7652	0.0436	No	*
D'Agostino-Pearson omnibus (K2)	6.184	0.0454	No	*
Shapiro-Wilk (W)	0.9459	0.0231	No	*
Kolmogorov-Smirnov (distance)	0.1237	0.0538	Yes	ns

Mann-Whitney test		
1	Table Analyzed	Petal length
2		
3	Column B	Virginica
4	vs.	vs.
5	Column A	Setosa
6		
7	<b>Mann-Whitney test</b>	
8	P value	<0.0001
9	Exact or approximate P value?	Exact
10	P value summary	****
11	Significantly different (P < 0.05)	Yes
12	One- or two-tailed P value?	Two-tailed
13	Sum of ranks in column A,B	325 , 950
14	Mann-Whitney U	0
15		
16	<b>Difference between medians</b>	
17	Median of column A	1.400, n=25
18	Median of column B	5.300, n=25
19	Difference: Actual	3.900
20	Difference: Hodges-Lehmann	3.900



- You want to know whether the mean petal length of iris flowers correlates with the mean sepal length in the virginica species.
- Again, we want to plot our data – there looks to be a strong outlier so we will test for this, by selecting nonlinear regression and fitting a straight line, so we can add the option to report outliers.



Parameters: Nonlinear Regression

Model Method Compare Constrain Initial values Range Output Confidence Diagnostics Flag

Choose an equation

- Binding - Saturation
- Binding - Competitive
- Binding - Kinetics
- Enzyme kinetics - Inhibition
- Enzyme kinetics - Velocity as a function of substrate
- Exponential
- Lines
  - Straight line
  - Line through point (X0, Y0)
  - Line through origin
  - Horizontal line
  - Semilog line -- X is log, Y is linear
  - Semilog line -- X is linear, Y is log
  - Log-log line -- X and Y both log
  - Segmental linear regression
  - Continuous hinge function. Segmental regression lines with gentle conr
  - Cumulative Gaussian -- Percentages
  - Cumulative Gaussian -- Fractions
  - Two intersecting lines. Fit the crossing point.

Same as linear regression, but with opportunity to compare models, apply weighting, use robust regression and automatically remove outliers.

Straight line  
Analytical derivatives [Learn about this equation](#)

Interpolate  
 Interpolate unknowns from standard curve. Confidence interval: None

Learn Cancel OK

Model Method Compare Constrain Initial values Range Output Confidence Diagnostics Flag

**Outliers**

No special handling of outliers

Detect and eliminate outliers

Report the presence of outliers

Q =  %  Create a table of clean data (with outliers removed)

**Fitting method**

Least squares regression. Used most commonly.

Robust regression. Outliers have little impact.

Poisson regression. Y values are counts of objects or events.

Don't fit the curve. Instead plot the curve defined by the initial values of the parameters.

**Convergence criteria**

How strict   Automatically switch to strict convergence when needed

Maximum number of iterations

**Weighting method**

No weighting. Minimize the sum-of-squares of the distances of the points from the curve. Choose when you expect the average distance between points and curve to be unrelated to the value of Y.

Weight by  $1/Y^2$ . Minimize the sum of the squares of the relative distance of the points from the curve. Choose when you expect the average distance between points and curve to be proportional to Y.

Weight by  $1/Y$

**Replicates**

Consider each replicate Y value as an individual point

Only consider the mean Y value of each point

**How to quantify goodness-of-fit?**

R squared  Sy.x  Sum-of-Squares

Adjusted R squared  RMSE  AICc

**Are residuals Gaussian (normal)?**

D'Agostino-Pearson omnibus normality test

Anderson-Darling test

Shapiro-Wilk normality test

Kolmogorov-Smirnov normality test with Dallal-Wilkinson-Lilliefors P value

**Are residuals clustered or heteroscedastic?**

Runs test  Replicates test  Test for appropriate weighting (homoscedasticity)

**Which residual graphs to create?**

Residual vs X plot

Residual vs Y plot

Homoscedasticity plot

QQ plot

Actual vs Predicted plot

**Are the parameters intertwined, redundant or skewed?**

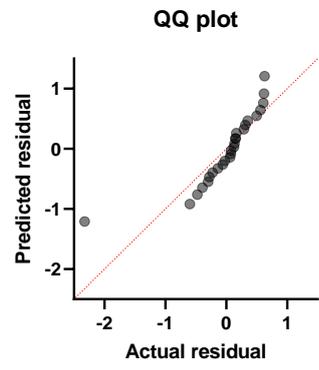
Covariance of parameters

Dependency

Hougaard's measure of skewness

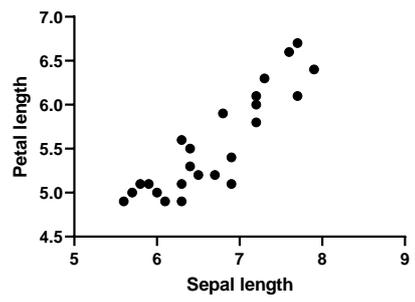
Normality of Residuals	
D'Agostino-Pearson omnibus (K2)	33.55
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****
Anderson-Darling (A2')	1.402
P value	0.0010
Passed normality test (alpha=0.05)?	No
P value summary	***
Shapiro-Wilk (W)	0.7607
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****
Kolmogorov-Smirnov (distance)	0.1568
P value	0.0994
Passed normality test (alpha=0.05)?	Yes
P value summary	ns

	X	A	B
	Sepal length	Petal length	
1	5.100	1.400	

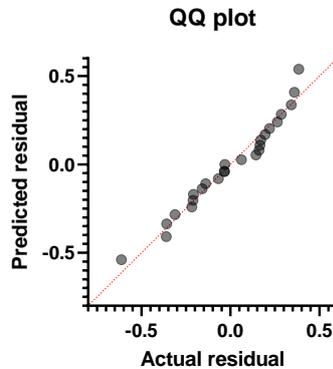


- As expected, the outlier has been detected, and the data are failing normality tests. In this case, the raw data is in the raw.data tab in the Excel file, and we can see that this outlier actually comes from another species of iris. We can comfortably remove that outlier, knowing it was there due to a copying error.

Sepal.Len	Sepal.Wid	Petal.Len	Petal.Wid	Species
5.8	2.7	5.1	1.9	virginica
6.3	2.9	5.6	1.8	virginica
7.6	3	6.6	2.1	virginica
7.3	2.9	6.3	1.8	virginica
7.2	3.6	6.1	2.5	virginica
6.4	2.7	5.3	1.9	virginica
5.7	2.5	5	2	virginica
6.4	3.2	5.3	2.3	virginica
7.7	3.8	6.7	2.2	virginica
6	2.2	5	1.5	virginica
5.6	2.8	4.9	2	virginica
6.3	2.7	4.9	1.8	virginica
7.2	3.2	6	1.8	virginica
6.1	3	4.9	1.8	virginica
7.2	3	5.8	1.6	virginica
7.9	3.8	6.4	2	virginica
6.3	2.8	5.1	1.5	virginica
7.7	3	6.1	2.3	virginica
6.4	3.1	5.5	1.8	virginica
6.9	3.1	5.4	2.1	virginica
6.9	3.1	5.1	2.3	virginica
6.8	3.2	5.9	2.3	virginica
6.7	3	5.2	2.3	virginica
6.5	3	5.2	2	virginica
5.9	3	5.1	1.8	virginica
5.1	3.5	1.4	0.2	setosa



Normality of Residuals	
D'Agostino-Pearson omnibus (K2)	1.116
P value	0.5723
Passed normality test (alpha=0.05)?	Yes
P value summary	ns
Anderson-Darling (A2*)	0.3579
P value	0.4256
Passed normality test (alpha=0.05)?	Yes
P value summary	ns
Shapiro-Wilk (W)	0.9578
P value	0.3720
Passed normality test (alpha=0.05)?	Yes
P value summary	ns
Kolmogorov-Smirnov (distance)	0.1475
P value	>0.1000
Passed normality test (alpha=0.05)?	Yes
P value summary	ns



Parameters: Correlation

**Compute correlation between which pairs of columns?**

Compute r for every pair of Y data sets (Correlation matrix).  
 When a value is missing or excluded, remove the entire row from the calculation

Compute r for X vs. every Y data set:

[X] Sepal length

Compute r between two selected data sets:

[X] Sepal length

[A] Petal length

**Assume data are sampled from Gaussian distribution?**

Yes. Compute Pearson correlation coefficients.

No. Compute nonparametric Spearman correlation.

**Options**

P value:  One-tailed  Two-tailed

Confidence interval: 95%

**Output**

Show this many significant digits (for everything except P values): 4

P value style: GP: 0.1234 (ns), 0.0332 (\*), 0.0021 (\*\*), N = 6

**Graphing**

Create a heatmap of the correlation matrix.

Make these choices the default for future analyses

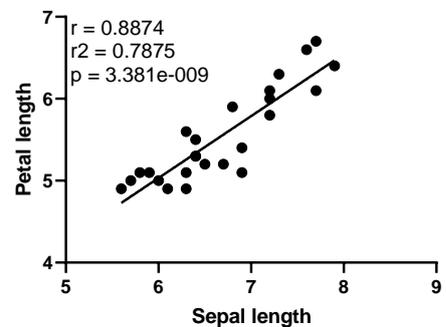
Learn Cancel OK

Correlation	
	Sepal length vs. Petal length
<b>Pearson r</b>	
r	0.8874
95% confidence interval	0.7581 to 0.9496
R squared	0.7875
<b>P value</b>	
P (two-tailed)	<0.0001
P value summary	****
Significant? (alpha = 0.05)	Yes
<b>Number of XY Pairs</b>	25

Hook Constant

Choose a value from an info or analysis constant. This will be a 'hot link' to be automatically updated in the future.

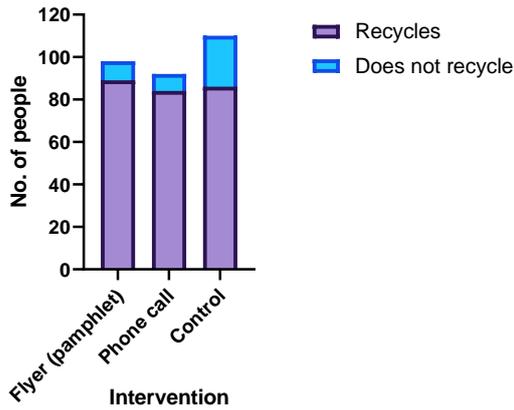
- Correlation of Virginia\_sepals\_petal
- Pearson r
  - # Sepal length vs. Petal length = 0.8874
  - Lower confidence limit
  - Upper confidence limit
  - R squared
  - # Sepal length vs. Petal length = 0.7875
  - P value
  - # Sepal length vs. Petal length = 3.381e-009
  - Number of XY Pairs



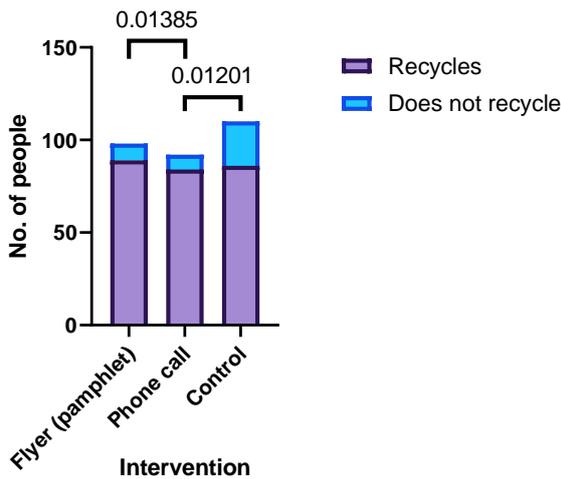
- After removing the outlier, the data look much more normal, and we can continue to calculate the correlation. **There is a significant relationship between petal length and sepal length ( $\rho=0.89$ ,  $p<0.001$ ).**

## Recycling

- *A city wants to encourage more of its residents to recycle their household waste. The city decides to test two interventions: an educational flyer (pamphlet) or a phone call. Which intervention should the city use to maximise household recycling?*



Contingency		A
1	Table Analyzed	Recycling
2		
3	P value and statistical significance	
4	Test	Chi-square
5	Chi-square, df	9.791, 2
6	P value	0.0075
7	P value summary	**
8	One- or two-sided	NA
9	Statistically significant (P < 0.05)?	Yes
10		
11	Data analyzed	
12	Number of rows	3
13	Number of columns	2

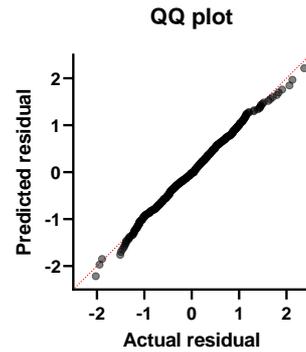
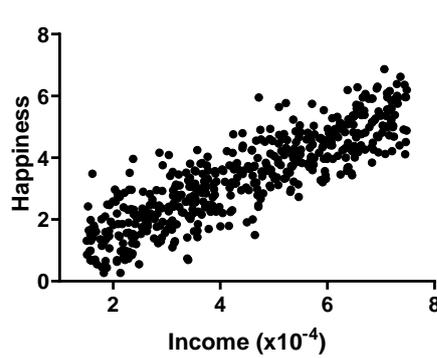


Adjusted P values		P small enough?	
	Below threshold?	P value	Adjusted P Value
1	Flyer v control	Yes	0.0138 0.0356
2	Phone call v contro	Yes	0.0120 0.0356
3	Flyer v phone call	No	0.9999 0.9999

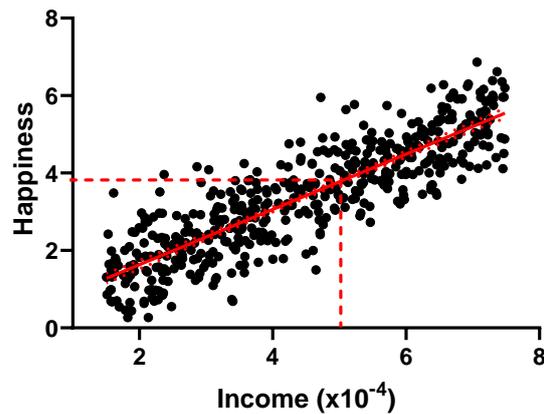
- **Answer:** There is a significant difference between our groups ( $p=0.0075$ ). Both interventions are significantly better than the control group (adjusted  $p=0.0356$  for both). Therefore, choice of intervention should come down to other considerations, incl. cost, resources, time, etc.

### Income data

- You are a social researcher interested in the relationship between income and happiness. You survey 500 people whose incomes range from 15k to 75k and ask them to rank their happiness on a scale from 1 to 10.
- Generate a linear model describing the relationship between income and happiness
- How happy would you expect someone earning 50k to be, based on your model?



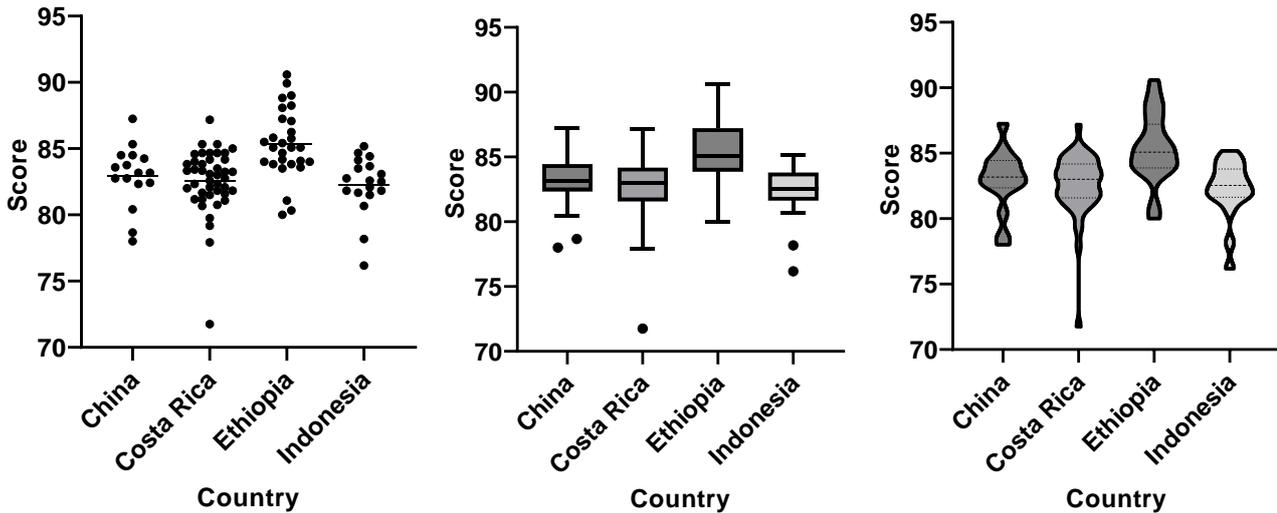
Nonlin fit		A
Table of results		happiness
1	<b>Straight line</b>	
2	<b>Best-fit values</b>	
3	YIntercept	0.2043
4	Slope	0.7138
5	<b>95% CI (profile likelihood)</b>	
6	YIntercept	0.02972 to 0.3788
7	Slope	0.6774 to 0.7502
8	<b>Goodness of Fit</b>	
9	Degrees of Freedom	496
10	R squared	0.7493
11	Sum of Squares	255.8
12	Sy.x	0.7181
13	<b>Normality of Residuals</b>	
14	<b>D'Agostino-Pearson omnibus (K2)</b>	1.396
15	P value	0.4977
16	Passed normality test (alpha=0.05)?	Yes
17	P value summary	ns
18	<b>Anderson-Darling (A2')</b>	0.3191
19	P value	0.5337
20	Passed normality test (alpha=0.05)?	Yes
21	P value summary	ns
22	<b>Shapiro-Wilk (W)</b>	0.9968
23	P value	0.4237
24	Passed normality test (alpha=0.05)?	Yes
25	P value summary	ns
26	<b>Kolmogorov-Smirnov (distance)</b>	0.02568
27	P value	>0.1000
28	Passed normality test (alpha=0.05)?	Yes
29	P value summary	ns



- Answer: From the linear regression  $\text{happiness} = 0.20 + 0.71 \cdot \text{income}(x10^{-4})$  [y-int + slope\*income] so if earning 50k, happiness =  $0.2 + 0.71 \cdot 5 = 3.75$ .

## Coffee – part 2

- You enjoyed your previous coffee so much that you decide to do some more research, this time looking at which country the beans came from. Your local coffee shop only sells beans from Ethiopia, Costa Rica, China, and Indonesia – which beans should you buy?

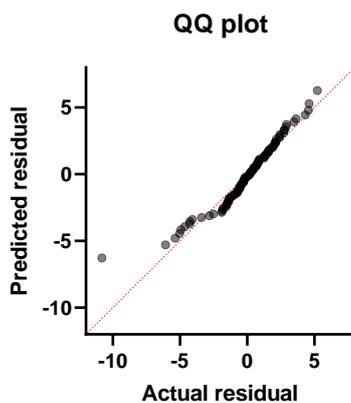


- This dataset again has a question of whether or not to remove outliers. In this case, it seems likely that the outliers are real results, as they are not implausible and it is based on a scoring system. Therefore, doing the analysis with the outliers included is probably preferable, unless there is a reason to exclude them or there was a plan to exclude any identified outliers set prior to data analysis.

**With outliers**

- Including the potential outliers, the data fail normality tests and, although the QQ plot doesn't look too far off, it doesn't seem to follow the line as well as we would like, particularly with the potential outlier. Also, given that this is based on a scoring system, where non-parametric tests are often used, I have opted for a Kruskal-Wallis test.

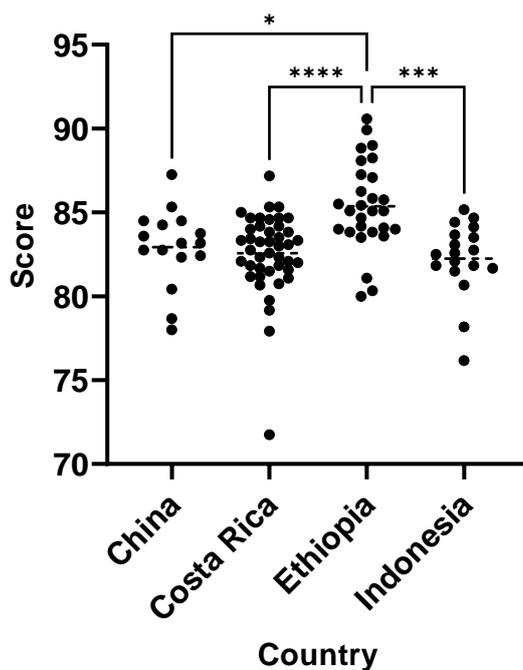
Normality of Residuals				
Test name	Statistics	P value	Passed normality test (alpha=0.05)?	P value summary
D'Agostino-Pearson omnibus (K2)	29.47	<0.0001	No	****
Anderson-Darling (A2*)	1.473	0.0008	No	***
Shapiro-Wilk (W)	0.9368	<0.0001	No	****
Kolmogorov-Smirnov (distance)	0.1053	0.0060	No	**



Kruskal-Wallis test ANOVA results	
Table Analyzed	Coffee_country
<b>Kruskal-Wallis test</b>	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Do the medians vary signif. (P < 0.05)?	Yes
Number of groups	4
Kruskal-Wallis statistic	23.65
<b>Data summary</b>	
Number of treatments (columns)	4
Number of values (total)	105

Dunn's multiple comparisons test	Mean rank diff.	Significant?	Summary	Adjusted P Value	
China vs. Costa Rica	4.688	No	ns	>0.9999	A-B
China vs. Ethiopia	-27.38	Yes	*	0.0246	A-C
China vs. Indonesia	9.160	No	ns	>0.9999	A-D
Costa Rica vs. Ethiopia	-32.07	Yes	****	<0.0001	B-C
Costa Rica vs. Indonesia	4.472	No	ns	>0.9999	B-D
Ethiopia vs. Indonesia	36.54	Yes	***	0.0004	C-D

- **Answer:** Coffee beans from Ethiopia receive significantly higher scores than those from China ( $p=0.025$ ), Costa Rica ( $p<0.0001$ ), and Indonesia ( $p<0.0001$ ). There are no significant differences between the scores for the other three countries.



Kruskal-Wallis test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Do the medians vary signif. ( $P < 0.05$ )?	Yes
Number of groups	4
Kruskal-Wallis statistic	23.65

### Without outliers

- For completeness, I have included the analysis including the outliers below, this time using a parametric approach as the data look more normal with the outliers removed. We get very similar results using both approaches, so can be confident that our approach is not impacting the results.
- **Answer:** Coffee beans from Ethiopia receive significantly higher scores than those from China ( $p=0.0037$ ), Costa Rica ( $p<0.0001$ ), and Indonesia ( $p<0.0001$ ). There are no significant differences between the scores for the other three countries.

Analyze Data

Use: Built-in analysis

Which analysis?

Analyze which data sets? Table: Coffee\_country

- A:Ethiopia
- B:Costa Rica
- C:China
- D:Indonesia

Recently used

- Transform, Normalize...
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of total
- XY analyses
- Column analyses
  - t tests (and nonparametric tests)
  - One-way ANOVA (and nonparametric or
  - One sample t and Wilcoxon test
  - Descriptive statistics
  - Normality and Lognormality Tests
  - Frequency distribution
  - ROC Curve
  - Bland-Altman method comparison
  - Identify outliers
  - Analyze a stack of P values

Select All Deselect All

Help Cancel OK

Parameters: Identify Outliers

Method

- ROUT (recommended; can find any number of outliers)
- Grubbs' (can only find one outlier)
- Iterative Grubbs' (can find several; not recommended)

How aggressive?

Remove definitive outliers  Remove likely outliers

Subcolumns

- Average the replicates in each row, and then perform the calculation for each column
- Perform the calculation for each subcolumn separately
- Treat all the values in all the subcolumns as one set of data

Make these choices the default for future analyses

Learn Cancel OK

	A	B	C	D	E	F
	China	Costa Rica	Ethiopia	Indonesia		
1	84.500	84.670	85.080	81.670		

	A	B	C	D
	China	Costa Rica	Ethiopia	Indonesia
1				
2				
3				
4				
5	16	43	28	18
6	0	1	0	0

Analyze Data

Use: Built-in analysis

Which analysis?

Analyze which data sets? Table: Identify outliers of Coffee\_country\_

- A:China
- B:Costa Rica
- C:Ethiopia
- D:Indonesia

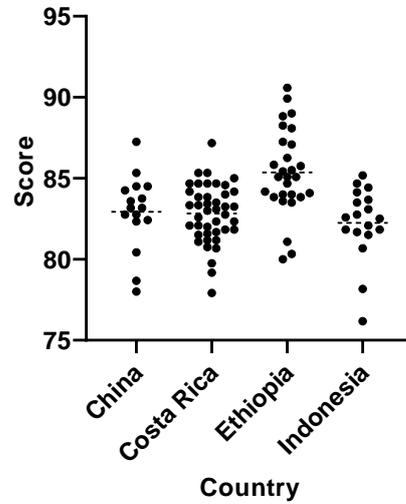
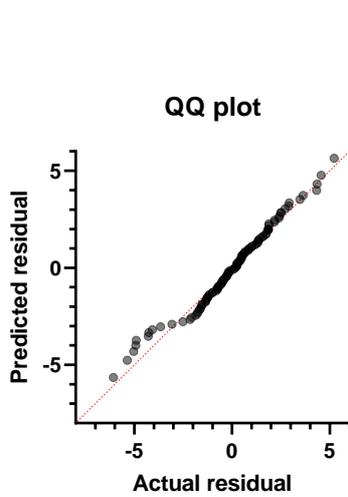
Recently used

- Transform, Normalize...
- XY analyses
- Column analyses
  - t tests (and nonparametric tests)
  - One-way ANOVA (and nonparametric or
  - One sample t and Wilcoxon test
  - Descriptive statistics
  - Normality and Lognormality Tests
  - Frequency distribution
  - ROC Curve
  - Bland-Altman method comparison
  - Identify outliers
  - Analyze a stack of P values

Select All Deselect All

Help Cancel OK

Normality of Residuals				
Test name	Statistics	P value	Passed normality test (alpha=0.05)?	P value summary
D'Agostino-Pearson omnibus (K2)	4.470	0.1070	Yes	ns
Anderson-Darling (A2*)	0.8174	0.0335	No	*
Shapiro-Wilk (W)	0.9751	0.0470	No	*
Kolmogorov-Smirnov (distance)	0.07259	0.1000	Yes	ns



ANOVA summary					
F	10.16				
P value	<0.0001				
P value summary	****				
Significant diff. among means (P < 0.05)	Yes				
R squared	0.2335				
<b>Brown-Forsythe test</b>					
F (DFn, DFd)	0.8353 (3, 100)				
P value	0.4776				
P value summary	ns				
Are SDs significantly different (P < 0.05)	No				
<b>Bartlett's test</b>					
Bartlett's statistic (corrected)	4.727				
P value	0.1929				
P value summary	ns				
Are SDs significantly different (P < 0.05)	No				
<b>ANOVA table</b>					
	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	149.1	3	49.70	F (3, 100) = 10.16	P<0.0001
Residual (within columns)	489.3	100	4.893		
Total	638.4	103			

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value	
China vs. Costa Rica	0.1063	-1.592 to 1.804	No	ns	0.9984	A-B
China vs. Ethiopia	-2.431	-4.243 to -0.6201	Yes	**	0.0037	A-C
China vs. Indonesia	0.6836	-1.302 to 2.669	No	ns	0.8051	A-D
Costa Rica vs. Ethiopia	-2.538	-3.948 to -1.128	Yes	****	<0.0001	B-C
Costa Rica vs. Indonesia	0.5773	-1.051 to 2.206	No	ns	0.7908	B-D
Ethiopia vs. Indonesia	3.115	1.369 to 4.861	Yes	****	<0.0001	C-D

