Introduction to Statistics with GraphPad Prism 8

Anne Segonds-Pichon
v2019-03

Outline of the course

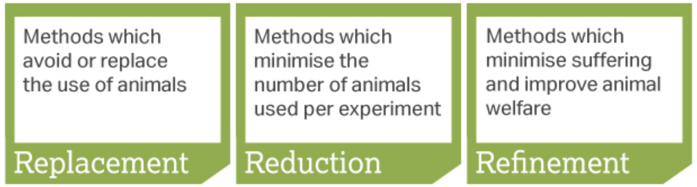
- Power analysis with **G*Power**
- Basic structure of a GraphPad Prism project
- Analysis of qualitative data:
  - **Chi-square test**
- Analysis of quantitative data:
  - **Student's *t*-test, One-way ANOVA, correlation and curve fitting**

Power analysis

- **Definition of power**: probability that a statistical test will reject a false null hypothesis (H_0).
    - **Translation**: the probability of detecting an effect, given that the effect is really there.

- **In a nutshell**: the bigger the experiment (big sample size), the bigger the power (more likely to pick up a difference).

- Main output of a **power analysis**:
    - Estimation of an appropriate **sample size**
        - **Too big**: waste of resources,
        - **Too small**: may miss the effect (p>0.05)+ waste of resources,
        - **Grants**: justification of sample size,
        - **Publications:** reviewers ask for power calculation evidence,
        - **Home office**: the 3 Rs: Replacement, **Reduction** and Refinement.

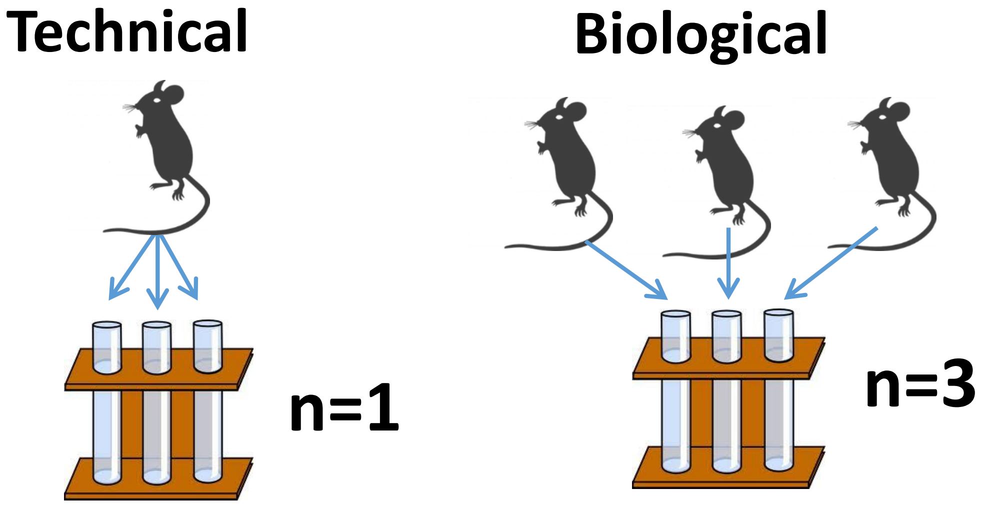


Replacement	Reduction	Refinement
Methods which avoid or replace the use of animals	Methods which minimise the number of animals used per experiment	Methods which minimise suffering and improve animal welfare

Experimental design

Think stats!!

- Translate the hypothesis into statistical questions:
    - What type of data?
    - What statistical test ?
    - **What sample size?**
- Very important: Difference between **technical** and **biological** replicates.



Power analysis

A power analysis depends on the relationship between 6 variables:

- the **difference** of biological interest
- the **variability** in the data (**standard deviation**)

} **Effect size**

- the significance level (5%)
- the desired power of the experiment (80%)
- the **sample size**
- the alternative hypothesis (ie one or two-sided test)

1 The difference of biological interest

- This is to be determined scientifically, not statistically.
  - minimum meaningful effect of biological relevance
  - the larger the effect size, the smaller the experiment will need to be to detect it.
- **How to determine it?**
  - Substantive knowledge, previous research, pilot study …

2 The Standard Deviation (SD)

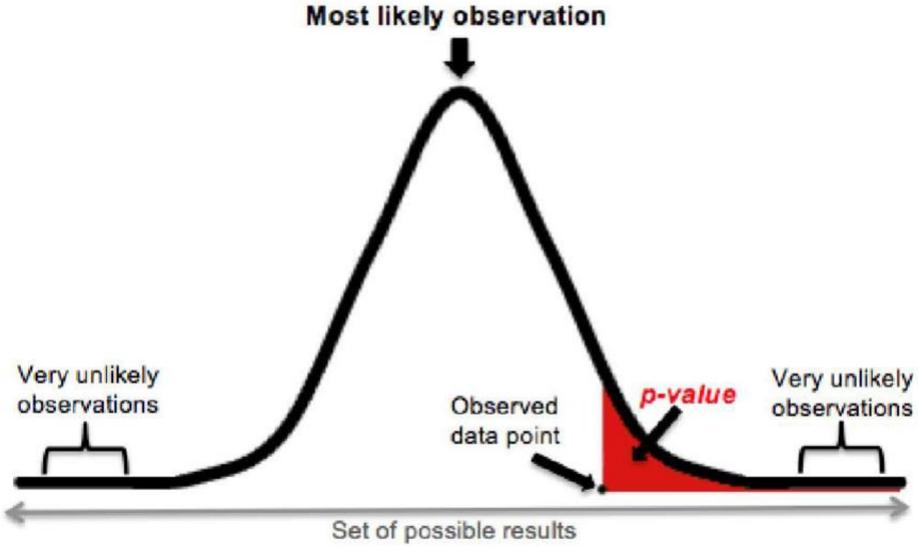
- Variability of the data
- **How to determine it?**
  - Substantive knowledge, previous research, pilot study …

- In ‘power context’: **effect size**: combination of both:
  - e.g.: **Cohen’s d** = (Mean 1 – Mean 2)/Pooled SD

3 The significance level

- usually **5%** (**$p<0.05$**), probability of the **Type I error α**

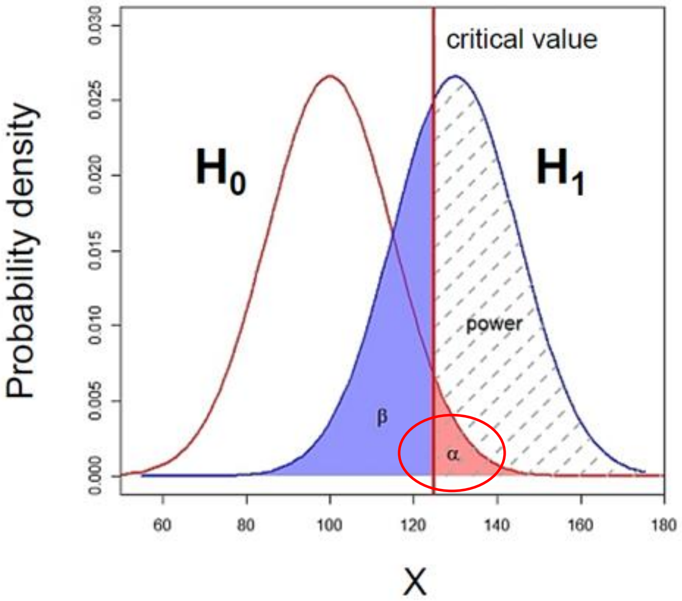
- **p-value** is the probability that a difference as big as the one observed could be found even if there is no effect.
  - Probability that an effect occurs by chance alone



A *p-value* (shaded red area) is the probability of an observed (or more extreme) result arising by chance

- Don’t throw away a p-value=0.051 !

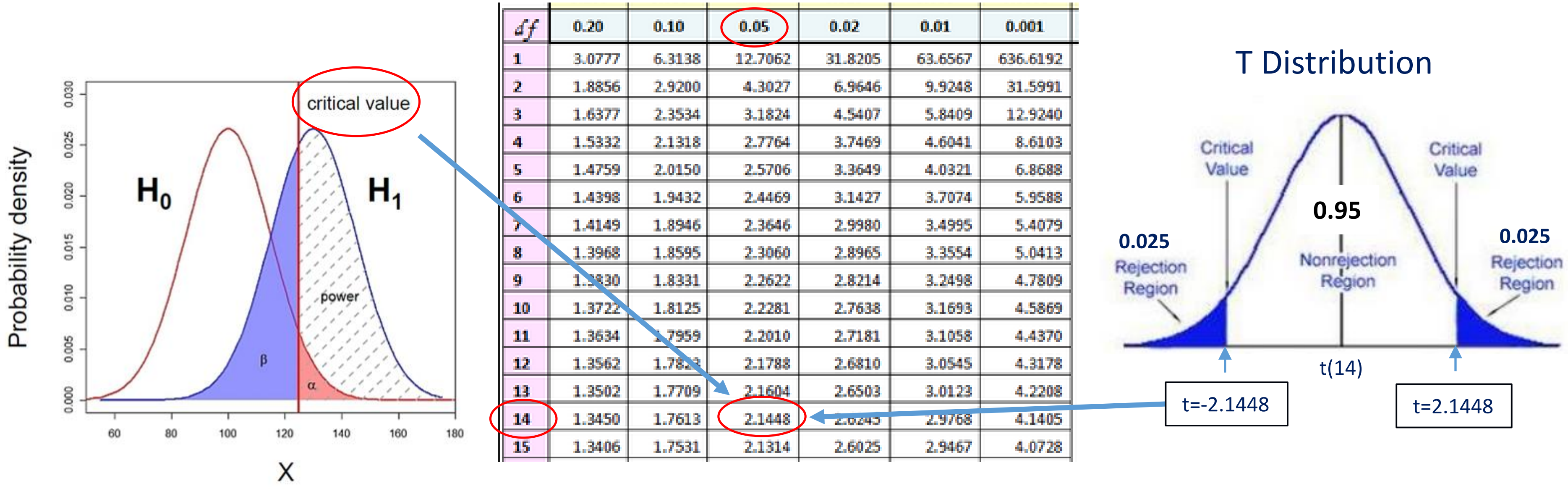
The significance level, critical value, α and β



- **α :** the threshold value that we measure p-values against.
  - For results with 95% level of confidence: $\alpha = 0.05$
  - = probability of **type I error**
- **p-value**: probability that the observed statistic occurred by chance alone
- **Statistical significance**: comparison between **α** and the **p-value**
  - p-value < 0.05: reject H_0 and p-value > 0.05: fail to reject H_0

The critical value

Example: 2-tailed t-test with n=15 (df=14)



df	0.20	0.10	0.05	0.02	0.01	0.001
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370
12	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178
13	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208
14	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405
15	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728

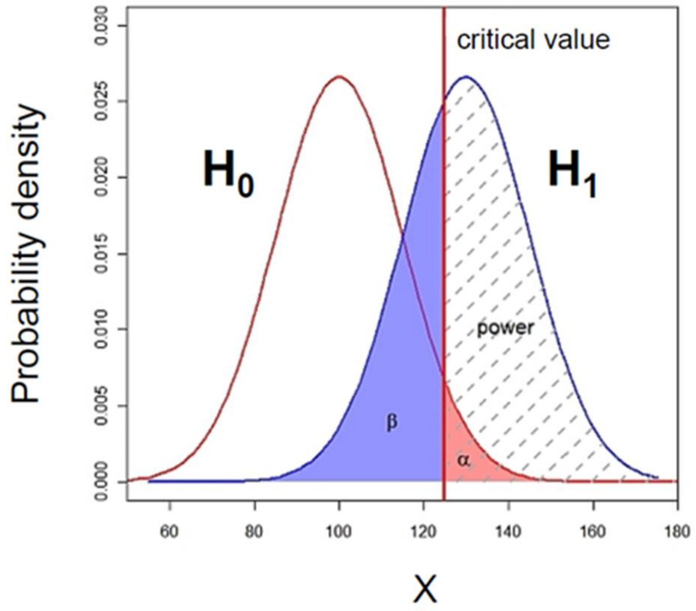
T Distribution

t=-2.1448

t=2.1448

- In **hypothesis testing**, a **critical value** is a point on the test distribution that is compared to the **test statistic** to determine whether to reject the null **hypothesis**
  - Example of test statistic: t-value
- If the absolute value of your **test statistic** is greater than the **critical value**, you can declare statistical significance and reject the null **hypothesis**
  - Example: t-value > critical t-value

4 The desired power: 80%



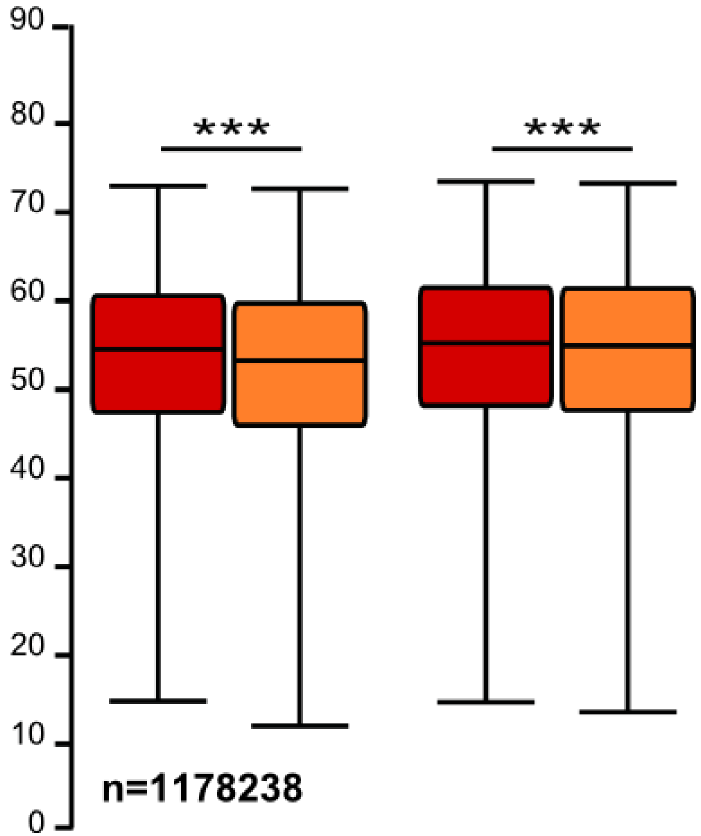
- **Type II error** (**β**) is the failure to reject a <u>false</u> H_0
  - Direct relationship between **Power** and type II error:
    - if β = 0.2 and **Power** = 1 – β = 0.8 (80%)
  - Hence a true difference will be missed 20% of the time
  - General convention: 80% but could be more or less
  - Cohen (1988):
    - For most researchers: Type I errors are four times more serious than Type II errors: 0.05 * 4 = 0.2
    - Compromise: 2 groups comparisons: 90% = +30% sample size, 95% = +60%

5 The sample size: the bigger the better?

- It takes huge samples to detect tiny differences but tiny samples to detect huge differences.

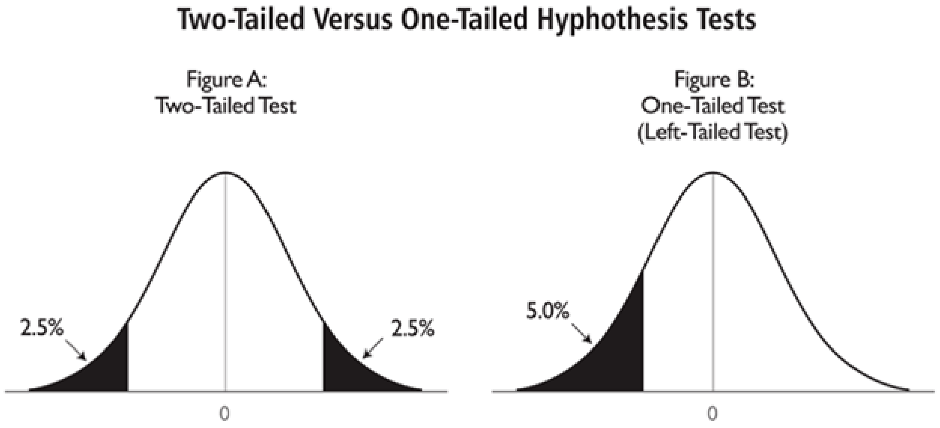
- What if the tiny difference is meaningless?
  - Beware of **overpower**
  - Nothing wrong with the stats: it is all about interpretation of the results of the test.

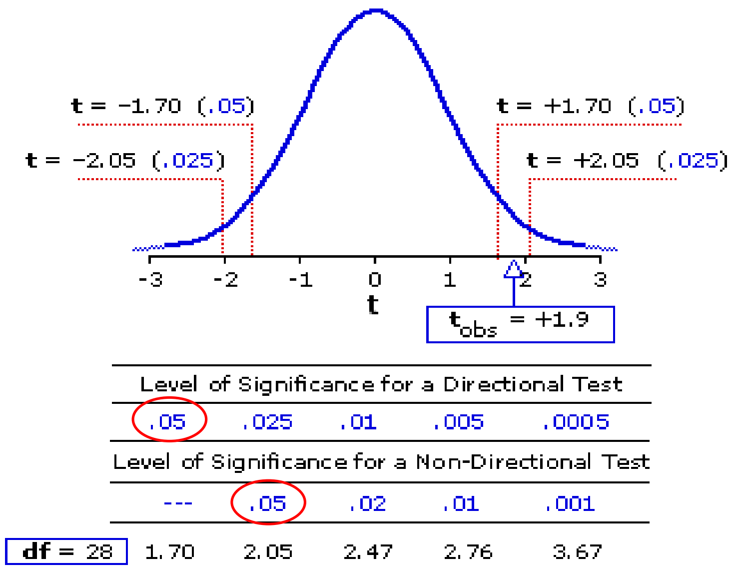
- Remember the important first step of power analysis
  - **What is the effect size of biological interest?**



6 The alternative hypothesis

- One-tailed or 2-tailed test? One-sided or 2-sided tests?

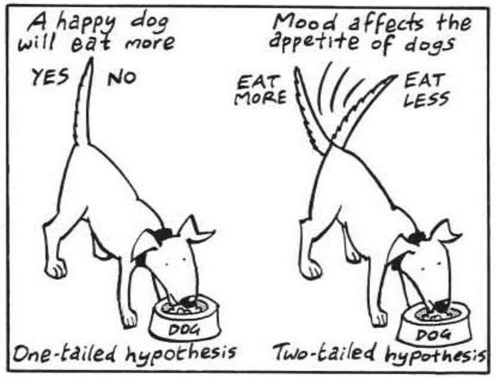

Two-Tailed Versus One-Tailed Hyphothesis Tests



Level of Significance for a Directional Test	.05	.025	.01	.005	.0005
Level of Significance for a Non-Directional Test	---	.05	.02	.01	.001
df = 28	1.70	2.05	2.47	2.76	3.67

- Is the question:
  - Is the there a difference?
  - Is it bigger than or smaller than?

- Can rarely justify the use of a one-tailed test
- Two times easier to reach significance with a one-tailed than a two-tailed
  - Suspicious reviewer!



To recapitulate:

- The null hypothesis (H_0): H_0 = no effect

- The aim of a statistical test is to reject or not H_0.

Statistical decision	True state of H_0	
	H_0 True (no effect)	**H_0 False (effect)**
Reject H_0	**Type I error α** **False Positive** ☹	Correct **True Positive** ☺
Do not reject H_0	Correct **True Negative** ☺	**Type II error β** **False Negative** ☹

- Traditionally, a test or a difference are said to be "**significant**" if the probability of type I error is: **$\alpha =< 0.05$**
- **High specificity** = low **False Positives** = low **Type I error**
- **High sensitivity** = low **False Negatives** = low **Type II error**

Hypothesis

↓

Experimental design

Choice of a Statistical test

↓

Power analysis: Sample size

↓

Experiment(s)

↓

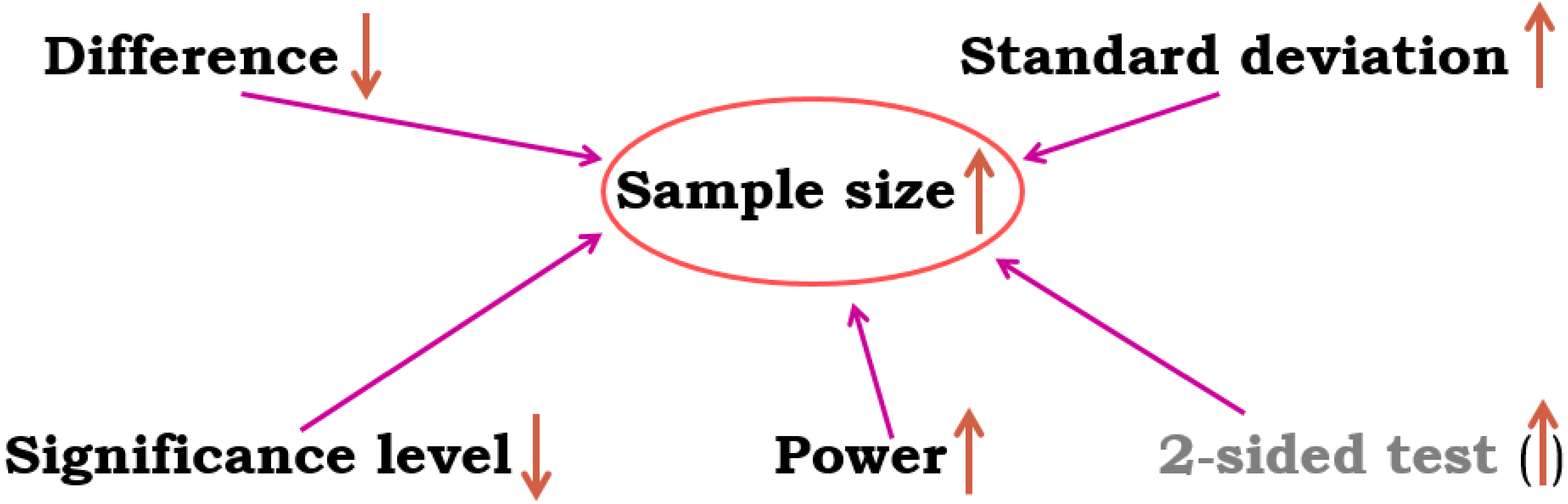
Data exploration

↓

Statistical analysis of the results

- **Fix any five of the variables and a mathematical relationship can be used to estimate the sixth.**

e.g. What sample size do I need to have a 80% probability (**power**) to detect this particular effect (**difference** and **standard deviation**) at a 5% **significance level** using a **2-sided test**?



Difference ↓

Standard deviation ↑

Sample size ↑

Significance level ↓

Power ↑

2-sided test (↑)

- **Good news**:

there are packages that can do the power analysis for you ... providing you have some prior knowledge of the key parameters!

difference + standard deviation = effect size

- **Free packages**:
  - R
  - **G*Power** and InVivoStat
  - Russ Lenth's power and sample-size page:
    - http://www.divms.uiowa.edu/~rlenth/Power/

- Cheap package: StatMate (~ $95)

- Not so cheap package: MedCalc (~ $495)

Sample →

Statistical inference

→ Population

Difference → Meaningful? → Yes → Real?

Statistical test

Statistic
e.g. t, F …

Big enough?

=

Difference + Noise + Sample

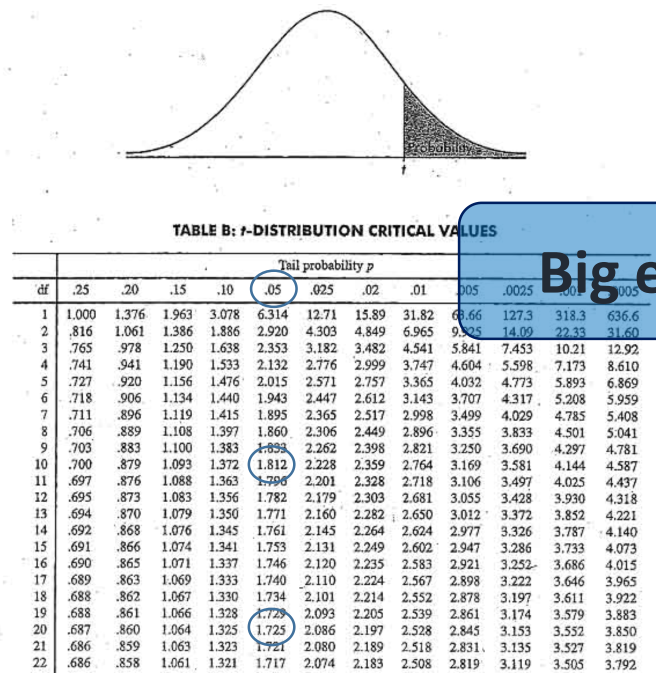


TABLE B: t-DISTRIBUTION CRITICAL VALUES

df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
	Tail probability p											
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792

Qualitative data

Qualitative data

- = not numerical

- = values taken = usually names (also *nominal*)
    - e.g. causes of death in hospital

- Values can be numbers but not numerical
    - e.g. group number = numerical label but not unit of measurement

- Qualitative variable with intrinsic order in their categories = *ordinal*

- Particular case: qualitative variable with 2 categories: ***binary*** or *dichotomous*
    - e.g. alive/dead or presence/absence

Fisher's exact and Chi^2

Example: **cats and dogs.xlsx**

- Cats and dogs trained to line dance
- 2 different rewards: food or affection
- **Question**: Is there a difference between the rewards?

- **Is there a significant relationship between the 2 variables?**
  - does the reward significantly affect the likelihood of dancing?

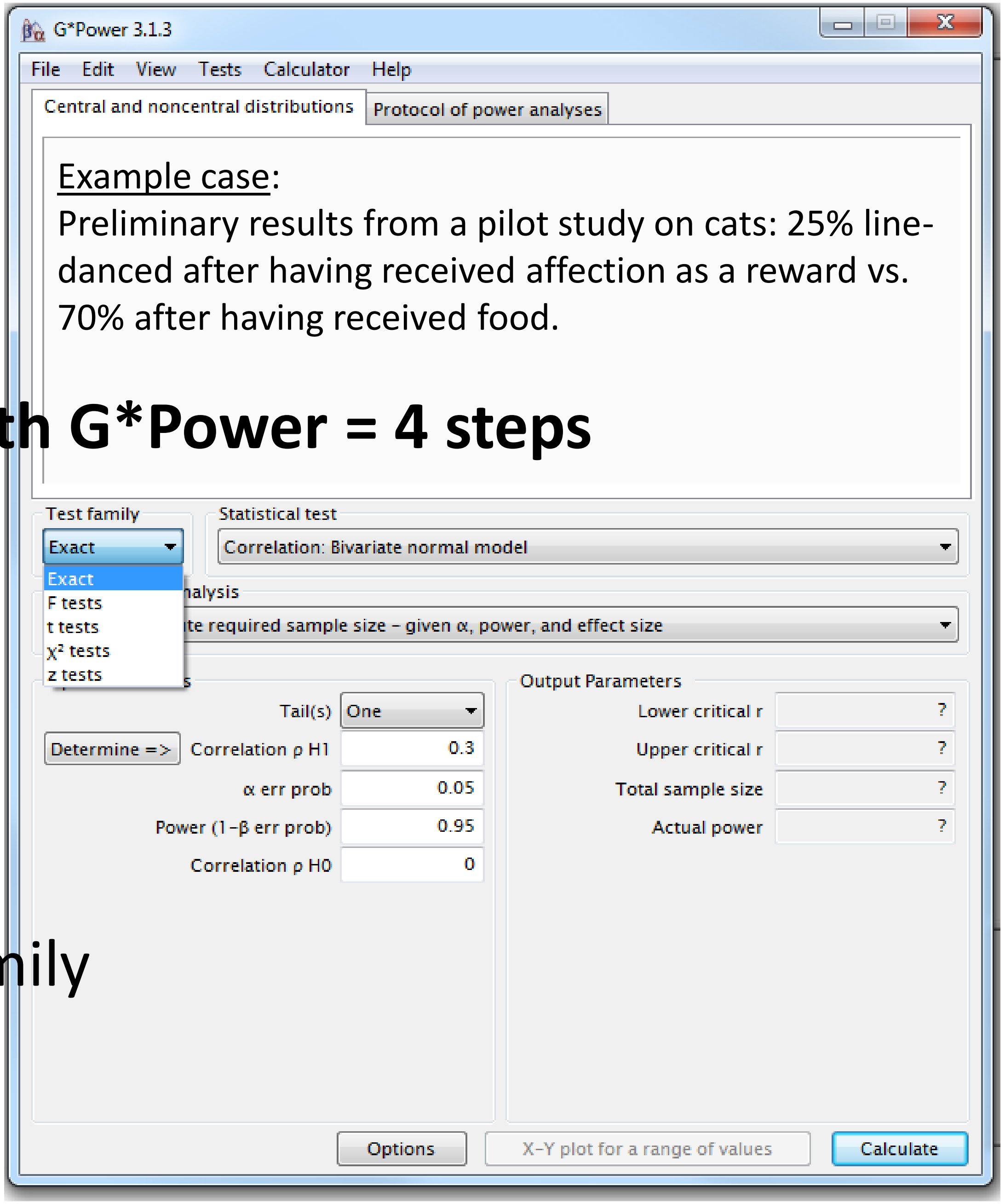
- To answer this type of question:
  - **Contingency table**
  - **Fisher's exact or Chi^2 tests**

	Food	Affection
Dance	?	?
No dance	?	?

But first: **how many cats** do we need?

G*Power

A priori Power Analysis



Example case:

Preliminary results from a pilot study on cats: 25% line-danced after having received affection as a reward vs. 70% after having received food.

Power analysis with G*Power = 4 steps

Test family: Exact

Statistical test: Correlation: Bivariate normal model

alysis: te required sample size – given α, power, and effect size

Tail(s): One

Correlation ρ H1: 0.3

α err prob: 0.05

Power (1–β err prob): 0.95

Correlation ρ H0: 0

Output Parameters

Lower critical r: ?

Upper critical r: ?

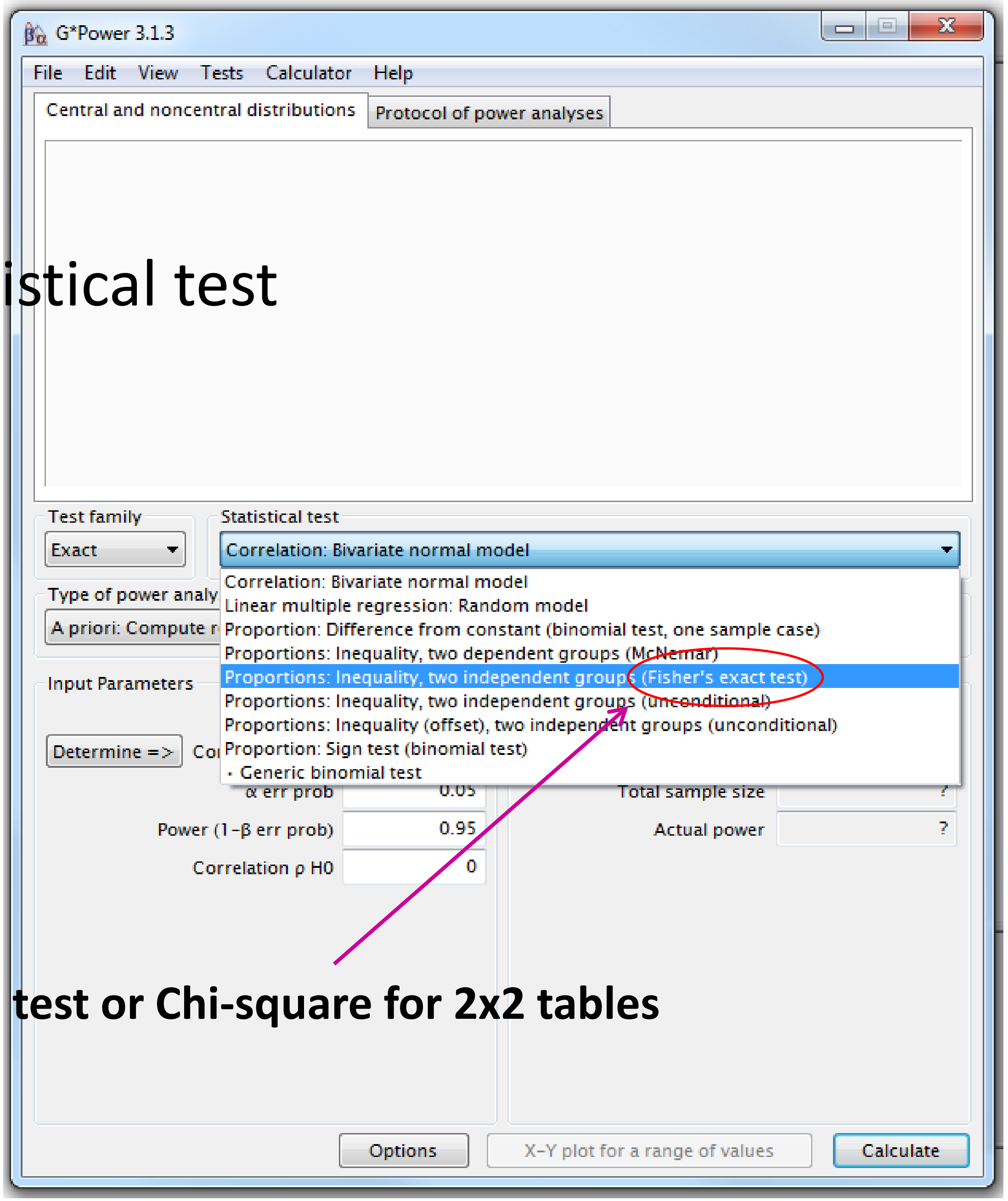
Total sample size: ?

Actual power: ?

Step1: choice of Test family

G*Power

Step 2 : choice of Statistical test



Test family: Exact

Statistical test: Correlation: Bivariate normal model

Type of power analy: A priori: Compute r

α err prob: 0.05

Power (1-β err prob): 0.95

Correlation ρ H0: 0

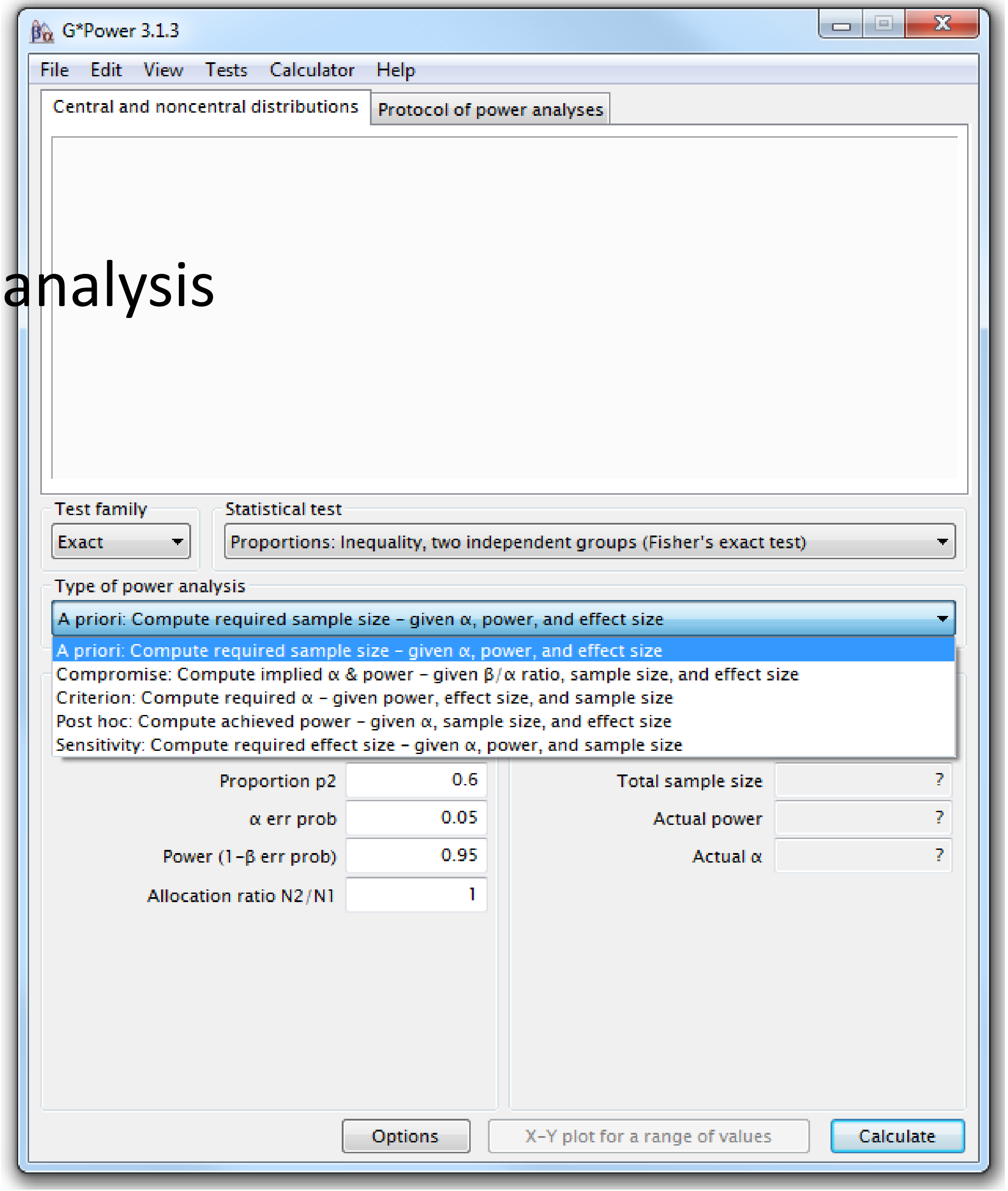
Total sample size: ?

Actual power: ?

Fisher's exact test or Chi-square for 2x2 tables

G*Power

Step 3: Type of power analysis



Test family: Exact

Statistical test: Proportions: Inequality, two independent groups (Fisher's exact test)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

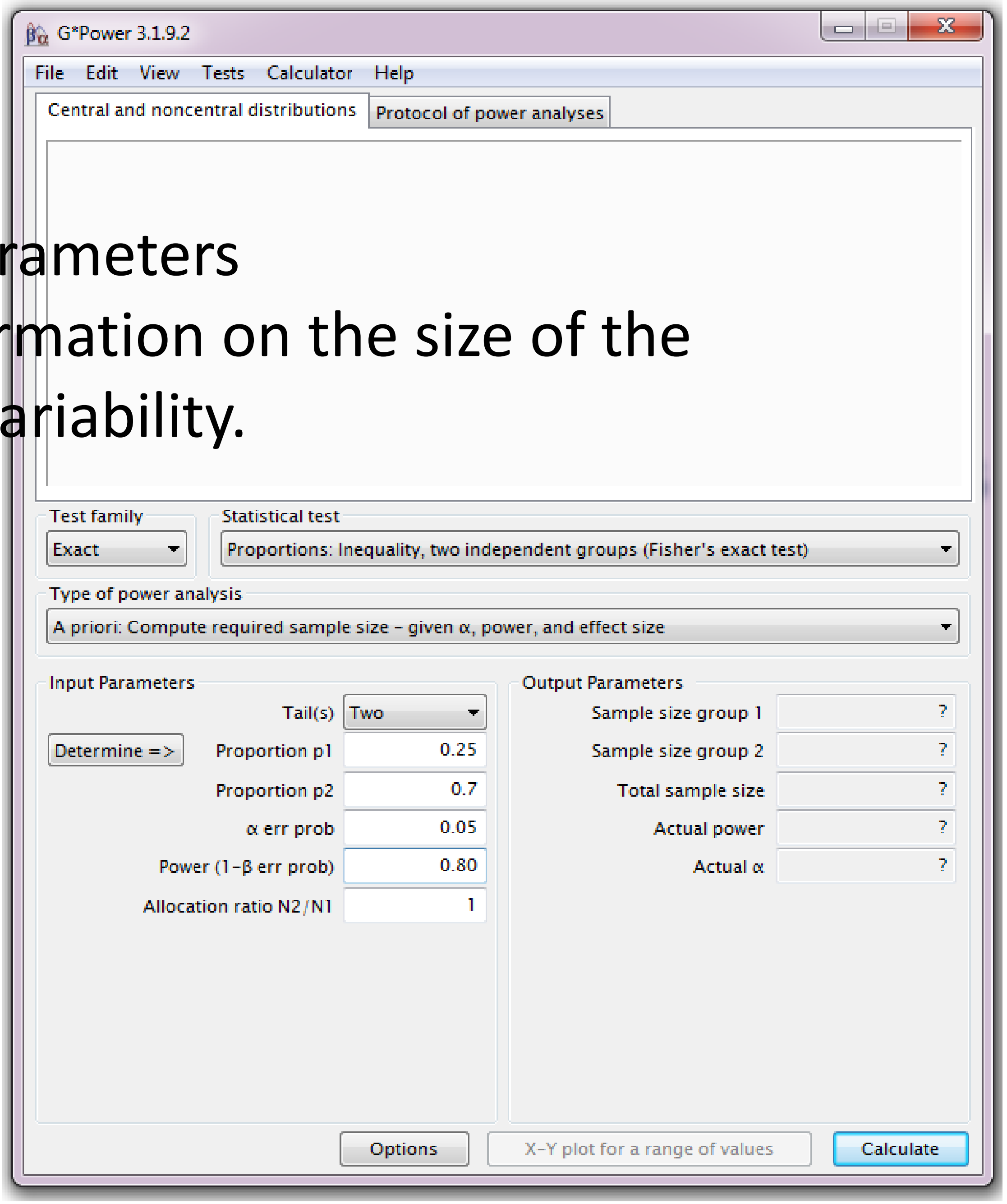
- A priori: Compute required sample size – given α, power, and effect size
- Compromise: Compute implied α & power – given β/α ratio, sample size, and effect size
- Criterion: Compute required α – given power, effect size, and sample size
- Post hoc: Compute achieved power – given α, sample size, and effect size
- Sensitivity: Compute required effect size – given α, power, and sample size

Proportion p2	0.6	Total sample size	?
α err prob	0.05	Actual power	?
Power (1−β err prob)	0.95	Actual α	?
Allocation ratio N2/N1	1		

G*Power

Step 4: Choice of Parameters

Tricky bit: need information on the size of the difference and the variability.



Test family: Exact

Statistical test: Proportions: Inequality, two independent groups (Fisher's exact test)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

Tail(s)	Two
Proportion p1	0.25
Proportion p2	0.7
α err prob	0.05
Power (1-β err prob)	0.80
Allocation ratio N2/N1	1

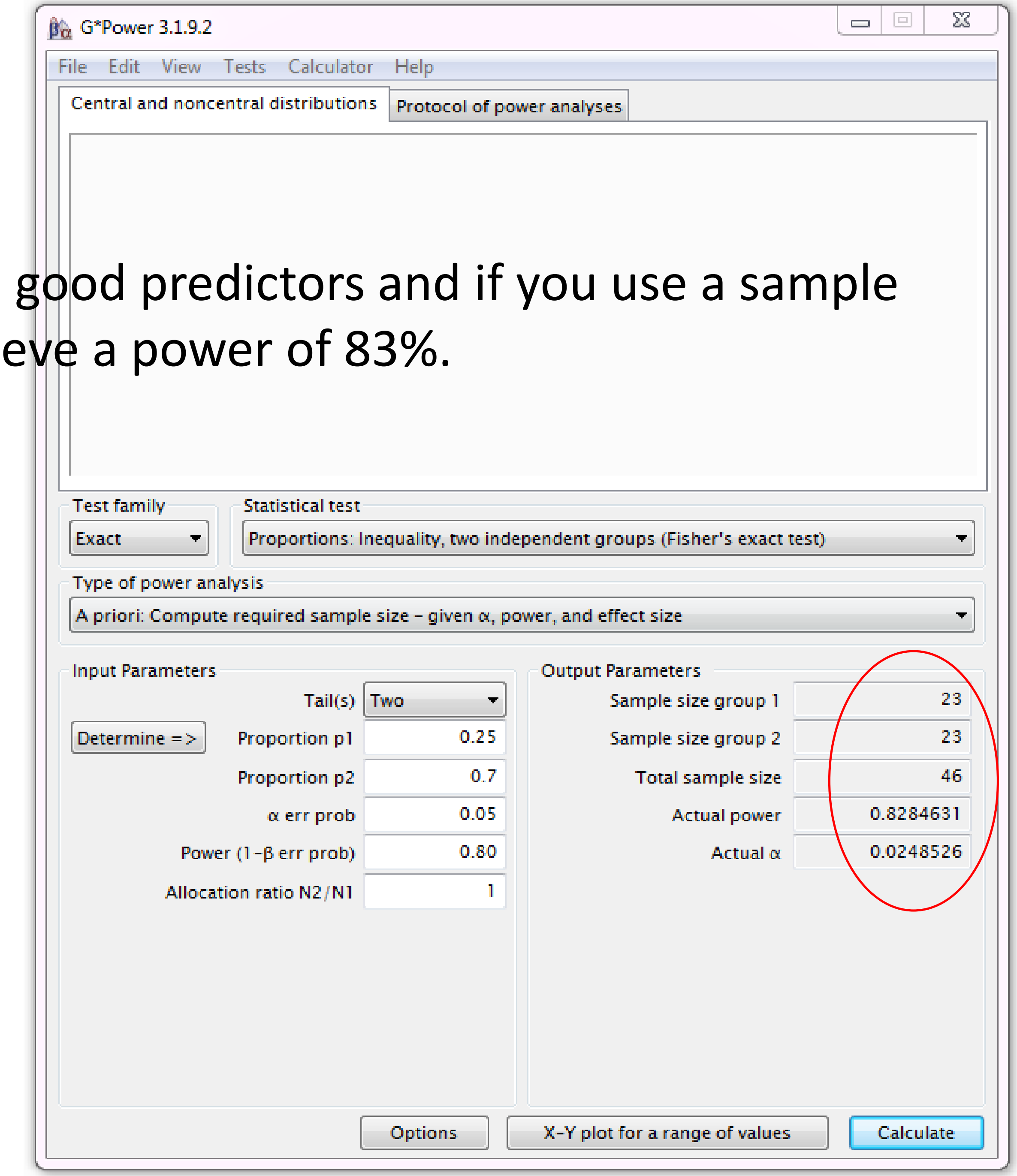
Output Parameters

Sample size group 1	?
Sample size group 2	?
Total sample size	?
Actual power	?
Actual α	?

G*Power

Output:

If the values from the pilot study are good predictors and if you use a sample of n=23 for each group, you will achieve a power of 83%.



Test family: Exact

Statistical test: Proportions: Inequality, two independent groups (Fisher's exact test)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

Tail(s)	Two
Proportion p1	0.25
Proportion p2	0.7
α err prob	0.05
Power (1-β err prob)	0.80
Allocation ratio N2/N1	1

Output Parameters

Sample size group 1	23
Sample size group 2	23
Total sample size	46
Actual power	0.8284631
Actual α	0.0248526

Chi-square and Fisher's tests

- Chi^2 test very easy to calculate by hand but Fisher's very hard

- Many software will not perform a Fisher's test on tables > 2x2

- **Fisher's test more accurate** than Chi^2 test on **small samples**
- **Chi^2 test more accurate** than Fisher's test on **large samples**

- Chi^2 test assumptions:
  - 2x2 table: no expected count <5
  - Bigger tables: all expected > 1 and no more than 20% < 5

- Yates's continuity correction
  - All statistical tests work well when their assumptions are met
  - When not: probability Type 1 error increases
  - Solution: corrections that increase p-values
    - Corrections are dangerous: no magic
    - Probably best to avoid them

Chi-square test

- In a chi-square test, the observed frequencies for two or more groups are compared with expected frequencies by chance.

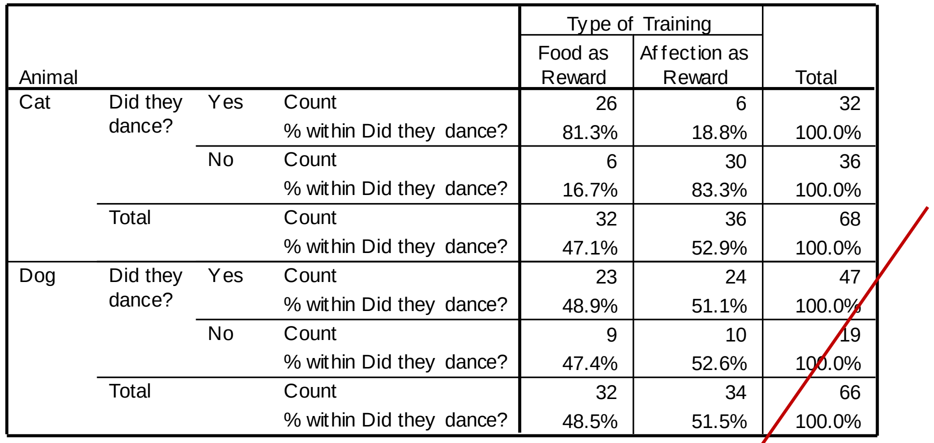
$$\chi^2 = \Sigma \frac{(Observed\ Frequency - Expected\ Frequency)^2}{Expected\ Frequency}$$

  - With observed frequency = collected data

- **Example with ‘cats and dogs’**

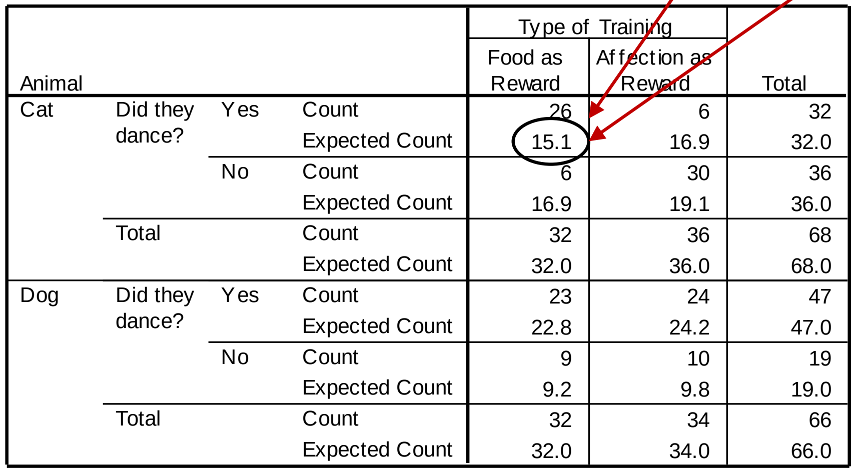
Chi-square test

Did they dance? * Type of Training * Animal Crosstabulation



Animal				Type of Training		Total
				Food as Reward	Affection as Reward	
Cat	Did they dance?	Yes	Count	26	6	32
			% within Did they dance?	81.3%	18.8%	100.0%
		No	Count	6	30	36
			% within Did they dance?	16.7%	83.3%	100.0%
	Total		Count	32	36	68
			% within Did they dance?	47.1%	52.9%	100.0%
Dog	Did they dance?	Yes	Count	23	24	47
			% within Did they dance?	48.9%	51.1%	100.0%
		No	Count	9	10	19
			% within Did they dance?	47.4%	52.6%	100.0%
	Total		Count	32	34	66
			% within Did they dance?	48.5%	51.5%	100.0%

Did they dance? * Type of Training * Animal Crosstabulation



Animal				Type of Training		Total
				Food as Reward	Affection as Reward	
Cat	Did they dance?	Yes	Count	26	6	32
			Expected Count	15.1	16.9	32.0
		No	Count	6	30	36
			Expected Count	16.9	19.1	36.0
	Total		Count	32	36	68
			Expected Count	32.0	36.0	68.0
Dog	Did they dance?	Yes	Count	23	24	47
			Expected Count	22.8	24.2	47.0
		No	Count	9	10	19
			Expected Count	9.2	9.8	19.0
	Total		Count	32	34	66
			Expected Count	32.0	34.0	66.0

Example: expected frequency of cats line dancing after having received food as a reward:

Direct counts approach:

Expected frequency=(row total)*(column total)/grand total = 32*32/68 = **15.1**

Probability approach:

Probability of line dancing: 32/68
Probability of receiving food: 32/68

Expected frequency:(32/68)*(32/68)=0.22: **22% of 68 = 15.1**

For the cats:

Chi^2 = (26-15.1)^2/15.1 + (6-16.9)^2/16.9 + (6-16.9)^2 /16.9 + (30-19.1)^2/19.1 = 28.4

Is 28.4 big enough for the test to be significant?

Results

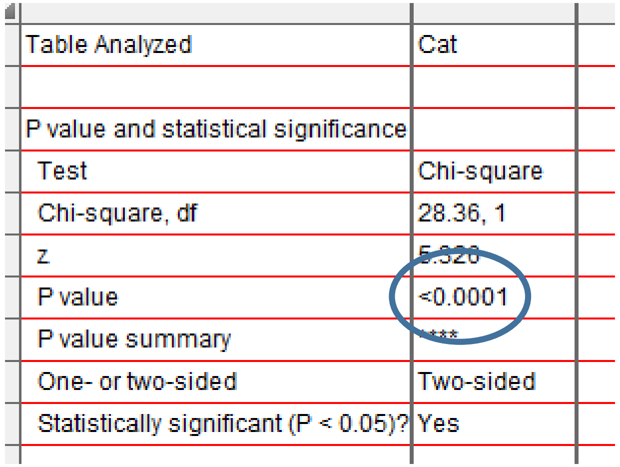
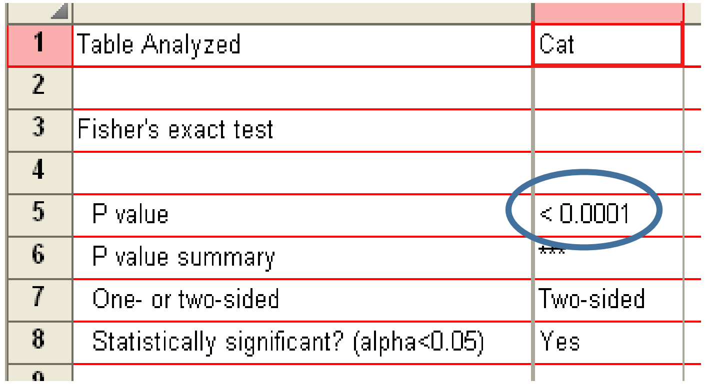


Table Analyzed	Cat
P value and statistical significance	
Test	Chi-square
Chi-square, df	28.36, 1
z	5.326
P value	<0.0001
P value summary	****
One- or two-sided	Two-sided
Statistically significant (P < 0.05)?	Yes



1	Table Analyzed	Cat
2		
3	Fisher's exact test	
4		
5	P value	< 0.0001
6	P value summary	***
7	One- or two-sided	Two-sided
8	Statistically significant? (alpha<0.05)	Yes

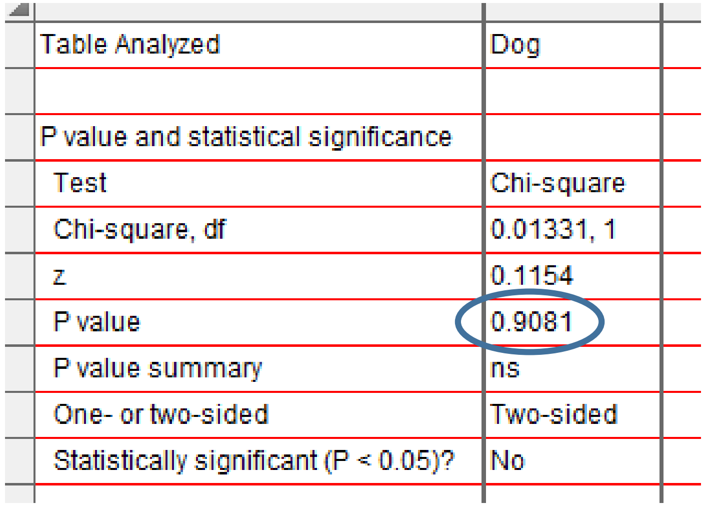


Table Analyzed	Dog
P value and statistical significance	
Test	Chi-square
Chi-square, df	0.01331, 1
z	0.1154
P value	0.9081
P value summary	ns
One- or two-sided	Two-sided
Statistically significant (P < 0.05)?	No

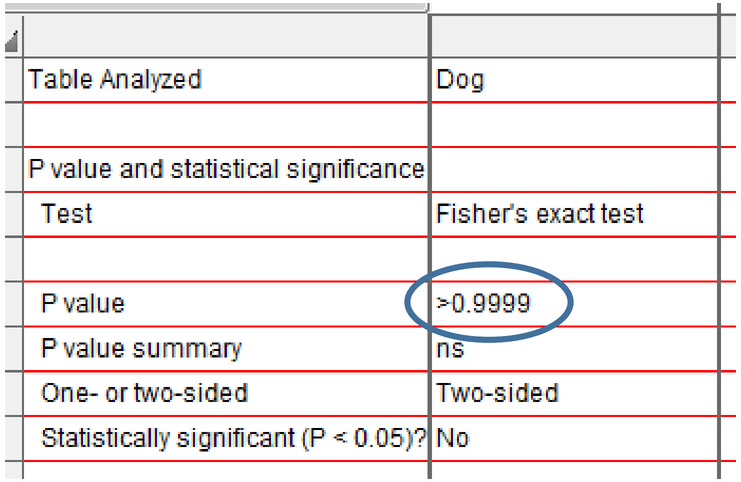
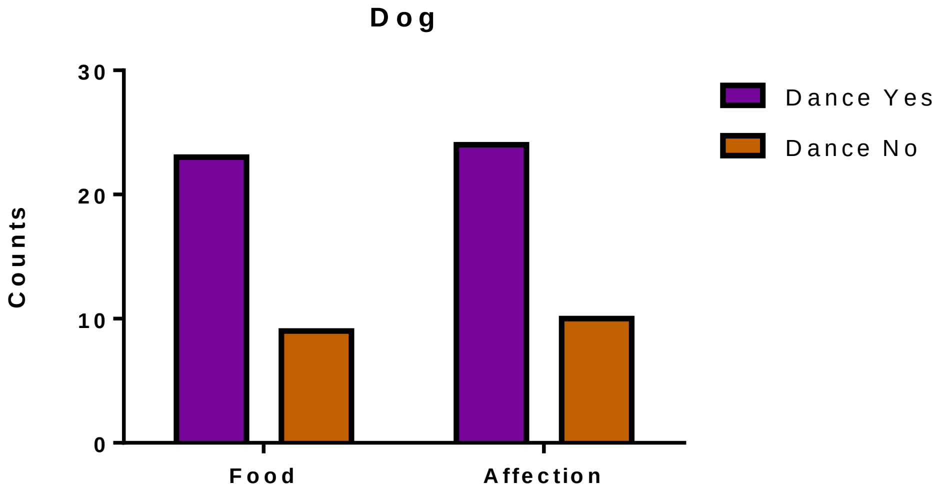
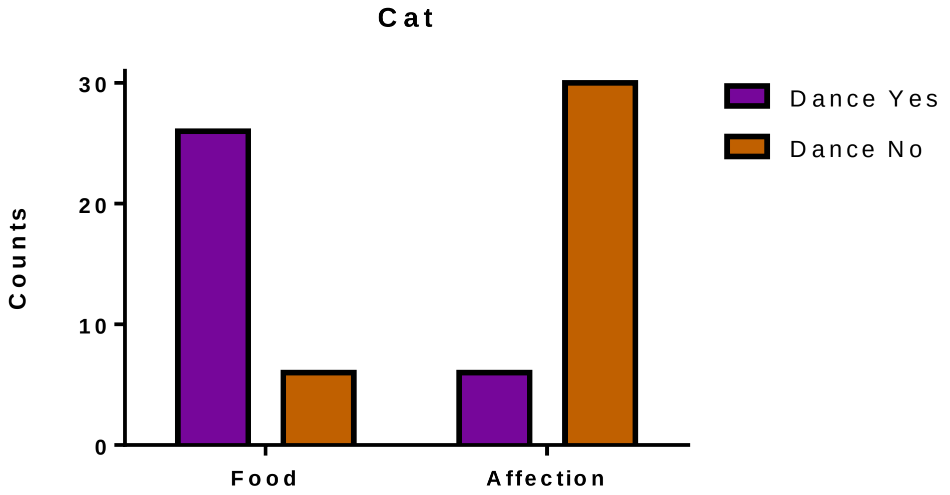


Table Analyzed	Dog
P value and statistical significance	
Test	Fisher's exact test
P value	>0.9999
P value summary	ns
One- or two-sided	Two-sided
Statistically significant (P < 0.05)?	No

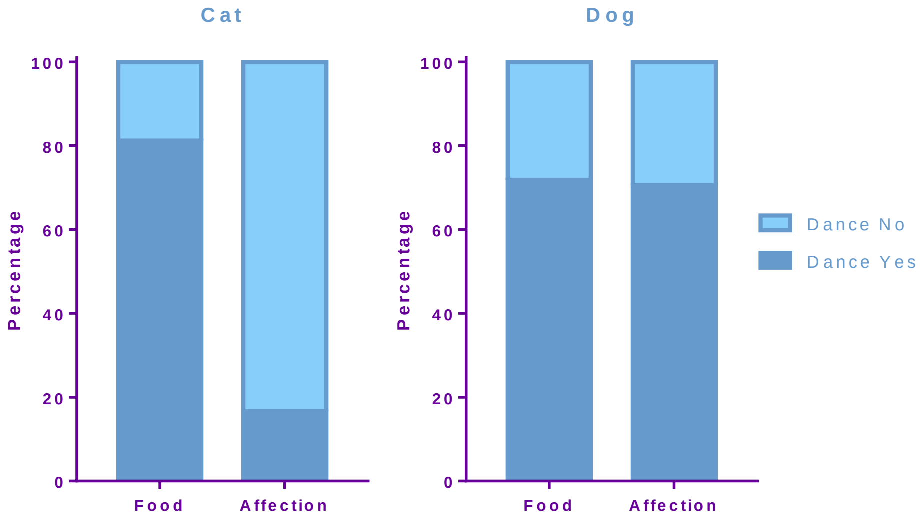
Fisher’s exact test: results





- **In our example:**

cats are more likely to line dance if they are given food as reward than affection ($p<0.0001$) whereas dogs don’t mind ($p>0.99$).



Quantitative data

Quantitative data

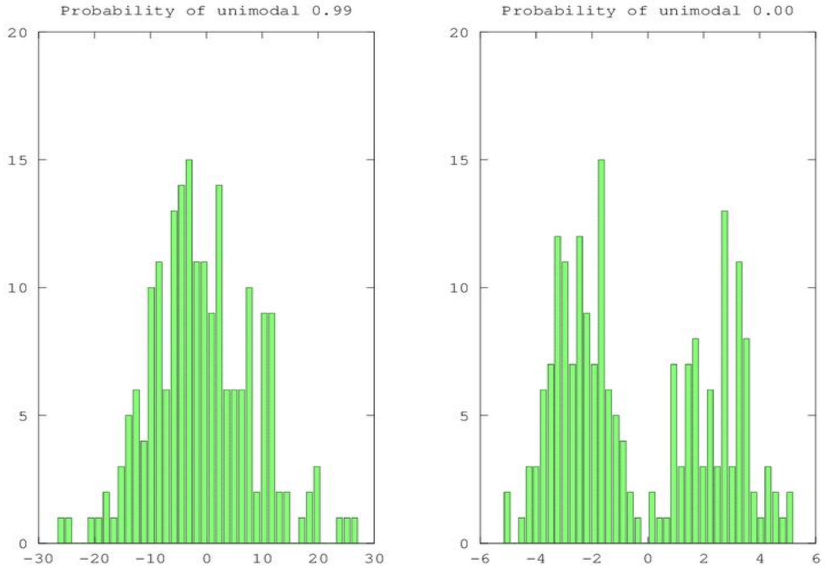
- They take numerical values (units of measurement)

- Discrete: obtained by counting
  - Example: number of students in a class
  - values vary by finite specific steps
- or continuous: obtained by measuring
  - Example: height of students in a class
  - any values

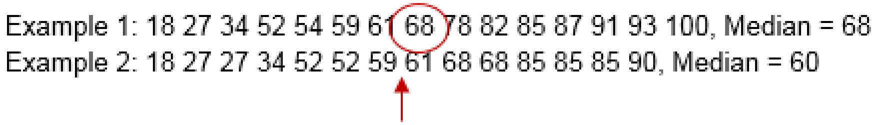
- They can be described by a series of parameters:
  - Mean, variance, standard deviation, standard error and confidence interval

Measures of central tendency
Mode and Median

- **Mode:** most commonly occurring value in a distribution



- **Median**: value exactly in the middle of an ordered set of numbers


Example 1: 18 27 34 52 54 59 61 68 78 82 85 87 91 93 100, Median = 68
Example 2: 18 27 27 34 52 52 59 61 68 68 85 85 85 90, Median = 60

Measures of central tendency

Mean

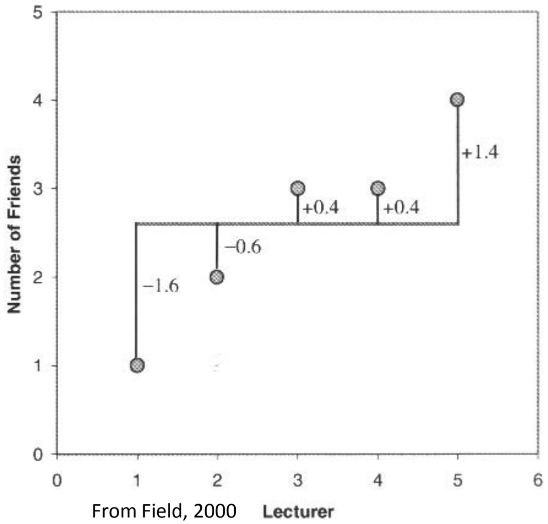
- Definition: average of all values in a column

- It can be considered as a model because it summaries the data
  - Example: a group of 5 lecturers: number of friends of each members of the group: 1, 2, 3, 3 and 4
    - Mean: (1+2+3+3+4)/5 = 2.6 friends per person
      - Clearly an hypothetical value

- How can we know that it is an accurate model?
  - Difference between the real data and the model created

Measures of dispersion

- Calculate the magnitude of the differences between each data and the mean:



From Field, 2000

- Total error = sum of differences

$= 0 = \Sigma(x_i - \overline{x})$ = (-1.6)+(-0.6)+(0.4)+(1.4) = 0

No errors !

- Positive and negative: they cancel each other out.

Sum of Squared errors (SS)

- To avoid the problem of the direction of the errors: we square them
  - Instead of sum of errors: sum of squared errors (SS):

$$(SS) = \Sigma(x_i - \overline{x})(x_i - \overline{x})$$

$= (1.6)^2 + (-0.6)^2 + (0.4)^2 + (0.4)^2 + (1.4)^2$

$= 2.56 + 0.36 + 0.16 + 0.16 + 1.96$

$= 5.20$

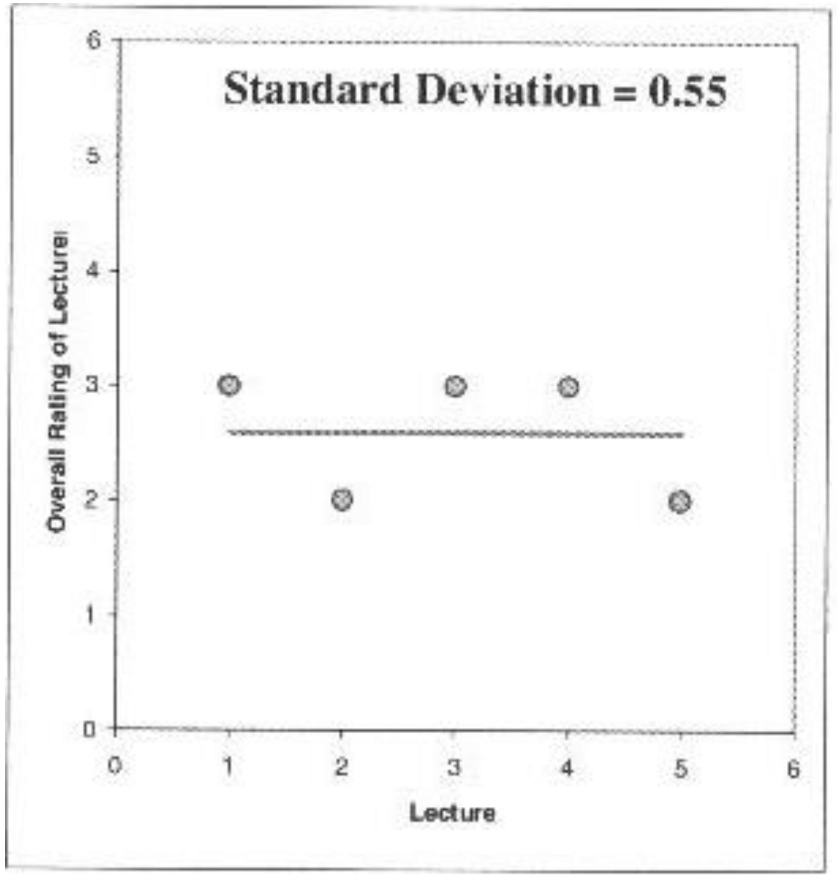
- SS gives a good measure of the accuracy of the model
  - But: dependent upon the amount of data: the more data, the higher the SS.
  - Solution: to divide the SS by the number of observations (N)
    - As we are interested in measuring the error in the sample to estimate the one in the population we divide the SS by N-1 instead of N and we get the **variance** (S^2) = SS/N-1

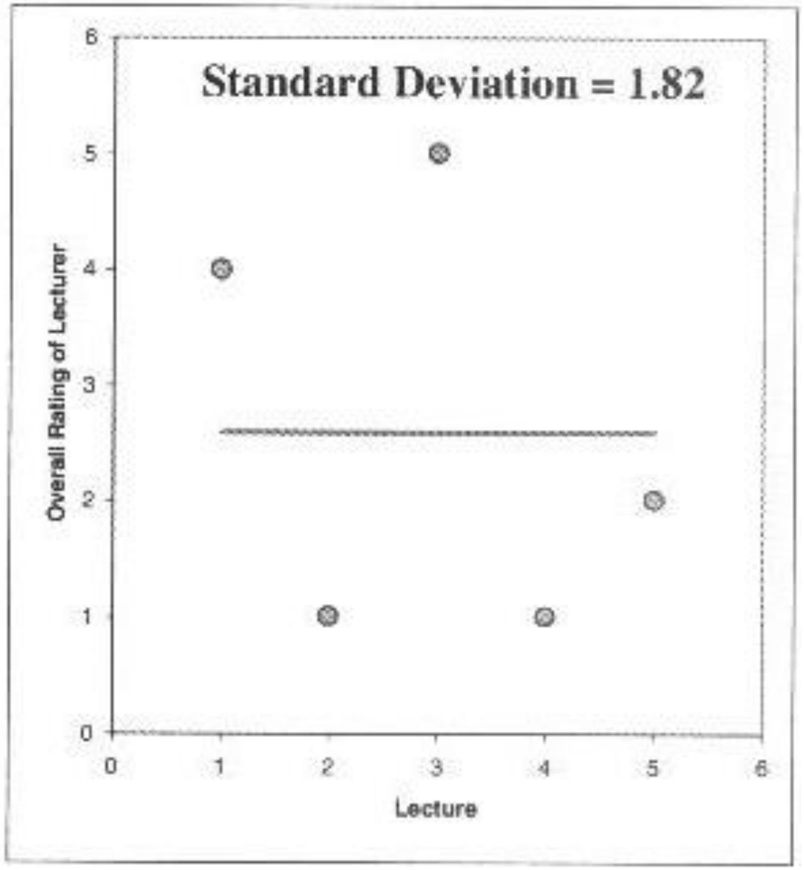
Variance and standard deviation

- $variance\ (s^2) = \frac{SS}{N-1} = \frac{\Sigma\,(x_i - \overline{x})^2}{N-1} = \frac{5.20}{4} = 1.3$

- Problem with variance: measure in squared units
  - For more convenience, the square root of the variance is taken to obtain a measure in the same unit as the original measure:
    - the **standard deviation**
      - S.D. = √(SS/N-1) = √(s^2) = s $= \sqrt{1.3} = 1.14$
    - The **standard deviation** is a measure of how well the mean represents the data.

Standard deviation





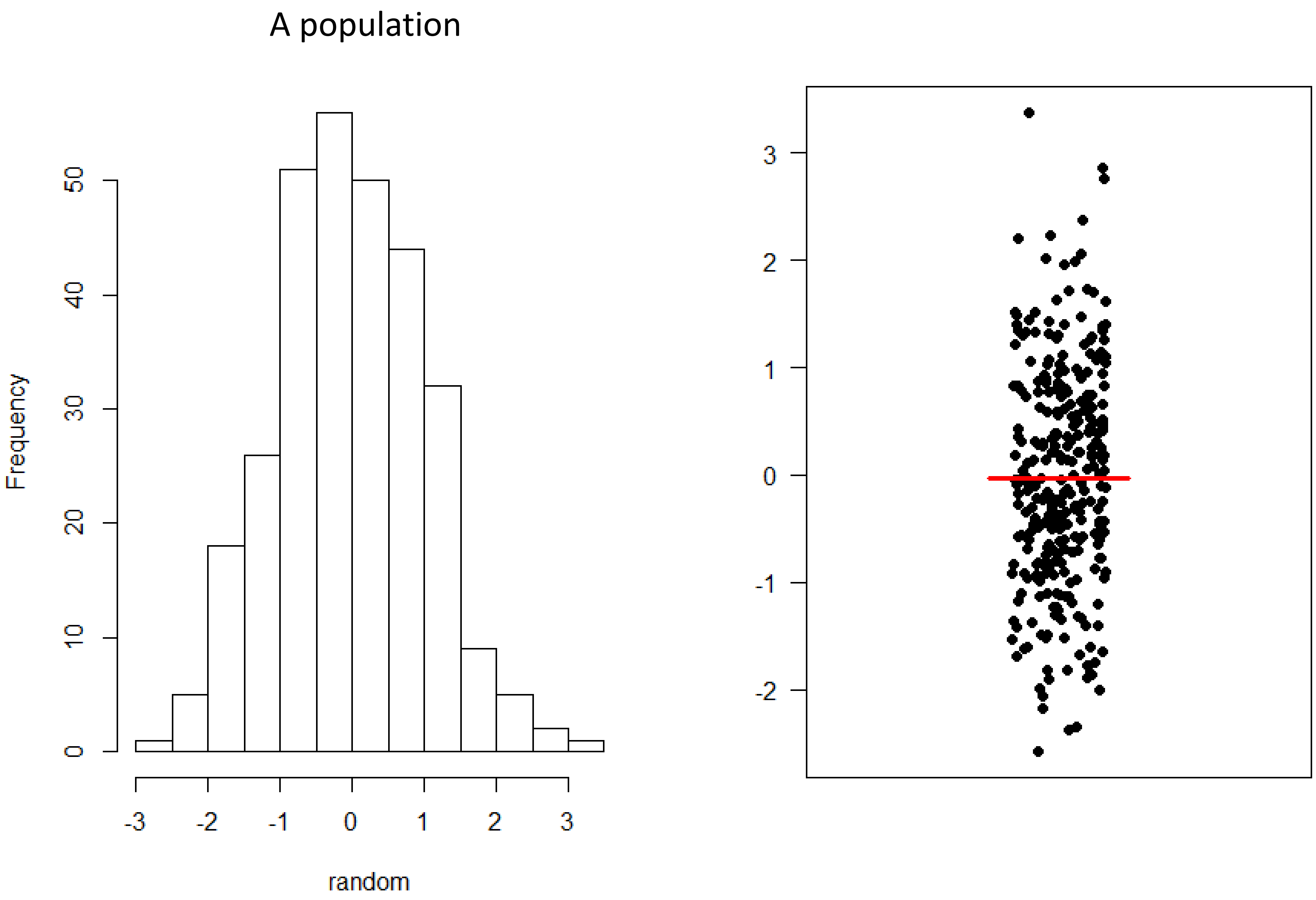
Small S.D.:
data close to the mean:
mean is a good fit of the data

Large S.D.:
data distant from the mean:
mean is not an accurate representation

SD and SEM (SEM = SD/√N)

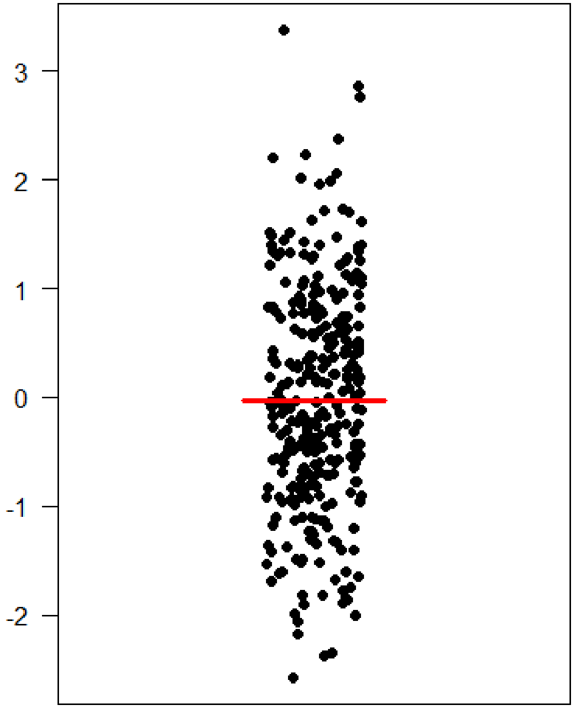
- What are they about?
  - The **SD** quantifies **how much the values vary** from one another: **scatter or spread**
    - The SD does not change predictably as you acquire more data.
  - The **SEM** quantifies **how accurately** you know the **true mean** of the population.
    - Why? Because it takes into account: **SD + sample size**
    - The SEM gets smaller as your sample gets larger
      - Why? Because the mean of a large sample is likely to be closer to the true mean than is the mean of a small sample.

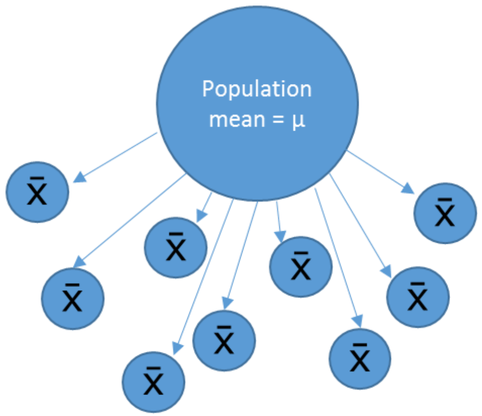
The SEM and the sample size



A population

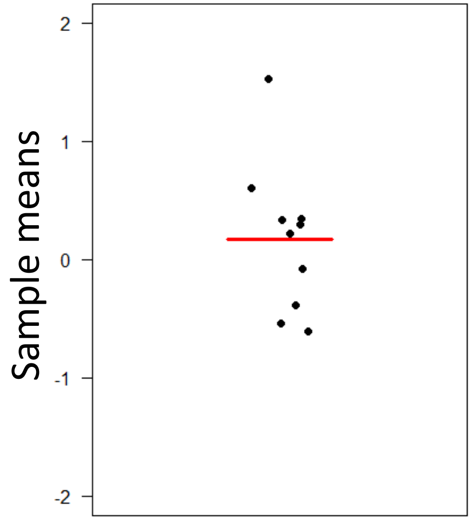
The SEM and the sample size



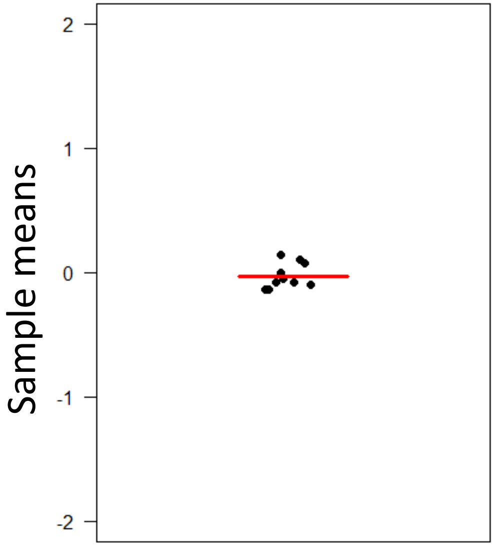


‘Infinite’ number of samples
Samples means = $\bar{x}$

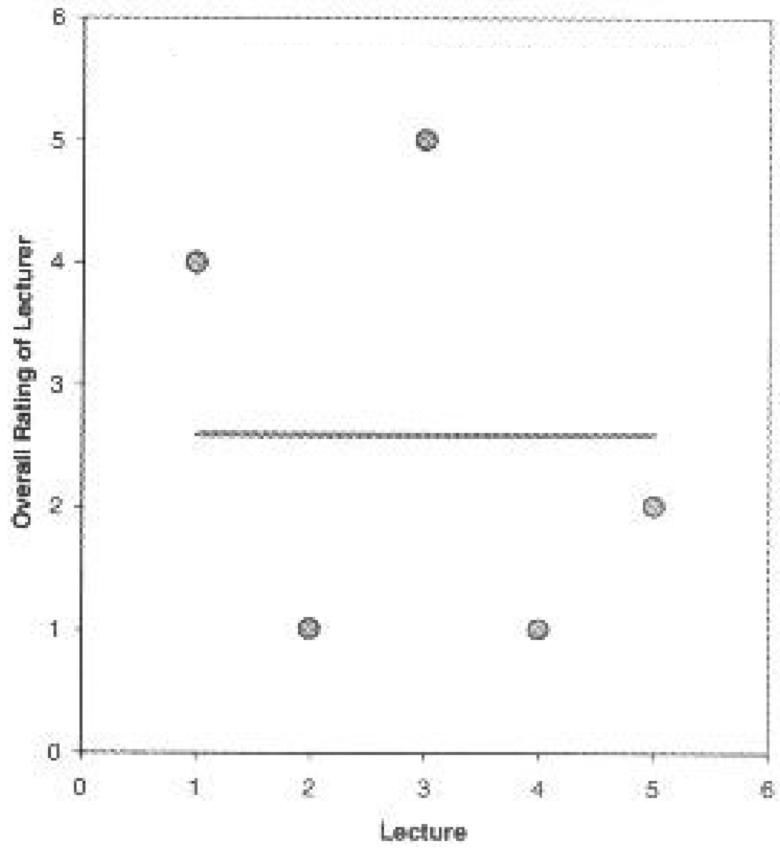
Small samples (n=3)



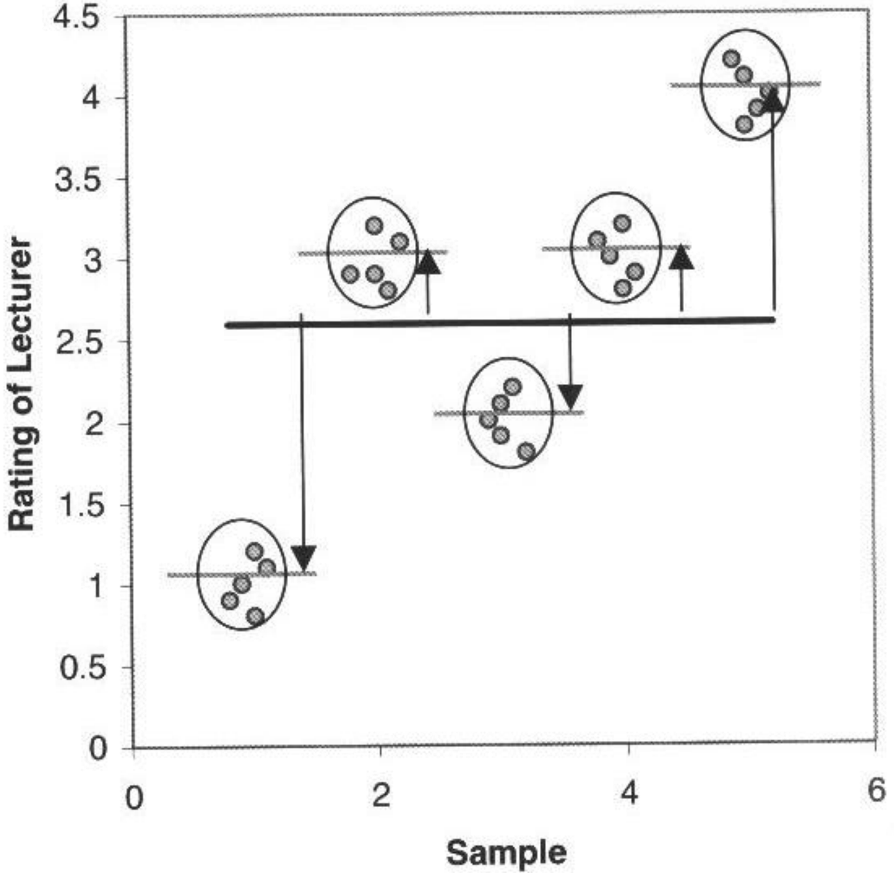
Big samples (n=30)



SD and SEM



The SD quantifies the scatter of the data.



The SEM quantifies the distribution of the sample means.

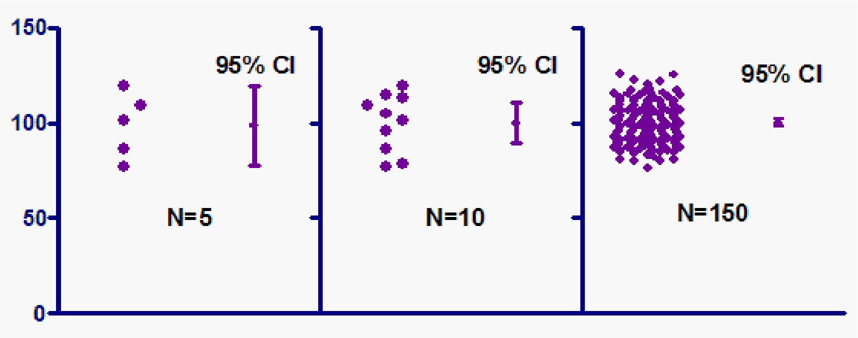
SD or SEM ?

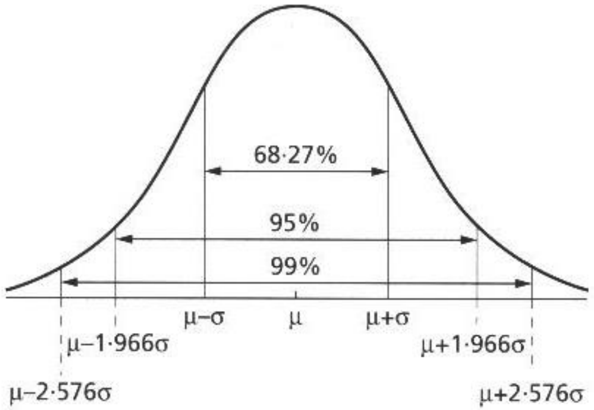
- If the scatter is caused by biological variability, it is important to show the variation.
  - Report the SD rather than the SEM.
    - Better even: show a graph of all data points.

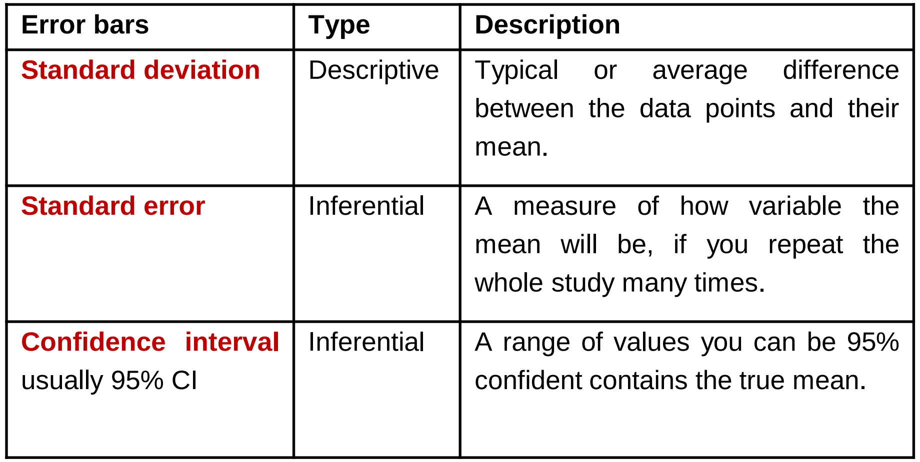
- If you are using an in vitro system with no biological variability, the scatter is about experimental imprecision (no biological meaning).
  - Report the SEM to show how well you have determined the mean.

Confidence interval

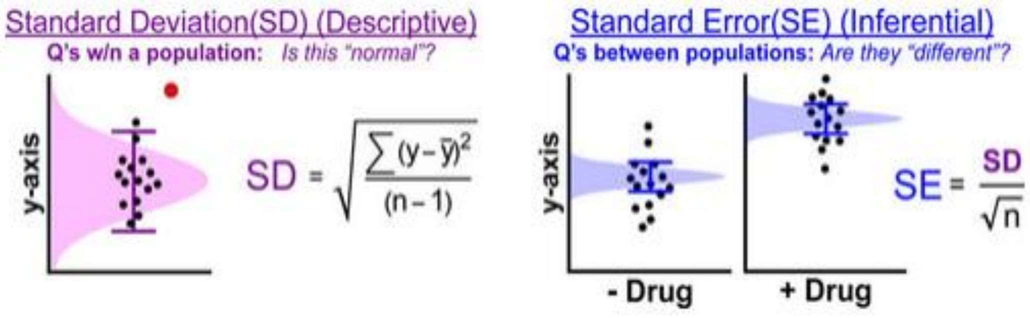
- Range of values that we can be 95% confident contains the true mean of the population.
  - So limits of 95% CI: **[Mean - 1.96 SEM; Mean + 1.96 SEM]** (SEM = SD/√N)







Error bars	Type	Description
Standard deviation	Descriptive	Typical or average difference between the data points and their mean.
Standard error	Inferential	A measure of how variable the mean will be, if you repeat the whole study many times.
Confidence interval usually 95% CI	Inferential	A range of values you can be 95% confident contains the true mean.



Analysis of Quantitative Data

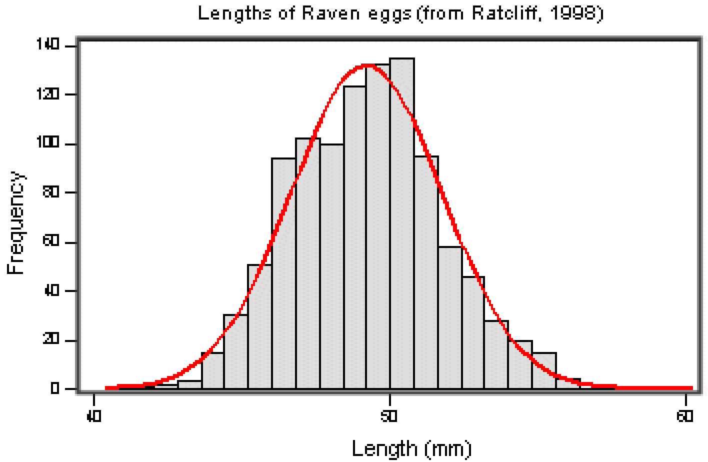
- Choose the correct statistical test to answer your question:
  - They are 2 types of statistical tests:
    - **Parametric tests** with 4 assumptions to be met by the data,
    - **Non-parametric tests** with no or few assumptions (e.g. Mann-Whitney test) and/or for qualitative data (e.g. Fisher's exact and χ^2 tests).

Assumptions of Parametric Data

- All parametric tests have 4 basic assumptions that must be met for the test to be accurate.

1) <u>Normally distributed data</u>

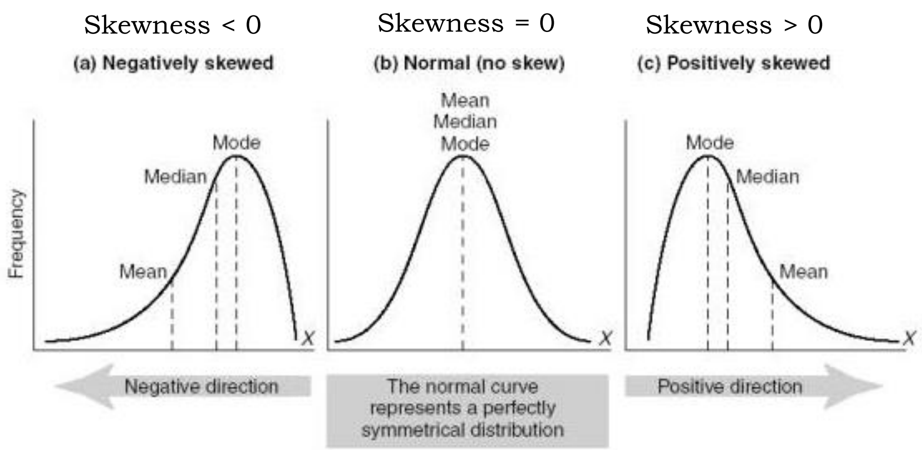
- Normal shape, bell shape, Gaussian shape



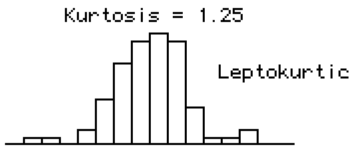
- Transformations can be made to make data suitable for parametric analysis.

Assumptions of Parametric Data

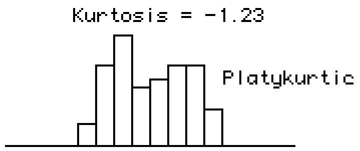
- Frequent departures from normality:
  - Skewness: lack of symmetry of a distribution



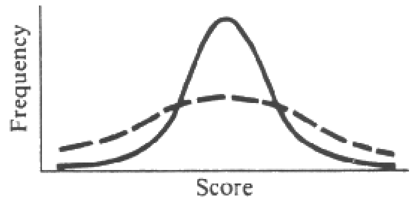
  - Kurtosis: measure of the degree of 'peakedness' in the distribution
    - The two distributions below have the same variance approximately the same skew, but differ markedly in kurtosis.



More peaked distribution: kurtosis > 0



Flatter distribution: kurtosis < 0



(e) Platykurtic and leptokurtic

Assumptions of Parametric Data

2) ***Homogeneity in variance***

- The variance should not change systematically throughout the data

3) ***Interval data (linearity)***

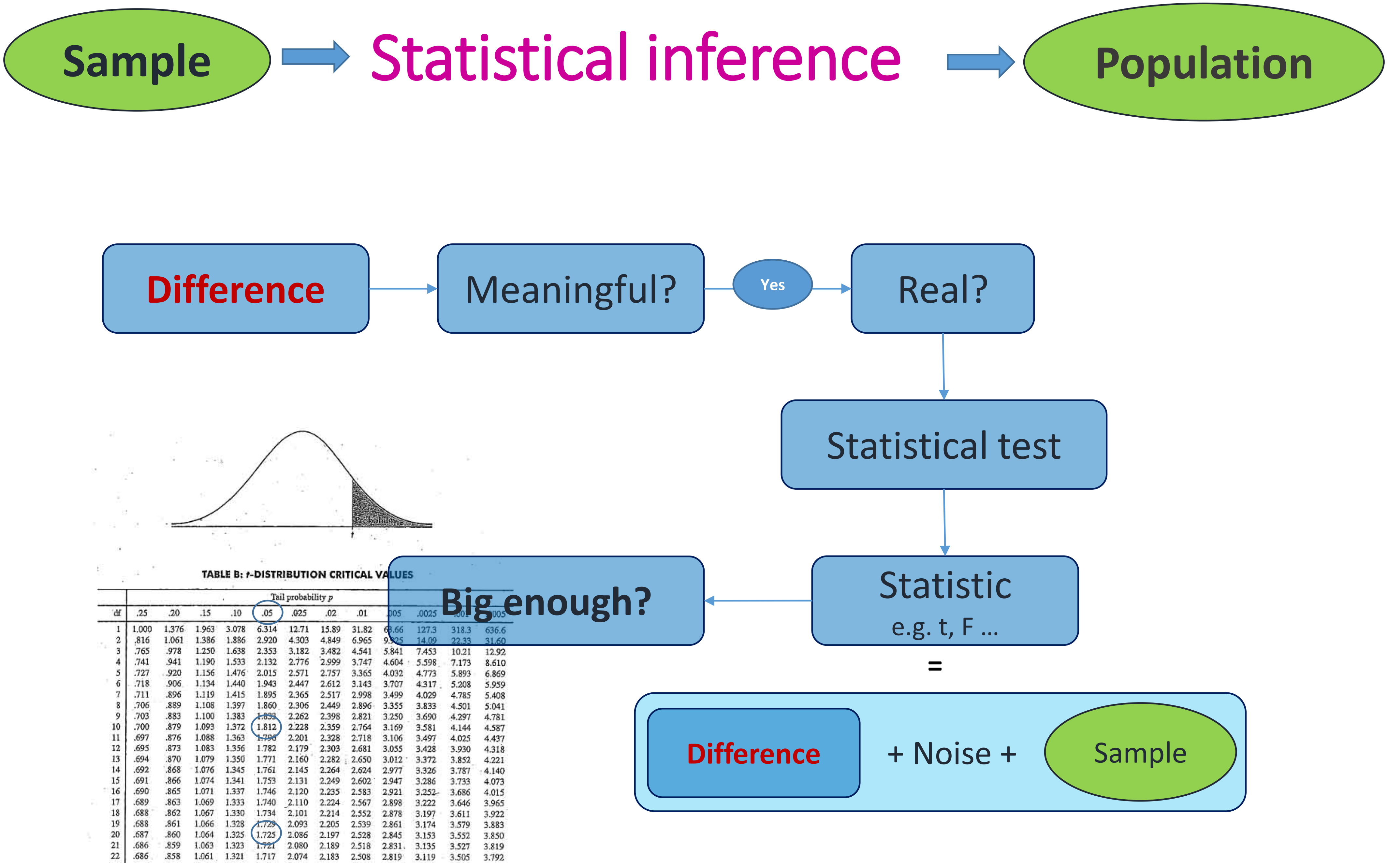
- The distance between points of the scale should be equal at all parts along the scale.

4) ***Independence***

- Data from different subjects are independent
    - Values corresponding to one subject do not influence the values corresponding to another subject.
    - Important in repeated measures experiments

Analysis of Quantitative Data

- **Is there a difference between my groups regarding the variable I am measuring?**
  - e.g. are the mice in the group A heavier than those in group B?
    - Tests with 2 groups:
      - Parametric: **Student's *t*-test**
      - Non parametric: Mann-Whitney/Wilcoxon rank sum test
    - Tests with more than 2 groups:
      - Parametric: **Analysis of variance (one-way ANOVA)**
      - Non parametric: Kruskal Wallis
- **Is there a relationship between my 2 (continuous) variables?**
  - e.g. is there a relationship between the daily intake in calories and an increase in body weight?
    - Test: **Correlation** (parametric) and **curve fitting**



Sample → Statistical inference → Population

Difference → Meaningful? → Yes → Real? → Statistical test → Statistic e.g. t, F … → Big enough?

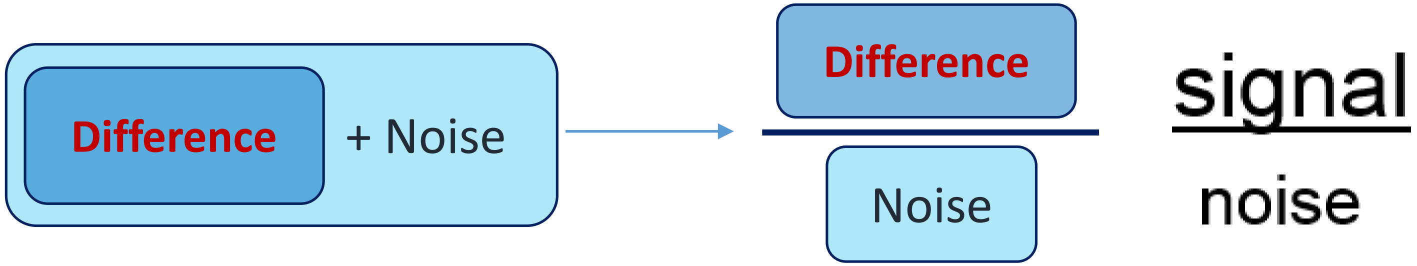
Statistic = **Difference** + Noise + Sample

TABLE B: *t*-DISTRIBUTION CRITICAL VALUES

	Tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792

Signal-to-noise ratio

- Stats are all about understanding and controlling variation.



Difference + Noise → $\dfrac{\text{Difference}}{\text{Noise}}$ $\dfrac{\text{signal}}{\text{noise}}$

$\dfrac{\text{signal}}{\scriptstyle\text{noise}}$ If the **noise is low** then the **signal is detectable** …

= statistical significance

$\dfrac{\text{signal}}{\text{\Large noise}}$ … but if the **noise** (i.e. interindividual variation) **is large** then the **same signal will not be detected**

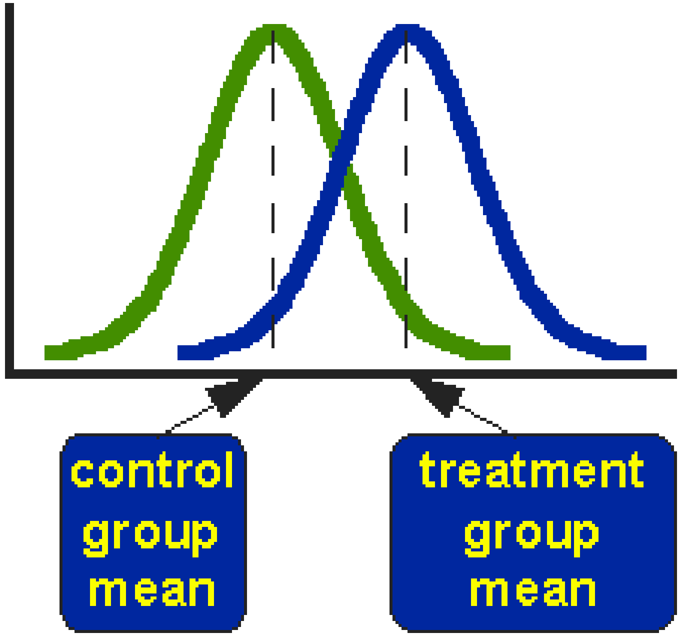
= no statistical significance

- In a statistical test, the ratio of signal to noise determines the significance.

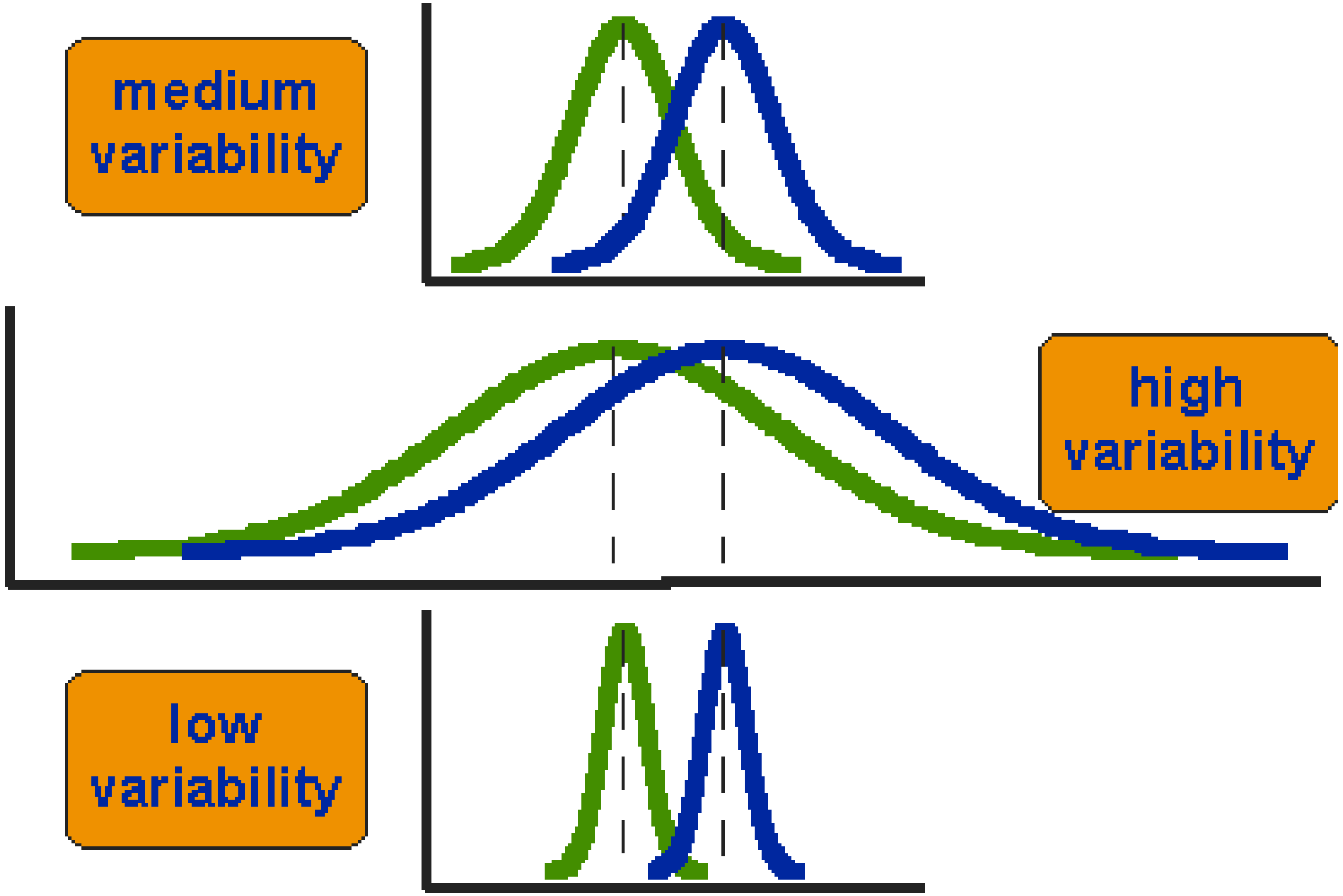
Comparison between 2 groups:

Student's *t*-test

- **Basic idea**:
  - When we are looking at the differences between scores for 2 groups, we have to judge the difference between their means relative to the spread or variability of their scores.
    - Eg: comparison of 2 groups: control and treatment



Student's *t*-test

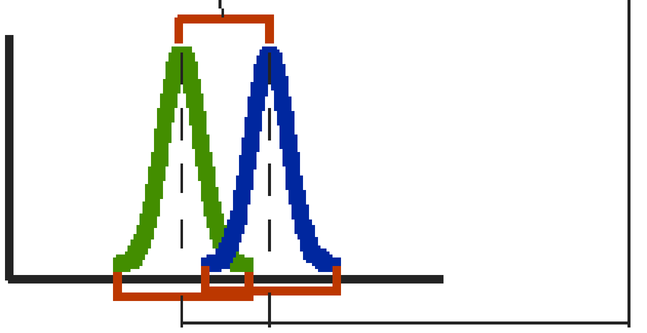


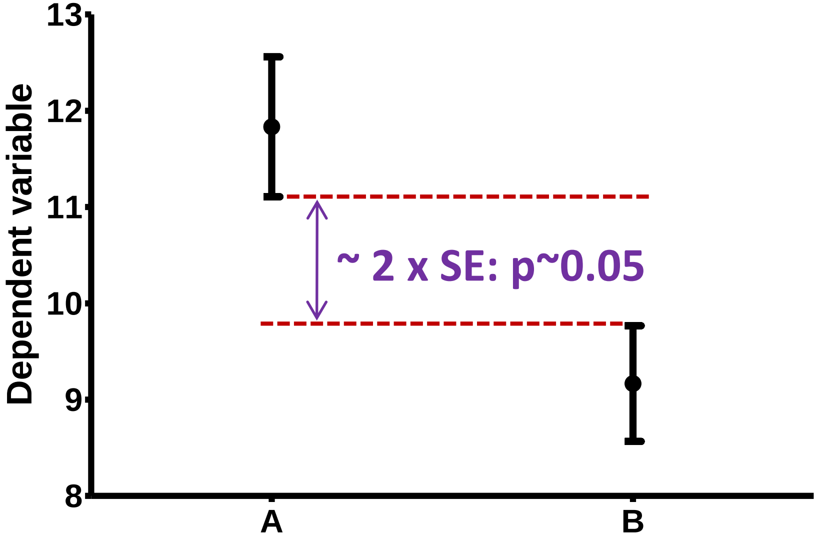
Student’s *t*-test

$$\frac{\text{signal}}{\text{noise}} = \frac{\text{difference between group means}}{\text{variability of groups}}$$

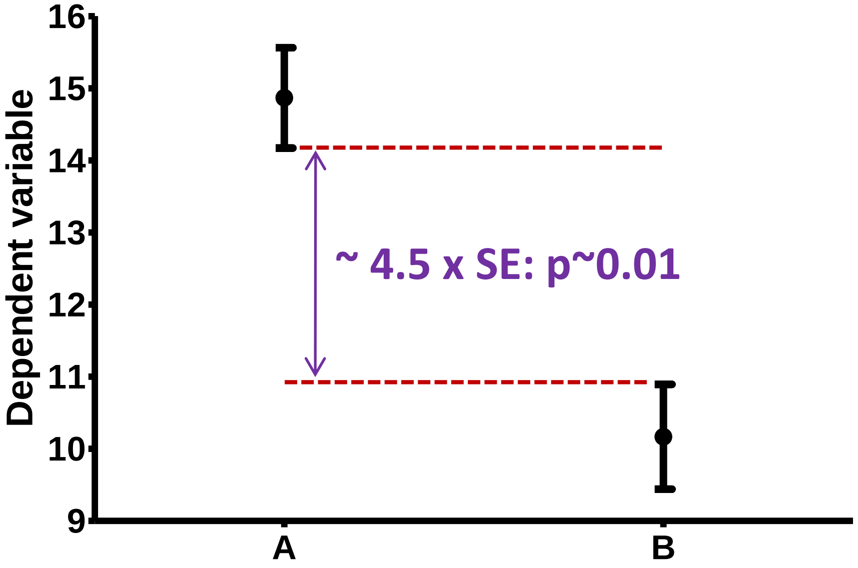
$$= \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{\text{var}_T}{n_T} + \frac{\text{var}_C}{n_C}}}$$

$$= \text{t-value}$$

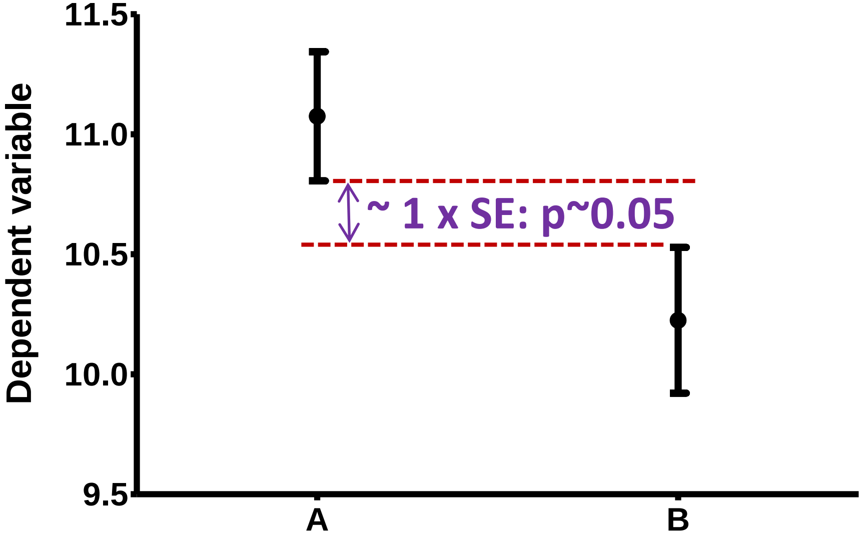




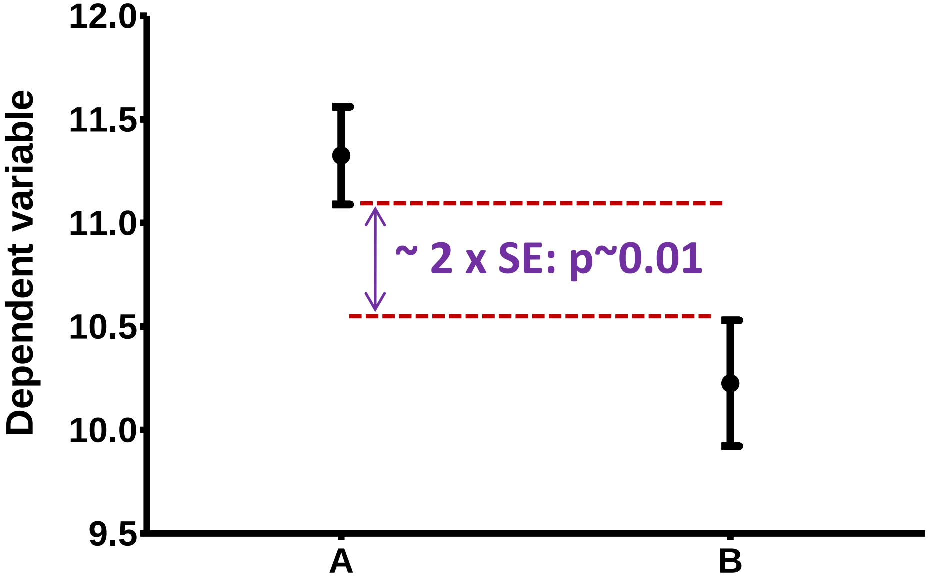
SE gap ~ 2 n=3



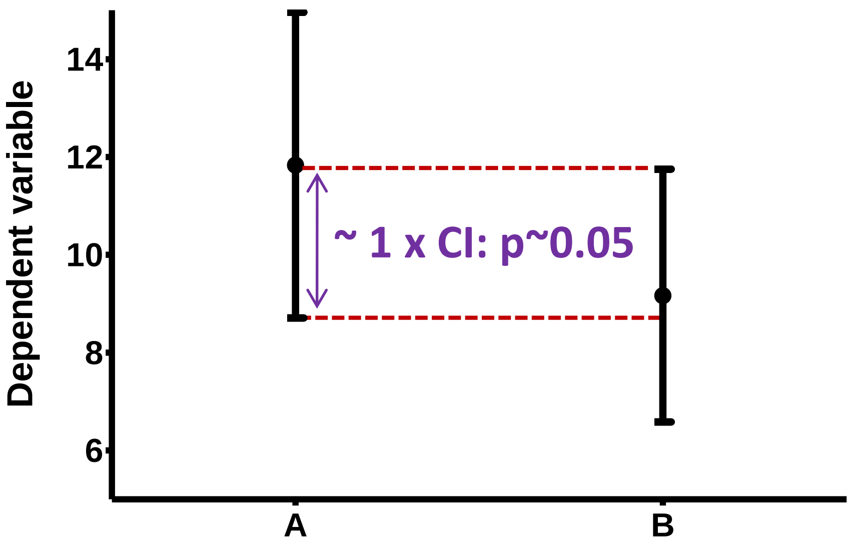
SE gap ~ 4.5 n=3



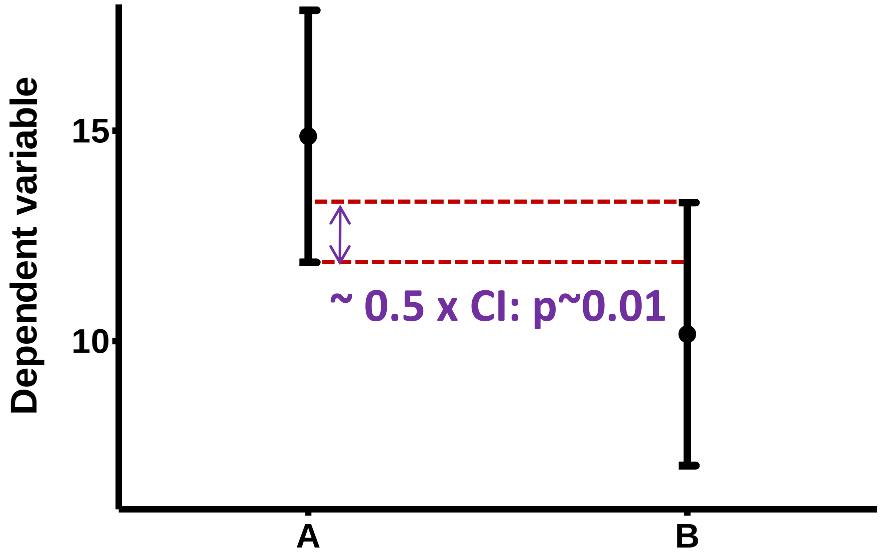
SE gap ~ 1 n>=10



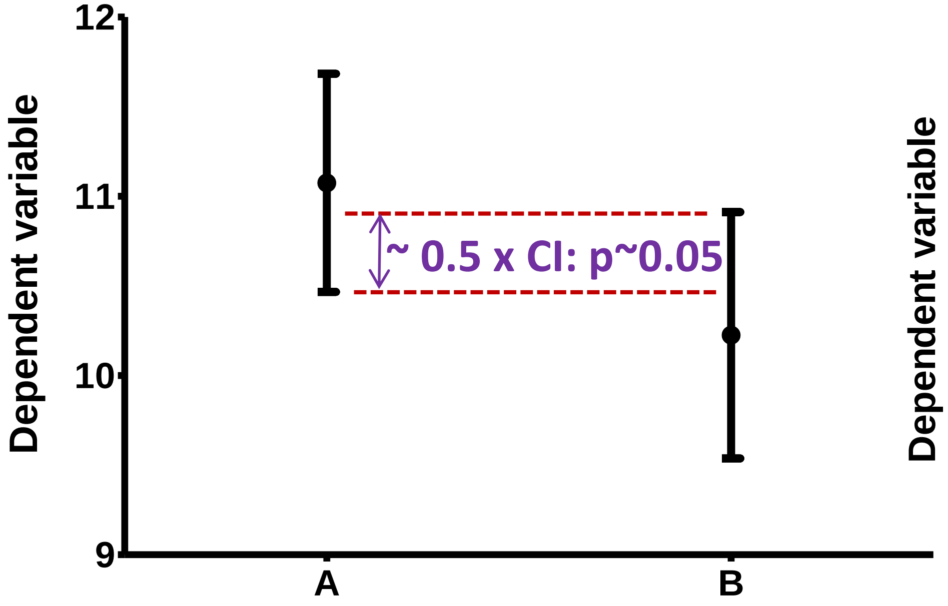
SE gap ~ 2 n>=10



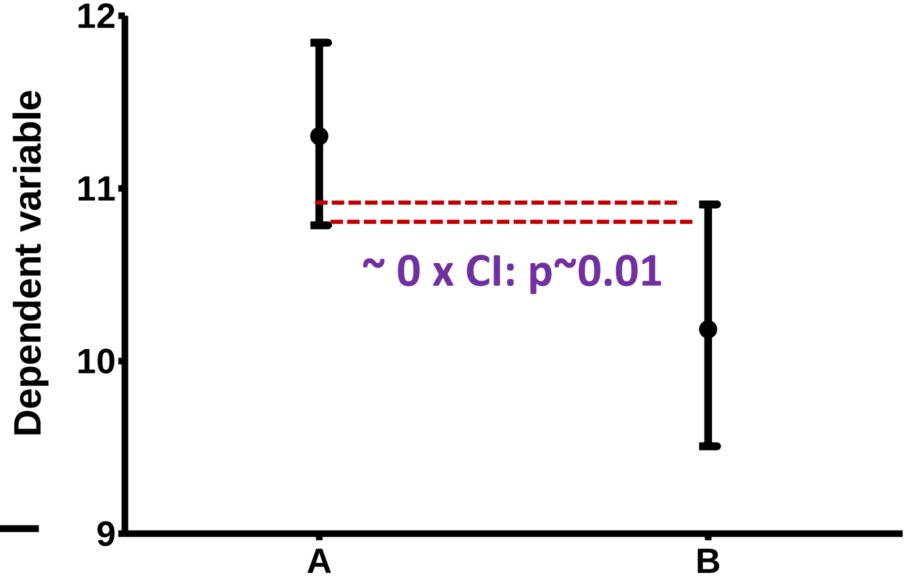
CI overlap ~ 1 n=3



CI overlap ~ 0.5 n=3



CI overlap ~ 0.5 n>=10



CI overlap ~ 0 n>=10

Student’s *t*-test

- 3 types:
  - **Independent t-test**
    - compares means for two independent groups of cases.
  - **Paired t-test**
    - looks at the difference between two variables for a single group:
      - the second ‘sample’ of values comes from the same subjects (mouse, petri dish …).
  - One-Sample t-test
    - tests whether the mean of a single variable differs from a specified constant (often 0)

Example: coyotes.xlsx

- Question: do male and female coyotes differ in size?

- **Sample size**
- **Data exploration**
- **Check the assumptions for parametric test**
- **Statistical analysis: Independent t-test**

Power analysis

- Example case:

No data from a pilot study but we have found some information in the literature.

In a study run in similar conditions as in the one we intend to run, **male coyotes** were found to measure: **92cm+/- 7cm (SD)**.

We expect a **5% difference** between genders.

- **smallest biologically meaningful difference**

G*Power

Independent t-test

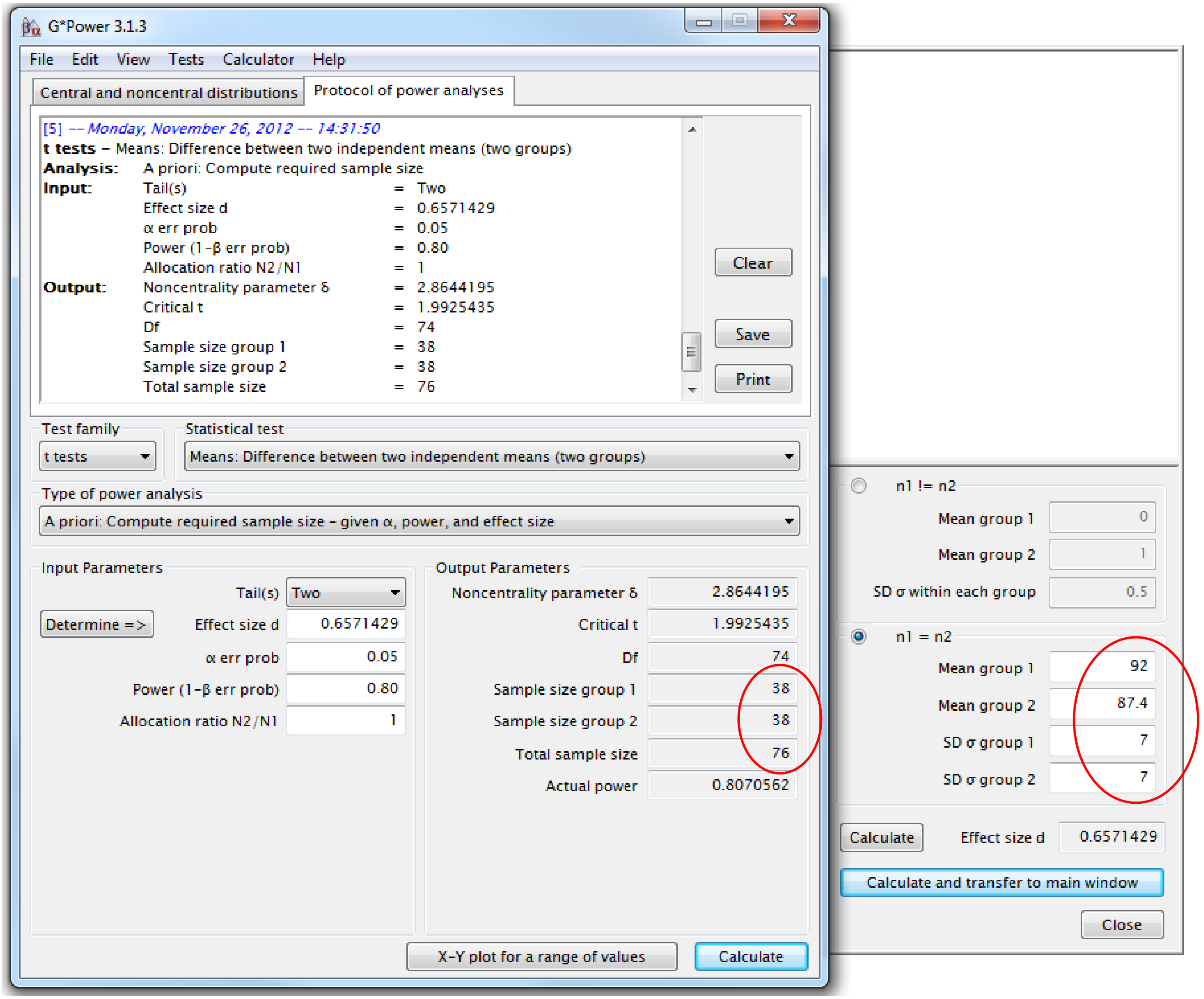
***A priori* Power analysis**

Example case:

You don't have data from a pilot study but you have found some information in the literature.

In a study run in similar conditions to the one you intend to run, male coyotes were found to measure:
92cm+/- 7cm (SD)

You expect a 5% difference between genders with a similar variability in the female sample.



```
[5] -- Monday, November 26, 2012 -- 14:31:50
t tests – Means: Difference between two independent means (two groups)
Analysis:   A priori: Compute required sample size
Input:      Tail(s)                          =  Two
            Effect size d                    =  0.6571429
            α err prob                       =  0.05
            Power (1-β err prob)             =  0.80
            Allocation ratio N2/N1           =  1
Output:     Noncentrality parameter δ        =  2.8644195
            Critical t                       =  1.9925435
            Df                               =  74
            Sample size group 1              =  38
            Sample size group 2              =  38
            Total sample size                =  76
```

Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters
Tail(s): Two
Effect size d: 0.6571429
α err prob: 0.05
Power (1-β err prob): 0.80
Allocation ratio N2/N1: 1

Output Parameters
Noncentrality parameter δ: 2.8644195
Critical t: 1.9925435
Df: 74
Sample size group 1: 38
Sample size group 2: 38
Total sample size: 76
Actual power: 0.8070562

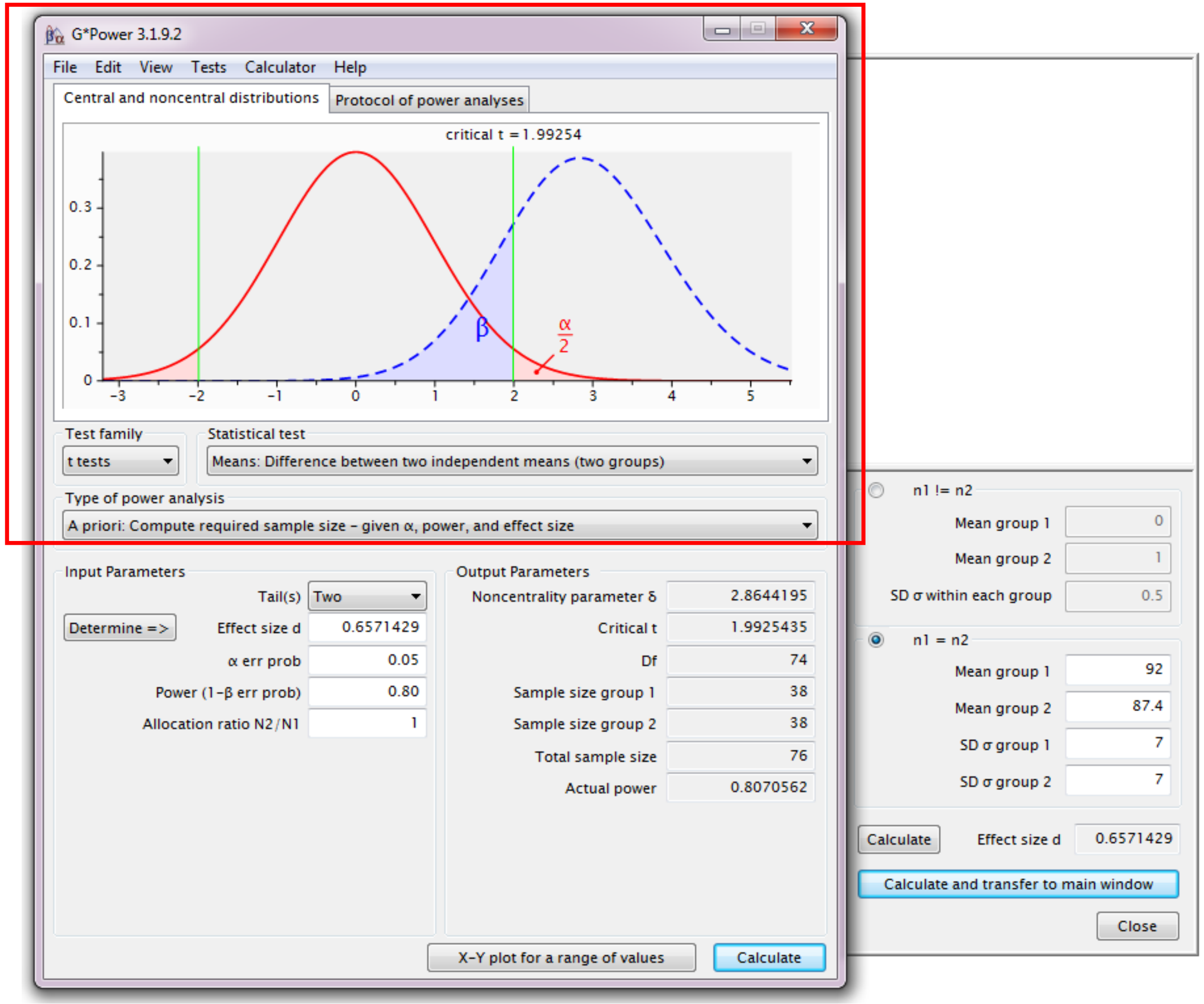
○ n1 != n2
Mean group 1: 0
Mean group 2: 1
SD σ within each group: 0.5

● n1 = n2
Mean group 1: 92
Mean group 2: 87.4
SD σ group 1: 7
SD σ group 2: 7

Effect size d: 0.6571429

You need a sample size of n=76 (2*38)

Power Analysis



Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

Tail(s): Two

Effect size d: 0.6571429

α err prob: 0.05

Power (1-β err prob): 0.80

Allocation ratio N2/N1: 1

Output Parameters

Noncentrality parameter δ: 2.8644195

Critical t: 1.9925435

Df: 74

Sample size group 1: 38

Sample size group 2: 38

Total sample size: 76

Actual power: 0.8070562

○ n1 != n2

Mean group 1: 0

Mean group 2: 1

SD σ within each group: 0.5

● n1 = n2

Mean group 1: 92

Mean group 2: 87.4

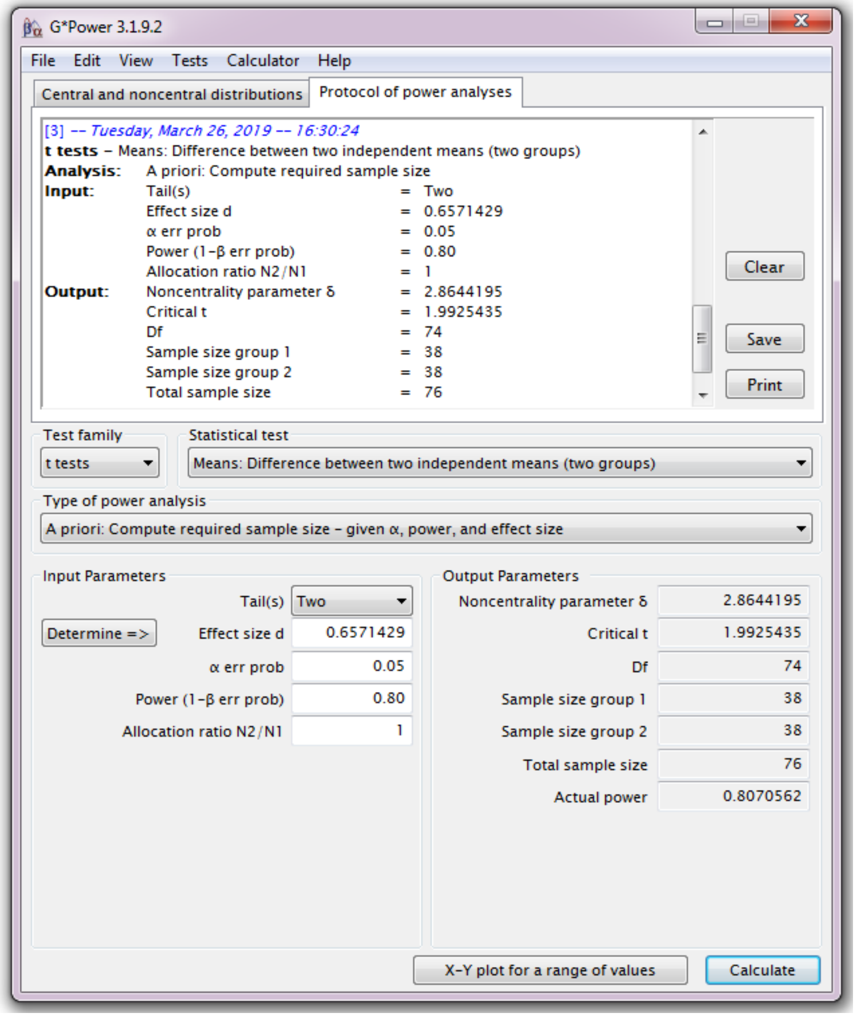
SD σ group 1: 7

SD σ group 2: 7

Calculate

Effect size d: 0.6571429

Calculate and transfer to main window



```
[3] -- Tuesday, March 26, 2019 -- 16:30:24
t tests - Means: Difference between two independent means (two groups)
Analysis:  A priori: Compute required sample size
Input:     Tail(s)                          =  Two
           Effect size d                    =  0.6571429
           α err prob                       =  0.05
           Power (1-β err prob)             =  0.80
           Allocation ratio N2/N1           =  1
Output:    Noncentrality parameter δ        =  2.8644195
           Critical t                       =  1.9925435
           Df                               =  74
           Sample size group 1              =  38
           Sample size group 2              =  38
           Total sample size                =  76
```

Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

Tail(s): Two

Effect size d: 0.6571429

α err prob: 0.05

Power (1-β err prob): 0.80

Allocation ratio N2/N1: 1

Output Parameters

Noncentrality parameter δ: 2.8644195

Critical t: 1.9925435

Df: 74

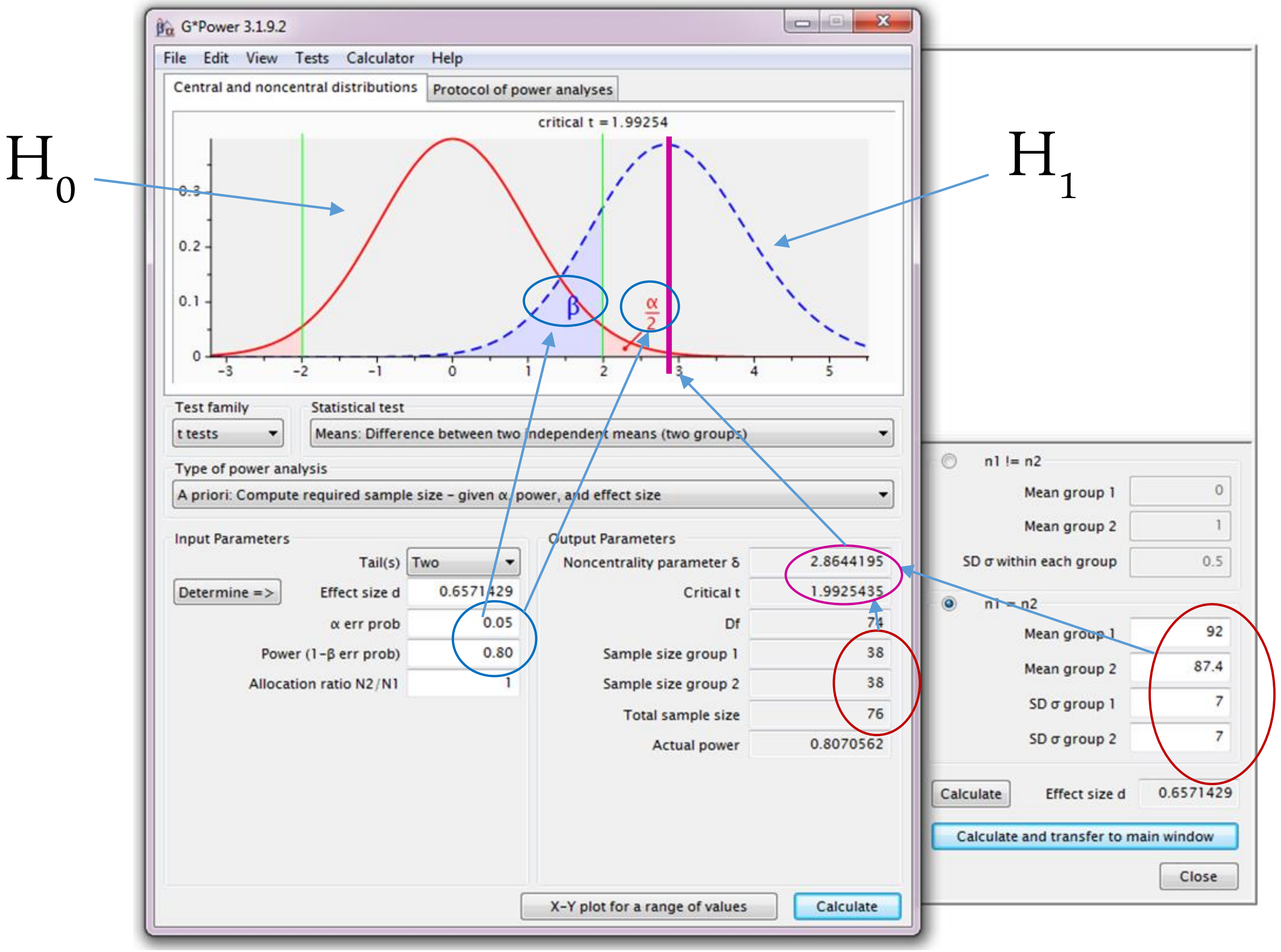
Sample size group 1: 38

Sample size group 2: 38

Total sample size: 76

Actual power: 0.8070562

Power Analysis



Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

Tail(s): Two
Effect size d: 0.6571429
α err prob: 0.05
Power (1-β err prob): 0.80
Allocation ratio N2/N1: 1

Output Parameters

Noncentrality parameter δ: 2.8644195
Critical t: 1.9925435
Df: 74
Sample size group 1: 38
Sample size group 2: 38
Total sample size: 76
Actual power: 0.8070562

○ n1 != n2

Mean group 1: 0
Mean group 2: 1
SD σ within each group: 0.5

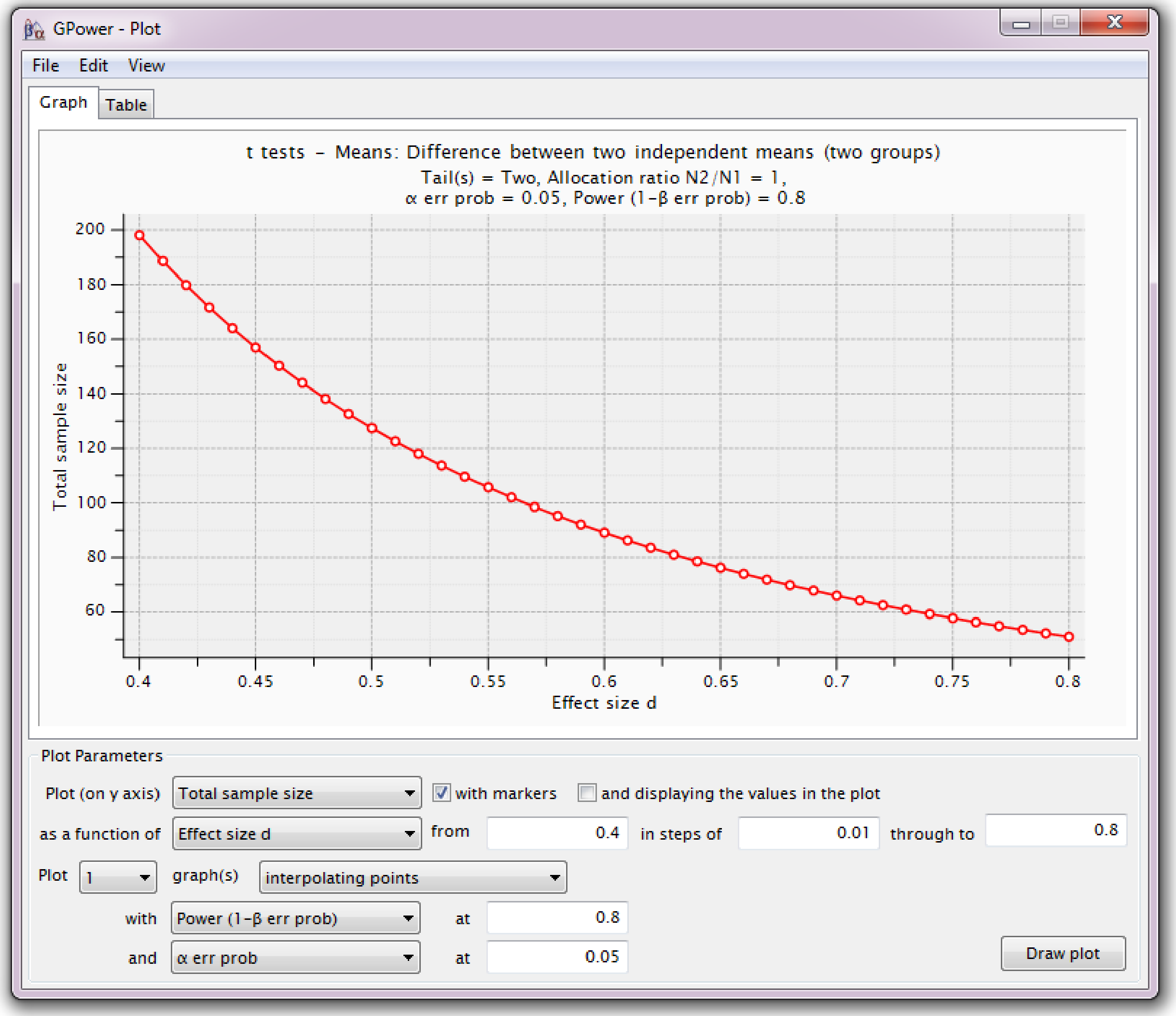
● n1 = n2

Mean group 1: 92
Mean group 2: 87.4
SD σ group 1: 7
SD σ group 2: 7

Effect size d: 0.6571429

Power Analysis

For a range of sample sizes:



t tests – Means: Difference between two independent means (two groups)
Tail(s) = Two, Allocation ratio N2/N1 = 1,
α err prob = 0.05, Power (1−β err prob) = 0.8

Plot Parameters

Plot (on y axis) Total sample size ☑ with markers ☐ and displaying the values in the plot

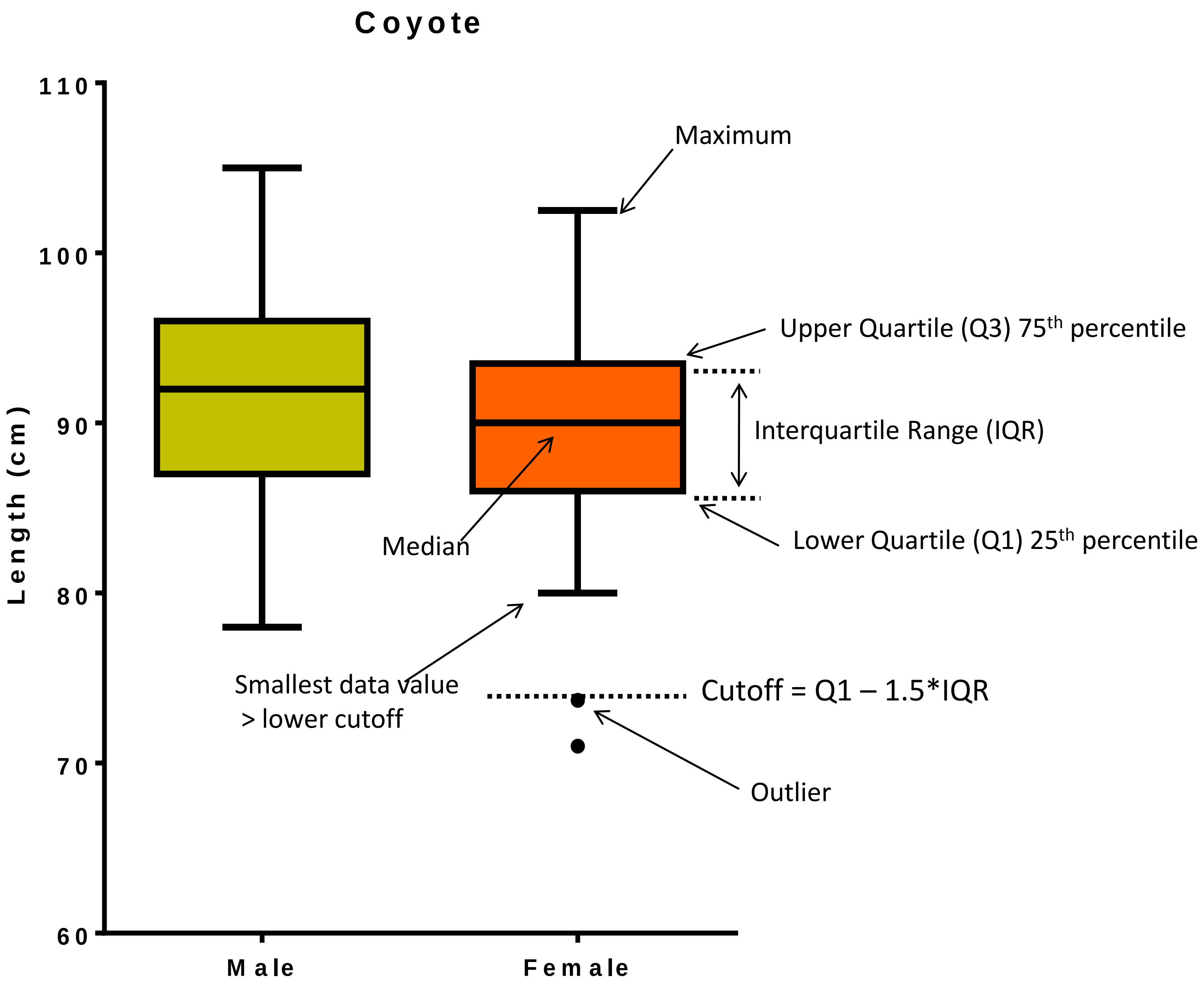
as a function of Effect size d from 0.4 in steps of 0.01 through to 0.8

Plot 1 graph(s) interpolating points

with Power (1−β err prob) at 0.8

and α err prob at 0.05

Data exploration ≠ plotting data



Maximum

Upper Quartile (Q3) 75th percentile

Interquartile Range (IQR)

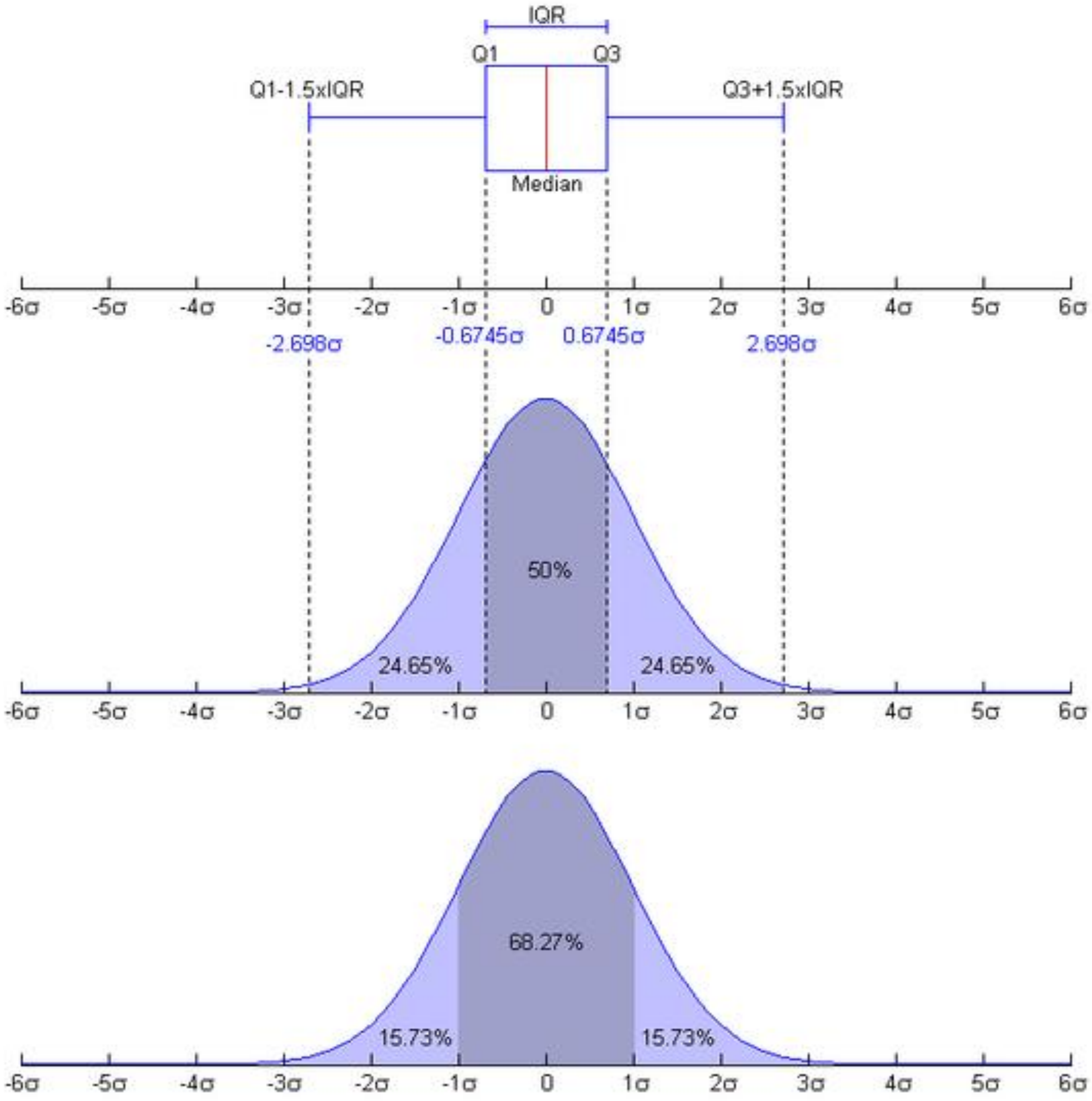
Median

Lower Quartile (Q1) 25th percentile

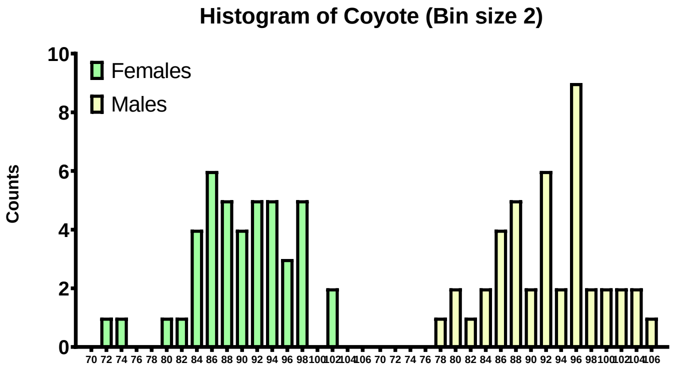
Smallest data value
> lower cutoff

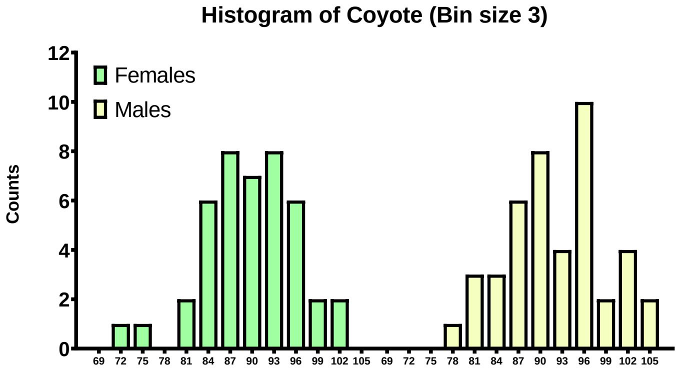
Cutoff = Q1 – 1.5*IQR

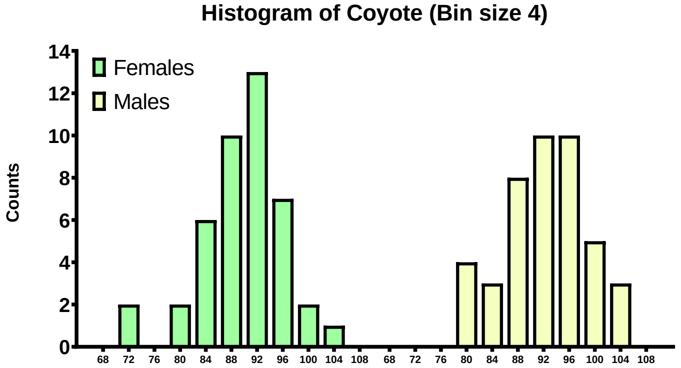
Outlier


IQR
Q1
Q3
Q1-1.5xIQR
Q3+1.5xIQR
Median
-6σ
-5σ
-4σ
-3σ
-2σ
-1σ
0
1σ
2σ
3σ
4σ
5σ
6σ
-2.698σ
-0.6745σ
0.6745σ
2.698σ
50%
24.65%
24.65%
-6σ
-5σ
-4σ
-3σ
-2σ
-1σ
0
1σ
2σ
3σ
4σ
5σ
6σ
68.27%
15.73%
15.73%
-6σ
-5σ
-4σ
-3σ
-2σ
-1σ
0
1σ
2σ
3σ
4σ
5σ
6σ

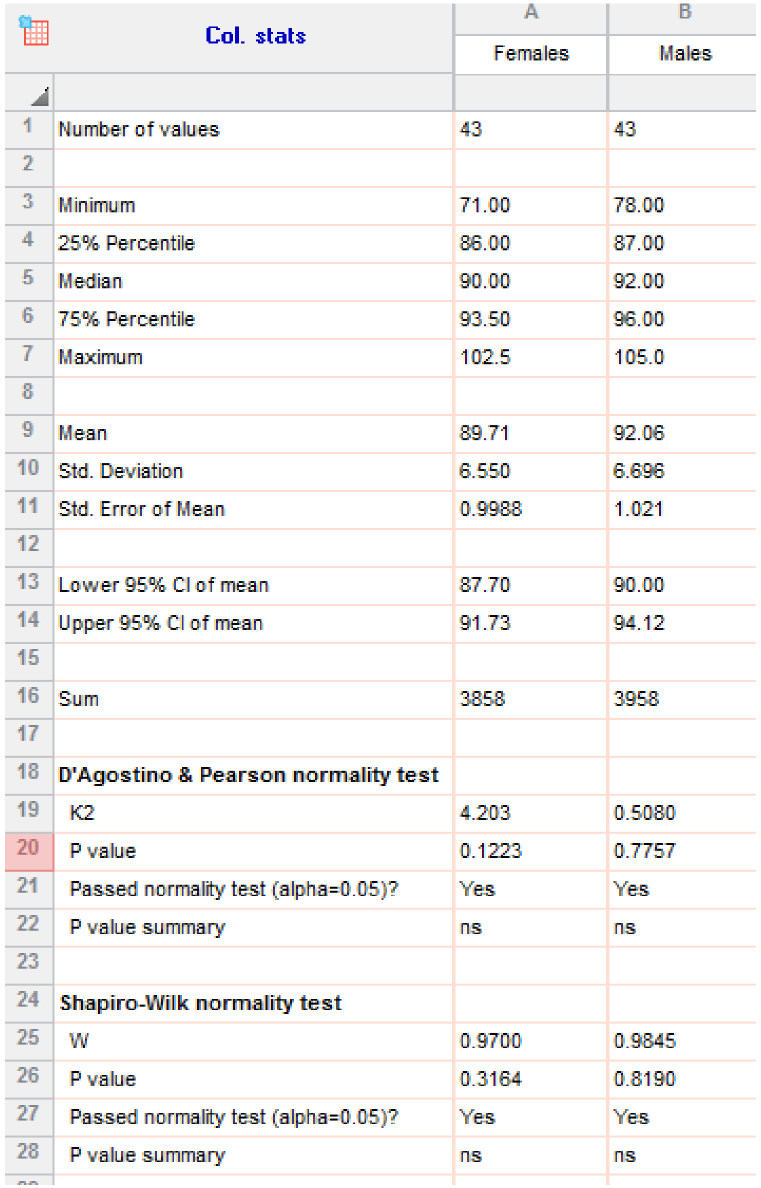
Assumptions for parametric tests





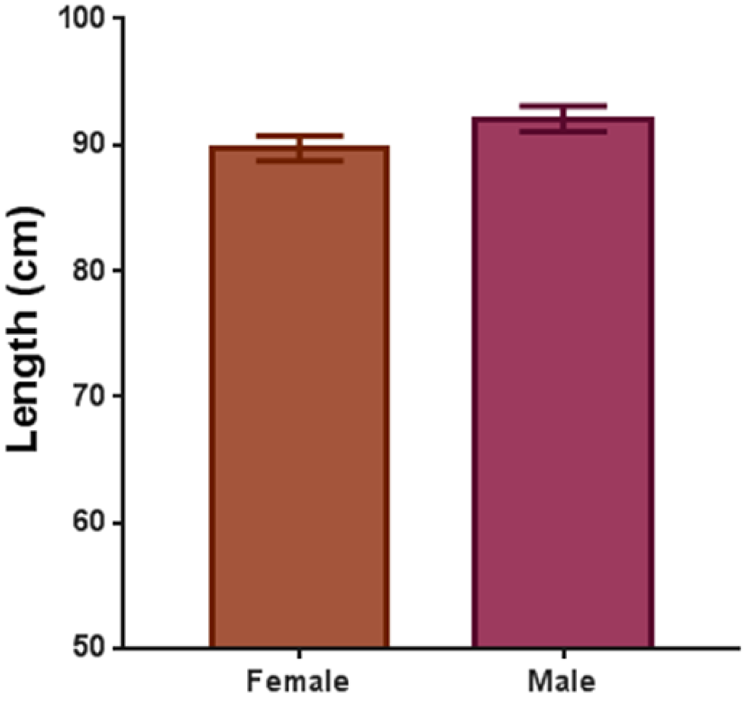


Normality ☑

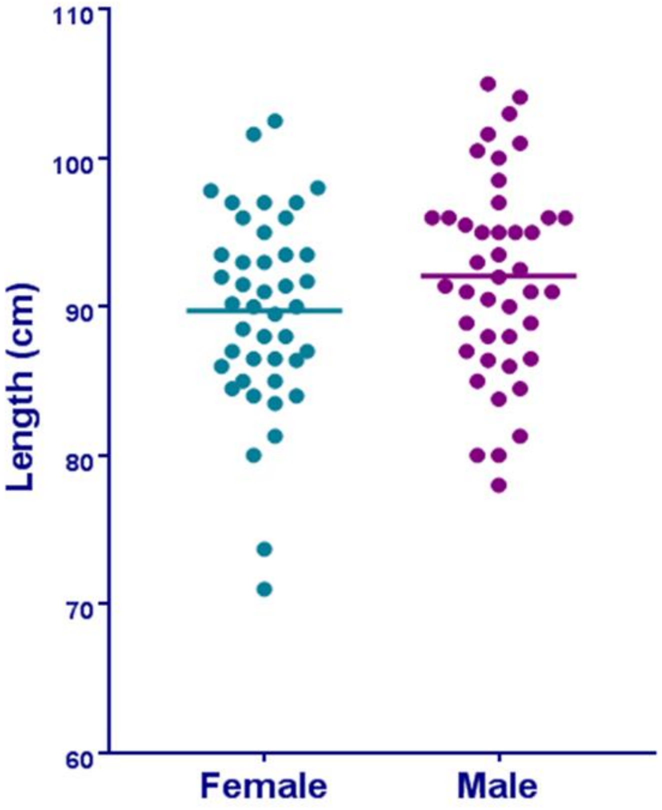


	Col. stats	A Females	B Males
1	Number of values	43	43
2			
3	Minimum	71.00	78.00
4	25% Percentile	86.00	87.00
5	Median	90.00	92.00
6	75% Percentile	93.50	96.00
7	Maximum	102.5	105.0
8			
9	Mean	89.71	92.06
10	Std. Deviation	6.550	6.696
11	Std. Error of Mean	0.9988	1.021
12			
13	Lower 95% CI of mean	87.70	90.00
14	Upper 95% CI of mean	91.73	94.12
15			
16	Sum	3858	3958
17			
18	**D'Agostino & Pearson normality test**		
19	K2	4.203	0.5080
20	P value	0.1223	0.7757
21	Passed normality test (alpha=0.05)?	Yes	Yes
22	P value summary	ns	ns
23			
24	**Shapiro-Wilk normality test**		
25	W	0.9700	0.9845
26	P value	0.3164	0.8190
27	Passed normality test (alpha=0.05)?	Yes	Yes
28	P value summary	ns	ns

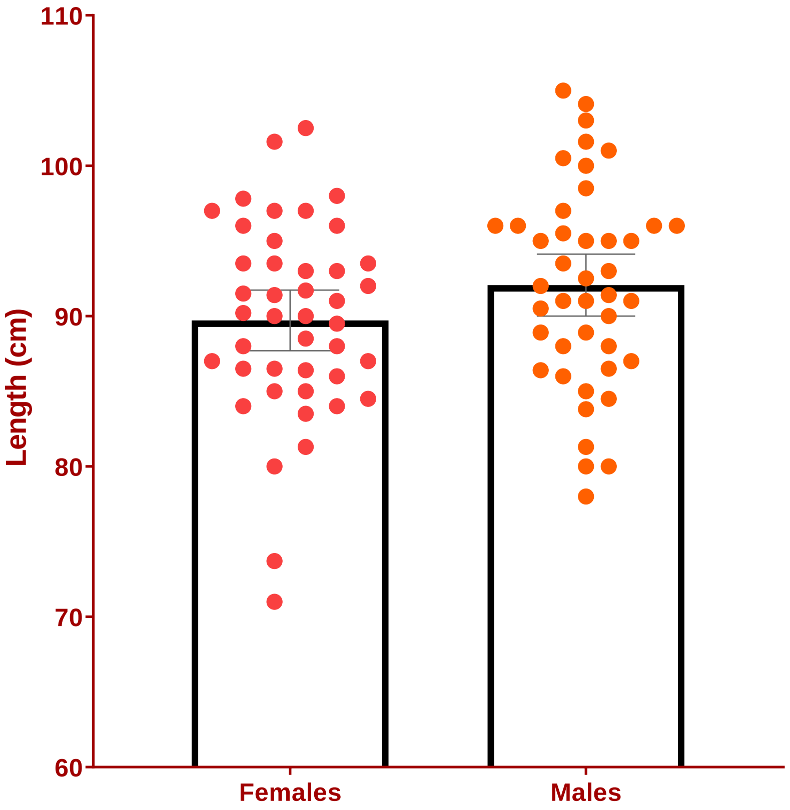
Coyotes



Length (cm)
100
90
80
70
60
50
Female
Male

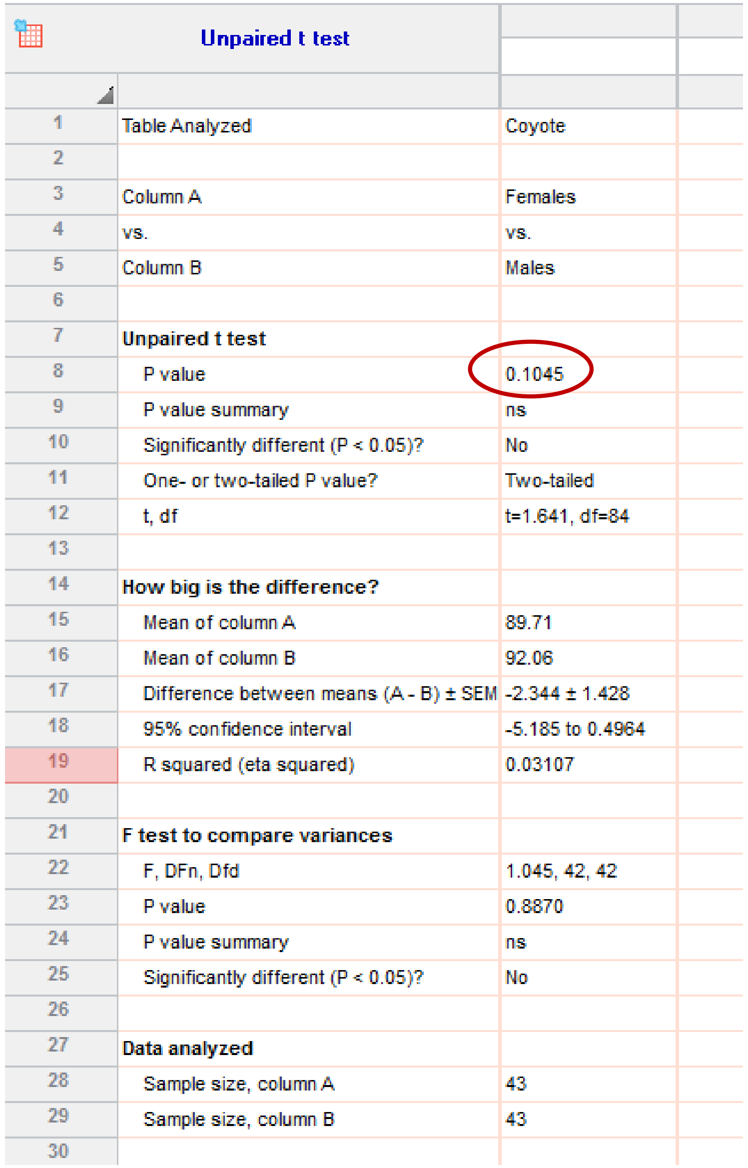


Length (cm)
110
100
90
80
70
60
Female
Male



Length (cm)
110
100
90
80
70
60
Females
Males

Independent *t*-test: results



	Unpaired t test	
1	Table Analyzed	Coyote
2		
3	Column A	Females
4	vs.	vs.
5	Column B	Males
6		
7	**Unpaired t test**	
8	P value	0.1045
9	P value summary	ns
10	Significantly different (P < 0.05)?	No
11	One- or two-tailed P value?	Two-tailed
12	t, df	t=1.641, df=84
13		
14	**How big is the difference?**	
15	Mean of column A	89.71
16	Mean of column B	92.06
17	Difference between means (A - B) ± SEM	-2.344 ± 1.428
18	95% confidence interval	-5.185 to 0.4964
19	R squared (eta squared)	0.03107
20		
21	**F test to compare variances**	
22	F, DFn, Dfd	1.045, 42, 42
23	P value	0.8870
24	P value summary	ns
25	Significantly different (P < 0.05)?	No
26		
27	**Data analyzed**	
28	Sample size, column A	43
29	Sample size, column B	43
30		

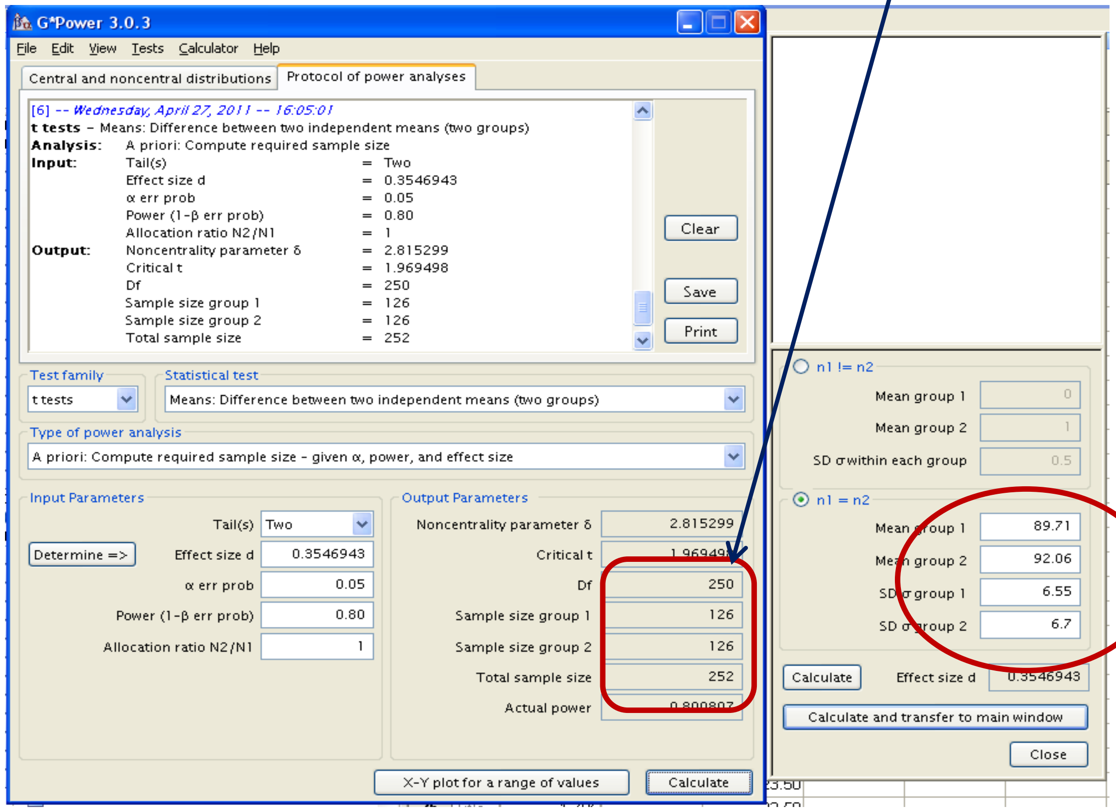
Males tend to be longer than females but not significantly so (p=0.1045)

Homogeneity in variance ☑

What about the power of the analysis?

Power analysis

You would need a sample 3 times bigger to reach the accepted power of 80%.



```
[6] -- Wednesday, April 27, 2011 -- 16:05:01
t tests - Means: Difference between two independent means (two groups)
Analysis:   A priori: Compute required sample size
Input:      Tail(s)                        =  Two
            Effect size d                  =  0.3546943
            α err prob                     =  0.05
            Power (1-β err prob)           =  0.80
            Allocation ratio N2/N1         =  1
Output:     Noncentrality parameter δ      =  2.815299
            Critical t                     =  1.969498
            Df                             =  250
            Sample size group 1            =  126
            Sample size group 2            =  126
            Total sample size              =  252
```

Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

- Tail(s): Two
- Effect size d: 0.3546943
- α err prob: 0.05
- Power (1-β err prob): 0.80
- Allocation ratio N2/N1: 1

Output Parameters

- Noncentrality parameter δ: 2.815299
- Critical t: 1.969498
- Df: 250
- Sample size group 1: 126
- Sample size group 2: 126
- Total sample size: 252
- Actual power: 0.800807

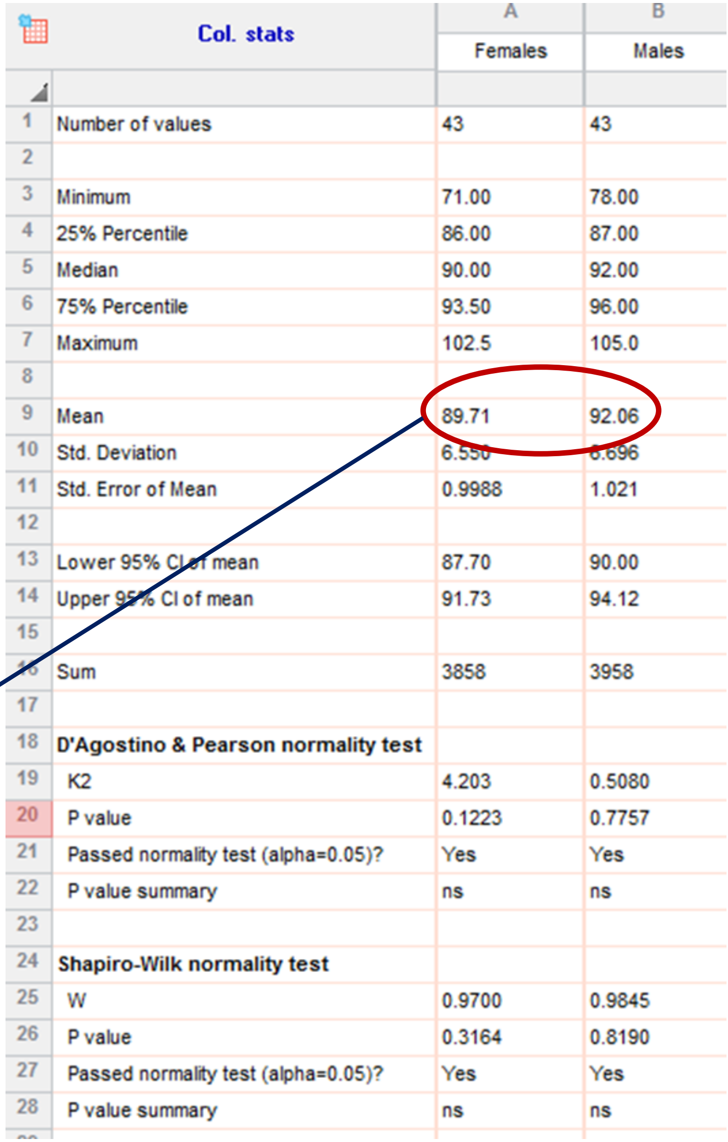
○ n1 != n2

- Mean group 1: 0
- Mean group 2: 1
- SD σ within each group: 0.5

● n1 = n2

- Mean group 1: 89.71
- Mean group 2: 92.06
- SD σ group 1: 6.55
- SD σ group 2: 6.7

Effect size d: 0.3546943



	Col. stats	A Females	B Males
1	Number of values	43	43
2			
3	Minimum	71.00	78.00
4	25% Percentile	86.00	87.00
5	Median	90.00	92.00
6	75% Percentile	93.50	96.00
7	Maximum	102.5	105.0
8			
9	Mean	89.71	92.06
10	Std. Deviation	6.550	6.696
11	Std. Error of Mean	0.9988	1.021
12			
13	Lower 95% CI of mean	87.70	90.00
14	Upper 95% CI of mean	91.73	94.12
15			
16	Sum	3858	3958
17			
18	D'Agostino & Pearson normality test		
19	K2	4.203	0.5080
20	P value	0.1223	0.7757
21	Passed normality test (alpha=0.05)?	Yes	Yes
22	P value summary	ns	ns
23			
24	Shapiro-Wilk normality test		
25	W	0.9700	0.9845
26	P value	0.3164	0.8190
27	Passed normality test (alpha=0.05)?	Yes	Yes
28	P value summary	ns	ns

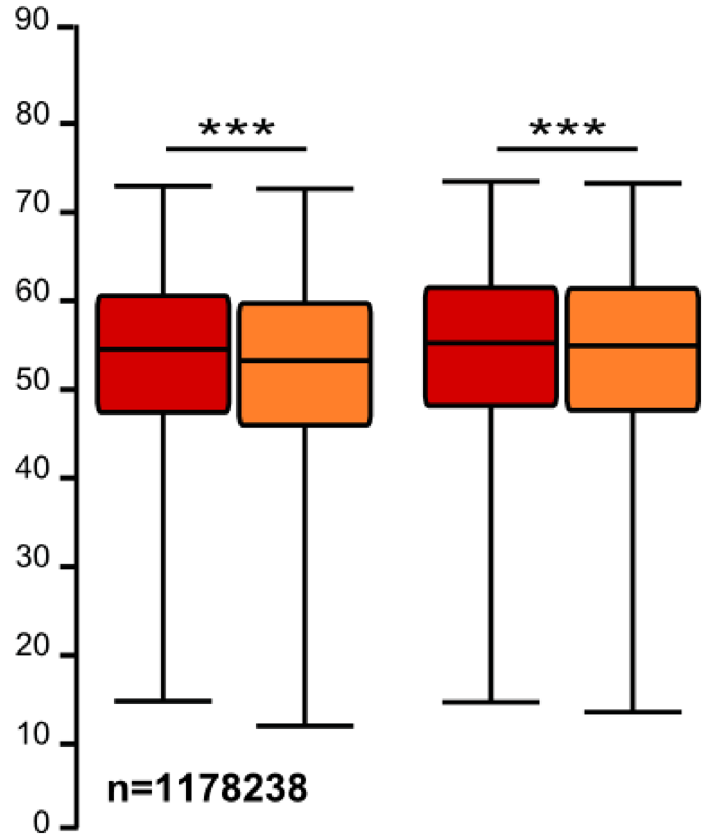
But is a 2.3 cm difference between genders biologically relevant (<3%) ?

The sample size: the bigger the better?

- It takes huge samples to detect tiny differences but tiny samples to detect huge differences.

- What if the tiny difference is meaningless?
  - Beware of **overpower**
  - Nothing wrong with the stats: it is all about interpretation of the results of the test.

- Remember the important first step of power analysis
  - **What is the effect size of biological interest?**



Another example of *t*-test:

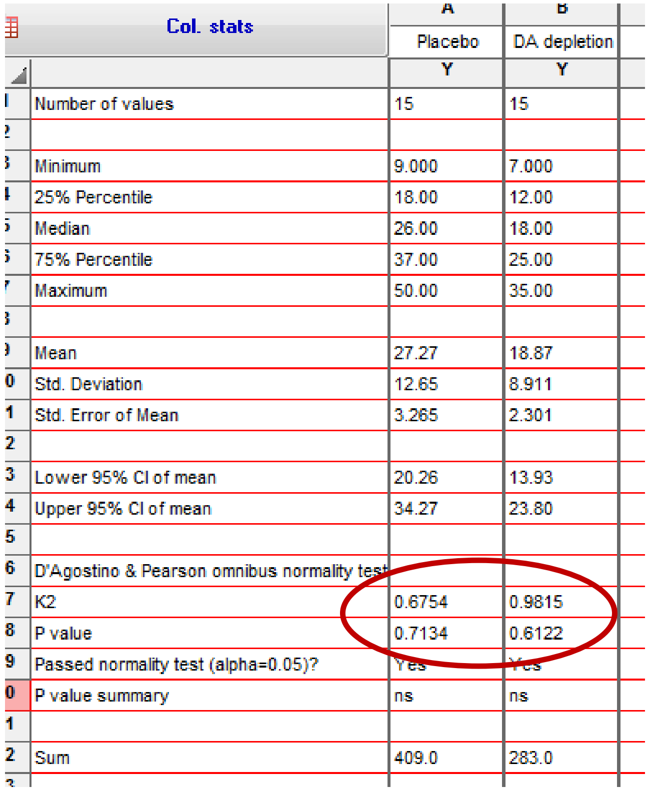
working memory.xlsx

A group of rhesus monkeys (n=15) performs a task involving memory after having received a placebo. Their performance is graded on a scale from 0 to 100. They are then asked to perform the same task after having received a dopamine depleting agent.

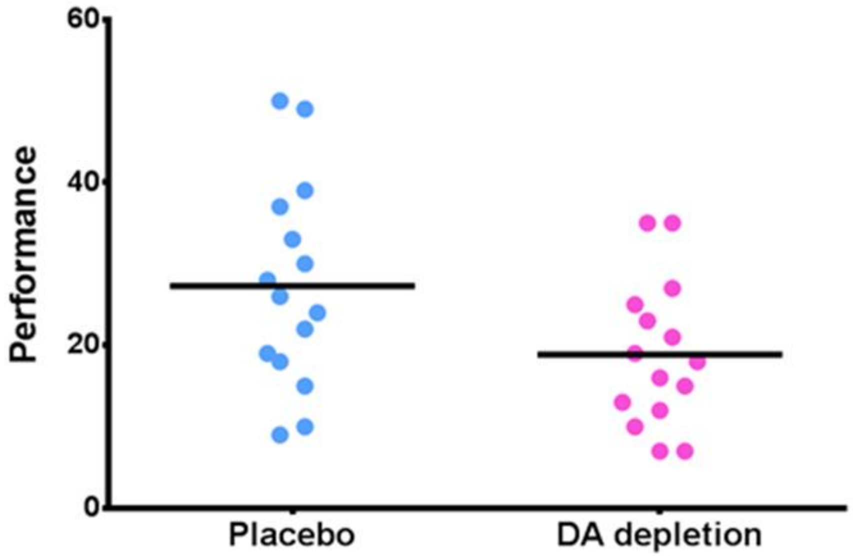
Is there an effect of treatment on the monkeys' performance?

Another example of *t*-test:

working memory.xlsx



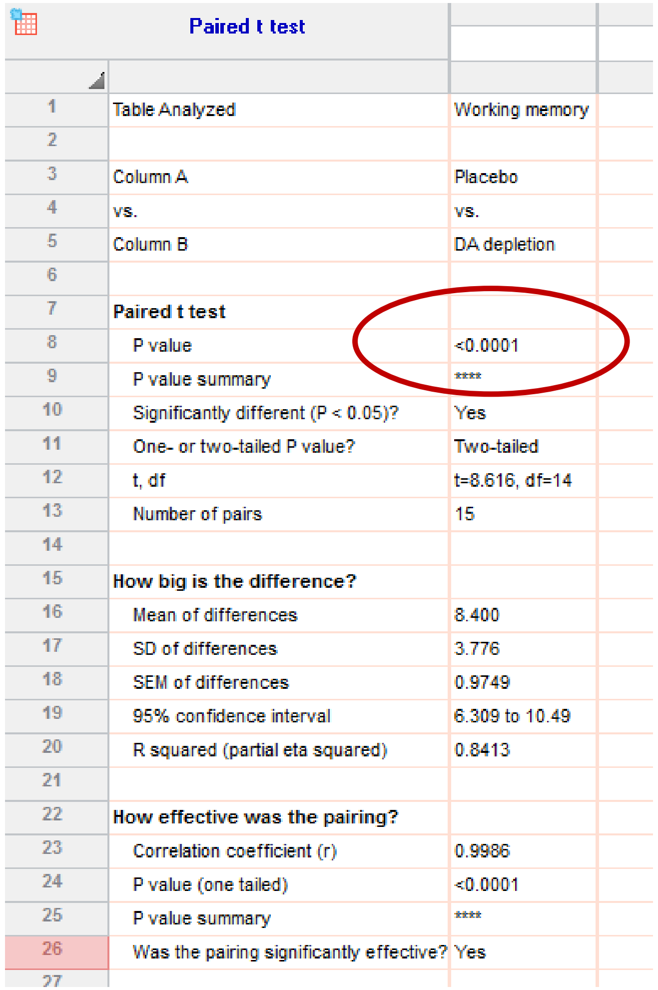
Col. stats	A Placebo Y	B DA depletion Y
Number of values	15	15
Minimum	9.000	7.000
25% Percentile	18.00	12.00
Median	26.00	18.00
75% Percentile	37.00	25.00
Maximum	50.00	35.00
Mean	27.27	18.87
Std. Deviation	12.65	8.911
Std. Error of Mean	3.265	2.301
Lower 95% CI of mean	20.26	13.93
Upper 95% CI of mean	34.27	23.80
D'Agostino & Pearson omnibus normality test		
K2	0.6754	0.9815
P value	0.7134	0.6122
Passed normality test (alpha=0.05)?	Yes	Yes
P value summary	ns	ns
Sum	409.0	283.0



Normality ☑

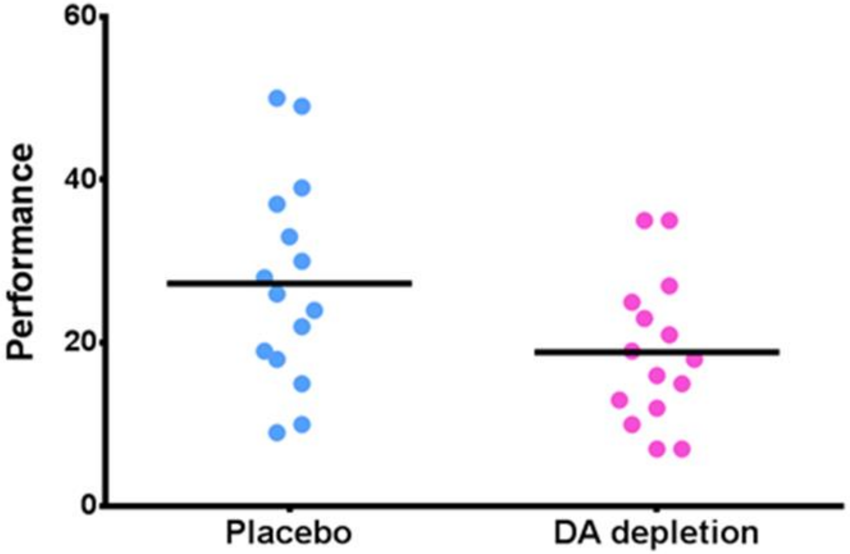
Another example of *t*-test:

working memory.xlsx



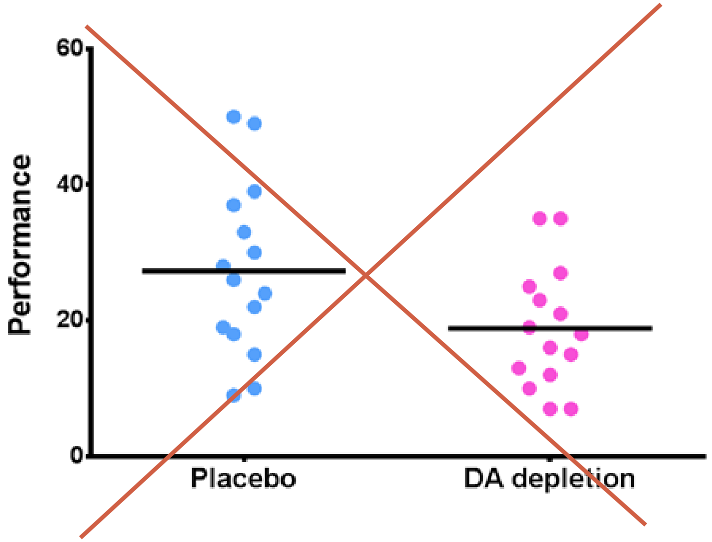
Paired t test

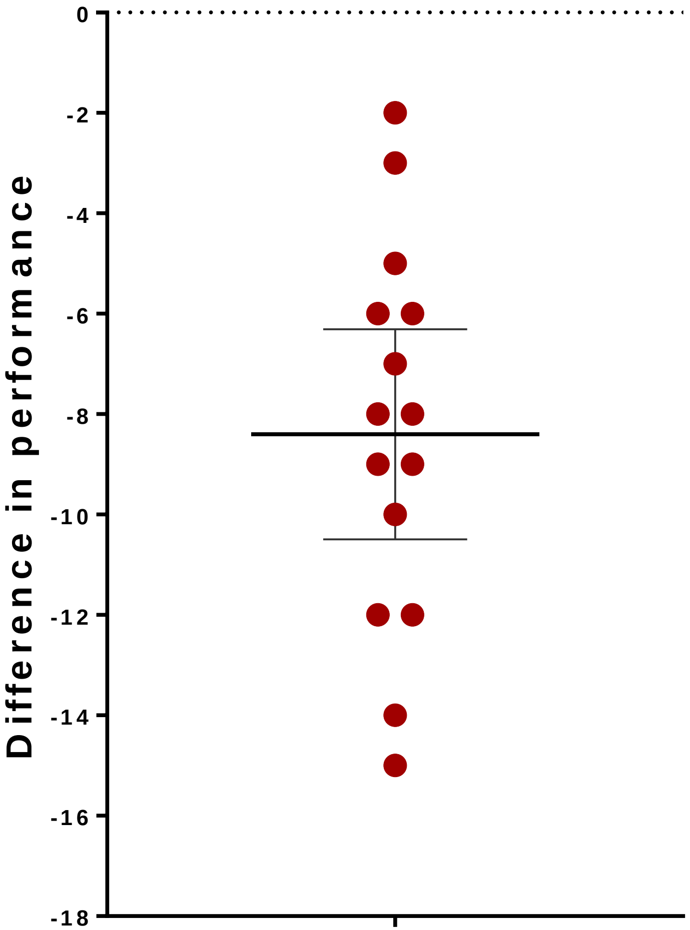
1	Table Analyzed	Working memory
2		
3	Column A	Placebo
4	vs.	vs.
5	Column B	DA depletion
6		
7	**Paired t test**	
8	P value	<0.0001
9	P value summary	****
10	Significantly different (P < 0.05)?	Yes
11	One- or two-tailed P value?	Two-tailed
12	t, df	t=8.616, df=14
13	Number of pairs	15
14		
15	**How big is the difference?**	
16	Mean of differences	8.400
17	SD of differences	3.776
18	SEM of differences	0.9749
19	95% confidence interval	6.309 to 10.49
20	R squared (partial eta squared)	0.8413
21		
22	**How effective was the pairing?**	
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		



Paired *t*-test: Results

working memory.xlsx





Comparison of more than 2 means

- Running multiple tests on the same data increases the **familywise error rate**.
- What is the familywise error rate?
  - The error rate across tests conducted on the same experimental data.
- One of the basic rules ('laws') of probability:
  - The Multiplicative Rule: The probability of the joint occurrence of 2 or more independent events is the product of the individual probabilities.

$$\boxed{P(A,B) = P(A) \times P(B)}$$

For example:

$$P(\text{2 Heads}) = P(\text{head}) \times P(\text{head}) = 0.5 \times 0.5 = 0.25$$

Familywise error rate

- **Example**: All pairwise comparisons between 3 groups A, B and C:
  - A-B, A-C and B-C

- Probability of making the Type I Error: **5%**
  - The probability of not making the Type I Error is 95% (=1 – 0.05)

- Multiplicative Rule:
  - Overall probability of no Type I errors is: 0.95 * 0.95 * 0.95 = 0.857

- So the probability of making at least one Type I Error is 1-0.857 = 0.143 or **14.3%**
    - The probability has increased from 5% to 14.3%

- Comparisons between 5 groups instead of 3, the familywise error rate is **40%** (=1-$(0.95)^n$)

Familywise error rate

- Solution to the increase of familywise error rate: correction for multiple comparisons
  - **Post-hoc tests**

- Many different ways to correct for multiple comparisons:
  - Different statisticians have designed corrections addressing different issues
    - e.g. unbalanced design, heterogeneity of variance, liberal vs conservative

- However, they all have **one thing in common**:
  - the more tests, the higher the familywise error rate: the more stringent the correction

- Tukey, Bonferroni, Sidak, Benjamini-Hochberg …
  - Two ways to address the multiple testing problem
    - **Familywise Error Rate** (FWER) vs. **False Discovery Rate** (FDR)

Multiple testing problem

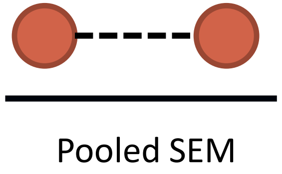
- **FWER**: **Bonferroni**: α_{adjust} = 0.05/n comparisons e.g. 3 comparisons: 0.05/3=0.016
  - Problem: very conservative leading to loss of power (lots of false negative)
  - 10 comparisons: threshold for significance: 0.05/10: 0.005
  - Pairwise comparisons across 20.000 genes ☹

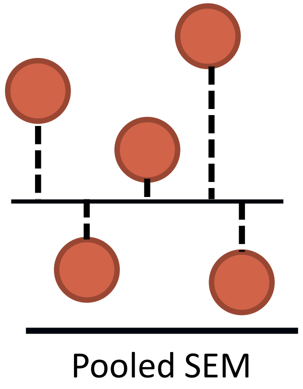
- **FDR**: **Benjamini-Hochberg**: the procedure controls the expected proportion of “discoveries” (significant tests) that are false (false positive).
  - Less stringent control of Type I Error than FWER procedures which control the probability of at least one Type I Error
  - More power at the cost of increased numbers of Type I Errors.

- **Difference between FWER and FDR**:
  - a p-value of 0.05 implies that 5% of all tests will result in false positives.
  - a FDR adjusted p-value (or **q-value**) of 0.05 implies that 5% of significant tests will result in false positives.

Analysis of variance

- Extension of the 2 groups comparison of a *t*-test but with a slightly different logic:
- *t*-test = $\frac{\text{mean1 – mean2}}{\text{Pooled SEM}}$





- ANOVA = $\frac{\text{variance between means}}{\text{Pooled SEM}}$
- ANOVA compares variances:
  - If variance between the several means > variance within the groups (random error) then the means must be more spread out than it would have been by chance.

Analysis of variance

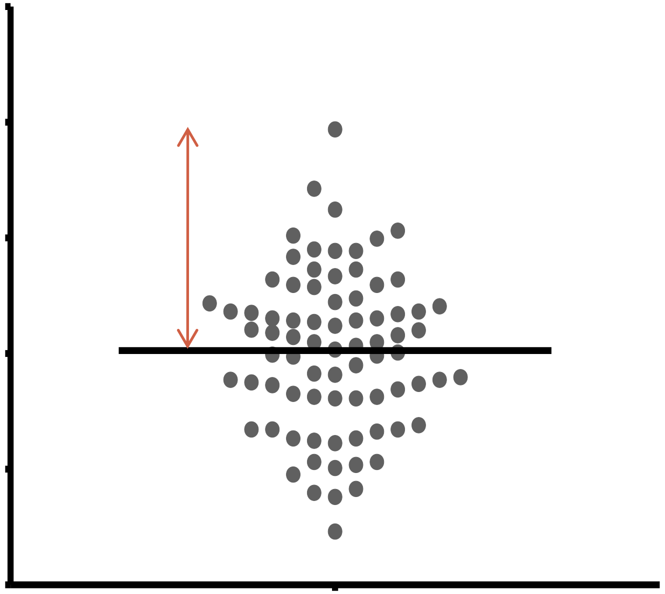
- The statistic for ANOVA is the **F ratio**.
- F = $\dfrac{\text{Variance between the groups}}{\text{Variance within the groups (individual variability)}}$
- F = $\dfrac{\text{Variation explained by the model (= systematic)}}{\text{Variation explained by unsystematic factors (= random variation)}}$
- If the variance amongst sample means is greater than the error/random variance, then F>1
  - In an ANOVA, we test whether F is significantly higher than 1 or not.

Analysis of variance

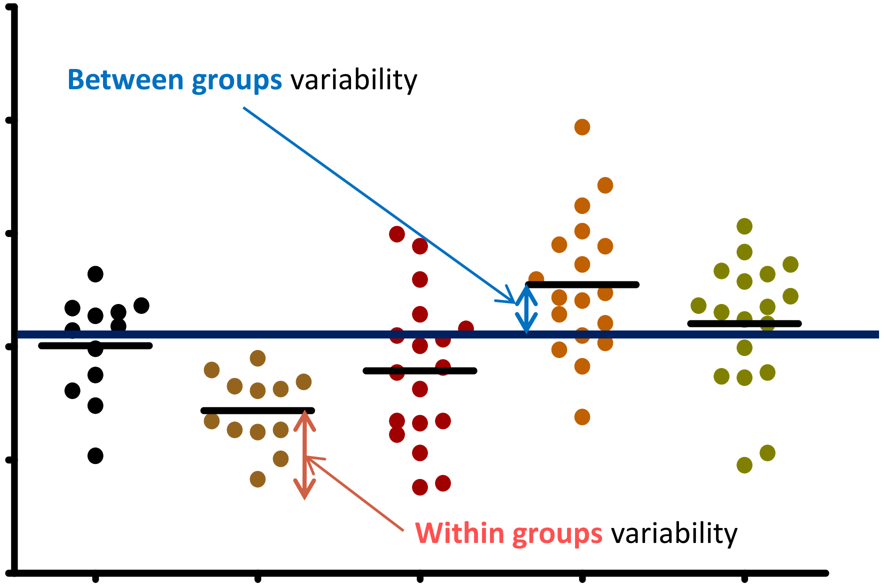
Source of variation	**Sum of Squares**	**df**	**Mean Square**	**F**	**p-value**
Between Groups	2.665	4	0.6663	8.423	<0.0001
Within Groups	5.775	73	0.0791		
Total	8.44	77			

In Power Analysis:
Pooled SD=√MS(Residual)

- Variance (= SS / N-1) is the mean square
  - df: degree of freedom with df = N-1



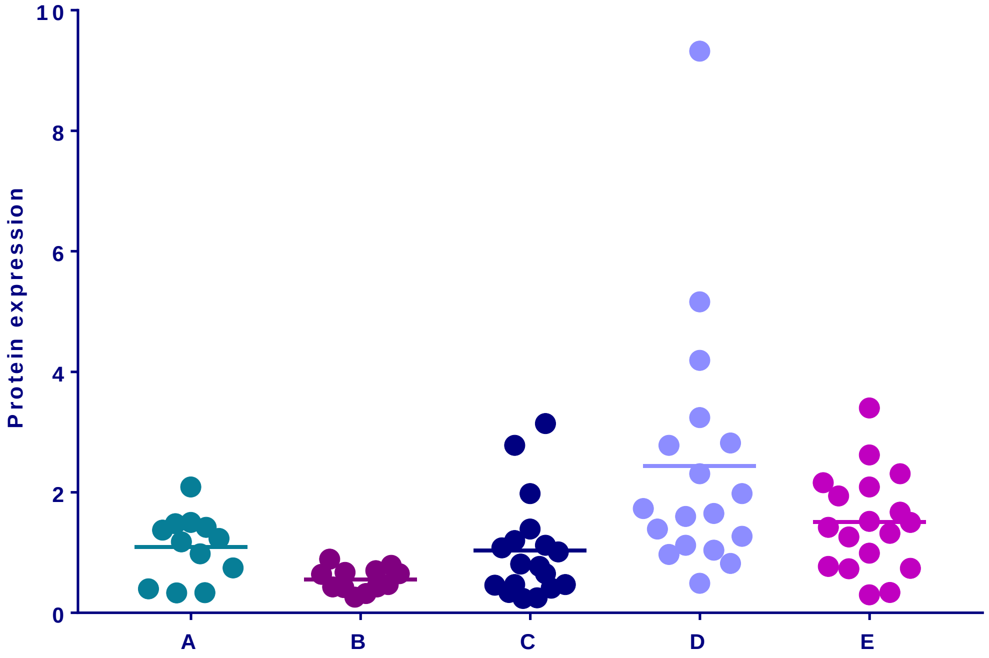
Total sum of squares



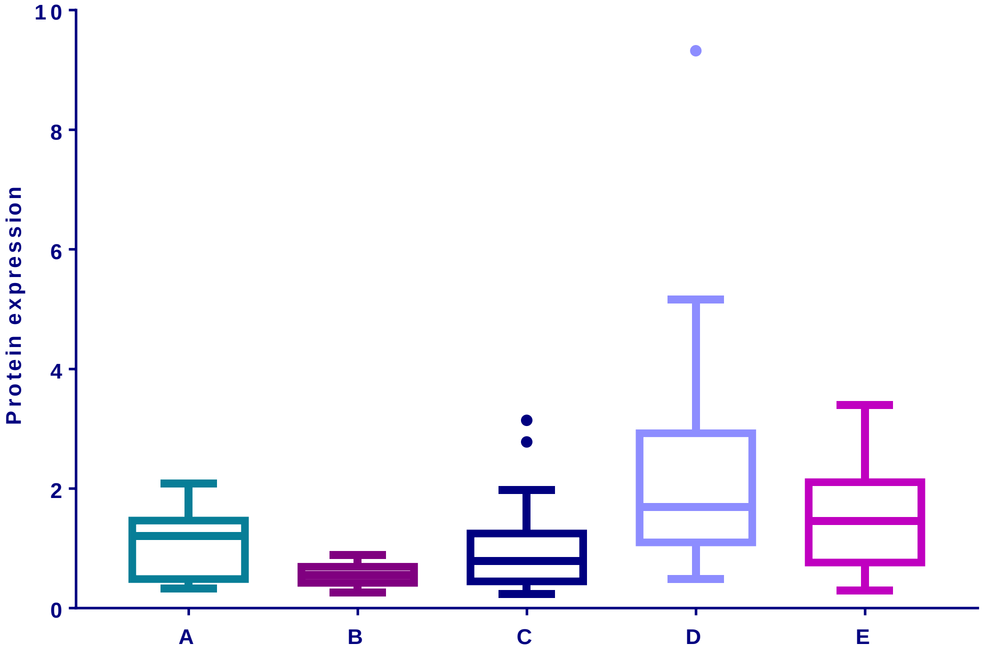
Example: protein.expression.csv

- **Question**: is there a difference in protein expression between the 5 cell lines?

- **1 Plot the data**
- **2 Check the assumptions for parametric test**
- **3 Statistical analysis: ANOVA**

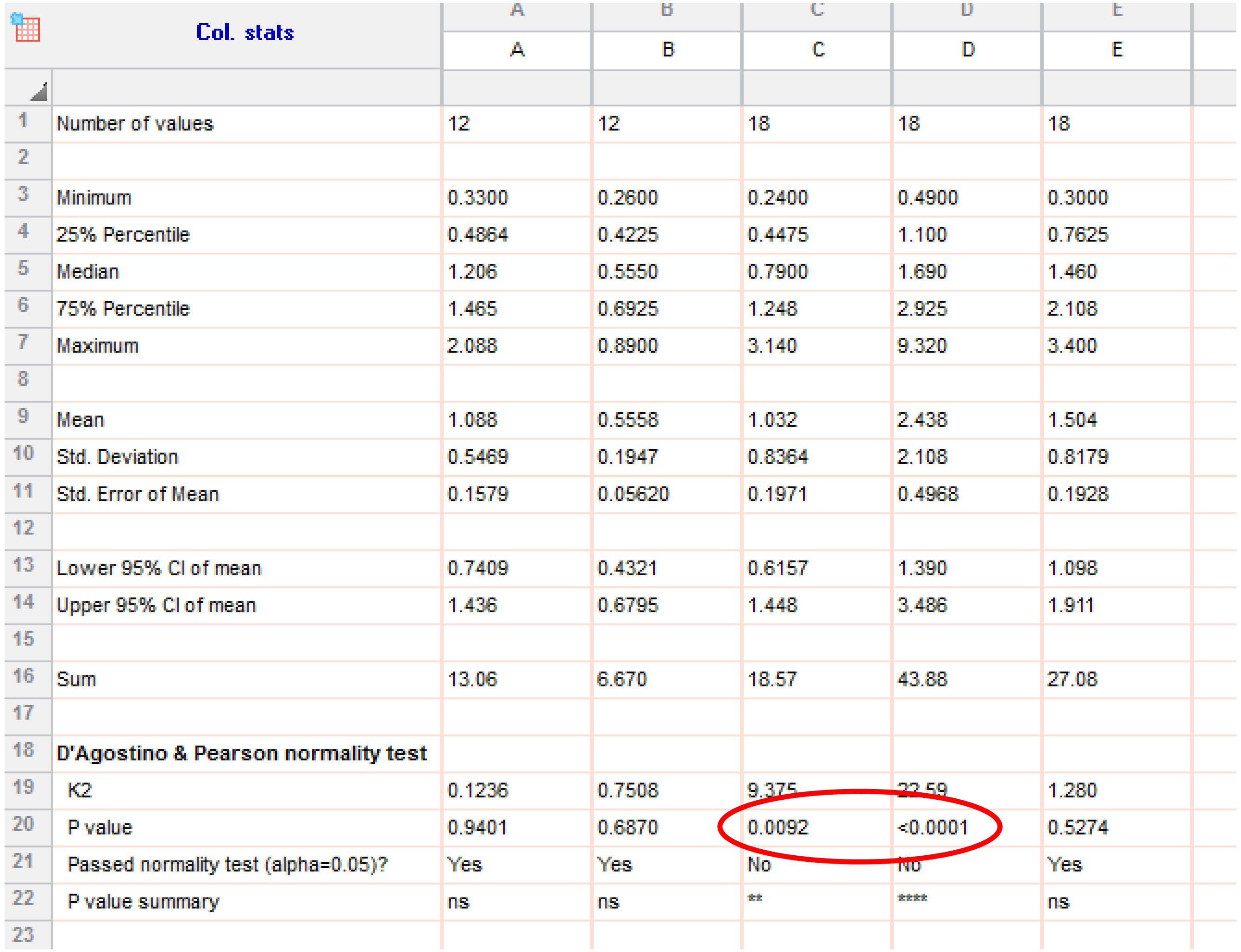


Protein expression
0
2
4
6
8
10
A
B
C
D
E

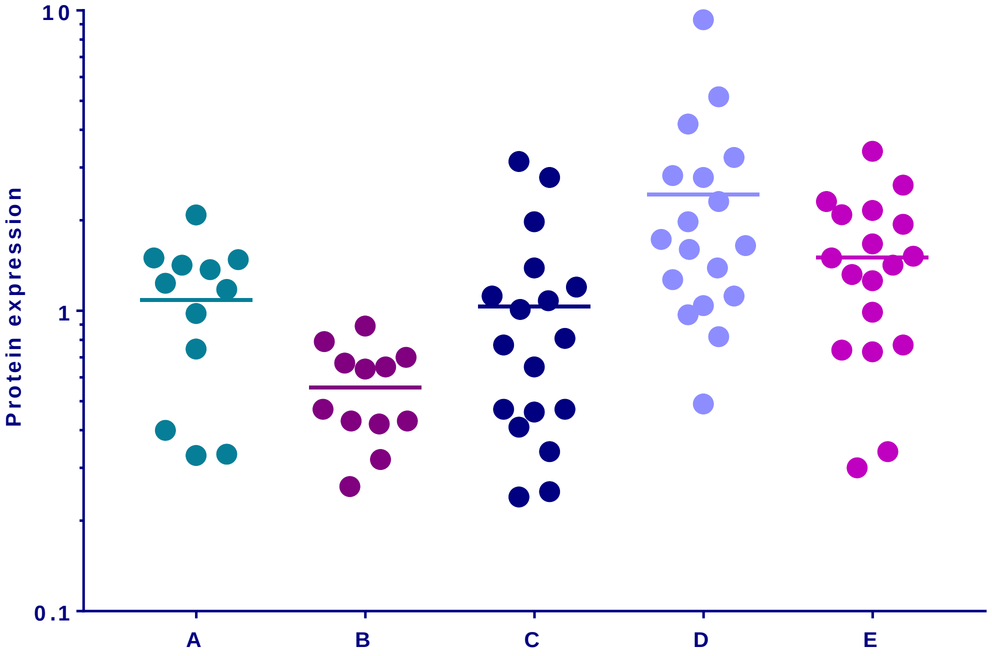


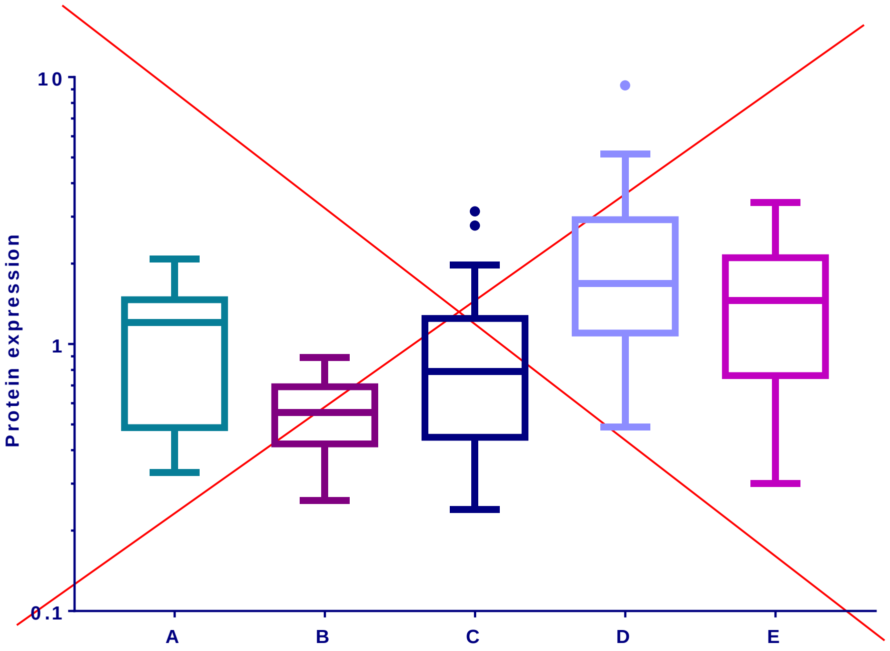
Protein expression
0
2
4
6
8
10
A
B
C
D
E

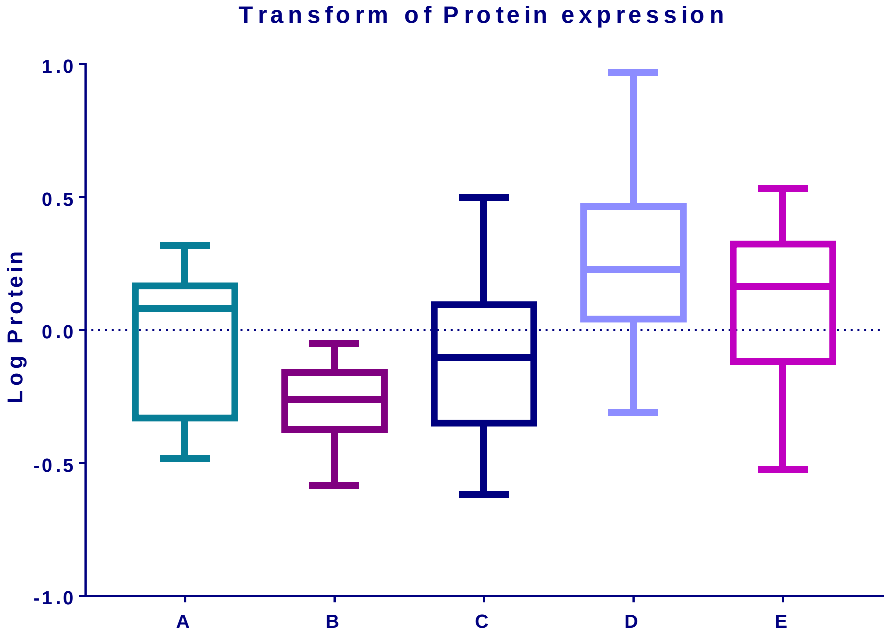
Parametric tests assumptions



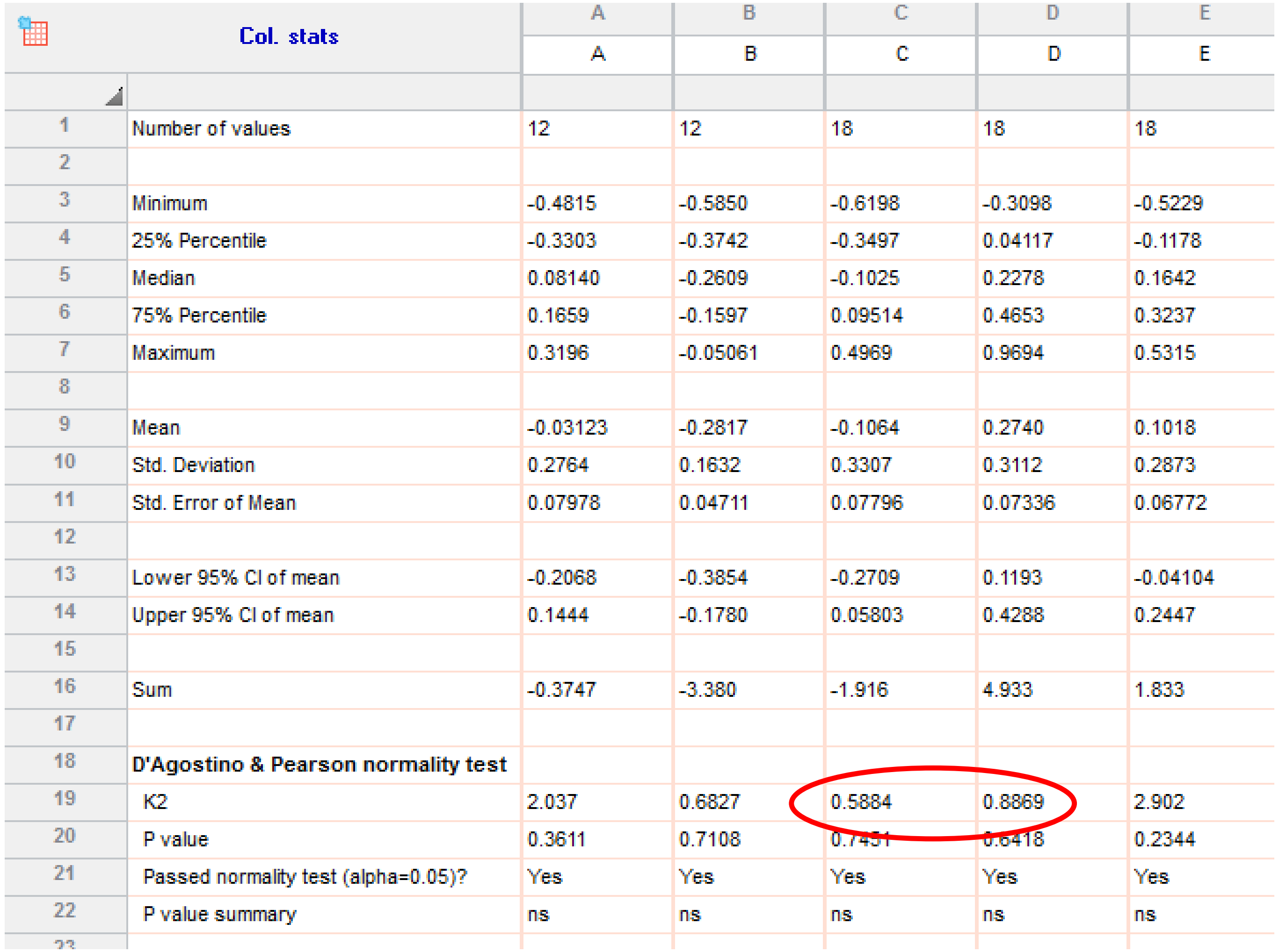
	Col. stats	A	B	C	D	E
		A	B	C	D	E
1	Number of values	12	12	18	18	18
2						
3	Minimum	0.3300	0.2600	0.2400	0.4900	0.3000
4	25% Percentile	0.4864	0.4225	0.4475	1.100	0.7625
5	Median	1.206	0.5550	0.7900	1.690	1.460
6	75% Percentile	1.465	0.6925	1.248	2.925	2.108
7	Maximum	2.088	0.8900	3.140	9.320	3.400
8						
9	Mean	1.088	0.5558	1.032	2.438	1.504
10	Std. Deviation	0.5469	0.1947	0.8364	2.108	0.8179
11	Std. Error of Mean	0.1579	0.05620	0.1971	0.4968	0.1928
12						
13	Lower 95% CI of mean	0.7409	0.4321	0.6157	1.390	1.098
14	Upper 95% CI of mean	1.436	0.6795	1.448	3.486	1.911
15						
16	Sum	13.06	6.670	18.57	43.88	27.08
17						
18	**D'Agostino & Pearson normality test**					
19	K2	0.1236	0.7508	9.375	22.59	1.280
20	P value	0.9401	0.6870	0.0092	<0.0001	0.5274
21	Passed normality test (alpha=0.05)?	Yes	Yes	No	No	Yes
22	P value summary	ns	ns	**	****	ns
23						


10
1
0.1
Protein expression
A
B
C
D
E


10
1
0.1
Protein expression
A
B
C
D
E


Transform of Protein expression
1.0
0.5
0.0
-0.5
-1.0
Log Protein
A
B
C
D
E

Parametric tests assumptions

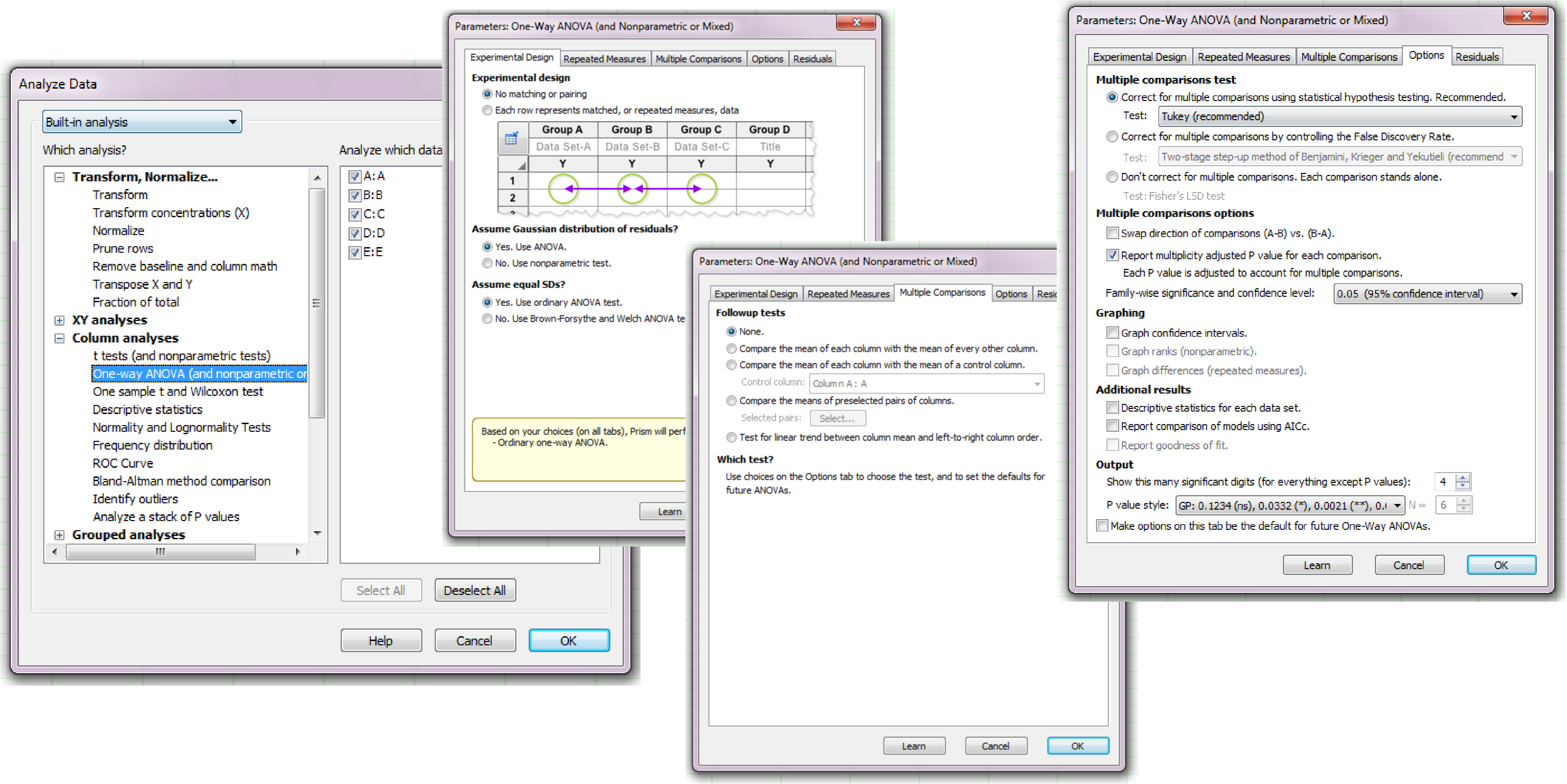


	Col. stats	A	B	C	D	E
		A	B	C	D	E
1	Number of values	12	12	18	18	18
2						
3	Minimum	-0.4815	-0.5850	-0.6198	-0.3098	-0.5229
4	25% Percentile	-0.3303	-0.3742	-0.3497	0.04117	-0.1178
5	Median	0.08140	-0.2609	-0.1025	0.2278	0.1642
6	75% Percentile	0.1659	-0.1597	0.09514	0.4653	0.3237
7	Maximum	0.3196	-0.05061	0.4969	0.9694	0.5315
8						
9	Mean	-0.03123	-0.2817	-0.1064	0.2740	0.1018
10	Std. Deviation	0.2764	0.1632	0.3307	0.3112	0.2873
11	Std. Error of Mean	0.07978	0.04711	0.07796	0.07336	0.06772
12						
13	Lower 95% CI of mean	-0.2068	-0.3854	-0.2709	0.1193	-0.04104
14	Upper 95% CI of mean	0.1444	-0.1780	0.05803	0.4288	0.2447
15						
16	Sum	-0.3747	-3.380	-1.916	4.933	1.833
17						
18	**D'Agostino & Pearson normality test**					
19	K2	2.037	0.6827	0.5884	0.8869	2.902
20	P value	0.3611	0.7108	0.7451	0.6418	0.2344
21	Passed normality test (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes
22	P value summary	ns	ns	ns	ns	ns

Analysis of variance: Post hoc tests

- The ANOVA is an “omnibus” test: it tells you that there is (or not) a difference between your means but not exactly which means are significantly different from which other ones.
  - To find out, you need to apply **post hoc** tests.
  - These post hoc tests should only be used when the ANOVA finds a significant effect.

Analysis of variancec



Analyze Data

Built-in analysis

Which analysis?

- ⊟ **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
- ⊞ **Grouped analyses**

Analyze which data

☑ A:A
☑ B:B
☑ C:C
☑ D:D
☑ E:E

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Experimental design

● No matching or pairing
○ Each row represents matched, or repeated measures, data

	Group A	Group B	Group C	Group D
	Data Set-A	Data Set-B	Data Set-C	Title
	Y	Y	Y	Y
1				
2				

Assume Gaussian distribution of residuals?

● Yes. Use ANOVA.
○ No. Use nonparametric test.

Assume equal SDs?

● Yes. Use ordinary ANOVA test.
○ No. Use Brown-Forsythe and Welch ANOVA te

Based on your choices (on all tabs), Prism will perf
- Ordinary one-way ANOVA.

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Followup tests

● None.
○ Compare the mean of each column with the mean of every other column.
○ Compare the mean of each column with the mean of a control column.
Control column: Column A : A
○ Compare the means of preselected pairs of columns.
Selected pairs: Select...
○ Test for linear trend between column mean and left-to-right column order.

Which test?

Use choices on the Options tab to choose the test, and to set the defaults for future ANOVAs.

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Multiple comparisons test

● Correct for multiple comparisons using statistical hypothesis testing. Recommended.
Test: Tukey (recommended)
○ Correct for multiple comparisons by controlling the False Discovery Rate.
Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommend
○ Don't correct for multiple comparisons. Each comparison stands alone.
Test: Fisher's LSD test

Multiple comparisons options

☐ Swap direction of comparisons (A-B) vs. (B-A).
☑ Report multiplicity adjusted P value for each comparison.
Each P value is adjusted to account for multiple comparisons.
Family-wise significance and confidence level: 0.05 (95% confidence interval)

Graphing

☐ Graph confidence intervals.
☐ Graph ranks (nonparametric).
☐ Graph differences (repeated measures).

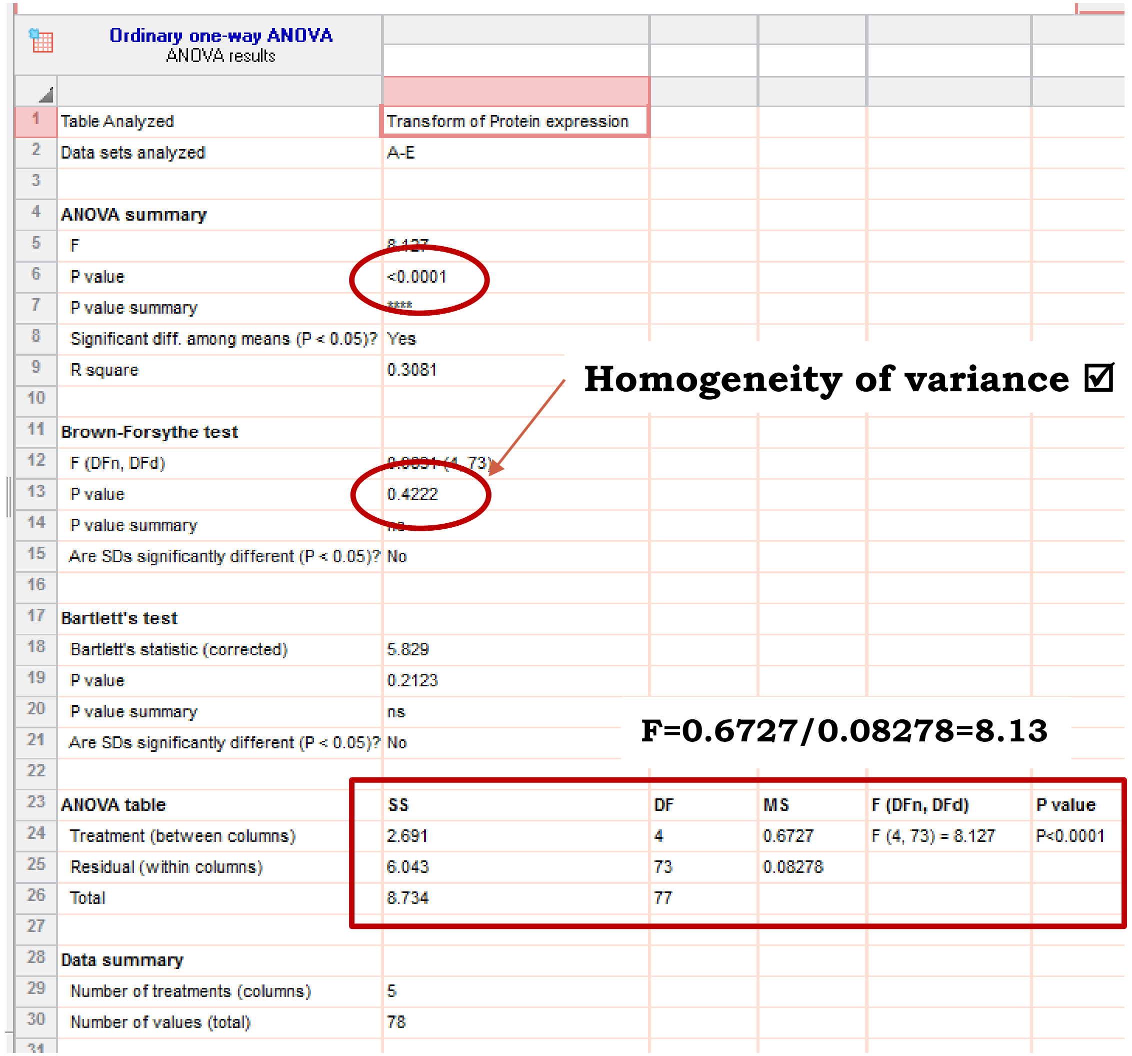
Additional results

☐ Descriptive statistics for each data set.
☐ Report comparison of models using AICc.
☐ Report goodness of fit.

Output

Show this many significant digits (for everything except P values): 4
P value style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.(N = 6
☐ Make options on this tab be the default for future One-Way ANOVAs.

Analysis of variance: results

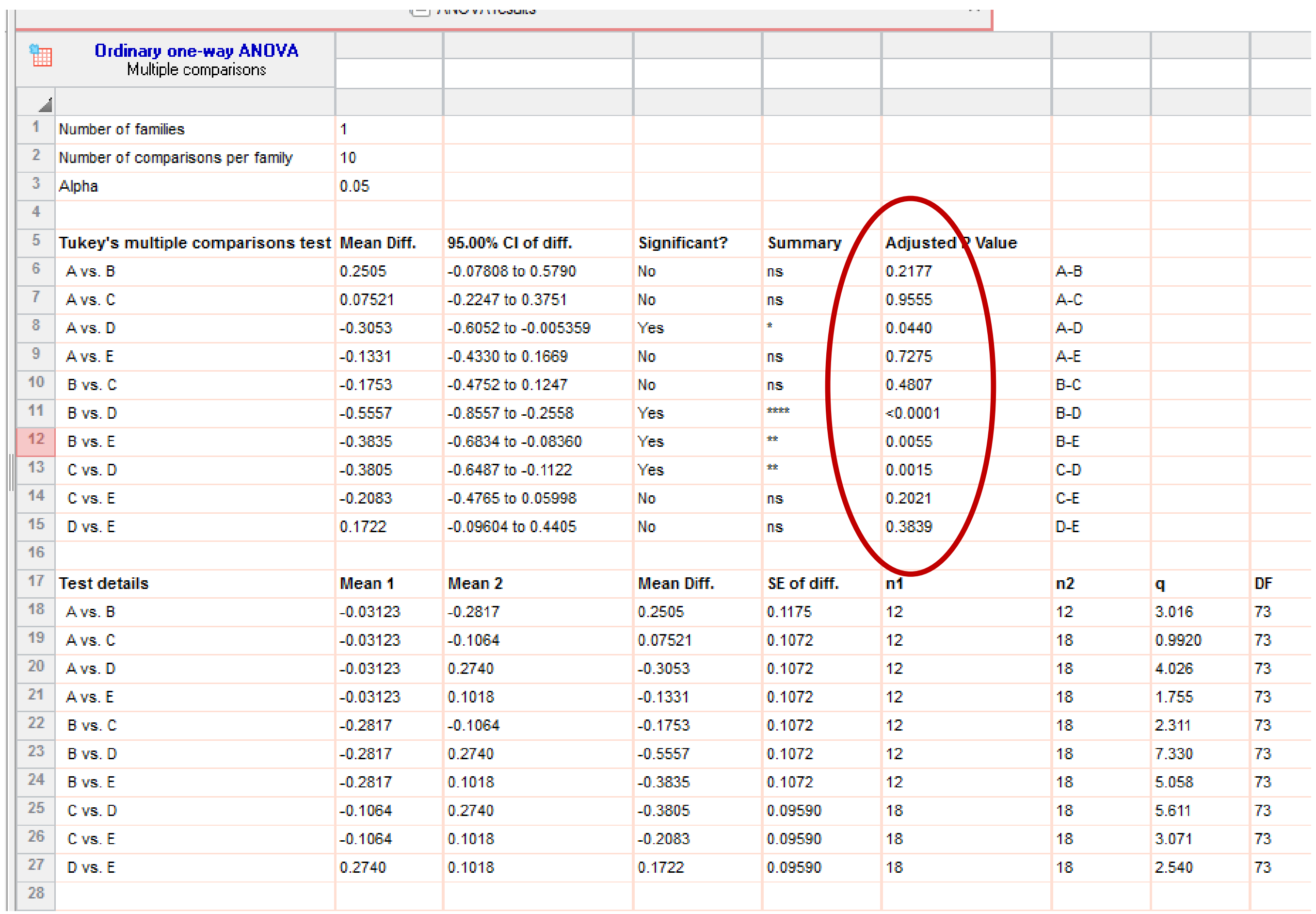


Ordinary one-way ANOVA
ANOVA results

		A				
1	Table Analyzed	Transform of Protein expression				
2	Data sets analyzed	A-E				
3						
4	**ANOVA summary**					
5	F	8.127				
6	P value	<0.0001				
7	P value summary	****				
8	Significant diff. among means (P < 0.05)?	Yes				
9	R square	0.3081				
10						
11	**Brown-Forsythe test**					
12	F (DFn, DFd)	0.9891 (4, 73)				
13	P value	0.4222				
14	P value summary	ns				
15	Are SDs significantly different (P < 0.05)?	No				
16						
17	**Bartlett's test**					
18	Bartlett's statistic (corrected)	5.829				
19	P value	0.2123				
20	P value summary	ns				
21	Are SDs significantly different (P < 0.05)?	No				
22						
23	**ANOVA table**	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
24	Treatment (between columns)	2.691	4	0.6727	F (4, 73) = 8.127	P<0.0001
25	Residual (within columns)	6.043	73	0.08278		
26	Total	8.734	77			
27						
28	**Data summary**					
29	Number of treatments (columns)	5				
30	Number of values (total)	78				

Homogeneity of variance ☑

F=0.6727/0.08278=8.13

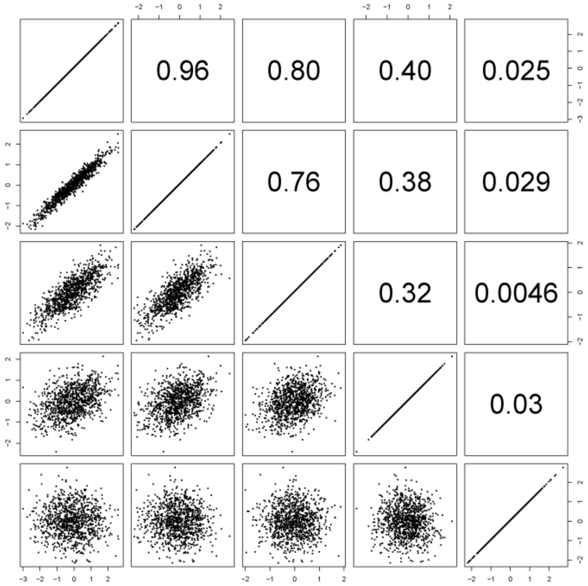


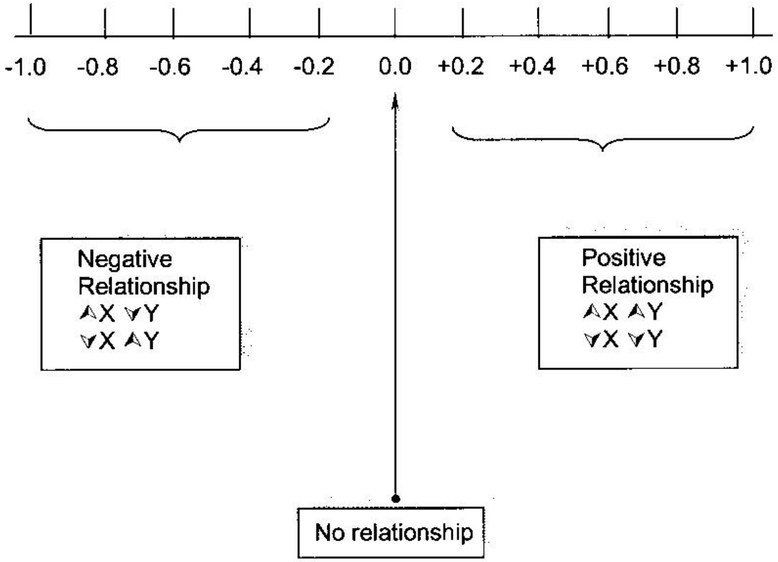
Ordinary one-way ANOVA
Multiple comparisons

1	Number of families	1							
2	Number of comparisons per family	10							
3	Alpha	0.05							
4									
5	**Tukey's multiple comparisons test**	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted P Value**			
6	A vs. B	0.2505	-0.07808 to 0.5790	No	ns	0.2177	A-B		
7	A vs. C	0.07521	-0.2247 to 0.3751	No	ns	0.9555	A-C		
8	A vs. D	-0.3053	-0.6052 to -0.005359	Yes	*	0.0440	A-D		
9	A vs. E	-0.1331	-0.4330 to 0.1669	No	ns	0.7275	A-E		
10	B vs. C	-0.1753	-0.4752 to 0.1247	No	ns	0.4807	B-C		
11	B vs. D	-0.5557	-0.8557 to -0.2558	Yes	****	<0.0001	B-D		
12	B vs. E	-0.3835	-0.6834 to -0.08360	Yes	**	0.0055	B-E		
13	C vs. D	-0.3805	-0.6487 to -0.1122	Yes	**	0.0015	C-D		
14	C vs. E	-0.2083	-0.4765 to 0.05998	No	ns	0.2021	C-E		
15	D vs. E	0.1722	-0.09604 to 0.4405	No	ns	0.3839	D-E		
16									
17	**Test details**	**Mean 1**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**n1**	**n2**	**q**	**DF**
18	A vs. B	-0.03123	-0.2817	0.2505	0.1175	12	12	3.016	73
19	A vs. C	-0.03123	-0.1064	0.07521	0.1072	12	18	0.9920	73
20	A vs. D	-0.03123	0.2740	-0.3053	0.1072	12	18	4.026	73
21	A vs. E	-0.03123	0.1018	-0.1331	0.1072	12	18	1.755	73
22	B vs. C	-0.2817	-0.1064	-0.1753	0.1072	12	18	2.311	73
23	B vs. D	-0.2817	0.2740	-0.5557	0.1072	12	18	7.330	73
24	B vs. E	-0.2817	0.1018	-0.3835	0.1072	12	18	5.058	73
25	C vs. D	-0.1064	0.2740	-0.3805	0.09590	18	18	5.611	73
26	C vs. E	-0.1064	0.1018	-0.2083	0.09590	18	18	3.071	73
27	D vs. E	0.2740	0.1018	0.1722	0.09590	18	18	2.540	73
28									

Correlation

- A correlation coefficient is an index number that measures:
  - The <u>magnitude</u> and the <u>direction</u> of the relation between 2 variables
  - It is designed to range in value between -1 and +1





Correlation

- Most widely-used correlation coefficient:
  - Pearson product-moment correlation coefficient “r”

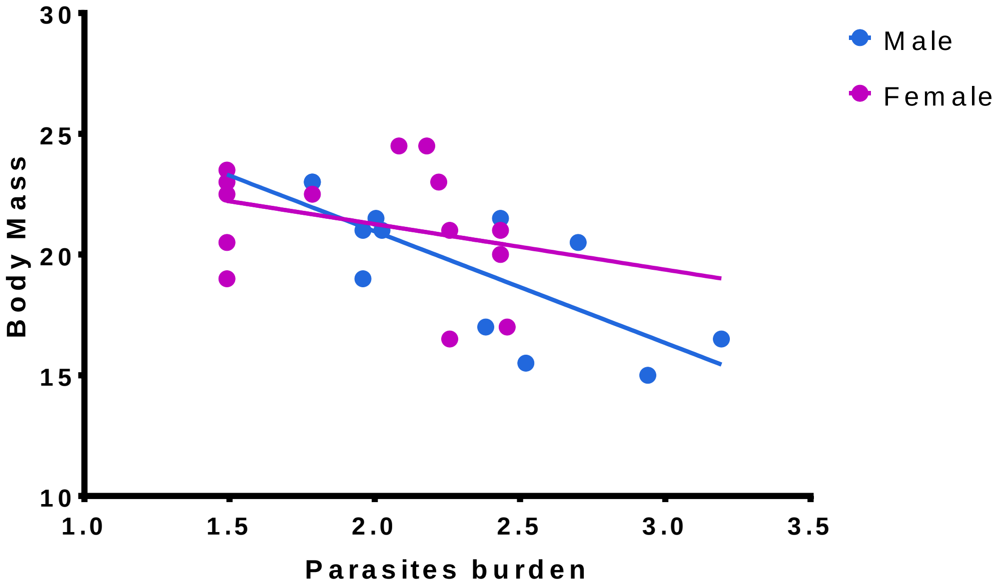
$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

    - The 2 variables do not have to be measured in the same units but they have to be proportional (meaning linearly related)
  - Coefficient of determination:
    - r is the correlation between X and Y
    - r^2 is the coefficient of determination:
      - It gives you the proportion of variance in Y that can be explained by X, in percentage.

Correlation

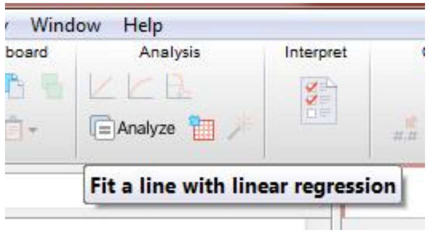
Example: roe deer.xlsx

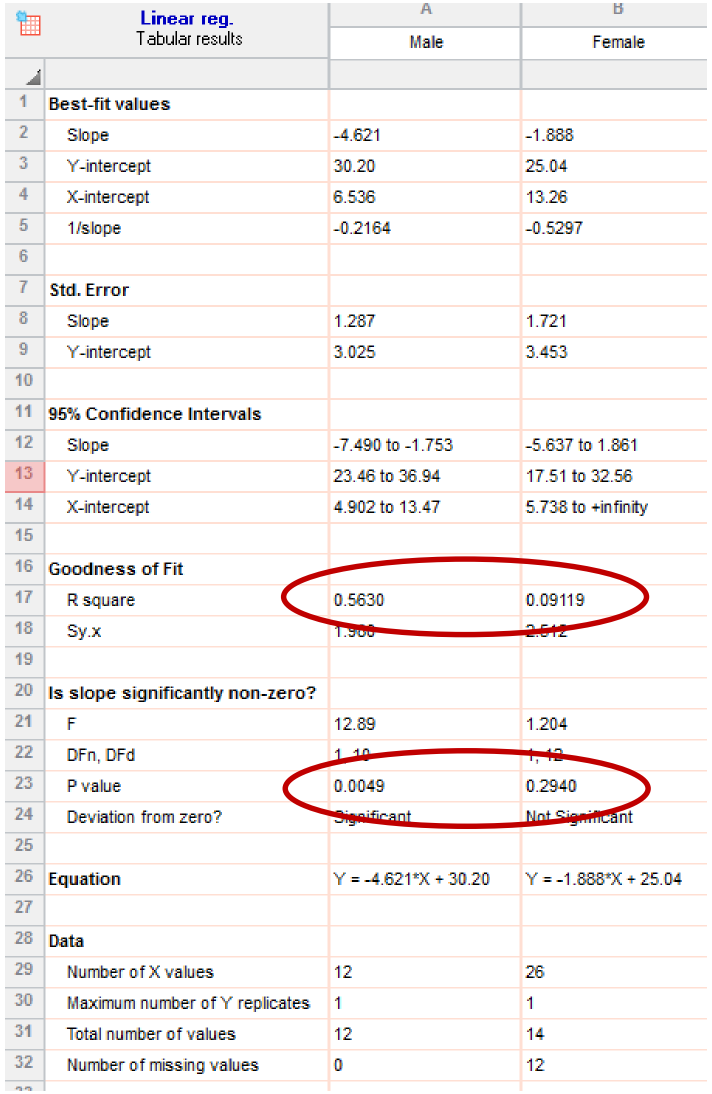
- Is there a relationship between parasite burden and body mass in roe deer?



Correlation

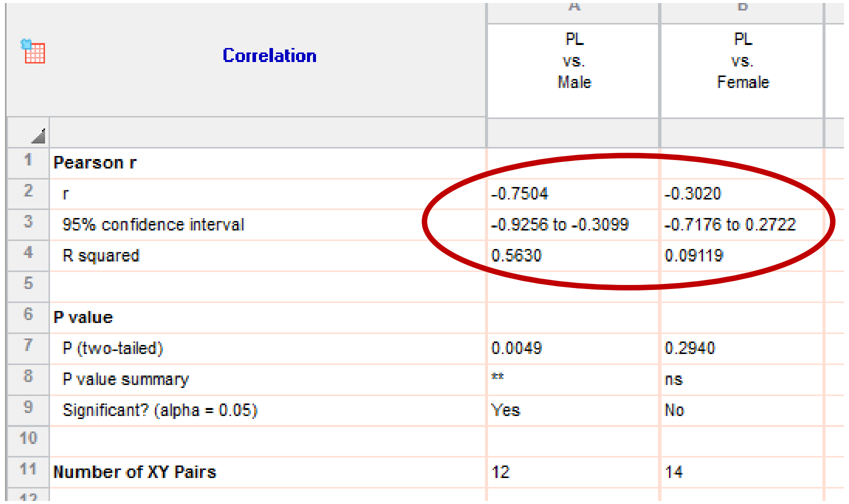
Example: roe deer.xlsx





	Linear reg. Tabular results	A Male	B Female
1	**Best-fit values**		
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	**Std. Error**		
8	Slope	1.287	1.721
9	Y-intercept	3.025	3.453
10			
11	**95% Confidence Intervals**		
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	**Goodness of Fit**		
17	R square	0.5630	0.09119
18	Sy.x	1.980	2.512
19			
20	**Is slope significantly non-zero?**		
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	**Equation**	Y = -4.621*X + 30.20	Y = -1.888*X + 25.04
27			
28	**Data**		
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

There is a negative correlation between parasite load and fitness but this relationship is only significant for the males(p=0.0049 vs. females: p=0.2940).

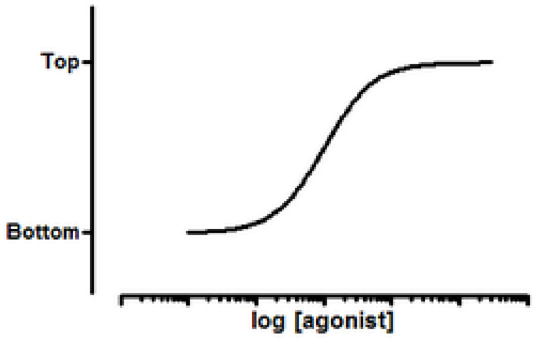


	Correlation	A PL vs. Male	B PL vs. Female
1	**Pearson r**		
2	r	-0.7504	-0.3020
3	95% confidence interval	-0.9256 to -0.3099	-0.7176 to 0.2722
4	R squared	0.5630	0.09119
5			
6	**P value**		
7	P (two-tailed)	0.0049	0.2940
8	P value summary	**	ns
9	Significant? (alpha = 0.05)	Yes	No
10			
11	**Number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

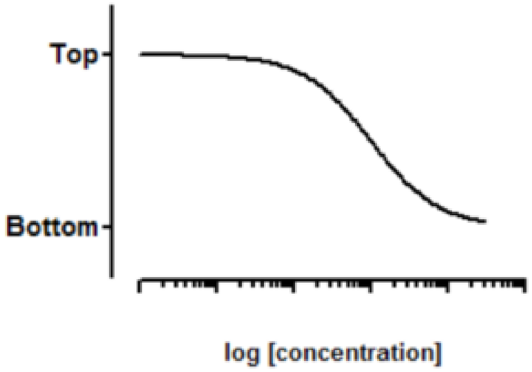
Curve fitting

- **Dose-response curves**
  - Nonlinear regression
  - Dose-response experiments typically use around 5-10 doses of agonist, equally spaced on a logarithmic scale
  - Y values are responses

- The aim is often to determine the **IC50** or the **EC50**
  - **IC50 (I=Inhibition)**: concentration of an agonist that provokes a response half way between the maximal (Top) response and the maximally inhibited (Bottom) response.
  - **EC50 (E=Effective):** concentration that gives half-maximal response



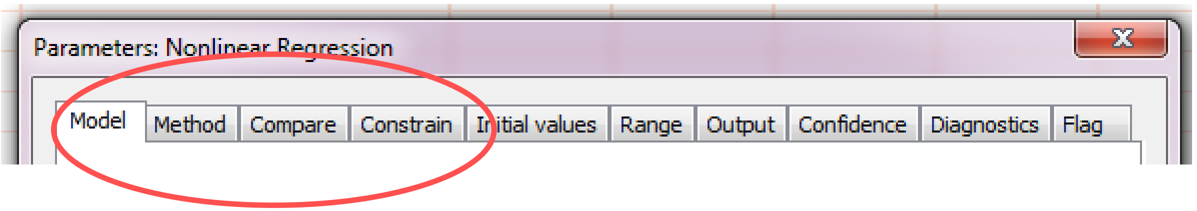
Stimulation:
Y=Bottom + (Top-Bottom)/(1+10^((LogEC50-X)*HillSlope))

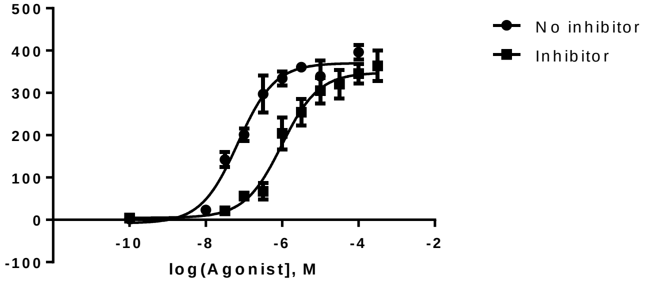


Inhibition:
Y=Bottom + (Top-Bottom)/(1+10^((X-LogIC50)))

Curve fitting

Example: Inhibition data.xlsx





Step by step analysis and considerations:

1- Choose a **Model**:

not necessary to normalise

should choose it when values defining 0 and 100 are precise

variable slope better if plenty of data points (variable slope or 4 parameters)

2- Choose a **Method:** outliers, fitting method, weighting method and replicates

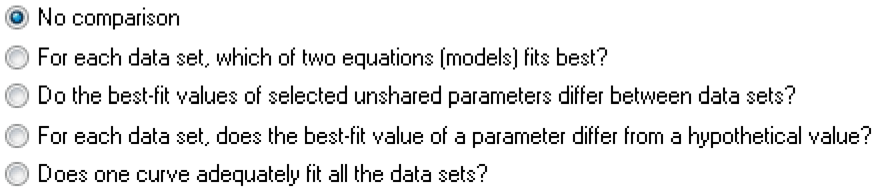
3- **Compare** different conditions:

Diff in parameters

Diff between conditions for one or more parameters →

Constraint vs no constraint

Diff between conditions for one or more parameters →



● No comparison
○ For each data set, which of two equations (models) fits best?
○ Do the best-fit values of selected unshared parameters differ between data sets?
○ For each data set, does the best-fit value of a parameter differ from a hypothetical value?
○ Does one curve adequately fit all the data sets?

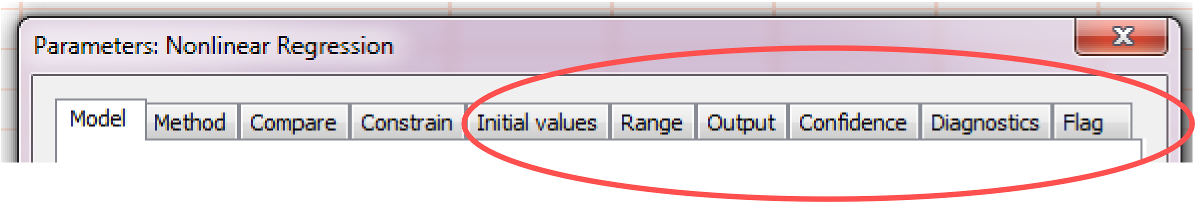
4- **Constrain**:

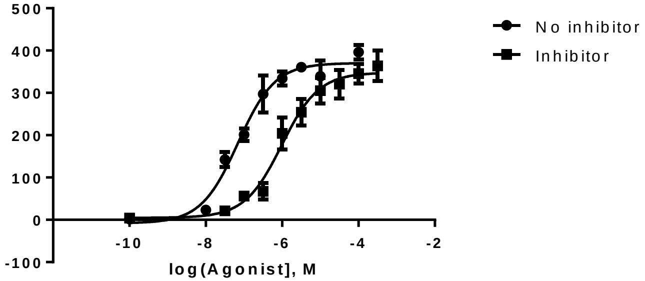
depends on your experiment

depends if your data don't define the top or the bottom of the curve

Curve fitting

Example: Inhibition data.xlsx





Step by step analysis and considerations:

5- **Initial values**:

defaults usually OK unless the fit looks funny

6- **Range**:

defaults usually OK unless you are not interested in the x-variable full range (ie time)

7- **Output**:

summary table presents same results in a … summarized way.

8 – **Confidence**: calculate and plot confidence intervals

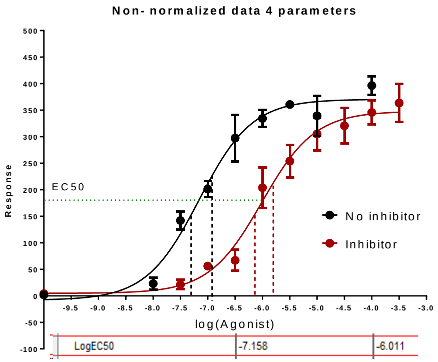
9- **Diagnostics**:

check for normality (weights) and outliers (but keep them in the analysis)
check Replicates test
residual plots

Curve fitting

Example: Inhibition data.xlsx

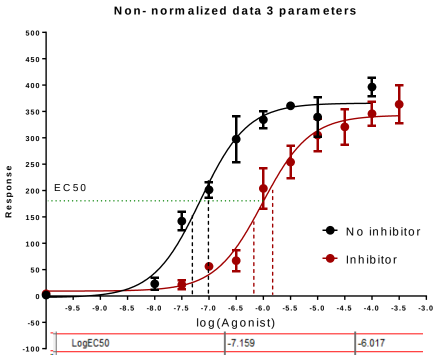
LogEC50 same for all data sets
LogEC50 different for each data set
< 0.0001
Reject null hypothesis
LogEC50 different for each data set
64.86 (1,48)



LogEC50	-7.158	-6.011

95% Confidence Intervals		
Bottom	-41.39 to 24.94	-22.15 to 31.56
Top	348.3 to 392.6	323.1 to 373.0
LogEC50	-7.324 to -6.991	-6.185 to -5.837
HillSlope	0.6347 to 1.159	0.6095 to 1.186
EC50	4.739e-008 to 1.020e-007	6.538e-007 to 1.455e-006

R square	0.9663	0.9653



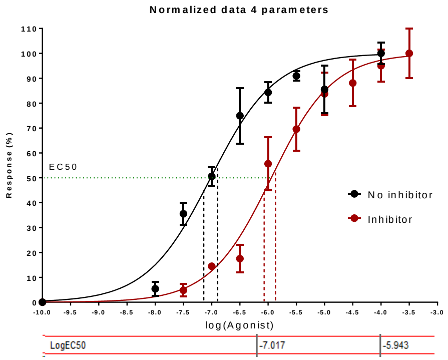
LogEC50	-7.159	-6.017

LogEC50 same for all data sets
LogEC50 different for each data set
< 0.0001
Reject null hypothesis
LogEC50 different for each data set
101.0 (1,50)

95% Confidence Intervals		
Bottom	-30.74 to 24.78	-11.65 to 30.07
Top	348.2 to 383.2	324.3 to 361.4
LogEC50	-7.312 to -7.006	-6.175 to -5.859
EC50	4.875e-008 to 9.858e-008	6.677e-007 to 1.385e-006

R square	0.9655	0.9648

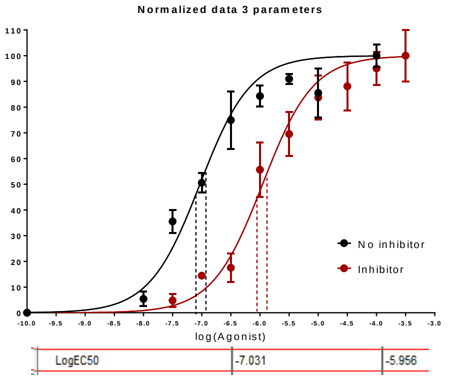
LogEC50 same for all data sets
LogEC50 different for each data set
< 0.0001
Reject null hypothesis
LogEC50 different for each data set
162.8 (1,52)



LogEC50	-7.017	-5.943

95% Confidence Intervals		
LogEC50	-7.137 to -6.897	-6.057 to -5.830
HillSlope	0.6094 to 0.9184	0.6467 to 0.9460
EC50	7.295e-008 to 1.268e-007	8.763e-007 to 1.481e-006

R square	0.9580	0.9635



LogEC50	-7.031	-5.956

One curve for all data sets
Different curve for each data set
< 0.0001
Reject null hypothesis
Different curve for each data set
175.0 (1,54)

95% Confidence Intervals		
LogEC50	-7.144 to -6.917	-6.064 to -5.848
EC50	7.179e-008 to 1.209e-007	8.633e-007 to 1.420e-006

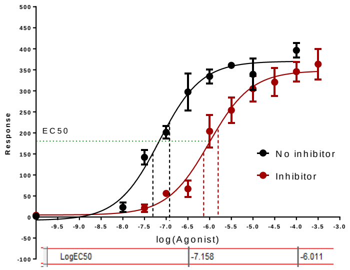
R square	0.9476	0.9568

Curve fitting

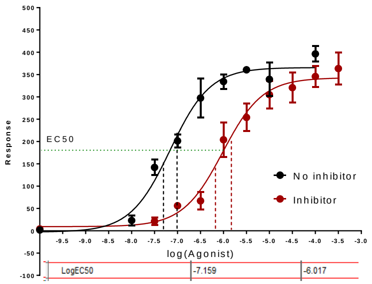
Example: Inhibition data.xlsx

	No inhibitor	Inhibitor
Replicates test for lack of fit		
SD replicates	22.71	25.52
SD lack of fit	41.84	32.38
Discrepancy (F)	3.393	1.610
P value	0.0247	0.1989
Evidence of inadequate model?	Yes	No
Replicates test for lack of fit		
SD replicates	22.71	25.52
SD lack of fit	39.22	30.61
Discrepancy (F)	2.982	1.438
P value	0.0334	0.2478
Evidence of inadequate model?	Yes	No
Replicates test for lack of fit		
SD replicates	5.755	7.100
SD lack of fit	11.00	8.379
Discrepancy (F)	3.656	1.393
P value	0.0125	0.2618
Evidence of inadequate model?	Yes	No
Replicates test for lack of fit		
SD replicates	5.755	7.100
SD lack of fit	12.28	9.649
Discrepancy (F)	4.553	1.847
P value	0.0036	0.1246
Evidence of inadequate model?	Yes	No

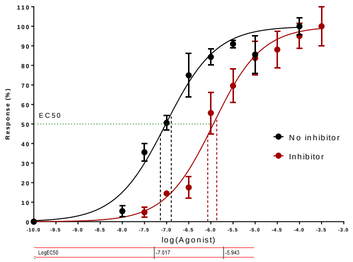
Non- normalized data 4 parameters



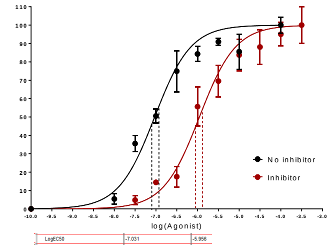
Non- normalized data 3 parameters



Normalized data 4 parameters



Normalized data 3 parameters



No inhibitor	Inhibitor
-7.158	-6.011
-7.159	-6.017
-7.017	-5.943
-7.031	-5.956

My email address if you need some help with GraphPad:

anne.segonds-pichon@babraham.ac.uk