Data and variables

DATA: the answers to questions or measurements from the experiment

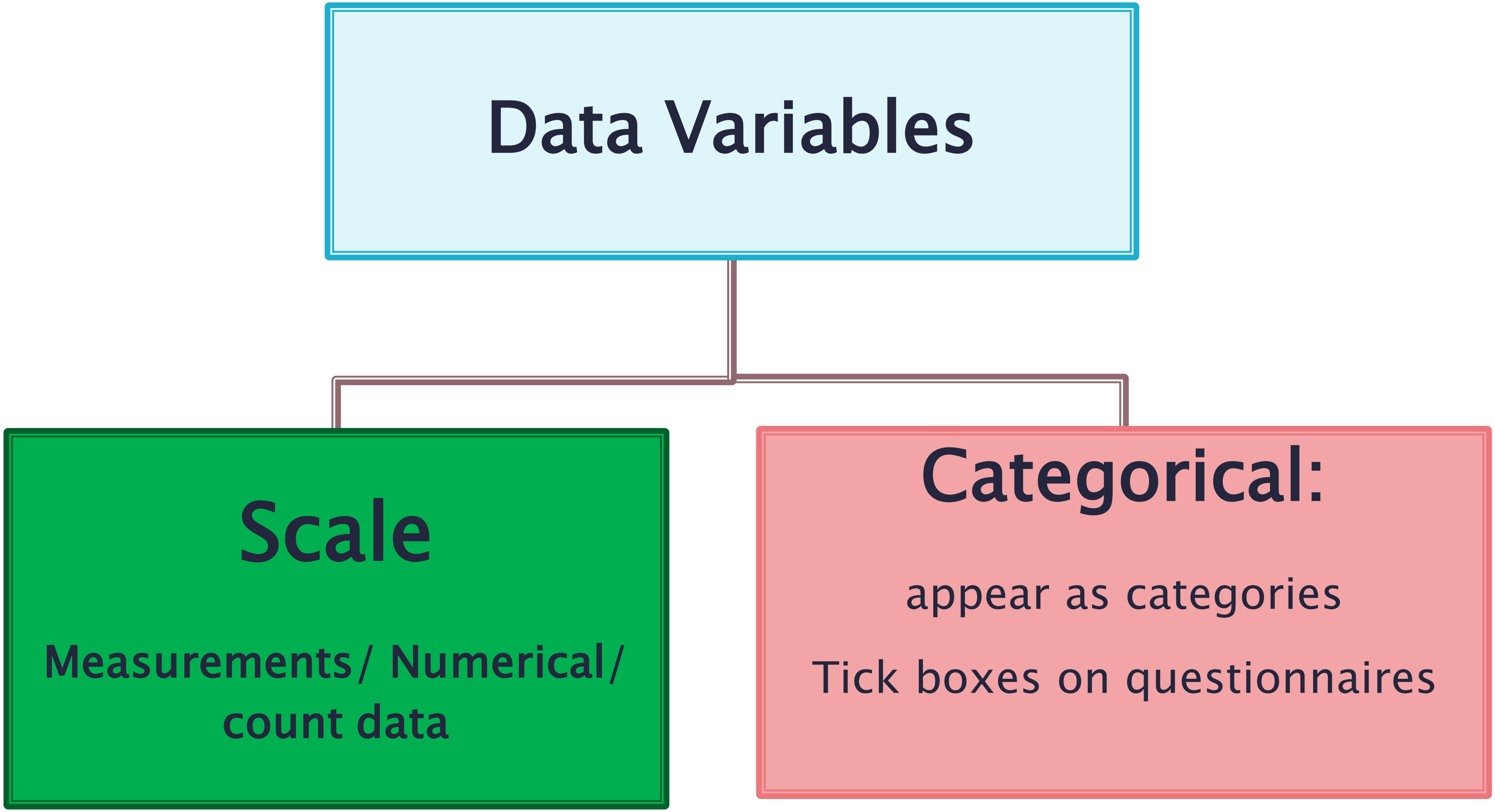
VARIABLE = measurement which varies between subjects e.g. height or gender

One variable per column

	A	B	C	D
1	Subject ID	Gender	Year of study	Height
2	1	Male	1	170
3	2	Female	2	160
4	3	Female	3	165
5	4	Male	PG	175
6	5	Female	3	168

One row per subject

Data types


Data Variables
Scale
Measurements/ Numerical/ count data
Categorical:
appear as categories
Tick boxes on questionnaires

Data types

- Variables
  - Scale
    - Continuous
      Measurements
      takes any value
    - Discrete:
      Counts/ integers
  - Categorical
    - Ordinal:
      obvious order
    - Nominal:
      no meaningful order

Questionnaire for GCSE Maths Pupils

What data types relate to following questions?

- Q1: What is your favourite subject?

Maths	English	Science	Art	French

- Q2: Gender:

Male	Female

- Q3: I consider myself to be good at mathematics:

Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree

- Q4: Score in a recent mock GCSE maths exam:

Score between 0% and 100%

Questionnaire for GCSE Maths Pupils

What data types relate to following questions?

- Q1: What is your favourite subject? — **Nominal**

Maths	English	Science	Art	French

- Q2: Gender: — **Binary/ Nominal**

Male	Female

- Q3: I consider myself to be good at mathematics: — **Ordinal**

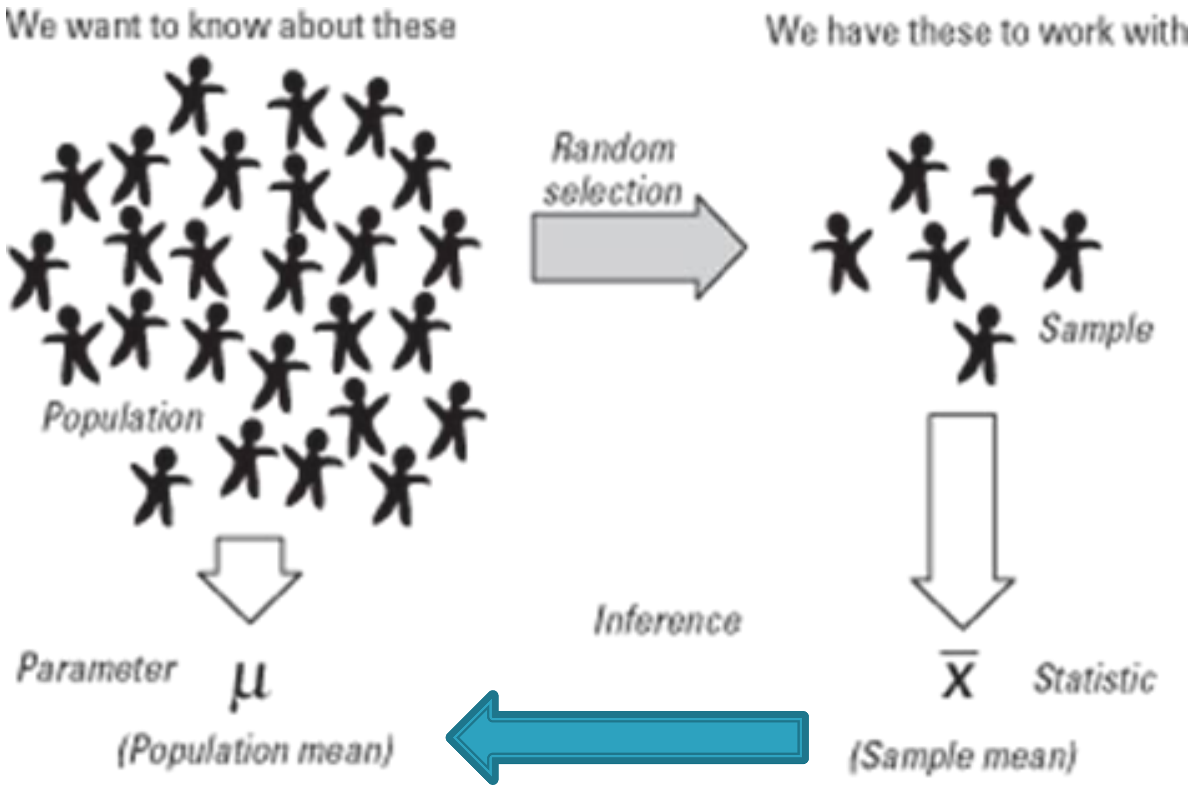
Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree

- Q4: Score in a recent mock GCSE maths exam: — **Scale**

Score between 0% and 100%

Populations and samples

- Taking a sample from a population



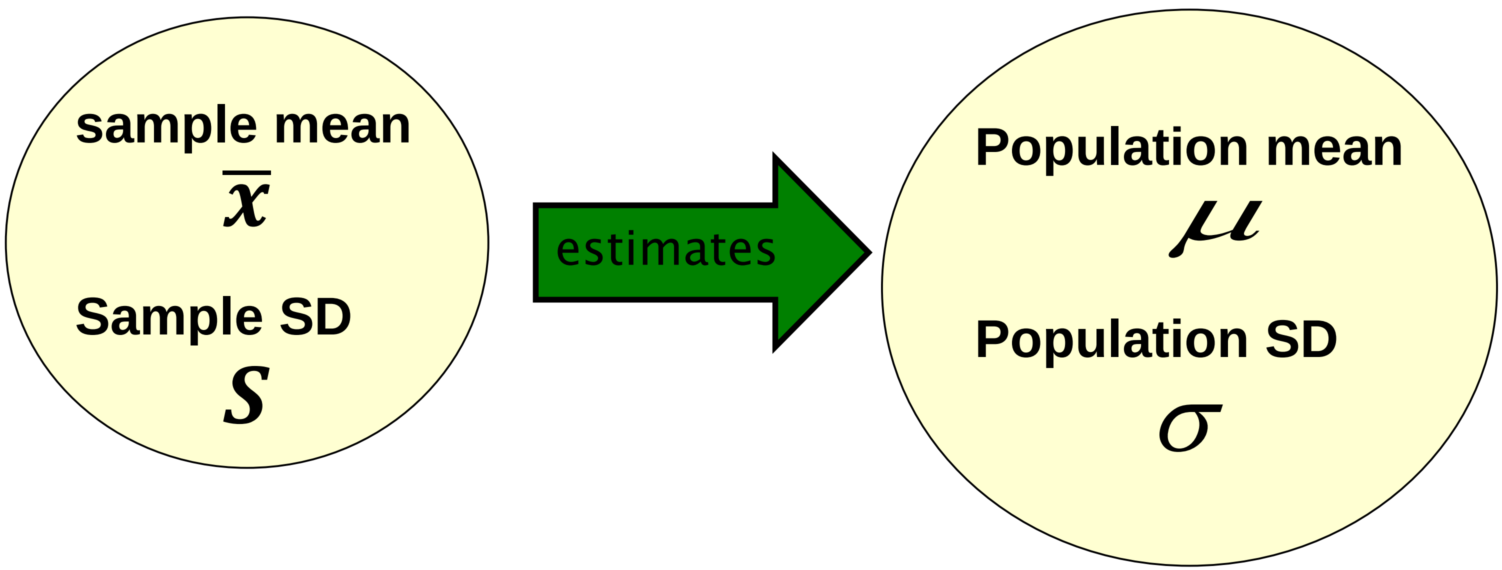
Sample data ‘represents’ the whole population

Point estimation

Sample data is used to estimate parameters of a population

Statistics are calculated using sample data.

Parameters are the characteristics of population data



How can exam score data be summarised?

Exam marks for 60 students (marked out of 65)

48	37	1	33	26	22	15	22	40	30	12	36
21	20	29	13	44	52	28	39	16	48	56	27
47	12	35	24	10	36	18	34	9	25	31	42
31	27	64	25	58	17	26	38	28	43	33	5
25	55	7	32	39	23	49	43	11	37	22	54

Summary statistics

- Mean = $$\frac{\sum_{i=1}^{n} x}{n} = \bar{x}$$

Standard deviation (s) is a measure of how much the individuals differ from the mean

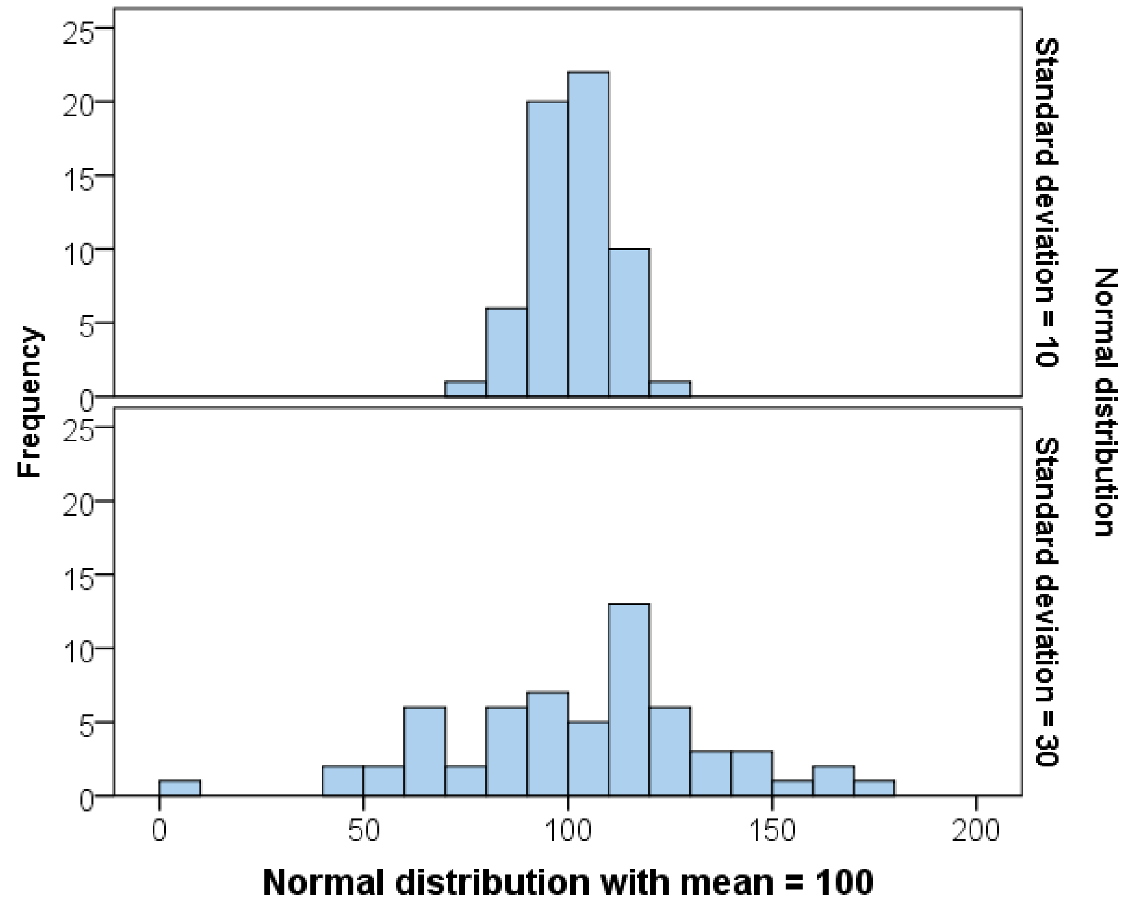
$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Large SD = very spread out data
Small SD = there is little variation from the mean

For exam scores, mean = 30.5, SD = 14.46

Interpretation of standard deviation

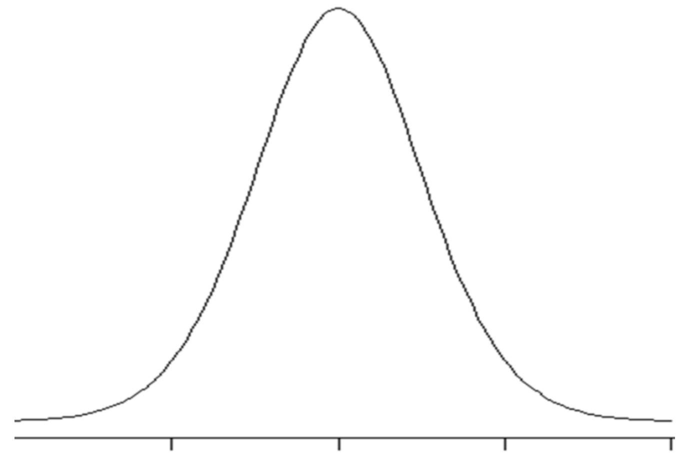
- The larger the standard deviation, the more spread out the data is.



Scale data

If have scale data

To analyse it we often assume it follows a Normal distribution



Normal

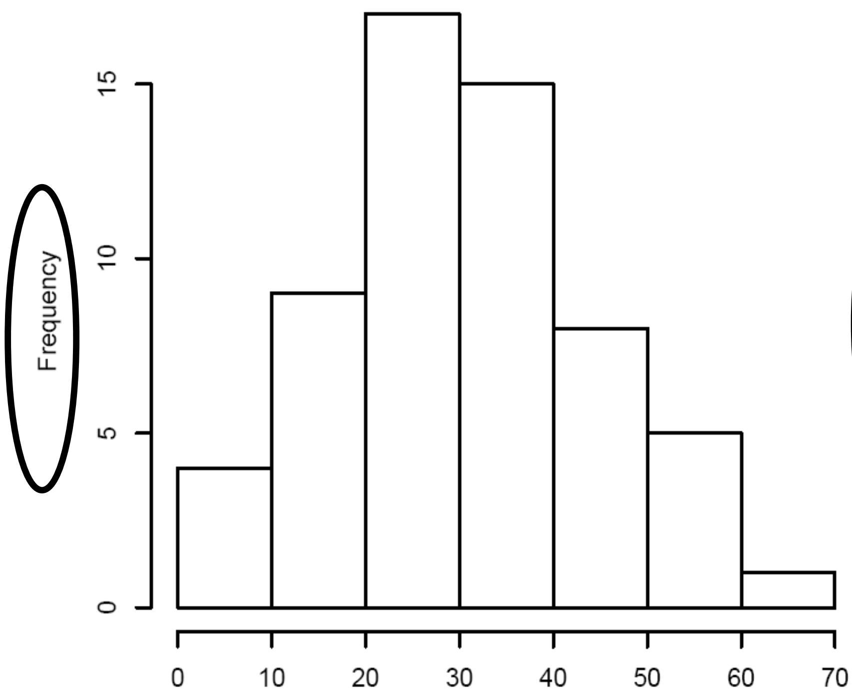
Discussion

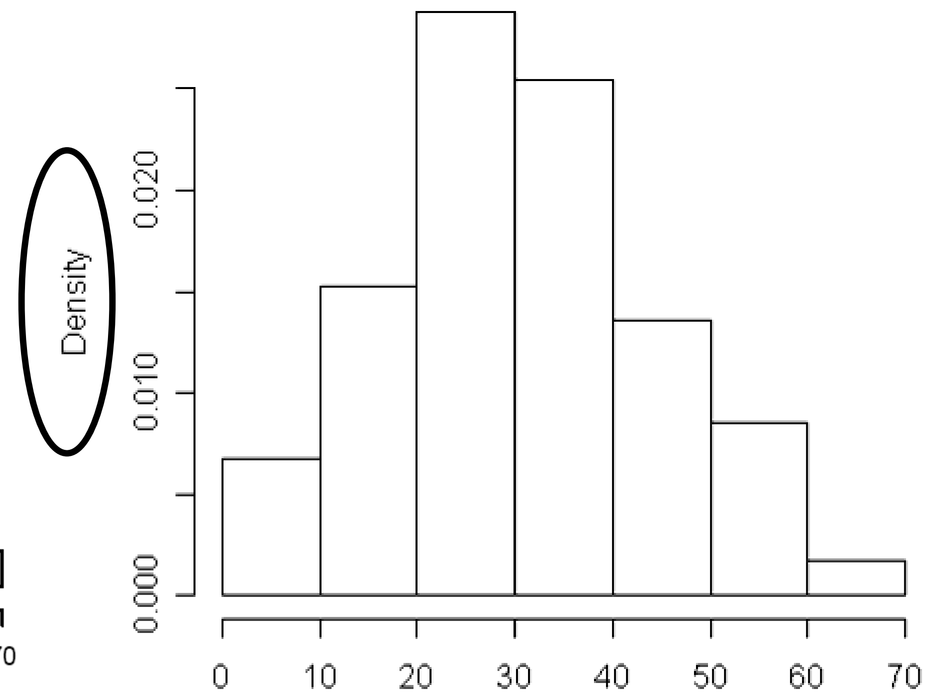
- How could you explain to a student what we mean by data being assumed to follow a Normal Distribution?

Group Frequency Table

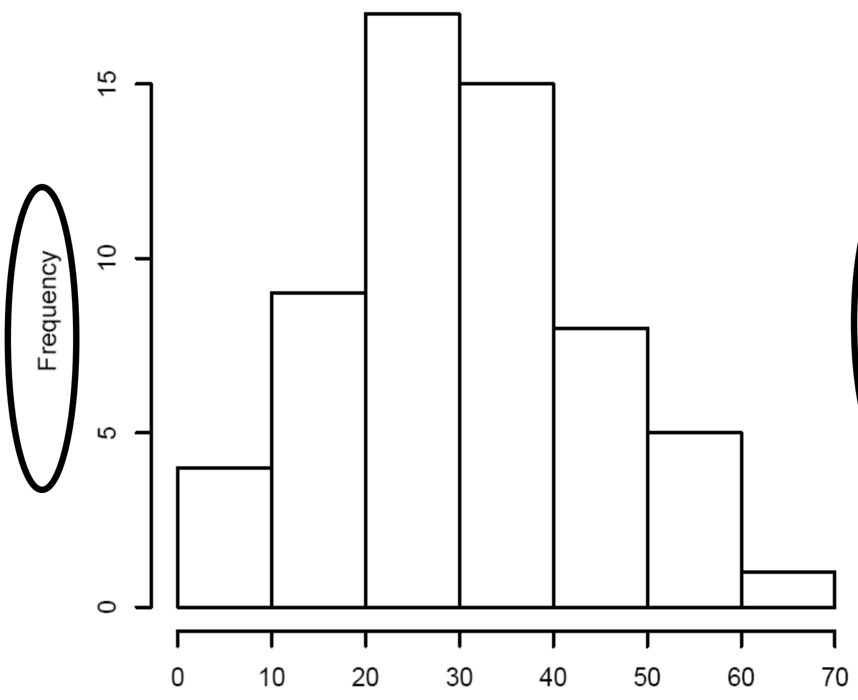
	Frequency	**Percent**
0 but less than 10	4	6.7
10 but less than 20	9	15.0
20 but less than 30	17	28.3
30 but less than 40	15	25.0
40 but less than 50	9	15.0
50 but less than 60	5	8.3
60 or over	1	1.7
Total	**60**	**100.0**

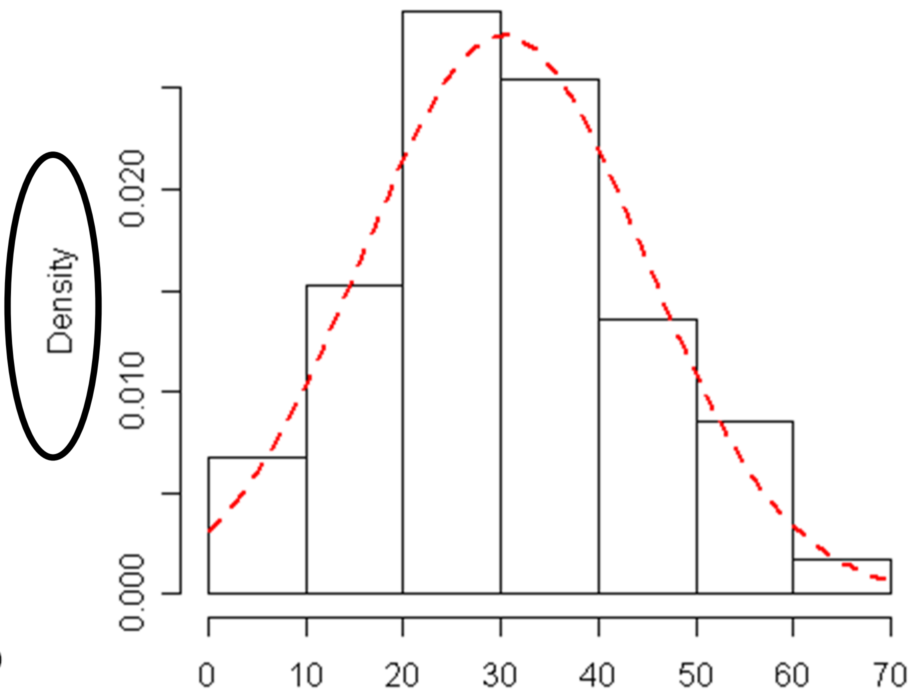
Histogram and Probability Distribution for Exam Marks Data



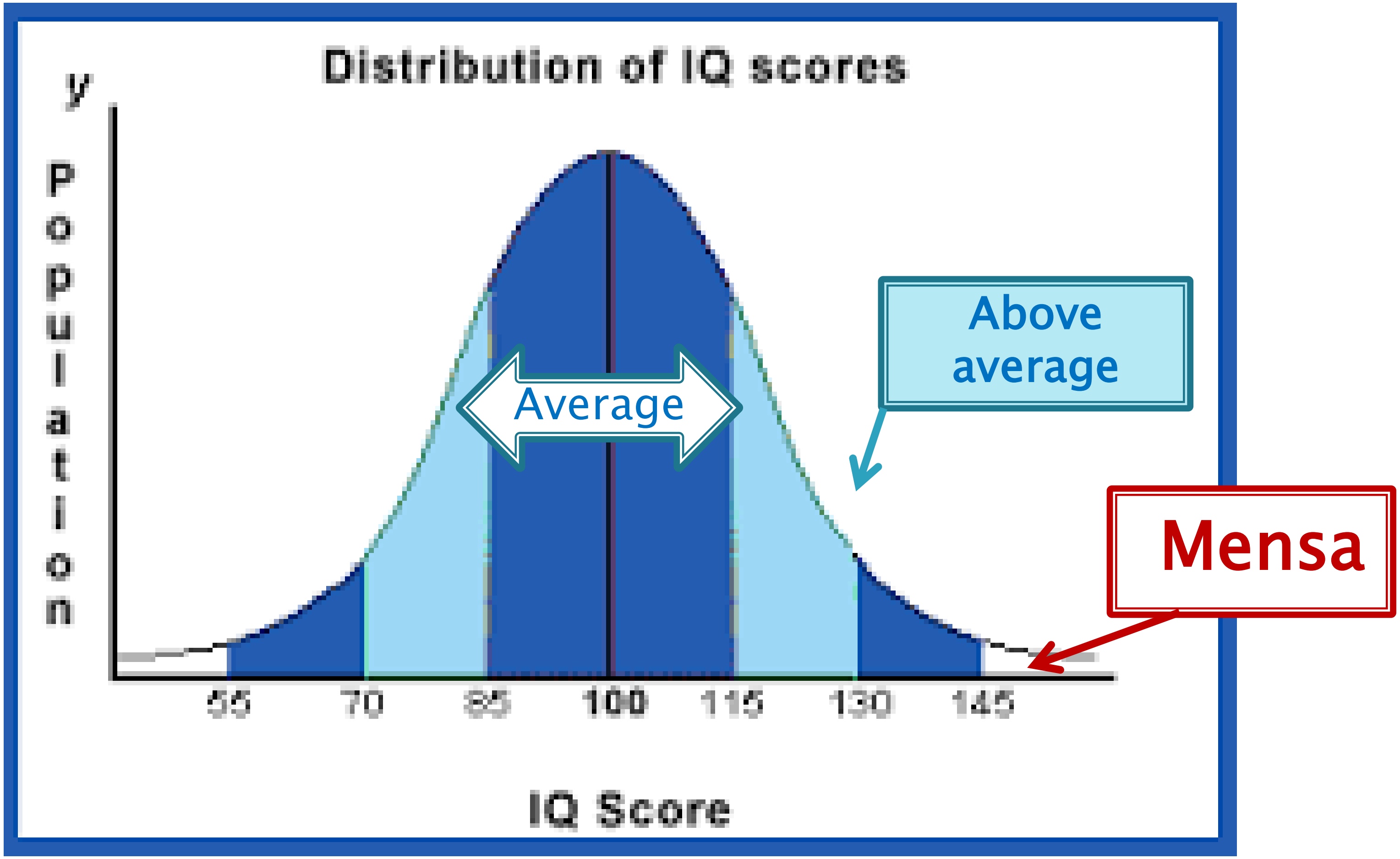


Histogram and Probability Distribution for Exam Marks Data



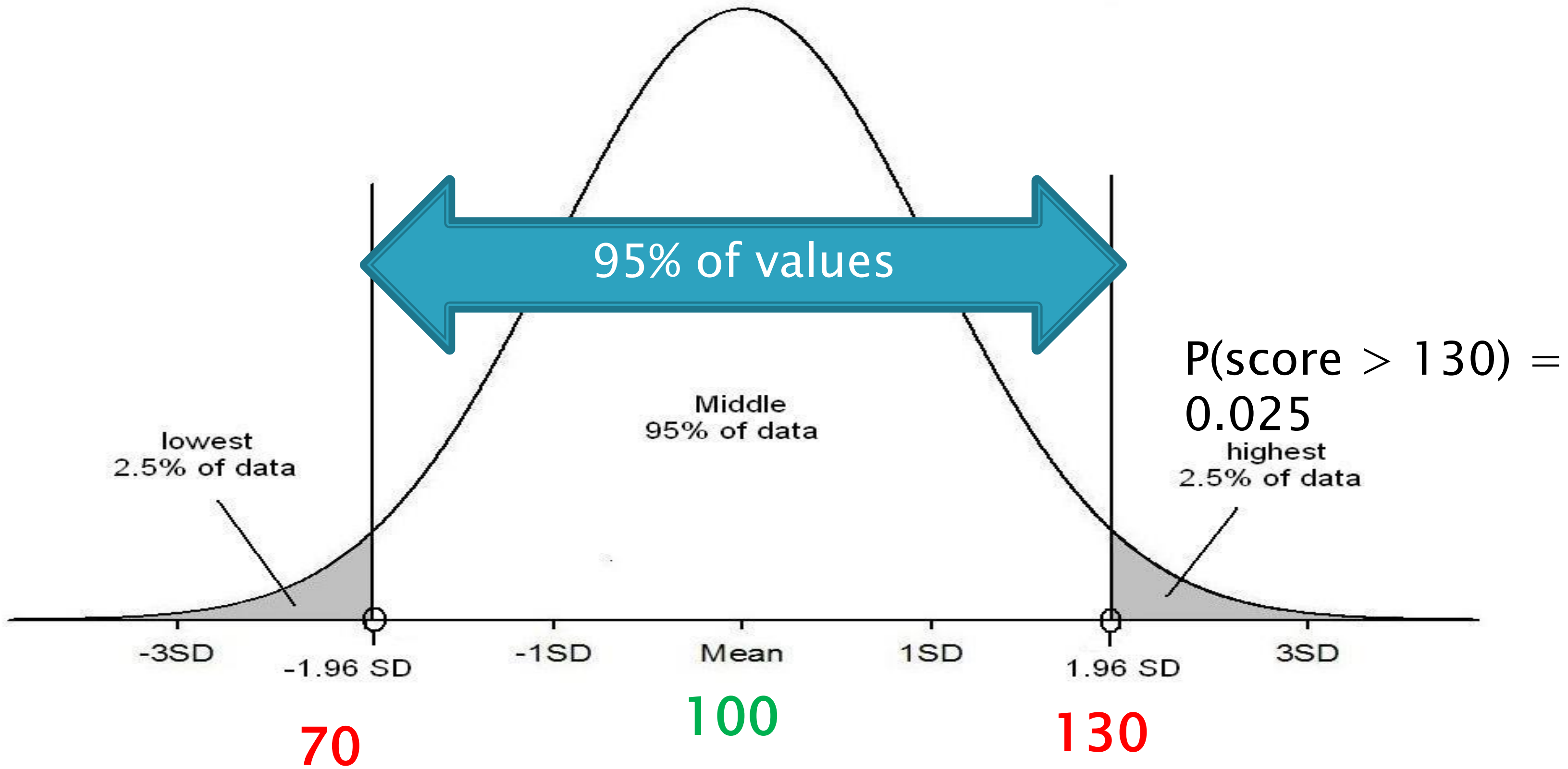


IQ is normally distributed



Mean = 100, SD = 15.3

95% 1.96 x SD's from the mean



$$mean - (1.96 \times SD)$$

$$100 - (1.96 \times 15.3) = 70$$

$$mean + (1.96 \times SD)$$

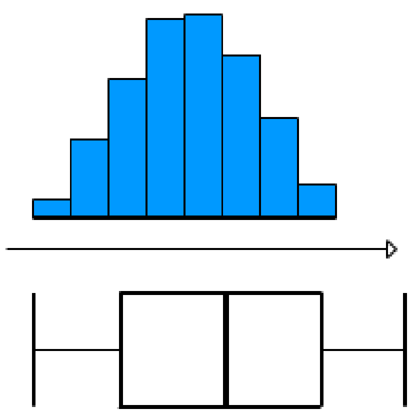
$$100 + (1.96 \times 15.3) = 130$$

95% of people have an IQ between 70 and 130

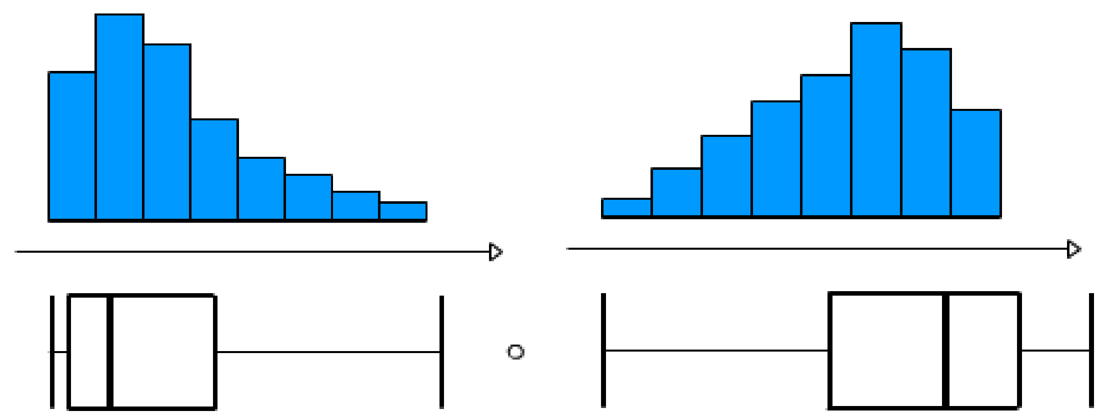
Assessing Normality

Charts can be used to **informally** assess whether data is:

Normally distributed



Or….Skewed



The mean and median are very different for skewed data.

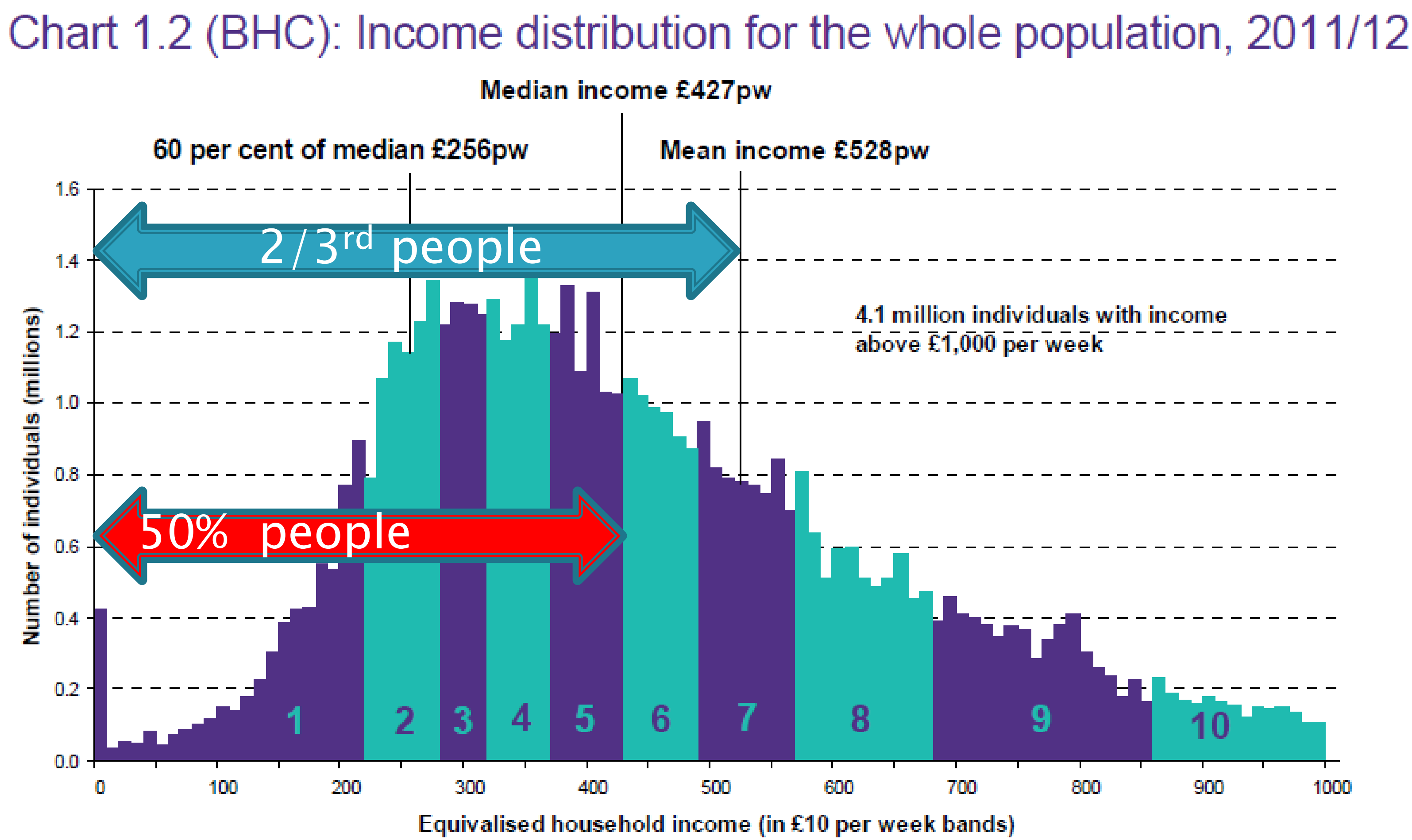
Discussion

Is the following statement:

“2 out of 3 people earn less than the average income”

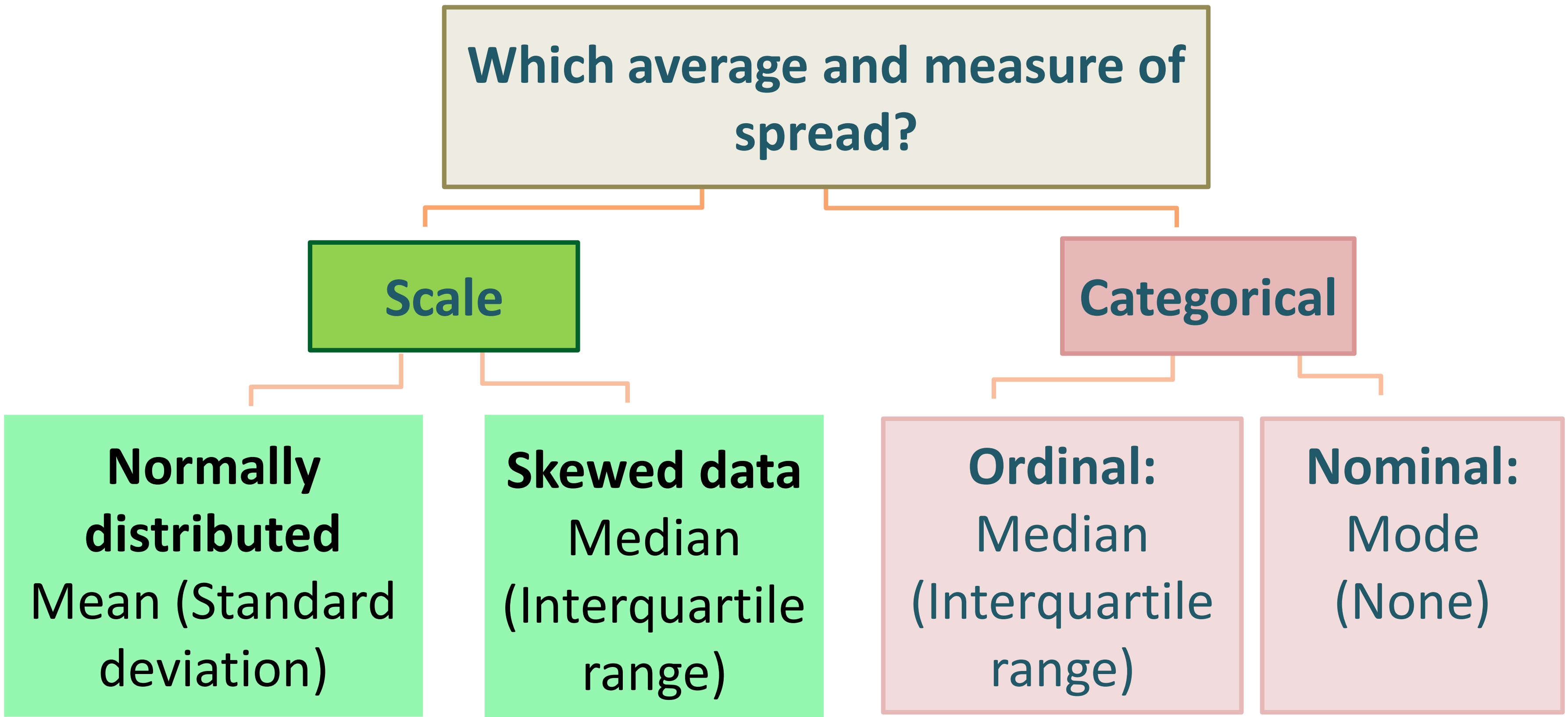
A. True
B. False
C. Don’t know

Sometimes the median makes more sense!



Source: Households Below Average Income: An analysis of the income distribution1994/95 – 2011/12, Department for Work and Pensions

Choosing summary statistics



Which graph? Exercise

1st variable	Only 1 variable	Scale	Categorical
Scale	Histogram	Scatter plot	Box-plot/ Confidence interval plot
Categorical	Pie/ Bar	Box-plot/ Confidence interval plot	Stacked/ multiple bar chart

Which graph would you use when investigating:

1) Whether daily temperature and ice cream sales were related?

2) Comparison of mean reaction time for a group having alcohol and a group drinking water

Exercise: Ticket cost comparison

Summary statistics for cost of Titanic ticket by survival

Cost of ticket	Survived?	
	Died	Survived
Mean	23.4	49.4
Median	10.5	26
Standard Deviation	34.2	68.7
Interquartile range	18.2	46.6
Minimum	0	0
Maximum	263	512.33

a) **Is there a big difference in average ticket price by group?**

b) **Which group has data which is more spread out?**

c) **Is the data skewed?**

d) **Is the mean or median a better summary measure?**

Which graph? Solution

1st variable	Only 1 variable	Scale	Categorical
Scale	Histogram	Scatter plot	Box-plot/ Confidence interval plot
Categorical	Pie/ Bar	Box-plot/ Confidence interval plot	Stacked/ multiple bar chart

Which graph would you use when investigating:

1) Whether daily temperature and ice cream sales were related? Scatter
2) Comparison of mean reaction time for a group having alcohol and a group drinking water Boxplot or confidence interval plot

Exercise: Ticket cost comparison Solution

a) Is there a big difference in average ticket price by group?

The mean and median are much bigger in those who survived.

b) Which group has data which is more spread out?

The standard deviation and interquartile range are much bigger for those who survived so that data is more spread out

c) Is the data skewed?

Yes. The medians are much smaller than the means and the plots show the data is positively skewed.

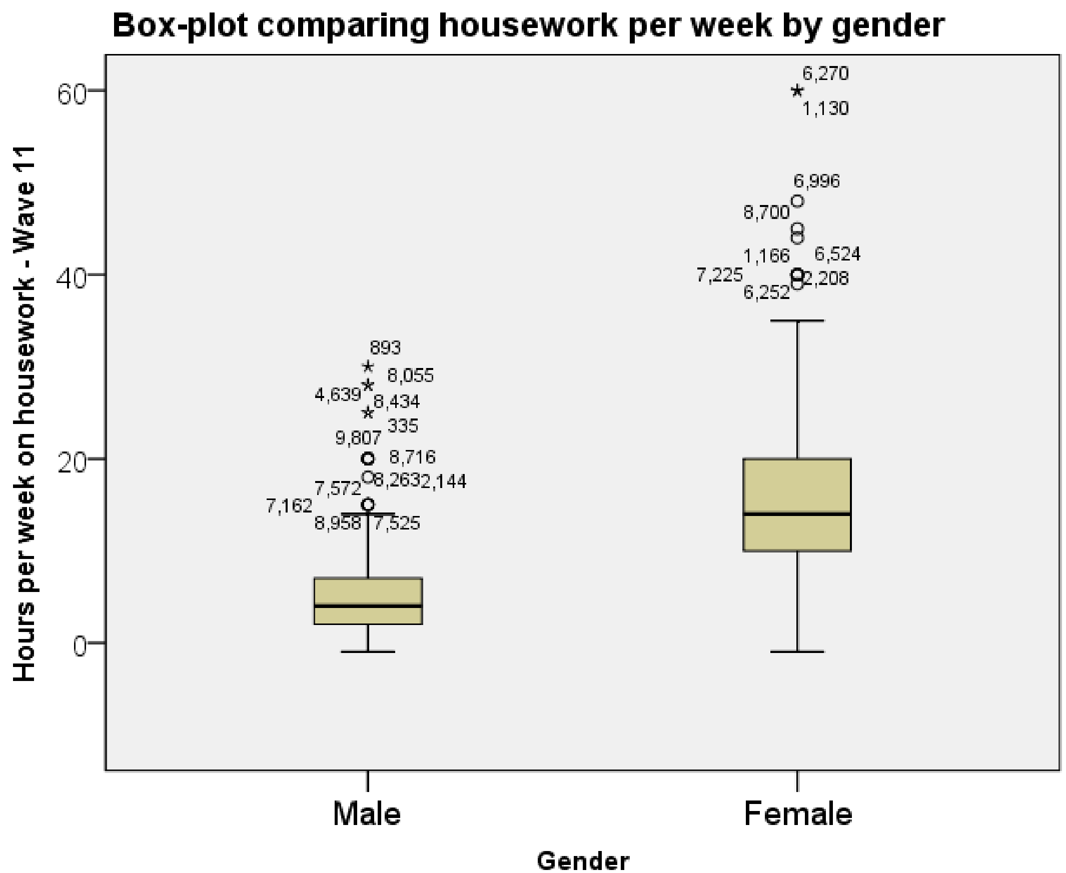
d) Is the mean or median a better summary measure?

The median as the data is skewed

Comparing means

Summarising means

- Calculate summary statistics by group
- Look for outliers/ errors
- Use a box-plot or confidence interval plot



T-tests
Paired or Independent (Unpaired) Data?

T-tests are used to compare two population means

- **Paired data:** same individuals studied at two different times or under two conditions PAIRED T-TEST
- **Independent:** data collected from two separate groups INDEPENDENT SAMPLES T-TEST

Comparison of hours worked in 1988 to today

Paired or unpaired?

If the same people have reported their hours for 1988 and 2014 have PAIRED measurements of the same variable (hours)

Paired Null hypothesis: The mean of the paired differences = 0

$$H_0 : \mu_d = 0$$

If different people are used in 1988 and 2014 have independent measurements

Independent Null hypothesis: The mean hours worked in 1988 is equal to the mean for 2014

$$H_0 : \mu_{1988} = \mu_{2014}$$

SPSS data entry

Paired Data

	Name	Hours1988	Hours2014
1	Joe Bloggs	35	38
2	Sam Smith	37	35
3	Joyce Jones	20	35
4			

Independent Groups

	Name	Hours	Year
1	Joe Bloggs	35	1988
2	Sam Smith	37	1988
3	Joyce Jones	20	1988
4	Li Yu	38	2014
5	Sally McGregor	35	2014
6	Balvinder Singh	35	2014
7			

What is the t-distribution?

- The t-distribution is similar to the standard normal distribution but has an additional parameter called degrees of freedom (df or v)

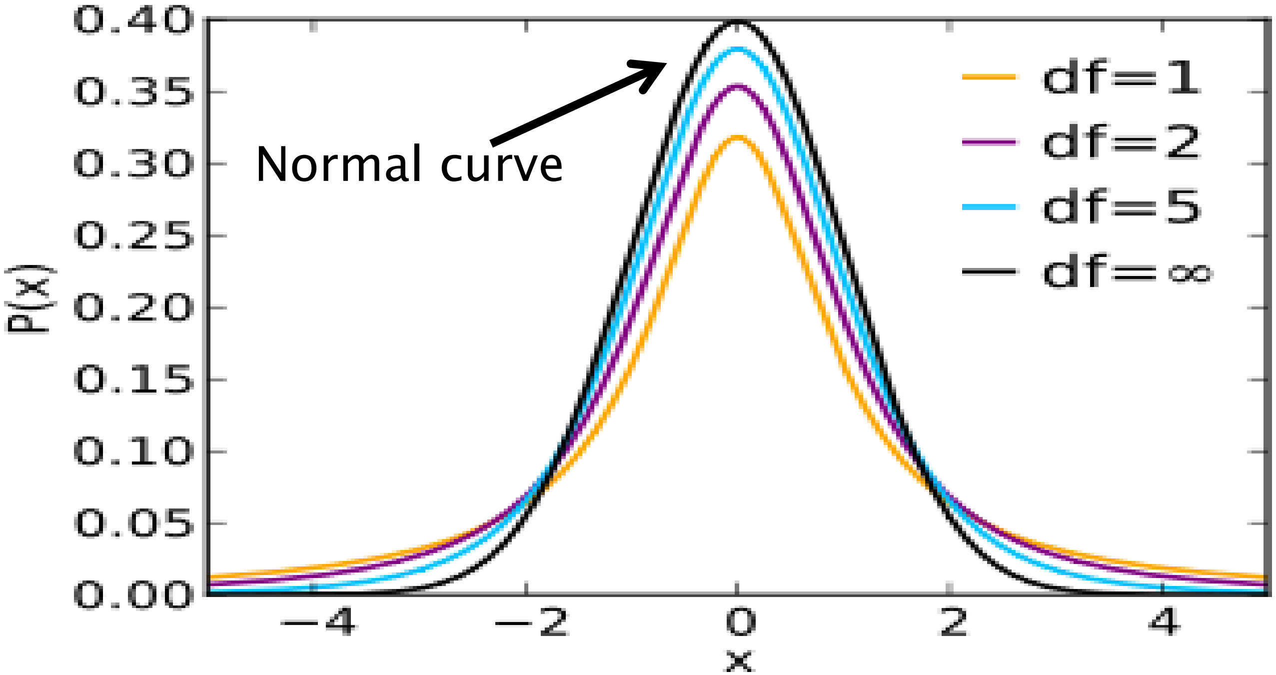
For a paired t-test, v = number of pairs – 1

For an independent t-test, $v = n_{group1} + n_{group2} - 2$

- Used for small samples and when the population standard deviation is not known
- Small sample sizes have heavier tails

Relationship to normal

- As the sample size gets big, the t-distribution matches the normal distribution



CAST e-books in statistics: t-distribution

http://cast.massey.ac.nz/collection_public.html

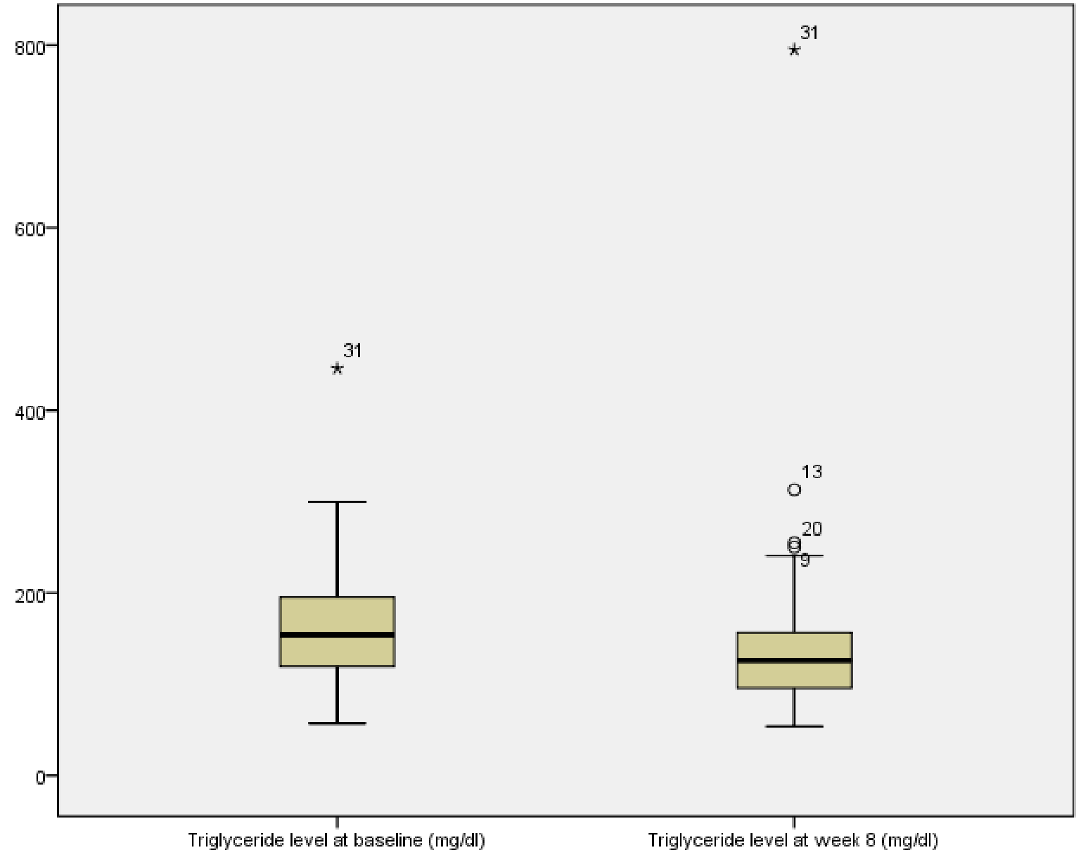
- Introductory e-book (general) and select
  - 10. Testing Hypotheses
    - 3. Tests about means
      - 4. The t-distribution
      - 5. The t-test for a mean

Exercise

- For Examples 1 and 2 (on the following four slides) discuss the answers to the following:
  - State a suitable null hypothesis
  - State whether it's a Paired or Independent Samples t-test
  - Decide whether to reject the null hypothesis
  - State a conclusion in words

Example 1: Triglycerides

- In a weight loss study, Triglyceride levels were measured at baseline and again after 8 weeks of taking a new weight loss treatment.



Example 1: t-Test Results

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Triglyceride level at week 8 (mg/dl) - Triglyceride level at baseline (mg/dl)	-11.371	80.360	13.583	-38.976	16.233	-.837	34	.408

Null Hypothesis is:

P-value =

Decision (circle correct answer): Reject Null/ Do not reject Null

Conclusion:

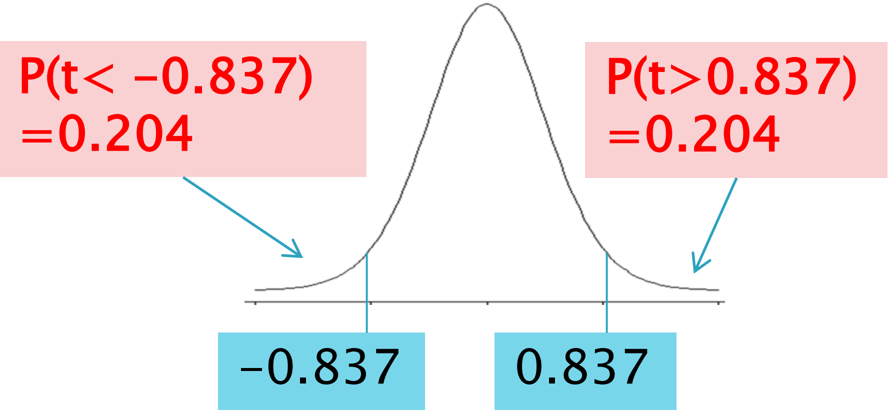
Example 1: Solution

						t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Triglyceride level at week 8 (mg/dl) – Triglyceride level at baseline (mg/dl)	-11.371	80.360	13.583	-38.976	16.233	-.837	34	.408

$$H_0 : \mu_d = 0$$

As p > 0.05, do NOT reject the null

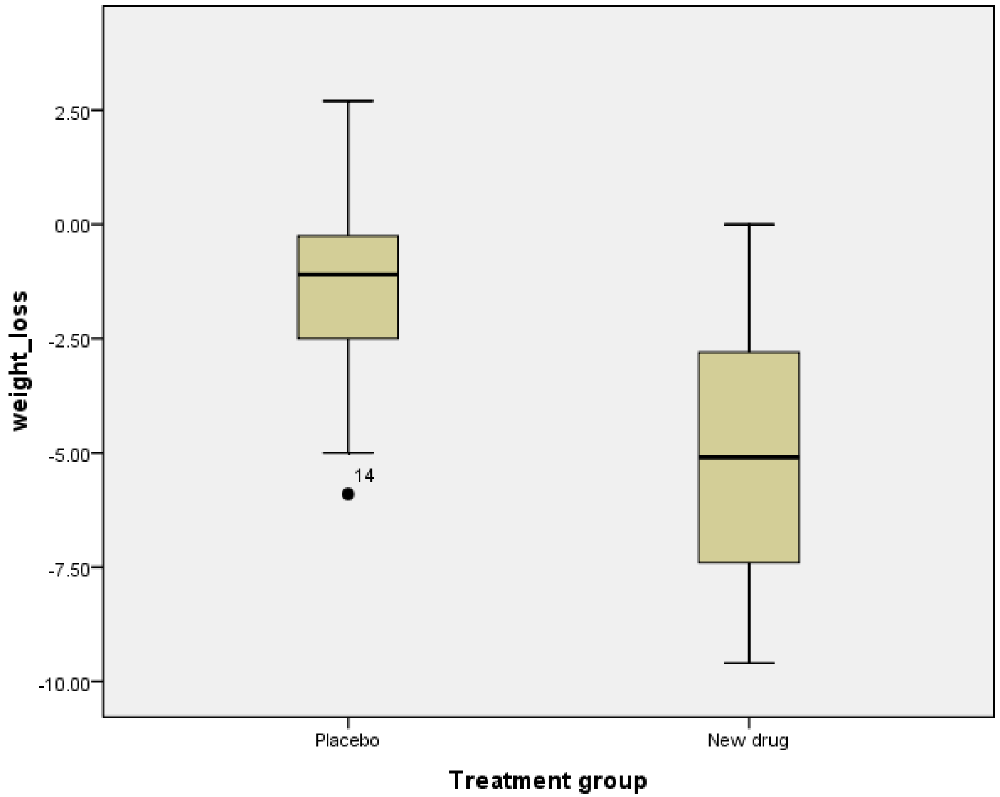
NO evidence of a difference in the mean triglyceride before and after treatment



Example 2: Weight Loss

- Weight loss was measured after taking **either** a new weight loss treatment **or** placebo for 8 weeks

Treatment group	N	Mean	Std. Deviation
Placebo	19	−1.36	2.148
New drug	18	−5.01	2.722



Ignore the shaded part of the output for now!

Example 2: t-Test Results

	Levene's Test for Equality of Variances		T-test results					95% CI of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	2.328	.136	4.539	35	.000	3.648	.804	2.016	5.280
Equal variances not assumed			4.510	32.342	.000	3.648	.809	2.001	5.295

Null Hypothesis is:

P-value =

Decision (circle correct answer): Reject Null/ Do not reject Null

Conclusion:

Ignore the shaded part of the output for now!

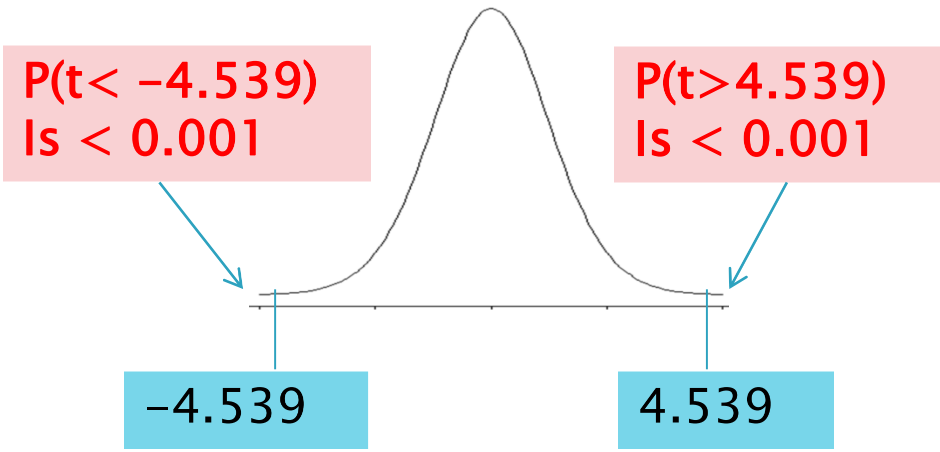
Example 2: Solution

	Levene's Test for Equality of Variances		T-test results					95% CI of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	2.328	.136	4.539	35	.000	3.648	.804	2.016	5.280
Equal variances not assumed			4.510	32.342	.000	3.648	.809	2.001	5.295

$H_0: \mu_{new} = \mu_{placebo}$

As $p < 0.05$, DO reject the null

IS evidence of a difference in weight loss between treatment and placebo



Assumptions

- Every test has assumptions
- Tutors quick guide shows assumptions for each test and what to do if those assumptions are not met

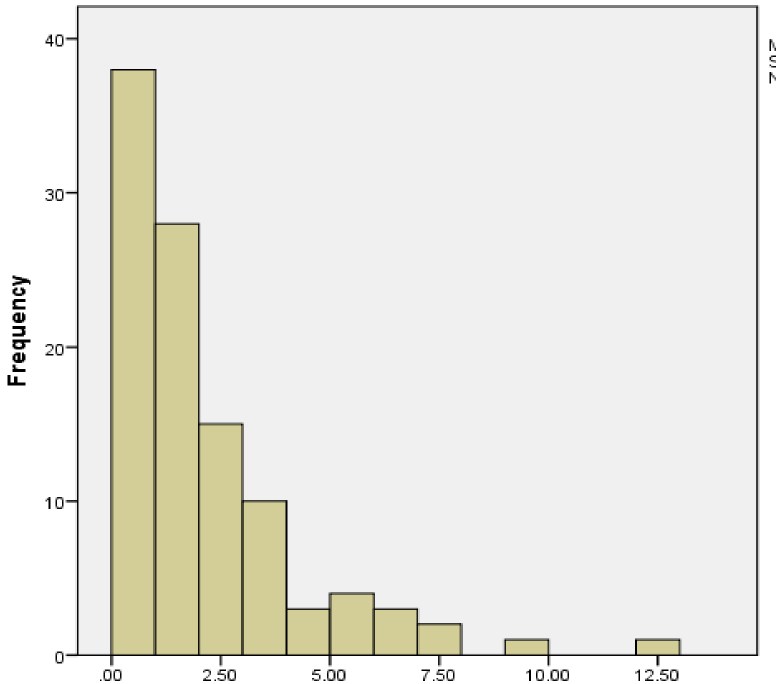
Assumptions in t-Tests

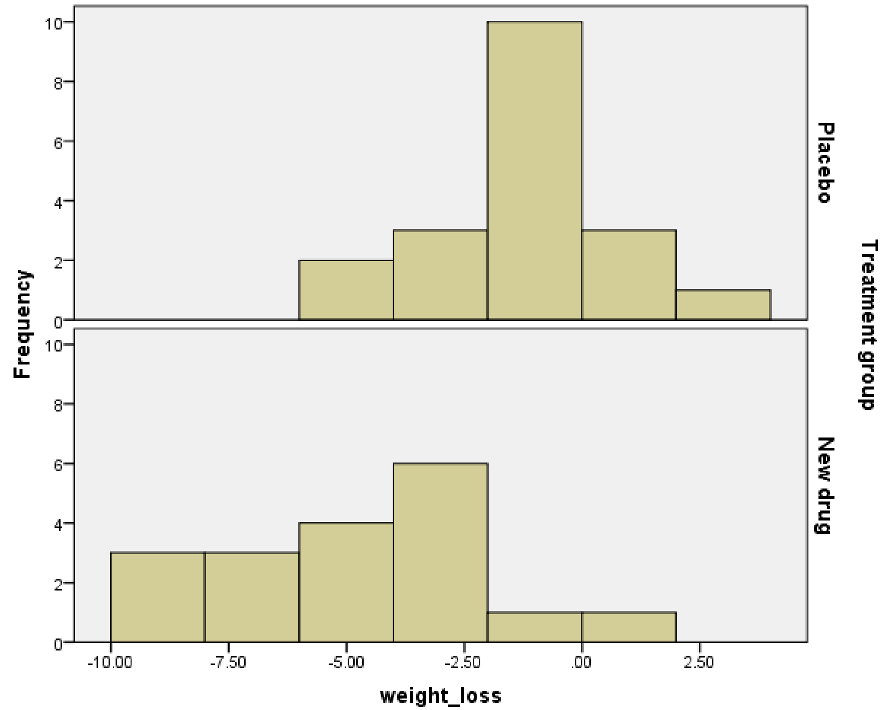
- **Normality:** Plot histograms
  - One plot of the paired differences for any paired data
  - Two (One for each group) for independent samples
  - Don’t have to be perfect, just roughly symmetric

- **Equal Population variances:** Compare sample standard deviations
  - As a rough estimate, one should be no more than twice the other
  - Do an F-test (Levene’s in SPSS) to formally test for differences

- However the *t*-test is very robust to violations of the assumptions of Normality and equal variances, particularly for moderate (i.e. >30) and larger sample sizes

Histograms from Examples 1 and 2





Do these histograms look approximately normally distributed?

Levene's Test for Equal Variances from Examples 2

	Levene's Test for Equality of Variances		T-test results					95% CI of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	2.328	.136	4.539	35	.000	3.648	.804	2.016	5.280
Equal variances not assumed			4.510	32.342	.000	3.648	.809	2.001	5.295

Null hypothesis is that pop variances are equal

i.e. H_0: $\sigma^2_{new} = \sigma^2_{placebo}$

Since $p = 0.136$ and so is >0.05 we do not reject the null

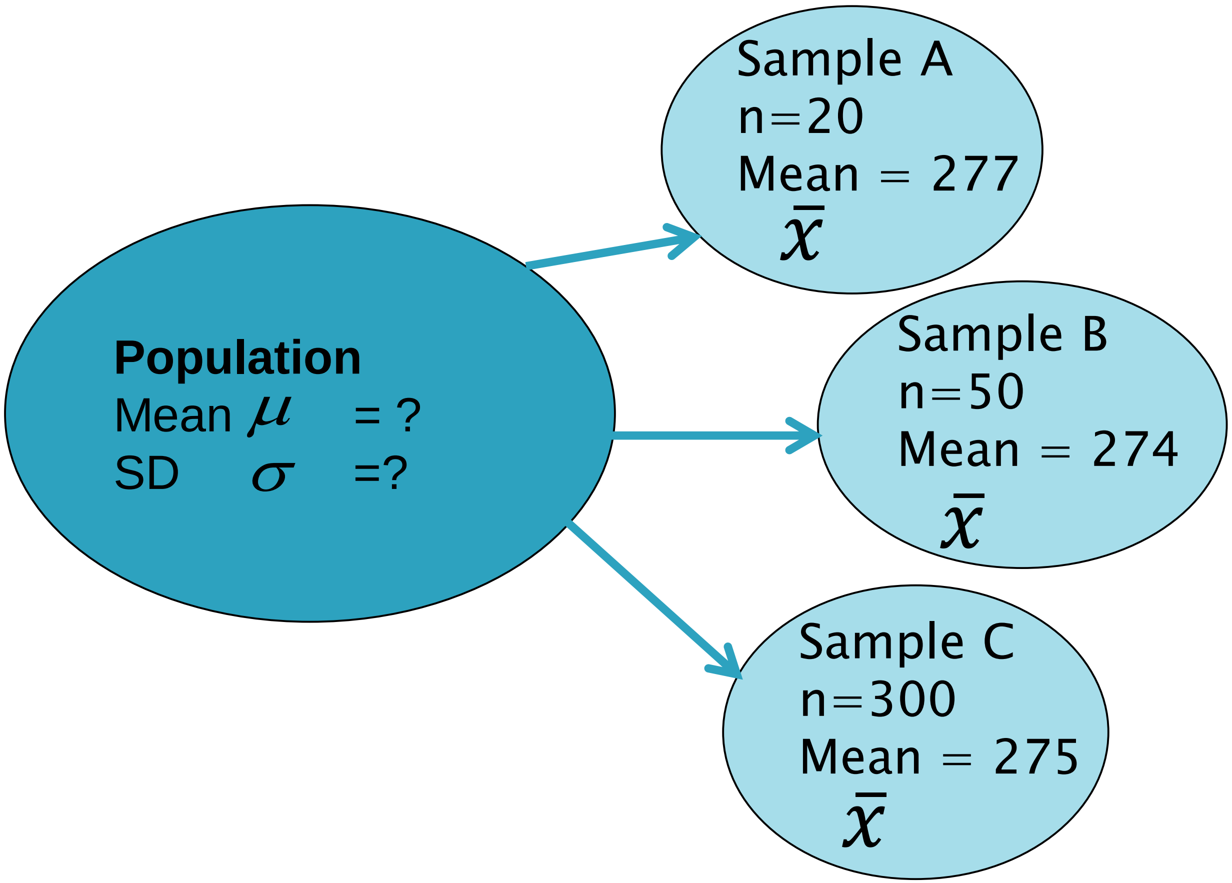
i.e. we can assume equal variances ☺

What if the assumptions are not met?

- There are alternative tests which do not have these assumptions

Test	Check	Equivalent non-parametric test
Independent t-test	Histograms of data by group	Mann-Whitney
Paired t-test	Histogram of paired differences	Wilcoxon signed rank

Sampling Variation



Every sample taken from a population, will contain different numbers so the mean varies.

Which estimate is most reliable?

How certain or uncertain are we?

Confidence Intervals

- A range of values within which we are confident (in terms of probability) that the true value of a pop parameter lies
- A 95% CI is interpreted as 95% of the time the CI would contain the true value of the pop parameter
- i.e. 5% of the time the CI would fail to contain the true value of the pop parameter

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Triglyceride level at week 8 (mg/dl) - Triglyceride level at baseline (mg/dl)	-11.371	80.360	13.583	-38.976	16.233	-.837	34	.408

CAST e-books in statistics: Confidence Intervals

http://cast.massey.ac.nz/collection_public.html

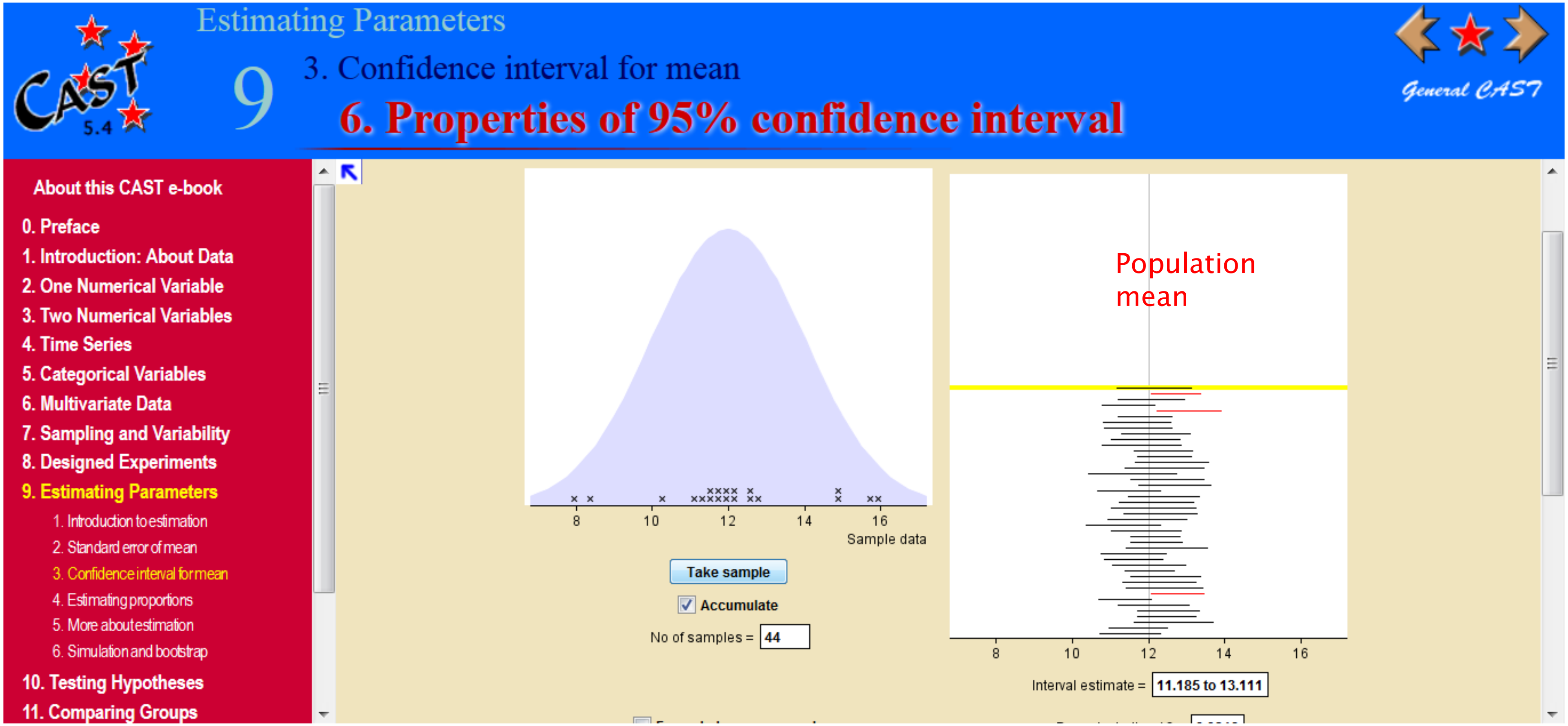
- Choose introductory e-book (general) and select

  9. Estimating Parameters

     3. Conf. Interval for mean

        6. Properties of 95% C.I.

Confidence Interval simulation from CAST



Seven do not include 141.1mmHg – we would expect that the 95% CI will not include the true population mean 5% of the time

Exercise

- Discuss what the interpretation is for the confidence interval from Example 2 (Weight loss was measured after taking **either** a new weight loss treatment **or** placebo for 8 weeks) highlighted below:

	Levene's Test for Equality of Variances		T-test results					95% CI of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	2.328	.136	4.539	35	.000	3.648	.804	2.016	5.280

Exercise: Solution

- Discuss what the interpretation is for the confidence interval from Example 2 highlighted below:

	Levene's Test for Equality of Variances		T-test results					95% CI of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	2.328	.136	4.539	35	.000	3.648	.804	2.016	5.280

The true mean weight loss would be between about 2 to 5 kg with the new treatment.

This is always positive hence the hypothesis test rejected the null that the difference is zero

Choosing the right test

Choosing the right test

- One of the most common queries in stats support is ‘Which analysis should I use’
- There are several steps to help the student decide
- When a student is explaining their project, these are the questions you need answers for

Choosing the right test

1) A clearly defined research question

2) What is the dependent variable and what type of variable is it?

3) How many independent variables are there and what data types are they?

4) Are you interested in comparing means or investigating relationships?

5) Do you have repeated measurements of the same variable for each subject?

Research question

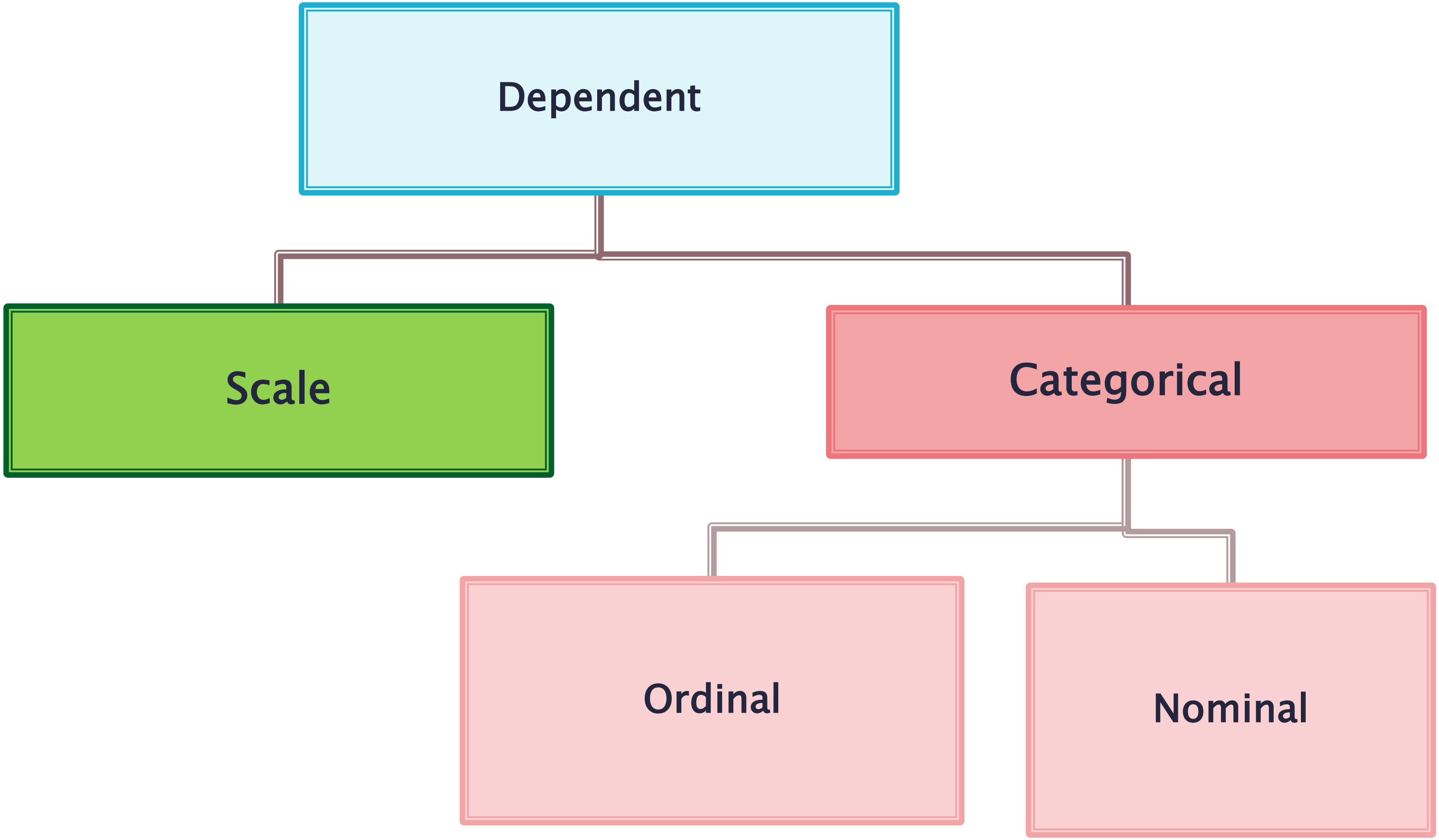
- Clear questions with measurable quantities
- Which variables will help answer these questions
- Think about what test is needed before carrying out a study so that the right type of variables are collected

Dependent variables

Does **attendance** have an association with **exam score**?

Do **women** do more **housework** than **men**?

What variable type is the dependent?



Are boys better at maths?

- How can 'better' be measured and what type of variable is it?

Exam score (Scale)

- Do boys think they are better at maths??

I consider myself to be good at maths (ordinal)

How many variables are involved?

- Two – interested in the relationship
- One dependent and one independent
- One dependent and several independent variables: some may be controls
- Relationships between more than two: multivariate techniques (not covered here)

Data types

Research question	Dependent/ outcome variable	Independent/ explanatory variable
Does attendance have an association with exam score?	Exam score (scale)	Attendance (Scale)
Do women do more housework than men?	Hours of housework per week (Scale)	Gender (binary)

Exercise:

How would you investigate the following topics? State the dependent and independent variables and their variable types.

Research question	Dependent/ outcome variable	Independent/ explanatory variable
Were Americans more likely to survive on board the Titanic?		
Does weekly hours of work influence the amount of time spent on housework?		
Which of 3 diets is best for losing weight?		

Exercise: Solution

How would you investigate the following topics?
State the dependent and independent variables and their variable types.

Research question	Dependent/ outcome variable	Independent/ explanatory variable
Were Americans more likely to survive on board the Titanic?	Survival (Binary)	Nationality (Nominal)
Does weekly hours of work influence the amount of time spent on housework?	Hours of housework (Scale)	Hours of work (Scale)
Which of 3 diets is best for losing weight?	Weight lost on diet (Scale)	Diet (Nominal)

Comparing means

- Dependent = Scale
- Independent = Categorical

- How many means are you comparing?

- Do you have independent groups or repeated measurements on each person?

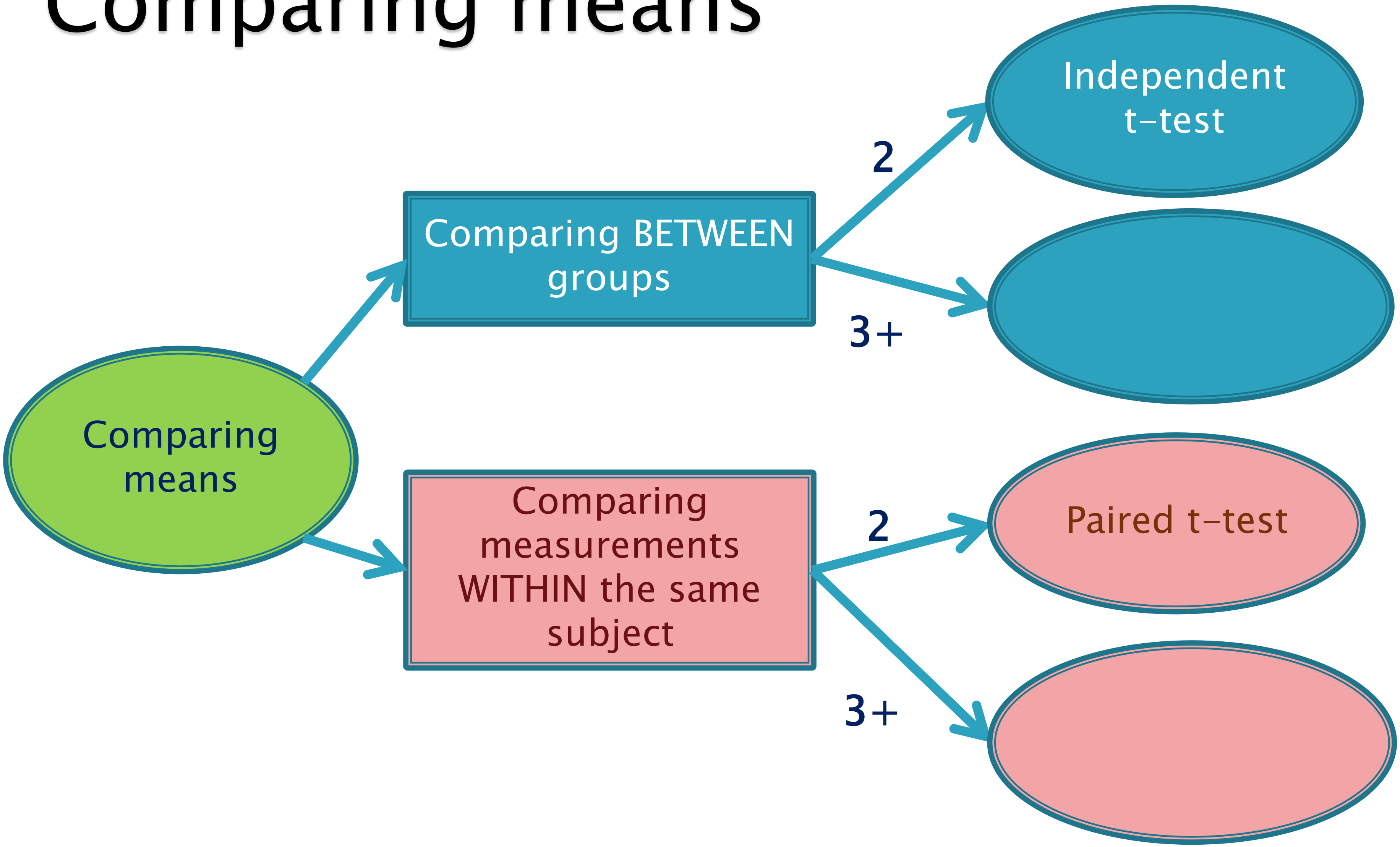
Comparing measurements on the same people

Also known as within group comparisons or repeated measures.

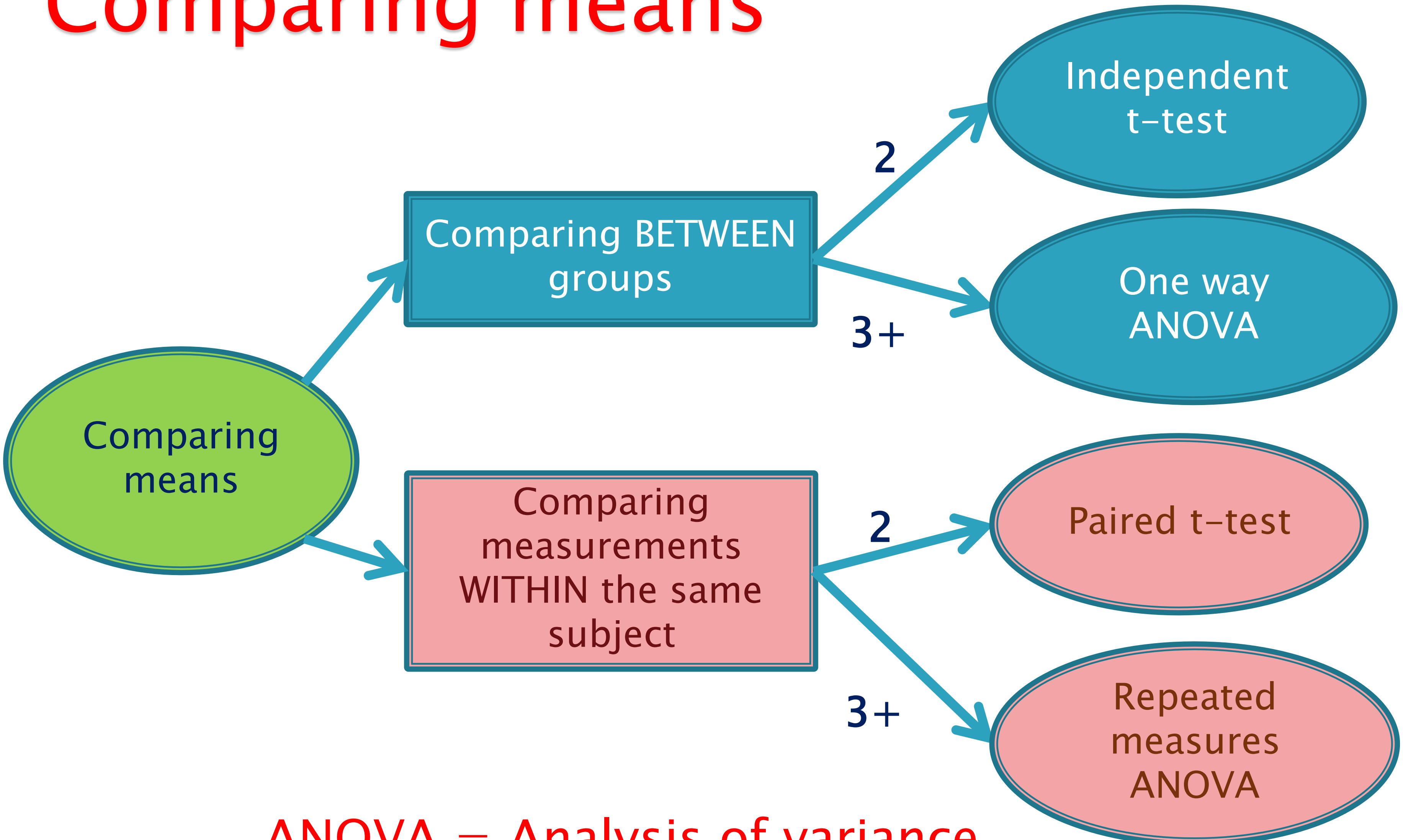
Can be used to look at differences in mean score:
(1) over 2 or more time points e.g. 1988 vs 2014
(2) under 2 or more conditions e.g. taste scores

Participants are asked to taste 2 types of cola and give each scores out of 100.
Dependent = taste score
Independent = type of cola

Comparing means



Comparing means



ANOVA = Analysis of variance

Exercise – Comparing means

Research question	Dependent variable	Independent variable	Test
Do women do more housework than men?	Housework (hrs per week) (Scale)	Gender (Nominal)	
Does Margarine X reduce cholesterol? Everyone has cholesterol measured on 3 occasions	Cholesterol (Scale)	Occasion (Nominal)	
Which of 3 diets is best for losing weight?	Weight lost on diet (Scale)	Diet (Nominal)	

Exercise: Solution

Research question	Dependent variable	Independent variable	Test
Do women do more housework than men?	Housework (hrs per week) (Scale)	Gender (Nominal)	Independent t-test
Does Margarine X reduce cholesterol? Everyone has cholesterol measured on 3 occasions	Cholesterol (Scale)	Occasion (Nominal)	Repeated measures ANOVA
Which of 3 diets is best for losing weight?	Weight lost on diet (Scale)	Diet (Nominal)	One-way ANOVA

Tests investigating relationships

Investigating relationships between	Dependent variable	Independent variable	Test
2 categorical variables	Categorical	Categorical	Chi-squared test
2 Scale variables	Scale	Scale	Pearson's correlation
Predicting the value of an dependent variable from the value of a independent variable	Scale	Scale/binary	Simple Linear Regression
	Binary	Scale/ binary	Logistic regression

Note: Multiple linear regression is when there are several independent variables

Exercise: Relationships

Research question	Dependent variable	Independent variables	Test
Does attendance affect exam score?	Exam score (Scale)	Attendance (Scale)	
Do women do more housework than men?	Housework (hrs per week) (scale)	Gender (Binary) Hours worked (Scale)	
Were Americans more likely to survive on board the Titanic?	Survival (Binary)	Nationality (Nominal)	
	Survival (Binary)	Nationality , Gender, class	

Note: There may be 2 appropriate tests for some questions

Exercise: Solution

Research question	Dependent variable	Independent variables	Test
Does attendance affect exam score?	Exam score (Scale)	Attendance (Scale)	Correlation/ regression
Do women do more housework than men?	Housework (hrs per week) (scale)	Gender (Binary) Hours worked (Scale)	Regression
Were Americans more likely to survive on board the Titanic?	Survival (Binary)	Nationality (Nominal)	Chi-squared
	Survival (Binary)	Nationality , Gender, class	Logistic regression

Note: There may be 2 appropriate tests for some questions

Non-parametric tests

Parametric or non-parametric?

Statistical tests fall into two types:

Parametric tests — Assume data follows a particular distribution e.g. normal

Non-parametric — Nonparametric techniques are usually based on ranks/ signs rather than actual data

Ranking raw data

- Nonparametric techniques are usually based on ranks or signs
- Scale data is ordered and ranked
- Analysis is carried out on the ranks rather than the actual data

Original variable → ReactionTime

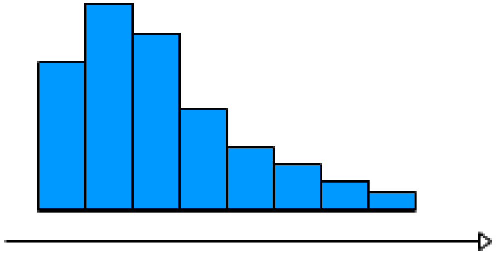
Rank of subject → Rank

Group	ReactionTime	Rank
Placebo	.37	1
Placebo	.38	2
Placebo	.61	3
Placebo	.78	4
Placebo	.83	5
Placebo	.86	6
Placebo	.90	7
Placebo	.95	8
Alcohol	.98	9
Alcohol	1.11	10
Alcohol	1.27	11
Alcohol	1.32	12
Alcohol	1.44	13
Alcohol	1.45	14
Alcohol	1.46	15
Placebo	1.63	16
Alcohol	1.76	17
Placebo	1.97	18
Alcohol	2.56	19
Alcohol	3.07	20

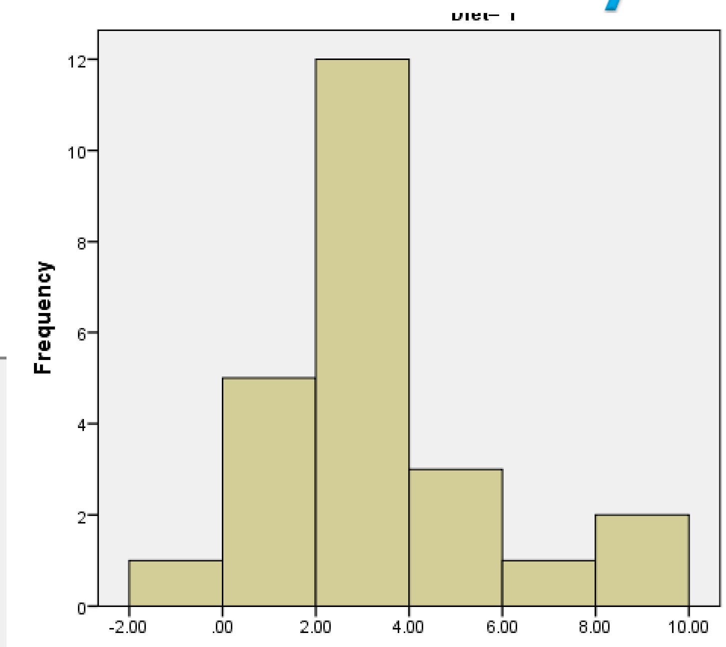
Non-parametric tests

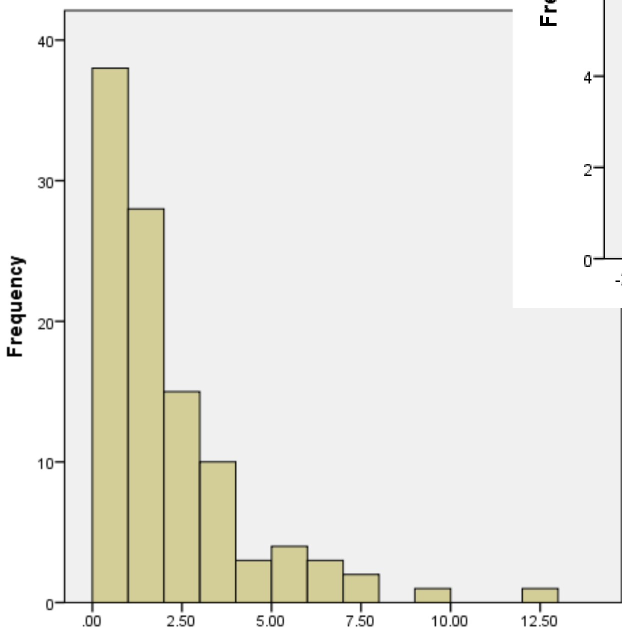
- Non-parametric methods are used when:
  - Data is ordinal
  - Data does not seem to follow any particular shape or distribution (e.g. Normal)
  - Assumptions underlying parametric test not met
  - A plot of the data appears to be very skewed
  - There are potential influential outliers in the dataset
  - Sample size is small

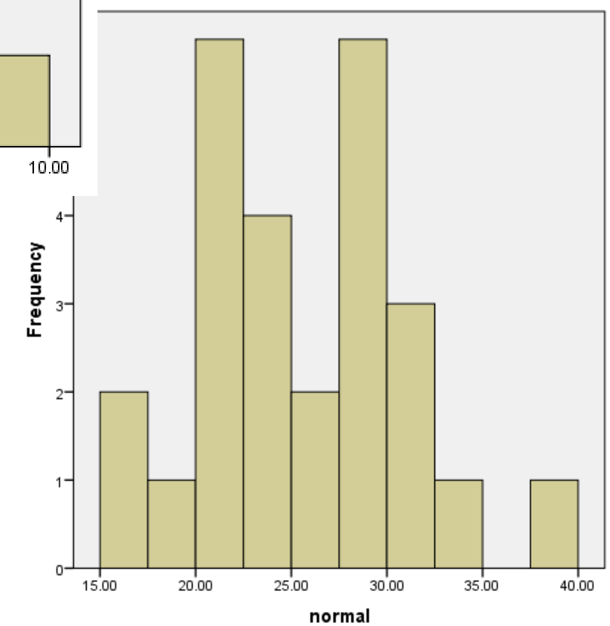
Note: Parametric tests are fairly robust to non-normality. Data has to be very skewed to be a problem



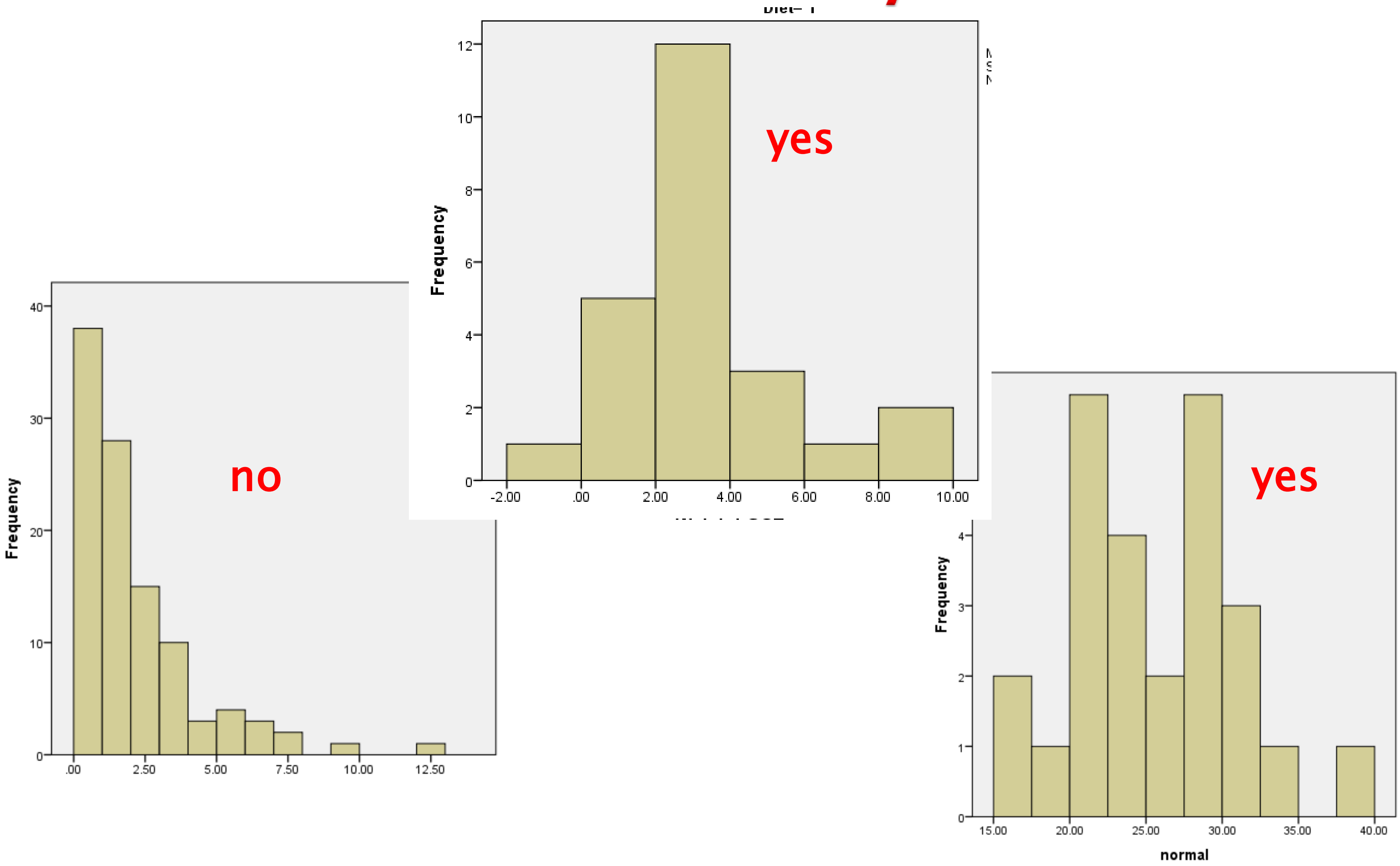
Do these look normally distributed?







Do these look normally distributed?

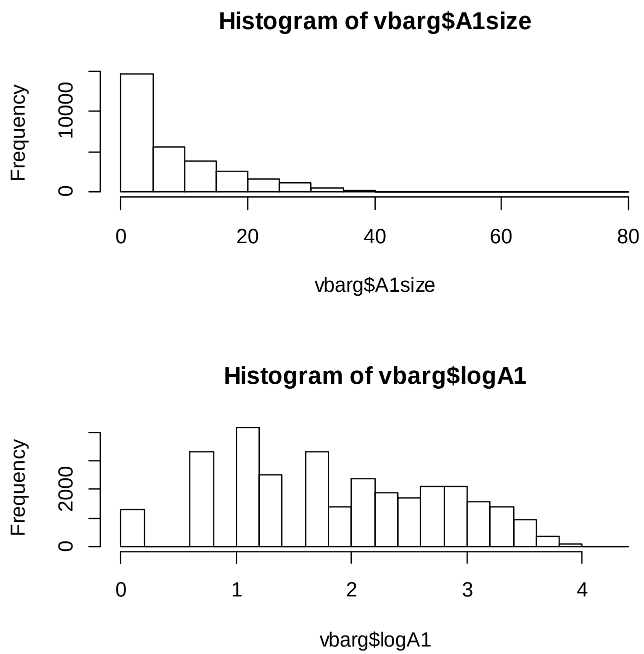


What can be done about non-normality?

If the data are not normally distributed, there are two options:

1. **Use a non-parametric test**
2. **Transform the dependent variable**

For positively skewed data, taking the log of the dependent variable often produces normally distributed values



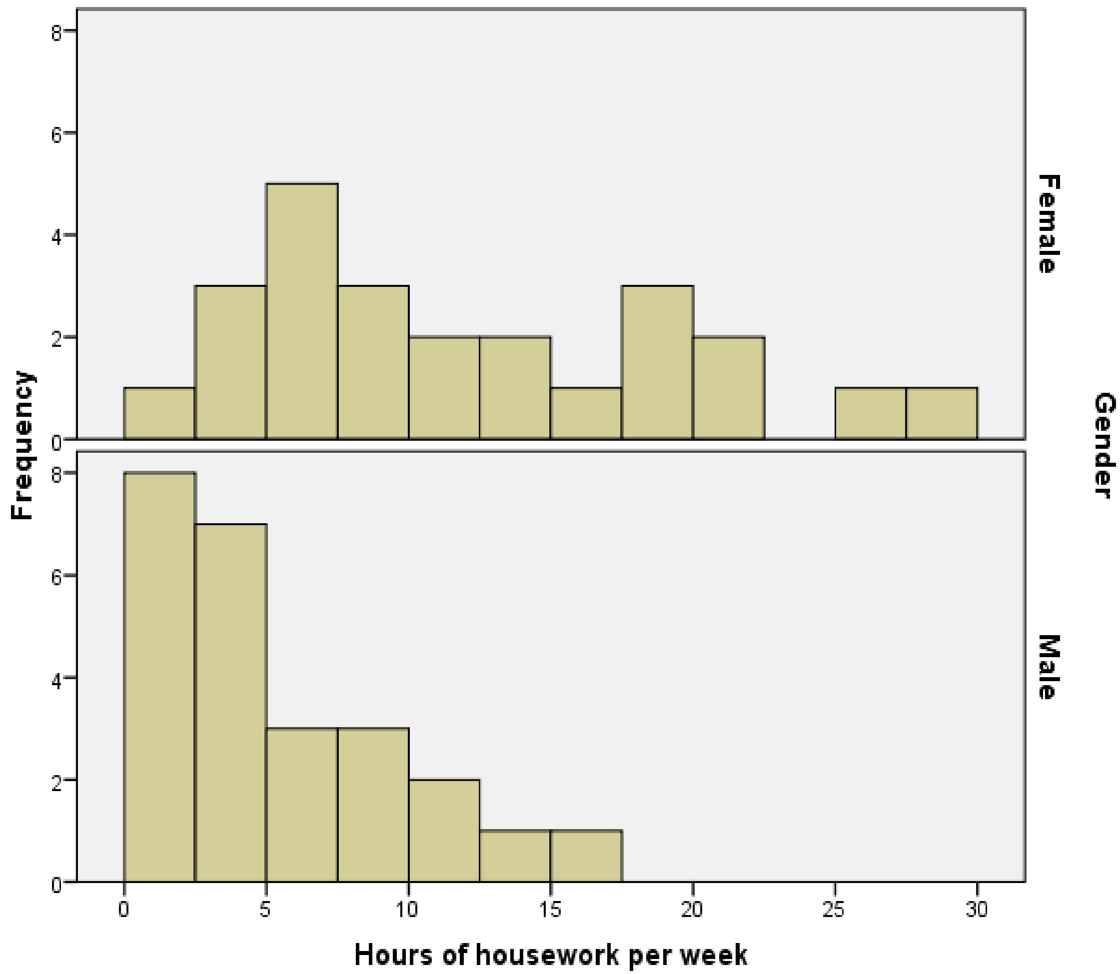
Non-parametric tests

Parametric test	What to check for normality	Non-parametric test
Independent t-test	Dependent variable by group	Mann-Whitney test
Paired t-test	Paired differences	Wilcoxon signed rank test
One-way ANOVA	Residuals/Dependent	Kruskal-Wallis test
Repeated measures ANOVA	Residuals	Friedman test
Pearson's Correlation Co-efficient	At least one of the variables should be normal	Spearman's Correlation Co-efficient
Linear Regression	Residuals	None – transform the data

Notes: The residuals are the differences between the observed and expected values.

Exercise: Comparison of housework

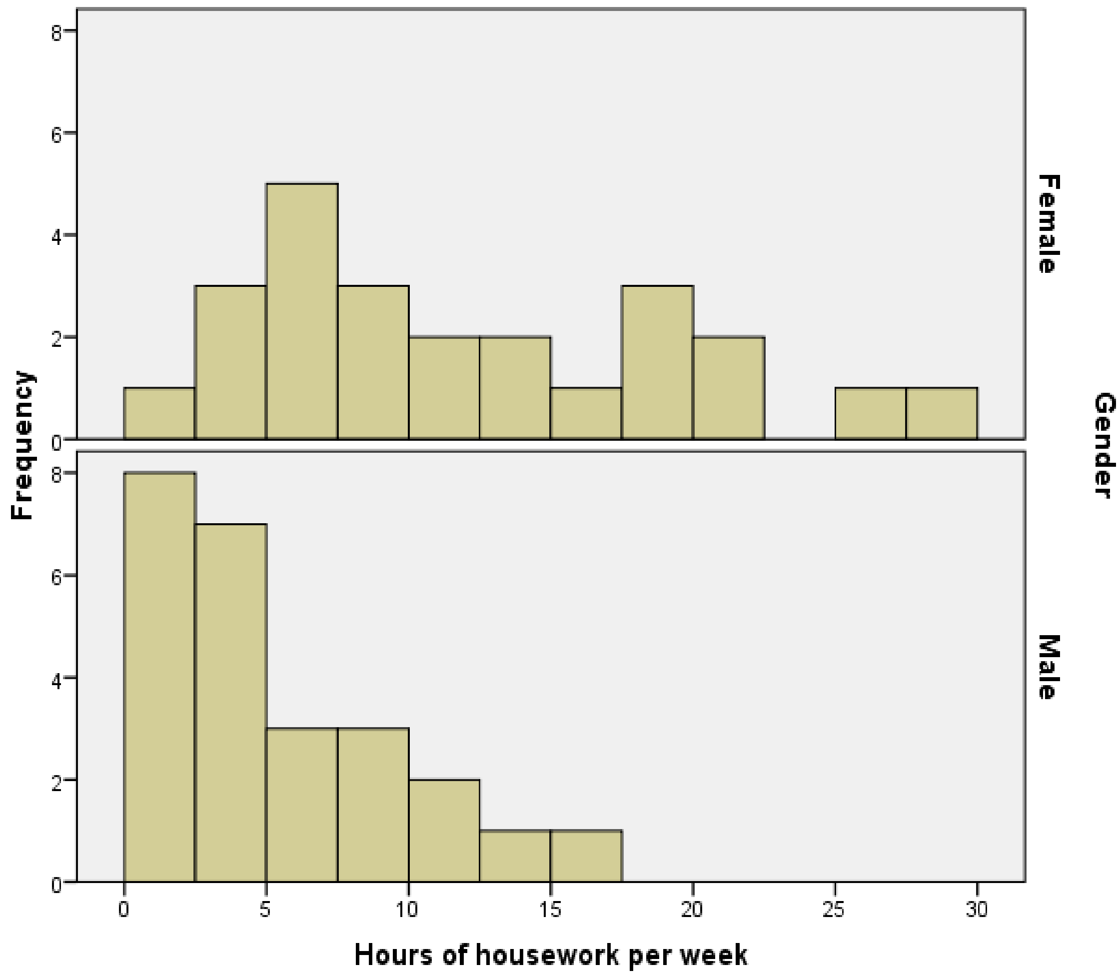
Which test should be carried out to compare the hours of housework for males and females? Look at the histograms of housework by gender to decide.



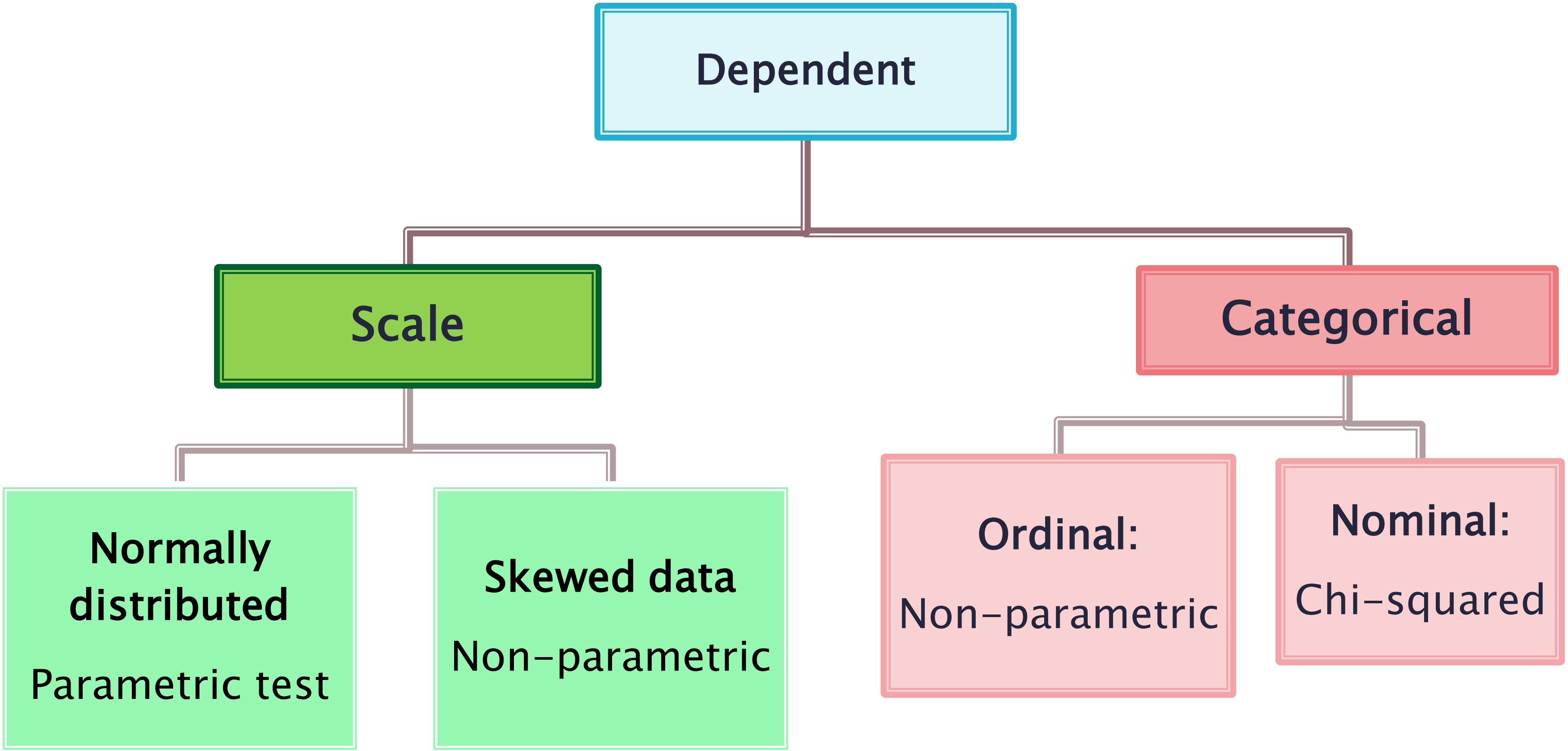
Exercise: Solution

Which test should be carried out to compare the hours of housework for males and females?

The male data is very skewed so use the Mann–Whitney.



Summary


Dependent
Scale
Categorical
Normally distributed
Parametric test
Skewed data
Non-parametric
Ordinal:
Non-parametric
Nominal:
Chi-squared

Scenario – questions to consider

Research question

Dependent variable and type:

Independent variables:

Are there repeated measurements of the same variable for each subject?

Friedman results

The student has never studied hypothesis testing before. Explain the concepts and what a p-value is. Is there a difference between the scores given to the preferences of the different criteria?

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.
1	The distributions of Visual Feedback, Price Update and Virtual use of Product are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.358

Asymptotic significances are displayed. The significance level is .05.

SPSS: Analyze → Non-parametric Tests → Related Samples

Resume

Do you feel that you

1. Have improved your ability to recall and describe some basic applied statistical methods?

2. Have developed ideas for explaining some of these methods to students in a maths support setting?

3. Feel more confident about providing statistics support with the topics covered?

Mann Whitney Test

Legal drink drive limits

Research question: Does drinking the legal limit for alcohol affect driving reaction times?

- Participants were given either alcoholic or non-alcoholic drinks and their driving reaction times tested on a simulated driving situations
- They did not know whether they had the alcohol or not

Placebo: 0.90 0.37 1.63 0.83 0.95 0.78 0.86 0.61 0.38 1.97
Alcohol: 1.46 1.45 1.76 1.44 1.11 3.07 0.98 1.27 2.56 1.32

Mann–Whitney test

- Nonparametric equivalent to independent t–test
- The data from both groups is ordered and ranked
- The mean rank for the groups is compared

Group	ReactionTime	Rank
Placebo	.37	1
Placebo	.38	2
Placebo	.61	3
Placebo	.78	4
Placebo	.83	5
Placebo	.86	6
Placebo	.90	7
Placebo	.95	8
Alcohol	.98	9
Alcohol	1.11	10
Alcohol	1.27	11
Alcohol	1.32	12
Alcohol	1.44	13
Alcohol	1.45	14
Alcohol	1.46	15
Placebo	1.63	16
Alcohol	1.76	17
Placebo	1.97	18
Alcohol	2.56	19
Alcohol	3.07	20

Drink driving reactions

H_0: There is no difference between the alcohol and placebo populations on reaction time

H_a: The alcohol population has a different reaction time distribution to the placebo population

Test Statistic = Mann–Whitney U

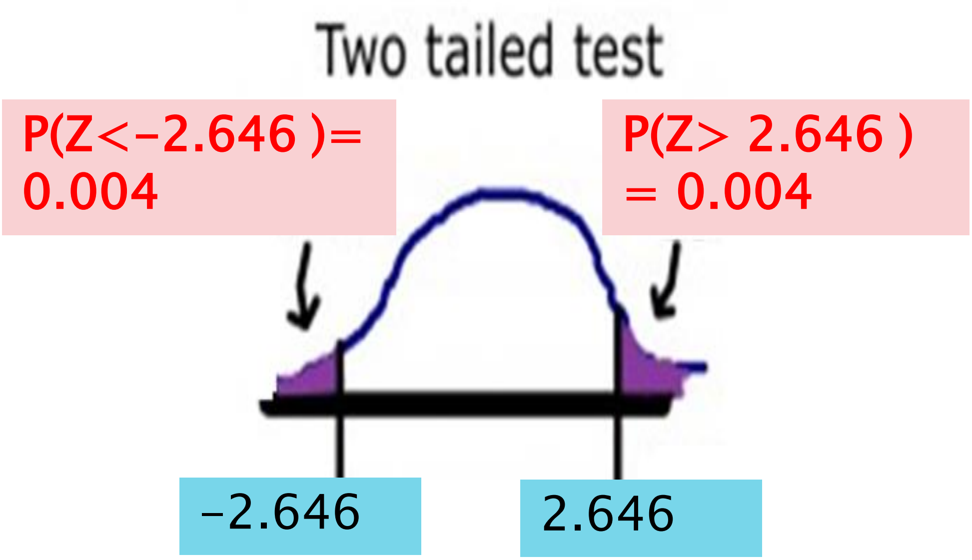
The test statistic U can be approximated to a z score to get a p-value

Mann-Whitney results

- Interpret the results

Test Statistics

	ReactionTime
Mann-Whitney U	15.000
Z	-2.646
p-value (2-tailed)	.008



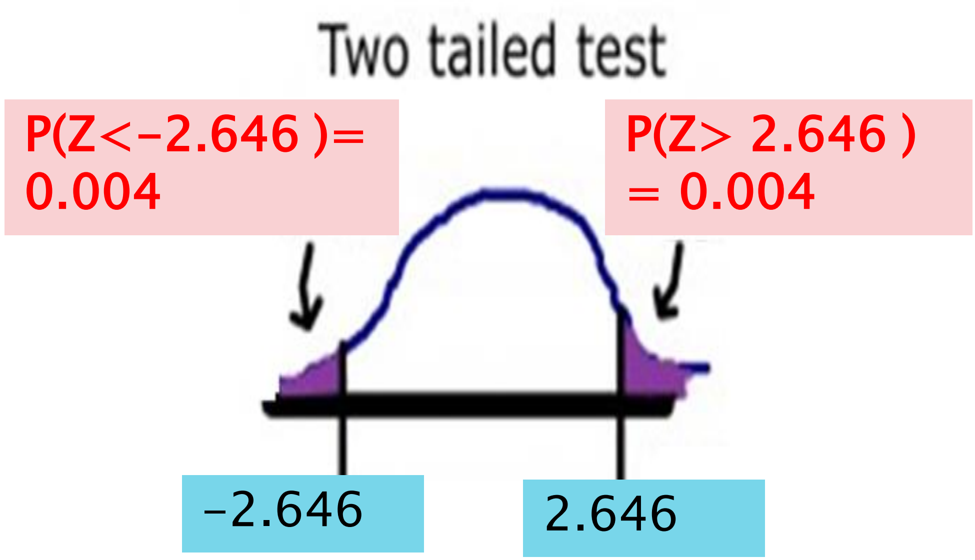
Mann–Whitney results – solution

- Interpret the results

- p =0.008
- Highly significant evidence to suggest a difference in the distributions of reaction times for those in the placebo and alcohol groups

Test Statistics

	ReactionTime
Mann-Whitney U	15.000
Z	-2.646
p-value (2-tailed)	.008



ANOVA

ANOVA

Compares the means of several groups

Which diet is best?
Dependent: Weight lost (Scale)
Independent: Diet 1, 2 or 3 (Nominal)

Null hypothesis: The mean weight lost on diets 1, 2 and 3 is the same

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Alternative hypothesis: The mean weight lost on diets 1, 2 and 3 are not all the same

Summary statistics

	Overall	Diet 1	Diet 2	Diet 3
Mean	3.85	3.3	3.03	5.15
Standard deviation	2.55	2.24	2.52	2.4
Number in group	78	24	27	27

- Which diet was best?
- Are the standard deviations similar?

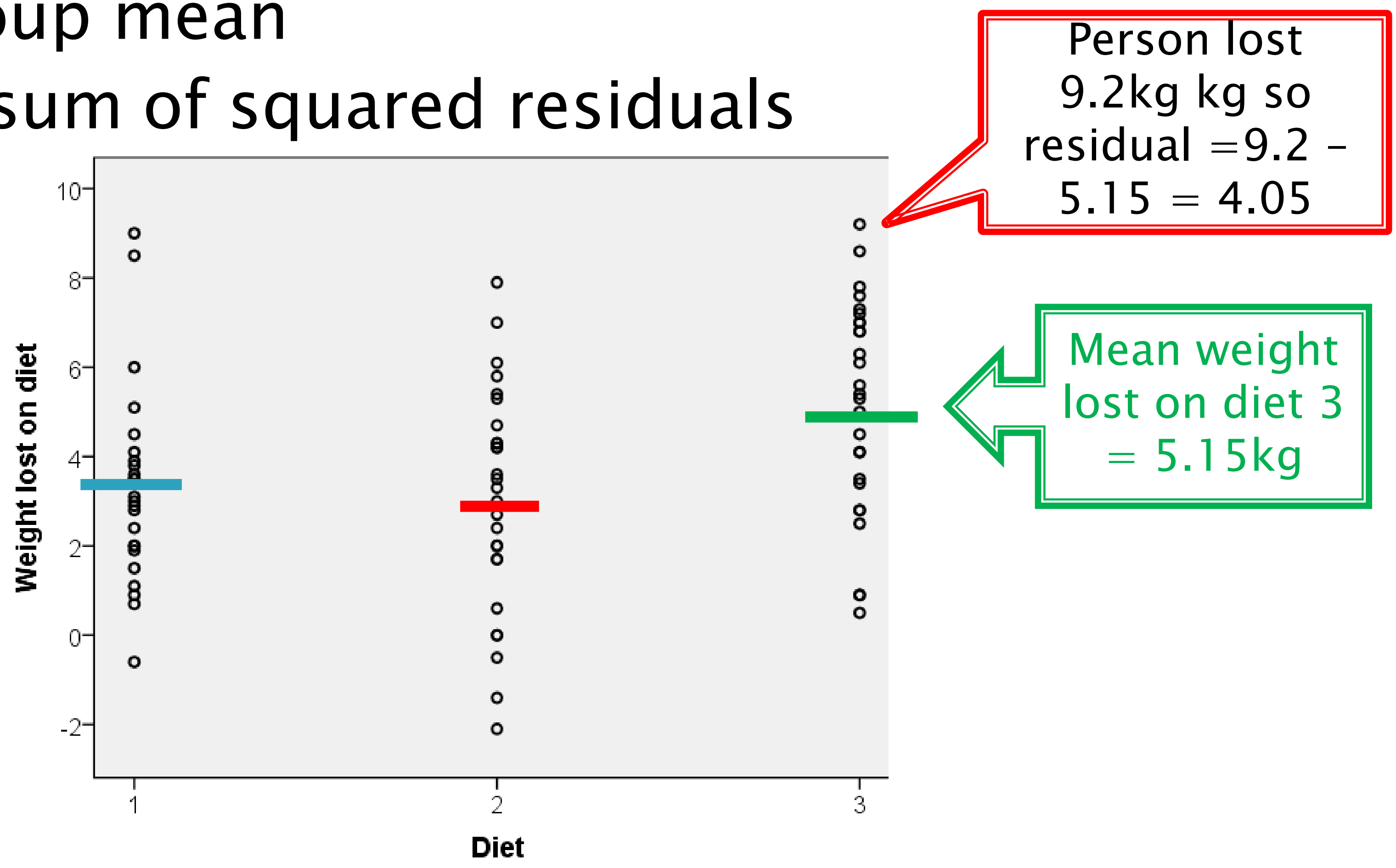
How Does ANOVA Work?

- ANOVA = Analysis of variance
- We compare variation **between** groups relative to variation **within** groups
- Population variance estimated in two ways:
  - One based on variation **between** groups we call the **M**ean **S**quare due to **T**reatments/ **MST**/ $\mathbf{MS_{between}}$
  - Other based on variation **within** groups we call the **M**ean **S**quare due to **E**rror/ **MSE**/ $\mathbf{MS_{within}}$

Within group variation

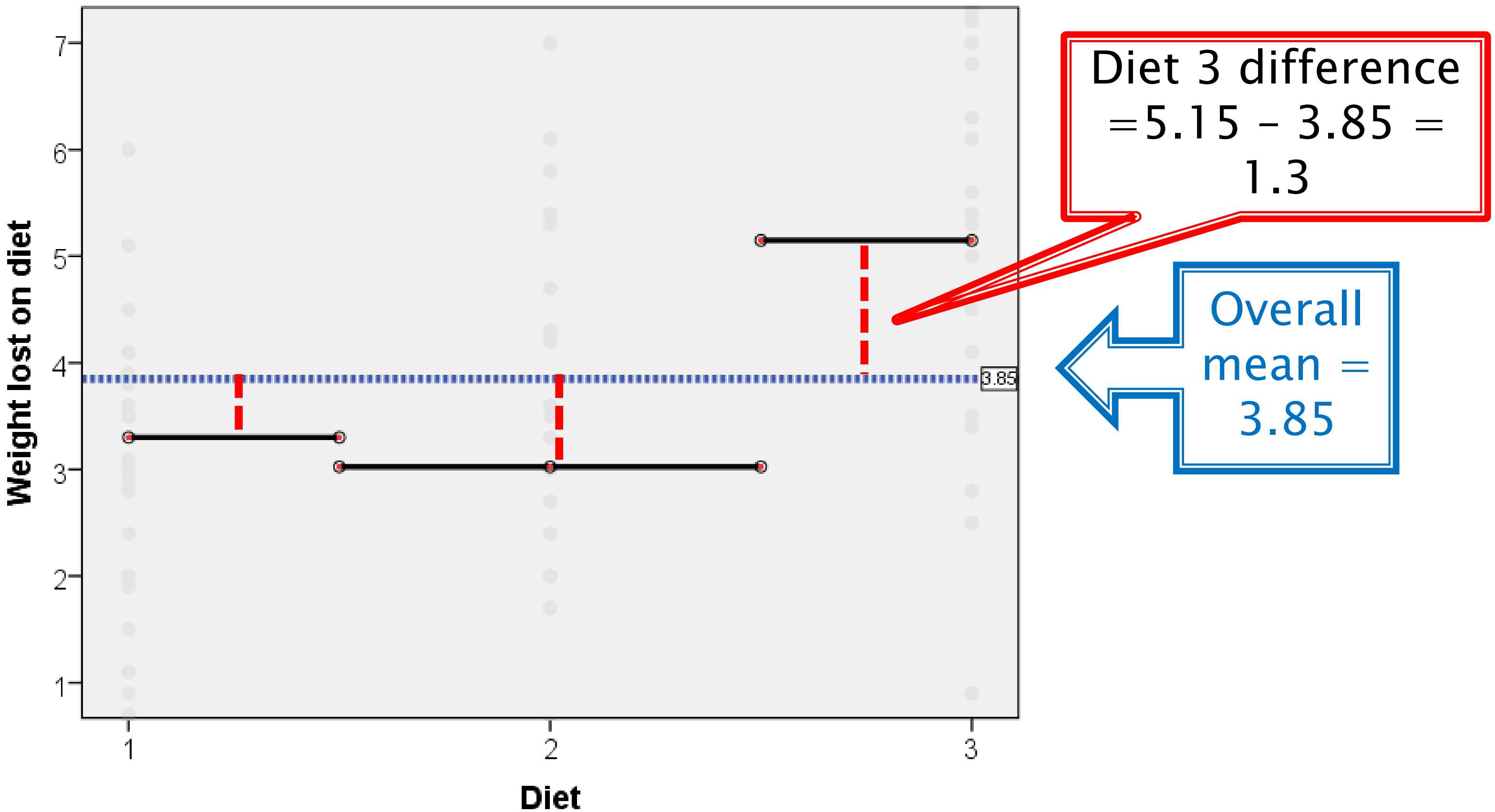
Residual =difference between an individual and their group mean

SS_{within}=sum of squared residuals



Between group variation

Differences between each group mean and the overall mean



Sum of squares calculations

- K = number of groups

$$SS_{within} = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2$$

$$= \sum_{i=1}^{24}(x_i - 3.3)^2 + \sum_{i=1}^{27}(x_i - 3.03)^2 + \sum_{i=1}^{27}(x_i - 5.15)^2 = 430.179$$

$$SS_{Between} = \sum_{j=1}^{k} n_j\left(\overline{x_j} - \overline{x_T}\right)^2$$

$$= 24(3.3 - 3.85)^2 + 27(3.03 - 3.85)^2 + 27(5.15 - 3.85)^2 = 71.094$$

ANOVA test statistic

Summary ANOVA

Source	Sum of Squares	Degrees of Freedom	Variance Estimate (Mean Square)	*F* Ratio
Between	SS_B	$K-1$	$MS_B = \frac{SS_B}{K-1}$	$\frac{MS_B}{MS_W}$
Within	SS_W	$N-K$	$MS_W = \frac{SS_W}{N-K}$	
Total	$SS_T = SS_B + SS_W$	$N-1$		

Test Statistic (usually reported in papers)

N = total observations in all groups,

K = number of groups

Test Statistic (by hand)

- Filling in the boxes

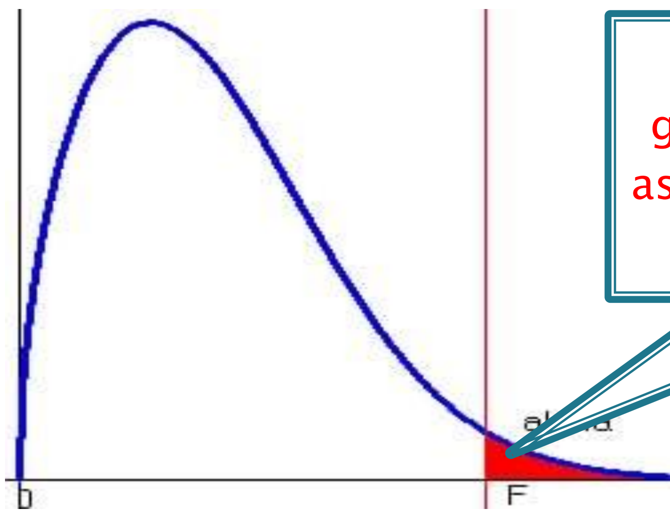
	Sum of squares	Degrees of freedom	Mean square	F-ratio (test statistic)
$SS_{between}$	71.045	2	35.522	**6.193**
SS_{within}	430.180	75	5.736	
SS_{total}	501.275	77		

$$F = \frac{\text{Mean between group sum of squared differences}}{\text{Mean within group sum of squared differences}}$$

If F > 1, there is a bigger difference between groups than within groups

P-value

- The p-value for ANOVA is calculated using the F-distribution
- If you repeated the experiment numerous times, you would get a variety of test statistics



p-value = probability of getting a test statistic at least as extreme as ours if the null is true

Test Statistic

One way ANOVA

$$\text{Test Statistic} = \frac{\text{between group variation}}{\text{within group variation}} = \frac{MS_{\text{Diet}}}{MS_{\text{Error}}} = 6.197$$

Tests of Between-Subjects Effects

Dependent Variable: Weight lost on diet (kg)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	71.094[a]	2	35.547	6.197	.003
Intercept	1137.494	1	1137.494	198.317	.000
Diet	71.094	2	35.547	6.197	.003
Error	430.179	75	5.736		
Total	1654.350	78			
Corrected Total	501.273	77			

a. R Squared = .142 (Adjusted R Squared = .119)

$MS_{between}$

MS_{within}

There was a significant difference in weight lost between the diets (p=0.003)

Post hoc tests

If there is a significant ANOVA result, pairwise comparisons are made

They are t–tests with adjustments to keep the type 1 error to a minimum

- Tukey's and Scheffe's tests are the most commonly used post hoc tests.
- Hochberg's GT2 is better where the sample sizes for the groups are very different.

Post hoc tests

- Which diets are significantly different?

Multiple Comparisons

Dependent Variable: Weight lost on diet (kg)

	(I) Diet	(J) Diet	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1	2	.2741	.67188	.912	-1.3325	1.8806
		3	-1.8481*	.67188	.020	-3.4547	-.2416
	2	1	-.2741	.67188	.912	-1.8806	1.3325
		3	-2.1222*	.65182	.005	-3.6808	-.5636
	3	1	1.8481*	.67188	.020	.2416	3.4547
		2	2.1222*	.65182	.005	.5636	3.6808

- Write up the results and conclude with which diet is the best.

Pairwise comparisons

- Results

Test	p-value
Diet 1 vs Diet 2	
Diet 1 vs Diet 3	
Diet 2 vs Diet 3	

- Report:

Pairwise comparisons

- Results

Test	p-value
Diet 1 vs Diet 2	P = 0.912
Diet 1 vs Diet 3	P = 0.02
Diet 2 vs Diet 3	P = 0.005

There is no significant difference between Diets 1 and 2 but there is between diet 3 and diet 1 ($p = 0.02$) and diet 2 and diet 3 ($p = 0.005$).

The mean weight lost on Diets 1 (3.3kg) and 2 (3kg) are less than the mean weight lost on diet 3 (5.15kg).

Assumptions for ANOVA

Assumption	How to check	What to do if assumption not met
Normality: The residuals (difference between observed and expected values) should be normally distributed	Histograms/ QQ plots/ normality tests of residuals	Do a Kruskall-Wallis test which is non-parametric (does not assume normality)
Homogeneity of variance (each group should have a similar standard deviation)	Levene's test	Welch test instead of ANOVA and Games-Howell for post hoc or Kruskall-Wallis

Ex: Can equal variances be assumed?

- Null:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: WeightLOST

F	df1	df2	Sig.
.659	2	75	.520

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diet

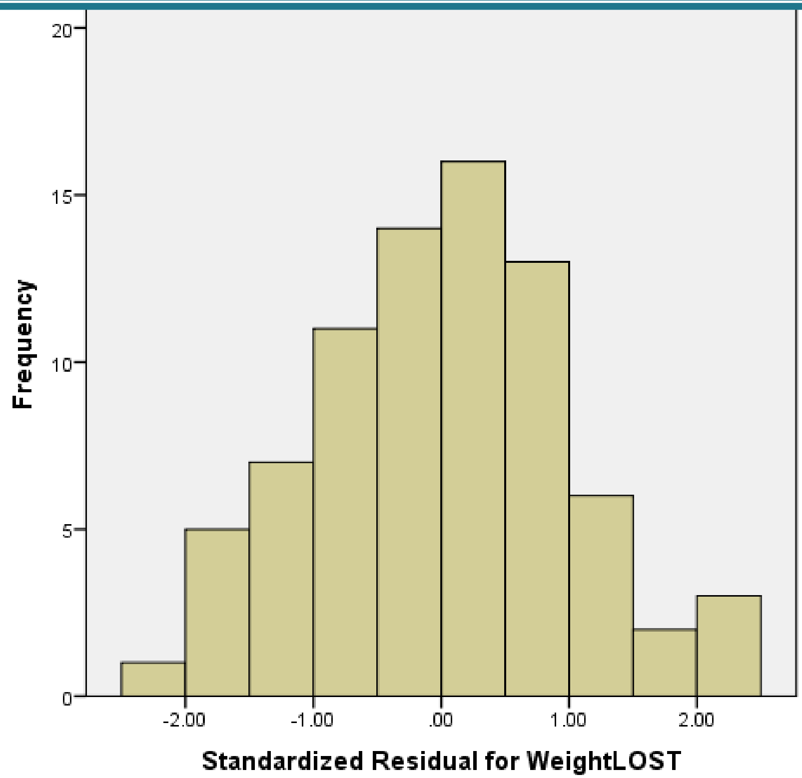
p =

Reject/ do not reject

- Conclusion:

Exercise: Can normality be assumed?

Histogram of standardised residuals



Can normality be assumed?

Should you:

a) Use ANOVA

b) Use Kruskall-Wallis

Ex: Can equal variances be assumed?

- Null:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: WeightLOST

F	df1	df2	Sig.
.659	2	75	.520

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diet

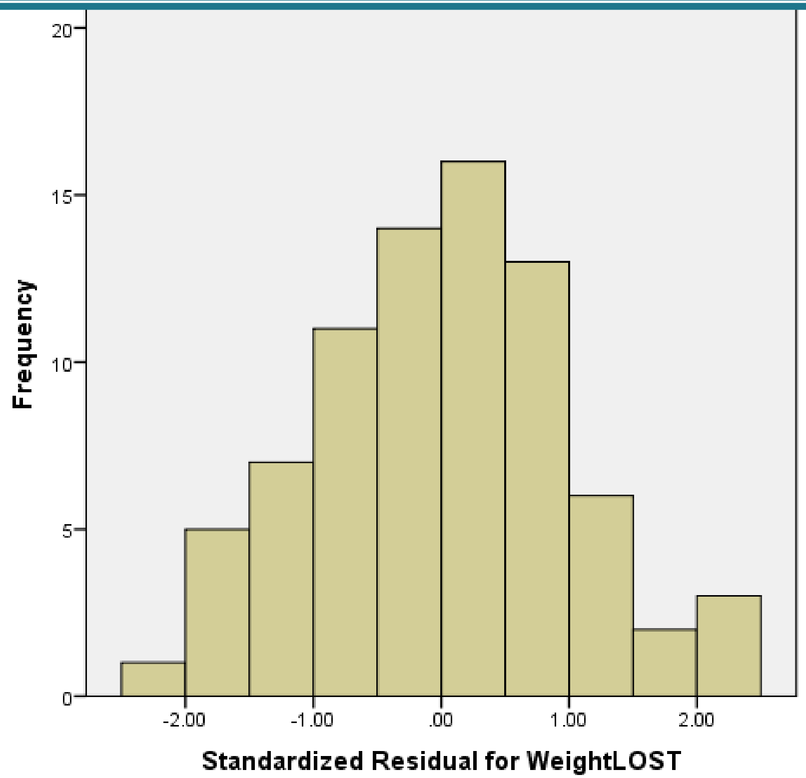
$p = 0.52$

Do not reject

- Conclusion: Equality of variances can be assumed

Ex: Can normality be assumed?

Histogram of standardised residuals



Can normality be assumed?
Yes

Use ANOVA

ANOVA

- Two-way ANOVA has 2 categorical independent between groups variables

e.g. Look at the effect of gender on weight lost as well as which diet they were on

Between groups factor

Between groups factor

	WeightLOST	Diet	gender
16	1.1	1	Male
17	1.5	1	Female
18	1.7	2	Female
19	1.7	2	Male
20	1.9	1	Female
21	2.0	2	Female
22	2.0	2	Female
23	2.0	1	Female
24	2.0	1	Female

Two-way ANOVA

- Dependent = Weight Lost
- Independents: Diet and Gender

- Tests 3 hypotheses:

1. Mean weight loss does not differ by diet
2. Mean weight loss does not differ by gender
3. There is no interaction between diet and gender

What's an interaction?

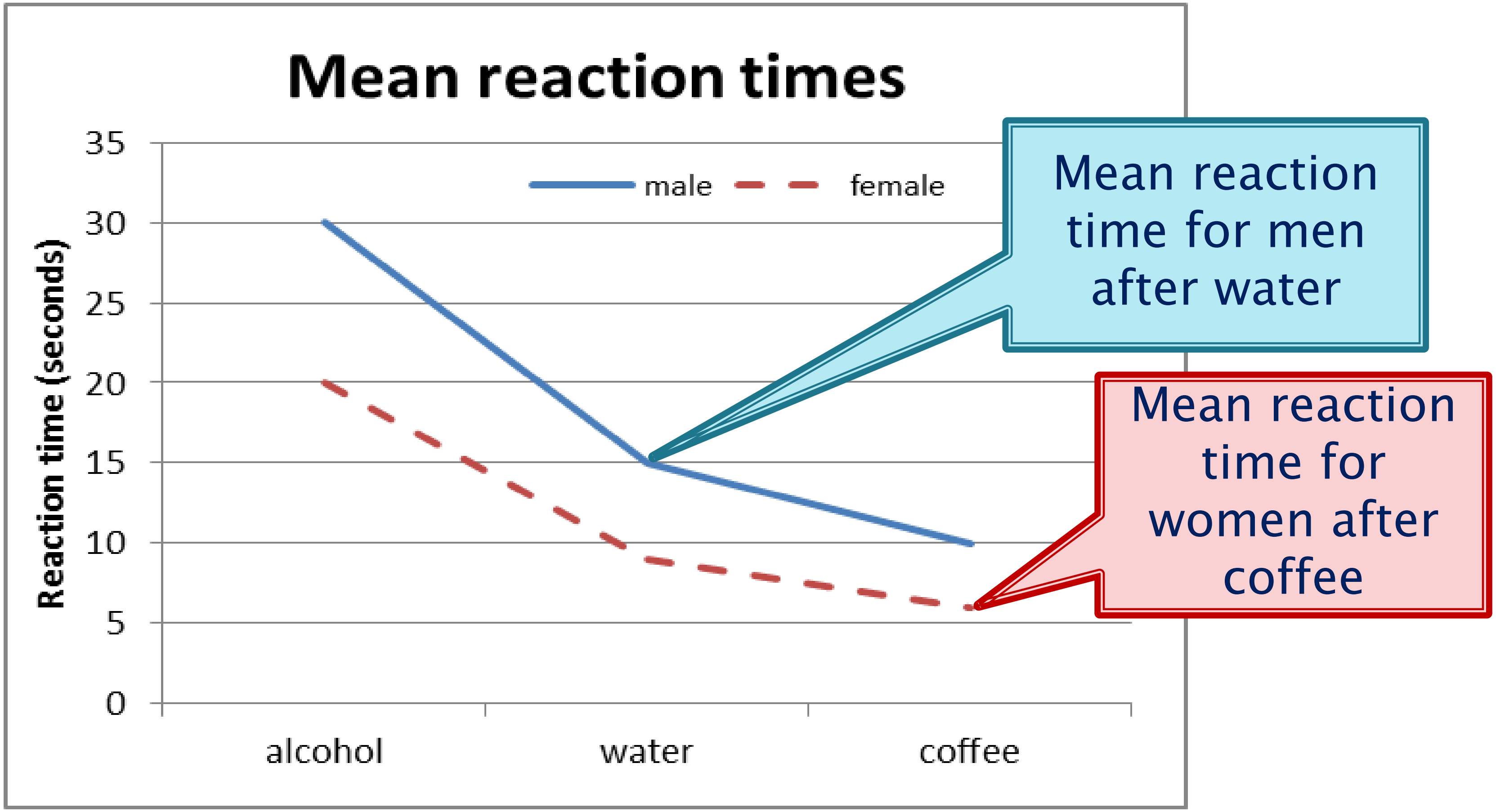
Means plot

- Mean reaction times after consuming coffee, water or beer were taken and the results by drink or gender were compared.

Mean Reaction times	male	female
alcohol	30	20
water	15	9
coffee	10	6

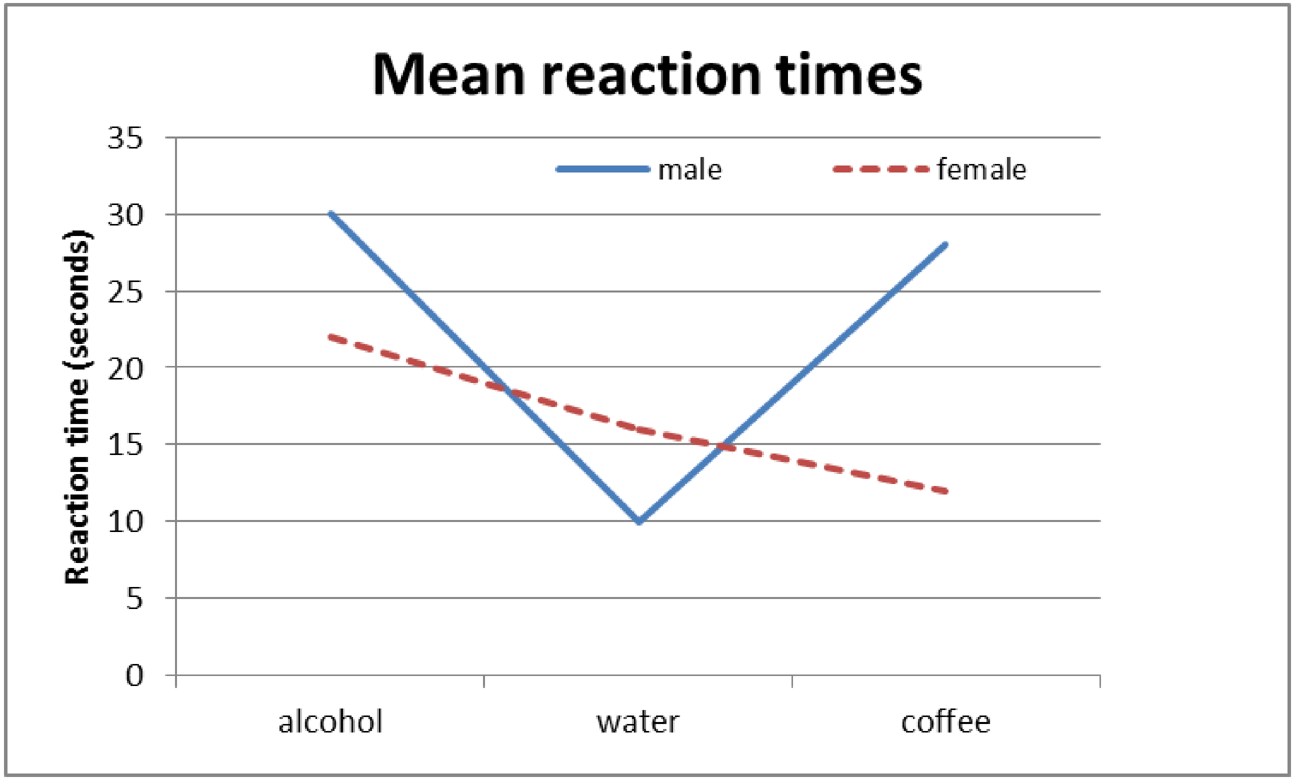
Means/ line/ interaction plot

- No interaction between gender and drink



Means plot

- Interaction between gender and drink



ANOVA

- Mixed between-within ANOVA includes some repeated measures and some between group variables

e.g. give some people margarine B instead of A and look at the change in cholesterol over time

Repeated measures

Between groups factor

ID	Cholesterol Before	Cholesterol after 4 weeks	Cholesterol after 8 weeks	Margarine
1	6.42	5.83	5.75	A
2	6.76	6.2	6.13	A
3	6.56	5.83	5.71	A
4	4.8	4.27	4.15	B
5	8.43	7.71	7.67	A
6	7.49	7.12	7.05	B
7	8.05	7.25	7.1	A
8	5.05	4.63	4.67	A

Related resources

Resources available on http://www.statstutor.ac.uk/

*Paper scenario :*Emmissions_scenario_Stcp-marshallowen-5b

*Video scenario: :*Mass_Customisation_Video_Stcp-marshallowen-1a

Top tips for stats support videos:

Careful_with_the_maths_Video_ Stcp-marshallowen-3a

Conjoint_Analysis_Video_Stcp-marshallowen-4a

Reference booklet: Tutors_quick_guide_to_statistics_7

SPSS training: Workbook: tutor_training_SPSS_workbook_6a, Solutions_for_tutor_training_SPSS_workbook_6b, Data_for_tutor_training_SPSS_workbook_6c