



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

200
1822 - 2022

Creativity and Value Creation Creadigrwydd a Chreu Gwerth

Level / Lefel 5

Research Methods / Dulliau Ymchwil

Kate Piper

*Man Geni Addysg Uwch yng Nghymru
Where Higher Education Began in Wales*

www.uwtsd.ac.uk

Introduction and Experiences

Analysis

- First step on understanding the data.
- Vital to give your research meaning
 - Trustworthiness/ Reliable
 - Quantitative or Qualitative
- Presented as results or findings.
- Step before discussion – why are these results/ findings important?



Quantitative Analysis

- Statistical Analysis
- Statistical package for Social Sciences (SPSS)
- Patterns
- Difference and relationships
- Significance



Significance

Statistical analysis relates to the probability that an outcome occurs and not to absolutes.

In the social science significance is normally set at $P < .05$ and $p < .01$

This means that we are confident that there is a relationship or difference in 95 out of a 100 (95%) cases, if $P < .01$ then 99% of the time.

If the Statistical values is greater than .05 then we say there is no relationship or differences in the data.



Types of Statistics

Descriptive

Describes the information in terms of raw data, means standard deviations etc.

No inference of what the data means is made.

Often used at the start of the results section or results sub-section to provide a summary of findings.

Inferential

Makes inferences about difference and or associations within and between data.

Compares between variables and groups.





Understanding quantitative data

Deall data meintiol

*Man Geni Addysg
Uwch yng Nghymru*

*Where Higher Education
Began in Wales*

Nominal

Categorical data

e.g.

- Gender
- Ethnicity
- Socio-economic group
- Age range
- Martial status

You are either in the group or not

Ordinal

Rank Order Data

e.g.

- Finishing place in a race
- Enjoyment of stats module – Likert scale
- Rate of perceived exertion

Although in order the exact difference between each increment is ambiguous

Interval

Data is ordered relative to each other.

e.g.

Temperature

Mean average

Tide heights

Dates

Zero points are arbitrary or meaningless

Ratio

- Data is relative not only to each other but to a defined zero point.

e.g.

Weight of a person

Length of a piece of string

Time to run a mile

The amount of energy in a Mars bar

Zero points are absolute as part of the scale
(TRUE ZERO)

Parametric v Non-parametric

Parametric

- Interval & Ratio data
- Normally distributed data
- Uses Mean averages
- Homogeneous

Greater rigour in the analysis and more confident in the outcomes and development of conclusions

Non- Parametric

- Ordinal & Nominal data.
- Non-normally distributed data or where no assumptions are made about distribution
- Uses Median averages
- Homogeneous and heterogeneous

Less affected by outliers in the data but less robust



Statistical Tests

Parametric

Relationships

- Pearson's moment correlations
- Linear regression

Differences

- T-tests
- ANOVA's

Relationships

Non- Parametric

- Spearman rank order

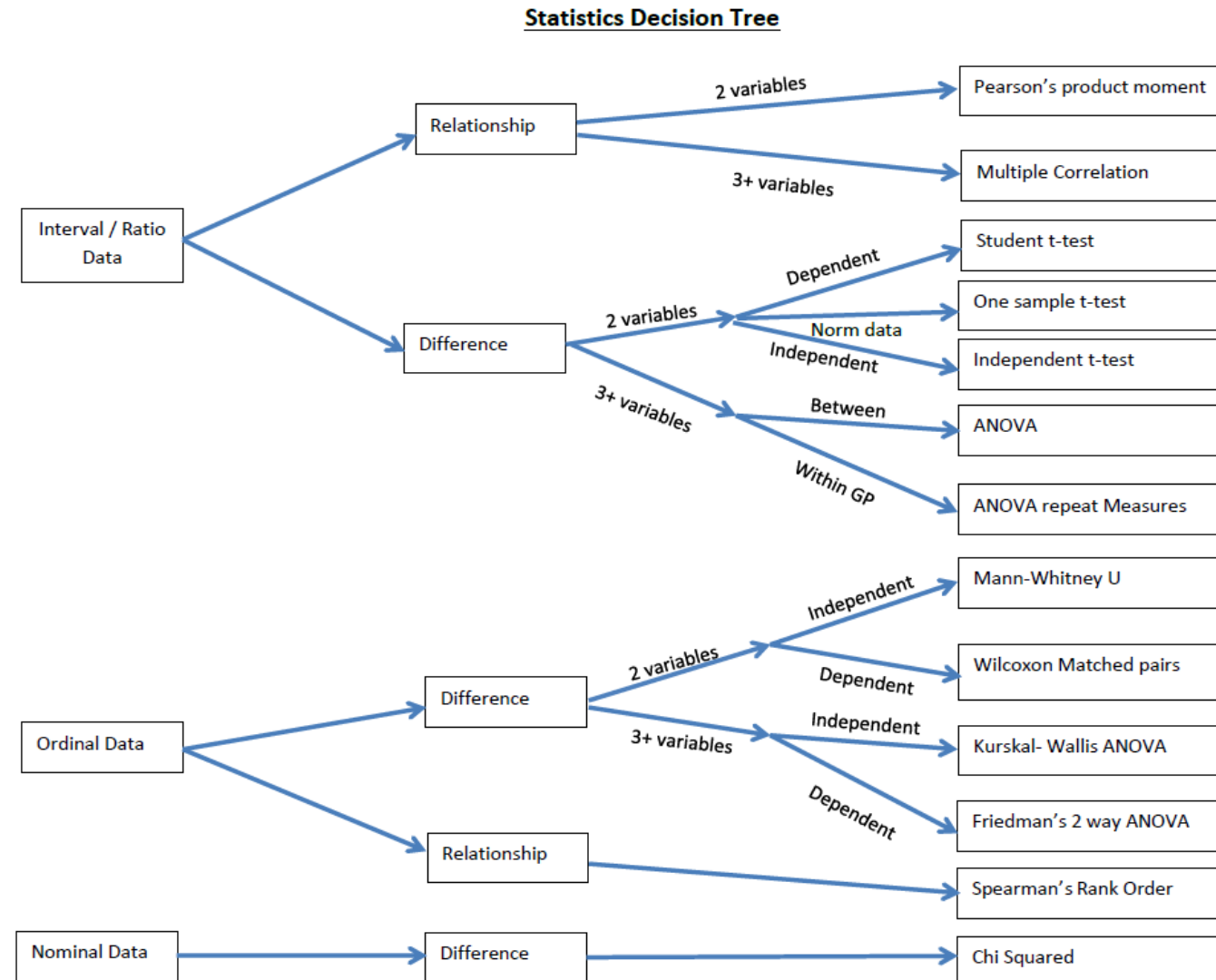
Differences

- Mann-Whitney U-test
- Wilcoxon matched pairs
- Kurskall Wallis ANOVA
- Friedman 2 way ANOVA
- Chi Sqared



Choice of statistical test

1. Type of Data
2. Analysis required
3. Number of variables
4. Relationship of variables



Descriptive Statistics

Raw and calculated data that helps you describe, show or summarise the data you have collected in a meaningful way.

Measures of central tendency:

MEAN, MEDIAN, MODE, CONFIDENCE INTERVAL

Measures of spread or distribution:

RANGE, QUARTILES, DEVIATION, VARIANCE, Z-scores



Mean, Median and Mode

The mean

The **mean** is most commonly used in numeric data collection and is the sum of all values divided by the number of values. In most fitness assessment data you will use the mean data and this will be associated with its standard deviation.

The mode

The **mode** is the most common value provided within a sample and is useful often when analysing survey or questionnaire data.

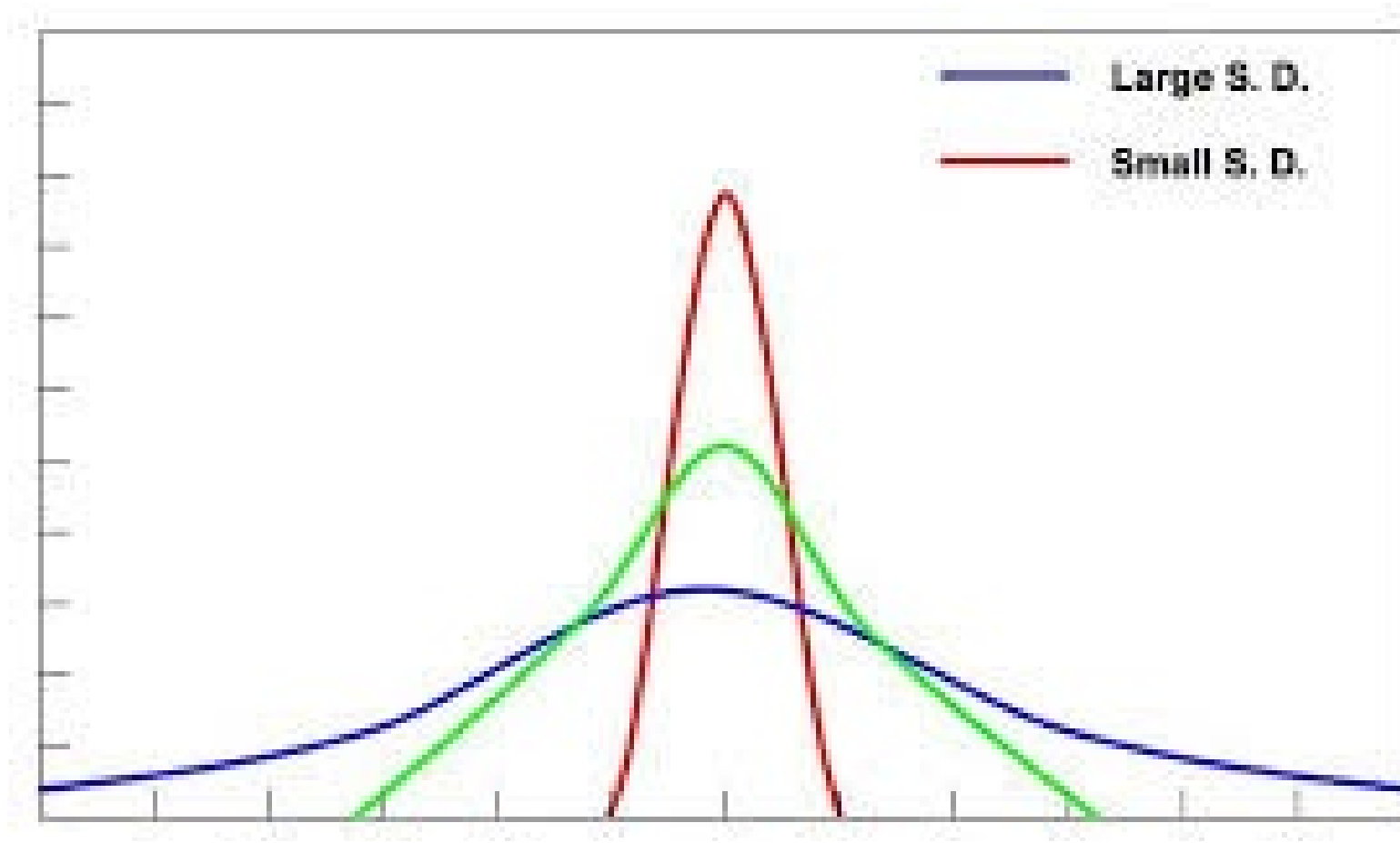
The median

The **median** is the middle point of your data when placed in order and equates to the 50th percentile. It is useful to use the median point when you have lots of outliers or the data is skewed.

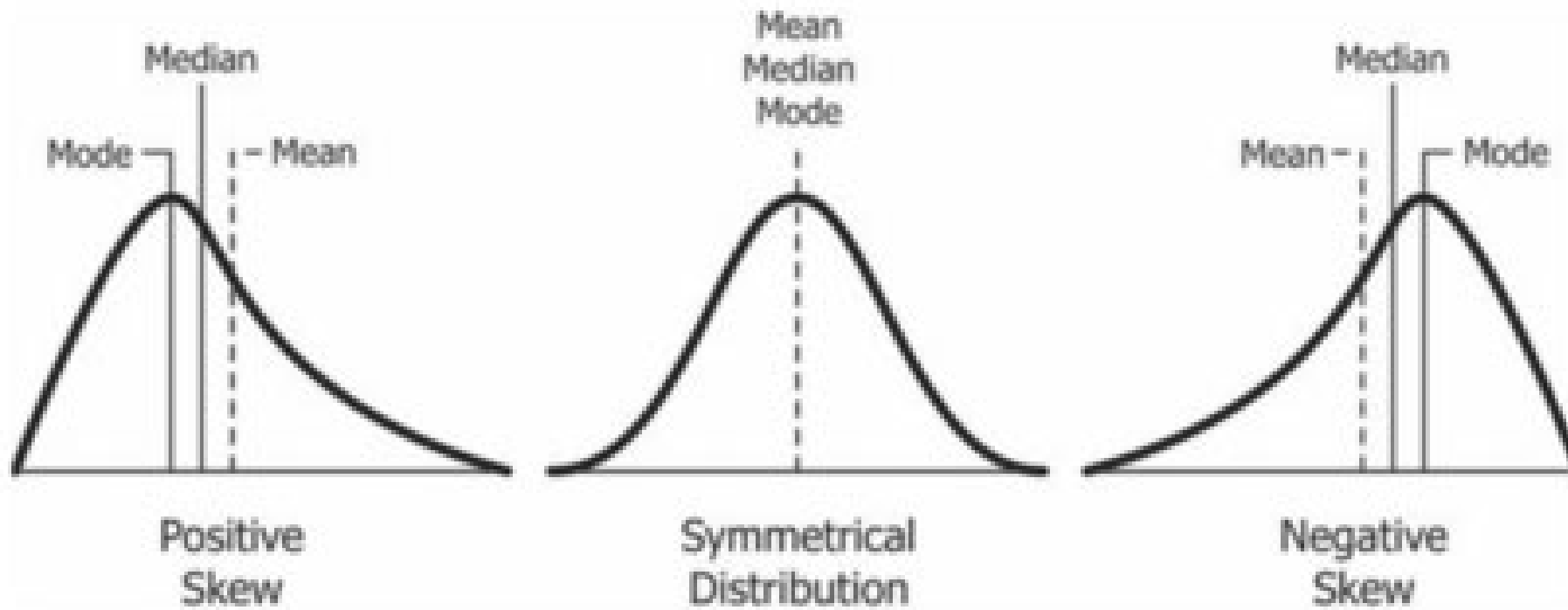


Distribution Normal

Where the distribution of values is symmetrical around the mean.



Skewed Data



Standard Deviation

The standard deviation (SD) is a measure of the spread of data or how much a variable differs from the mean value.

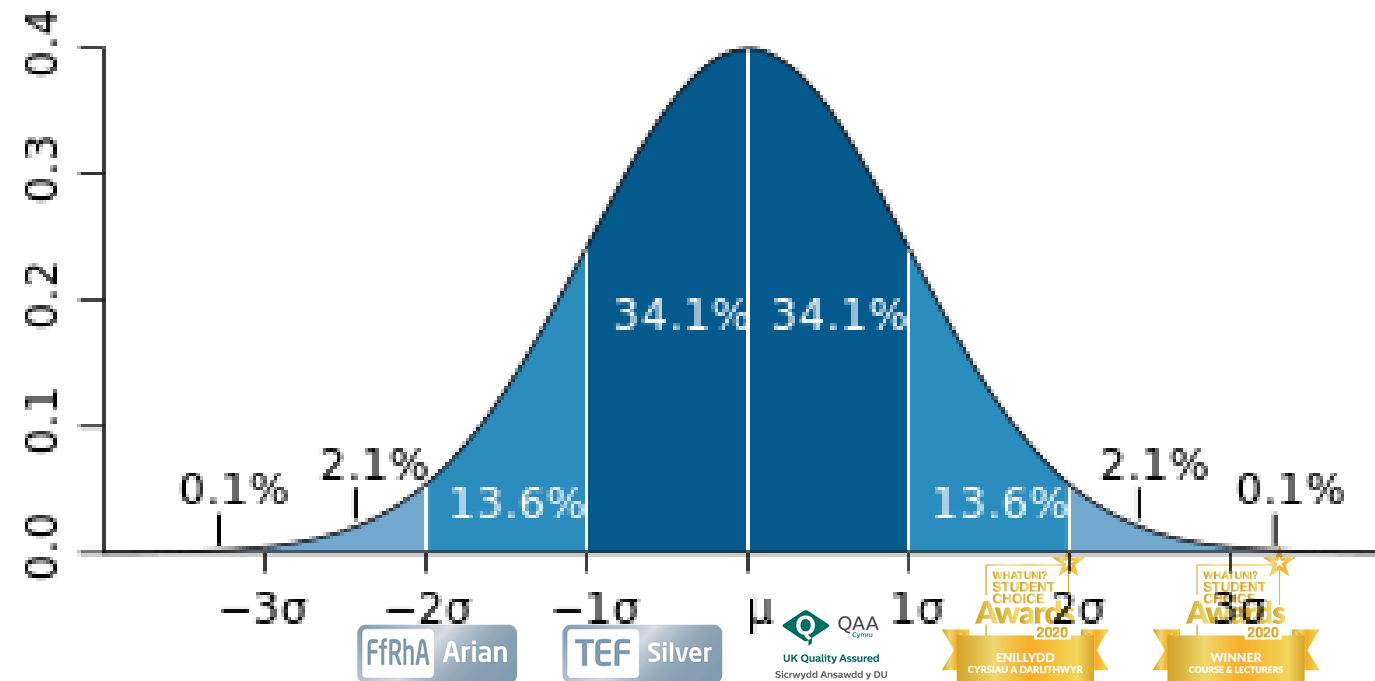
A large deviation means that there is a wide variance in the data whereas a small deviation means all values are quite similar.

SPSS output normally provides the mean value and one standard deviation. This means that the 68.2% of the data lies between the mean score and $\pm$ the SD.

A measure of distribution

mean $\pm$ 1 Standard deviation (68.2% of data points)

mean $\pm$ 2 Standard deviations (95.4% of data points)



Standard Deviation Task

Data

You conduct an evaluation of grip strength on athletes v non-athletes

Athletes	Non-Athletes
45	41
41	23
51	51
46	31
39	34
48	46
46	52
44	37
50	41

Questions

What is the mean and standard deviation for athletes and non athletes?

Calculate the range of the data.

Write a statement describing the data.



Calculating Mean, Standard Deviation & Range

Most SPSS output provide mean and SD

You can just calculate in using descriptives.

1. Input data into SPSS
2. Analyse
3. Descriptives statistics
4. Descriptives
5. Move the variables across
6. Options – choice those you want
7. OK



Confidence Intervals

Calculating the mean of a group only allows you to describe the mean of **that** group.

Calculating the confidence interval from this mean and its standard deviation (normally with a certainty of 95% probability) allows you to predict the mean score within the population as a whole.

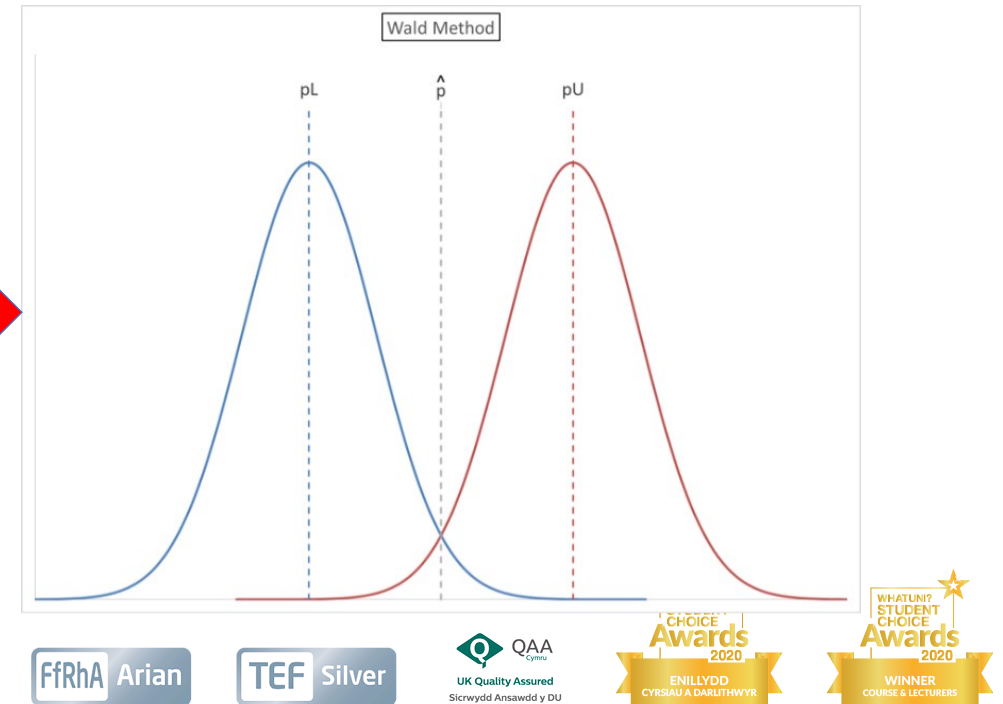
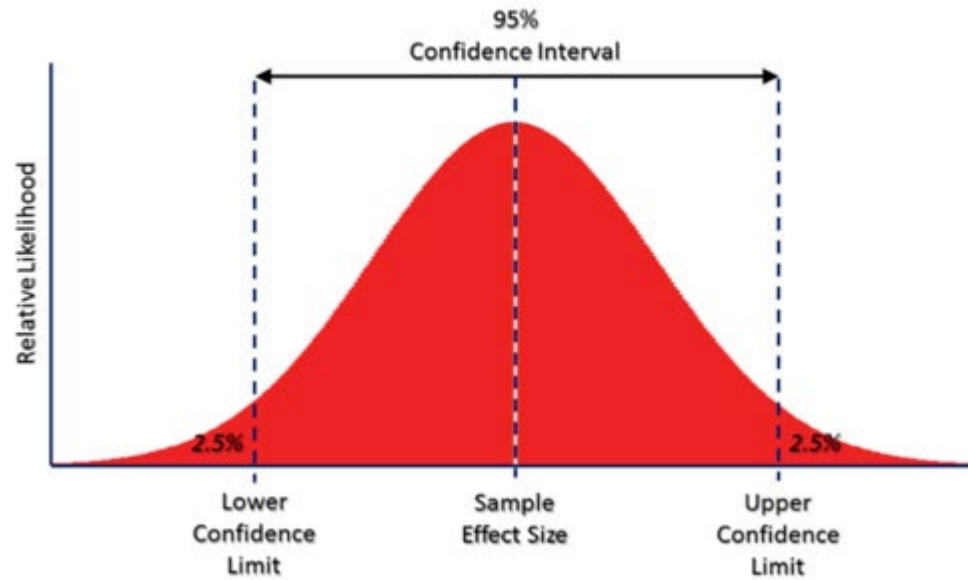
However this calculation only provides a range score that the mean lies between.

i.e. The mean strength of the group was $34.5 \pm 9.5\text{kg}$ (95% CI 29.5, 39.5)

This means that the mean strength of the whole population from which the group mean was derived is between 29.5kg and 39.5kg.



Why Is Confidence Interval Important



Calculating the Confidence Interval

The confidence intervals is a standard output in SPSS but can be calculated.

$$\bar{x} \pm z \frac{s}{\sqrt{n}}$$

Where

X is the mean value

Z is 1.96

S is the standard deviation

N is the number of data points

Using SPSS determine the mean, standard deviation and confidence interval of the variables within data set 1.



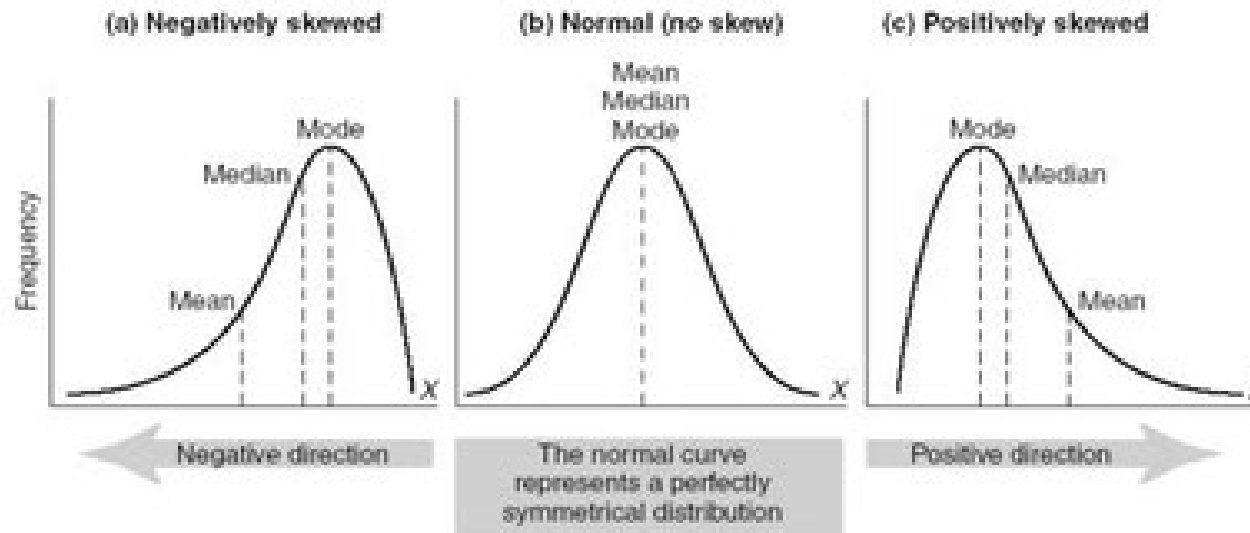
Distribution

Parametric

Statistical use is based on the groups mean.
Has a normal distribution.
Sample size greater than 12.
Outliers can be removed.
Type of data must be scale (interval or ratio)

Nonparametric

Statistical use is based on the groups median.
May not have a normal distribution.
Sample size may be less than 12
Outliers can't be removed
Type of data – nominal or ordinal



Examples of normal and skewed distributions:



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

200
1822 - 2022



Man Geni Addysg Uwch yng Nghymru
ere Higher Education Began in Wales

T-test

- A T- test is used when you want to determine whether two sets of data are the same or different (compare the means).
- However you can only use a t-test when the data is interval or ratio.

1.The effect of a certain diet on body fat

2.The impact of plyometric training on leg power

3.The impact of providing playground equipment on children's activity levels.

4.The impact of coaching on climbing skills development



Types of T-test

- There are several type of t-test but you will only need 2 variations;
- **Independent t-test** – when you compare the means of two distinct groups e.g. males and females, obese and not obese, old and young.
- **A Paired sample (Student) T-test** – when you compare the means of one group that has been measure on 2 occasions e.g body weight before and after a diet, strength before and after a training programme.

You will also be asked if it is one or two tailed. To make life easy always use a two tailed test.



Example data

Do men consume more calories than women?

gender	energy
m	1725
f	1724
m	1442
m	1943
m	1726
m	1198
m	1853
m	2915
f	2420
f	1532
m	3416
f	2419
f	1045
m	2540
m	2800
m	2500
m	2030
f	1900
f	1580
f	2080
f	1600
m	1986
m	2135
m	2172
m	2359
f	2270
f	1734
f	1545
f	2712



Conducting an Independent sample t-test

We may want to determine if there is a difference in calorific consumption between men and women?

Therefore the H_0 is that there no difference in calorie intake between men and women

(1) Set up your variable view information

- In Name type in gender
- In Values type in 1=m; 2=f (this means that any value 1 read under hand refers to a male and any number 2 refers to female– thus we differential between our groups)
- Second row enter strength

(2) Go back to data view

(3) Copy across from the data from excel spread sheet



Conducting the assessment

- Analyse
- Compare means
- Independent t-test

You have two gender groups so move this to the grouping variable box – you then define the groups (in this example, group 1=m and group 2=f because in the data the genders are labeled 1 and 2)

Energy intake is your test variable to move this to the test variable box

Press Ok and you get descriptive stats and the t-test scores



SPSS Output tables

Group Statistics					
	gender	N	Mean	Std. Deviation	Std. Error Mean
energy	m	16	2171.2500	569.36040	142.34010
	f	13	1889.3077	466.39797	129.35552

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
energy	Equal variances assumed	.263	.612	1.435	27	.163	281.94231	196.43997	-121.11921	685.00383
	Equal variances not assumed			1.466	26.993	.154	281.94231	192.33709	-112.70543	676.59005



Reporting the Information

- As the 2 tailed significance is 0.163 i.e greater than 0.05 then the two sets of data are not significantly different.
- So we conclude there is no difference in calorific intake between men and women

The data would be reported;

- A comparison of energy intake identified that men ($n=16$, 2171 ± 569 Kcals) consumed more calories than women ($n= 13$, 1889 ± 466 Kcals) though this was not statistically different ($t(27)=1.435$, $p=0.163$)
- Where $t(\text{degrees of freedom}) = \text{score}$, $p = \text{value}$.

Outcome

- The data therefore accepts the null hypothesis and concludes there is no difference in calorific consumption between men and women.



Dependent T-test

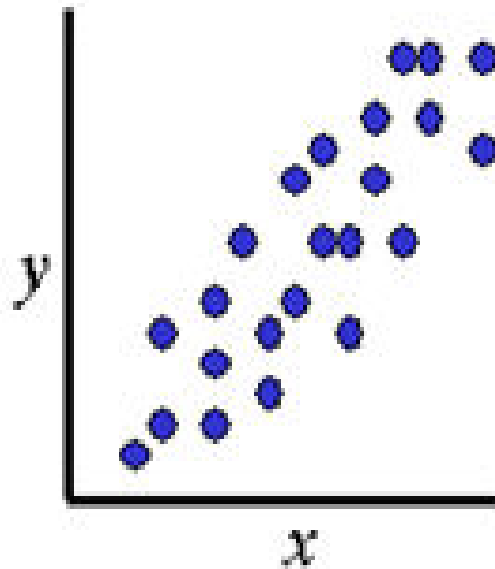
To compare two set of measures when one population is measured on two separate occasions (the type of data must be exactly the same).

- **Strength before and after training**
- **Mountain biking speed after coaching**
- **VO2max tested using bleep test and Metamax**

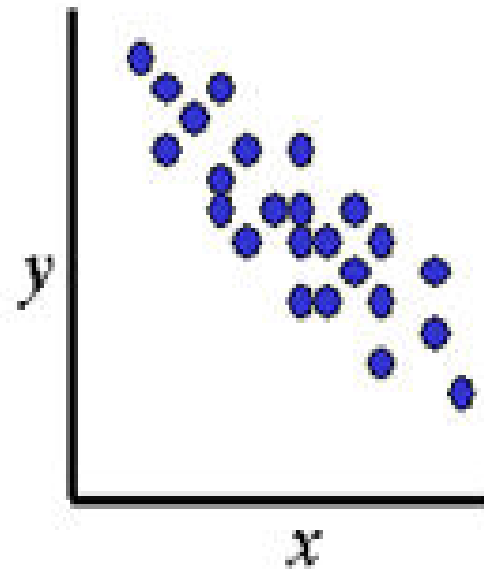


Correlations

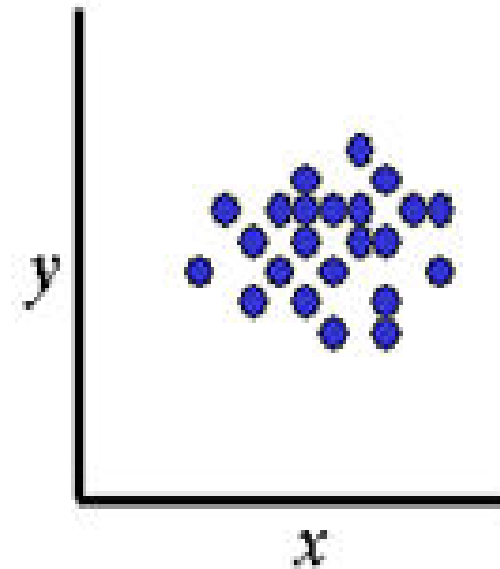
Correlations determine the relationship between variables



Positive correlation



Negative correlation



No correlation



Use of Correlations

To determine how one variable changes in relationship to another variable.

HOWEVER

The relationship may not be causal!



Calculating a Correlation - r

If the data is interval then use **Pearson's moment correlation**.

If the data is ordinal then use **spearman rank order correlation**



Examples of Correlation

Parametric

Pearson's Moment Correlation

- BMI to percentage body fat
- Leg strength and jump height
- Calorific intake and weight
- Child age and coordination
- Arm length and climbing speed
- Training intensity and injury

Non Parametric

Spearman's Rank Order

- Order a person scores compared to other participants (not normally distributed).
- Likert Scale Questionnaires
- Any categorical data comparisons i.e.
 - levels of body mass
 - Age range



Is there a relationship between grip strength and arm girth?

Data

Grip St	Forearm
45	24
43	23
31	23
43	25
26	21
15	20
60	30
34	22
41	26
36	23
51	27
53	29
29	20

Arian



Conducting a Bivariate Correlation in SPSS

Enter variable view

Under name

Row 1 - grip

Row 2 - forearm

In measure for each select scale data.

Enter data view



You'll see grip and forearm now as column headings.

Copy data across from spread sheet.

Click

1.Analyse

2.correlate

3.Bivariate

4.Pearsons (as data is scale / interval) and tick 2 tailed and flag significance

5.In options tick means and SD's- continue

6.Move the variables across into variables box

Press ok



SPSS output

Descriptive Statistics			
	Mean	Std. Deviation	N
grip	39.0000	12.23383	13
forearm	24.0769	3.20056	13

Correlations

grip	grip	forearm		
	Pearson Correlation	1		.919**
forearm	Sig. (2-tailed)			.000
	N	13	13	
forearm	Pearson Correlation	.919**	1	
	Sig. (2-tailed)	.000		
	N	13	13	

** . Correlation is significant at the 0.01 level (2-tailed).

The values required are;

the correlation (r-value) = 0.919

Significance (p) $p < 0.000$

** indicates this is significant at the $p < 0.1$ level i.e the two values are correlated



Probability values(p)

Statistical analysis works on probabilities that outcomes are related or not.

In social sciences these are determined to be;

$$P < 0.05 \quad \& \quad P < 0.01$$

This means that we are confident 95 / 99 time out of a hundred that the scores are related.



Addition use of the r values

The correlation, r score provides a indication of relationship.

r^2 is and indication of variance.

If the r score is 0.919 then r^2 is 0.84.

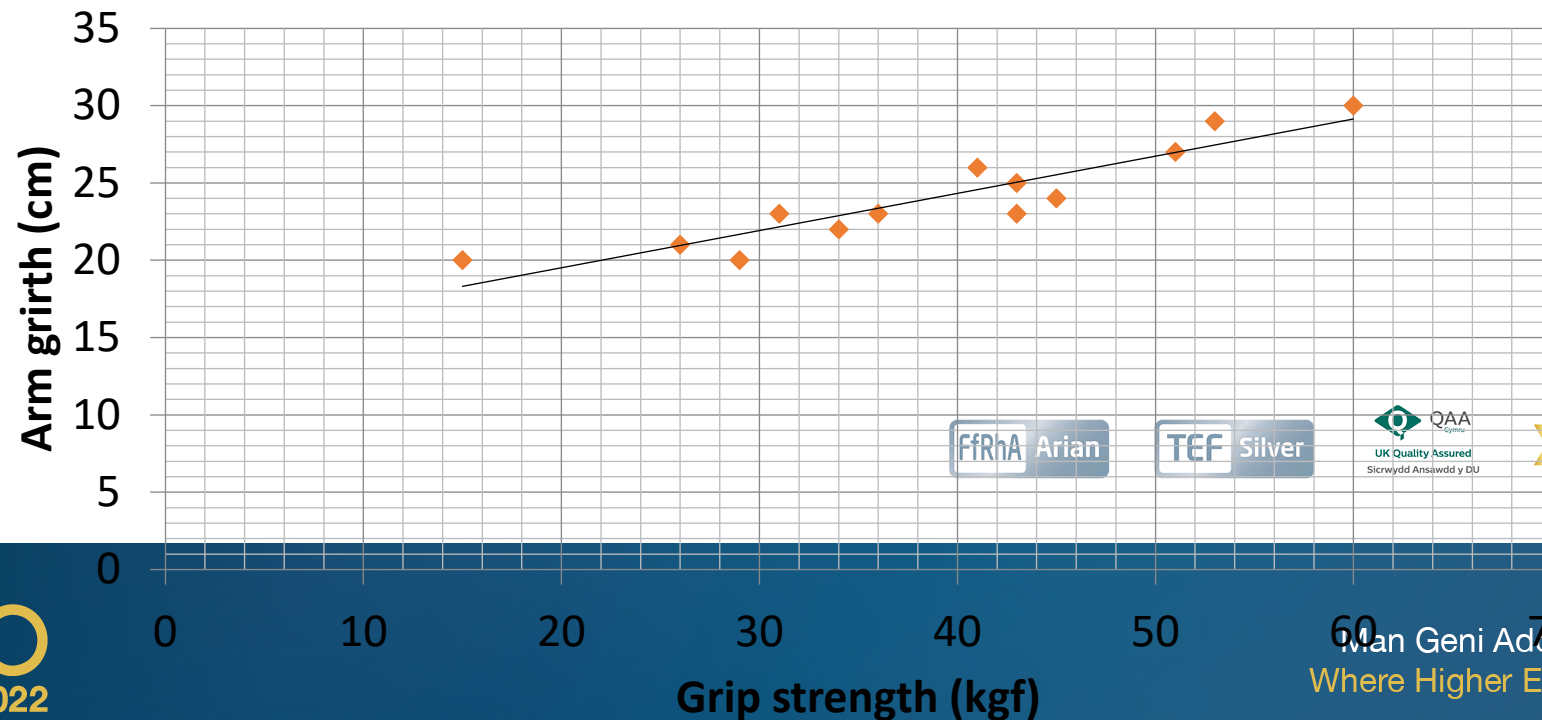
This means that 84% of the variability between the two variable is accounted for and 16% is something else.



Plotting a correlation

- Often a very good way to see and interpret the data

Graph 1 Relationship between grip strength and arm girth



Effect Size

The p-value of a relationship only infers whether two variable change relative to each other, or not and not how strong the relationship is.

The effect size quantifies the strength of the relationship.

In a Pearson's moment correlation the effect size is calculated as r^2 and ranges between -1 and +1.

So, in the example above the r value is .919 and r^2 is .844. This mean that 84.4% of the variable in grip strength data can be accounted for by measuring of forearm girth.

Strength of Association	Coefficient, r	
	Positive	Negative
Small	.1 to .3	-0.1 to -0.3
Medium	.3 to .5	-0.3 to -0.5
Large	.5 to 1.0	-0.5 to -1.0



Predictive correlations

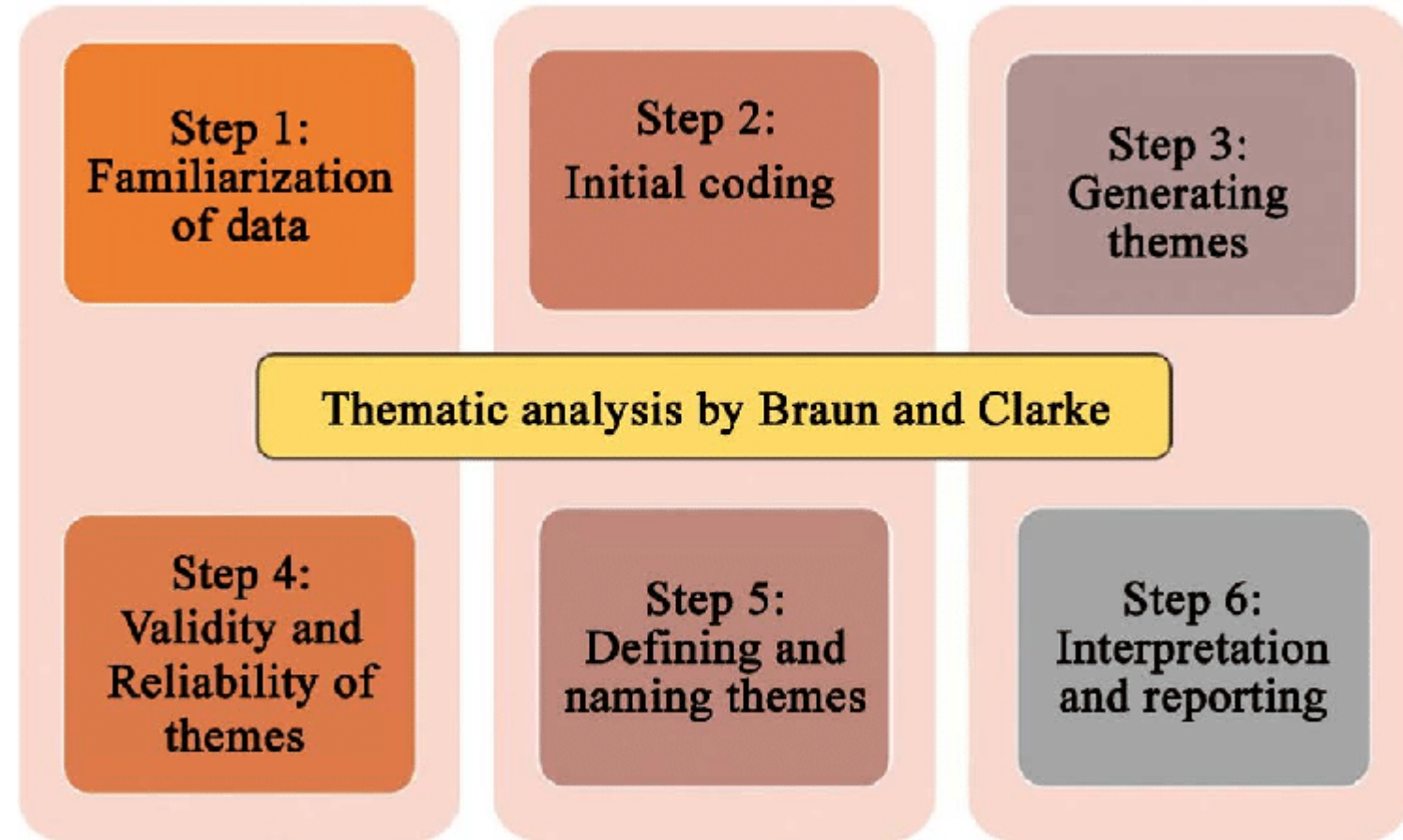
- If the correlation and r^2 is high then the use of **regression analysis** may be useful;

i.e. the ability to predict one score from the other – regression equation.



Qualitative Analysis

- Many ways
- Numerical Analysis
 - Word Frequency
 - Numerical Observations
- Thematic Analysis
- Braun and Clarke



Upcoming weeks

Date	Information
14 th Nov (wk7)	Methodology for A1
21 st Nov (wk8)	Methods for A1
28 th Nov (wk9)	Tutorials
5 th Dec (wk10)	A2 Focus
12 th Dec (wk11)	A2 Focus





Task for next week:

Bring everything created for your assessment.

Introduction/Rationale for your research

- Why is it a hot topic – what is important about it?
- What do you already know about the topic from research (how do you know it?)
- What are the gaps / what is interesting about what you know?
- What research needs to be done?



Research Questions and Aim

Methodology

- What approach/ methodology are you using? Why?
- What methods are you using?
- Why?
- How?
- How would you intend the data to be analysed [if the study was actually being done]?

Man Geni Addysg Uwch yng Nghymru
Where Higher Education Began in Wales

www.uwtsd.ac.uk



Research question

We are going to determine whether the measure of body fat by sum of skinfolds (siri) provides the same information as body fat measured by impedance?

H_0 there is no difference in body fat measured by impedance compared to using the siri equation method



Setting up SPSS

- (1) Set up your variable view information
 1. skinfold
 2. Imped
- (2) Go back to data view
- (3) Copy across from the excel spread sheet



Conducting the Analysis

- (5) Analyse
- (6) Compare means
- (7) Paired samples t-test
- (8) Move over skinfold to variable 1 and impedance to variable 2.
- (9) Press Ok and you get descriptive stats and the t-test scores.



SPSS Outputs

- We get 3 outputs now;
- Descriptive statistics
- Correlations
- T-test



SPSS Output Tables

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	skinfold	28.6037	27	5.91208	1.13778
	imped	29.7926	27	6.02679	1.15986

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	skinfold & imped	27	.992	.000

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	skinfold - impeded	-1.18889	.77872	.14987	-1.49694	-.88084	-7.933	26	.000



Reporting the Data

In a comparison of methods used to determine percentage body fat it was found that the skinfold method ($28.4 \pm 6\%$) resulted in significantly lower values ($t(20) = -6.728$, $p < 0.01$) than the impedance method ($29.7 \pm 6\%$).

Outcome

The null hypothesis is rejected and it is concluded that the skinfold method and impedance method provide different measures of body composition.



However it should also be notes that the two sets of data are highly correlated!