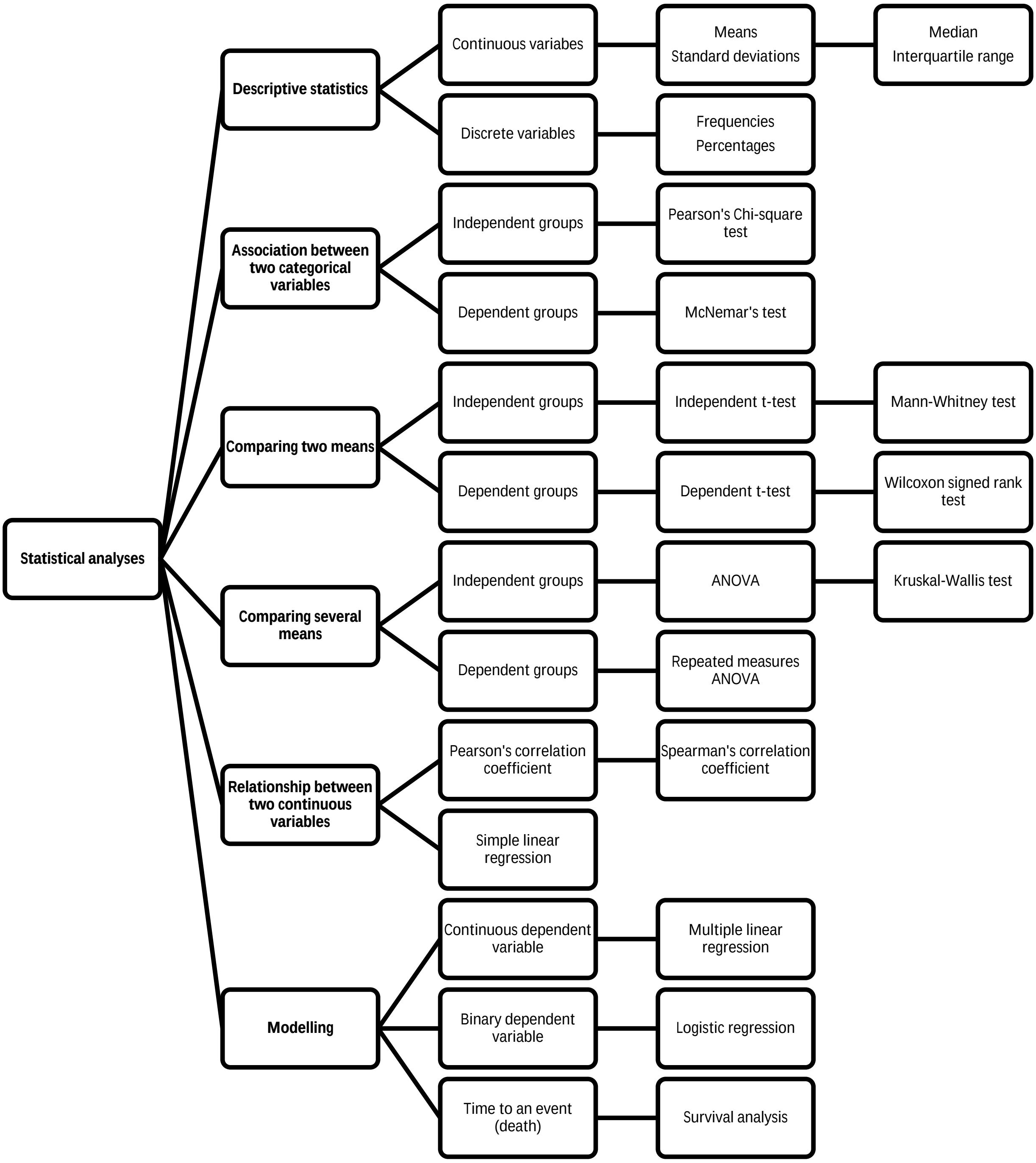
Introduction to statistical techniques applied to drug utilization research



Introduction to statistical techniques applied to drug utilization research

Marike Cockeran

MURIA symposium

25 July 2016

It all starts here

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

Introduction

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

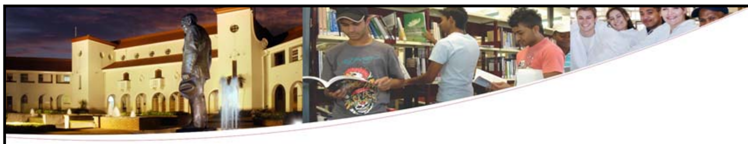
- Every quantitative study yields a set of data.
- A complete set of data will not necessarily provide a researcher with information that can easily be interpreted.
- Between the raw data and the reported results of the study lies some intelligent and imaginative manipulation of the numbers, carried out using statistics.
- Statistics explores the
  - collection
  - organisation
  - analysis
  - and interpretation of numerical data

Introduction

- Parameter
  - A numerical measurement describing some characteristic of a population.
- Statistics
  - A numerical measurement describing some characteristic of a sample.
- One of the fundamental goals of statistics is to describe some characteristics of a population using the information contained in a sample of observations.

Outline of the session

- Measurement scales
- Descriptive statistics
- Basic terminology used in inferential statistics
- Comparing two means
  - Independent t-test
  - Dependent t-test
- Comparing several means
  - One-way ANOVA
- Determining the relationship between two continuous variables
  - Pearson's correlation coefficient



Measurement scales

Marike Cockeran
MURIA symposium
25 July 2016

It all starts here

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

Measurement scales: Introduction

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

- Measurement scales are used to classify variables or types of data:
  - Nominal scale
  - Ordinal scale
  - Interval scale
  - Ratio scale
- It is important to know on what type of scale a variable is measured since certain statistical analysis is only applicable on variables measured on a certain measurement scale.
- Example:
  - It does not make sense to calculate the average gender of a sample.
  - However you can calculate the average weight or height of a sample.

Measurement scales: Nominal scale

- Values (numbers) are assigned to different categories of a variable.
- Example:
  - The variable gender has two categories: male and female.
  - The number one is assigned to the male category.
  - The number two is assigned to the female category.
- The sequence of the values is not important.
- The numbers serve as labels for the different categories.
- The categories do not overlap.
- If a variable is measured on a nominal scale and takes on one of two distinct values, the variable is called a dichotomous or binary variable.

Measurement scales: Ordinal scale

- Values (numbers) are assigned to different categories of a variable, but categories now have an ordered relationship to one another.
- Example: Variable measuring physical activity
  1 – low physical activity
  2 – medium physical activity
  3 – high physical activity
- The categories do not overlap.

Measurement scales: Interval scale

- Each value on the scale has a unique meaning.
- Values have an ordered relationship to one another.
- Scale units along the scale are equal to one another.
- The scale does not have a true zero point.
  - Zero does not represent the absolute lowest value.
  - It is a point on the scale with numbers both above and below it.
- Example: Temperature

Measurement scales: Ratio scale

- Each value on the scale has a unique meaning.
- Values have an ordered relationship to one another.
- Scale units along the scale are equal to one another.
- The scale has a true zero point.
  - Zero presents the absolute lowest value.
- Example: Weight

Measurement scales: Additional remarks

- Discrete data
  - Data are restricted to taking on only specified values – often integers.
  - Fractional values are not possible.
  - Example: The number of new cases of tuberculosis reported.

- Continuous data
  - Data are not restricted to taking on certain specified values.
  - Fractional values are possible.
  - Example: Serum cholesterol level of a patient.

Descriptive statistics

Marike Cockeran
MURIA symposium
25 July 2016

It all starts here

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

Descriptive statistics: Outline

- Frequencies and percentages
- Measures of location
  - Arithmetic mean
  - Median
- Measures of spread
  - Range
  - Interquartile range
  - Variance and standard deviation
- Possible distributions of data values
- Box-and-Whiskers plot
- Histogram

Frequencies and percentages

Outcome	Number (%) of patients with outcome	
	Ciprofloxacin n=60	Pivmecillinam n=60
Clinical success	48 (80%)	39 (65%)
Bacteriological success	60 (100%)	54 (90%)

Arithmetic mean

- The mean is calculated by summing all the observations in a set of data and dividing by the total number of measurements.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Properties of the mean:
  - The mean takes into consideration the magnitude of every observation in a set of data.
  - This causes the mean to be extremely sensitive to unusual values.

Arithmetic mean: Example

- HDL cholesterol values of 7 women:

Dataset 1	Dataset 2
1.30	1.30
1.38	1.38
1.42	1.42
1.58	1.58
1.61	1.61
1.45	1.45
1.57	5.17

- Average value of Dataset 1: $\bar{X} = 1.47$
- Average value of Dataset 2: $\bar{X} = 1.99$

Median

- If a list of observations is ranked from smallest to largest, half the values are greater than or equal to the median, whereas the other half are less than or equal to it.

- The median is the middle value of an ordered dataset.

3 5 21 25 31 33 36 40 42 43 45

- Properties of the median:
  - The median takes into consideration only the ordering and relative magnitude of observations.
  - The median is less sensitive to unusual data points.

Median: Example

- HDL cholesterol values of 7 women:

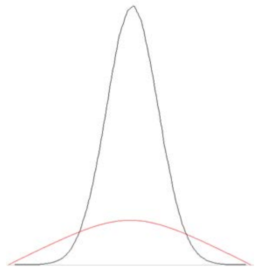
Dataset 1	Dataset 2
1.30	1.30
1.38	1.38
1.42	1.42
1.45	1.45
1.57	1.58
1.58	1.61
1.61	5.17

- Average value for Dataset 1: $\bar{X} = 1.47$
- Median value for Dataset 1: $m = 1.45$

- Average value for Dataset 2: $\bar{X} = 1.99$
- Median value for Dataset 2: $m = 1.45$

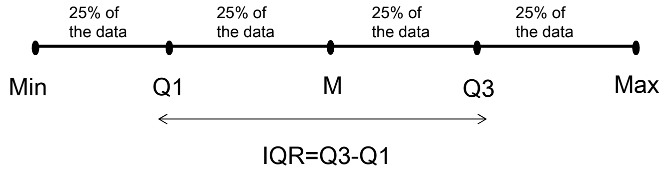
Measures of spread

- To know how good our measure of central tendency is, we need to have some idea about the variation among the measurements.
- Do all the observations tend to be quite similar and therefore lie close to the center, or are they spread out across a broad range of values?



Measures of spread: Range and IQR

- The range is defined as the difference between the largest observation and the smallest observation.
- The usefulness of the range as a measure of spread is limited since it considers only the extreme values rather than the majority of the observations.
- The interquartile range is calculated by subtracting the 25th percentile of the data from the 75th percentile.
- The interquartile range includes the middle 50% of the observations.



Measures of spread: Variance

- The variance quantifies the amount of variability, or spread, around the mean of the measurements.
- The variance of a set of observations is defined as:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Measures of spread: Standard deviation

- The standard deviation of a set of observations is the square root of the variance.

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

- The standard deviation has the same units of measurement as the mean, rather than squared units.
- In a comparison of two groups of data, the group with the smaller standard deviation has the more homogeneous observations and the group with the larger standard deviation exhibits a greater amount of variability.
- The magnitude of the standard deviation depends on the values in the dataset – what is large for one group of data may be small for another.
- Since the standard deviation has units of measurement, it is meaningless to compare standard deviations for two unrelated quantities.

Measures of spread: Example

Patient	Heart rate (beats per minute)
1	167
2	150
3	125
4	120
5	150
6	150
7	40
8	136
9	120
10	150

Example 1:

$\bar{X} = 130.8$

$m = 143$

$sd = 35.47$

Example 2:

$\bar{X} = 140.89$

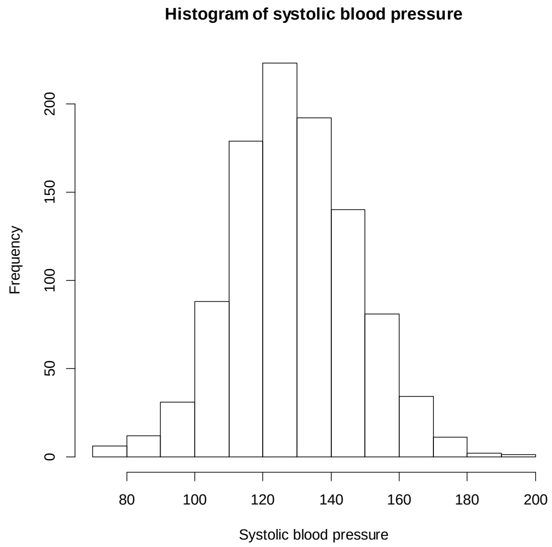
$m = 150$

$sd = 16.44$

Histogram

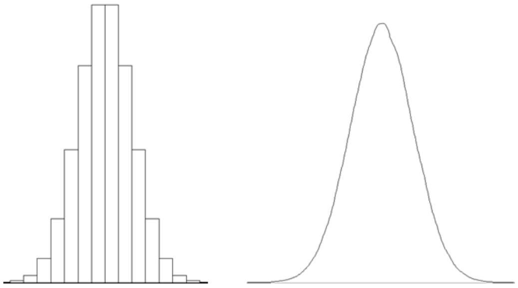
- The horisontal axis shows possible intervals for the values of the variable.
- The vertical axis shows either the frequency or the relative frequency of observations within each interval.
- This is also known as a frequency distribution.

Histogram: Example



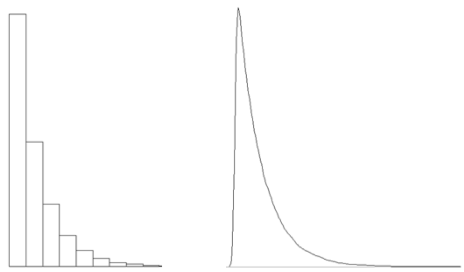
Symmetric and unimodel distributions

- The best measure of central tendency depends on the way in which the values are distributed.
- If they are symmetric and unimodel then the mean, median and the mode should all be roughly the same.



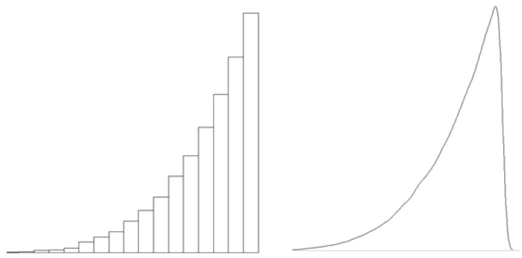
Right skewed distributions

- If the data are skewed to the right, the mean lies to the right of the median.
- When the data are not symmetric, the median is often the best measure of central tendency.
- Since the mean is sensitive to extreme observations, it is pulled in the direction of the outlying data.



Left skewed distributions

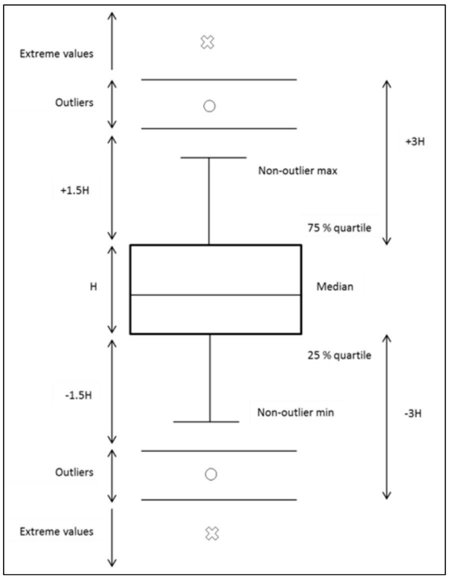
- If the data are skewed to the left, the mean lies to the left of the median.
- When the data are not symmetric, the median is often the best measure of central tendency.
- Since the mean is sensitive to extreme observations, it is pulled in the direction of the outlying data.



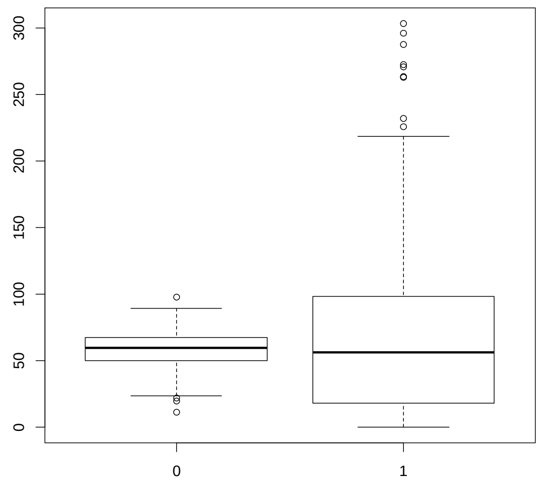
Box-and-Whiskers plot

- The box-and-whiskers plot is a graphical representation of the numerical summary measures calculated in the previous section.
- The box plot shows the centre, spread and skewness of a dataset.
- The median (50th percentile) is indicated by a vertical line, within the box.
- The lower and upper quartiles (25th and 75th percentiles) are indicated by the corresponding vertical ends of the box.
- The box thereby encloses the interquartile range, sometimes referred to as the *H* spread.
- The minimum and maximum non-outlier values are indicated by the whiskers.
- An outlier is beyond the whisker but less than three interquartile ranges from the box edge and finally an extreme value is more than three interquartile ranges from the box edge.
- It should be noted that the preceding explanation is merely one way to define a Box-and-Whisker plot.

Box-and-Whiskers plot



Box-and-Whiskers plot: Example



Basic terminology

Marike Cockeran
MURIA symposium
25 July 2016

It all starts here

Types of variables

- Independent variable
  - Thought to be cause of effect – experimental research – has been manipulated.
- Dependent variable
  - Thought to be affected by independent – outcome.
- Predictor variable
  - Independent variable
  - Thought to predict outcome variable.
- Outcome variable
  - Dependent variable
  - Thought to change due to changes in predictor.

Hypothesis testing

- Null hypothesis
  - Predicted effect does not exist
- Alternative hypothesis
  - Predicted effect does exist

P-values

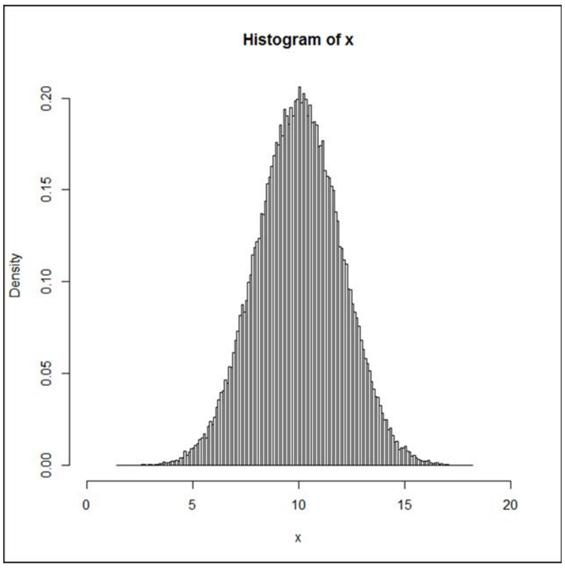
- p-value = probability of test statistic (result) occurring by chance, given that null hypothesis is true.
- Small p-value → reject null hypothesis
- Significant result ($p<0.05$) does not necessarily imply that null hypothesis is false.
- Non-significant result ($p>0.05$) does not imply that null hypothesis is true. In fact: Null hypothesis can **never** concluded to be true.

Statistical significance ($p<0.05$) does not necessarily imply that the effect is important in practice.

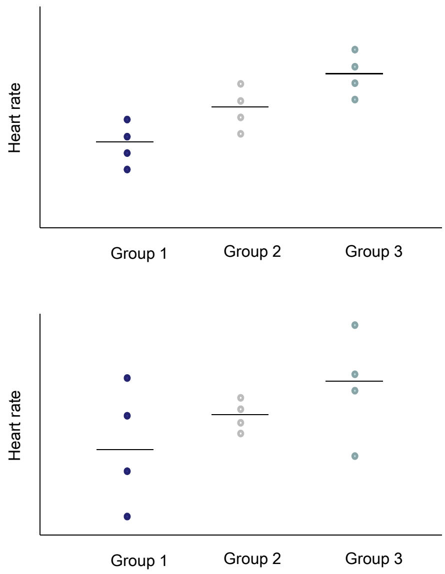
Effect sizes

- Importance of statistical significant effect in practice.
- Effect size is an objective and standardised measure of magnitude of effect.
- Can compare it over different studies, different variables, different scales of measurement.
- Effect sizes are independent of measuring units.
- Many different types of effect sizes e.g. Cohen's d value, Cramer's v value and Pearson's correlation coefficient (r).

Normality



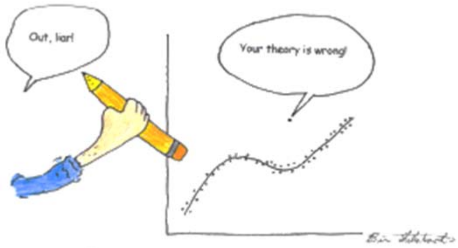
Homogeneity of variance



Outliers

- Definition
  - An outlier is an observation that lies an abnormal distance from the other values in your sample.

- Identifying outliers using graphs
  - Descriptive statistics
  - Box-and-Whiskers plot
  - Histogram



Inferential statistics

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

Statistical methods

- Parametric statistical methods
- Nonparametric / Robust statistical methods

Inferential statistics

Question	Parametric test	Nonparametric test
Is there a difference between two unrelated groups?	Independent t-test	Mann-Whitney test/ Wilcoxon rank-sum test
Is there a difference between two related groups?	Dependent t-test	Wilcoxon signed-rank test
Is there a difference between several unrelated groups?	ANOVA	Kruskal-Wallis test
Is there a relationship between two continuous variables?	Pearson Correlation	Spearman correlation

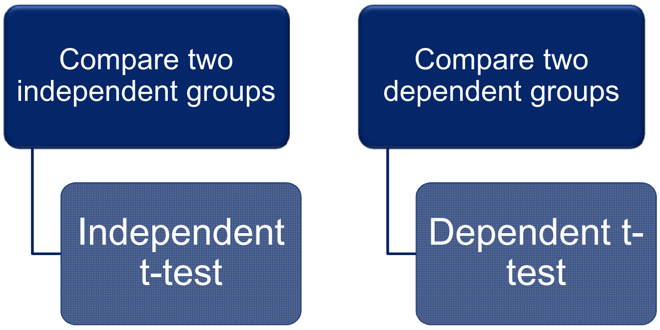
The focus of this session is on parametric statistical techniques

Comparing two means

Marike Cockeran
MURIA symposium
25 July 2016

It all starts here

Comparing two means: Outline



Methods of data collection

- Two methods of data collection:

- Independent design
  - Use different participants in different groups to take part in the study.
  - Sample a group of boys and a group of girls to test for gender differences in balance development.
  - Also called between-subjects design.

- Dependent design
  - Use the same participants in different groups to take part in the study.
  - Sample a group of children to measure their fine motor skills before and after a fine motor skills improvement program to evaluate the effectiveness of the program.
  - Also called within-subject or repeated measures design.

Independent t-test: Introduction

- The independent samples t-test compares the means between two unrelated groups on the same continuous, dependent variable.
- For example, you could use an independent t-test to compare blood pressure levels between smokers and non-smokers.

Independent t-test: Hypothesis

- Null hypothesis:

$$H_0: \mu_1 = \mu_2$$

- Alternative hypothesis:

$$H_A: \mu_1 \neq \mu_2$$

- If the sample means differ a lot, H_0 is rejected. The researcher can conclude that the two population means differ.
- If the sample means do not differ a lot, H_0 is not rejected. The researcher cannot conclude that the two population means differ.

Independent t-test: Effect size

- A measure of practical significance (effect size) is Cohen's d-value:

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{\max(s_1, s_2)}$$

- The effect size, Cohen's d-value is calculated manually.
- Guideline values for interpreting Cohen's d-value:

$|d|$ ≈ 0.2 Small effect / No practically significant difference

$|d|$ ≈ 0.5 Medium effect / Practically visible difference

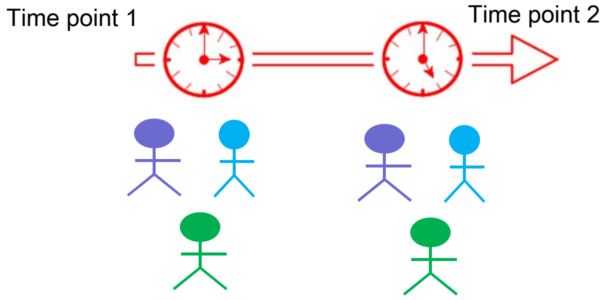
$|d|$ ≈ 0.8 Large effect / Practically significant difference

Independent t-test: Reporting of results

- **Practical significance** (effect size) guideline values:
  - |d| ≈ 0.2 small effect
  - |d| ≈ 0.5 medium effect
  - |d| ≈ 0.8 large effect
- **Statistical significance** guideline value:
  - Usually when a p-value is smaller than 0.05 the result is viewed as statistically significant.
- **Reporting of the results:**
  - An independent samples t-test was conducted to compare BMI values in males and females. There was a significant difference in the BMI values for males (M=24.12, SD=6.02) and females (M=27.47, SD=7.43), t(198)=2.89, p=0.02.

Dependent t-test: Introduction

- The dependent t-test (paired t-test) compares the means between two related groups on the same continuous dependent variable.
- Measurements are taken on a single subject at two distinct points in time.



Dependent t-test: Hypothesis

- Denote the difference in population means by:

$$\delta = \mu_1 - \mu_2$$

- Null hypothesis:

$$H_0 : \delta = 0$$

- Alternative hypothesis:

$$H_A : \delta \neq 0$$

Dependent t-test: Effect size

- A measure of practical significance (effect size) is Cohen's d-value:

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{s_1}$$

- The effect size, Cohen's d-value is calculated manually.

- Guideline values for interpreting Cohen's d-value:

  $|d| \approx 0.2$ Small effect / No practically significant difference

  $|d| \approx 0.5$ Medium effect / Practically visible difference

  $|d| \approx 0.8$ Large effect / Practically significant difference

Dependent t-test: Reporting of results

- Practical significance (effect size) guideline values:
  - |d| ≈ 0.2 small effect
  - |d| ≈ 0.5 medium effect
  - |d| ≈ 0.8 large effect
- Statistical significance guideline value:
  - Usually when a p-value is smaller than 0.05 the result is viewed as statistically significant.
- Reporting of results:
  - On average the triceps measurement was lower before the intervention (M=21.73, SD=9.39) compared to the triceps measurement after the intervention (M=21.47, SD=9.94). This difference was not statistically significant, *p* = 0.409.

Comparing several means

Marike Cockeran
MURIA symposium
25 July 2016

It all starts here

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

Analysis of variance: Introduction

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

- In the previous section we looked at techniques for determining whether a difference exists between the means of two independent populations.
- In some situations we would like to test for differences among three or more independent means rather than just two.
- The extension of the independent two-sample t-test to three or more groups is known as the analysis of variance.

ANOVA: Hypothesis

- Null hypothesis

$$\mu_1 = \mu_2 = \cdots = \mu_k$$

- Alternative hypothesis
  - At least one of the population means differs from one of the others.

- The One-way ANOVA is an omnibus test and cannot tell you which specific groups were significantly different from each other.

- The omnibus test is referred to as the F-test.

- To determine which groups differ from each other you need to use post hoc tests.

Bonferroni multiple comparison test

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

- F-test: A p-value of 0.03 is obtained. We can reject the null hypothesis. At least one of the population means differs from one of the others.

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

- You need to adjust the significance level used for each test to have an overall significance level of $\alpha = 0.05$.

- Bonferroni adjustment: $\frac{\alpha}{\#\ tests} = \frac{0.05}{3} = 0.0167$

- Bonferroni adjustment: p-value x # tests

Reporting of results

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

- Statistical significance guideline value:
  - Usually when a p-value is smaller than 0.05 the result is viewed as statistically significant.
- Reporting of results:
  - There was a statistically significant effect of educational level on nutrition knowledge, F (15.05, 2), $p < 0.001$. Tukey's test revealed that the nutrition knowledge of participants with no educational level differed statistically significantly from participants with any educational level, both $p < 0.001$. However, the nutrition knowledge of participants with a low educational level did not differ statistically significantly from participants with a medium educational level, $p = 0.413$.

Correlation analysis

Marike Cockeran
MURIA symposium
25 July 2016

It all starts here

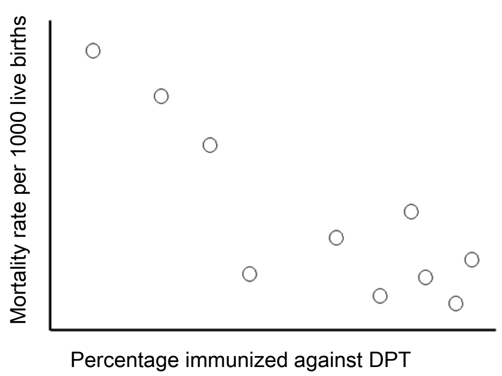
NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

Introduction

- The aim of this section is to investigate the relationships that can exist among continuous variables.
- One statistical technique often employed to measure such a relationship is known as correlation analysis.
- Correlation is defined as the quantification of the degree to which two random variable are related, provided that the relationship is linear.

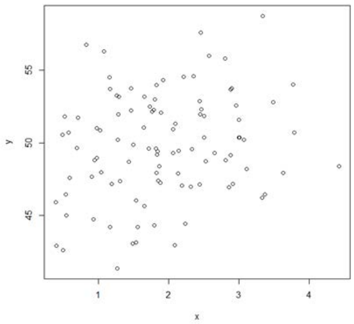
Two-way scatter plot

- Place the outcomes of the X variable along the horizontal axis and the outcomes of the Y variable along the vertical axis.
- Each point on the graph represents a combination of values (X_i, Y_i).



Correlation analysis

- Correlation is defined as the quantification of the degree to which two random variables are related, provided that the relationship is linear.
- Graphic presentation of the correlation between two variables is called a two-way scatter plot.

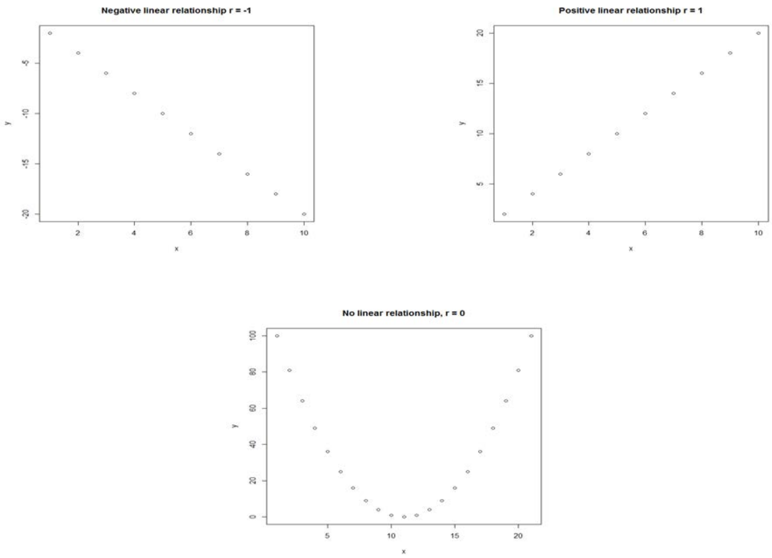


Pearson's correlation coefficient

- The parametric estimator of the correlation between two variables is known as Pearson's correlation coefficient (r).
- The maximum value of r is 1; the minimum value of r is -1.
- If $r = 1$ or $r = -1$ then an exact linear relationship exists between the two variables.
- If $r = 0$ there is no linear relationship between the two variables and the variables are uncorrelated.
- If the values of the first variable increase as the values of the second variable increase, then the two variables are positively correlated.
- If the values of the first variable decrease as the values of the second variable increase, then the two variables are negatively correlated.

Pearson's correlation coefficient

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS



Pearson's correlation: Hypothesis testing

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

- Null hypothesis:

$$H_0:\ \rho = 0$$

- Alternative hypothesis:

$$H_0:\ \rho \neq 0$$

Correlation coefficient as an effect size

- The effect size can be used to measure the strength of the relationship between the two continuous variables.
- $|r| = 0.1$ small effect
- $|r| = 0.3$ medium effect
- $|r| = 0.5$ large effect

Summary of the session

- Measurement scales
- Descriptive statistics
- Basic terminology used in inferential statistics
- Comparing two means
  - Independent t-test
  - Dependent t-test
- Comparing several means
  - One-way ANOVA
- Determining the relationship between two continuous variables
  - Pearson's correlation coefficient