# Biometrics

**August 2018 – January 2019**

Edition: 2018

Radboudumc Health Academy
Geert Grooteplein 21
Postbus 9101
6500 HB  NIJMEGEN
Tel: (024) 361 70 50

# Contents

COURSE ORIENTATION ..... 1

Introduction ..... 1

Objectives and topics ..... 2
- Course aim ..... 2
- Topics ..... 2
- Prerequisites ..... 2
- Literature ..... 3

Attendance, examination, and evaluation ..... 3
- Attendance ..... 3
- Examination ..... 3
- Evaluation ..... 3

COURSE ORGANIZATION ..... 4

Course committee and instructors ..... 4

Course schedule ..... 4

Study Path Biometrics 2018 – 2019 ..... 5

Preparation schedule ..... 7

STUDY MANUAL ..... 9

Meeting 1 ..... 9
- Introduction: Getting acquainted with each other, organisational remarks and Introduction to and use of the book by Andy Field ..... 9
- Lecture 1: Variation ..... 9
- Preparation for Meeting 2 ..... 13
- Computer Session 1: Descriptive statistics ..... 14

Meeting 2 ..... 23
- Lecture 2: Confidence intervals and hypothesis testing ..... 23
- Preparation for Meeting 3 ..... 23
- Computer Session 2: T-test, confidence interval and P-value ..... 24

Meeting 3 ..... 29
- Lecture 3: Linear regression analysis ..... 29
- Preparation for Meeting 4 ..... 29
- Computer Session 3: Linear regression ..... 30

Meeting 4 ..... 35
- Lecture 4: 2 x 2 table and logistic regression ..... 35
- Logistic regression analysis. ..... 36
- Preparation for Meeting 5 ..... 40
- Computer Session 4: 2 x 2 table and logistic regression ..... 41

Meeting 5 ..... 47
- Lecture 5: Sample size and power ..... 47
- Preparation for Meeting 6 ..... 52
- Computer Session 5: Sample size and power ..... 54

Meeting 6 ..... 59

Lecture 6: Comparing more than two means (ANOVA) ..... 59
Preparation for Meeting 7 ..... 59
Computer Session 6: Comparing more than two means (ANOVA) ..... 61

Meeting 7 ..... 67
Lecture 7: Analysis of Covariance (ANCOVA) ..... 67
Preparation for Meeting 8 ..... 67
Computer Session 7: ANCOVA ..... 68

Meeting 8 ..... 73
Lecture 8: Experimental Designs ..... 73
Preparation for Meeting 9 ..... 73
Computer Session 8: Experimental Designs ..... 74

Meeting 9 ..... 83
Lecture 9: Repeated Measurements ..... 83
Preparation for Meeting 10 ..... 84
Computer Session 9: Mixed models for Repeated Measurements ..... 85

Meeting 9 ..... 91
Lecture 10: Survival analysis ..... 91
Preparation for Meeting 11 ..... 94
Computer Session 10: Survival analyses ..... 95

Meeting 11 ..... 101
Lecture 11: Logistic regression analysis: prediction ..... 101
Preparation for Meeting 12 ..... 101
Computer Session 11: Logistic regression analysis ..... 102

Meeting 12 ..... 107
Lecture 12: Multilevel models ..... 107
Preparation for Meeting 13 ..... 107
Computer Session 12: Multilevel models ..... 108

Meeting 13 ..... 113
Lecture 13: 'A gentle introduction to imputation of missing values' ..... 113
Preparation for Meeting 14 ..... 113
Computer Session 13: Missing data ..... 114

Meeting 14 ..... 117
Lecture 14: Introduction to meta-analysis ..... 117
Preparation for Meeting 15 ..... 117
Computer Session 14: Introduction to meta-analysis ..... 118

Meeting 15 ..... 123
Lecture 15 Data visualisation ..... 123
Preparation for next week ..... 123
Computer Session 15 ..... 124

Additional assignments ..... 129
Overview ..... 129

Evaluation form ..... 135

# COURSE ORIENTATION

## Introduction

This course book consists of three chapters. This first chapter describes the position of the Biometrics course within all the post-academic education (Radboudumc Health Academy) courses, its relevance, learning objectives, prerequisites, and the evaluation.

The second chapter presents the organization of the course: the instructors, the structure, a description of the different types of learning forms (lectures, computer exercises), and the course schedule.

The third chapter contains the learning forms and the assignments for the course biometrics.

### *Position of the course within the post-academic courses*

The Biometrics course is part of the post-academic education of the Department for Health Evidence. This postgraduate course is particularly aimed at those who have previously followed an introductory course in (bio)statistics and now intend to do research. It is especially intended for junior researchers from Radboud university medical center, but also open to other participants. The Biometrics course was developed for this target group and has been taught for a number of years, originally by late prof. dr. M. van 't Hof.

The course was redesigned at several times. The course builds upon the content and view of the (second year) biostatistics courses given by the Department for Health Evidence as part of the curriculum for the study Biomedical and Health Sciences, along with selected contents from the courses on Experimental Design, Diagnostic Research and Measurement Error. The choice of the book *Discovering Statistics Using IBM SPSS STATISTICS* by Andy Field was motivated by a desire to give the course a 'user-friendly' character without losing its solidity. In order to provide practical experience, every session ends with computer exercises. In total, the course is an 100 hour program, approximately 1/3 of which is in the form of instructional and computer hours on Thursday afternoons in the months August 2018 - January 2019.

### *Relevance*

The training for researcher in the medical sciences aims to train junior researchers to become professionals who can contribute to the improvement of public health and the health system through scientific research. This training contains the basic research skills needed for all medical researchers. More depth is developed via the in-service training at the department where the researcher is appointed.

Knowledge of the basics of biostatistics is essential for both the analysis of medical research data and for a critical appraisal of research articles. In your undergraduate studies you learned about basic statistical research methods and the essential elements of the research process. In addition, you learned to assess the use of statistical methods by others (in research papers, etc.). You also learned to interpret the most important statistical procedures that appear in scientific publications. However, much of this statistical and methodological training was covered in the first two years of your undergraduate studies. Now that you are preparing to do your own research, the time is ripe to refresh your statistical and methodological knowledge and to learn essential elements that were not covered in your educational program.

This Biometrics course will cover statistical methods to adequately explore, describe, display and analyze data. The emphasis will be on understanding how to interpret statistical results and how observed effects in a sample of patients can be generalized to the population. You will learn how to design an analysis plan for medical-scientific data yourself, how to perform several commonly used techniques and how to interpret the results.

## Objectives and topics

### *Course aim*

This course aims to (re)introduce the basics of biostatistics from an applied biomedical perspective, give a greater understanding of statistical concepts, and help researchers understand the most commonly used biostatistical methods, including when and when not to use them. Upon completion of this course, the student should be able to:

1. design an analysis plan for medical-scientific data,
2. perform several commonly used statistical analysis techniques using SPSS,
3. interpret the results of these analyses,
4. to appreciate more advanced statistical techniques such as repeated measures analyses

### *Topics*

1. Describing and displaying data
2. Distributions
3. Hypothesis testing
4. Estimation using confidence intervals
5. Interaction and Confounding
6. Analysis of Variance (ANOVA)
7. Analysis of Covariance (ANCOVA)
8. Regression analysis (linear, logistic, Cox proportional hazards)
9. Analyses of Repeated Measurements
10. Multilevel analyses
11. Missing value analysis
12. Design and analysis of experiments
13. Meta-analysis

In addition, you will be given the opportunity to analyze your own data set with the help of an experienced biostatistician.

### *Prerequisites*

The course assumes that the student has previously had an introduction to statistics. Because the knowledge and skills obtained at that time may have been lost over time, this course will start with a short refresher course of basic statistical topics that are relevant for a medical-scientific researcher.

### *Literature*

The following book is mandatory for this course:
Andy Field, *Discovering Statistics Using IBM SPSS*, SAGE Publications, 5th edition, 2018. ISBN-978-1-5264-1952-1

## Attendance, examination, and evaluation

### *Attendance*

The learning process in this Biometrics course is strongly based on the interaction between students. The student-instructor contact moments are chosen such that the questions/topics that are difficult for the students can be covered. Therefore, you are strongly advised to be present at all lectures, and computer exercises. When you cannot be present, please notify the coordinator (Ton de Haan, Ton.deHaan@radboudumc.nl). Attendance will be registered.

### *Examination*

You will be awarded a certificate upon completion of the course. There will be a written exam (2 hours, open-end questions, open book) on 17 January, 2019.

Every student will receive written notification of the exam result. If the exam result is unsatisfactory, the exam can be re-taken at a later date. If the exam result is satisfactory, the student will receive a certificate showing that the course has been completed with a satisfactory exam result.

### *Evaluation*

At the end of the course, students will receive an evaluation form for comments on the course. The coordinator also welcomes suggestions and comments during the course.

# COURSE ORGANIZATION

## Course committee and instructors

- Ton de Haan, MSc tel. (024-36)67346 (Coordinator)
- Rana Dandis, MSc tel. (024-36)17669
- Dr. Joanna in ‘t Hout tel. (024-36)17666
- Dr. Steven. Teerenstra, tel. (024-36)17679
- Jordache Ramjith, MSc tel. (024-36)68364
- Dr. Rogier Donders tel. (024-36)17794

All from the Department for Health Evidence, Section Biostatistics

## Course schedule

The course will begin on the 30th of August 2018 and will meet every Thursday afternoon. The course consists of 15 afternoons with lectures and computer sessions, 2 afternoons for analyzing your own data, a question hour one week before the exam and the two hour exam.

## Study Path Biometrics 2018 – 2019

| Date | Meeting | Time | Educational Method | Topic |
|---|---|---|---|---|
| 30/8 | 1 | 13.30 – 14.00 | Introduction | Getting acquainted with each other, organizational remarks and Introduction to and use of the book by Andy Field |
| | | 14.00 – 15.15 | Lecture 1 | Variation |
| | | 15.15 – 17:00 | Computer Session 1 | Descriptive Statistics |
| 6/9 | 2 | 13:30 – 14.30 | Lecture 2 | Confidence intervals and hypothesis testing |
| | | 14.30 – 16.30 | Computer Session 2 | T-test, confidence interval and P-value |
| 13/9 | 3 | 13:30 – 14.30 | Lecture 3 | Linear Regression |
| | | 14.30 – 16.30 | Computer Session 3 | Linear Regression |
| 20/9 | 4 | 13:30 – 14.30 | Lecture 4 | The 2x2 table; Logistic regression: correction for confounding |
| | | 14.30 – 16.30 | Computer Session 4 | The 2x2 table; Logistic regression: correction for confounding |
| 27/9 | 5 | 13:30 – 14.30 | Lecture 5 | Sample size and power |
| | | 14.30 – 16.30 | Computer Session 5 | Sample size and power |
| 4/10 | 6 | 13:30 – 14.30 | Lecture 6 | Comparing more than two means |
| | | 14.30 – 16.30 | Computer Session 6 | Comparing more than two means |
| 11/10 | 7 | 13:30 – 14.30 | Lecture 7 | ANCOVA |
| | | 14.30 – 16.30 | Computer Session 7 | ANCOVA |
| 18/10 | 8 | 13:30 – 14.30 | Lecture 8 | Experimental Design |
| | | 14.30 – 16.30 | Computer Session 8 | Experimental Design |
| 25/10 | 9 | 13:30 – 14.30 | Lecture 9 | Repeated measurements |
| | | 14.30 – 16.30 | Computer Session 9 | Mixed models |
| 1/11 | 10 | 13:30 – 14.30 | Lecture 10 | Survival Analysis |
| | | 14.30 – 16.30 | Computer Session 10 | Survival Analysis |

| Date | Meeting | Time | Educational Method | Topic |
|---|---|---|---|---|
| 8/11 | 11 | 13:30 – 14.30 | Lecture 11 | Logistic Regression Analysis: prediction |
| | | 14.30 – 16.30 | Computer Session 11 | Logistic Regression Analysis: prediction |
| 15/11 | 12 | 13:30 – 14.30 | Lecture 12 | Multi level models |
| | | 14.30 – 16.30 | Computer Session 12 | Multi level models |
| 22/11 | 13 | 13:30 – 14.30 | Lecture 13 | Imputation of Missing values |
| | | 14.30 – 16.30 | Computer Session 13 | Imputation of Missing values |
| 29/11 | 14 | 13:30 – 14.30 | Lecture 14 | Introduction to Meta analyses |
| | | 14.30 – 16.30 | Computer Session 14 | Introduction to Meta analyses |
| 6/12 | 15 | 13:30 – 14.30 | Lecture 15 | Data visualisation |
| | | 14.30 – 16.30 | Computer Session 15 | Data visualisation |
| 13/12 | 16 | 13.00 – 17.00 | | Analyze your own data |
| 20/12 | 17 | 13.00 – 17.00 | | Analyze your own data |
| 10/1 | 18 | 15.30 – 17.00 | Question Time | |
| 17/1 | 19 | 13.30 – 16.30 | Examination | |

## Preparation schedule

| Date | Preparation for Meeting | Reading in Field | Self Diagnostics |
|---|---|---|---|
| 30/8 | 1 | Read your own textbook on basic statistics | |
| 6/9 | 2 | Chapters 1 and 2; Section 6.4<br>Scan Chapters 4 and 5 so you can use them for reference | Basic SPSS<br>Descriptive Statistics |
| 13/9 | 3 | Sections 9.1, 9.2, 9.3.1, 9.4, 9.6 - 9.8 | Confidence intervals and hypothesis testing |
| 20/9 | 4 | Sections 19.1, 19.2, 19.3.1-19.3.3<br>Remark: Never use (Yates') continuity correction.<br>Chapter 20 all parts that are marked ||| or ||||, but not 20.3.6.<br>Section 20.3.5<br>Smart Alex's Tasks, especially task 5 and 6 are instructive (in combination with the elaborations on the book's website). | Linear regression |
| 27/9 | 5 | Section 2.9.8; (if time left 3.1-3.7) | Chi-square test and logistic regression |
| 4/10 | 6 | Chapter 12 of Field (Comparing Several Means: ANOVA (GLM 1)) contains many technical details that may be omitted, such as sections: 12.2.2-11.2.6, 12.4-.12.11<br>12.14 (Key Terms That We've Discovered): omit all except:<br>Analysis of Variance (ANOVA)<br>Bonferroni correction<br>Pairwise comparisons<br>Post hoc tests<br>Sections 14.1 – 14.3 till 14.3.1 (not 14.3.1) | Sample size calculations and power |
| 11/10 | 7 | Chapter 13 Analysis of Covariance. The sections marked |||| and the following sections may be omitted when reading: 13.5.5 - 13.5.7, 13.6.2 – 13.6.3; 13.9 – 13.11.<br>Do read 13.7 | Comparing Means (ANOVA) |
| 18/10 | 8 | Sections 1.7; Chapter 13 Factorial ANOVA the following sections:<br>Sections 14.2, 14.3 till 14.3.1, 14.5 (not 14.5.6) 14.6, 14.7 | ANCOVA |
| 25/10 | 9 | Sections 21.1-20.4, 21.5.1 | Experimental Design |
| 1/11 | 10 | Please read the text of Petrie and Sabin on survival analyses (page 92 of this manual)<br>Spruance, S. L., J. E. Reid, et al. (2004). "Hazard Ratio in Clinical Trials." Antimicrobial Agents and Chemotherapy **48**(8): 2787-2792. (Link on Brightspace) | Linear Mixed Models |
| 8/11 | 11 | Chapter 20 all parts that are marked |||| or ||||, but not 20.3.6, 20.5.12, and 20.5.13. | Survival analysis |
| 15/11 | 12 | Sections 21.1-21.4 | Logistic regression (prediction) |

| Date | Preparation for Meeting | Reading in Field | Self Diagnostics |
|---|---|---|---|
| 22/11 | 13 | Please read the article "*Review: A gentle introduction to imputation of missing values*" written by ART Donders, GJMG van de Heijden, T Stijnen, and KGM Moons in Journal of Clinical Epidemiology 59 (2006) 1087-1091. Link on Brightspace | Multilevel Models |
| 29/11 | 14 | Please read the article *"Introduction to systematic reviews and meta-analysis"* written by JE McKenzie, EM Beller, and AB Forbes in Respirology, 21(2016): 626–637. doi: 10.1111/resp.12783. Link on Brightspace | Missing data |
| 6/12 | 15 | Next week we are going to use R for data visualisation. Please read the instruction on Brightspace how to install R and R-studio on your own computer. | Meta analysis |

# STUDY MANUAL

## Meeting 1

***Introduction:*** ***Getting acquainted with each other, organisational remarks and Introduction to and use of the book by Andy Field***

### *Lecture 1: Variation*

Teacher: Jo Ramjith

First, we'll look at three applications of biostatistics:

a. describing the data and the study results
b. assessing the role of chance in the obtained study results
c. assessing multi-dimensional relationships

Note that the word 'statistics' has two meanings.

- Firstly, the word 'statistics' refers to the science that develops and investigates techniques and quantities to describe and draw inferences from data.
- Secondly, it is the plural of the word 'statistic', which means 'a quantity that can be calculated from the data and that takes a different value depending on the sample (because each sample has different data)'. Examples are (distributional) statistics such as parameters of location (e.g., mean, median) that describe the center of the distribution or parameters of dispersion (e.g., standard deviation, range, interquartile range) that describe the 'width' or amount of variation in the distribution. Later on, we will also encounter (inferential) statistics such as the t-test.

**a. Describing the data and the study results**

Descriptive statistics play an important role in all aspects of data analysis. In the first place, quantities such as the mean and the standard deviation can be used to describe a set of data. During data analysis, graphical displays of the descriptive statistics may give a useful insight in the relations between various phenomena, which enables a motivated choice for further analysis. After the final data analysis is complete, descriptive statistics are used to provide an orderly presentation of the results in figures or tables for a publication.

**b. Assessing the role of chance in the obtained study results**

We may have a research question that pertains to patients with cardiovascular disease, for example. In order to answer the research question directly, we would have to test all people with cardiovascular disease (the 'population'). Of course we can't do that. So, we test a 'sample' of people with cardiovascular disease and, based on the results in the sample, we infer something about the population of all people with cardiovascular disease.

Inferential statistics takes the role of chance into account when interpreting the results found in a sample. If Medication A is better than Medication B at preventing heart attacks in a 'sample' of people with cardiovascular disease, how confident can we be that this will also be the case in the 'population' of all people with cardiovascular disease? Inferential statistics help us to distinguish between chance findings (results that are specific to this particular sample) and systematic trends (results that can reasonably be expected to be found in the population).

The act of drawing a general conclusion (pertaining to the population of interest) on the basis of results found in a sample is called *statistical inference*. The most important techniques of statistical inference are testing hypotheses (using p-values) and estimating confidence intervals.

**c. Assessing multi-dimensional relationships**

Multi-dimensional relationships exist when many variables or factors are associated with the outcome. For example, whether or not a person with cardiovascular disease has a heart attack may depend not only on the type of medication used (A or B) but also on age, gender, smoking habits, activity level, etc. The statistical techniques used in such situations are called 'multivariate techniques' and include analyses such as multiple (multivariate) regression, analysis of variance (ANOVA), and logistic regression. These techniques play an important role in the correction for 'confounders'.

**d. Descriptive statistics**

Techniques to summarize the observations are called descriptive statistics. The two main types of descriptive statistics are:

- Numerical methods, including statistics such as the mean, median, standard deviation, correlation coefficient, percentages, etc.
- Graphical method such as the histogram, scatter plot, pie-chart, bar-chart, box-and-whisker plot.

Descriptive statistics play an important role in all aspects of data analysis. Descriptive statistics are used prior to the actual data analysis to detect erroneous data and outliers. During data analysis, graphical displays of the data may give a useful insight in the relationships between different phenomena, which can enable a motivated choice for the further analysis. After the data analysis has been finished, descriptive statistics are used to provide an orderly presentation of the results in figures or tables for a publication.

## Math revision

## What is a Logarithm (log)?

Any number can be written as a base raised to a power. For example, we can write 100 as $10^2$, where 10 is the base and 2 is the power (or exponent).

The **log** of a number (to a certain base) is the power that the base needs to be raised to in order to produce that number. Let's look at a simple example.

The most commonly used bases are 10 for the *common logarithm* and *e* for the *natural logarithm*. [Remember that *e* is a mathematical constant that has a value of 2.71828 (rounded to 5 decimals).]

Since base 10 may be a little more intuitively understandable, we'll look at the common logarithm (i.e. base 10) first.

Example 1. The *common logarithm* of 100 is 2, because 100 can be written as 10 to the power of 2 (i.e. $10^2 = 100$). So, in general, for a number *x*, which can be expressed as base *b* raised to the power of *y*:

$$\text{If} \quad x = b^y, \quad \text{then} \quad {}^{b}\log(x) = y$$

Specifically, for the example above, $x = 100$, $b = 10$ and $y = 2$. So, because $100 = 10^2$, the common log of 100 is 2, which can be written in the following notation:

$${}^{10}\log(100) = 2$$

Example 2. The *natural logarithm* has the base *e*. The natural logarithm of 100 is 4.61, because 100 is equal to *e* to the power of 4.61:

$$e^{4.61} = 100 \qquad \text{So,} \qquad {}^{e}\log(100) = 4.61$$

The natural log is often abbreviated LN. Instead of ${}^{e}\log(x)$, the natural log is often written LN(*x*). So, for the above example,

$$\text{LN}(100) = 4.61$$

Note that we cannot take the log of zero or of negative numbers.

**Math revision**

(continued from previous page)

**Back-transformation**

Remember that the log of a number is the power that the base needs to be raised to in order to produce that number. So, if we have log transformed data and we want to transform the data back to the original scale, we do this by raising the base to the power of the log.

So, we can convert the natural log of x (where LN(x) = y) back to the original scale by raising *e* to the power of y ($e^y$). This is called 'exponentiation' or 'taking the antilog'. Note that $e^y$ can also be written as $e^{LN(x)}$ or EXP(y) or EXP(LN(x))

So, for the natural log of 100, LN(100) = 4.61, the original number (100) is returned by raising *e* to the power of 4.61 ($e^{4.61} = 100$).

**Log transformation of positively skewed data**

The distributions of many biochemical measurements are not normally distributed, but are positively skewed (i.e. skewed to the right). This means that there are outliers on the high end of the scale, but not on the low end. Many commonly used statistical procedures assume normally distributed data. Using these procedures on skewed data can lead to unreliable conclusions. However, a logarithmic transformation can often help in such cases.

The logarithmic transformation 'squeezes' large numbers closer together, while small numbers are pulled away from each other. Because of this property a positively skewed distribution of data becomes a more symmetrical distribution after log transformation (i.e. more 'normal'). If a variable is positively skewed, the log transformation of that variable will be approximately normally distributed. Statistical procedures that assume normally distributed data can then be performed on the transformed data.

The LN transformed data is no longer on the scale of the original data, but statistics resulting from the LN transformed data (such as the mean) can be easily transformed back to the original scale by taking the antilog ($e^{LN(x)}$), making the interpretation straightforward.

**Convenient properties of logs**

The LN transformation has a number of convenient properties. Because the rank-order of the scores is preserved in the LN transformation, the middle score of the original data (i.e. sample median) is also the middle score of the transformed data. If we transform the mean of the log transformed data (also called the geometric mean) back to the original scale this value will be *approximately* equal to the median of the original data. This value will usually be a smaller than the original mean and a bit smaller than the original median. Percentiles, such as the median and quartiles, are (largely) preserved under log transformation.

Another important feature of logs is that they *reduce multiplication to addition*, by the formula:

$$^b\log(c*d) = {}^b\log(c) + {}^b\log(d)$$

So, the log of a product is the sum of the two logs. In other words, the log of the product of two numbers is equal to the sum of the logs of those numbers.

Similarly, logs *reduce division to subtraction* by the formula:

$$^b\log(c/d) = {}^b\log(c) - {}^b\log(d)$$

So, The log of a quotient is the difference of the logs.

### *Preparation for Meeting 2*

In preparation for next week:

- Please read (Field):
  - Chapters 1 and 2
  - Section 6.4
  - Scan Chapters 4 and 5 so you can use them for reference
- Self diagnosis on Brightspace
  - Basic SPSS
  - Descriptive Statistics

## *Computer Session 1: Descriptive statistics*

### *Background*
One of the first steps in data analysis is to make a good summary of the data that have been collected in the study. This gives insight in the distribution and variation of the outcomes. In fact, these descriptions are also required in the final research publication. To summarize the data, three types of descriptive statistics are available: figures, tables and statistical measures (of location and spread).

### *Objective*
After completing the assignments, the student should be able to:
- Use SPSS to calculate derived variables and recode variables
- Use SPSS to display the observations of an investigation in a frequency table or a figure (histogram, box plot)
- Use SPSS to summarize the observations using summary statistics such as the mean, median, mode, percentiles, standard deviation, and inter-quartile range.

### *Literature*
Chapters 3 and 4 of Field.

### *Instruction*
Make the assignments. In Chapters 3 and 4 of Field you can find details on how to use SPSS.

### *Product*
Written answers and computer output. Please fill in the evaluation form at page 135.

### *Discussion*
Results can be discussed with your peers and with the tutor during the computer session.

### *Assignments*

**Assignment 1 overweight**
The data in dataset '*NBS-obesitas.sav*' are a selection from the "Nijmegen Biomedische Studie" (NBS). This computer session we focus on a sample of participants using body measurements, diet and alcohol use.

Look at the table on the next page. The table describes the content of the dataset NBS-obesitas.sav. The variables *AGE_cl, BMI, BMI_cl, WEIGHT_CH, and D_SUCCES* are not in the dataset and need to be calculated.

1) GETTING FAMILIAR WITH SPSS
   a) Start SPSS and open the file NBS_OBESITAS.SAV. SPSS opens with the data window (left under: *Data View*). The structure of the file is as follows: the rows contains the data of different subjects, the columns contain the data of different variables. How many subjects are in the dataset and how many variables?

   b) By clicking on *Variable View* (see left under) you see detailed information on the variables: the label, coding measurement scale etc.

c) Derive the variables BMI and Weight_ch. (Transform – Compute variable)

d) Derive the variables *AGE_cl, BMI_cl, D_SUCCES.* (*Transform – Recode into different variables*)

e) What scale of measurement do the variables have?

| **Variable Name** | **Variable Label (description)** | **Value Labels (meaning of the values)** | **Scale of measurement (type data)** |
|---|---|---|---|
| ID | Identification number | number range 1-195 | |
| GENDER | gender | 1 = male<br>2 = female | |
| BIRTHDAY | birthday | Date from 12 march 1909<br>to 11 may 1983 | |
| AGE | Age at start of study | 19 – 93 years | |
| LGT | length in cm | 146 – 198 cm | |
| WEIGHT | weight in kg | 44 – 125 kg | |
| WHR | Waist-Hip-Ratio | 0,68 – 1,11 | |
| DIET | Atkins-diet followed? | 0 = no, other diet<br>1 = yes, Atkins-diet | |
| WEIGHT_AFTER | weight in kg after 3 months of diet | 41 – 118 kg | |
| ALC | drinks alcohol? | 0 = no<br>1 = yes | |
| ALC_HH | How much alcohol per week | 0 = NA, drinks no alcohol<br>1 = 1 to 2 days per week<br>2 = 3 days or more per week | |
| AGE_cl | age in classes | 1 = less than or equal to 30 years<br>2 = from 30 till 50 years<br>3 = from 50 till 70 years<br>4 = 70 years of older | |
| BMI | Body Mass Index in kg/$m^2$ | 16.9 – 38.9 kg/$m^2$ | |
| BMI_cl | Body Mass Index in classes | 1 = less than or equal to 20 kg/$m^2$<br>2 = from 20 to 25<br>3 = from 25 to 30 (overweight)<br>4 = 30 of more (obesity) | |
| WEIGHT_CH | Change in body weight (weight after diet) – (start weight) | range +5 to -16 kg | |
| D_SUCCES | Was the diet successful? | 0 = no, not yet 5 kg weight loss<br>1 = yes, at least 5 kg weight loss | |

f) Make a frequency table of the variables Gender and AGE_cl. (*Analyze – Descriptive statistics – Frequencies.* Select the variables ‘Gender and ‘AGE_cl’. Click on [OK].
The output window pops up.

i) How many males participants are in the dataset?

ii) What percentage of the participants is 70 years or older?

g) For continuous variables such as AGE you can calculate the mean and the standard deviation. You can do this by *Analyze – Descriptive statistics – Descriptives*. Analyse the variables, 'Length' en 'Weight' in this way and see what the output gives.

   i) What is the mean and standard deviation of length rounded to cm?

   ii) What is the minimal en maximal weight?

h) The Body Mass Index is a measure for bodyweight corrected for length. Let the computer program give an extended description of the variable BMI. To do so, choose *Analyze – Descriptive statistics – Explore*. Put the variable 'BMI' in the box '*Dependent list*'. Mark after clicking the button '*Statistics'* also the option '*Outliers'*. Click on '*Continue'* and mark after clicking button '*Plots'* also the option '*Histogram'*, followed by [*Continue*] and [*OK*].
   We see in the output the summary measures, the extreme values, the histogram, a stem-and-leaf plot and a boxplot. Study the output.

   i) Give the median BMI and its interpretation.

   ii) Give the inter quartile range and its interpretation.

   iii) Which participant has the lowest BMI? What is his/her BMI value?

   iv) Do the data support a Normal distribution of BMI?

   v) What is the meaning of the two T-ends? How many outliers are present in the boxplot?

   (NB. Use the option 'Explore' only for continuous variables. See for instance what happens if you use 'BMI_cl').

2) RELATION BETWEEN GENDER AND OVERWEIGHT

a) How many subjects do have obesity? What are the percentage obese people in this group? (Variable BMI-cl; BMI $\geq$ 30.0 kg/m2)?
   Explore this by using *Analyze – Descriptive statistics – Frequencies*. What percentage has a too high body weight?

b) People say that obesity is more prevalent in women than in men. You can verify this by making a cross tabulation: *Analyze – Descriptive statistics – Crosstabs*. Choose the variable 'Gender' for the rows and 'BMI_cl' for the columns. You get the percentage by clicking '*Cells'* and marking under the heading '*Percentages'* your choice. Now you can fill in the next table.

| Gender | Obesity | | |
|---|---|---|---|
| | Present | Absent | Total |
| Male | .. | .. | |
| Female | | | |
| Total | .. | .. | |

i) Does this table confirm the hypothesis that obesity is more prevalent in women?

ii) Make also a box-whisker-plot for BMI to confirm your findings (see instruction earlier).

c) The BMI and the waist-hip ratio (WHR) are both measures for obesity and the amount of body fat

i) Show the relation between WHR and BMI by means of a scatterplot: *Graphs – Chart builder - Scatter/Dot – Simple Scatter (double click) (or Graphs - Legacy Dialogs - Scatter/Dot – Simple Scatter, Define).* Move 'BMI' to the X-axis and 'WHR' to the Y-axis. What is your conclusion about the strength of the relation?

ii) Calculate the correlation coefficient r for the relation between WHR and BMI using (*Analyze – Correlate – Bivariate*). How do you interpret this correlation coefficient?

3) THE EFFECT OF DIET ON OVERWEIGHT

a) Some participants followed the so called Atkins-diet (not many carbohydrates, much protein and fats) to lose weight.

i) How many participants followed the Atkins-diet?

ii) How many subjects had a weight loss of 5 kg or more?

iii) What percentage of subjects in the two diet groups has a successful diet (weight loss of 5 or more kg)?
*(Hint: make a cross tabulation with percentages)*

*iv)* What is your conclusion concerning the effectiveness of the Atkins-diet compared with the other diets?

*b)* The Atkins-diet is apparently better than the other diets. The consumption of alcohol could diminish this effect. Is support for this hypothesis available in the dataset? To investigate this, please fill in the next table. *(Hint: make a cross tabulation of diet against success, separately for alcohol users and non-alcohol users. You can arrange that using 'Layer' in the 'Crosstabs' procedure).*

| **Alcohol** | **Diet** | **% successful weight loss** |
|---|---|---|
| Drinks no alcohol | Atkins | … % |
| | Other diet | … % |
| | | |
| Drinks alcohol | Atkins | … % |
| | Other diet | … % |

| **How much alcohol per week** | **Diet** | **% successful weight loss** |
|---|---|---|
| Drinks no alcohol | Atkins | … % |
| | Other diet | … % |
| | | |
| 1 to 2 days | Atkins | … % |
| | Other diet | … % |
| | | |
| 3 days or more | Atkins | … % |
| | Other diet | … % |

**Assignment 2 Paediatric surgery**

The SPSS dataset *surgery.sav* contains data of 141 newborns that were sent to the academic hospital for surgery.

a. Open in SPSS the dataset *surgery.sav*

In this assignment we consider the variables *birth weight*, *gestational age* and *length of stay.* We want to get insight in the distribution of these variables.

b. What distributions do you expect for *birth weight*, *gestational age* and *length of stay?*

c. First we want a general description per variable.
To do so click: *Analyze – descriptive statistics- explore*
Mark *Exclude cases pairwise*, located under the button O*ptions.* If you explore more than one variable, the default SPSS handling of missings is listwise deletion. This means that the descriptive is given for the subjects with ALL variables non-missing.

d. Get a first impression about whether or not you have a normal distribution is by comparing the median to the average. In the ideal (normal) case, these two values are equal. A second indication is given by the **skewness**. A skewness between -1 and 1 is an indication of a symmetrical distribution, values between -3 and -1 and between 1 and 3 are an indication for skewed distributions. Values below -3 and above 3 are a good indication for a non-normal distribution. The third option is to look at the **kurtosis**. A value above 0 is an indication for a stronger peak than for a Normal distribution, values below 0 indicates a flatter distribution than the Normal distribution. Values between -1 and 1 are no contra indication for a Normal distribution, but values outside the interval -3 to 3 are a strong indication for a non-normal distribution.
What is your opinion about the distribution of the variables *birth weight*, *gestational age* en *length of stay?*

e. Different graphical methods are available (histogram, boxplot) to get an impression of the distribution of a variable.

To create a <u>histogram</u> do as follows:
*Graphs – Chart Builder – histogram – Simple histogram (double click)* - move the variable *birth weight* to the box *x-axis.* (You can also use *Graphs - Legacy Dialogs- histogram* – move the variable *birth weight* to the box *variable* and click on *OK*.*)*
Repeat this procedure for the other two variables.

To create a <u>boxplot</u> do as follows:
*Graphs – Chart builder – boxplot- 1D-Boxplot (double click)* move the variable *birth weight* to the box *x-axis (*You can also use *Graphs – Legacy Dialogs - boxplots – simple – Mark summaries of separate variables – define* – and move the variable *birth weight* to the box *Boxes Represent.)*
Repeat this procedure for the other two variables.
Find out using this method whether the three variables act like a Normal distribution.

To create a boxplot separate for boys and girls in one figure, use *Graphs – Chart builder – boxplot- Simple Boxplot (double click)* move the variable *birth weight* to the box y-*axis* and *Gender* to the x-axis.

f. Does an LN-transformation normalize the distribution? Argue which variables could be normalized with an LN-transformation and which certainly not. Apply the LN-transformation for those variables and find out whether the LN-transformated values follow a Normal distribution.

**Assignment 3 Kidney transplants**
The dataset *kidneytransplantation.sav* contains measurements made in 1228 patients who received a new kidney either from a living donor or a dead donor.

| **Variable Name** | **Variable Label (description)** | **Value Labels (coding values)** |
|---|---|---|
| ID | Patient identification number | unique number from 1863 to 110465 |
| SEX | gender | 1 = male<br>2 = female |
| AGE | Age at transplantation | 18 – 76 year |
| AGE_CAT | Age categories | 1: ≤ 40 years<br>2: between 40 and 50 years<br>3: ≥ 50 years |
| SBP | systolic blood pressure before transplantation | 80 – 240 mmHg<br>" . " is 'missing value' |
| DBP | Diastolic blood pressure before transplantation | 30 – 140 mmHg<br>" . " is 'missing value' |
| HPT | hypertension before transplantation | 1 = yes (SBP>140 and/or DBP>90)<br>0 = no<br>" . " is 'missing value' |
| DONOR | Type of donor | 1 = deceased donor<br>2 = living donor |
| DM | diabetes mellitus | 1 = present<br>0 = absent |
| VIT_STAT | vital status | 0 = alive<br>1 = deceased |
| FU | follow-up in years | 0 – 21 years<br>Until death<br>or until GF if GF precedes death<br>or till end of follow-up if no GF and no death |
| REJECTION | Rejection transplant / Graft Failure (GF) | 1 = yes<br>0 = no |
| TimeToRejection | Time to rejection | |
| PERIOD | period of transplantation | 1 = 1984 - 1988<br>2 = 1989 - 1993<br>3 = 1994 - 1997 |

a) Start SPSS and open *kidneytransplantation.sav.*

b) Inspect all variables.
For categorical variables, you can do this by going to **Analyze – Descriptive statistics – Frequencies**.
For continuous variables, you can compute the means and standard deviations: **Analyze – Descriptive statistics – Descriptives**.
- How many males are there among the participants of this study?

- What is the percentage of patients suffering from diabetes mellitus (DM)?

- What is the mean age at transplantation and what is the mean blood pressure?
- What is the frequency of rejection?
- How many subjects survive a graft failure? (crosstabs)
- How many subjects die without experiencing graft failure?
- What is the percentage of living donors?
- Is the number of living donors increasing over time?

Please fill in the evaluation form at page 135.

## Meeting 2

### *Lecture 2: Confidence intervals and hypothesis testing*

Teacher: Ton de Haan

An important role of statistics is drawing inferences (statistical inferences): making general statements (applying to a population) on the basis of a limited number of observations (study sample). To do so, inferential statistics quantify probabilities and uncertainty.

In the medical literature, two approaches of statistical inference are used: (hypothesis) testing and (parameter) estimation.
Estimating a (population) parameter (such as a mean, an effect, an odds ratio, a correlation or a regression coefficient) involves calculating an (sample) estimate of the parameter based on the sample data. To describe the uncertainty in the sample estimate (every sample will likely give another estimate), the error in this estimate is provided, which is called the standard error (SE). The standard error describes the typical deviation of sample estimates from the population parameter. To give an impression of the range in which the population parameter lies, confidence intervals (CI) are calculated from the standard error and the sample estimate.

Comparative research focuses on the question whether a difference in outcome exists between two or more conditions (or whether a correlation or association exists between the condition and the outcome). The basic question is whether the difference observed in the sample is a chance finding or strong indication of a systematic difference. This is quantified by the p-value.

### *Preparation for Meeting 3*

In preparation for next week:

- Please read (Field):
  - Sections 9.1, 9.2, 9.3.1, 9.4, 9.6 - 9.8
- Self diagnosis on Brightspace
  - Confidence intervals and hypothesis testing

## *Computer Session 2: T-test, confidence interval and P-value*

### *Background*

In medical research the research question often involves the comparison of means. This comparison is influenced by chance fluctuation. Different (random) samples will give different results. The influence of chance on the results can be taken into account by means of calculating confidence intervals and hypothesis testing.

### *Objective*

After completing the assignment, the student should be able

- to test for differences in population parameters in unpaired and paired designs, when the outcome variable is metric, and applies the t-test, the paired t-test, the Mann-Whitney U-test and the (Wilcoxon) Signed-Rank-test and indicates the conditions when these tests are valid.
- to calculate a 95%-confidence interval (CI) for the difference between two means.
- to interpret a confidence interval and a P-value.

### *Instructions*

This assignment consists of five topics.

### *Product*

Written answers and computer output. Please fill in the evaluation form at page 135.

### *Discussion*

Results can be discussed with your peers and with the tutor during the computer session.

### *Assignments*

**Diagnostic testing of dental caries**

Diagnosing dental caries by visual inspection provides many errors. Therefore, new diagnostic methods are being investigated, such as measuring electrical resistance of tooth enamel. To obtain a gold standard for dental caries, tooth enamel is removed to verify whether caries is present or not. For ethical reasons, this procedure is only admitted for molars that have to removed anyway, e.g. because of orthodontic reasons. A pilot study provided the following data on 16 occlusal surfaces of molars, just before extraction of these molars (*caries.sav*):

| Diagnosis after removal | Electrical resistance of the occlusal surface | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Caries | 4.3 | 5.1 | 2.6 | 4.7 | 3.8 | | | | | | |
| No carries | 4.3 | 8.6 | 7.2 | 6.7 | 5.1 | 6.6 | 6.3 | 6.4 | 7.9 | 6.9 | 5.8 |

a. Compare the caries and no caries groups using a parametric test for the difference in mean. Give the 95% confidence intervals.

b. Does measuring electrical resistance of the occlusal surface have added value in diagnosing caries or can the difference in the two groups be largely attributed to chance?

**HLA and the seriousness rheumatism.**
Carriers of the HLA-DR4 gene do have an elevated risk for developing rheumatism compared to humans not having this gene. A small study is performed to investigate whether carriers of this gene also develop a more serious form of rheumatism than non-carriers. In 18 patients the visible erosions in the hand joints were counted three years after start of treatment. The following numbers of erosions were found (*HLAandRheumatism.sav*):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Patient with HLA-DR4 | 8 | 12 | 0 | 51 | 15 | 9 | 8 | 22 | | |
| Patient without HLA-DR4 | 6 | 12 | 18 | 0 | 0 | 7 | 9 | 10 | 1 | 10 |

a. Calculate relevant summary measures for both groups.

b. Using graphics, give the distribution of the erosion of both groups (if possible in one figure).

c. What statistical test would you use to compare the number of erosions in both groups? Perform this test with SPSS

d. Give the 95% confidence interval for the difference?

**Parkinson-trial**
The well-known rule: 95%-CI = sample estimate ± 2 x SE is only valid if the observations are coming from an approximately normal distribution (this assumption doesn't matter too much when the samples are large). The LN-transformation often helps to arrive at normality, if the original distribution is skewed to the right (positively skewed). But what is the interpretation of the transformed values? To this end, consider the following example:
Parkinson patients receive DOPA when indicated. To avoid complications, the DOPA dosage is preferably kept low. In a clinical trial, patients are randomly allocated to the standard dose or 75% of the standard dose. After one month follow up, the handicap of the patient is assessed using the time needed to perform a certain locomotory task. The obtained (in seconds) results are in *parkinson.sav*.

a. Use SPSS to fill in the following table:

| UNTRANSFORMED | | | | | | LN-TRANSFORMED | | |
|---|---|---|---|---|---|---|---|---|
| Dose | n | mean | sd | skewness | coefficient of variation | mean | sd | skewness |
| Standard | | | | | | | | |
| 75% | | | | | | | | |

b. What shows that the LN-transformation has been successful?

c. After LN-transformation the mean is 2.96. Transforming to the original scale gives exp[2.96]=$e^{2.96}$=19.3. This is called the geometric mean. Give an interpretation of the geometric mean

d. Differences after LN-transformation can be interpreted as relative differences on the original scale.

Remember: $\ln(a)-\ln(b)=\ln\left(\frac{a}{b}\right)=\ln\left(1-\frac{(b-a)}{b}\right)\approx\frac{a-b}{b}$

The sd after LN-transformation can therefore be interpreted as the relative sd. What is a more common name for the relative sd?

e. Transforming back the difference in mean after LN-transformation gives $e^{3.22 - 2.96}$ = $e^{0.26}$ =1.29.
Give the interpretation of 1.29.

Calculate the 95%-CI for the ratio of the medians.

**KRIS trial**

Data from a trial to investigate the effect of clofibrate are available in dataset *kris.sav*. Male subjects were randomized to placebo or clofibrate. The effect parameter is change in cholesterol (mg/dL).

a. Compare the mean change in cholesterol between the two treatment groups. Give a 95% confidence interval for the effect of clofibrate.

b. Is there a difference in change in BMI between the two treatment groups? Do you need to correct the effect of clofibrate for the change in BMI?

c. Is there a difference in compliance rate between the two groups? Do you need to correct the effect of clofibrate for the difference in compliance rate?

**Tennis elbow**

Right-handed top tennis players often have a more powerful right arm than left arm. This can be observed from the difference in duration time (time-to-give-up in seconds) during stress tests for the separate arms. The following duration times were measured during a severe stress test in the best Dutch (right handed) tennis players (*Tennisplayer.sav and Tennisplayer_tr.sav*).

| Tennis player | Duration time right (seconds) | Duration time left |
|---|---|---|
| 1 | 72 | 61 |
| 2 | 41 | 38 |
| 3 | 51 | 47 |
| 4 | 64 | 55 |
| 5 | 57 | 49 |
| 6 | 46 | 40 |
| 7 | 42 | 38 |
| 8 | 68 | 59 |
| 9 | 55 | 48 |

a. Is there a difference left-right when using a test for independent observations?

b. Is there a difference left-right when using a test for paired observations?

c. Explain the difference between both tests. Which test do you prefer?

Please fill in the evaluation form at page 135.

## Meeting 3

### *Lecture 3: Linear regression analysis*

Teacher: Rogier Donders

The aim of many medical studies is to explain or predict a disease variable (dependent variable Y) from some other variable (independent variable X). Regression equations of the type Y = a+bX (a = intercept, b = slope) are the basis for such analyses. The lecture presents the assumptions and applications of regression analysis in several research situations as preparation for the assignments.

Field discusses the topic of regression in much more depth than is necessary for this course.
This lecture discusses the most important topics of multiple linear regression:

- Model assumptions and how to check them
- The use of dummy variables when categorical variables are involved
- Interpretation of the analysis results
- Multicollinearity
- Sample size

### *Preparation for Meeting 4*

In preparation for next week:

- Please read (Field):
  - Sections 19.1, 19.2, 19.3.1-19.3.3
    Remark: Never use (Yates') continuity correction.
  - Chapter 20 all parts that are marked or, but not 20.3.6.
    Section 20.3.5
    Smart Alex's Tasks, especially task 5 and 6 are instructive (in combination with the elaborations on the book's website).
- Self diagnosis on Brightspace
  - Linear regression

## *Computer Session 3: Linear regression*

### *Background*

To investigate whether two variables are related to each other one can calculate the correlation coefficient. This does however not indicate how the relationship goes, i.e. how the dependent value (y) can be predicted from the independent value (x). Such relation is given by the regression equation. The regression line is found using regression analysis (e.g. least squares method). What kind of regression analysis and the interpretation of the regression coefficients depend, of course, on the measurement levels of the variables.

### *Objectives*

- The student is able to indicate the level of association between two continuous measured variables using regression coefficients.
- The student is able to calculate a correlation coefficient between two variables using SPSS.
- The student is able to perform linear regression analysis using SPSS.

### *Instruction*

Make the assignments and make notes and save SPSS output.

### *Product*

Answers on the questions

### *Discussion*

Results can be discussed with your peers and with the tutor during the computer session.

### *Assignments*

**Calibration**

Regression analysis can be used to calibrate two measurements with each other. An example is the quantification of bone density using digitalized x-ray pictures. The average 'gray level' of x-rays of a number of bones with a known bone density (mg Calcium per $mm^2$) is calculated. Gray level ranges from 0 (white) to 256 (black). A high average gray level (e.g. 210) means that a lot of radiation hit the photographic plate, which means that little absorption of x-rays has taken place, i.e. when the bone density is low. A low gray level (e.g. 60) corresponds to a high bone density. The true bone density (= golden standard) is obtained by making an x-ray picture of a pig bone and subsequently determining the actual chemical composition of the bone. The needed variation in bone density is obtained by using bones of different density.

Dataset *calibration.sav* contains the outcome of a calibration experiment using 40 x-rays.

Answer the following questions:

a) Fill in the next table:

| | Mean | SD |
|---|---|---|
| Gray level | | |
| Bone density ($mg/mm^3$) | | |

b) The line of calibration have the following equation:
$$\text{Bone density} = \ldots\ldots + \ldots\ldots \times \text{Gray scale}$$
r=…….

c) What percentage of variance in bone density is explained by gray level (i.e. explained by the calibration line)?

d) An x-ray of a broken hip has an average gray level of 105 (after digitalization by computer). What is the estimated density of this bone? Approximately how large is the error in this estimation of bone density?

e) The oldest x-ray picture is a picture of Dr. Röntgen's wife's hand. This picture can also be digitalized. Because her hand had thin bones, the gray level was high: 170. What is the estimated density of the bones of her hand bones? What do you think of the error in this measurement?

**Osteoporosis**

The bone density (g/cm$^3$) of the lower arm of 600 middle-aged women in Apeldoorn was measured and their Calcium-intake (g/day) was measured using a diet-history-method. Regression analysis was used to quantify the relationship between Ca-intake and bone density. Data can be found in *osteo.sav*.

Answer the following questions:

a) What is the regression line of Bone density on Ca-Intake?

$$\text{Bone density} = \ldots\ldots + \ldots\ldots \times \text{Ca-Intake}$$

b) What is the interpretation of the value of the regression coefficient for Ca-Intake? What is the interpretation of the intercept?

c) The advice to the women of Apeldoorn to reduce osteoporosis is to drink an extra cup of low-fat milk (250 mg Ca). What effect on bone density would this advice have in due course according to the regression line?

d) What is the precision of the estimation of this effect?

e) Does the regression line give a valid estimate for the effect of drinking an extra cup of low-fat milk a day? Please provide your argument.

**Regression with a dichotomous predictor (independent variable)**

Linear regression is usually used to explain a quantitative variable (y) from a quantitative variable (x). However, linear regression can also be used with a dichotomous predictor variable.

For example: $\text{height} = a + b \times \text{gender}$ (gender: male = 0 or female = 1)

Roede & v.Wieringen measured the height of Dutch young adults in 1980:

| | Height | |
|---|---|---|
| | **Mean** | **SD** |
| Males: | 182.0 | 6.7 |
| Females: | 168.3 | 6.2 |
| Humans: | 175.2 | 9.3 |

a) Derive the point estimates for a and b in: $\text{height} = a + b \times \text{gender}$

b) What percentage of variance is explained by this regression line?

c) What is the simplest interpretation of the intercept a?
What is the simplest interpretation of the slope b?

**Weight of babies of one month old**

The SPSS dataset *weights.sav* contains data of 550 babies that were born in time. Besides weight (at one month), also gender, length, cranium circumference, education level and parity of the mother were recorded.
The variable descriptions and label values can be viewed in the *variable view* box.

Questions:

a) Is there a difference in the mean weight of boys compared to girls at one month? Provide a 95%-CI for the difference (hint: use an appropriate t-test).

b) Answer question a) using univariate linear regression analysis. Compare the result t with that obtained in a).

c) Explore the relation between length and weight using a scatter diagram, correlation coefficient and univariate linear regression.

d) Which percentage of the variance in weight at one month can be explained by differences in length?

e) Explore the association between cranium circumference and weight at one month.

f) Build a linear regression model to predict the weight at one month from length and cranium circumference. What is the variance explained by this model?

g) Examine the added value of including parity in your linear regression model. I.e. what does including parity as a predictor variable add to the predictive value of the model that already contains length and cranium circumference as predictors.

Please fill in the evaluation form at page 135.

## Meeting 4

### *Lecture 4: 2 x 2 table and logistic regression*

Teacher: Rogier Donders

*2 x 2 table*

***Probability, Odds, Odds Ratio***

**Proportion**: The proportion of patients with a certain characteristic is the number of patients with the characteristic divided by the total number of patients.

**Probability**: The probability that a patient will have a certain characteristic is the number of patients with the characteristic divided by the total number of patients. So, proportion = probability and

$$\text{Probability} = \frac{\text{events}}{\text{events} + \text{non-events}}$$

For example, imagine we have 10 subjects, 2 whom had a myocardial infarction. These subjects would have the scores [0,0,0,0,0,0,0,0,1,1] on the binary outcome variable 'myocardial infarction'. The average of these scores is 0.2. Likewise, the proportion of subjects with a myocardial infarction in our sample is 0.2 and the probability that a subject in our sample would have a myocardial infarction is 0.2.

**Odds**: The odds are just an alternative way of expressing the likelihood of an event. Whereas the probability is the number of patients with the characteristic divided by the total number of patients, the odds would be the number of patients with the characteristic divided by the number of patients without the characteristic.

$$\text{Odds} = \frac{\text{events}}{\text{non-events}}$$

**Odds ratio**: The odds ratio is a widely used statistic to compare the frequency of exposure to risk factors in epidemiological studies. The odds ratio can be used to compare the odds (of an event) between groups of patients. The odds ratio is simply the odds of the event occurring in group 1 divided by the odds of the event occurring in group 2.

$$\text{Odds ratio} = \frac{\text{group 1 odds}}{\text{group 2 odds}}$$

For example, suppose we had two groups of patients: one with a family history of heart disease and the other without.

| | | Myocardial infarction | | |
|---|---|---|---|---|
| | | yes | no | total |
| Family history | with | 4 | 6 | 10 |
| | without | 2 | 8 | 10 |

Odds of myocardial infarction in 'with family history' group = 4/6 = .67
Odds of myocardial infarction in 'without family history' group = 2/8 = .25
The odds ratio is .67/.25 = 2.68.

Interpretation: So, in this fictitious example, the odds of a myocardial infarction were

2.68 times greater in patients with a family history of heart disease than in those without a family history.

**Relative Risk and Risk Difference**
We can also calculate the Relative Risk (RR): The risk in the family history group is 0.4 and in the other group it is 0.2, so the RR=0.4/0.2=2.
Yet another way to present the results is to calculate the Risk Difference (RD). Obviously RD=0.4-0.2=20%.

**Relation between Relative Risk and Odds ratio**
It can be shown that the following relation holds:

$$\text{Relative Risk}=\frac{\text{risk in the exposed group}}{\text{risk in the control group}}=\frac{p_1}{p_0}=\frac{\text{Odds ratio}}{(1-p_0)+p_0(\text{Odds ratio})}$$

So, if the disease is rare, meaning that $p_0$ is small, that the denominator is almost 1. In other words, the odds ratio is an approximation of the relative risk. For common diseases with known $p_0$ the relative risk can be calculated from the odds ratio and $p_0$ using the formula above.

**Conversion between Odds and probabilities**
For the group with a family history, the odds were the number of events divided by the number of non-events, i.e. 4/6=0.67.

But, as the number (or percentage) of events corresponds to the chance of an event and as the number of non-events corresponds to the chance of no event, the odds are also equal to the probability of an event divided by the probability of no event. The probability of an event is often indicated by p (p between 0 and 1), so the probability of no event is 1-p.
Hence:

$$\text{Odds}=\frac{\text{probability of an event}}{\text{probability of no event}}=\frac{p}{1\text{-}p}$$

The probability p of MI in the family history group was 4/10=0.4, so the odds are 0.4/0.6=0.67, as we already saw.

As we shall see, the results of some statistical methods are expressed as odds. In that situation, you may want to calculate p. Using some algebra, it can be shown that:

$$p=\frac{\text{odds}}{1+\text{odds}}.$$

For the MI example, we can verify this: in the family history group, the odds were 0.67, so p=0.67/(1+0.67)=0.4 and this indeed was the risk of MI.

## *Logistic regression analysis.*

Last week we used linear regression to examine the relationship between a quantitative outcome variable (e.g., temperature, blood pressure, IQ) and a set of predictors that were either quantitative of binary . However, in many studies the outcome variable of interest is the presence or absence of some condition, such as responding to treatment or the occurrence of a myocardial infarction (MI). We cannot use ordinary multiple (linear) regression for such data, but instead we can use a similar approach known as multiple logistic regression.

The basic principle of logistic regression is much the same as for ordinary multiple regression. The main difference is this:

- In multiple regression, we developed a model that uses a combination of values from a group of explanatory variables to predict the ***value*** of a (quantitative) outcome (e.g., age + weight + pulse rate = SBP).
- In logistic regression, we will develop a model that uses a combination of values from a group of explanatory variables to predict the ***probability*** of a particular (binary) outcome (e.g., age + weight + pulse rate = likelihood of a myocardial infarction, where 0 = no MI and 1 = MI).

Remember that probabilities range from 0 (the event will definitely not occur) to 1 (the event will certainly occur). One reason that we cannot use ordinary linear regression is that linear regression might predict impossible probabilities outside the range 0 to 1.

If we have a binary variable, it is advisable to assign the categories numerical values of 0 and 1 (representing 'No' and 'Yes' respectively). In that case, the mean of the variable in a sample of individuals is the same as the *proportion* of individuals with the characteristic (= the probability of having the characteristic).

We might expect, therefore, that the appropriate regression model would predict the proportion of subjects with the feature of interest (or, equivalently, the probability of an individual having that feature) based on the combination of explanatory variables in the model. Unfortunately, the standard method is to estimate the odds, or, more precisely, the natural logarithm of the odds. The later value is usually called the logodds, or the 'logit'. The following example illustrates this.

**Example**
When use logistic regression to analyse the relationship between nodal involvement, age and tumor size in the prostate cancer data base, the results are:

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | age | -,053 | ,051 | 1,092 | 1 | ,296 | ,948 | ,858 | 1,048 |
| | size | 1,682 | ,640 | 6,918 | 1 | ,009 | 5,377 | 1,535 | 18,835 |
| | Constant | 1,703 | 3,018 | ,318 | 1 | ,573 | 5,489 | | |

## Interpretation of the coefficients: odds and percentages

Based on the column B (the regression coefficients), we have:

Logit=log(odds)=1.703 -0.053*age + 1.682*size

Hence, for a 60 year old man with a large tumor:

Logit=log(odds)=1.703 -0.053*60 + 1.682*1=0.2,

So the odds are exp(0.2)=1.22 and the risk of nodal involvement is

p=odds/(1+odds)=1.22/(1+1.22)=0.55=55%

For a 60 year old man with a small tumor:

Logit=log(odds)=1.703 -0.053*60 + 1.682*0=-1.5,

So the odds are exp(-1.5)=0.22 and the risk of nodal involvement is

p=odds/(1+odds)=0.22/(1+0.22)=0.18=18%

**Graphs**

The Figures 1-3 illustrate the above calculations.
We had: Logit=log(odds)=1.703 -0.053*age + 1.682*size, so in fact, we have two lines, one for the patients with a small tumor (size=0):

Logit= 1.703 -0.053*age + 1.682*0=1.703 -0.053*age,

and one for the patients with a large tumor:

Logit= 1.703 -0.053*age + 1.682*1=3.385 -0.053*age.

Figure 1 shows these (parallel) lines. The values for age=60 are highlighted.
The Figures 2 and 3 show the odds and the p, respectively.

The curves for the p value are parts of so called logistic curves, or sigmoid curves. Figures 4 and 5 show examples of more complete logistic curves with the values for p ranging from almost 0 to almost 1. As shown, the curves may differ in steepness and they may be increasing or decreasing.

**Figure 1**

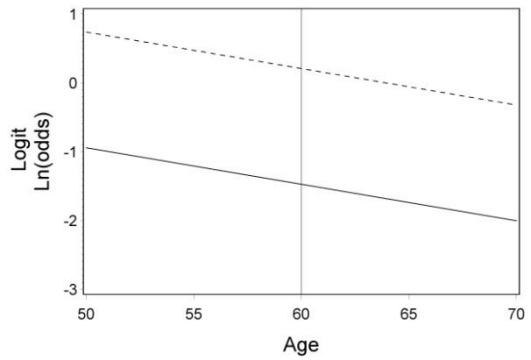



**Figure 2**

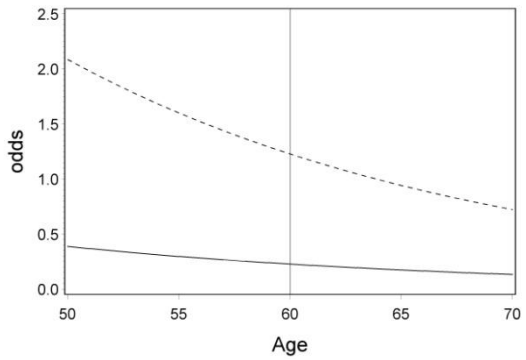



**Figure 3**

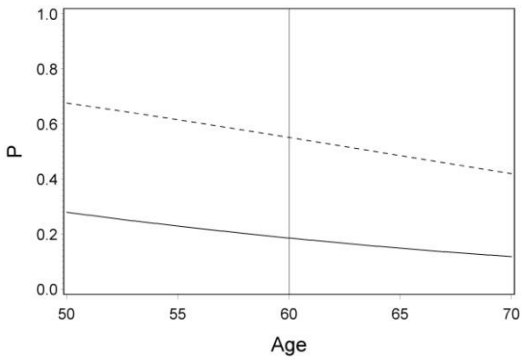


**Figure 4**

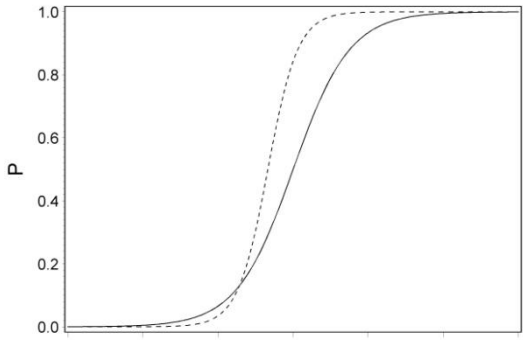
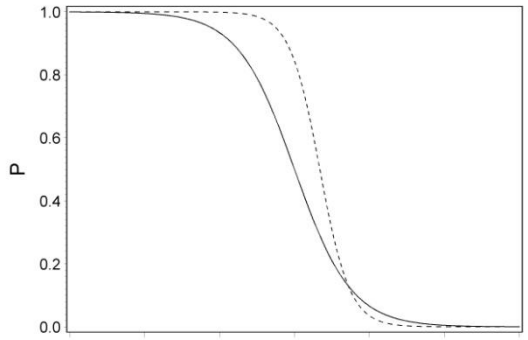



**Figure 5**




## Interpretation of the coefficients: odds ratios

The coefficient B of a variable is the logarithm of the odds ratio when the variable increases one unit. Hence exp(B) is the odds ratio. According to the results of the logistic regression, Exp(B) for tumor size is approximately 5.4, so the odds ratio for a large tumor is 5.4.
We can verify this from our earlier calculations: We saw that for a 60 year old with and without a large tumor, the odds were 1.22 and 0.22, respectively. Hence the odds ratio based on these findings should be 1.22/0.22=5.5.
Apart from a small difference due to rounding off, this is exactly exp(B).

The coefficient B for age is the slope of the curves in fig. 1. As B=-0.53, the logodds decreases by 0.053 when the variable on the horizontal axis, age, increases by 1. Similarly, Exp(B) for age is 0.948, so when a man gets one year older, this corresponds to an odds ratio of 0.95. (This seems to suggest that older men have less risk, but this is far from sure, because the upper limit of the confidence interval for Exp(B) is 1.048, so the odds ratio may be as high as 1.05).

To calculate the effect of 10 years difference, you have to use the coefficient B: B for age is -0.053, so the slope is -0.053: if the value on the horizontal axis (age) increases by one, the value on the vertical logOdds axis changes by -0.053 (so it decreases). A difference of 10 years then corresponds to 10*-0.053=-0.53 on the logOdds axis. As exp(-0.53)=0.58, the odds ratio associated with an age difference of 10 years is 0.58.

### Events versus non-events

Finally, when the odds of an event are 2, for example, the odds of no event are the inverse: 1/2=0.5. Similarly, the odds ratio for no events is the inverse of the odds ratio for events. Unfortunately, statistical packages (including SPSS) do not always correctly select what is an event and what is not. Even when events are coded 1 and non-events are coded 0, the package may ‘decide’ that code 0 represents an event. (When events are coded in a different way, for example ‘yes’ versus ‘no’ or 1 versus 3, most statistical packages will still carry out the analysis, but it may not be clear what was considered an event.)
Usually there are options to control this and to specify which code presents an event, but sometimes this is not straightforward.

Therefore, you should always check your analysis results by means of cross tables. Trends of increasing risk should result in odds ratios larger than 1 in the analysis (and positive Bs). If this is not the case, you may either rerun the analysis, after having made sure that the statistical package uses the correct event, or you may simply recalculate the odds ratios by hand, by taking the inverse: 1/odds ratio instead of odds ratio (and –B instead of B).

Of course, the most important reason to verify your results by cross tables is that it will give you some idea about the fit of your model and whether you have not made any serious mistakes. The tables should globally reflect the results of your analysis.

## *Preparation for Meeting 5*

In preparation for next week:

- Please read (Field):
  - Section 2.9.8; (if time left 3.1-3.7)
- Self diagnosis on Brightspace
  - Chi-square test and logistic regression

## *Computer Session 4: 2 x 2 table and logistic regression*

### *Background*

If the independent / predictor variable is dichotomous, a 2*2 table is the simplest way to represent the relation. For example, if we want to compare the outcomes (success/failure) of two treatments, or if we want to determine whether there is an association between an exposure factor of interest and the occurrence of a particular disease. In such cases, the association can be described using *risk differences*, *risk ratios*, *relative risks* or *odds ratios* (OR). The OR is not the most intuitive measure, but it is the only one that can be estimated when a case-control design has been used. Furthermore, the odds ratio is very suitable for statistical modeling. To study the relation between a binary outcome (i.e. recovered or nor recovered) and a predictor (i.e. age of a patient, blood pressure) logistic regression is most suitable.

### *Objectives*

Upon completion of this assignment, the student should be able to

- Analyse a 2x2 table using SPSS
- Calculate risk differences, risk ratios, relative risks and odds ratios.
- perform a logistic regression analysis using SPSS in order to quantify the relation between a dichotomous dependent variable and an explanatory variable
- interpret the results
- formulate the conclusions

### *Instruction*

Make the assignments and make notes and save SPSS output.

### *Product*

Answers on the questions

### *Discussion*

Results can be discussed with your peers and with the tutor during the computer session.

### *Assignments*

**The relationship between clinical characteristics and the risk of metastasis for patients with prostate cancer**

The treatment regimen to be adopted for patients who have been diagnosed as having cancer of the prostate is crucially dependent upon whether or not the cancer has spread to the surrounding lymph nodes. Indeed, a laparotomy, a surgical incision into the abdominal cavity, may be performed to ascertain the extent of this nodal involvement. There are a number of variables that are indicative of nodal involvement and that can be measured without the need for surgery, and the aim of an observational study reported by Brown (1980) was to determine whether a combination of five variables could be used to determine whether or not the cancer has spread to the lymph nodes. The five variables were:

- age of patient at diagnosis (in years)
- level of serum acid phosphatase (in King-Armstrong unites)
- the result of an X-ray examination (0 = negative, 1 = positive)
- the size of the tumor as determined by a rectal examination (0 = small, 1 = large)
- a summary of the pathological grade of the tumor determined from a biopsy (0 = less serious, 1 = more serious, 2 = very serious)

The values of each of these variables were obtained for a group of patients presenting

with prostatic cancer who had undergone a laparotomy. The result of the laparotomy is a binary response variable where zero signifies the absence of, and unity the presence of nodal involvement. The data are stored in the SPSS file *ProstaticCancer.sav*.

Later on in the course, we will investigate in detail whether these variables indeed predict nodal involvement. This week we try to answer two more simple questions:

- is there a relationship between size of the tumor and nodal involvement
- is there a relationship between the pathological grade of the tumor and nodal involvement

1. Which percentage of the patients has *nodal involvement*?

2. Use a cross-table to assess whether the percentage of patients with *nodal involvement* is larger in the group patients with a large size tumor. Calculate the Odds Ratio (OR), the Relative Risk (RR) and the Risk Difference (RD). carry out a chi-squared test (use the options: statistics>chisquared and >risk). Is there a strong relationship? Why? Is tumor size a good predictor of nodal involvement? Why? What would you use to report the result of the analysis: OR, RR, RD, or something else?

3. In order to assess whether the percentage of patients with *nodal involvement* is larger in the group patients with a more serious or very serious biopsy results compared to patients with less serious results a new variable (a 'dummy' variable) has been created in the database: 'grade_serious'. It has value 0 when the biopsy result was less serious and it has value 1 when it was either more serious or very serious.
   Construct a cross table and carry out a chi-squared test. Is there a strong relationship? Why? Is a more serious or very serious biopsy result a good predictor of nodal involvement? Why?

4. Assess whether the percentage of patients with *nodal involvement* is larger in the group of patients with a very serious biopsy results compared to the other patients:
   - First create a dummy variable that indicates whether a patient had very serious

biopsy results.
- Construct a cross table to check whether your dummy is correct.
- Analyze the relationship between the dummy and nodal involvement: carry out a test and calculate the OR, the RR and the RD. Is there a strong relationship? Why? Is a very serious biopsy result a good predictor of nodal involvement? Why? What do you conclude based on this analysis?

5. The chi-squared test can also be used for larger tables.
   Construct a 2X3 table for the variables biopsy result and nodal involvement. Carry out a test.

   It would be interesting to investigate whether there is a trend: is the risk of nodal involvement larger when the results of the biopsy are worse. Does the chi-squared test or the Fisher exact test investigate this? Why? Which type of analyses is required to investigate trends? (Hint: How is the approach when the outcome is continuous: when is analysis of variance used and when is linear regression used?)

**Analysis of the Compliance Trial**

In order to evaluate the efficacy of a program for promoting compliance with compression therapy and prescribed exercise in leg ulcer patients, a randomized trial has been carried out. Half of the patients received usual care, the other half received special support according to the program.
The data set *DetailedCompliance.sav* contains the main results of the trial.

1. Is there a statistically significant difference between the groups? (The data set does not contain individual patient data, but frequencies. Hence, use data>weight cases> by frequency.)

2. In the previous example, we reduced the 2X3 table for biopsy results to two 2X2 tables. For example, we compared ‘very serious’ to ‘less serious’ and ‘more serious’ combined.
   - In a similar way, compare the percentage of patients with sufficient compliance in each treatment group to the percentage of patients with at best poor compliance. Calculate the Odds Ratio.
   - In a similar way, calculate the odds ratio for the percentage of patients with ‘sufficient’ compliance in each treatment group and the percentage of patients with ‘poor’ compliance (Use the option: data>select cases>if condition satisfied). What is your interpretation of this analysis? Is it meaningful for the evaluation of the treatment?

The statistical paragraph of the protocol of the trial specified that the primary outcome was the percentage of patients that had ‘sufficient compliance’. The odds ratio and its 95% confidence interval had to be calculated.

The data base *Compliance.sav* contains individual patient data for the compliance study.

3. Logistic regression can be used to construct confidence intervals for the OR: In SPSS, select ‘regression’; ‘binary logistic’; dependent variable ‘compliance’ and covariate ‘group’. In ‘options’ tick ‘CI for exp(B)’.
   In the output, exp(B) is the OR, so the OR for the treatment is 0.94, with confidence interval from 0.47 to 1.88, so this is the result of the primary analysis of the trial. Note that the confidence interval is identical to the confidence interval produced by ‘crosstabs’method. What is the interpretation of this result?

4. What is the compliance percentage for men (irrespective of the group)? And for women?

   Also calculate the odds ratio (including confidence interval) for compliance versus sex.

5. Would it be interesting or worth wile to repeat the analyses separately for men and women? Why?

6. How would you report your findings?

In an additional analysis, the impact of age and sex had to be analyzed by adding these variables to the model.

7. Carry out the additional analysis. Compare the result to the result of the primary analysis. Conclusion? What would you report in your article?

8. Interpretation of the coefficients:
   -What is the interpretation of the regression coefficient B and the estimate exp(B) for the variable ‘sex’?

   -What is, according to this analysis, the odds ratio between the treatment groups for men? And for women?

   -What is the interpretation of the regression coefficient B and the estimate exp(B) for the variable ‘age’?

9. According to the logistic model, what percentage of the women aged 60 who follow the compliance program is sufficiently compliant?

   You may wonder how good the model fits the data. How precise and accurate is it?

   Would it be possible to check whether this estimated percentage for women of 60 years old who follow the program according to the model is approximately equal to the percentage that is *actually found* in the study? How? Can you do it?

10. Make a new variable age_decade=age/10 and repeat the additional analysis. What are B and exp(B) for the variable age_decade. What is their interpretation? Why is it sometimes useful to change the scale of a variable: from years to decades, from cm to m, from mg to kg, etc.?

Please fill in the evaluation form at page 135.

# Meeting 5

## *Lecture 5: Sample size and power*

Teacher: Ton de Haan

*Summary*

The inclusion of new patients in scientific research in patients (or subjects/animals) must be properly regulated. Medical ethical committees (CWOM Commission scientific research in humans) and the Ethics Committee for laboratory animals (DEC) more and more ask for a justification of the chosen sample size. The sample size is indeed an ethical issue: Unnecessarily many patients included in research should be avoided because of unnecessarily exposure to risks and burden, but too small samples leave the investigator with unanswered questions. (The study can still be used as a part of a larger study or meta analyses)

The ethical committees are often not satisfied with a sample size, that is based on habits in the literature (everyone does it this way) or on the availability of resources (such as time and money). The committees ask for good statistical arguments for the proposed sample size. There are two methods possible to achieve a meaningful sample size:

:

A: During the course of the study it will be determined when enough information is collected. Regularly statistically testing determines whether the study can be terminated (interim analysis). Such a procedure is called a sequential procedure. A key issue in sequential testing is the prevention of false significant results because of multiple testing. A sequential procedure is only applicable if the statistical analysis can be quickly executed. In research with a long follow-up period, for example, this is not the case. The intake of the patients is finished when the first results are becoming available to the researcher.

B: The sample size will be determined according to the objectives of the study (clinical relevance) and the random fluctuation of the outcome. The sample size should be such that statistically significant also means clinically relevant.

This course is limited to situation B: calculating in advance the necessary sample size when comparing two treatments.

A variation is the so called power analysis. Given the sample size, the probability to find a significant difference is calculated, assuming a 'real' difference between the populations.

To be able to calculate a sample size, one need to know what statistical test will be used. Furthermore, one needs to make a choice for the four parameters below. A simple formula than gives you the sample size.

$\alpha$ = significance level of the test.
Mostly used $\alpha$ = 0.05, two sided. This results in: $Z_{\alpha/2}$ = 1.96

1-$\beta$= power of the test.
The power is the probability that the test is significant when the assumed effect exists. Popular values for the power are 0.8 or 0.9, resulting in $Z_\beta$ = 0.84 or $Z_\beta$ = 1.282.

$\delta$ = Relevant effect (assumed effect).
If in reality the effect is smaller than $\delta$, the investigator doesn't care that no significant result is found, but if in reality the effect is larger than $\delta$ the test should be statistically significant.

$\sigma$ = noise = unexplained variation = residual standard deviation.
A reasonable value for $\sigma$ is often available in the literature. Otherwise, a pilot study should be carried out resulting in a reasonable estimate of $\sigma$.

If all parameters are chosen and the statistical procedure is known, the sample size

follows from the correct formula.
The next text gives an overview for simple designs (based on Chapter 8, Sample Size Determination from: Leslie E Daly & Geoffrey J Bourke, *Interpretation and uses of Medical Statistics*, fifth edition, Blackwell Science, Oxford, 2000).

## *Sample Size Determination*

When planning any investigation, one of the most common problems is determining what sample size is required. Unfortunately, the question 'How many animals should I study?', though simple, does not have an easy and quick answer. A lot of information is required before a sensible answer can be provided.
This chapter describes some of the methods used to determine the best sample size for single- group descriptive studies and for two-group comparative studies and gives some simple formulas that might help the investigator. An understanding of the concepts of power and significance in hypothesis tests and the ideas underlying estimation with confidence intervals are prerequisites for the material that follows.

## *Sample sizes for single-group studies*

If the objective of a study is simply the estimation of the value of some parameter in terms of, say, a percentage or a mean, then an appropriate sample size can be determined on the basis of the precision of the sample estimate or, in other words, on the width of its confidence interval. It must be stressed that the following approach is only suitable if no comparisons are involved.

### *Estimating a single mean:*
A 95% CI for a mean is given by:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

where *n* is the sample size and $\sigma$ is the population standard deviation. If the '$\pm$' part of the 95%CI is required to be $\pm\delta$, then the sample size needed is easily seen to be:

$$n \geq \left(1.96 \frac{\sigma}{\delta}\right)^2 .$$

The largest whole number greater than the expression is taken. For a 99%CI, 1.96 is replaced by 2.576. Of course, it is necessary to have an estimate of the population standard deviation $\sigma$, but this can be based on a pilot study or on prior investigations with the variable in question.
*Example:*

> *Suppose that it is required to estimate diastolic blood pressure in a population to within $\pm$2mmHg (using a 95% confidence interval) and the standard deviation of systolic blood pressure was known to be 15mmHg.*
>
> *Then:* $n \geq \left(1.96 \frac{15}{2}\right)^2 = 216.1.$
>
> *The next highest integer is taken, giving a requirement of 217 subjects.*

### *Estimating a single proportion or percentage:*
An identical approach is used for determining sample size to estimate a proportion with a pre-specified degree of precision.
The 95%CI for a proportion (*p*) is:

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}} .$$

Pre-specification of a precision of $\pm\delta$ at a 95% confidence level gives the sample size:

$$n \geq (1.96)^2 \frac{p(1-p)}{\delta^2}.$$

For a 99%CI, the 1.96 would be replaced by 2.576. Notice that, unlike the situation with estimating a mean, no prior estimate of a standard deviation is required. The precision is instead affected by the observed proportion and thus the sample size depends only on the likely result and the desired precision.

*Example:*

> *Suppose it is thought that there are about 28% smokers in the population and it is required to estimate the percentage of smokers within $\pm$ 3% (in absolute terms), using a 95% confidence interval.*
>
> *Converting everything to proportions:*
>
> $$n \geq (1.96)^2 \frac{0.28(1-0.28)}{(0.03)^2} = 860.5$$
>
> *so that a survey of 861 persons is required, even before non-response is allowed for.*

### *Sample sizes specifications for two-group comparisons*

When interest lies in the comparison of two groups, it is not sufficient to base sample sizes on the precision of any estimates. Usually, a two-group study is designed to detect if there is a real difference between the populations from which the groups were sampled. This real difference should be distinguished from any difference that might be actually observed in a study based on samples from the populations. A null hypothesis of no group difference in the population is postulated and, after study completion, a significance test is performed to indicate how the observed (sample) difference between the groups should be interpreted. Either the observed result is likely to reflect a real difference between the populations and the null hypothesis is rejected – a statistically significant result – or the result is explicable by sampling variation and the null hypothesis is not rejected – a non-significant result. The sample size in a comparative study should be large enough for the study to be capable of detecting important (real) differences between the groups by finding them statistically significant. In what follows we discuss the input needed for sample estimation in a two-group comparative study.

<u>The significance level:</u>

The chances of obtaining a statistically significant result (at a given significance level) depend on the magnitude of the real difference between the groups (in the population) and on the sample size. If there is a real difference, the investigator hopes for a significant result and, if there is no real difference, a non-significant result is required. Obviously, however, if the sample size is too small, a non-significant result may be obtained in either situation and the study may be unable to distinguish between a possible real effect and sampling variation.

The investigator protects herself against falsely deciding there is a difference (making a Type I error) by setting an appropriate significance level for the test. Typically, the investigator accepts a 5% chance that she might wrongly reject the null hypothesis (when it is true) and concludes that a real difference exists, when in fact there is none. Thus the significance level of the test is also referred to as the chance of making a Type I error.

<u>The minimum difference to be detected:</u>

The investigator must also guard against missing a real difference if it exists by falsely declaring a non-significant result. This is called a Type II error. From a clinical

perspective, however, if the real difference between the groups is very small, it is probably not worth knowing about in the first place and a non-significant result would not matter. This is perhaps the most difficult part of sample size determination. The investigator must decide what size of difference she wants to be able to detect. This is variously called 'the minimum worthwhile difference' or 'the minimum difference to be detected'. It is chosen, not on any statistical grounds, but on the basis of clinical importance. If the true difference (in the population) is this large or larger, the study must be big enough to detect it (with a statistically significant result). If the true difference is smaller than this value, it is accepted that the study might miss it (with a non-significant result).

The power of the study:
Of course, one can never guarantee that a study will detect a particular difference between groups. All that can be done is to set a level of probability that this difference will be detected. This is called the power of the study – the chance of detecting a particular real difference, if it exists. Power is the complement of the Type II error. If the power to detect a real difference is set at 80%, then there is a 20% chance that this real difference could be missed. To calculate the required sample size, the investigator must specify the power she wants to have to detect this minimum worthwhile difference. If the true difference is greater than that specified, the chances of detecting it are greater than the chosen power; if the true difference is smaller than that specified, the power to detect it is lower. This is acceptable, since it is the *minimum* difference to be detected that is specified.
While 5% and 1% levels of significance have become standard in medical research, there is less universal agreement on what levels of power to choose. Generally, power levels of 80%, 90% and 95% are used, with a useful rule of thumb that the size of the Type II error should be around four times that of the (two-sided) significance level chosen. Thus, 80% power and 5% significance is the most common combination in sample size determination, while 95% power is often paired with 1% significance.

*Sample sizes for the comparison of two independent groups.*

*Comparing two means*
For the one sided comparison of means in two independent samples:

$$n \geq \frac{2\left(Z_\alpha + Z_\beta\right)^2 \sigma^2}{\delta^2}$$

where
*n* is the sample size required in each of the groups,
$\alpha$ is the significance level,
$\beta$ such that $1-\beta$ is the power,
$Z_\alpha$ ( $Z_\beta$ ) the standardized normal deviate exceeded with probability $\alpha$ ( $\beta$ ), (see figure)
$\sigma$ the population standard deviation of the outcome variable (by assumption the same in each group),
$\delta$ the minimum difference to be detected.

In case of two sided comparison $Z_\alpha$ has to be replaced by $Z_{\alpha/2}$.

*Example:*
*Suppose a sample size is required for a randomized trial to compare two*

*different treatments for blood pressure reduction. The researchers want to have a large chance of detecting, as statistically significant, a real (population) difference between the mean systolic blood pressures in the two groups of 10mmHg or greater. The researchers decide on a two-sided significance test at a 5% level and a power to detect the treatment effect of 80%. An estimate of the standard deviation of systolic blood pressure is 20mmHg.*

*Then:* $n \geq \frac{2(1.96+0.84)^2 * 20^2}{10^2} = 62.8$

*This is the sample size required in each group and should be rounded up. Thus a total of 126 are required for the study – 63 in each group.*

*Comparing two proportions or percentages:*

For the comparison of proportions in two independent groups:

$$n \geq \frac{\left(Z_\alpha + Z_\beta\right)^2 \left(\pi_1(1-\pi_1) + \pi_2(1-\pi_2)\right)}{\left(\pi_1 - \pi_2\right)^2}$$

where

*n* is the sample size required in each of the groups,

$\alpha$ is the significance level,

$\beta$ such that $1-\beta$ is the power,

$Z_\alpha$ ( $Z_\beta$ ) the standardized normal deviate exceeded with probability $\alpha$ ( $\beta$ ),

$\pi_1$ ($\pi_2$) are the proportions in the two groups which define the minimum difference to be detected.

In case of two sided comparison $Z_\alpha$ has to be replaced by $Z_{\alpha/2}$.

(This formula is somewhat inaccurate with small *n*, particularly if the smallest of $\pi_1, \pi_2, 1-\pi_1$ and $1-\pi_2$ multiplied by *n* is less than 5).

**Sample sizes for the comparison of two dependent groups.**

*Sample sizes for the comparison of paired means*

The usual formula given for comparing means in a paired situation is:

$$n \geq \frac{\left(Z_\alpha + Z_\beta\right)^2 \sigma_d^2}{\delta^2},$$

*n* is the number of pairs,

$\sigma_d$ is the population standard deviation of the differences,

$\delta$ the minimum difference in means to be detected.

Example:

*To investigate the effect of feeding* $10\mu g$ *of vitamin* $B_{12}$ *per pound of ration to growing swine an experiment is designed in which 2 animals from each of n litters of swine is taken. The animals of each litter are assigned at random to vitamin* $B_{12}$ *or not. The researchers want to have a large chance of detecting, as statistically significant, a real (population) effect of vitamin* $B_{12}$ *on daily weight gain of 0.1 lb or greater. The researchers decide on a two-sided significance test at a 5% level and a power to detect the vitamin effect of 80%. An estimate of the standard deviation of the difference in daily weight gain per pair of swine comes from another comparable experiment and can be taken as 0.2 lb/day.*

*The required sample size is:*

$$n \geq \frac{(1.96+0.84)^2 0.2^2}{0.1^2} = 31.4$$

*Thus a total of 64 swine is required, coming from 32 litters.*

*Figure: Relation between $\alpha$ and $Z_\alpha$*

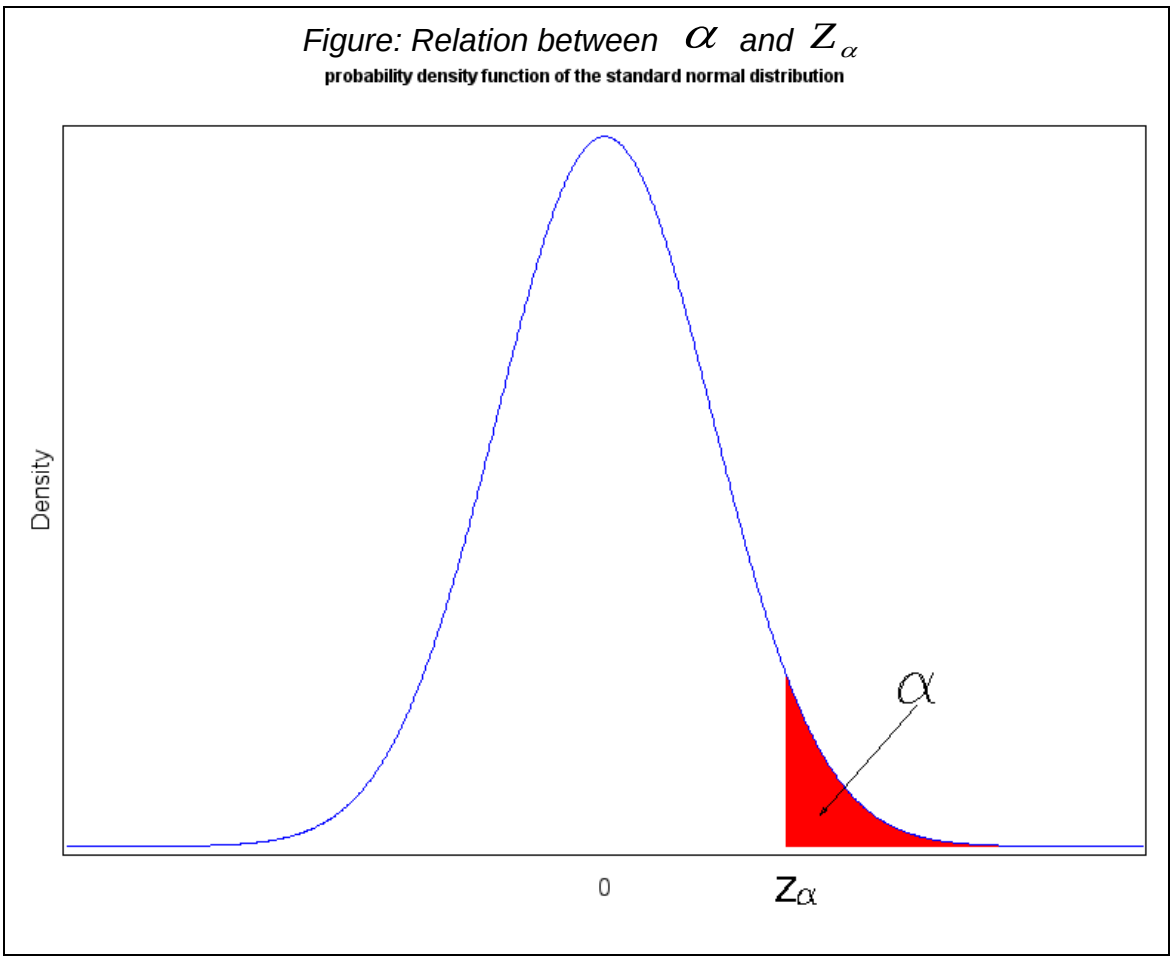



Table: some frequently used standardized normal deviates

| $\alpha$ | $Z_\alpha$ |
|---|---|
| 0.005 | 2.58 |
| 0.025 | 1.96 |
| 0.05 | 1.65 |
| 0.10 | 1.28 |
| 0.20 | 0.84 |

## *Preparation for Meeting 6*

In preparation for next week:

- Please read Field:
  - Chapter 12 of Field (Comparing Several Means: ANOVA (GLM 1)) contains many technical details that may be omitted, such as sections: 12.2.2-11.2.6, 12.4-.12.11
  - 12.14 (Key Terms That We've Discovered): omit all except:
    Analysis of Variance (ANOVA)
    Bonferroni correction

Pairwise comparisons
Post hoc tests
- Sections 14.1 – 14.3 till 14.3.1 (not 14.3.1)

❖ Self diagnosis on Brightspace
- Sample size calculations and power

## *Computer Session 5: Sample size and power*

*Background*
The calculation of the sample size in clinical trials is simple if the parameters α, β, σ, and δ are known. But to find reasonable values for these parameters is often difficult.

*Objectives*
The student can calculate the sample size for the most common study designs.
The student is able to handle the most important parameters:

- The clinical (or biological) relevant effect
- The natural variation in the outcome measurements
- The design of the study (paired/parallel group)

*Instruction*
Make the assignments and make notes of the arguments used for the choice of the parameters. Make use of Russ Lenth's website:
http://homepage.stat.uiowa.edu/~rlenth/Power/index.html.

*Product*
Answers on the questions of the assignments.

*Discussion*
Results can be discussed with your peers and with the tutor during the computer session.

*Assignments*

**Laser beams**

After surgical removal of a wisdom tooth the patient often suffers from post-surgery pain, bleeding or inflammation. The idea exists that one minute radiation with a so called soft laser would reduce the duration of the adverse events. Using a clinical trial this idea should be examined.
To gain experience with this kind of research a pilot study was performed. In 6 patients with two removable wisdom teeth in the same jaw, the extraction of the teeth was performed with and without the laser treatment. The results of the duration of complaints are as follows.

| | | Duration of complaints (hours) | | |
|---|---|---|---|---|
| | Patient | with laser | without laser | difference |
| | 1 | 50 | 48 | -2 |
| Results | 2 | 12 | 20 | 8 |
| Of | 3 | 16 | 10 | -6 |
| Pilot | 4 | 60 | 68 | 8 |
| Study | 5 | 24 | 28 | 4 |
| | 6 | 34 | 34 | 0 |
| | Mean | 32.7 | 34.7 | 2.0 |
| | SD | 19.1 | 20.8 | 5.7 |

The investigators have to decide between a parallel group design (from each patient one wisdom tooth enters the experiment) or a split mouth design (each patient two wisdom

teeth are used, one with laser treatment, one without laser treatment).
Answer the next questions:

a) If the parallel group design is chosen, what are reasonable values for $\alpha$, $\beta$ (or power=1-$\beta$), $\delta$ and $\sigma$?
   Calculate for this situation the required sample size
   Remember: - one or two sided testing
   - what test should be used?

b) Calculate the power of the study with $\delta$ and $\sigma$ as chosen in a), but with sample sizes n1=300 and n2=300, n1=200 and n2=400, for n1=150 and n2=450, n1=120 and n2=480, for n1=100 and n2=500 (Note that the total sample size is 600 for each combination). Discuss reasons why one of the groups should contain more observations than the other.

c) If the split mouth design is chosen, what are reasonable values for $\alpha$, $\beta$ (or power=1-$\beta$), $\delta$ and $\sigma$?
   Calculate for this situation the required sample size
   Remember: - one or two sided testing
   - what test should be used?

d) It is clear that the split mouth design needs much less patients. What reason can you think off to choose nonetheless for the parallel group design instead of the split mouth design?

**Improvement of kidney transplantation**

In 4% of the transplantations the kidney transplant failed within one week due to technical problems. Hopefully, the failure rate can be reduced by using an additional plastic vessel that has to be removed surgically at a later time with of course additional risks. A researcher wants to study the effectiveness of this technique in a clinical trial.
Answer the next questions:

a) How many patients are needed for such a trial?
   Remember: - one or two sided testing
   - what test should be used?

b) Sample size calculations often result in huge numbers of patients. How do you react to a situation where the calculated number of patients is considerably larger than expected?

**ß-carotene supplementation in CF-patients**

The article: 'Neutrophil Elastase/$\alpha$1-Proteinase Inhibitor Complex Levels Decrease in plasma of Cystic Fibrosis Patients during longterm oral ß-carotene supplementation' (Winklhofer-Roob et al. 1996, see next page) describe the reduction of elevated MDA levels and NE/$\alpha$1-PI complex levels as a result of ß-carotene supplementation in CF-patients in an observational clinical investigation.
To validate these good results, a randomized trial needs to be conducted. The question is of course how many patients should be included in such trial.
Please carefully read the information in the article.

a) Which design is the most obvious of these questions? A parallel group design or a cross-over design?

b) Calculate with respect to the NE/$\alpha$1-PI complex level the number of patients needed.

c) Calculate with respect to the MDA level the number of patients needed.

d) What is the final choice for the number of patients in the trial? Please give arguments.

e) The article is based on patients with a good vitamin E status. If a good vitamin E status of the patient is not an inclusion criterion, how will this influence the required sample size?

Please fill in the evaluation form at page 135.

0031-3998/96/4001-0130$03.00/0
PEDIATRIC RESEARCH

Vol. 40, No. 1, 1996
*Printed in U.S.A.*

# Neutrophil Elastase/$\alpha_1$-Proteinase Inhibitor Complex Levels Decrease in Plasma of Cystic Fibrosis Patients during Long-Term Oral $\beta$-Carotene Supplementation[1]

BRIGITTE M. WINKLHOFER-ROOB, SUSANNA E. SCHLEGEL-HAUETER, GHOLAMALI KHOSCHSORUR, MARTIN A. VAN'T HOF, SUSANNE SUTER, AND DAVID H. SHMERLING

*Department of Pediatrics, University of Zurich, Switzerland [B.M.W.-R., D.H.S.], Department of Pediatrics, University of Geneva, Switzerland [S.E.S.-H., S.S.], Department of Surgery, University of Graz, Austria [G.K.], and Medical Statistics Department, University of Nijmegen, The Netherlands [M.A.V.H.]*

## ABSTRACT

Lung inflammation in cystic fibrosis (CF) is associated with an increased release from activated neutrophils of oxidants and proteinases. Free radical generation is not efficiently neutralized, and the major anti-proteinase, $\alpha_1$-proteinase inhibitor ($\alpha_1$-PI) is thought to be oxidatively inactivated. We hypothesized that enhanced antioxidant protection could represent an additional long-term strategy to attentuate the host inflammatory response. The effect on plasma neutrophil elastase/$\alpha_1$-PI (NE/$\alpha_1$-PI) complex levels (as a marker of lung inflammation) and plasma malondialdehyde concentrations (as a marker of lipid peroxidation) of additional oral $\beta$-carotene supplementation was studied in 33 CF patients who had already received long-term vitamin E supplementation. In the presence of a more than 10-fold increase in plasma $\beta$-carotene concentrations (mean ± SEM) (0.09 ± 0.01 to 1.07 ± 0.19 $\mu$mol/L; $p < 0.0001$), a small increase in plasma $\alpha$-tocopherol concentrations (23.8 ± 1.31 to 28.4 ± 1.81 $\mu$mol/L; $p = 0.02$), and a more than 50% decrease in plasma malondialdehyde concentrations (1.00 ± 0.07 to 0.46 ± 0.03 $\mu$mol/L; $p < 0.0001$), plasma NE/$\alpha_1$-PI complex levels decreased from 102.2 ± 16.0 to 83.0 ± 10.4 $\mu$g/L; ($p = 0.02$). Plasma retinol concentrations increased (1.05 ± 0.06 to 1.23 ± 0.07 $\mu$mol/L; $p = 0.0001$) due to conversion of $\beta$-carotene to retinol, which could have contributed to the decrease in NE/$\alpha_1$-PI complex levels. Based on these results, we speculate that efficient antioxidant supplementation could attenuate lung inflammation in CF. (***Pediatr Res* 40: 130–134, 1996**)

### Abbreviations

**$\alpha_1$-PI,** $\alpha_1$-proteinase inhibitor
**CF,** cystic fibrosis
**ELF,** epithelial lining fluid
**MDA,** malondialdehyde
**NE,** neutrophil elastase
**TNF-$\alpha$,** tumor necrosis factor-$\alpha$

**Table 1.** *Changes in NE/$\alpha_1$-PI complexes, MDA and antioxidant concentrations, and weight gain during $\beta$-carotene supplementation*

| | Patients ($n$ = 33) | | | Controls ($n$ = 34) | |
|---|---|---|---|---|---|
| Test | Baseline | 16 months | Baseline vs 16 mo* $P$ value | | Patients at 16 mo vs controls† $P$ value |
| NE/$\alpha_1$-PI ($\mu$g/L) | 102.2 ± 16.0 | 83.0 ± 10.4 | 0.02 | 45.6 ± 3.2 | 0.0003 |
| MDA ($\mu$mol/L) | 1.01 ± 0.07 | 0.46 ± 0.03 | <0.0001 | 0.61 ± 0.04 | 0.004 |
| $\alpha$-Tocopherol ($\mu$mol/L) | 23.8 ± 1.3 | 28.4 ± 1.8 | 0.02 | 28.7 ± 1.1 | NS |
| $\beta$-Carotene ($\mu$mol/L) | 0.09 ± 0.01 | 1.07 ± 0.19 | <0.0001 | 1.02 ± 0.07 | NS |
| Retinol ($\mu$mol/L) | 1.05 ± 0.06 | 1.23 ± 0.07 | 0.0001 | 1.99 ± 0.09 | <0.0001 |
| Weight (z score) | −0.55 ± 0.16 | −0.39 ± 0.17 | 0.005 | n.d. | |
| White blood cells ($10^9$/L) | 8.47 ± 0.52 | 8.52 ± 0.57 | NS | n.d. | |
| Bands (%) | 21.1 ± 1.9 | 22.1 ± 1.7 | NS | n.d. | |
| $\alpha_1$-Acid glycoprotein (g/L) | 1.06 ± 0.06 | 1.02 ± 0.06 | NS | n.d. | |

Values are mean ± SEM. NS = not significant; n.d. = not determined.
* Wilcoxon matched pairs signed ranks tests.
† Mann-Whitney $U$ tests.

# Meeting 6

## *Lecture 6: Comparing more than two means (ANOVA)*

Teacher: Ton de Haan

**Summary**
Medical research often aims to assess the effect of an intervention in subgroups of the patient population or in several patient populations. For example: the effect of treatment (methotrexate or with gold salts) on the erythrocyte sedimentation rate in rheumatoid patients with positive or negative rheumatoid factor. In this case, rheumatoid factor is possibly an effect modifier (interaction).
Analysis of variance is a powerful statistical technique to compare different groups. The name “analysis of variance” suggests that the interest is the variances. This is not the case: the interest is in detecting differences in (sub)population means by analyzing variances.
The outcome (dependent) variable has to be METRIC (continuous) and the predictor variable(s) (independent variables) are NOMINAL. The situation that only one independent (nominal) variable is present i.e. if the patients are classified according to one characteristic (factor), is referred to as one-way ANOVA. A special case is the t-test which compares the means of two (sub)groups.
Research is not limited to the study of one factor. Often one wishes to investigate whether treatment effects are different for sub populations. For example, does the dosing of a drug depend on the age of the patient? To answer this question one can design the experiment as follows. The subjects are divided in Adults (age <50) and Elderly (age 50+). Both groups are randomized to two dose regiments (low, high). In such a way four groups of subjects are made, or taking into account the structure of the study 2x2 groups. The two-way ANOVA is the statistical technique to handle data that comes from this kind of experiments. One can easily extend the research question to involve gender as well. Or combinations of treatments, or different compounds or….

**Assumptions**
1) The assumptions for the analysis of variance:
   a) The observations are independent.
   b) For each combination of levels of the design factors (“cells”), the dependent variable is normally distributed.
   c) The observations in each of the cells have the same standard deviation (homoscedasticity).
2) How can these assumptions be checked?
   a) When observations come from different units (subjects) the assumption of independency is met.
   b) Normality can be checked by a (visual) inspection of the residuals. Even if there is only one observation per cell the residuals can be explored.
   c) Homoscedasticity can be checked when there is more than one observation per cell.

## *Preparation for Meeting 7*

In preparation for next week:
- Please read in Field:
  - Chapter 13 Analysis of Covariance.
    The sections marked ||| and the following sections may be omitted when reading: 13.5.5 - 13.5.7, 13.6.2 – 13.6.3; 13.9 – 13.11.
  - Do read 13.7
- Self diagnosis on Brightspace
  - Comparing Means (ANOVA)

## *Computer Session 6: Comparing more than two means (ANOVA)*

### *Background*

The comparison of two groups in a non-randomized study may be biased by confounding variables (so-called “confounders”). A possible solution is to adjust for confounding variables in the analysis. (Two-way) analysis of variance is a method to do this.

In a randomized trial, groups are comparable for confounding variables due to the randomization of subjects to the different treatments. Nevertheless, adjusting for prognostic factors can be advantageous in a randomized trial too: it can increase efficiency (i.e. precision).

Besides two-way ANOVA, also one-way ANOVA exists. The latter is a generalization of the t-test to the situation where more than 2 groups are compared. One-way ANOVA applied to two groups is identical to the t-test for two (independent) groups.

Research interest is often in the effect of several factors simultaneous. For example, does combined administering of two drugs have a different effect than expected on the effect of each drug separately? The effect of one drug may be hampered when also given another drug (antagonism) or the effect may be enhanced when given the other drug (synergism). Another example: Which combinations of pre-surgery treatment and after surgery care are optimal in surgery? Both examples are about statistical interaction between two factors.

A factorial design is not only useful to study interactions, but may also be more efficient than designing separate experiments for each factor, because it enables to study several factors simultaneously in one experiment.

### *Objectives:*

After working through the assignments and reading material, the student should be able to:

- perform a statistical test to compare more than 2 group means and to interpret the results (one-way ANOVA)
- recognize situations where two-way ANOVA is indicated (confounding, efficiency and interaction)
- The student is able to analyze data from a factorial design using SPSS. The student can perform multi-way ANOVA (Factorial ANOVA) with interaction and interpret the interaction terms.

### *Literature*

In Field Section 12.3

*Instruction:*

Make the assignments, take notes and save your SPSS output.

*Product*:

The written answers to the questions. Please fill in the evaluation form at page 135.

### *Discussion*

Results can be discussed with your peers and with the tutor during the computer session.

*Assignments*

**In-company Fitness programs**

Effects of different in-company fitness programs were assessed in a study that included 40 middle-aged males employed at the Katholieke Universiteit van Nijmegen (nowadays known as Radboud University Nijmegen). These volunteers were randomly assigned to 4 groups. The first group was assigned to a standard jogging program, the second group to a bicycle training program, the third group to a muscle power training program, and the fourth group of 10 volunteers acted as control group,

At the start of the fitness program, all volunteers were subjected to a maximal stress test in a treadmill, in order to measure the maximal oxygen intake. $VO_2$-max was measured again at the end of the fitness programs (i.e. after 20 weeks) and the results are recorded in dataset *vo2max.sav*.

a. The means and standard deviations of the VO2-max, expressed in ml/(kg.min), of the 4 groups at the start of the program and after 20 weeks were as follows:

| | | At start program | | After 20 weeks | |
|---|---|---|---|---|---|
| | n | mean | sd | mean | sd |
| Jogging | 10 | | | | |
| Bicycle | 10 | | | | |
| Power | 10 | | | | |
| Control | 10 | | | | |

b. Analyze the differences in mean $VO_2$-max at the end of the 4 fitness programs using one-way ANOVA. Is the difference statistically significant? What is your conclusion?

c. A closer look at the tables reveals that the mean VO2-max of the control group did not increase, whereas the mean of, for instance, the bicycle group increased. Therefore, we decide to explore the change of VO2-max.

| | | Difference Week 20 – Baseline | |
|---|---|---|---|
| | n | Mean | sd |
| Jogging | 10 | | |
| Bicycle | 10 | | |
| Power | 10 | | |
| Control | 10 | | |

d. Does the mean change in VO2-max differ significantly between the 4 programs? Explain the difference in conclusion between analysing the VO2-max at Week 20 and the change in VO2-max.

e. Which programs do significantly improve fitness? (α= 0,05). What problem do you face if the tests are performed at α= 5%? How can you address this problem?

**Stimulation of blood circulation in patients with a spinal cord lesion**
Decreased blood circulation and, consequently, low blood pressure are a problem during rehabilitation of patients with a spinal cord lesion due to an accident. To stimulate blood circulation, the idea arises to use so-called anti-G pants (pants inflatable by high air pressure that are used by pilots to deal with G-forces) instead of the usual methods such as bandages or elastic stockings.
The three treatments are compared in a randomized experiment which is stratified for the level of the lesion (high/low). The blood pressure in sitting position is measured after a 3-weeks training of sitting exercises. The measurements of the diastolic blood pressure (in mmHg) are available in dataset *stimbloodcirc.sav*.

a. Complete the table below.

| | | elastic stockings | Bandages | anti-G pants | total |
|---|---|---|---|---|---|
| high lesion | mean | | | | |
| | sd | | | | |
| | n | | | | |
| low lesion | mean | | | | |
| | sd | | | | |
| | n | | | | |
| total | mean | | | | |
| | sd | | | | |
| | n | | | | |

b. Are the differences between the overall means of the three treatments significant?

c. Are for the low lesions the differences between the means of the three treatments significant? And for the high lessions?

d. Are the differences between the means of the three treatments significant taken into account the level of the lesion?

e. Give 95% (bonferroni corrected) confidence intervals for the differences in diastolic

bloodpressure between the three treatment groups.

f. Why is the two way ANOVA more efficient than the one way ANOVA? Which is the preferred analysis?

**Lung function measurement**

To evaluate the effect of dust on lung function in industrial workers in a file factory, the lung function of these workers was measured. Lung function was expressed as the residual forced vital capacity (FVC) which is the difference between the actual FVC and the predicted FVC i.e. predicted on the basis of body length. The lung function of the 119 workers of the file factory was compared to that of 84 office workers. Moreover, smoking status of all workers was recorded. The data are available in dataset *lungfunction.sav*.

a. The majority of these industrial workers have a positive residual forced vital capacity. Would you have expected this for a random sample from the general population within the same age range?

b. What is the difference in mean lung function between factory and office workers? Also provide the 95%-CI.

c. Do the factory workers have a similar smoking habit as the office workers?

d. Argue why smoking is a potential confounder. Check whether smoking is a confounder.

e. Fill in the two by two table (smoking * working) with per cell the mean FVC. Is smoking an effect modifier for the relation between working place and lung function? (The effect of working place on lung function is different between smokers and non-smokers) In other words, depends the difference in lung function between factory workers and office workers on the smoke status?

| | Smoker | Non-smoker |
|---|---|---|
| Factory worker | | |
| Office worker | | |

f. Formulate your conclusion regarding the relation between working place and lung function.

**A mouse experiment**

A $2^2$ factorial design was carried out to assess the effects of castration and removal of the adrenal gland (adrenalectomy) on the size of the thymus (=zwezerik) in mice. The data are available in the SPSS file *thymus.sav* and is also displayed in the table below:

| Weight thymus (mg) | Adrenalectomy | |
|---|---|---|
| | no | yes |
| No castration | 25 | 44 |
| | 26 | 35 |
| | 31 | 26 |
| | 24 | 39 |
| | 27 | 32 |
| | 20 | 25 |
| | 27 | 21 |
| | 19 | 31 |
| castration | 31 | 44 |
| | 34 | 41 |
| | 41 | 38 |
| | 32 | 39 |
| | 39 | 34 |
| | 36 | 45 |
| | 43 | 55 |
| | 41 | 48 |

a. Eight mice were sacrificed for each combination of castration (yes/no) and

adrenalectomy (yes/no). Calculate the mean and standard deviation of the thymus weights for each of these four combinations. Fill in the results in the table below.

| Weight thymus (mg) | Adrenalectomy | |
|---|---|---|
| | no | yes |
| No castration | Mean:<br><br>SD: | Mean:<br><br>SD: |
| Castration | Mean:<br><br>SD: | Mean:<br><br>SD: |

b. Look at this table: is there any indication of an interaction between castration and adrenalectomy with regard to the weight of the thymus? Give argumentation.

c. Analyse the data using an analysis of variance with two factors (two-way ANOVA). (Hint: general linear models – univariate – enter two fixed factors – model: full factorial - etc.)

   Is there a significant interaction between castration and adrenalectomy?

Fit a model without interaction between castration and adrenalectomy.

d. What is the magnitude of the effect of castration on the thymus weight in male mice? Does this effect depend on whether the adrenal gland is removed or not?

   Provide a 95% confidence interval for this effect.

   What is the magnitude of the effect of adrenalectomy on the thymus weight in male mice? Does this effect depend on whether castration is performed or not?

   Also provide a 95% confidence interval for this effect

Please fill in the evaluation form at page 135.

# Meeting 7

## *Lecture 7: Analysis of Covariance (ANCOVA)*

Teacher: Ton de Haan

To compare the difference in diastolic blood pressure between males and females, we can randomly sample males and females and then compare the sample means using an independent samples t-test. However, blood pressure and weight are associated, so that the crude difference between males and females is partly due to the fact that males are on average heavier than females. To correct for the confounding variable "BMI", only males and females of the same BMI should be included in the samples. A less restrictive strategy is to form weight strata (e.g. 40-50 kg, 50-60 kg, and so on) and perform a separate analysis for each stratum or perform a combined analysis using 2-way ANOVA.
Another option, which avoids forming strata, is analysis of covariance (ANCOVA). Analysis of covariance can correct for a continuous variable. For example, the difference in blood pressure between males and females can be adjusted for BMI using analysis of covariance.
Besides correcting for a covariate, analysis of covariance can also in increase the precision of the estimated difference between groups for a continuous outcome variable. In the example above: if there is no difference in mean BMI between males and females, then the difference in mean blood pressure is not (partially) caused by a differences in mean BMI (so there is no need to correct for BMI). However, analysis of covariance with BMI as covariate will result in a smaller 95%-confidence interval for the difference in mean blood pressure between males and females.
Analysis of covariance can be generalized such that more than two groups can be compared and such that more than one covariate is adjusted for.
Analysis of covariance can be performed using multiple linear regression where dummy variables are used to specify the group classification.

Field uses the Viagra example to explain analysis of covariance. We will comment on his approach and make some distinctions.

## *Preparation for Meeting 8*

In preparation for next week:

- Please read of in the book (Field):
  - Sections 1.7;
  - Chapter 13 Factorial ANOVA the following sections:
    Sections 14.2, 14.3 till 14.3.1, 14.5 (not 14.5.6) 14.6, 14.7
- Self diagnosis on Brightspace
  - ANCOVA

## *Computer Session 7: ANCOVA*

### *Background*

In a randomized study the outcome variable may depend on the baseline value of this variable. Randomization helps to balance the treatment groups with respect to this baseline value. Controlling for this baseline value would still be recommended because it reduces the 'noise' in the data and therefore increase the power. ANCOVA is a statistical technique to accomplish this. In non-randomized studies the groups to be compared can differ in a covariate that is associated with the outcome of interest. In that situation the ANCOVA can correct for the difference in the covariate and see whether after correction a difference exists between the groups.

*Objectives:*

Upon completion of this assignment the student should be able to analyze the relation between a continuous dependent variable and a nominal independent variable while controlling for a continuous variable (for reasons concerning validity and/or efficiency).

*Instruction:*

Make the assignments, take notes and save your SPSS output.

*Product*:

The written answers to the questions. Please fill in the evaluation form at page 135.

### *Discussion*

Results can be discussed with your peers and with the tutor during the computer session.

### *Assignments*

**Mercuhydrin and sodium in dogs**

Urine samples were analyzed for sodium content for each of two collection periods, one before and one after administration of Mercuhydrin, for each of 30 dogs. The experimenter used 7 dogs, administered a placebo, as a control group for the study. The dogs were randomly assigned to the two treatment groups. The data are stored in the Excel file *dogs.xls*.

X: sodium content mM/L (pre-treatment period)
Y: sodium content in mM/L (after treatment period)
Z=1 for the experimental group, Z=0 for the control group

a. Read the data in SPSS
b. What is the crude difference between the experimental and control groups in sodium content after treatment?

c. Make a scatter plot of the after treatment sodium content versus pre-treatment sodium content, such that each group has its own regression line.

d. Use ANCOVA, with the pre-treatment measure as the covariate, to find the adjusted difference between the experimental and control groups in mean sodium content in the second collection period. Is this difference statistically significant?

e. An alternative analysis compares the change in sodium between the two treatment groups. Perform this analysis.

**Blood pressure, body size, age and smoking**

The Excel dataset *BP01.xls* gives the systolic blood pressure (SBP), body mass index (BMI), age (AGE) and smoking history (SMK = 0 if non-smoker, SMK = 1 if a current or previous smoker) for a hypothetical sample of 32 white males over 40 years old from the town Angina. The main interest is in the relation between smoking and SBP.

a. Read the data in SPSS.
b. What is the crude difference in SBP between smokers and non-smokers?

c. Make a scatter plot of SBP (vertical axis) versus BMI (horizontal axis) values in which smokers and non-smokers are shown with different symbols and in which the two regression lines are displayed.

d. Determine the regression line of SBP on BMI separately for smokers and non-smokers. (Two formulas)

Assume that the regressions lines of SBP on BMI for smokers and non-smokers could be considered parallel.

e. Give the formula for the appropriate analysis of covariance (ANCOVA) regression model to compare the mean SBP in the two smoking categories controlling for BMI.

f. What is the difference in mean SBP between smokers and non-smokers adjusted for BMI? Is this difference statistically significant? Compare this difference with the unadjusted difference (crude difference). Obtain a 95% confidence interval for the difference in adjusted SBP means.

g. Is it necessary to control for both Age and BMI as opposed to just one of the two covariates?

**Brain development in children with autism**

Autism is a neurobiological developmental disorder with symptoms in three domains: social interaction, communication and rigid behavior patterns. Despite many years of research it is still far from clear what causes the disease. Indeed, it is still not clear which brain structures are involved in the development of autism. Normally, the brain thickness decrease as it develops. It is not entirely known why it decreases, but it is in any case a sign of maturation. Probably it is related to the removal of irrelevant links (synaptic pruning).

How the brain develops in adolescents with autism has been studied in the FC Donders Center of UMC St Radboud. A large number of MRI scans are collected of children with and without autism to compare the development of the brain in autistic children to the normal development. The adolescents were all right handed and in the same IQ-range. Thanks to FC Donders Centre for providing the research data. There are 22 scans of children with autism (ASD = autism spectrum disorders) and 29 scans of children without autism (TD = Typically Developing controls) made. The thickness of the cortex of right and left hemispheres (LHthickness and RHthickness) could thus be determined. See *CortexThickness.sav* for the data. For this session we analyze data from the left hemisphere.

a. Give the mean thickness (with standard deviation) of the cortex in the autistic group and in the 'normal' group. Is there a statistically significant difference in mean cortical thickness?

b. Describe the age distribution of the adolescents with autism and without autism.

c. Explore, separately for the autistic group and the normal group, the relationship between cortex thickness and age. Make one scatter plot with the two regression lines. Give the formulas for the regression equations. Is there a significant relationship within each group?

d. Check whether the residuals follow a normal distribution.

e. Is the relation between cortex thickness and age during adolescence similar for the two groups?

**Model with interaction**

To test for interaction (difference in slope between the two groups) both regression lines are incorporated in one model:

$$\text{Cortex thickness} = a + b \times \text{age} + c \times \text{diagnosis} + d \times \text{diagnosis} \times \text{age}$$

with diagnosis = 0 for the normal group (TD) and diagnosis = 1 for the autistic group (ASD)

f. Perform this regression analysis.

g. Give the regression lines for the autistic group and for the normal group according to this model. Show that these lines are the same as the lines earlier. What is different between the two methods?

h. What parameter shows whether the lines are parallel? Is this parameter significantly different from 0?

i. Why can no conclusions be drawn from this study regarding the development of the brain?

j. Improve the design of the study so the conclusions can be drawn regarding the development of the cortex.

k. Report the differences found in the relation of the cortex thickness and age between adolescents with and without autism.

Please fill in the evaluation form at page 135.

# Meeting 8

## *Lecture 8: Experimental Designs*

Teacher: Ton de Haan

The variation in the outcome measurements is often composed from several sources of variation. The research question however pertains to only a few of those sources. Only for those sources the question whether the level of that source influences the outcome is relevant. An example is a study to assess the effect of three blood pressure lowering drugs and possible sources of variation include:

treatment (type of drug)
gender
age
severity of disease
body weight

A simple study design would be: take sufficient subjects with high blood pressure, randomize which drug each subject will receive for a period of 3 months, and measure their blood pressure at the end of the three months. This design only controls for one source of variation: the treatment. The randomization will hopefully distribute age and also gender evenly over the three treatment groups. In fact, randomization makes a balanced distribution most likely, but does not guarantee this. Bias due to an unbalanced distribution of sources of variation could be corrected for by a suitable statistical model.
An alternative and preferable method to account for sources of variation is in the design of the experiment. This lecture and the assignments will illustrate how to control for sources of variation in the design.
Many of the experimental designs that we will discuss can be analyzed using n-way analysis of variance (factorial ANOVAs). Field discusses this technique in Chapter 13. The computer exercises will show some real life applications.

## *Preparation for Meeting 9*

In preparation for next week:

- Please read the text of Petrie and Sabin on survival analyses (page 92 of this manual)
- Spruance, S. L., J. E. Reid, et al. (2004). "Hazard Ratio in Clinical Trials." Antimicrobial Agents and Chemotherapy **48**(8): 2787-2792. (Link on Brightspace)
- Self diagnosis on Brightspace
  - Experimental Design

## *Computer Session 8: Experimental Designs*

### *Background*

The factors in an experimental design may have more than two levels. Suppose that three factors, each with 4 levels are of interest. A full factorial design should have 4x4x4 = 64 different combination tested. For this reason, statisticians have sought to reduce full factorial designs incorporating factors with more than two levels around the 1950s. An example of such a reduced design is the *Latin Square*. This design can be considered if there are 3 factors each having the same number of levels. In the above mentioned example of three factors with each 4 levels, the latin square design needs only 16 smartly chosen combinations to study main effects.

In multi factor experiments testing all the possible interactions between those factors may not be meaningful. Especially higher order interactions (between three or more factors) are very difficult to interpret. In addition, testing all interactions may introduce chance findings: finding one or more interactions to be significant by chance. Indeed, testing with a significance level of 5% $(\alpha = 0.05)$ means that we expect one out of every 20 tests to be significant, even if for each test the null hypothesis holds.

### *Objectives*

- The student can assess the advantages and disadvantages of a factorial design versus a series of experiments for each factor separately.
- The student is able to recognize some experimental designs (Randomized block, Factorial, *Latin squares*) and can analyze these designs adequately using SPSS.

*Instruction:*

Make the assignments, take notes and save your SPSS output.

*Product*:

The written answers to the questions. Please fill in the evaluation form at page 135.

### *Discussion*

Results can be discussed with your peers and with the tutor during the computer session.

*Assignments*

**Vitamin D**

The data below are from a study that investigated the effect of vitamin D in rats. Four treatments were investigated (2 dosages, 2 preparation methods). Eight rats were taken from each of six litters (48 rats in total). The outcome measure is the anti-rickets activity (Rickets is an English disease) of vitamin D measured on a scale from 0 to 12 using a bone test.

*Anti- rickets activity in rats by treatment.*

| | Standard Preparation | | Experimental Preparation | |
|---|---|---|---|---|
| Litter | Low Dose | High dose | Low dose | High dose |
| 1 | 3 | 4 | 4 | 5 |
| | 3 | 7 | 6 | 7 |
| 2 | 2 | 4 | 4 | 5 |
| | 2 | 4 | 5 | 6 |
| 3 | 2 | 5 | 4 | 5 |
| | 3 | 5 | 4 | 5 |
| 4 | 3 | 4 | 4 | 5 |
| | 3 | 5 | 5 | 7 |
| 5 | 3 | 4 | 3 | 6 |
| | 3 | 4 | 2 | 5 |
| 6 | 3 | 4 | 3 | 4 |
| | 4 | 5 | 6 | 7 |

The above data are available in the file *VitamineD.sav*.

a. Read the data from that file.
b. What is the design of the experiment?

Only for the sake of presentation we forget for a moment the litter.

c. Fill in the table below:

| | | Preparation | |
|---|---|---|---|
| | | Standard | Experimental |
| Dose | Low | Mean:<br>SD: | Mean:<br>SD: |
| | High | Mean:<br>SD: | Mean:<br>SD: |

Look at this table and answer the following question intuitively.

d. Does the effect of preparation depend on dose?

e. Does the effect of dose depend on the preparation method?

f. Why does this table not fully take into account all the available information?

g. Perform an analysis of variance with factors: litter, preparation method, dose and interaction between preparation method and dose.

h. Is a full factorial model sensible? Motivate your answer.

i. Do the preparation methods differ? Motivate your answer.

j. Do the dose levels differ? Motivate your answer.

k. Does the optimal dose depend on the preparation method? Motivate your answer.

**Meat production and sheep**

Already in 1788 an experiment was carried out to investigate the effect of four different diets on weight gain of sheep. The research questions were:

- which diet is the best,
- which month is the best to slaughter the sheep,
- do the results depend on the type of sheep?

Four sheep of four breeds were included in the experiment, therefore 16 sheep in total. These sheep were weighted before the start of the experiment and 1, 2, 3 and 4 months after the start of the diet. At the end of each month, four sheep were slaughtered directly after weighting the sheep. The corresponding data are collected in the table below.

| Weights in lb, 1 lb = 0,373 kg | | | | | | |
|---|---|---|---|---|---|---|
| **Diet** | **Sheep breed** | **Baseline weight** | **Weight at 1 month** | **Weight at 2 months** | **Weight at 3 months** | **Weight at 4 months** |
| 1 | 1 | 69.75 | 79.75 | . | . | . |
| 1 | 2 | 70.75 | 82.50 | 90.25 | 93.00 | 95.00 |
| 1 | 3 | 69.25 | 83.00 | 82.50 | 84.00 | . |
| 1 | 4 | 88.00 | 95.00 | 101.00 | . | . |
| 2 | 1 | 69.00 | 86.00 | 87.00 | . | . |
| 2 | 2 | 71.00 | 86.00 | . | . | . |
| 2 | 3 | 68.50 | 78.50 | 82.50 | 84.00 | 84.50 |
| 2 | 4 | 79.00 | 95.50 | 97.50 | 97.50 | . |
| 3 | 1 | 72.00 | 83.25 | 90.50 | 94.00 | . |
| 3 | 2 | 70.75 | 80.75 | 86.00 | . | . |
| 3 | 3 | 77.25 | 90.50 | . | . | . |
| 3 | 4 | 80.00 | 93.50 | 98.50 | 100.50 | 101.00 |
| 4 | 1 | 74.00 | 91.00 | 95.50 | 102.00 | 106.00 |
| 4 | 2 | 73.50 | 84.25 | 91.50 | 96.00 | . |
| 4 | 3 | 71.00 | 86.25 | 93.00 | . | . |
| 4 | 4 | 71.00 | 87.00 | . | . | . |

Diet: 1= potatoes, 2 = turnips, 3 = beets, 4 = oats, barley, and grey peas.
Breed: 1= Isle de France, 2 = Beauce, 3 = Champagne, 4 = Picardy

a. Relevant is the weight gain between start of the experiment and the moment of slaughter and, therefore, this is chosen as the dependent variable. What is the minimal number of sheep needed if a full factorial design is chosen with the factors diet, duration of diet (in months) and sheep breed?

In fact, fewer sheep were used in this experiment (16). Nonetheless, it is possible to test the effects of the three factors separately. Moreover, the differences in effect between the three diets can be estimated as well as the differences in effects between the breeds and between the several durations. It is the elegant design that makes all these comparisons possible. The scheme of the design can be summarized as follows:

| | Month 1 | Month 2 | Month 3 | Month 4 |
|---|---|---|---|---|
| Diet 1 | S1 | S4 | S3 | S2 |
| Diet 2 | S2 | S1 | S4 | S3 |
| Diet 3 | S3 | S2 | S1 | S4 |
| Diet 4 | S4 | S3 | S2 | S1 |

S = sheep breed

Each cell in the scheme above corresponds to exactly one observation of weight gain. E.g. the cell in the upper left corner corresponds to the observation of the Isle de France sheep (S1) that received the potato diet (Diet1) and was slaughtered after one month (Month 1).
Note that each breed occurs once in each row and each column.

b. Give the name of the design.

c. Part of the data (omitting the in-between weights) is recorded in the SPSS data file *sheep.sav.* Analyse the data and calculate for each significant effect the difference between levels in mean weight gain (with 95%-confidence interval). Hint: use the option post hoc / equal variances *assumed / Tukey.*

An important assumption for applying a Latin Square design without replicates is the absence of interaction between the three factors with regard to the outcome variable (= weight gain).

d. Test for interaction between breed and diet with regard to their effect on weight gain. Can the model assumption of absence of interaction be verified?

**Rheumatoid arthritis**

Rheumatoid arthritis is characterized by a chronic inflammation of the synovial joints resulting in a progressive breakdown of cartilage and bone tissue. The effects of a (experimental) treatment on cartilage breakdown can be studied using animal models.

Cartilage consists of a matrix of collagen fibers and proteoglycanes. Proteoglycanes are strongly sulfatated polysaccharides that form large aggregates in the collagen network. Proteoglycanes generate an osmotic pressure (swelling) by their negative electrical charge from which the matrix derives its elasticity. Proteoglycanes have a high turnover and during arthritis the breakdown of molecules is increased and the polyglycane production of the cells decreases strongly. The latter can be measured by monitoring the uptake of radioactive sulfate.

The experimental set up is that arthritis is introduced in the right knee of a mouse, while the left knee is left unaffected and acts as control. After two days, the synthesis in the cartilage of the kneecap is measured using the uptake of $^{35}$S-sulfate.

The research question is whether proteoglycane synthesis (uptake of $^{35}$S-sulfaat) is inhibited by the inflammation process. Radio activity is measured as the number of counts per minute (liquid scintillation counter) per kneecap.

After some time, the experiment is repeated with other mice. The results of this experiment are on average smaller due to the use of a different strain. The results of both experiments are shown in the table below.

**Results of arthritis on the uptake of $^{35}S$ –sulfate in cartilage (counts/minute)**

| | | Leg | |
|---|---|---|---|
| | Mouse | Left | Right |
| Experiment 1 | 1 | 3482 | 2500 |
| | 2 | 3306 | 2468 |
| | 3 | 2480 | 3104 |
| | 4 | 2549 | 2678 |
| | 5 | 3109 | 2100 |
| | 6 | 2912 | 2394 |
| Experiment 2 | 1 | 1920 | 1702 |
| | 2 | 1918 | 1595 |
| | 3 | 2374 | 1499 |
| | 4 | 1651 | 1965 |
| | 5 | 1885 | 1300 |
| | 6 | 2231 | 1766 |

a. In *rheuma.xls* the data of the experiment are available. Please read the data in SPSS and perform for Experiment 1 and Experiment 2 separately an appropriate analysis to answer the research question.

b. Is the treatment effect different for the two strains? How would you report the treatment effect?

c. Can the setup of these experiments be improved? (Please comment).
   i) regarding the number of mice (n=6)?
   ii) regarding the design?

**Pain relief**

The data[1] are from an experiment examining the effects of codeine and acupuncture on post-operative dental pain in male subjects. The codeine levels are a codeine capsule (codeine=2) or a sugar capsule (codeine=1). The acupuncture levels are two inactive acupuncture points (acupuncture=1) or two active acupuncture points (acupuncture=2). There are four distinct treatment combinations due to the factorial treatment structure. The 32 subjects are assigned to eight blocks of four subjects having the same pain tolerance score. The four subjects in a block are randomly assigned to one of the four treatments. Pain relief scores were obtained for all subjects two hours after dental treatment. The data are available in *painrelief.sav* and displayed below.

[1] From Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied linear statistical models. 4 ed: Irwin Chicago; 1996.pages 1111-1112

**Pain relief scores**

| | Inactive points | | Active points | |
|---|---|---|---|---|
| PainLevel | Sugar capsule | Codeine capsule | Sugar capsule | Codeine capsule |
| 1 | .0 | .5 | .6 | 1.2 |
| 2 | .3 | .6 | .7 | 1.3 |
| 3 | .4 | .8 | .8 | 1.6 |
| 4 | .4 | .7 | .9 | 1.5 |
| 5 | .6 | 1.0 | 1.5 | 1.9 |
| 6 | .9 | 1.4 | 1.6 | 2.3 |
| 7 | 1.0 | 1.8 | 1.7 | 2.1 |
| 8 | 1.2 | 1.7 | 1.6 | 2.4 |

a. What is the design of the experiment?

b. Analyze the data and formulate your conclusions regarding the effects of codeine and acupuncture on pain relief.

**Muscles**

The following example is from a well-known reference text on experimental designs[2]. Previous experiments have shown that electrical stimulation of muscles, whose nerves are blocked, helps to prevent breakdown of those muscles. A factorial experiment with rats was performed to find out what is the best way to apply electrical stimulation. The following table provides the factors and their levels:

| Number of treatments per day | Duration of treatment (minutes) | Type of current |
|---|---|---|
| 1 | 1 | Galvanic |
| 3 | 2 | Faradic |
| 6 | 3 | Alternating 60 Hz |
| | 5 | Alternating 25 Hz |

Treatments started the third day after nerve blockade and lasted for 11 days in total. The total number of different treatments is 48 ($3x4^2$ factorial design) and each of these treatments was applied to a different rat. The number of replicates is two, so 96 rats

[2] Experimental Designs by W.G. Cochran and G.M. Cox, second edition, John Wiley & Sons, Inc. Page 176-177

were used in total.
The muscles used in the experiment are the *gastrocnemius-soleus* group on one side of the animal and the nerve blockade was introduced by removal of a small part of the *sciatic* nerve. The weight of the blocked muscle at the end of the treatment was used to quantify the effect of the different treatments. However, this weight also depends on the size of the rat, and therefore the weight of the corresponding muscle on the other side of the rat was measured in addition. To correct for the size of the rat, the ratio of these two muscles (weight of the blocked muscle divided by the weight of the control muscle) was calculated. The data are provided in the table (next page) and electronically available in the SPSS file *electric.sav*. The dependent variable is the ratio (weight of the blocked muscle divided by the weight of the control muscle) multiplied by 100.

*Weights of denervated (y) and corresponding normal (x) muscle (unit = 0,01 gram)*

| | | | Number of treatment periods daily | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | One | | Three | | Six | |
| | Length of treatment (minutes) | Type of current | y | x | y | x | y | x |
| Rep I | 1 | Galvanic | 72 | 152 | 74 | 131 | 69 | 131 |
| | | Faradic | 61 | 130 | 61 | 129 | 65 | 126 |
| | | 60 cycle alternating | 62 | 141 | 65 | 112 | 70 | 111 |
| | | 25 cycle alternating | 85 | 147 | 76 | 125 | 61 | 130 |
| | 2 | Galvanic | 67 | 136 | 52 | 110 | 62 | 122 |
| | | Faradic | 60 | 111 | 55 | 180 | 59 | 122 |
| | | 60 cycle alternating | 64 | 126 | 65 | 190 | 64 | 98 |
| | | 25 cycle alternating | 67 | 123 | 72 | 117 | 60 | 92 |
| | 3 | Galvanic | 57 | 120 | 66 | 132 | 72 | 129 |
| | | Faradic | 72 | 165 | 43 | 95 | 43 | 97 |
| | | 60 cycle alternating | 63 | 112 | 66 | 130 | 72 | 180 |
| | | 25 cycle alternating | 56 | 125 | 75 | 130 | 92 | 162 |
| | 5 | Galvanic | 57 | 121 | 56 | 160 | 78 | 135 |
| | | Faradic | 60 | 87 | 63 | 115 | 58 | 118 |
| | | 60 cycle alternating | 61 | 93 | 79 | 126 | 68 | 160 |
| | | 25 cycle alternating | 73 | 108 | 86 | 140 | 71 | 120 |
| Rep II | 1 | Galvanic | 46 | 97 | 74 | 131 | 58 | 81 |
| | | Faradic | 60 | 126 | 64 | 124 | 52 | 102 |
| | | 60 cycle alternating | 71 | 129 | 64 | 117 | 71 | 108 |
| | | 25 cycle alternating | 53 | 108 | 65 | 108 | 66 | 108 |
| | 2 | Galvanic | 44 | 83 | 58 | 117 | 54 | 97 |
| | | Faradic | 57 | 104 | 55 | 112 | 51 | 100 |
| | | 60 cycle alternating | 62 | 114 | 61 | 100 | 79 | 115 |
| | | 25 cycle alternating | 60 | 105 | 78 | 112 | 82 | 102 |
| | 3 | Galvanic | 53 | 101 | 50 | 103 | 61 | 115 |
| | | Faradic | 56 | 120 | 57 | 110 | 56 | 105 |
| | | 60 cycle alternating | 56 | 101 | 56 | 109 | 71 | 105 |
| | | 25 cycle alternating | 56 | 97 | 58 | 87 | 69 | 107 |
| | 5 | Galvanic | 46 | 107 | 55 | 108 | 64 | 115 |
| | | Faradic | 56 | 109 | 55 | 104 | 57 | 103 |
| | | 60 cycle alternating | 64 | 114 | 66 | 101 | 62 | 99 |
| | | 25 cycle alternating | 59 | 102 | 58 | 98 | 88 | 135 |

a. Read the data and calculate the outcome variable.
b. Use analysis of variance (ANOVA with three factors) to determine whether there are significant interactions. Hint: check in the model options whether the option “full factorial” is set.

c. Repeat the ANOVA, but remove the non-significant interactions (step by step

starting with the highest order interactions).
Does duration of the treatment influence the ratio?

Idem for the number of treatments?

Idem for the type of current?

d. Calculate the mean difference (contrast) between the two types of current *Galvanic* and *Faradic* and the corresponding 95%-confidence interval. Hint: use the 'post hoc' option to calculate contrasts. Should you correct for multiple testing?

e. Is there a significant difference between Faradic current and alternating current 25Hz?

Please fill in the evaluation form at page 135.

# Meeting 9

## *Lecture 9: Repeated Measurements*

Teacher: Ton de Haan

Many medical studies focus on how a variable changes over time, rather than only observing a variable at a given instant. This could be because the investigator wishes to observe how a variable evolves over time, such as the height of a growing child, or to observe the natural variation that occurs in a clinical measurement, such as the blood pressure of a volunteer on successive days. Often the goal is to observe the time course of some intervention, such as a treatment, for example, the respiratory function of a patient at a series of time points after the administration of a bronchodilator, or the blood glucose of a diabetic patient in the two hours after a glucose challenge.

Data collected successively on each of several units, whether patients, volunteers, animals or other units, are variously referred to as *longitudinal data* or *repeated measurements*. Data can be collected on several groups of individuals and with each subject measured irregular.
Most studies attempt to make observations at preplanned times. Those that do not – for example observations taken opportunistically, or perhaps when some clinical event occurs – are likely to present formidable interpretation problems. For preplanned observations, there is no requirement that they be taken at regular intervals, and in fact it may often not be sensible to do so. For example, observations may need to be taken more frequently when the response is changing rapidly, provided, of course, that this aspect of the response is of interest. For example, in the study of the profile of the blood level of a short acting drug, measurements may be made every 10 or 15 minutes in the initial stages when the profile is changing rapidly, but then less frequently, perhaps at 1, 2 and 3 hours post-administration. In many medical studies, the reasons behind the timing of observations are seldom discussed and it may be that this aspect of research involving the collection of longitudinal data would benefit from greater reflection.

Often a study will be intended to measure individuals at the same set of times, but this is not achieved in every case. Such *missing data* give rise to two separate problems, which are often not distinguished from one another as clearly as they might be. The first, which is largely technical, is that the varying number of observations per individual may influence the type of analysis performed, as some methods are less tractable, or even impossible, when the number of observations varies between individuals. The second problem occurs when the reasons that data are missing related to the purpose of the study, so an analysis of only the available data may well be biased and possibly misleading. This problem is more subtle, but is potentially more serious because it can evade an unwary analyst.
As with any statistical analysis, it is important when dealing with longitudinal data that the structure of the data is respected. The two most important aspects for longitudinal data are that the method should:

1 take the link between successive measurements on a given individual into account and
2 recognize that successive measurements on an individual will not, in general, be independent.

Despite warnings to the contrary, both aspects appear to be frequently overlooked in the medical literature, where it is common to see separate analyses performed at the different times when measurements were made. Such analyses ignore the fact that the same individuals are present in successive analyses and, as a result, make no allowance for within-individual correlations.

Appropriate methods for analysis of longitudinal data have been studied intensively in recent years, and there is now a very large statistical literature on the subject. In this course we will confine ourselves to studies where the dependent variable is continuous, and we will analyze these data with the aid of so called *linear mixed models.*

The Chapters 14 and 15 of Field deal with repeated measurements in an old fashion way. There are several objections against this type of analysis:

- it cannot handle missing data efficiently.
- it cannot incorporate irregular measured data because it cannot incorporate the time of measurement as a continuous variable
- it only can use a simple correlation structure.

However, the mixed models as described in Chapter 20 of Field are also applicable for the analyses of repeated measures.

(This introductory text is adapted from Armitage, Berry, Matthews, Statistical Methods in Medical Research, Blackwell Science, fourth edition, 2002, pages 430-431)

## *Preparation for Meeting 10*

In preparation for next week:

- Please read in Field:
  - Chapter 20 all parts that are marked  or  , but not 20.3.6, 20.5.12, and 20.5.13.
- Self diagnosis on Brightspace
  - Linear mixed models

## *Computer Session 9: Mixed models for Repeated Measurements*

*Background:*
A simple way of dealing with repeated measurements is by summarizing the data. For instance take per subject the mean of the measurements or the maximum. Another way is fitting line or curve and use the parameter(s) as outcome per subject.
The two-period cross-over design is regularly used in clinical research. In cross-over designs each unit is its own control. Therefore, with this design we have automatically controlled for (unforeseen) confounders, and the random variation (error) is less than in a parallel groups design. But the data are dependent and the statistical method should handle this dependence of the data. Also in non-clinical research, in which treatments/experimental conditions are performed successively, it is wise to perform it according to the cross-over design. In that way, possible period effects (systematic differences between the first and second measurement, for example a learning effect by subjects) can be determined and corrected for. Cross-over designs can only be used in special situations where treatment effects are temporary and carry-over effects can be excluded.

*Objectives*
Upon completing this assignment, the student should be able to:

- Describe the difference between ordinary linear regression and longitudinal data analysis
- Analyze a repeated measurement design.
- Consider when a 'cross-over design' is indicated.
- Analyse data from a 'cross-over' design.
- Design a cross-over experiment.

*Instruction*
This assignment consists of 3 questions.
Read these questions carefully and write down your answers.

*Product*
The product will be the written answers. Please fill in the evaluation form at page 135.

*Discussion*
Results can be discussed with your peers and with the tutor during the computer session.

*Assignments*

**Growth in girls and boys**
For 11 girls and 16 boys the distance from the center of the pituitary to the pterygomaxillary fissure was recorded at the age of 8, 10, 12 and 14 years. See SPSS datasets *potthoff_miss.sav* and the table below.

There are at most 4 observations for each subject.
a. Make graphics that shows each individual growth curve.

How can we see the observations are correlated?

The dataset needs to be restructured to be able to perform statistical analyses. (You can try to restructure the dataset yourself or use *potthoff_miss_tr.sav)*

b. If you ignore the dependence between the data of a subject, what analysis can be done to see whether the growth in girls differs from the growth in boys? (No specific growth pattern is assumed, linear growth will be investigated in Computer Session 13) Perform this analysis.

c. Perform the correct analysis to see whether the growth in girls differs from the growth in boys? (No specific growth pattern is assumed, linear growth will be investigated in Computer Session 13). Look at the within subject tests (change in time) and between subjects test (Gender difference). Are the results comparable with the outcome of question b? Explain your findings.

d. The old fashion analyses handle the missing values by deleting the subject with a missing value, the so called complete cases method. Use the *select cases* option of SPSS to restrict the data to complete cases and perform the last analysis on these selected data. Fill in the Table below the means and standard errors using both analyses. A trick to do this is using the *EM means* options of the Mixed models.

| Mean Distance<br>SE | Age | | | |
|---|---|---|---|---|
| | 8 | 10 | 12 | 14 |
| Girls<br>All data | | | | |
| Complete cases | | | | |
| Boys<br>All data | | | | |
| Complete cases | | | | |

*Growth data for 11 girls and 16 boys (mm)*.

| | | Age in years | | | |
|---|---|---|---|---|---|
| Gender | idnr | 8 | 10 | 12 | 14 |
| Girl | 1 | 21.0 | 20.0 | 21.5 | 23.0 |
| | 2 | 21.0 | 21.5 | 24.0 | 25.5 |
| | 3 | 20.5 | 24.0 | | 26.0 |
| | 4 | 23.5 | 24.5 | 25.0 | 26.5 |
| | 5 | 21.5 | | 22.5 | 23.5 |
| | 6 | 20.0 | 21.0 | 21.0 | 22.5 |
| | 7 | 21.5 | 22.5 | 23.0 | 25.0 |
| | 8 | 23.0 | 23.0 | 23.5 | 24.0 |
| | 9 | 20.0 | 21.0 | 22.0 | 21.5 |
| | 10 | 16.5 | 19.0 | 19.0 | 19.5 |
| | 11 | 24.5 | 25.0 | | |
| Boy | 12 | 26.0 | | 29.0 | 31.0 |
| | 13 | 21.5 | 22.5 | 23.0 | 26.5 |
| | 14 | 23.0 | 22.5 | 24.0 | |
| | 15 | 25.5 | 27.5 | 26.5 | 27.0 |
| | 16 | 20.0 | 23.5 | 22.5 | 26.0 |
| | 17 | 24.5 | 25.5 | 27.0 | 28.5 |
| | 18 | 22.0 | | 24.5 | 26.5 |
| | 19 | 24.0 | 21.5 | 24.5 | |
| | 20 | 23.0 | | 31.0 | 26.0 |
| | 21 | 27.5 | | 31.0 | 31.5 |
| | 22 | 23.0 | 23.0 | 23.5 | 25.0 |
| | 23 | 21.5 | 23.5 | 24.0 | 28.0 |
| | 24 | 17.0 | 24.5 | 26.0 | 29.5 |
| | 25 | 22.5 | | 25.5 | 26.0 |
| | 26 | 23.0 | 24.5 | 26.0 | |
| | 27 | 22.0 | 21.5 | 23.5 | 25.0 |

Sources: Potthoff and Roy (1964), Jennrich and Schluchter (1986), Verbeke and Molenberghs (1997)

**Chemotherapy**

In patients with lung cancer, some bone marrow parameters (WBC, Hb, blood platelets) decline due to the chemotherapy. During a second treatment these parameters decline even further. A theory is that a growth factor given preceding the treatment will decrease the decline, so that the dose of the chemotherapy may be increased.
The effect of the growth factor G-CSF was studied in twelve patients in a 'cross-over design'. All patients received two chemo treatments with an interval of 4 weeks. Group 1 received G-CSF preceding the second chemo treatment, Group 2 received G-CSF preceding the first chemo treatment. The effect was measured by means of the amount of days with a WBC less than $3.0*10^9$/l. The results are presented in the table.

**Number of days with WBC less than $3.0*10^9$/l.**

| | Period 1 | Period 2 |
|---|---|---|
| Group 1 | 4 | 4 |
| | 13 | 9 |
| | 11 | 11 |
| | 14 | 17 |
| | 7 | 1 |
| | 9 | 6 |
| Group II | 7 | 7 |
| | 1 | 8 |
| | 11 | 10 |
| | 0 | 11 |
| | 6 | 13 |
| | 1 | 11 |

**Theory**

a. What is a carry-over effect?

   What are the assumptions underlying a valid cross-over design?

   Can these assumptions be verified? If yes, how? If no, why not?

   Is it possible to analyse the data in case of a treatment-period interaction? Please explain.

   Explain why it is not advisable to perform a 'cross-over trial' when a treatment-period interaction cannot be excluded in advance.

b. Do you expect a carry-over effect for the chemotherapy trial?

c. Which alternative design is possible? Is this design better than the cross-over design?

**Analysis of the data**

Enter the data in SPSS using an appropriate structure. Consider what variables you need. Analyse these data using an appropriate mixed model.

d. Is there a significant treatment-period interaction?

e. Calculate the treatment effect with the 95% confidence interval.

**Three period cross over trial**

The table below is based on data considered in Senn and Hildebrand (1991) which come from a single centre in the multi-centre trial reported by Tsoy et al. (1990). The data present values of forced expiratory volume in one second (FEV1) obtained after an exercise challenge in a three-period three-treatment double-blind cross-over trial comparing the protective effect of a single dose of an experimental treatment, formoterol solution aerosol (12 µg), to a single dose of a standard therapy, salbutamol suspension aerosol (100 µg) and placebo for patients suffering from exercise-induced asthma. In a titration study carried out at the beginning of the trial an appropriate exercise test was established for each patient. The patient was then asked to perform this exercise test two hours after treatment on each of the three treatment days. The values reported are the lowest of a number of determinations of FEV1 made in the period after the exercise test. All six possible sequences of formoterol, salbutamol and placebo were used. Patients were allocated at random to the sequences.

| Patient | Sequence | Period 1 | Period 2 | Period 3 |
|---|---|---|---|---|
| 1 | FSP | 3500 | 3200 | 2900 |
| 10 | FSP | 3400 | 2800 | 2200 |
| 17 | FSP | 2300 | 2200 | 1700 |
| 21 | FSP | 2300 | 1300 | 1400 |
| 23 | FSP | 3000 | 2400 | 1800 |
| 4 | SPF | 2200 | 1100 | 2600 |
| 8 | SPF | 2800 | 2000 | 2800 |
| 16 | SPF | 2400 | 1700 | 3400 |
| 6 | PFS | 2200 | 2500 | 2400 |
| 9 | PFS | 2200 | 3200 | 3300 |
| 13 | PFS | 800 | 1400 | 1000 |
| 20 | PFS | 950 | 1320 | 1480 |
| 26 | PFS | 1700 | 2600 | 2400 |
| 31 | PFS | 1400 | 2500 | 2200 |
| 2 | FPS | 3100 | 1800 | 2400 |
| 11 | FPS | 2800 | 1600 | 2200 |
| 14 | FPS | 3100 | 1600 | 1400 |
| 19 | FPS | 2300 | 1500 | 2200 |
| 25 | FPS | 3000 | 1700 | 2600 |
| 28 | FPS | 3100 | 2100 | 2800 |
| 3 | SFP | 2100 | 3200 | 1000 |
| 12 | SFP | 1600 | 2300 | 1600 |
| 18 | SFP | 1600 | 1400 | 800 |
| 24 | SFP | 3100 | 3200 | 1000 |
| 27 | SFP | 2800 | 3100 | 2000 |
| 5 | PSF | 900 | 1900 | 2900 |
| 7 | PSF | 1500 | 2600 | 2000 |
| 15 | PSF | 1200 | 2200 | 2700 |
| 22 | PSF | 2400 | 2600 | 3800 |
| 30 | PSF | 1900 | 2700 | 2800 |

The data can be found in *sen164.sav*. Analyse this trial. Look for indication of carry-over. Estimate the differences in mean FEV1 between the treatments.

Please fill in the evaluation form at page 135.

## Meeting 9

### *Lecture 10: Survival analysis*

Teacher: Steven Teerenstra

The time until a certain event occurs is often of interest in clinical or epidemiological follow-up studies. Examples of such time-to-event endpoints are the time from the first heart attack till the second heart attack or the time from inclusion into the trial till death. A typical problem in the analysis of such time-to-event data is that some patients will have experienced the event of interest (sometimes referred to as “failure”) before the end of the trial, while others will not have had this event. For the latter, the time-to-event is unknown, but we do know that the event has not occurred in the time frame the subject was followed i.e. the time-to-event for such subjects will exceed the their follow-up time. The (length of the) follow-up period of such subjects are called censored times-to-event and an appropriate statistical analysis has to account for the fact that data of such subjects are not actual times-to-event, but lengths of the period that the subject was followed and did not experience the event.

*Kaplan-Meier curves*
The method developed by Kaplan and Meier will be introduced and discussed. The Kaplan-Meier curve is an estimate of the ‘survival function’ with few assumptions.

*Cox Proportional Hazards Model*
The Cox proportional hazards regression model is a flexible tool for assessing the relationship of multiple predictors to a right-censored, time-to-event outcome, and has much in common with linear and logistic models.

*Literature:*
Petrie & Sabin. Medical Statistics at a Glance. Second edition, Chapter 44 Survival analysis. Blackwell Publishing Ltd, 2005. (See next pages)
Kleinbaum & Klein. Survival Analysis, A Self-Learning Text. Second Edition. Springer, 2005.

# 44 Survival analysis

Survival data are concerned with the time it takes an individual to reach an endpoint of interest (often, but not always, death) and are characterized by the following two features.

- It is the **length of time** for the patient to reach the endpoint, rather than whether or not s/he reaches the endpoint, that is of primary importance. For example, we may be interested in length of survival in patients admitted with cirrhosis.
- Data may often be **censored** (see below).

Standard methods of analysis, such as logistic regression or a comparison of the mean time to reach the endpoint in patients with and without a new treatment, can give misleading results because of the censored data. Therefore, a number of statistical techniques, known as **survival methods**[1], have been developed to deal with these situations.

## Censored data

Survival times are calculated from some baseline date that reflects a natural 'starting point' for the study (e.g. time of surgery or diagnosis of a condition) until the time that a patient reaches the endpoint of interest. Often, however, we may not know when the patient reached the endpoint, only that s/he remained free of the endpoint while in the study. For example, patients in a trial of a new drug for HIV infection may remain AIDS-free when they leave the study. This may either be because the trial ended while they were still AIDS-free, or because these individuals withdrew from the trial early before developing AIDS, or because they died of non-AIDS causes before the end of follow-up. Such data are described as **right-censored**. These patients were known *not* to have reached the endpoint when they were last under follow-up, and this information should be incorporated into the analysis.

Where follow-up does not begin until after the baseline date, survival times can also be **left-censored**.

## Displaying survival data

- A separate horizontal line can be drawn for each patient, its length indicating the survival time. Lines are drawn from left to right, and patients who reach the endpoint and those who are censored can be distinguished by the use of different symbols at the end of the line (Fig. 44.1). However, these plots do not summarize the data and it is difficult to get a feel for the survival experience overall.
- **Survival curves**, usually calculated by the **Kaplan–Meier** method, display the cumulative probability (the **survival probability**) of an individual remaining free of the endpoint at any time after baseline (Fig. 44.2). The survival probability will only change when an endpoint occurs, and thus the resulting 'curve' is drawn as a series of steps. An alternative method of calculating survival probabilities, using a **lifetable** approach, can be used when the time to reach the endpoint is only known to within a particular time interval (e.g. within a year). The survival probabilities using either method are simple but time-consuming to calculate, and can be easily obtained from most statistical packages.

## Summarizing survival

We often summarize survival by quoting survival probabilities (with confidence intervals) at certain time points on the curve, for example, the 5 year survival rates in patients after treatment for breast cancer. Alternatively, the median time to reach the endpoint (the time at which 50% of the individuals have *progressed*) can be quoted.

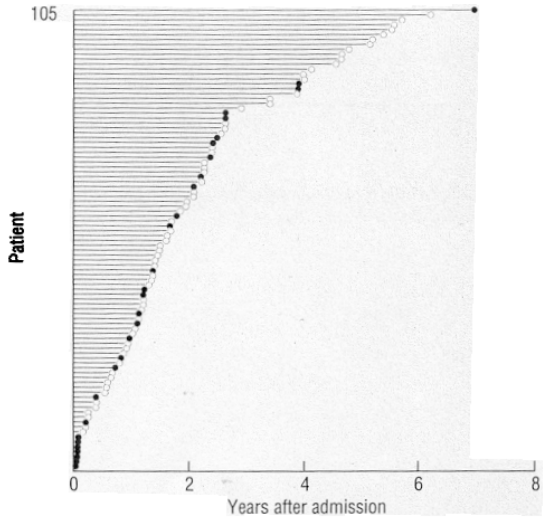



**Figure 44.1** Survival experience of 105 patients following admission with cirrhosis. Filled blank circles indicate patients who died, open circles indicate those who remained alive at the end of follow-up.

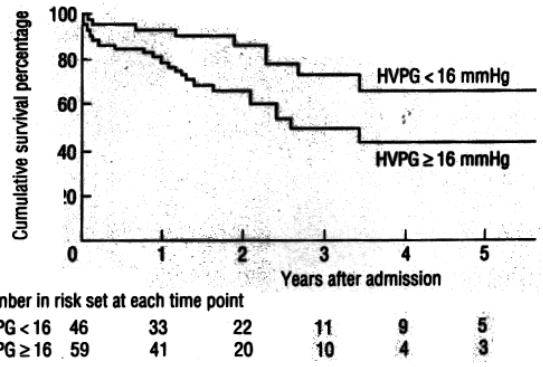



**Figure 44.2** Kaplan–Meier curves showing the survival probability, expressed as a percentage, following admission for cirrhosis, stratified by baseline HVPG measurement.

[1] Collett, D. (2003). *Modelling Survival Data in Medical Research*. 2nd Edn. Chapman and Hall/CRC, London.

## Comparing survival

We may wish to assess the impact of a number of factors of interest on survival, e.g. treatment, disease severity. Survival curves can be plotted separately for subgroups of patients; they provide a means of assessing visually whether different groups of patients reach the endpoint at different rates (Fig. 44.2). We can test formally whether there are any significant differences in progression rates between the different groups by, for example, using the log-rank test or regression models.

### The log-rank test

This non-parametric test addresses the null hypothesis that there are no differences in survival times in the groups being studied, and compares events occurring at all time points on the survival curve. We cannot assess the independent roles of more than one factor on the time to the endpoint using the log-rank test.

### Regression models

We can generate a regression model to quantify the relationships between one or more factors of interest and survival. At any point in time, $t$, an individual, $i$, has an instantaneous risk of reaching the endpoint (often known as the **hazard**, or $\lambda_i(t)$), given that s/he has not reached it up to that point in time. For example, if death is the endpoint, the hazard is the risk of dying at time $t$. This instantaneous hazard is usually very small and is of limited interest. However, we may want to know whether there are any systematic differences between the hazards, over all time points, of individuals with different characteristics. For example, is the hazard generally reduced in individuals treated with a new therapy compared with those treated with a placebo, when we take into account other factors, such as age or disease severity?

We can use the **Cox proportional hazards model** to test the independent effects of a number of explanatory variables (factors) on the hazard. It is of the form:

$$\lambda_i(t) = \lambda_0(t)\exp\{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k\}$$

where $\lambda_i(t)$ is the hazard for individual $i$ at time $t$, $\lambda_0(t)$ is an arbitrary baseline hazard (in which we are not interested), $x_1, \ldots, x_k$ are explanatory variables in the model and $\beta_1, \ldots, \beta_k$ are the corresponding coefficients. We obtain estimates, $b_1, \ldots, b_k$, of these parameters using a form of maximum likelihood known as **partial likelihood**. The exponential of these values (e.g. $\exp\{b_1\} = e^{b_1}$) are the estimated **relative hazards** or **hazard ratios**. For a particular value of $x_1$, the hazard ratio is the estimated hazard of disease for $(x_1 + 1)$ relative to the estimated hazard of disease for $x_1$, while adjusting for all other $x$'s in the equation. The relative hazard is interpreted in a similar manner to the odds ratio in logistic regression (Chapter 30) or the relative rate in Poisson regression (Chapter 31); therefore values above one indicate a raised hazard, values below one indicate a decreased hazard and values equal to one indicate that there is no increased or decreased hazard of the endpoint. A confidence interval can be calculated for the relative hazard and a significance test performed to assess its departure from 1.

The relative hazard is assumed to be constant over time in this model (i.e. the hazards for the groups to be compared are assumed to be *proportional*). It is important to check this assumption either by using graphical methods or by incorporating an interaction between the covariate and log(time) in the model and ensuring that it is non-significant[1].

Other models can be used to describe survival data, e.g. the **Exponential, Weibull** or **Gompertz** models, each of which assumes a specific probability distribution for the hazard function. However, these are beyond the scope of this book[1].

## Example

Height of portal pressure (HVPG) is known to be associated with the severity of alcoholic cirrhosis but is rarely used as a predictor of survival in patients with cirrhosis. In order to assess the clinical value of this measurement, 105 patients admitted to hospital with cirrhosis, undergoing hepatic venography, were followed for a median of 566 days. The experience of these patients is illustrated in Fig. 44.1. Over the follow-up period, 33 patients died. **Kaplan–Meier curves** showing the cumulative survival percentage at any time point after baseline are displayed separately for individuals in whom HVPG was less than 16 mmHg (a value previously suggested to provide prognostic significance) and for those in whom HVPG was 16 mmHg or greater (Fig. 44.2).

The computer output for the **log-rank test** contained the following information:

| Test | Chi-square | *df* | *P*-value |
|---|---|---|---|

Thus there is a significant difference ($P = 0.02$) between survival times in the two groups. By 3 years after admission, 73.1% of those with a low HVPG measurement remained alive, compared with 49.6% of those with a higher measurement (Fig. 44.2).

A **Cox proportional hazards regression model** was used to investigate whether this relationship could be explained by differences in any known prognostic or demographic factors at baseline. Twenty variables were considered for inclusion in the model, including demographic, clinical and laboratory markers. Graphical methods suggested that the proportional hazards assumption was reasonable for these variables. A stepwise selection procedure (Chapter 33) was used to select the final optimal model, and the results are shown in Table 44.1.

The results in Table 44.1 indicate that raised HVPG remains independently associated with shorter survival after adjusting for other factors known to be associated with a poorer outcome. In particular, individuals with HVPG of 16 mmHg or higher had 2.46 (= exp{0.90}) times the hazard of death compared with those with lower levels ($P = 0.04$) after adjusting for other factors. In other words, the hazard of death is increased by 146% in these individuals. In addition, increased prothrombin time (hazard increases by 5% per additional second), increased bilirubin level (hazard increases by 5% per 10 additional mmol/L), the presence of ascites (hazard increases by 126% for a one level increase) and previous long-term endoscopic treatment (hazard increases by 246%) were all independently and significantly associated with outcome.

**Table 44.1** Results of Cox proportional hazards regression analysis.

| Variable (and coding) | *df* | Parameter estimate | Standard error | *P*-value | Estimated relative hazard | 95% CI for relative hazard |
|---|---|---|---|---|---|---|
| HVPG* (0 = <16, 1 = ≥16 mmHg) | 1 | 0.90 | 0.44 | 0.04 | 2.46 | (1.03–5.85) |
| Prothrombin time (secs) | 1 | 0.05 | 0.01 | 0.0002 | 1.05 | (1.02–1.07) |
| Bilirubin (10 mmol/L) | 1 | 0.05 | 0.02 | 0.04 | 1.05 | (1.00–1.10) |
| Ascites (0 = none, 1 = mild, 2 = moderate/severe) | 1 | 0.82 | 0.18 | 0.0001 | 2.26 | (1.58–3.24) |
| Previous long-term endoscopic treatment (0 = no, 1 = yes) | 1 | 1.24 | 0.41 | 0.003 | 3.46 | (1.54–7.76) |

HVPG*, Height of portal pressure.

Data kindly provided by Dr D. Patch and Prof. A.K. Burroughs, Liver Unit, Royal Free Hospital, London, UK.



## *Preparation for Meeting 11*

In preparation of next week:

- In Field:
  - Sections 21.1-20.4, 21.5.1
- Self diagnosis on Brightspace
  - Survival analyses

## *Computer Session 10: Survival analyses*

### *Background*

The time to a certain event is often of interest in clinical or epidemiological follow-up studies. Examples of such time-to-event endpoints are the time from the first heart attack till the second heart attack or the time from inclusion into the trial till death. A typical problem in the analysis of such time-to-event data is that some patients will have experienced the event of interest (sometimes referred to as “failure”) before the end of the trial, while others will not have had this event. For the latter, the time-to-event is unknown, but we do know that the event has not occurred in the time frame the subject was followed i.e. the time-to-event for such subjects will exceed the their follow-up time. The (length of the) follow-up period of such subjects are called censored times-to-event and an appropriate statistical analysis has to account for the fact that data of such subjects are not actual times-to-event, but lengths of the period that the subject was followed and did not experience the event.

### *Objectives*

The student should be able to interpret descriptive survival analyses. This includes:

- interpreting a survival curve that describes (censored) times-to-event;
- producing a survival curve using the Kaplan Meier method ;
- sketching the hazard function for the risk of failure (i.e. the event of interest).

Furthermore, the student should be able to perform a simple proportional hazard analyses (Cox-regression).

### *Instruction*

Make the assignments and make notes of the arguments used for the choice of the parameters.

### *Product*

Answers on the questions of the assignments.

### *Discussion*

Results can be discussed with your peers and with the tutor during the computer session.

### *Assignments*

**Survival-Analyse bladder carcinoma**

In 1983 urologists in the region east of the Netherlands began to create a database containing the long-term treatment results of bladder cancer. The aim of the project is to ultimately compare patients undergoing only Trans Urethral Resection (TUR) of the tumor with patients undergoing TUR and an additional therapy (chemotherapy or radiotherapy). The question was whether these therapies differ in survival of the patient. Data are in *bladder.sav* containing the follow variables:

| | |
|---|---|
| Id | Identification number |
| Gender | Gender (1=Male, 2=Female) |
| Birthday | Date of birth |
| Date_diag | Date of diagnose |
| Stage | Stage of tumor (2=pTa, 3=pT1) |
| Grade | Grade of tumor (1=good, 2=moderate, 3=bad) |
| solitaire | 1=solitaire, 2 = multiple |
| Nareas | Number of areas with abbreviations |
| Biopsy | Outcome of biopsy |
| Therapy | 0=TUR only, 1=TUR+Chemo, 3=TUR+Radio |
| Recidive | 0=no, 1=yes |
| Died | 0=no, 1=yes |
| date_rec | Date of recidive or last follow up |
| date_ovl | Date of death or last follow up |

a) Study the meaning of the variables in the file. Which variables are required in order to be able to carry out a survival analysis?

For performing a survival analysis SPSS uses a variable that represents the time until the event (in this case, the death of the patient). You will need to make this variable. Use Date and Time wizard in the Transform tab to create a variable suitable for survival analyses.

b) Calculate the Kaplan-Meier curve for the survival time since diagnosis for each of the three therapy groups. Plot the survival function for the therapy groups.

c) On the basis of the figure, estimate the median survival duration for the therapies. Do these estimates correspond to the values that SPSS calculates?

d) What percentage of the patients in the groups is still alive after 2, 4 and 6 years respectively?

| After | TUR only | TUR + chemo | TUR + radio |
|---|---|---|---|
| 2 years | | | |
| 4 years | | | |
| 6 years | | | |

e) Test whether the survival curves differ between the three therapy groups using the log rank test.

f) Use Cox-regression (proportional hazard regression) to assess which covariates are prognostic for the survival. Is age at diagnosis prognostic for survival? What is the hazard ratio of a patient diagnosed at age 65 compared with a patient of age 55?

g) What is your conclusion with respect to the therapy? Motivate your answer.

**Kidney transplants**

The dataset *kidneytransplantation.sav* contains measurements made in 1228 patients who received a new kidney either from a living donor or a dead donor.

| **Variable Name** | **Variable Label (description)** | **Value Labels (coding values)** |
|---|---|---|
| ID | Patient identification number | unique number from 1863 to 110465 |
| SEX | gender | 1 = male<br>2 = female |
| AGE | Age at transplantation | 18 – 76 year |
| AGE_CAT | Age categories | 1: ≤ 40 years<br>2: between 40 and 50 years<br>3: ≥ 50 years |
| SBP | systolic blood pressure before transplantation | 80 – 240 mmHg<br>" . " is 'missing value' |
| DBP | Diastolic blood pressure before transplantation | 30 – 140 mmHg<br>" . " is 'missing value' |
| HPT | hypertension before transplantation | 1 = yes (SBP>140 and/or DBP>90)<br>0 = no<br>" . " is 'missing value' |
| DONOR | Type of donor | 1 = deceased donor<br>2 = living donor |
| DM | diabetes mellitus | 1 = present<br>0 = absent |
| VIT_STAT | vital status | 0 = alive<br>1 = deceased |
| FU | follow-up in years | 0 – 21 years<br>Until death<br>Or until GF if GF precedes death<br>or till end of follow-up if no GF and no death |
| REJECTION | Rejection transplant / Graft Failure (GF) | 1 = yes<br>0 = no |
| TimeToRejection | Time to rejection | Same as FU |
| PERIOD | period of transplantation | 1 = 1984 - 1988<br>2 = 1989 - 1993<br>3 = 1994 - 1997 |

Start SPSS and open *kidneytransplantation.sav.* During Computer Session 1 you inspected the variables.

*Survival*

a) Determine the survival in patients after kidney transplantation. Use Analyze – Survival - Kaplan-Meier. Select the appropriate variables for Time and Status respectively. Use Options…., Plots, Survival .

b) The mean age at transplantation for man and women in this study is 45 and 47 years, respectively. We know that in the Netherlands life expectancy for men at the age of 45 years is 30 years. For females of age 47 years the life expectancy is about 35 years. Compared to this, what can you say about the life expectancy for subjects that received a new kidney?

*Success probability of kidney transplantation*

c) What percentage of subjects in our sample experienced a graft failure (GF; rejection of the new kidney)?

d) Calculate the chance of rejection, using the procedure "One minus survival" :
**Analyze – Survival – Kaplan-Meier**. Select the correct variables for Time and Status respectively. Use Options... , Plots, *One minus survival .*
How does this differ from your answer to c)?

*Hypertension risk factor for graft failure?*

e) Is hypertension a prognostic factor with respect to graft failure?
Use crosstabs: **Analyze – Descriptive statistics – Crosstabs**. Perform a ***Chi-square analysis***. Ask for percentages in the cells. Which percentages do you base your conclusion on?

f) Is there a difference in mean blood pressure (systolic and/or diastolic) between patients who experience graft failure and those who don't? Is blood pressure a prognostic factor? Why or why not?

**g)** Is hypertension prior to transplantation a prognostic factor for 'rejection-free survival' (i.e. time to GF)? Make one graph displaying risk curves for patients with and without hypertension: **Compare factor... , log-rank.**

*Prognostic factors for graft failure*

h) Is there a difference in 'rejection-free survival' between patients who received a kidney from a living donor and patients who received a kidney from a dead donor?

i) Is the age at which the patient receives the transplant associated with rejection-free survival? Because age is a continuous variable we cannot apply the Kaplan-Meier approach (unless we decide to categorize age, but this entails some loss of information). ***Cox regression analysis*** can be used to control for this potentially confounding continuous variable: **Analyze – survival – Cox regression:** Time: Follow up; Status: rejection; Define event: 1; Covariate: age at transplantation; Options: CI for exp(b). What is the estimated hazard rate ratio (HR) for two patients who differ ten years in age at time of transplant? Provide also a 95%CI for this HR.

j) We can incorporate several (independent, prognostic) factors in a Cox regression model. Set up a model that uses age, type of donor (as categorical variable), hypertension, gender and diabetes. Interpret the resulting HRs.

k) Try to come to a more parsimonious prediction model for rejection-free survival time. (hint: use the selection procedure Backward LR) Which factors turn out to be the most important predictors for rejection-free survival time?

Please fill in the evaluation form at page 135.

## Meeting 11

### *Lecture 11: Logistic regression analysis: prediction*

Teacher: Rana Dandis

In clinical practice, there is often doubt between two potential diagnoses (e.g. an operable or an inoperable tumor, a mild or aggressive form of rheumatoid arthritis). A diagnostic test or a series of those tests is then used to throw light on this question. A good diagnostic decision rule can be developed in three steps.

Step 1: Gathering the data.

A so called 'learning set' is to be formed containing patients with known disease (gold standard) and known test results.

Step 2: Statistical analysis.

Statistical analysis of the learning set is needed to get an end-score that discriminates well between the two potential diagnoses. A prediction rule can be obtained by means of logistic regression.

Step 3: Formulation of the decision rule.

For a new patient who is not a member of the learning set, the diagnosis for this patient can be assessed by comparing her/his prediction with a threshold value (= cut-off point).

### *Preparation for Meeting 12*

In preparation for next week:

- Please read (Field):
  - Sections 21.1-21.4
- Self diagnosis on Brightspace
  - Logistic regression (prediction)

## *Computer Session 11: Logistic regression analysis*

*Background:*
To study the relation between a binary outcome (i.e. recovered or not recovered) and a predictor (i.e. age of a patient, blood pressure) logistic regression is most suitable.

*Objectives:*
Upon completion of this assignment the student should be able to apply logistic regression analysis in a diagnostic context.

*Instruction:*
Carry out the assignments with your partner.

*Product:*
The written answers to the questions. Please fill in the evaluation form at page 135.

*Discussion:*
Results can be discussed with your peers and with the tutor during the computer session.

*Assignment*

**A prediction model for metastasis for patients with prostate cancer**

Brown carried out a study to determine whether a combination of five variables could be used to forecast whether or not the cancer has spread to the lymph nodes. The original five variables were:

- age of patient at diagnosis (in years)
- level of serum acid phosphatase (in King-Armstrong unites)
- the result of an X-ray examination (0 = negative, 1 = positive)
- the size of the tumor as determined by a rectal examination (0 = small, 1 = large)
- a summary of the pathological grade of the tumor determined from a biopsy (0 = less serious, 1 = more serious, 2=very serious)

However, as only few patients had the pathological grade very serious, Brown decided to combine serious and very serious. This led to the variable grade_serious (0 = less serious, 1 = more serious or very serious). We will also use the variable grade_serious instead of the variable grade.

The result of the laparotomy is a binary response variable where zero signifies the absence of, and unity the presence of nodal involvement. The data are stored in the SPSS file *ProstaticCancer.sav*.

In this example, the five explanatory variables or *prognostic factors* are a mixture of dichotomous factors (X-ray result, size and grade of tumor) and continuous variables (age at diagnosis and level of serum acid phosphatase). The problem here is to identify whether all or just a subset of these variables are required in a model that could be used to predict nodal involvement.

An initial examination of the data reveals that the distribution of the values of *acid* is rather skew, and so the larger values (or outliers) may have a disproportionate influence on the outcome of the analysis (influential observations). This suggests that LN(Acid) should be taken to be the explanatory variable in modelling variation in the binary response variable.

1. Assess the association between the dichotomous variable X-ray and nodal involvement using logistic regression. Estimate and interpret the odds ratio (including the 95% CI).

2. Check that the LN-transformed variable *LNacid* is indeed less skew than the original variable acid.

In the rest of this assignment we will only use the log transformed values of acid (*LNacid*).

3. Assess graphically using box-plots whether or not there is an important difference in *LNacid* values between the group of patients with and without *nodal involvement*.

4. Perform a logistic regression analysis to assess the association between LNacid and the presence of nodal involvement.

   Estimate the odds ratio regarding nodal involvement of male A in comparison with male B, where LN(acid) of A is 0.1 larger than that of male B. Provide also a 95% confidence interval for this odds ratio.

   A difference of 0.1 on the logarithm of the acid level corresponds to a ratio of exp(0.1)=1.1 on the acid level, i.e. a difference of 10%.

   Can you now formulate your findings? What is the interpretation of the OR in this case? (Optional: what would be the odds ratio for a difference of 50%?)

   Are the odds ratio and the confidence interval relevant for the research question?

5. Determine the area under the ROC curve (AUC, also named concordance index or c-index) for nodal involvement versus *LN_acid*.

   Take the following steps:
   Run the previous analysis and save the predicted probabilities (= predicted values of the probabilities) as a separate variable in the de dataset (hint: see the options of the *Save* button).
   Then, plot a ROC curve and use the following option to obtain the AUC:
   Under analyse, ROC curve, test variable predicted probability, state variable = nodes, value of state variable = 1, display ROC curve with diagonal reference line.

Is the AUC (or c-statistic) relevant for the research question?

In the analyses of the Compliance Trial (page 43), you were not asked to calculate the AUCs. In fact, AUCs are never evaluated when clinical trials are analyzed. Why not?

6. Investigate which combination of the 5 variables is most suitable to predict *nodal involvement*. To this end, set up a model with all 5 predictors and perform an analysis with an automated selection procedure. (Use only the forward LR method and the backward LR method.)
   Which model is preferred and why?

   Determine the AUC.

   Does this model predict substantially better than a model with only LN_acid? Why?

7. Determine the AUC for the model with the variables LN_acid, Xray and size.

8. Compare the results of the previous three analyses.
   Please explain the paradoxical (?) results.

   Which model do you prefer (and why), based on these analyses?

   What additional analyses could you think of?

9. For the model that you prefer, check the predictive value. An easy way to do is to make use of the Hosmer Lemeshow test (use the SPSS button ‘options‘ in the logistic regression menu). The Hosmer Lemeshow test makes no sense, so do not pay any attention to it, but only inspect the ‘contingency table’.

What do you think of the accuracy of the predictions?

The model was derived on the Prostatic Cancer data set. After that, its performance (AUC) was measured on the same data set. As a result of this, the AUC may be too optimistic: the estimate may be too high. Why? What are your recommendations for further research to find a good prediction model for *nodal involvement* in patients with a prostate carcinoma?

Please fill in the evaluation form at page 135.

# Meeting 12

## *Lecture 12: Multilevel models*

Teacher: Ton de Haan

Experiments can have a hierarchical structure. For instance, in testing drugs often more than one centre is used and within centre the treatments are randomized. Such data can be handled using ANOVA. The situation is however more complex if within a hospital different wards are randomized to different treatments (education programs for instance) and then within each ward patients are followed to measure the results. Now the patients within a ward are no longer independent. Another example comes from animal experiments. Rodents are mostly housed in cages in small groups. Each cage is randomized to a treatment. The cages are situated in a rack and racks are situated in climate rooms. Here we can distinguish a four layer hierarchy: First layer consists of the different climate rooms, the second layer consists of the different racks within each climate room, the third layer consists of the cages within a rack, and the fourth layer are the animals within a cage. If we perform repeated measures on each animal we even have a 5 layer hierarchy.

## *Preparation for Meeting 13*

- Please read the article "*Review: A gentle introduction to imputation of missing values*" written by ART Donders , GJMG van de Heijden, T Stijnen, and KGM Moons in Journal of Clinical Epidemiology 59 (2006) 1087-1091. Link available on Brightspace
- Self diagnosis on Brightspace
  - Multilevel models

## *Computer Session 12: Multilevel models*

*Background:*
Longitudinal studies are defined as studies in which the outcome variable is repeatedly measured; i.e. the outcome variable is measured in the same individual on several different occasions. These forms of studies are examples of multilevel studies. Another example of such studies is cluster randomized trials or studies in families.

*Objectives*
The student should be able to:

- Perform simple repeated measure analyses and simple multilevel analyses using SPSS

*Instruction*
Read the questions carefully and write down your answers.

*Product*
The product will be the written answers. Please fill in the evaluation form at page 135.

*Discussion*
Results can be discussed with your peers and with the tutor during the computer session.

*Assignments*

**Growth of boys and girls**
For 11 girls and 16 boys the distance (mm) from the center of the pituitary to the pterygomaxillary fissure was recorded at the age of 8, 10, 12 and 14 years. See SPSS datasets *potthoff_miss.sav*. This data were also used and presented in Computer session 10. Use *potthoff_miss_tr.sav* for the analyses.

a. Make graphics that shows each individual growth curve, separate for boys and girls.

b. Compare the linear growth curves between boys and girls using a random effect model with only random intercepts. What fixed effects should you include? What is the mean growth (in mm/year) in boys and what in girls?

c. Compare the linear growth curves between boys and girls using a random effect model with random intercepts and slopes. What fixed effects should you include? What is the mean growth (in mm/year) in boys and what in girls?

d. If you compare both models (only random intercept versus both random intercept and random slope), what model do you prefer?

**The danger of daycare; pneumococcus in families**

The data in Table 1 are taken from a survey on the prevalence of upper respiratory tract infection. The variable to be analysed is the number of swabs positive for Pneumococcus during a certain period. Observations were made on 18 families, each consisting of a father, a mother and three children, the youngest of whom was always a preschool child. The children are numbered 1, 2 and 3 in descending order of age. Six families were a random selection of such families living in 'overcrowded' conditions, six were in 'crowded' conditions and six were in 'uncrowded' conditions[3].

**Numbers of swabs positive for Pneumococcus during fixed periods.**

| | | Family status | | | | |
|---|---|---|---|---|---|---|
| | | | | Child | | |
| Crowding category | Family serial number | Father | Mother | 1 | 2 | 3 |
| Overcrowded | 1 | 5 | 7 | 6 | 25 | 19 |
| | 2 | 11 | 8 | 11 | 33 | 35 |
| | 3 | 3 | 12 | 19 | 6 | 21 |
| | 4 | 3 | 19 | 12 | 17 | 17 |
| | 5 | 10 | 9 | 15 | 11 | 17 |
| | 6 | 9 | 0 | 6 | 9 | 5 |
| Crowded | 7 | 11 | 7 | 7 | 15 | 13 |
| | 8 | 10 | 5 | 8 | 13 | 17 |
| | 9 | 5 | 4 | 3 | 18 | 10 |
| | 10 | 1 | 9 | 4 | 16 | 8 |
| | 11 | 5 | 5 | 10 | 16 | 20 |
| | 12 | 7 | 3 | 13 | 17 | 18 |
| Uncrowded | 13 | 6 | 3 | 5 | 7 | 3 |
| | 14 | 9 | 6 | 6 | 14 | 10 |
| | 15 | 2 | 2 | 6 | 15 | 8 |
| | 16 | 0 | 2 | 10 | 16 | 21 |
| | 17 | 3 | 2 | 0 | 3 | 14 |
| | 18 | 6 | 2 | 4 | 7 | 20 |

Open the SPSS file A*&B_270_long.sav* and look at the structure of the dataset. Each family (= unit) has 5 repeated measurements ( 5 observations within each unit). You can analyse this data as follows: **Analyse – Mixed Models – Linear. >** Specify *Family serial number* in the subjects > continue .

a. If we want to investigate the relation between crowding and Family status (both categorical), how do we have to specify the model so that we can test whether the differences between the levels of Family status depend on the crowding category (for instance is the mean number of positive swabs between fathers and the third childs for Overcrowded families smaller/larger than for Uncrowded families)?

b. Build your model with random intercept for each family. Are the differences between the family status levels dependent on the crowding category? If not, fit a simpler model.

c. Describe the crowding effect on the number of positive swabs. Describe the family

[3] Data taken from Statistical Methods in Medical Research, third edition, Armitage & Berry, 1994, page 270.

status effect on the number of positive swabs.

d. Do you agree with the authors, Armitage and Berry, who said: *'There are clearly no significant differences between the father, mother and eldest child, but the two youngest children have significantly higher means than the other members of the family'* ?
   Carefully re-read the authors' statement.
   Take another careful look at the table of pairwise comparisons for the effect of FamilyMember.
   Do you agree with the authors?

e. Do we have enough data to judge the accuracy of a young mother's statement: "*Young children in daycare pick up a lot of infectious diseases from each other and bring them home to share with the rest of their family*"?

f. Do Moms, Dads and older children seem to be at greater risk because they have young children in the family?

g. What comparison(s) would you have to make to be able to properly judge the accuracy of her statement?

h. Based on Armitage and Berry's data, would you expect the parents of young children to get respiratory infections as often as their young children?

i. Based on Armitage and Berry's data, would you expect families with young children who live in a small apartment to get respiratory infections more or less frequently than families with young children who live in a villa?

**Improvement of the management of depression in nursing home residents of somatic and dementia**

In a 19-month longitudinal controlled study using a stepped wedge design, a new care program was implemented in 17 somatic and 16 dementia special care units.
The aim of this study was to evaluate the effectiveness of this care program to improve the management of depression in nursing home residents of somatic and dementia special care units. The care program is an evidence based standardization of the management of depression, including standardized use of measurement instruments and diagnostical methods, and protocolized psychosocial, psychological and pharmacological treatment.
Primary outcomes are the frequency of depression on the units and quality of life of residents on the units. Data of the quality of life outcome are available in *SteppedWedge.sav*.

a) What is the raw (not corrected for covariates) effect of the new care program. Hint: two types of random effects are needed: random unit and random patient within unit. The SPSS commands are:

   Analyze>Mixed models>Linear> Enter Unit and patnr to the 'Subjects' panel. In the 'Random' tab mark 'Include intercept' and move Unit to the 'Combination' panel. The click on the 'next' tab. Mark 'Include intercept' and move Unit and patnr together to the 'Combination' panel.

   Which factors and interactions do you need to include in the model to be consistent with the design? Perform a backward elimination by hand (i.e. fit several models starting with a complicated model including all kind of interactions and eliminate non-significant model components one-by-one starting with the highest order interaction. Be aware to keep hierarchy: include always lower interactions and main effects when a higher order interaction is included in the model)

b) What is the effect of the new care program, corrected for gender, age and region? Follow the same strategy as before.

Please fill in the evaluation form at page 135.

# Meeting 13

## *Lecture 13: 'A gentle introduction to imputation of missing values'*

Teacher: Rogier Donders

Missing data are a common problem in all types of medical research. There are various methods of handling missing data. Simple and frequently used methods include complete or available case analysis, the missing-indicator method, and overall mean imputation. However, these methods lead to inefficient analyses and, more seriously, commonly produce severely biased estimates of the association(s) investigated. There are more sophisticated (imputation) techniques to handle missing data, such as multiple imputation, that give much better results. With these techniques, missing data for a subject are imputed by a value that is predicted using the subject's other, known characteristics. Presently, these sophisticated techniques are easy accessible and available in standard software such as SAS and SPSS. In this lecture I will start with a brief introduction on different types of missing data and the principles of imputation in general, followed by explaining single and multiple imputation, and why frequently used methods fail. I will not go into technical details. Instead, I aim to clarify in simple wording why (more sophisticated) imputation is a better, more valid method than the simple and frequently used techniques for handling missing data.

## *Preparation for Meeting 14*

- *McKenzie, J. E., Beller, E. M., and Forbes, A. B. (2016) Introduction to systematic reviews and meta-analysis. Respirology, 21: 626–637. doi: 10.1111/resp.12783. (Link on Brightspace)*

- Self diagnosis on Brightspace
  - Missing data

## *Computer Session 13: Missing data*

*Background:*
Under the assumption of Missing At Random (or Missing Completely at Random), imputation can be used as a flexible manner of handling missing data

*Objectives:*
Upon completion of this assignment the student should be able to:
- perform a multiple imputation using SPSS
- use the imputed dataset in an analysis

*Instruction:*
Carry out the assignments with your partner.

*Product:*
The output, data files and written answers to the questions.

*Discussion:*
Results can be discussed with your peers and with the tutor during the computer session.

*Assignments*

**Deep venous thrombosis**

Data were obtained from a large cross-sectional study among adult patients with a suspicion of deep venous thrombosis (DVT)[4]. In brief, patients with a suspicion of DVT were consecutively included when they visited one of 110 participating primary care physicians in The Netherlands.
Suspicion of DVT was primarily based on the presence of at least one of the following symptoms or signs of the lower extremities: swelling, redness, or pain in one of the legs.
After informed consent, the primary care physician systematically documented the patient's history and physical examination. Subsequently, venous blood was drawn to measure the D-dimer level. Finally, all patients were referred to the hospital to undergo the reference test (repeated compression ultrasonography of the lower extremities) to determine the presence or absence of DVT.
For this exercise, we specifically selected two dichotomous variables (**kver3**: difference in calf circumference of 3 cm or more and **atraum**: history of a leg trauma) and one continuous variable (**dim**: D-dimer level). The data are in the data set 'dvtcomplete.sav'.

1. Please inspect the data? Can you see why in the final analysis dim was log transformed?

[4] For specific details and main results of the study, we refer to the literature (e.g. Oudega R, Moons KG, Hoes AW. Ruling out deep venous thrombosis in primary care. A simple diagnostic algorithm including D-dimer testing. Thromb Haemost 2005;94:200-5).

2. The purpose of the study was to construct a prediction model for DVT. Please fit such a prediction model using all predictors (Note: NOT all variables, use log transformed dim). Save the predicted probabilities. Create an ROC curve and calculate the AUC (area under the curve).

**Imputing an data set**
Now read in the data set 'dvtmisMAR.sav'

This is the same data set as you analysed in the previous assignment, but now with missing values for the logarithm of dim.

1. Inspect these data and fit the same model as in the previous assignment. Do you note any differences?

Next you can create imputation using SPSS's multiple imputation routine. Continuas variables should be set to '*scale'*. Start by inspecting the measurement level of the variables and correct these if necessary. Start by imputing using the default values.
Analyse the imputed data set using the same model again.

2. Do you note any differences between this analysis and the previous two?

Finally, if you have any time left, you can 'play' with custom imputation model. Note for example that Predictive Mean Matching gives very bad results. If you are feeling ambitious, you might even try to obtain an estimated AUC. Note however that this is not possible using SPSS in one go. You need to collect the estimated AUC's (and SE's) and do the pooling 'by hand' using for example a spreadsheet (MI.XLSX on Brightspace can do the calculations for you).

If there is still time left, you might even try it out on your own data, GIVEN that it would be sensible to do so…

Please fill in the evaluation form at page 135.

# Meeting 14

## *Lecture 14: Introduction to meta-analysis*

Teacher: Joanna in 't Hout

Nowadays, more and more knowledge on medical therapies is available, e.g. as papers on clinical studies. However, results of different studies will show variation, and even worse: studies can get contradictory or misleading along the way. It is too simple to just perform a head count, e.g. 3 studies saying yes minus 1 saying no means a positive conclusion. The one that says "no" might outweigh the others in validity and power.
You need to do a formal meta-analysis, to combine the information that is available in the studies. Meta-analysis is combining and analyzing data from more than one study at once, and has its own statistical techniques. Now that we're increasingly flooded with data and contradictory studies, you'll see meta-analyses more often. But whether it is appropriate to combine a set of studies is not always clear, especially if there is a lot of heterogeneity between the studies. Also the statistical technique will have its effect on the final conclusion: a random effects model may lead to a somewhat different conclusion than a fixed effects model. Selective reporting can bias the conclusion seriously. These issues will be discussed this afternoon, and with Excel we will get some practice.

Note: this text was inspired by Hilde Bastian, on "Absolutely maybe – evidence and uncertainties about medicine and life", a really funny and informative website on statistics (http://blogs.plos.org/absolutely-maybe/2014/01/20/5-key-things-to-know-about-meta-analysis/).

## *Preparation for Meeting 15*

In preparation for next week:

- Next week we are going to use R for data visualisation. Please read the instruction on Brightspace how to install R and R-studio on your own computer.
- Self diagnosis on Brightspace
  - Meta-analysis

## *Computer Session 14: Introduction to meta-analysis*

*This afternoon we use the add-on MetaXL for MS-Excel to do the excercises.*

*Background:*
Before the statistical meta-analysis can be done, the important figures have to be extracted from the selected articles. After entering these figures in an appropriate software tool, the meta-analysis results become easily available.

*Objectives:*
Upon completion of this assignment the student will be able to:
- Extract relevant data for meta-analysis from articles.
- Perform meta-analysis using the MetaXL-tool.

*Instruction:*
Carry out the assignment with your partner.

*Product:*
The output, datafiles and written answers to the questions.

*Assignments*

**Extracting data**
You will use two studies that have been conducted to assess the effectiveness of high doses of corticosteroids in patients with severe head injury. The references of the articles can be found on Brightspace and are listed below:

1. Giannotta S L, Weiss M H, Apuzzo M L J, Martin E. High dose glucocorticoids in the management of severe head injury. Neurosurgery 1984; **15**:497-501.

2. Grumme T, Baethmann A, Kolodziejczyk D, Krimmer J, Fischer M, Eisenhart Rothe B v, et al. Treatment of patients with severe head injury by triamcinolone: a prospective, controlled multicenter trial of 396 cases. Research in Experimental Medicine 1995; **195**:217-29.

From these papers, extract the data that are relevant to our meta-analysis. To this end, develop a form that allows you to record the following data:
- Reference ID: name of the first author, year of publication
- Study design: randomization, blinding
- Study population: what type of patients
- Treatment of patients in the experimental group
- Treatment of patients in the control group
- Follow-up duration
- Results: total number of patients in the experimental group; number of patients in the experimental group who were dead or in a vegetative state at the end of follow up; total number of patients in the control group; number of patients in the control group who were dead or in a vegetative state at the end of follow up.

You have 30 minutes to finish this assignment.
Send part of your results to the following e-mail account:
**32869205.biometrics@e.linoit.com**

**The mail should be formatted as follows:**

| | | | | |
|---|---|---|---|---|
| **Id1** | **dead/veg 1** | **Tot 1** | **dead/veg 0** | **Tot 0** |
| **Id2** | **dead/veg 1** | **Tot 1** | **dead/veg 0** | **Tot 0** |

**No additional information needed. Put your name as subject of the e-mail.**

After sending your results, we will discuss in plenary the results.

**Performing a meta-analysis**

The magnesium.xlsx file contains results from 16 randomized clinical trials, designed to investigate the effect of magnesium administration on mortality in patients with acute myocardial infarction. (Nuesch, E. and P. Juni (2009). "Commentary: Which meta-analyses are conclusive?" Int J Epidemiol 38(1): 298-303.)
The file is described below:

| Study Name | tot1 | dead1 | nodead1 | tot0 | dead0 | nodead0 |
|---|---|---|---|---|---|---|
| Morton 1984 | 40 | 1 | 39 | 36 | 2 | 34 |
| Rasmussen 1986 | 135 | 9 | 126 | 135 | 23 | 133 |
| Smith 1986 | 200 | 2 | 198 | 200 | 7 | 198 |
| Abraham 1987 | 48 | 1 | 47 | 46 | 1 | 44 |
| Feldstedt 1988 | 150 | 10 | 140 | 148 | 8 | 146 |
| Bertschat 1989 | 22 | 0 | 22 | 20 | 1 | 19 |
| Ceremuzynski1989 | 25 | 1 | 24 | 23 | 3 | 21 |
| Schechter 1989 | 59 | 1 | 58 | 56 | 9 | 54 |
| Pereira 1990 | 27 | 1 | 26 | 27 | 7 | 25 |
| Singh 1990 | 76 | 6 | 70 | 75 | 11 | 73 |
| Golf 1991 | 23 | 5 | 18 | 33 | 13 | 31 |
| Schechter 1 1991 | 89 | 2 | 87 | 80 | 12 | 78 |
| Thogersen 1991 | 130 | 4 | 126 | 122 | 8 | 120 |
| LIMIT-2 1992 | 1159 | 90 | 1069 | 1157 | 118 | 1155 |
| Schechter 2 1995 | 107 | 4 | 103 | 108 | 17 | 106 |
| ISIS-4 1995 | 29011 | 2216 | 26795 | 29039 | 2103 | 29037 |

As seen from the above list, there is a trial where the number of deaths in one of the two arms of the study is 0. In these cases it is recommended to add 0.5 to all cells of the standard 2x2 matrix (and 1 in the totals).

1. Use Excel to 'manually' calculate the risk ratio(RR), Ln(RR), and the SE of LN(RR). Below you can find the formula's.

   The risk ratio is calculated as the division of the Risk of Death in the Treatment and the control group.
   In the 2x2 matrix

| | Event | No event | Total |
|---|---|---|---|
| **GROUP 1** | a | b | a+b |
| **GROUP 2** | c | d | c+d |

   The RR is calculated as $\frac{\frac{a}{a+b}}{\frac{c}{c+d}}$ , so we create the RR for each study in the excel by entering the calculation in the function section.
   Ln(RR) is calculated with the Ln function.
   Var of Ln(RR) $= \frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}$
   SE= $\sqrt{Var}$

2. Do manually a fixed effect analysis using Excel and the calculated Ln(RR)'s and Var's or SE's.

We are now going to use MetaXL to answer the following questions. Add somewhere below the table the commands

```
=MAInputTable("Magnes_RR_IV","NumRR","IV",A2:G17)
=MAInputTable("Magnes_RR_RE","NumRR","RE",A2:G17)
```

3. Create the corresponding forest-plots for the fixed and random effects approaches.

4. What are Cochran's Q, $\boldsymbol{\tau^2}$, $\boldsymbol{I^2}$ ?

5. What are the pooled Risk Ratios of death using the Inverse variance method and the random effects method?

6. Explain the difference between the results of the above approaches.

7. Locate the study that seems to have the greatest impact on the assessment of the pooled Risk Ratios. Exclude this study from the analysis. What are now the pooled Risk Ratios of death using the Inverse variance method and the random effects method?

8. Are there indications of publication bias?

9. Make a Subgroup meta-analysis for small studies (N<200) and large studies respectively. Therefore you have to create in excel something like this:

| Small | Large |
|---|---|
| Bertschat 1989 | Schechter 2 1995 |
| Ceremuzynski 1989 | Thogersen 1991 |
| Pereira 1990 | Rasmussen 1986 |
| Golf 1991 | Feldstedt 1988 |
| Morton 1984 | Smith 1986 |
| Abraham 1987 | LIMIT-2 1992 |
| Schechter 1989 | ISIS-4 1995 |
| Singh 1990 | |
| Schechter 1 1991 | |

And then in one cell the command `=MASubGroups("Subgroup IV", A19, M14:N23).`
In this command the first parameter is just a name, the second is referring to one of the analysis already done, and the array points to the cells just created. Be sure that the 'colnames' (Small and large) are part of the selection.

Please fill in the evaluation form at page 135.

## Meeting 15

### *Lecture 15 Data visualisation*

Teacher: Jordache Ramjith

Recent technological developments and the onset of the world of big data, has changed the ways in which we visualize data. Data can often be understood better when represented in clever and intuitive ways. Many statistical software are able to produce graphs of high quality but it is usually more difficult to manipulate details in these graphs in an easily reproduceable way. The grammar of graphics (ggplot2) package in R provides a flexible way in which we can create good quality graphs that are easily amended to suit specifications. We will learn how to create simple charts (bar, line, points) and how to change the desired specifications like: reordering values in the x-axis, changing scales of an axis, relabeling axis and legends, flipping the coordinates, changing the background, changing colours, including text figures in the graph and joining multiple graphs.

### *Preparation for next week*

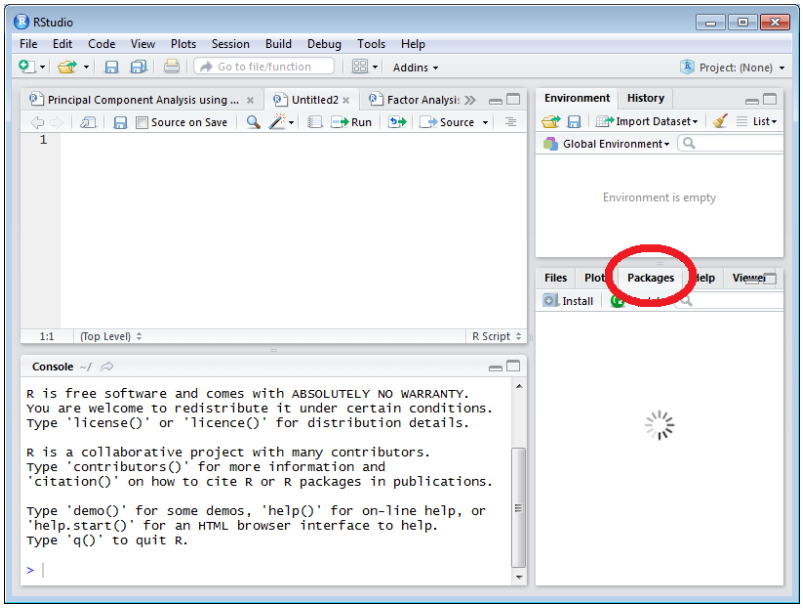
The next two weeks can be used to analyse your own data with the help of a biostatistician. A list will be made to assign a statistician to each participant. **You are supposed to contact the statistician by email or phone to make an appointment to discuss your study.** In collaboration with the statistician an analyses plan will be made and you can carry out the analyses yourself. The statistician can be called in case of emergency.

Preparation for examination: one example of an examination excercise can be found on Brightspace.

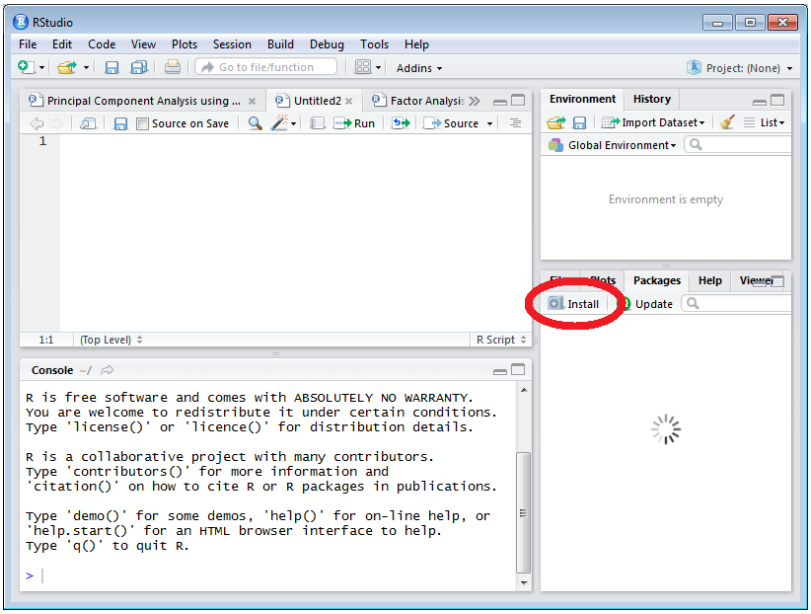
## *Computer Session 15*

### *Background*

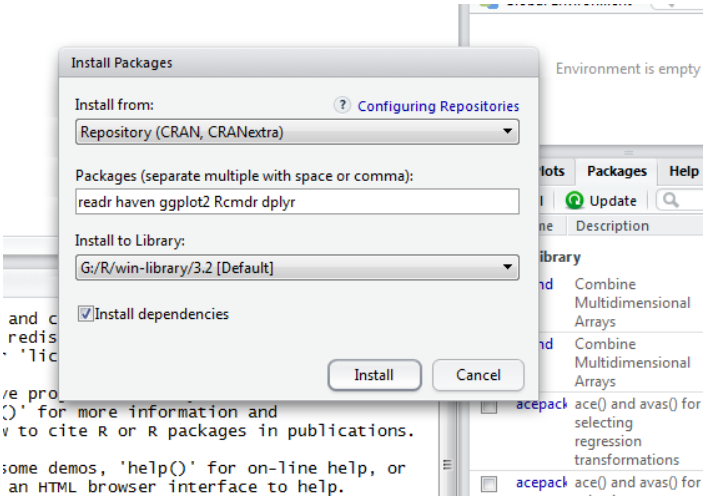
The required data needs to downloaded and saved in a folder. You will need to install specific packages. When you open RStudio, click on Packages.




Then click on install:

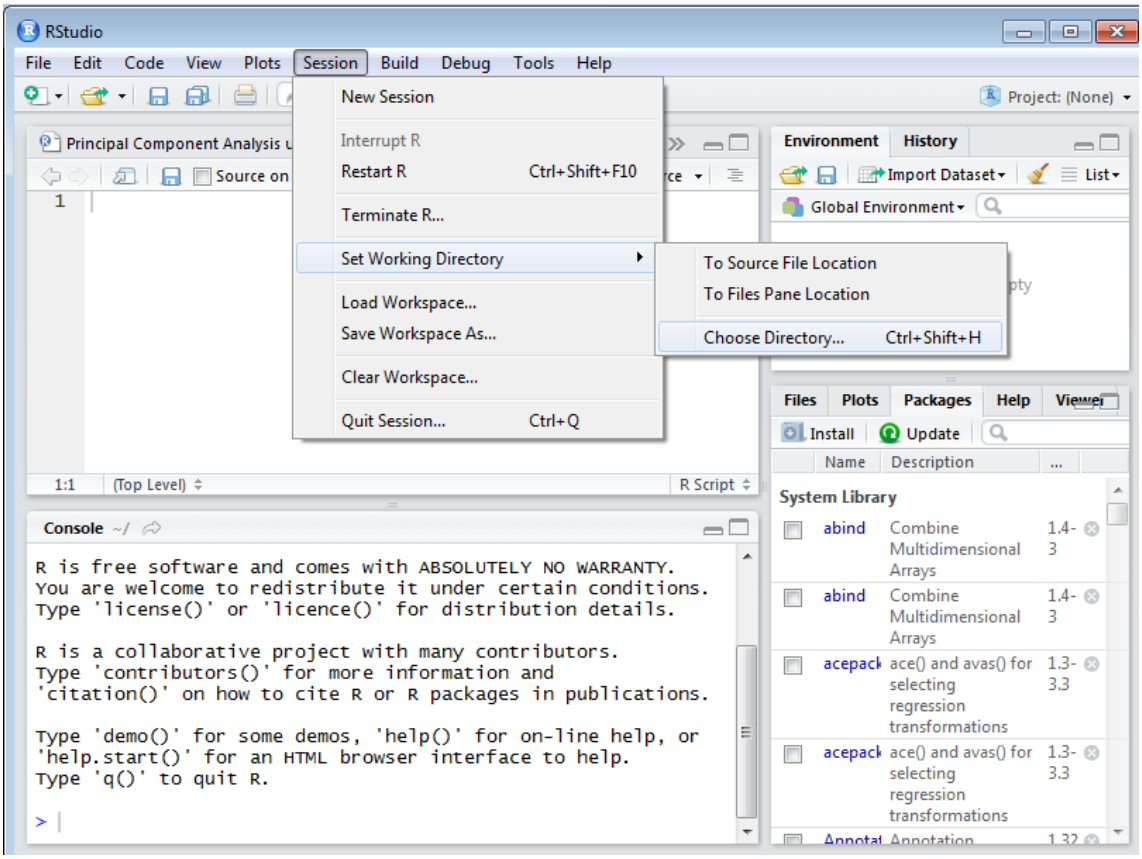



Then type in the names of the following packages we will need with spaces separating the names of these packages: readr haven ggplot2 Rcmdr dplyr




Finally you may click on install.

Also set up your working directory to the folder you have saved your data in. This ensures everything you work with will be saved in this folder and your data will be read from this folder without always needing to specify the folder directory. To do this, you may click on Session -> Set Working Directory -> Choose Directory and then proceed to the data we have worked from.




*Objectives:*

Upon completion of this assignment the student will be able to:

- Be able to plot basic plots using ggplot2 in R.

- Be able to perform some additional enchancements using ggplot2 in R.

*Instruction:*
Carry out the assignment on your own.

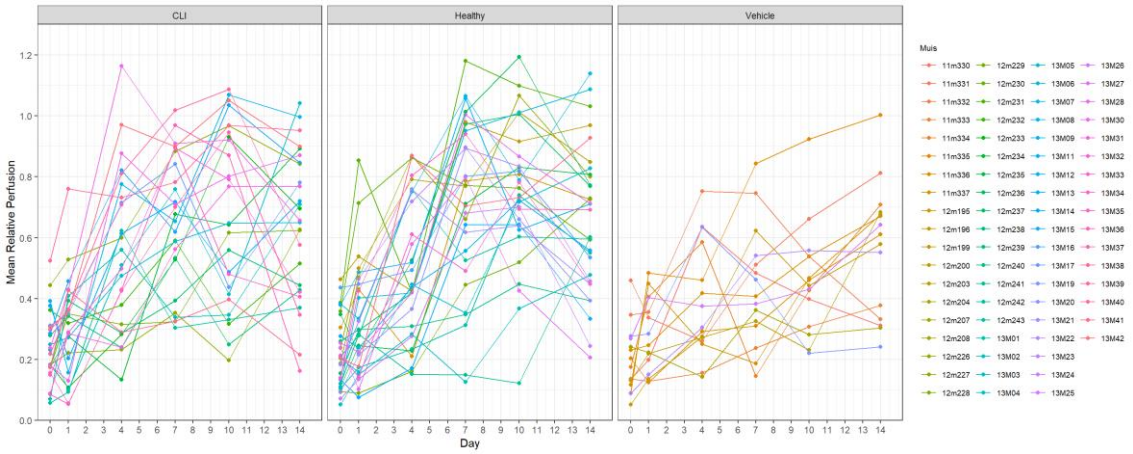
*Product:*
The R code and the saved images.

*Assignments:*

Assignment 1

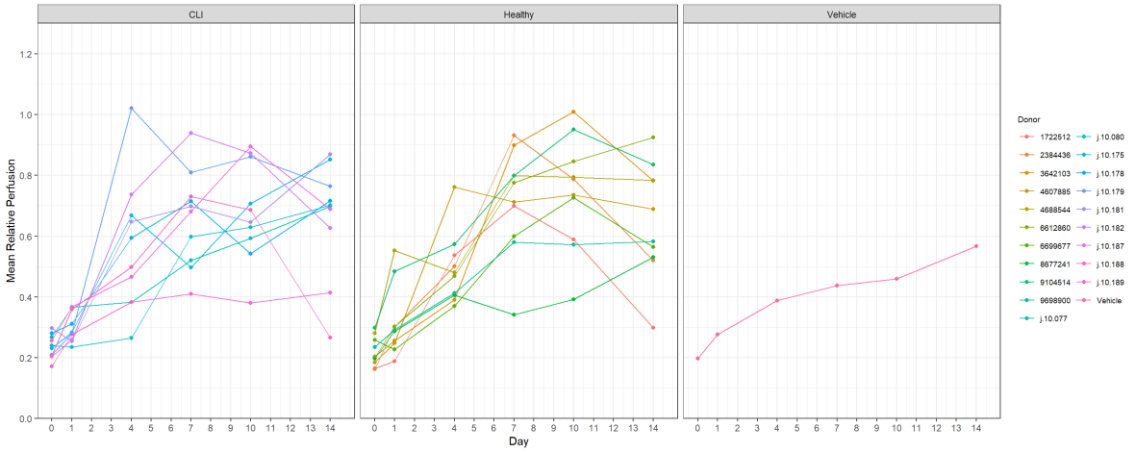
1. Replicate the following graph. The data can be found in `CLI_data_long.sav`

```
library(haven)
CLI_data_long <- read_sav("CLI_data_long.sav")
View(CLI_data_long)
```




2. In order to get the plot above by Donor instead of by muis:
   a) we could use the dplyr package to summarize the data into the raw means. Here are the code to determine the raw means in R, you may copy and paste this into R to get the summary dataset. Use this summary dataset to replicate the following plot:

```
library(dplyr)
mean_data <- group_by(CLI_data_long, Donor, Day, Group) %>%
dplyr::summarise(RelativePerfusion = mean(RelativePerfusion,
    na.rm = TRUE))
View(mean_data)
```

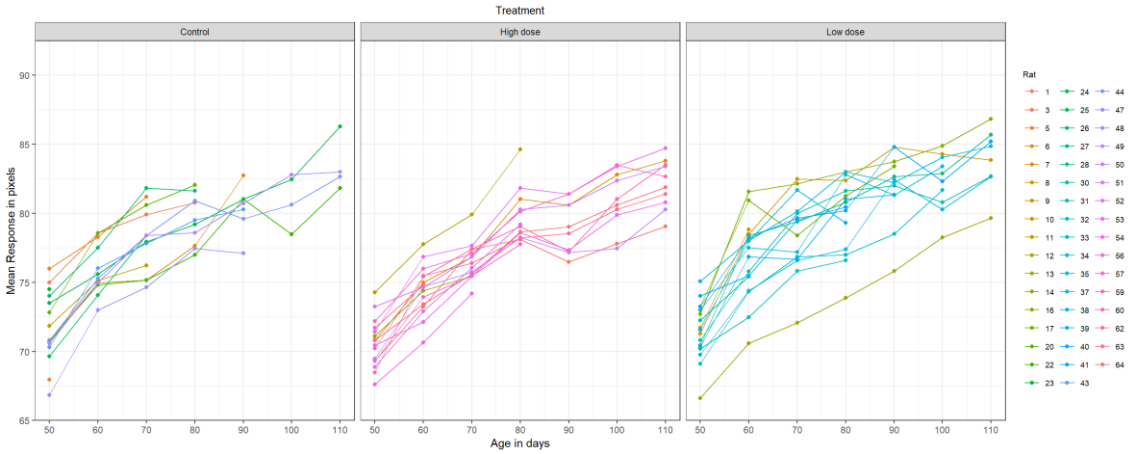



b) or we could create a new data set for the modelled mean. Please repeat the plot in (a) using the modelled means instead of the raw means. Create the model means with SPSS.

Assignment 2

1. Replicate the following graph. The data can be found in `SkullRats.sav`

```
library(haven)
SkullRats <- read_sav("SkullRats.sav")
View(SkullRats)
```



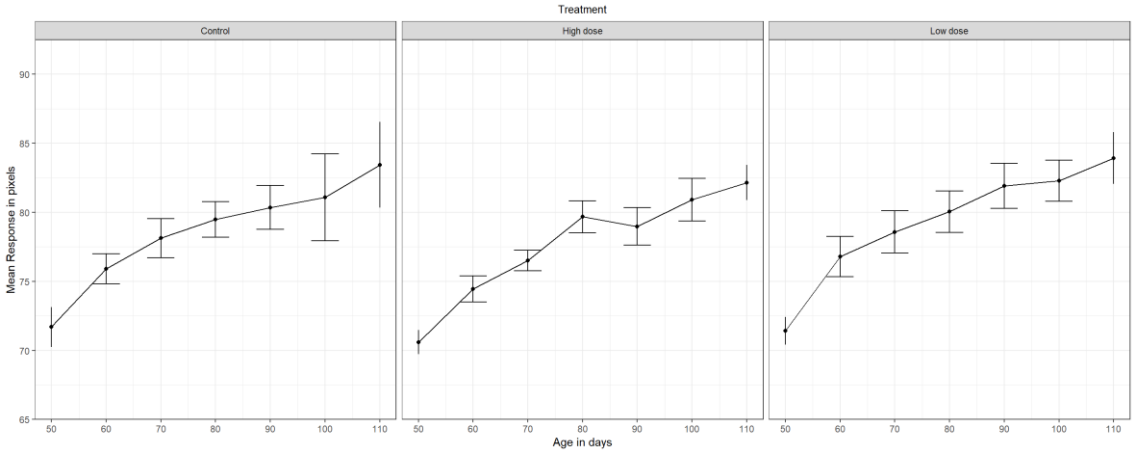

2. Create a graph indicating the mean growth of each group (average over the rats) We would be first required to find the summary data. i.e. find the mean and the confidence intervals.

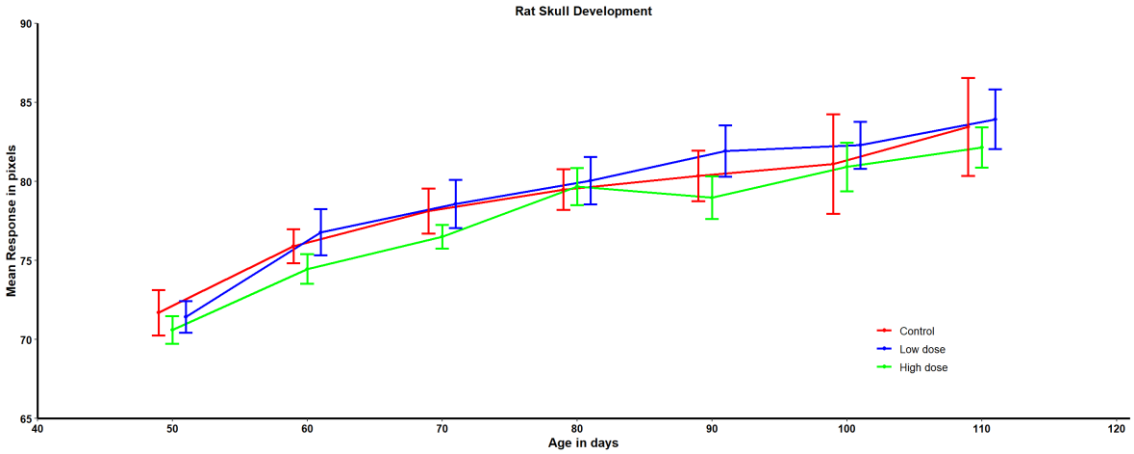
   a) First we will do this using raw estimates. We have used the dplyr package to previously find the means, we will now extend it to find the standard deviation and the sample size for each timepoint/treatment combination. We then calculate the SE's and we use the critical points of the t-distribution (we don't naively use 1.96). The code for programming this in R is:

```
mean_data <- group_by(SkullRats, age, treat) %>%
dplyr::summarise(mean.res = mean(response,
    na.rm = TRUE), sd.res = sd(response, na.rm = TRUE), n.res =
n()) %>% mutate(se.res = sd.res/sqrt(n.res),
    LCI = mean.res - qt(0.975, n.res - 1) * se.res, UCI = mean.res
+ qt(0.975, n.res -
        1) * se.res)

View(mean_data)
```

Now replicate the two different plots you could make from this data:




b) Now create a new dataset using SPSS containing the modelled means and confidence intervals, and then replicate the above two plots using this data.

## Additional assignments

### *Overview*

The choice for a statistical method depends mainly on the character of the outcome variable. Do we have a binary outcome (like cured or not), then the choice will contingency table analyses or logistic regression analyses. Contingency table analyses will be done when the other variables are nominal, such as treatment or gender, or ordinal, such as severity of the disease. If the interest is also in a continuous variable, like age or body weight, then logistic regression analyses is appropriate. Is the outcome variable
Is the outcome numerical continuous then depending on the structure of the study or experiment several statistical methods can be used. The main distinction is between studies with independent measurements and studies with dependent measurements like repeated measurements or multilevel models. Models with independent measurements can all be handled with multiple regression analyses. Models with dependent measurements can be handled using mixed models. However is the outcome is only partially known, such as the time to death with could be censored, then survival analyses should be applied.

*Background:*
A lot of statistical techniques and models are available for different kind of research questions. For non-statistician the choice of the appropriate statistical method is often unclear. The right choice depends on the outcome variable, research question and the design of the study.

*Objectives*
The student should be able to:

- Chose the right statistical method for commonly performed research in the biomedical field.

*Instruction*
Read the questions carefully and write down your answers.

*Product*
The product will be the written answers.

*Discussion*
Results can be discussed with your peers and with the tutor during the computer session.

*Assignments*

**Determinants of physical activity in Parkinson patients**
Parkinson's disease (PD) limits the performance of all kinds of physical activities. As a consequence patients are heavily inclined towards a sedentary lifestyle. A sedentary lifestyle is leading to lack of fitness, and patients may enter a downward spiral of inactivity. Limited research is however available identifying factors associated with physical activity in PD patients.
In this study 699 patients with PD are included. The level of physical activity is determined with the LAPAQ in minutes per day. The LAPAQ is an interview-based physical activity questionnaire which covers frequency and duration of different activities during the last week. Activities covered in the LAPAQ are: walking outside, bicycling, gardening, light and heavy household activities, and sport activities. The LAPAQ score is LN-transformed because of skewed distribution (LNLAPAQ).
Potential factors associated with time spent to physical activities are:

- HY, disease severity Hoehn and Yahr stage
- UPDRS, unified Parkinson disease rating scale
- LNCIRS, co-morbidity
- MMSE, mini mental state examination
- LNHADStotal, anxiety and depression
- TUG, mobility, performance on the timed up and go test
- LNFOGQ, freezing of gait
- FALLEN, fallen last year?
- LNFES, fear of falling
- LNSPDDS, disabilities in daily living

Background data included Gender ( 1= male, 2= female), Age, Education_level (1=low, 2=medium, 3=high) and Partner (1=single, 2= living with partner/children, 3=otherwise). The data are stored in the SPSS file Parkinson_LAPAQ.sav.

a. Assess the difference in LAPAQ between male and female PD patients. Give an interpretation on the original scale.

   Does the difference in LAPAQ between male and female PD patients depend on the age of the patient?

   Is age a confounder for the relation between gender and LAPAQ?

b. Identify the factors that are simultaneously associated with physical activity (LAPAQ). Choose your own strategy.

c. What is the performance of your model (model I)?

d. Address other models with the same number of factors that perform (approximately) equally well as model I , based on $R^2$.

e. Can model I be reduced without seriously affecting the performance of the model?

f. State your conclusion(s) regarding determinants of physical activity in patients with Parkinson disease.

**The role of the Mycobiome in human acute graft-versus-host disease**[5]

We are interested if Candida colonization can have a role of in the pathogenesis of acute GVHD in recipients of an allogeneic stem cell transplantation (SCT). These patients suffer from barrier dysfunction resulting from conditioning-induced intestinal mucositis, and changes in the gut mycobiome might have an impact on GVHD incidence.

We preformed a retrospective analysis in a very homogenous group of patients (N=153) receiving a matched-related partially T-cell depleted allogeneic SCT following myeloablative conditioning. All these patients had been treated similarly and received only cyclosporine for GVHD prophylaxis and standardized antimicrobial prophylaxis during the first weeks of SCT. Candida colonization was established in the first week after admission. Acute GVHD and GI-GVHD were scored according to the definitions of Przepiorka et al..

The main question is whether the presence of Candida species is related to the occurrence of acute GVHD or GI-GVHD.

Data can be found in the dataset *CandidaGvHD.sav*. Description of the dataset:

| **Variable** | **Code** | **SPSS name** |
|---|---|---|
| Conditioning | No TBI<br>TBI (total body irradiation) | Conditioning |
| Diagnosis | Other diagnosis<br>AML/MDS/ALL | Diagnosis |
| Gender combination | Other<br>M patient/F donor | Gendercomb |
| T-cell depletion | Elutriation<br>CD34<br>CD3/CD19 | Tcellmeth |
| Age, years | < 50 years<br>≥ 50 years | Age50 |
| SC source | BM (bone marrow)<br>PB (Peripheral Blood) | SCsource |
| Colonisation with Candida sp. | No<br>Colonisation | Colonisation |
| GI-GVHD | No<br>Yes, gut | GIGVHD |

[5] Data from Dr WJFM van de Velden, Haematology, UMC St Radboud

a. Is the colonisation of Candida related to the development of GI-GVHD? Find out which covariates are confounding this relation. A confounder is a covariate that is correlated with the colonisation of Candida and with the development of GI-GVHD. By taking a confounder into account the corrected relation between colonisation of Candida and the development of GI-GVHD is different from the raw relation.

b. Find out which covariates, except from Candida colonisation, describes the probability of developing GI-GVHD best? Can the model be improved by using the presence of Candida?

Please fill in the evaluation form at page 135.

## Evaluation form

| Subject | Subject was | | | | | Subject was | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Too Difficult | Difficult | Ok | Easy | Too easy | A waste of time | Not inter-esting | No opinion | Inter-esting | A must to known |
| 1. Variation | | | | | | | | | | |
| 2. Confidence intervals hypothesis testing | | | | | | | | | | |
| 3. Linear Regression | | | | | | | | | | |
| 4. The 2x2 table, Logistic Regression Analysis | | | | | | | | | | |
| 5. Sample size and power | | | | | | | | | | |
| 6. Comparing more than two means | | | | | | | | | | |
| 7. ANCOVA | | | | | | | | | | |
| 8. Experimental design | | | | | | | | | | |
| 9. Survival Analysis | | | | | | | | | | |
| 10. Repeated measurements | | | | | | | | | | |
| 11. Logistic regression analysis: prediction | | | | | | | | | | |
| 12. Multi level models | | | | | | | | | | |
| 13. Imputation of Missing values | | | | | | | | | | |
| 14. Meta-analysis | | | | | | | | | | |
| 15. Data visualisation | | | | | | | | | | |