Statistics and R

InterTU Study Day 2018 4TU.AMI

June 29, 2018; Maikel Verouden









Biometris Quantitative methods brought to life



Applied Mathematics in Education

- Several continuation courses assume knowledge on Mathematics, e.g.:
 - Introductory Physics (i.a. BBW, BMW, BAT)
 - Cell Biology and Advanced Imaging Technologies (i.a. BBI)
 - Models for Ecological Systems (i.a. MFN)
 - Functional Zoology (i.a. MBI, MAS)
- Statistics is part of Applied Mathematics and at Wageningen University & Research every student gets (an) introductory Statistics course(s).





\bigcirc

Brief Overview Contents Statistics Courses

Statistics 1

• Descriptive Statistics

- Probability calc. rules
- Normal Dist.
- Binomial Dist.
- Combining Dist. (μ and σ^2)
- Testing

Statistics 2

- Normality
- Sampling Dist.
- Student's t Dist.
- t-testing and CI
- SLR with CI, PI

Advanced Statistics

- Chi-square testing
- MLR
- Mixed Models
- AN(C)OVA
- Nonparametric

R for Statistics

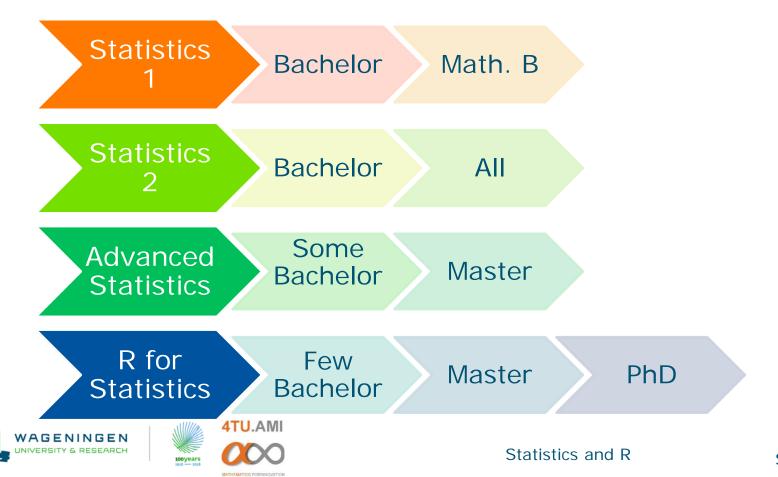
- Structures
- Sub setting
- Loop Functions
- Programming
- Visualisation
- Rep. Research





Which Students in Statistics Courses

 \bigcirc



Slide 4

\bigcirc

Statistics Course Material

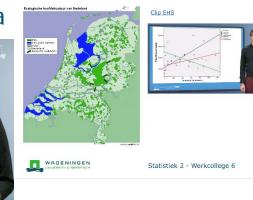
Prijs van landbouwgrond in relatie tot EHS

- Tailored with local Research Examples and Data
- Videos on YouTube
 - Refresh
 - Introduce
 - Motivate
 - Expand
- Self-tests on Blackboard
- Pen and Paper exercises (PPP)











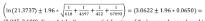


QUESTION 1 What is the research hypothesis? Formulate the alternarive hypothesis and cofine the parameters you use

Statistics and R

2 Mortality and birthweight, odds ratio

- a Estimate = 618 / 5215 = 0.1185 (round to 0.12), standard error = $\sqrt{0.1185 * (1 0.1185)/5215} = 0.004476$ (round to 0.0045).
- b 0.95-confidence interval: (0.1185 ± 1.96 * 0.004476) = (0.110, 0.127)
- c Estimate odds ratio = ⁶¹⁸⁺⁶⁷⁰⁷³/₄₂₂₊₄₉₇ = 21.3737 (round to 21.4). R gives a different estimate (21.36889), a so-called conditional maximum likelihood estimate. Details are beyond the scope of this course. In this case you can see that the difference is quite small.
- d First, construct an approximate 0.95- interval for ln(OR)



(2.935, 3.190). Second, take the exponential function of the lower and upper bound to get the approximate 0.95-confidence interval for OR: (exp(2.9347), exp(3.1896)) = (18.8, 24.3). This

2



Latest Change: Introduction of

What is it?

A language and environment for statistical computing

Why?

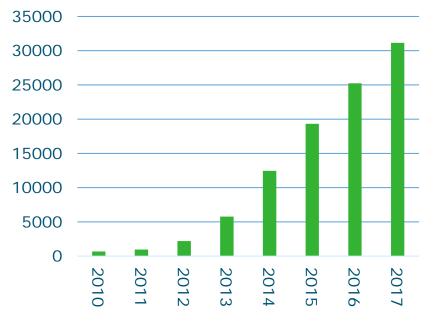
From a didactic point of view

- R has a steep learning curve
- One software for all courses
- By popular demand
 - Programme Directors
 - Students





R Core Team Citations



Source: Google Scholar search 'R Core Team'

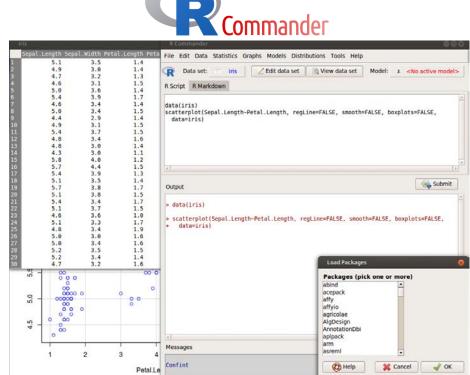
Statistics and R

Statistics 1

- Microsoft Excel used in computer practical
- Academic Year 2018-2019 a switch to R Commander will be made
 - A Point-and-Click Interface for R
 - Focus on understanding output, not learning R







Statistics and R

Statistics 2

- IBM SPSS Statistics
 - Output:
 - Lectures, exercises
 - Exam questions
 - Computers practicals

| When | introduction | of R | Commander | in Statistics 1 | successful |
|------|--------------|------|-------------|-----------------|------------|
| | | | Communation | | 3400033141 |

Possible switch to R (Commander) in Academic Year 2019-2020





| | | | 1 | | Direct Markel | ing <u>G</u> raphs (| tilites Add-on | | | 0 | - | | | | | |
|---------|-------|---------------|--------|------|---------------|----------------------|--|------------|------------------|------|-----------|------|-------|-----|------|-----------------|
| | | | | | - | | | | | | | | | | Visi | ble: 7 of 7 Var |
| _ | _ | id | school | male | math | langarts | daysatt | daysab | a var | VDF | var | var. | . var | Var | var | Var |
| 1 | | 1001 | 1 | 1 | 56.99 | 42.45 | | | 4 | | | | | | | |
| 2 | | 1002 | 1 | 1 | 37.09 | 46.82 | | | 4 | | | | | | | |
| 3 | | 1003 | 1 | 0 | 32.28 | 43.57 | 76 | | 2 | | | | | | | |
| 4 | | 1004 | 1 | 0 | 29.06 | ta One- | Dne-Sample T Test | | | | | | | | | |
| 5 | | 1005 | 1 | 0 | 6.75 | | | | | | | | | | | |
| 6 | | 1006 | 1 | 0 | | 1d | | - | Test Variable(s) | | Options | | | | | |
| 7 | | 1007 | 1 | 0 | 56.99 | | s math nce lmat | N | an school | - | Bootstrap | | | | | |
| 8 | | 1008 | 1 | 1 | 10.39 | | as lang noè (lang | | and there | | | | | | | |
| 9 | | 1009 | 1 | 1 | 50.53 | da da | days attend (daysat) days absent (daysat) | | | | | | | | | |
| 10 | | 1010 | 1 | 1 | 49.47 | da da | | | | | | | | | | |
| 11 | | 1011 | 1 | 0 | 39.56 | | | | | | | | | | | |
| 12 | | 1012 | 1 | 1 | 33.74 | | | | | | | | | | | |
| 13 | | 1013 | 1 | 0 | | | | | Test Value: 0 | 0 | | | | | | |
| 14 | | 1014 | 1 | 0 | 65.56 | | ОК | Paste | Reset Cancel | Help | | | | | | |
| 15 | | 1015 | 1 | 1 | 23.01 | | | Chancel of | Trans Constant | | | | | | | |
| 16 | | 1016 | 1 | 1 | 75.83 | 61.04 | | | 0 | | | | | | | |
| 17 | | 1017 | 1 | 0 | 41.31 | 49.47 | | | 1 | | | | | | | |
| 18 | | 1018 | 1 | 0 | 41.89 | 65.56 | | | 0 | | | | | | | |
| 19 | | 1019 | 1 | 1 | 65.56 | 46.82 | | | 2 | | | | | | | |
| 20 | | 1020 | 1 | 1 | 13.13 | 6.75 | | | 24 | | | | | | | |
| 21 | | 1021 | 1 | 0 | 33.02 | 42.45 | 75 | | 2 | - | | | | | | |
| - | _ | 4 | - | | | | | | | | | | | _ | | |
| Data Vi | Yew 3 | Variable Viev | | | | | | | | | | | | | | |

Advanced Statistics

- IBM SPPS Statistics
 - Output:
 - Lectures, exercises
 - Exam questions
 - Practicals
- R (Rgui/Rstudio)
- Some DIY exercises during computer practicals
- Output in Pen and Paper Practicals and exam questions
- In two years full switch to R





SPSS



R for Statistics

- Working with R in previous courses
 - With a focus on Statistics



- R is required/useful in many graduation/PhD projects
 - Need to focus on R
- 'R for Statistics' as optional course
 - With a focus on R with Rstudio (IDE around R)
 - Application by repetition of previous acquired Statistics knowledge







Teaching 'R for Statistics'

- R can not be learned from lectures
- An intensive practical approach is used:
 - Small introduction R internals
 - Exercises



- Introduction application previously learned statistical analyses in R
 - Exercises
- Assessment: end-of-the-week $(2 * \frac{1}{6})$ and final assignment $(\frac{4}{6})$ reports





\bigcirc

Student Survey Results (after 2 runs)

| Opinions | Weighted Average | Pooled Stand. Dev. |
|------------------|------------------|--------------------|
| Course Materials | 4.35 | 0.56 |
| Motivation | 4.26 | 0.85 |
| Workload | 3.15 | 1.25 |
| Challenge | 4.55 | 0.60 |
| Satisfaction | 4.35 | 0.65 |





Student Evaluation Remarks



"The workload is too high for a 3 ECTS course. Additional time is required in the afternoon to finish all assignments and prepare for the next day."

"The course is intensive, but very useful to get a basic understanding of of R in a relative short time. I would recommend it to everyone wanting to learn R."





Biometris collaborated statistical courses using R

- Modern Statistics for the Life Sciences (Animal Breeding and Genetics, ABG-30806)
- Data Science MSc, Leiden Universiteit
- Graduate Schools:
 - Production Ecology and Resource Conservation (PE&RC)
 - Voeding, Levensmiddelentechnologie, AgroBiotechnologie en Gezondheid (VLAG)
 - Wageningen Institute of Animal Sciences (WIAS; not yet fully in R, but will be completely in the near future)









Biometris Quantitative methods brought to life







Statistics and R

Slide 15