EXAM Basic Skills - Part 1: Theory Stats

Johnny van Doorn & Abe Hofman
UvA

March 2021

Instructions

- Write your name and student number in the upper left corner of the first page
- Include a photo of your student card on the first page of your answers
- Number your pages in the lower right corner
- Clearly indicate which question each of your answers refers to
- Provide a single answer to each question. If you provide multiple possible answers we will only mark your first answer and ignore all others
- When you have completed the theory part of the exam, take a photo of your answers and upload a PDF scan of your answers. Your file name should be *Surname_FirstName_StudentNumber.PDF*
- After you have uploaded your answers, remain seated and wait until the invigilator gives you the signal that you can take a break
- **If you are done early**: Upload your answers and wait for the invigilator to check your upload. You have to **remain seated and in view of the camera until the end of the theory part** of the exam. You may read a book or engage in other quiet activities but **you are not allowed to use any electronic devices until the end of the theory part**.
- Time for this part of the exam: 30 min
- Submit your answers through Canvas before 11:10
- The number in front of the question indicates how many points can be obtained (8 total)

1. **P-values.**

   (a) (1 pt) Explain what a p-value means in general (definition).

   (b) (1 pt) Explain why a p-value is a conditional probability.

2. **Three Researchers.** A team of three researchers conduct a two-sided correlation test ($n = 30$). They observe a correlation coefficient of 0.41, with a corresponding p-value of 0.026 and a 95% confidence interval ranging from 0.055 to 0.67. Each researcher attempts to interpret these results:

   (a) If we use a one-sided alternative hypothesis (correlation > 0) instead of a two-sided alternative hypothesis (correlation $\neq 0$), our result is significant for $\alpha = 0.025$.

   (b) A 90% confidence interval will not include 0.

   (c) If we would repeat this experiment, in 95% of the cases the correlation would be between 0.055 and 0.67.

   (3 pts) For each of the three researchers (a, b, c), say whether they are right or wrong, and explain why.

3. **Over-fitting.**

   (a) (2 pts) Explain what over-fitting is AND why you want to prevent over-fitting

   (b) (1 pt) Describe two distinct methods to prevent over-fitting, and briefly explain how these work.