Statistical Methods for Bioinformatics

June 16, 2024.

Instructions:

Practical exam: The second part is an **open book** practical exam, so you can use additional information sources. Any form of communication is forbidden; online communication will be monitored and will lead to immediate exclusion from the exam. You will have an additional two hours for the practical part. Write your answers on the provided pages. For the multiple choice questions, on the accompanying answer sheet check off the answer that in your opinion is most appropriate by drawing a (filled) circle on your chosen option. Only one answer should be checked for each question, if more than one answer is correct then select the most precise one (Question 9 to 13). Once you are ready (or after two hour) you can hand in the practical exam. **Please keep all sheets stapled**.

The data set for the assignment can be download from Toledo:

Part I> Exam-Practical

IMPORTANT NOTE: After the exam you should hand in the all exam questions.

Educational study:

In an educational research study a new and standard arithmetic training programs were compared. The objective of the study was to evaluate if, as an average, the new program can produce a faster increase in arithmetic ability over time than the standard program for third year primary school children. To that end a total of 200 children were randomized to follow the new or standard program for 5 months and their arithmetic ability was assessed using a validated test (SAT test) with larger values indicating higher arithmetic ability. All the students were assessed at the beginning of the study (after randomization but just before starting the training) and then monthly during the duration of the study. No missing values were present in the data. The final data set had the following variables

- id: a number identifying the student
- y: response variable containing the scores of the SAT test
- time: evaluation moments, 0 (before starting the training), 1, 2, 3, 4 and 5 months after starting the training
- prog: indicator variable for the program taking values 1 for the new training program and 0 for the standard one

ID-number:..... Name:.....

Practical Exam

Fit the following hierarchical model to the data using maximum likelihood:

$$Y_{ij} = \pi_{0i} + \pi_{1i}\text{time}_j + \varepsilon_{ij}$$

where i denotes the student, j the measurement and $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$ and

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} PROG_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{11} PROG_i \end{cases}$$

and $b_{0i} \sim N(0, \sigma^2_0)$

9. Based on the point estimates obtained from the previous model which of the following statements is correct?
   a. The variability between individuals is larger than the variability within individuals
   b. The variability between individuals is smaller than the variability within individuals
   c. The variability between individuals is equal to the variability within individuals
   d. The previous model does not allow to answer this question

10. Based on the previous model which of the following statements is correct?
   a. As an average, the new program is significantly better than the standard program
   b. As an average, the new program is significantly worse than the standard program
   c. As an average, based on the point estimates, the new program seems to be worse than the standard program but the result is not significant
   d. As an average, based on the point estimates, the new program seems to be better than the standard program but the result is not significant

11. Based on the previous model which of the following statements is true?
   a. The point estimate for the main effect of program is negative and significant
   b. The point estimate for the main effect of program is negative and not significant
   c. The point estimate for the main effect of program is positive and significant

d. The point estimate for the main effect of program is positive and not significant

12. Based on the previous model what is the average SAT score for children in the standard program one month after starting the experiment?

    a. 6.32
    b. 5.65
    c. 0.48
    d. 6.8

13. Based on the previous model which of the following statements is correct?

    a. The p-value associated with γ_{01} is smaller than 5%
    b. The p-value associated with γ_{01} is between 5% and 80%
    c. The p-value associated with γ_{01} is larger than 90%
    d. None of the above