

Medical Statistics 2: STAT0015

In-course Assessment: 2021-22

Information

- Your solutions should be your own work and are to be uploaded on Moodle by **Tuesday 22nd February 2022 at 5pm.**
- You are required to hand in your ICA report as a single **PDF** file.
- Name your report as: **studentid_report.pdf**. For example, if your ID is 123456 the file should be named **123456_report.pdf**.
- Do not include your name on your work. Please include only your student number (SRN, as printed on the front of your UCL identity card).
- This ICA contributes 20% of your overall mark for this module.
- You will receive a provisional grade for this ICA in the new year. All grades are provisional until confirmed at the Statistics Examiners' Meeting in June 2022.

Medical Statistics 2: STAT0015

In-course Assessment: 2021-22

Plagiarism and collusion declaration

A copy of the guidelines on plagiarism and collusion are available in the student handbook. Please ensure that you have read these.

In submitting your report online, you will be prompted to affirm that you have read the guidelines on what constitutes plagiarism and collusion and that the work being submitted is entirely your own.

The plagiarism and collusion declaration is as follows:

“I am aware of the UCL Statistical Science Department's regulations on plagiarism and collusion for assessed coursework. I have read the guidelines in the student handbook and understand what constitutes plagiarism and collusion.”

In Course Assessment

You will be using data from an asthma trial. Briefly, 157 children with asthma were randomised to either salmeterol xinafoate (SM) or placebo to investigate whether the intervention could increase peak expiratory flow rate (PEFR). PEFR was measured every three months, for a year.

Dataset

I have created an individual dataset for each student, which is a random subset of the original dataset. This has the name **ica-data-number**, where *number* is your student number. You will find your dataset in the folder 'ICA Datasets'.

Your dataset is in long format and contains the following variables.

id	patient ID number
drug	drug indicator: 0 = placebo; 1 = SM
month	month: 3, 6, 9 and 12
baseline	baseline PEFR (l/min)
pefr	PEFR measurement (l/min)

Your Task

Answer the following questions clearly in the form of a short report, i.e. use sentences and paragraphs, rather than bullet points.

There is no page limit, but you are not expected to write more than two pages of text. You can present as many tables and figures as you think is appropriate.

Please include your Stata do-file as an appendix but do not include your log-file.

Please note that raw Stata output (except figures) should not be copy and pasted straight into your answers. Select which output is necessary and format it appropriately. Also, please do not display p-values as e.g. $P = 0.000$. This should be changed to $P < 0.001$.

Questions

[The relative weight of each part is given in the parentheses]

1. Summarise the data, including its structure. Use both numerical and graphical methods as appropriate. **[7]**
2. Investigate the effect of salmeterol xinafoate on PEFR using a summary statistics analysis. Check all model assumptions and interpret the results. **[4]**
3. Investigate the effect of salmeterol xinafoate on PEFR using a hierarchical regression model. Then re-fit the model and adjust for month and baseline PEFR.

Check all model assumptions and interpret the results. Comment on how the results change in the adjusted model (with a focus on the intervention and variance estimates). **[6]**

4. Investigate the effect of salmeterol xinafoate on PEFR, adjusted for month and baseline PEFR, using an appropriate marginal model fitted using GEE. Interpret the results and compare them to the results for the corresponding hierarchical model. **[3]**