

You are working as a product developer in a pharmaceutical company in the field of ophthalmology. The product you are working on consists of a novel therapy for elevated ocular pressure based on eye drops. The eye drops are supposed to lower ocular pressure. To evaluate the product, you need to run a series of experiments

- a) First, a pilot study on rats is conducted. In this study, the objective is to compare the ocular pressure of rats who received no eye drops with that of rats who received the eye drops treatment. Data on pressure measurements in the group who did not receive eye drops is in the file `rat_pilot_data_43.dat`, where, in the file, the numbers represent pressure measurements in mmHg.
 - i) Determine the confidence interval for mean ocular pressure in rats in the absence of eye drops.
 - ii) Determine the minimum number of rats that needs to undergo the eye drop treatment in order to be able to detect a difference of 3 mmHg with a statistical power of 85%. You may assume the underlying standard deviation $\sigma = 3.5$ and the number of rats who did not receive the eye drops to be 22.
- b) Second, a pilot study on rabbits was conducted. In this study, ocular pressure was measured with and without the eye drops treatment on two distinct groups of rabbits. Data from this study are in two files. The file `rabbit_no_drops_43.dat` contains the data of rabbits who did not undergo the eye drops treatment. In this file, the first column is an identifying rabbit number and the second column is ocular pressure for this group of rabbits. The file `rabbit_drops_43.dat` contains the data of rabbits who did undergo the eye drops treatment. In this file, the first column is an identifying rabbit number and the second column is ocular pressure after the eye drops treatment. Your data analysis will aim at unveiling any change in ocular pressure between the two groups. You may assume equal variance.
- c) The third study is a clinical trial where the new eye drops are compared against the three best selling currently available eye drops for the same condition. For this study, you are given ocular pressure data from glaucoma patients using the three available eye drops on the market. Data are in the files `competitor_1_data_43.dat`, `competitor_2_data_43.dat` and `competitor_3_data_43.dat` respectively for three distinct groups of glaucoma patients. The file `drops_data_43.dat`, instead, contains ocular pressure data on glaucoma patients using the new eye drops that you are working on. Here, the required data analysis is intended to unveil whether the new eye drops improve on currently available alternatives in terms of lowering ocular pressure.

The data analysis of all studies should be conducted at the 97% confidence level, i.e., $\alpha = 0.03$.

INSTRUCTIONS

You need to submit to LumiNUS one Python file and one pdf file.

- A file called `Jennifer_43.py` containing your commented code to solve all the questions. In your code you may choose the name of the variables. However, you are expected to use the following variable names
 - `upper_bound_rats` and `lower_bound_rats` for the requested upper and lower bounds of the confidence interval in *a*),*i*
 - `n_rats_drops` requested answer in *a*),*ii*
 - `p_value_rabbits` for the *p* value associated with what you think is the most appropriate analysis in *b*)
 - `p_value_trial` for the *p* value associated with what you think is the most appropriate overall analysis of the clinical trial data in *c*)

Please do not submit any given data file (I have them) nor any other data file. I will place your code in the same directory as your assigned data files and your code is expected to work in such conditions.

- A pdf file called `Jennifer_43.pdf` containing your answers to the problems. The answers are expected to be the required data analysis reports for each of the three studies i.e., the key results as well what you can and can't claim after performing the necessary calculations and your reasons. An overall assessment of the new eye drops based on all collected data is also expected (one or two sentences). Your answers can include any graph (exported from Python) and/or text explanation you think necessary. **No more than two pages in total (one should be enough).**

You may treat this assignment as a realistic scenario where any questions/considerations/comments on the data and the the analysis should be included in your data analysis report (the pdf file).

This assignment aims at testing your ability to code the correct solution as well as the understanding of the data analysis we covered in this module so far. As such, your code must not contain any built-in Python functionality that was not covered in the tutorials or in the primer document.

The deadline is March 6th at 05:00pm. Late submissions will be penalized according to the amount of delay. Failure to follow the instructions above will also cause a penalty to be applied. Plagiarism will not be tolerated.