Name:
Roll Number:

BM5033 Statistical Inference Methods in Bioengineering

Final Exam

Instructions

1. This exam involves statistical analysis of a hypothetical experimental scenario. With the question, a file containing the data is also provided.
2. Read the description of the experiment carefully.
3. State any assumptions you make about the data explicitly.
4. You have to submit the final report in the typed format (PDF). Handwritten reports will not be accepted.
5. Append the R scripts at the end of the report in an appendix.
6. **Deadline: 17:00hrs 19th October 2023.**
7. Please include the list of resources you have taken help from. If you have taken help from any individual that is okay and you should mention that without naming the individual.
8. Please also score each group member out of 10 proportional to their contribution to this exam.
9. Whoever in the group emails the final report should not forget to copy other group members in the email.

In a study, researchers seek to identify the role of two cadherin proteins P_1 and P_2 on the cell shapes and sizes. The experimental capabilities of the research lab permit them to perform microscopy to clearly see the boundaries of cells, and also the expression levels of the two proteins in cells. They can also genetically transform the cells to increase and/or decrease the expression levels of the two proteins in the cells.

Toward this, they performed several experiments (listed below) and quantified the following from each experiment

1. expression levels of P_1 in each cell (in arbitrary units)
2. expression levels of P_2 in each cell (in arbitrary units)
3. major axis length of each cell (approximating the shape of the cell cross-section to be elliptical), (in μm)
4. minor axis length of each cell (in μm)

With this experimental setup, they measured the above-mentioned quantities under the following conditions. Each condition was replicated thrice.

1. Normal setup (WT) without any modifications (Sheet E0X.dat)
2. The expression levels of the protein P_1 were reduced while keeping that of P_2 intact (file E1X.dat).
3. The expression levels of the protein P_2 were reduced while keeping that of P_1 intact (file E2X.dat).
4. The expression levels of the protein P_1 were increased while keeping that of P_2 intact (file E3X.dat).
5. The expression levels of the protein P_2 were increased while keeping that of P_1 intact (file E4X.dat).

The data for each experiment are stored in separate files with file as mentioned above where X stands for the replication number. Each file contains four columns of data (i) level of P_1, (ii) level of P_2, (iii) cell area (in μm^2), (iv) major axis (in μm), and (v) minor axis (in μm) of the cell.

With this information,

1. You are expected to explore the data and identify patterns, if any, and report the data in a graphical form.

2. Perform appropriate statistical tests to infer the exact role of the two proteins, if any, in the regulation of the cell size and shape.

3. Characterize the statistical power for this experimental setup.

4. Can you suggest any changes in the experimental methodology to answer the original question of the "role of the two proteins in cell size and shape regulation"?