BIOE97142 – Computation and Statistical Methods for Research

Study Session 5: Inferential Statistics with Pandas

Lecturers: Dr. Rylie Green, Dr. Christopher Chapman
GTAs: Martina Genta, Hendrick Beck, Mikolaj Kegler, Sergio Mena, Alexey Novikov, Frederik Puffel

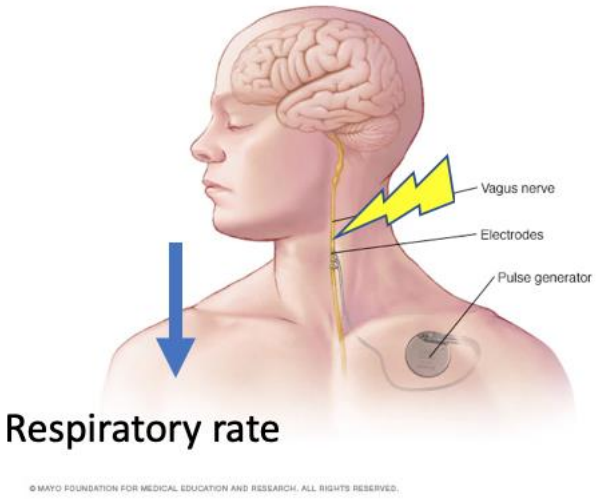

Figure 1 – Schematic of nerve stimulator

Problem:
You have been hard at work making clinically ready nerve electrodes from the newly synthesized materials. You have fabricated nerve electrodes out of each material you have synthesized (grain sizes: 30, 50, 100, 150, 200 µm) and have tested their charge storage capacities (CSC) *in vitro* to assess their overall performance.

Since stimulation of the vagus nerve has been shown to decrease respiratory rate due to eliciting activity in parasympathetic fibers innervating the lungs (Figure 1), you have designed a small pilot clinical study with your electrodes. In this study you will implant 5 groups of 5 patients with stimulators of each electrode type. After implantation you measure the average respiratory rate (RR) for one hour before stimulation, then you measure the average RR after stimulation for one hour in each patient.

The resulting data set is in the "clinical data.xlsx" spreadsheet. Using Python you must analyze this data for any significant differences between the *in vitro* performance of the electrode as well as significant effects seen in the clinical trial. In order to fully analyze this data set you will need to provide plots of the data and conduct appropriate inferential statistics to determine if any significant differences were seen. Specifically, the following aspects should be included:

- A line plot of the CSC means with appropriate y standard deviations
- A bar plot of the clinical data (RR before and after) means with appropriate standard deviations
- Tests of normality for all the data set
- ANOVA and (correct) post-hoc tests for both the CSC and clinical data

Please provide justifications for the statistical tests that you choose based off of the properties of the data (i.e. normality, sample size, etc.). These justifications should be made in your Jupyter Notebook. At the end of this study session please comment on aspects of this study design that you think could be improved. Also, what are some potential challenges that could make this data very hard to interpret and how could you better incorporate those variables into this analysis?