Assessing Reproducibility of 'Deceased Donor Hyperglycemia'

Evan Kim (STA 440 with Dr. Jiang)

2023-12-08

Overview

In this project, "Assessing Reproducibility of 'Deceased Donor Hyperglycemia'," I aimed to replicate crucial elements of a research paper using R. This involved recreating a set of box plots (Figure 1) and a detailed statistical table (Table 2). While tackling Figure 1, I encountered and resolved a few graphical challenges, like matching the original paper's format and data scale. Replicating Table 2 was more complex, involving statistical adjustments and a deep understanding of the original study's methodology, especially in terms of data transformation and analysis. This exercise sharpened my data handling and R programming skills, and underscored the significance of clear reporting in research studies.

Reproducing Figure 1

Referring to `ddh_DD.pdf`, I was able to match variable descriptions from the original paper with variable names from the data to create three box plots to be placed side-by-side with the identical figure caption as shown under Figure 1 in the original paper. While I was able to display axis labels, title, and all data-related features just as shown in the original paper (including logarithmic scale for the y-axis), one minor challenge was that I could not get the y-axis of the third box plot to show 300, despite setting the maximum (`ylim()`) of the y-axis for that specific plot to be 300. Another minor challenge (that led to learning) was that I was not initially able to show the minimum and maximum values with the box plots but rather the 1.5 IQR range, resulting in not showing the same information as the original paper did. I later realized, after reading through `boxplot()` documentation within the R container, that box plots in R would not extend to the minimum and maximum of the data unless I specify that `range=0` as one of the parameters. Overall, reproducing Figure 1 was relatively a smooth process, as creating such EDA-like box plots (by manually creating "filtered" data frames from the original raw data `ddh.csv`) is something I was fortunate to have gotten a lot of practice on throughout the semester via multiple case studies and projects.

Reproducing Table 2

Overall, reproducing Table 2 was a bit more challenging that it was for reproducing Figure 1. The key realization I had was that I had to perform log-2-transformation on `glutwa` to account for the paper having stated "odds ratios corresponding to a relative doubling in TWA". One interesting finding is that, to get the odds-ratio 95% confidence interval bounds identical to those seen in Table 2 from the original paper, when creating the "Adjusted" model that account for variables other than `glutwa`, while the paper states (in the Statistical Analysis section, on page 107) that the model "adjusted" for donor characteristics such as "age, cause of death, calculated Model for End-Stage Liver Disease score, and hemodynamic instability), I actually had to leave out Model for End-Stage Liver Disease score (`r_meld_calc`) from the list of predictor variables in the"Adjusted" model! Additionally, in Table 2 from the original paper, in expressing the odds ratio confidence interval, for instance, "1.46 ([0.91, 2.36)" gives rise to two ambiguities. First, the "1.46" is not necessarily the center of the confidence interval (in fact it is not the center of the interval for any of the figures in Table 2), and it is unclear and unspecified what that figure represents. Second, the "([" part leads me to guess that the table was hard-coded in the paper, which may be the reason behind the ambiguous "1.46" figure.