Torrey Capobianco

Bellevue University

Fall 2019

Women Who Exercise

Hypothesis

Women who have children under the age of 18 exercise less than women who do not have children under the age of 18.

EDA Outcome

Looking at if females who have children exercise less than those that do not have children, through comparisons of CDF, hypothesis testing, and simple regression, the analysis points to the null hypothesis being statistically significant. The null hypothesis will be rejected and will accept the alternative that females with children exercise less than those that do not have children. Comparing the two CDF's of the groups and their exercise durations, women who do not have children had an overall higher probability of exercising longer. There were exercise durations that the two had the same CDF, at 30 minutes and 60 minutes. Through regression, having children alone does not account for the majority of reasons why females exercise less. Having children only accounts for 0.9% of the difference in exercising between the two groups. Other reasons as to why they exercise less were not found in this study. When adding other variables such as age, hours the respondent worked, and the hours their spouse worked, it increased the accountability for exercise duration to 4.1%. However, having children or no children did not affect the duration of exercise when mixed with the other variables. In this case, the null hypothesis would be accepted and the alternative would be rejected. There are many other reasons for women's exercise habits that were not discovered.

Disconnect

Once the analysis had begun, the realization that the exercise data only captured if the respondent participated in the activity of exercise was made apparent. It did not include if the respondent did not partake in that activity throughout their diary day. Thus, the data only represented women with children under 18 who actually exercised, and not those that didn't exercise if they had children. This, I feel would change the outcome if all entries for women who had children and whether they exercised at all or not at all in comparison to women who did not have children. Therefore, this analysis only looked at those that actually participated in exercising.

Variables

There was no correlation between exercise duration and the variables chosen for this analysis. Other variables that would have been helpful but were not available was household income, and weight of the female respondent. These could have perhaps contributed to their exercise habits.

Assumptions

I didn't see any assumptions I thought were incorrect. However, I thought that having children would account for more of the variability in exercise habits.

Challenges

There were some challenges in reading the graphs that I did not understand that required further research for interpretation. One was a comparison of CDFs between two groups. Secondly, understanding which analytical model to choose was a challenge. At first, I thought my data would be represented in an exponential distribution since it related to time, but I realized after the time series analysis section in the course that it was a different type of time and could not be modeled that way.

Resources

DeJesus, John. (2019). What, Why, and How to Read Empirical CDF. *Towards Data Science*. Retrieved from https://towardsdatascience.com/what-why-and-how-to-read-empirical-cdf-123e2b922480

Downey, A. B. (2015). *Think Stats: Exploratory Data Analysis* (2nd ed.). Sebastopol, CA: O'Reilly Media Inc.

Minitab Blog Editor. (2013). How to Interpret Regression Analysis Results: P-values and Coefficients. Retrieved from https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-interpret-regression-analysis-results-p-values-and-coefficients

Pathak, Manish. (2018). Joining DataFrames in Pandas. *DataCamp*. Retrieved from https://www.datacamp.com/community/tutorials/joining-dataframes-pandas

U.S. Bureau of Labor Statistics. (2017). American Time Use Survey: Multi-Year Survey Microdata Files from 2003-2015. Retrieved from https://www.kaggle.com/bls/american-time-use-survey/data

VanderPlas, J. (2016). *Python Data Science Handbook: Essential tools for Working with Data (1st ed.).* O'Reilly Media Inc.