**Exploratory Data Analysis on Body Measurements- Term Project Summary**

Guruprasad Velikadu Krishnamoorthy

College of Science and Technology, Bellevue University

[DSC530-T301: Data Exploration and Analysis](https://cyberactive.bellevue.edu/webapps/blackboard/execute/courseMain?course_id=_512542_1)

Professor. Cary Jim

June 03, 2023

**Exploratory Data Analysis on Body Measurements- Term Project Summary**

**Introduction**

In this project, the exploratory data analysis concepts were applied to the Body measurements dataset, and the summary of the outcome is discussed in this paper.

**Statistical/Hypothetical Questions**

* Are Men healthier compared to Women?
* Are Young Men and Women healthier than their Adult counterparts?
* Are Men taller in general compared to Women?
* Are Men Heavier than Women?
* Do Men have broader shoulders and Chest Girth compared to Women?

**The outcome of EDA**

The outcome of the Project by analyzing the PDF, CDF, Statistical summary, Hypothesis testing, and Correlation and Regression analysis, below are the findings:

* Women are healthier compared to Men overall, though the differences are much more noticeable in adult men compared to young men.
* Most young men and women are healthier than their adult counterparts, though there can be some exceptions. Young women in general are more likely to be healthy.
* Men are more likely to be taller and Heavier than women and the differences are statistically significant.
* Men are likely to have bigger shoulder and chest girth compared to women.

**What was missed during the analysis?**

Aside from body measurements, other external factors such as Income, poverty level, and Education status could have helped to enhance the research to analyze the reasoning behind a person being healthy or not, which was missed.

**Variables that could have helped with the Analysis.**

The dataset did not contain information about Marital status, Family size, poverty, Income status, Homeownership, Education level, and Race to name a few. All these variables could have helped to build a prediction model to predict if a person is likely to be healthy.

**Assumptions made.**

The project was conducted with the assumption that the population only contained Adults and Young Adults. The senior population and Children were excluded from the study. With the world demographics changing drastically, the senior population is expected to increase significantly in future decades. Hence the study may not represent the entire population but only a subset. Also, there was no information on when the dataset was collected. The assumption was made that the data was collected recently which may be incorrect if the data was decades old. The dataset had a limited number of observations and hence representation of the dataset to the entire population was an assumption.

**Challenges faced.**

Choosing variables and creating additional variables during the initial step that can help with the study involved a lot of research and was challenging.

The presence of Outliers in the data led to erroneous results during the development. Hence the outliers were cleansed using Statistical formulas and by applying Domain knowledge.

Understanding some statistical concepts such as estimation, sampling, and Hypothesis testing and applying them to the dataset was a challenge.