Engineering Mathematics II (ED 121)

Implementation and Visualization of Basic Statistics

March 12, 2024

Instruction: Please feel free to use Python or any other tool of your choice to generate the datasets, perform the statistical analysis, create visualizations, and draw conclusions based on the data.

1. Generate a dataset with at least 5000 data points from a random distribution representing exam scores.
2. Randomly sample 10 data points from the dataset and calculate the variance for this sample.
3. Repeat the sampling process for larger sample sizes (e.g., 50, 100, 500, 1000, 5000) and calculate the variance for each sample.
4. Create a histogram for the sample of 10 data points to the entire dataset.
5. Visualize the histograms and variances for each sample size to observe the impact of sample size on variance estimation.
6. Discuss the Central Limit Theorem and show how the distribution of sample means approaches normality with increasing sample sizes.
7. Analyze the plots and draw conclusions on how variance changes with different sample sizes and the implications for statistical inference.
8. Compare and contrast the formulas for calculating population standard deviation and sample standard deviation using the exam dataset.
9. Highlight the specific differences in the formulas, specifically addressing the adjustment of using $n-1$ in the sample standard deviation formula's denominator, and explain the rationale behind this adjustment when working with sample data, emphasizing why $n-1$ is used instead of N in the denominator.
10. Generate a dataset with at least 5000 data points from a random distribution representing height and another 5000 data points representing weights.
11. Create a box plot to visualize the distribution of heights and identify any outliers.
12. Your objective is to calculate the correlation between height and weight.You are tasked with analyzing the relationship between height and weight in a dataset of 5000 individuals.

Click here for the Colab file link
Note:- Make your own functions for mean, variance and covariance. Try making plots different than the ones used in the tutorials. Extra efforts will be appreciated.