

Statistics Mini Project

1. Perform the following statistical analysis on the fifa world cup dataset. The dataset consists of player information which includes height, weight overall rating, potential etc. Data Cleaning and basic EDA process needs to be performed.

- a) Assuming age as population, perform the central limit theorem with sample size of 200. Find the standard error and the mean of sampling distribution. Compare the results with population mean and standard deviation. Explain in non-technical terms why the estimated standard error of a sample mean tends to decrease with an increase in sample size
- b) Is there any evidence that the players overall rating is greater than 80?
- c) Perform the statistical test to identify whether preferred foot has any impact on potential. Also, perform the graphical analysis.
- d) Perform the statistical test to identify the relation between the overall rating and potential. [Assume both are continuous data points].
- e) Does skill move have any effect in overall rating. Perform the graphical analysis and the statistical analysis.

Note:

- 1) Perform all the Statistical analysis with 5%(Significance level).**
- 2) Perform the test of normality and test of variance in the appropriate places.**

2. Perform the following statistical analysis on the mobile network dataset given. The given data set contains information about the internet, calls, messages of various network. Data Cleaning and basic EDA process needs to be performed.

- a) With 95% & 99% Confidence Interval find the range estimate of average duration calls made by all the network. Which of the Intervals are wider? Interpret the results.
- b) Perform statistical analysis to check whether the networks are equally distributed in the dataset.
- c) Plot the distribution of duration with respect to each network in the same graph and Interpret the results from the graph.
- d) Perform statistical tests to identify the relation between item and duration. Detailed explanation of relation between the classes is required.
- e) Perform statistical tests to identify the relation between network and network_type.

Note:

- 1) Perform all the Statistical analysis with 5%(Significance level).**
- 2) Perform the test of normality and test of variance in the appropriate places.**