Excel Project: Random Data Analysis and Linear Regression

Project Description:

The project consisted in generating realistic random data and structuring it into an Excel workbook to consolidate statistical concepts through hands-on manipulation. The initial dataset represented the ages of a sample population of Luggnagg, following a normal distribution based on user-defined parameters such as mean, standard deviation and probability. The main objective was to convert theoretical concepts learned in the course into concrete numerical evidence and analytical outputs.

Key Responsibilities and Tasks:

I created the dataset generation sheet (Parameters), generated 250 normally distributed age values, randomly assigned each individual to one of four groups, and built a sample extraction sheet filtering data based on a chosen group. Then, I created a Statistical Insight sheet summarising descriptive statistics (mean, standard deviation, confidence rate, p-value estimation and confidence interval) including text explanation of results. Finally, I designed a sheet to experiment with correlation (between age, number of cats and age of partner) and a sheet for linear regression, with scatterplot visualisation and interpretation of the regression output.

Outcome:

The final workbook was a multi-layer analytical tool that not only generated synthetic data but also delivered statistical reasoning, inferential metrics and correlation/regression evidence. It provided a full pipeline from random data generation → sample selection → descriptive statistics → correlation analysis → regression modelling. This enabled a complete practical application of the statistical topics studied, validating assumptions and interpreting results in a reproducible and documented spreadsheet.

Technologies and Tools Used:

Microsoft Excel (normal distribution functions, conditional formulas, descriptive statistics, confidence interval calculation, correlation functions, scatter plots and regression), spreadsheet design.