**Correlation Analysis of Start-Up Profits and Business Expenses**

The present report summarizes the results of statistical research conducted on a sample of 50 start-ups to investigate the correlation between profit and expenses related to Research and Development (R&D), administrative, and marketing expenditures.

**Methodology**:

 The first step involved data cleaning to address missing values and inaccuracies, done for one missing data with the cold-deck method (numerical mean). Then we search for outliers within the data set with the scatter plot. As you can see from the image there were not outliers for none of the variables involved into the statistical research.

 

 

 Subsequently, statistical methods were used to assess the correlation between the variables (R&D expenditure, administrative expenses, and marketing expenses) and the responsive variable (profit). In particular, this statistical research addressed the profit variation when one of the other variables changed. The further step consists of finding the best predictor for the profit, meaning the “best” independent variable that explains the profit variation in the best way. The last step consists of finding the best subset to balance the model's goodness of fit and the complexity with the highest adjusted R-squared value and the lowest AIC value.

**Findings:**

1. Correlation with R&D Expenses: The research revealed a strong positive correlation between profit and R&D expenses as shown by the graph. This suggests that an increase in R&D expenses is generally associated with a substantial increase in profit, highlighting the critical importance of investments in research and development for start-up success. Besides that, it can be said that to explain this relationship, it has been used three models: linear regression (red line), second-(green line) and third-order( blue line) polynomial regression. According to the findings, the best model that explains the relationship between the profit and the R&D expenditure is the third-order polynomial regression model with the highest adjusted R-squared (95%) while both other two models adjusted R-squared was around 94%.
2. Correlation with Administrative Expenses: Statistical analysis showed a positive but minimal correlation between profit and administrative expenditures. linear regression (red line) and second- (green line) and third-order (blue line) polynomial regression models were used to explain the relationship between the profit variable (responsive one) and the administration expenditures. The findings show that the second-order polynomial regression model is the most suitable model to explain it with an adjusted R-squared equal to 3%. Despite the minimal correlation, administrative expenses still have the potential to influence start-up profits.



1. Correlation with Marketing Expenses: A positive correlation was found between profit and marketing expenses. This indicates that investments in marketing can positively impact start-up profits, emphasizing the significance of effective marketing strategies. The same methodology was used also in this case but you can see only the linear regression line (red line) and second-order polynomial regression curve( green one) that’s because the third-order polynomial regression adjusted R-squared was lower than the second-order one, 50% and 51% respectively. So it has been decided to show only the first two lines, to see better the correlation between the profit variable and the marketing expenditures graphically.



**Best Predictor:**

In the statistical research, R&D expenditure emerged as the best predictor for profit. This conclusion was based on the highest adjusted R-squared value, which accounts for the model's predictive ability and complexity. Indeed, the R&D expenditures adjusted R-squared is the highest (95%) compared to administration expenditures with the highest adjusted R-squared being 3% and the marketing expenditures being 51%.

**Determining the Best Subset:**

using the AIC criterion, we identified the best model to predict the company’s profit. The model only considers the R&D expenditures variable, with the AIC= 916.9, because more complex models, including other marketing and administration expenditures, reported the AIC values higher. The estimated coefficient for the considered variable is 0.8543, with a standard error of 0.02931. it means that if there is an increase in R&D expenditures by one unit there will be a profit increase of approximately 0.8543 units, holding other variables constant. Finally, by prioritizing R&D expenditures the start-up profits will be maximized.