**ALY6010: Probability Theory and Introductory Statistics**

**Analysis of Adidas US Sales Dataset** Module5ROutput\_ZihanMa\_03.26.2023

Logo

Description automatically generated

**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Boston, MA, 02215.**

**Submitted by:**

Zihan Ma

**Instructor**

 Prof. Tom Breur

**Analysis of Adidas US Sales Dataset**

**I. Introduction:**

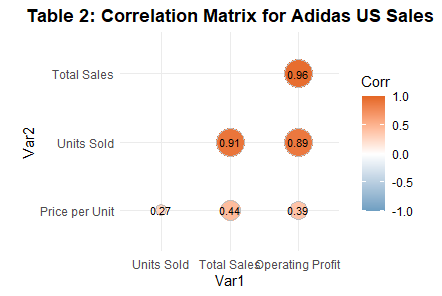
This analysis examines the Adidas US Sales dataset to identify relationships between price, units sold, and total sales.

**II. Data Description**

The data set for this homework is from data.world names Adidas US Sales Datasets, and no missing cells.  
This dataset includes 9,648 observations and 11 variables, and we will only use some values in this data set. Below is the first fine line of variables we will use.  
 Table

Description automatically generated

**III. Correlation Analysis:**

The correlation matrix between the variables of interest is presented in the following correlation plot:   


**Observation**:

The correlation matrix suggests strong positive relationships between Units Sold and Total Sales (0.91) and Total Sales and Operating Profit (0.96), indicating that as Units Sold and Total Sales increase, there is a corresponding increase in Total Sales and Operating Profit. There is also a moderate positive relationship between Price per Unit and Total Sales (0.44), which suggests that as Price per Unit increases, there is a moderate increase in Total Sales. Overall, the correlations suggest that Units Sold and Total Sales are the strongest drivers of Total Sales and Operating Profit.

The observation between Price per Unit and the Var1 group is more valuable for this data. Since Operating Profit is calculated using Total Sales and Units Sold, it is expected to correlate strongly with both variables. In contrast, Price per Unit is an independent variable and is not involved in calculating Operating Profit that much.

**Why No More Than 5?**

Limiting the correlation chart to no more than five variables is essential to improve readability and interpretability. Including too many variables can lead to clutter and make it difficult to identify the critical relationships between the variables.

**IV. Regression Analysis**

We performed a linear regression analysis with Total Sales as the outcome variable and Price per Unit and Units Sold as predictor variables. The regression results are presented in Table 3.1&3.2.

Table

Description automatically generated



**Observation**:

The regression analysis shows that both Prices per Unit and Units Sold have a significant positive relationship with Total Sales. Specifically, for every one-unit increase in Price per Unit, Total Sales increase by $2,003.802. Likewise, for every one-unit increase in Units Sold, Total Sales increase by $568.472. The adjusted R-squared value of 0.874 suggests that the model explains a high proportion of the variance in Total Sales.

But, for the intercept, even when the Price per Unit and Units Sold are zero, there is still an expected Total Sales of -$143,389.300. It indicates that the model overestimates the total sales when the values of price per unit and units sold are zero. This may be due to some missing variables that affect the total sales, which were not included in the model. Alternatively, it could be due to some outliers in the data pulling the model away from the actual expected value. (Frost, 2023)

**How does regression analysis differ from correlation analysis?**Regression analysis involves finding the relationship between a dependent variable and one or more independent variables, while correlation analysis examines the strength of the relationship between two or more variables. In regression analysis, a model is created to predict the dependent variable's value based on the independent variables' values, while correlation analysis measures the degree of association between the variables.

**References:**

Stellabigail. (2022, October 12). Adidas US sales datasets - dataset by Stellabigail. data.world. Retrieved March 26, 2023, from <https://data.world/stellabigail/adidas-us-sales-datasets>

Frost, J. (2023, March 21). *How to interpret the constant (Y intercept) in regression analysis*. Statistics By Jim. Retrieved March 26, 2023, from https://statisticsbyjim.com/regression/interpret-constant-y-intercept-regression/