Kamaal Bartlett

Analysis of Pay Disparity Between Me and Women at a Bank

Introduction

This study investigates the presence of a gender pay gap at a bank using data from the Sleuth3 library in R (case1202). The dataset contains information on 93 employees, including their starting salary (Bsal), salary in 1977 (Sal77), sex (Sex), seniority (Senior), age (Age), education level (Educ), and years of experience (Exper). The primary research question is whether there is a significant difference in starting salaries between men and women. Additionally, the study explores how the inclusion of gender as a variable in a regression model affects the predictability of starting salaries.

Data Source and Definitions

The data set used for this analysis is from the Sleuth3 library (case 1202) in R. It contains information on starting salaries (Bsal), previous year's salary (Sal77), gender (Sex), seniority (Senior), age (Age), education level (Educ), and years of experience (Exper) of 93 employees.

Exploratory Data Analysis

Exploratory data analysis (EDA) was conducted to summarize the main characteristics of the dataset. The overall mean starting salary was approximately 5420, with a minimum of 3900 and a maximum of 8100. A notable difference was observed between genders, with males having a higher mean starting salary of 5957.875 compared to females at 5138.852. This initial observation suggests a potential gender pay gap.

Figure 1: Boxplot of Starting Salaries by Sex

Research Question

The main research question is to determine if there is a significant difference in starting salary between men and women at the bank and to evaluate how the inclusion of gender as a variable affects the predictability of starting salaries.
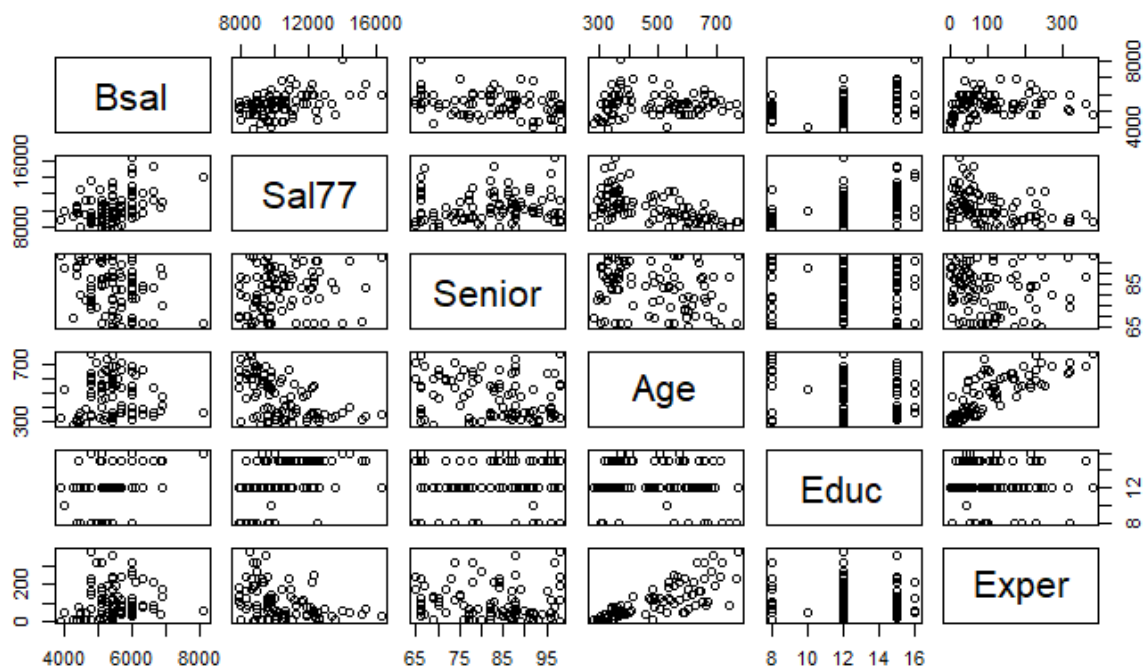

Figure 2: Scatterplot Matrix

Methods

Two-Sample T-Test

A two-sample t-test was conducted to assess whether the difference in mean starting salaries between men and women is statistically significant. This test is appropriate for comparing the means of two independent groups and assumes that the data are approximately normally distributed and that the variances are equal. The t-test indicated a significant difference in starting salaries between men and women (t = -5.83, df = 51.329, p-value = 3.71e-07), with men having higher average starting salaries.

Power Analysis

A power analysis was performed to ensure the sample size was sufficient to detect a significant difference. The effect size was calculated using Cohen's d, and the power test confirmed that the sample size was adequate, with a power value close to 1. This indicates that the study had a high probability of detecting a true effect if it exists.

Regression Analysis

Regression models were built to predict starting salaries, both with and without the gender variable. Forward and backward stepwise selection methods were used to identify the best predictors. Both methods resulted in models that identified Sal77, Senior, Educ, and Exper as significant predictors. The inclusion of the gender variable significantly improved the model's predictive ability, as evidenced by an increase in the R-squared value from 0.467 to 0.5527. The coefficient for the SexMale variable was positive and statistically significant, suggesting that men have a higher starting salary than women, even after controlling for other factors.
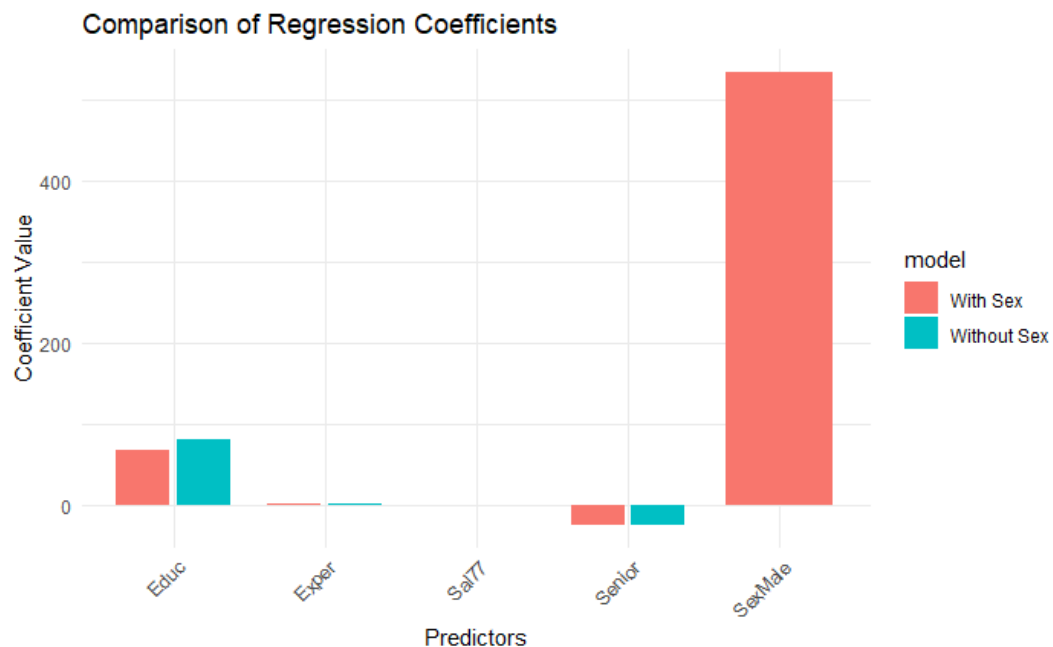

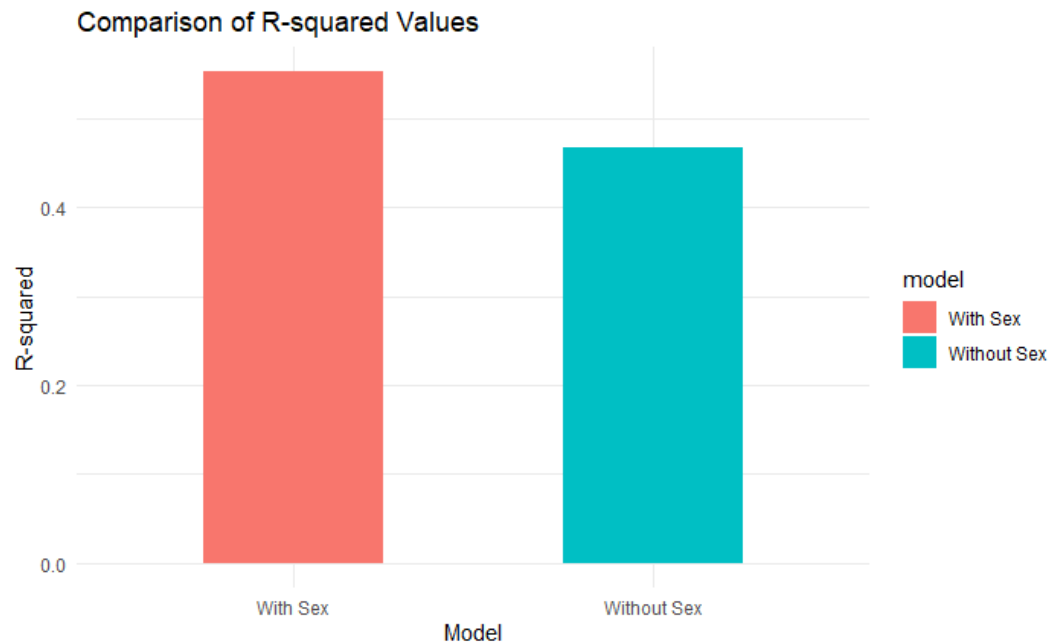Figure 3: Comparison of Regression Coefficients

Figure 4: R-Squared Values Comparison

Exploratory Graphs and Transformations

To address potential non-linearity in the relationships between predictors and the outcome variable, transformations were explored. A log transformation was applied to the response variable (Bsal), improving the relationships' linearity.

Interaction Terms

Two-way interaction terms were explored to examine whether the relationship between two predictors and the response variable depends on the level of a third variable. While interaction terms were included in the analysis, not all of them were found to be statistically significant.

Results

The statistical analysis provides strong evidence of a gender pay gap at the bank, favoring men. The results of the t-test and the regression analysis consistently show that men have higher starting salaries than women, even after controlling for other factors such as education, experience, and seniority. The inclusion of the gender variable significantly improves the model's predictive power, further supporting the existence of a gender pay gap.

Discussion

The findings of this study have important implications for the bank's policies and practices regarding pay equity. The significant gender pay gap suggests that there may be systemic biases or discriminatory practices in the bank's compensation system. The bank should consider conducting further investigations to understand the underlying causes of this disparity and take corrective actions to ensure fair compensation for all employees.

Limitations

This study has some limitations. The data is from a single bank and may not be generalizable to other organizations. Additionally, the dataset is relatively small, which may limit the statistical power of the analysis. Further research with larger and more diverse samples is needed to confirm these findings.

Conclusion

In conclusion, the statistical analysis provides compelling evidence of a gender pay gap at the bank, favoring men. This finding highlights the need for the bank to address pay equity issues and ensure fair compensation practices for all employees.

References

Ramsey, F.L., & Schafer, D.W. (2013). The statistical sleuth: A course in methods of data analysis (3rd ed.). Brooks/Cole.