How Your Salary Impacts Benefits You Get

Ziheng (Tony) Fang

Abstract

This study looks into how an employee's total benefits received is related to the employee's total salary. Data regarding the total compensation of different employees was studied. The objective was the identify any correlation between employees' salary and the benefits that they received. Methods utilized included building a linear regression model as well as utilizing a scatterplot. Results concluded that the amount of benefits an employee receives is strongly correlated to the total salary they receive.

Introduction

The amount of compensation employees obtain is crucial when it comes to workplace dynamics as well as employee retention. How well an employee is compensated as is directly related to how satisfied they are with their job. Compensation is not just about the base salary an employee obtains. Benefits, such as health insurance, retirement, and other benefits are also a major factor when it comes to the compensation amount each employee receives. The total salary an employee receives clearly shows how much an employee's labor is valued, understanding how it relates to the benefits they receive would provide another insight into pay structures.

Understanding the relationship between total salary and total benefits would reveal how corporations allocate resources across employees of different roles. If we can see that benefits and salary are not correlated with each other, we can conclude that corporations give benefits to employees equally, regardless of their level or role in the organization. In contrast, if the amount of benefits an employee obtains is dependent on the total salary they receive and there is a positive correlation, this would point to the fact that organizations not only give more value to employees of certain levels or roles through their salary, but the recognition is reflected in their benefits as well. This paper addresses this gap by providing statistical insight on whether employees who receive a higher salary also receive more benefits.

When deciding on a job offer, many prospective employees focus primarily on the salary as the key factor in evaluating the offer. Most job posts would list a range for the salary, while

completely leaving out details on benefits received. Little emphasis is placed on how the benefits an employee receives are correlated to the salary they receive. Understanding the relation would be very helpful for employees evaluating offers to see how well they are actually being compensated.

To help us look into the question, we pulled a dataset, Employee Compensation, from DataSF. The dataset provides details on different salaries adding up to a total salary and different benefits adding up to total benefits on numerous amounts of employees for the City of San Francisco.

Data

The dataset, Employee Compensation (*Employee Compensation* 2013), from the Open Data Portal on municipal website of San Francisco provides data on the salary and benefits of city employees since the fiscal year of 2013. There is a total of 1.05 million data points in this dataset, where each row reflects a city employee. The data is obtained through the San Francisco Controllers Office where they have a database of all salary and benefits of city employees.

Methods

A linear regression model was fitted based on the dataset using the lm() function in R (R Core Team 2024).

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

In our model, the total salary that each employee, i, received is represented as the independent variable (X), and the dependent variable (Y) was the total benefits each employee received.

Analyzing the model would show us whether or not there is a correlation between the dependent and independent variables. However, we are not able to conclude causation between the two variables. For example, the department a city employee works in could simply pay higher and give more benefits than another. In that sense, the benefits received is not dependent on the total salary received, but is directly related to the department the employee works in. Another example would include that some city workers are unionized, which would increase both their salary and benefits. In that sense, the benefits they receive is not caused by their salary, but their union membership. Furthermore, our dataset only contains a employees of the City of San Francisco. Trends may very likely differ if we look at data in another city or a private organization, a confounding variable.

Results

Within our regression model, β_0 was the intercept, β_1 was the slope, and ϵ_i was the error residual. Y_i was the independent variable representing the amount of benefits received by each city employee i. X_i was the dependent variable representing the total amount of salary employee i received.

The model estimated a slope. b_1 of 0.269 for the salary variable. Our null hypothesis concluded no relationship between the dependent variable Y, total benefits, and independent variable X, total salary. Obtaining p-value of less than $2 * 10^{-16}$, we reject the null hypothesis on a .95 confidence interval and conclude there is a relationship between the two variables. Thus, based on the the rejection of the null hypothesis and the estimated slope we conclude that the amount of benefits an employee receives is positively correlated with the total salary they receive. Our R^2 value of 0.7594, also indicates that total salary explains a lot of the variability in the total compensation received. For every dollar of salary an employee gets, they get about \$0.27 in benefits.

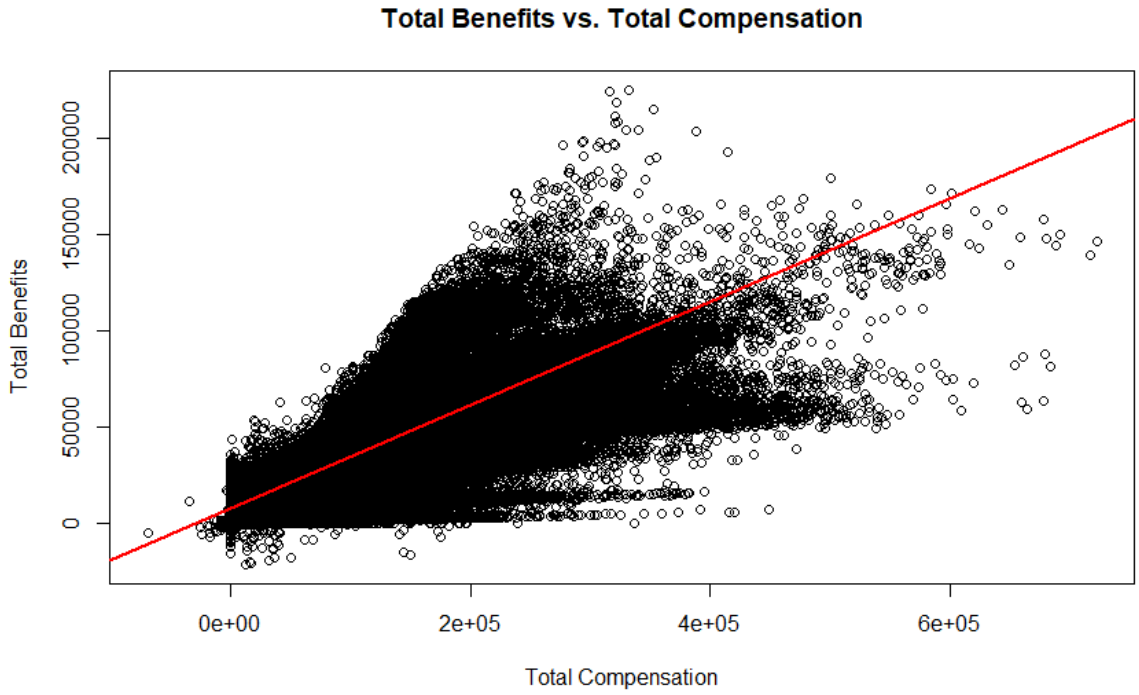

Figure 1: Fig 1: Total Benefits vs. Total Compensation for City of San Francisco Employees

Discussion

We analyzed the total annual salary and total benefit value of 1.05 million employees working for the City of San Francisco. We considered the total salary to be the independent variable and the total benefit value to be our response variable when fitting our regression model.

Analyzing our model, we found an estimated slope of 0.269 for our total annual salary. We considered a null hypothesis where there is no relation between our dependent and independent variables. In our case, our null hypothesis would be concluding that the amount of annual salary an employee gets has no relation to the benefits they receive. Considering a two-sided t-test with a confidence interval of 0.95, we observed a p-value that is significantly lower than 0.05, thus reject the null hypothesis and conclude that our variables are related with a positive correlation. A high R^2, shows that variance within the benefits received is explained by the salary received, especially in a large sample size of 1.05 million city employees.

However, this research has limitations in the sense of causal relationships. Although we can clearly see that total benefits an employee receives is positively correlated with the total annual salary they receive, we cannot say a high salary "causes" a high amount of benefits. As mentioned above, a particular department an employee works for could have significantly more funding. Should an employee work for that particular department, they would have higher annual salary and benefits received. Many city employees are members of unions. Unions would generally negotiate better pay and benefits for their members. Lastly, our data focuses on employees who work for the city of San Francisco. Trends that we see here may not be applicable for employees who work for other cities or non-governmental workers.

To make our findings more conclusive, instead of using a simple linear regression model where we only consider one predictor variable, the total salary received, we could utilize a multiple linear regression model where we have multiple predictor variables. Variables we mentioned above (e.g. union membership and department) could be one-hot encoded and included as additional predictor variables. The data on the estimated slopes of these predictor variables could be used to determine which predictor variable is more highly correlated with our response variable, the total benefits received. In addition, we could look into data on compensation from other cities or private organizations to get a sample that is more representative and is not limited to employees for the city of San Francisco.

References

Employee Compensation. 2013. 1 Dr Carlton B Goodlett Pl # 316, San Francisco, CA 94102: Controller's Office, City of San Francisco. https://data.sfgov.org/City-Management-and-Ethics/Employee-Compensation/88g8-5mnd/about_data.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.