**Analyzing Microsoft’s Quarterly Valuation: Using Correlation and Linear Regression**

Project Phase-III

March 25, 2025

**Introduction.**

Microsoft has consistently set the benchmark for success in software development and digital services. In this study, we delve into the relationship between time and two critical financial metrics, revenue, and profit in the context of Microsoft's performance. Using Correlation and Linear Regression analysis, we aim to uncover whether temporal factors independently drive fluctuations in Microsoft's revenue and profit trends. Furthermore, this study employs visual aids such as scatter plots to enhance the interpretation of the regression results. Through our exploration, we seek to offer insights that can inform strategic decision-making processes and contribute to Microsoft's long-term sustainability in the dynamic tech industry.

**Results And Discussion.**

The table below shows the independent variable being time as years(quarterly) and the dependent variable as revenue and profit simultaneously.

|  |  |
| --- | --- |
| Independent Variable (Time) | Dependent Variable |
| Years (Quarterly) | Revenue (in Billions of Dollars) |
| Years (Quarterly) | Profit (in Billions of Dollars) |

**Methodology**

Using Microsoft Excel data analysis function, regression analysis was performed on revenue and profit differently, scatter plots, and forecasting methods.

**Analysis of Findings**

From our regression analysis, we can find the following,

For Revenue:

* **Linear Relationship**

Using the hypothesis framework, we either reject or fail to reject the null by comparing the significance level F (p-value) and alpha (0.05), where the null hypothesis is that there is no linear relationship between the independent variable (time) and the dependent variable (revenue) and the alternative hypothesis being, there is a linear relationship between the independent variable (time) and the dependent variable (revenue). Based on the results provided, there is a linear relationship between the independent variable (year) and the dependent variable (income). This is supported because the p-value is less than alpha, which means that rejecting the null hypothesis leads us to accept the alternative hypothesis (there is a linear relationship between the two variables).

* **Strength and Direction of the correlations**
* The correlation between X Variable 1 and Y appears to be strong. as indicated by the high value of the correlation coefficient (R) of approximately 0.9178. The correlation coefficient (Multiple R) is approximately 0.918, indicating a very strong positive correlation between the "Years (Quarterly)" variable and the "Profit (Billions Dollar)" variable.
* **Reliability of the estimations provided by these regression models.**

The R-squared value (0.842) indicates that approximately 84.2% of the variability in profits can be explained by the linear relationship with time (Years). The ANOVA results also indicate that the regression model is significant at a 0.005 alpha level.

* **Regarding the population parameters,**

The point estimates for the slope (β1) and y-intercept (β0) are approximately 0.353 and 2006.54, respectively. The 95% confidence intervals for these parameters are provided in the regression output as well:

For the intercept (β0), the 95% confidence interval is (2005.25, 2007.84).

For the slope (β1), the 95% confidence interval is (0.311, 0.395).

These confidence intervals provide a range of values within which we can be 95% confident that the true population parameters lie.

FOR-PROFIT

* **Linear Relationship**

Using the hypothesis framework, we reject the null by comparing the significance F (p-value) and the alpha (0.05). The null hypothesis is that there is no linear relationship between the independent variable (time) and the dependent variable (profit). The alternative hypothesis test is there is a linear relationship between the independent variable (time) and the dependent variable (profit).

Based on the provided result, there is a linear relationship between the independent variable (years) and the dependent variable (profit). This is supported because the p-value is less than the alpha, which means rejecting the null hypothesis leading us to accept the alternative hypothesis. (there is a linear relationship between both variables).

* **Strength and Direction of the correlations**

The correlation coefficient (Multiple R) is 0.909555946, suggesting a strong positive linear relationship between the variables. The R-squared value (0.827) indicates that approximately 82.7% of the variability in profits can be explained by the linear relationship with time (Years).

* **Reliability of the estimations provided by these regression models.**

The R-squared value (0.827) indicates that approximately 82.7% of the variability in profits can be explained by the linear relationship with time (Years). The ANOVA results also indicate that the regression model is significant at a 0.005 alpha level.

* **Regarding the population parameters:**

The point estimate for the intercept (β0) is approximately 2007.63.

The point estimate for the slope (β1) is approximately 0.444.

The 95% confidence intervals for these parameters are:

For the intercept (β0), the 95% confidence interval is 2006.39, 2008.87).

For the slope (β1), the 95% confidence interval is (0.388, 0.499).

These confidence intervals provide a range of values within which we can be 95% confident that the true population parameters lie.

**Using the obtained regression lines, we estimate the value of a data point we have among our datasets.**

**For Revenue**

To estimate the value of a data point using the obtained regression line and calculate the error (residual), we would follow these steps:

Identify the data point: Choose the data point we want to estimate the value for, which is the 5th data point (2011).

Calculate the predicted Y value: using the regression equation,

**Y=b0+b1×X**

Substitute the coefficients you have obtained into this equation:

Y = 2006.5434 + 0.35308138 x X.

For the 5th data point, X = 2011

Y= Y = 2006.5434 + 0.35308138 x 2011.

Y = 2716.59005

The predicted Y value for the 5th data point is approximately 2716.59005.

Calculate the residual (error): The residual is the difference between the actual Y value and the predicted Y value.

For the 5th data point:

**Residual =Actual Y−Predicted Y**

Residual = 2011−2716.59005

Residual = -705.59005

So, the error (residual) for the 5th data point is approximately -705.59005.

**For Profit**

To estimate the value of a data point using the obtained regression line and calculate the error (residual), we would follow these steps:

Identify the data point: Choose the one you want to estimate the value for, which is the 5th data point.

Calculate the predicted Y value: using the regression equation:

**Y=b0+b1×X**

Substitute the coefficients we have obtained into this equation:

Y = 2007.62845 + 0.44364659 x X.

For the 5th data point, X = 2011.

Y = 2007.62845 + 0.44364659 x X.

Y = 2899.80174

Calculate the residual (error): The residual is the difference between the actual Y value and the predicted Y value.

For the 5th data point:

Residual =Actual Y−Predicted Y

Residual = 2011− 2899.80174

Residual = -888.80174

So, the error (residual) for the 5th data point is approximately -888.80174

**Forecasted values of your dependent variables for just one additional time interval**

Our forecasted value for revenue in the first quarter of 2024 is 73.445 billion dollars and for profit is 45.14 billion dollars.

**The Statement “Correlation does not imply causation”**.

In our case where we are analyzing the valuation of Microsoft corporation, it’s important to consider the statement “correlation does not imply causation”. This implies that just because two variables are correlated doesn’t mean that one variable makes the other change. In our analysis, we see both variables are core aspects of the company valuation.

Time may be correlated with revenue and profit such as product launch time or season. However, this demonstration does not automatically imply that time causes changes in Microsoft’s profit or revenue. Yes, time(years) and the dependent variable (revenue & profit) are related variables, but it doesn't necessarily mean one affects the other and other factors such as leadership board, market demand, market trend, economic conditions, management decisions, and market policies might impact them. While our analysis reveals a correlation between time and Microsoft’s valuation, further research is necessary to determine if time directly causes changes in Microsoft’s valuation or if other factors do.

**Conclusion**

From the research and data analysis by exploring the relationships between time and dependent variables (profit, revenue) through regression analysis, scatter plots, and forecasting methods, we sought to uncover potential correlations between time and dependent variables related to Microsoft valuation. Our findings yielded several insights, noting that the correlation doesn’t imply causation, emphasizing the need to look at other factors affecting it aside from time.

In conclusion, our research contributes to understanding Microsoft’s valuation dynamics, noting the complex interplay between time and performance metrics.

**Reference(s)**

Kennaway, R. (2020). When causation does not imply correlation. In *Elsevier eBooks* (pp. 49–72). <https://doi.org/10.1016/b978-0-12-818948-1.00004-6>

M. V. (1, January 30). *12 facts about Microsoft*. Facts. Retrieved February 19, 2024, from [www.facts.com](http://www.facts.com).

Zack Investment Research (2023, May 5). Microsoft Gross Profit and Revenue 2010-2023. Macrotrends. Retrieved March 19, 2024, from <https://www.macrotrends.net/stocks/charts/MSFT/microsoft/gross-profit>.