# The Surname Edge

1
1
2
2
3
3
3
3
4
4
· ·

# The Question

Do companies with names of common surnames have an advantage when it comes to stock returns?

#### **Hypothesis Statement**

- **Null Hypothesis (H0)**: There is no advantage for companies whose names match common U.S. surnames in terms of average stock returns compared to companies whose names do not match common surnames.
- Alternative Hypothesis (H1): Companies whose names match common U.S. surnames have an advantage in average stock returns compared to companies whose names do not match common surnames.

# Approach

This experiment will compare each company's name exactly to the common surnames found within 'Names.csv.' This exact matching technique was decided because the question at hand is customer loyalty according to *personal connection*. The feeling a person would have seeing their true name far outweighs a coincidental occurrence. By splitting the companies with matches into one portfolio and companies without name matches into another portfolio, the average returns can between the two portfolios can be compared. This was delineated for each client by populating a new column (name\_match) as 0 or 1/True or False. If companies with matching names do provide a

significantly higher return (alpha level ( $\alpha$ ) = 0.05) the alternative hypothesis is true; otherwise, the null hypothesis is true.

The earliest record of a name match occurs is 2012-05-31. This means we can trim the mret7023.sasbdat file into a 13-year dataset called 'stock\_data.'

```
> earliest_matches <- stock_data_filtered %>%
+ filter(name_match == 'True') %>%
+ summarise(earliest_date = min(DATE))
>
> print(earliest_matches)
# A tibble: 1 × 1
earliest_date
<date>
1 2012-05-31
```

The strategy to slowly increase the sample size of 5 years stock returns, to 10 years stock return, to the full dataset starting May 31<sup>st</sup>, 2012, can show how the data trends over a longer period time. If the data is consistent over time, it shows the results of this experiment are robust.

### **Research Design**

Two types of testing were used. First, the Welch Two Sample T-Test is used to compare the means of two independent groups to see if there is a statistically significant difference. The Welch t-test is more flexible than the standard t-test because it does not assume groups have the same variance. This makes it good for financial data that can be volatile and varying dispersion. Second, the Linear Regression Analysis is used to understand the relationship between a dependent variable (stock return RET) and an independent variable (name\_match). This can help determine if there is a positive effect of name matching in stock returns.

#### Testing

- <u>Welch t-test:</u> If the p-value from the test is less than the significance level ( $\alpha = 0.05$ ), it means there is enough evidence to reject the null hypothesis and confirm the alternative hypothesis which is that there is a statistical advantage of higher returns for companies whose names match common surnames.
- <u>Linear Regression Analysis:</u> If the coefficient of 'name\_match' is positive, it means there is a positive correlation, but the p-value must still be statistically significant ( $\alpha = \alpha = 0.05$ ). If both of those conditions are met, there is enough evidence to reject the null hypothesis and confirm the alternative hypothesis which is that there is a statistical of higher returns for companies whose names match common surnames.

## Findings

#### Welch Two-Sample T-Test

The t-test was conducted to determine whether there was a significant difference in average returns between companies whose names match common surnames (name\_match = TRUE) and those that do not (name\_match = FALSE).

The p-value obtained from the t-test is 0.221, which is greater than the significance level ( $\alpha$ =0.05). This indicates that there is not enough evidence to reject the null hypothesis. The average monthly return for companies without matching surnames is 0.69%, while the companies with matching surnames is 1.8%. Although the returns are higher, they are not statistically significant, which gestures to the higher returns stemming from a random variation rather than a systematic advantage.

#### Linear Regression Analysis

The linear regression was built to assess the relationship between having a matching surname ('name\_match') and stock returns (RET).

The coefficient for 'name\_match' is 0.0117, indicating a slight positive relationship between having a matching surname and stock returns. However, the p-value for this coefficient is 0.397, which is much greater than the significance level ( $\alpha$ =0.05). This means the effect of 'name\_match' on stock returns is not statistically significant. The multiple R-Squared value is 6.65e-07, which means the comparison model explains almost none of the variance in stock returns. The adjusted R-squared value is negative, further indicating that adding 'name\_match' as a predictor does not improve the model's prediction. The F-statistic for the model is 0.7188 with a p-value of 0.3965, confirming the model is overall not significant.



### Trends Throughout Sample Sizes

Looking at the graphs, the trends continue in the same direction as the sample size increases. This confirms the findings are robust and consistent across differing sample sizes.

#### Summary of Findings

Based on none of the analyses concluding a significant relationship, <u>we fail to reject the null</u> <u>hypothesis</u>. There is no evidence to suggest that companies whose names match common U.S. surnames have an advantage in average stock returns compared to companies without matching surnames.

### Recommendation

Based on the thorough analysis conducted, there is no statistically significant evidence to support the hypothesis that companies with names matching common U.S. surnames have an advantage in average stock returns. The t-tests consistently yielded p-values greater than the significance level, and regression analysis showed no meaningful positive relationship between matching surnames and returns.

Although there was a slight trend towards higher returns for companies with matching names, these differences were not statistically significant, suggesting they are likely due to random variation rather than a systematic advantage. The robustness of these findings was confirmed through increasing sample sizes, which consistently showed no evidence of an advantage.

Therefore, it is recommended <u>not</u> to pursue an investment strategy based solely on the presence of a common surname in a company's name. Other factors, such as industry performance, company fundamentals, and broader market conditions, are likely to have a far greater impact on stock returns than name familiarity alone.