**Assignment 2: Statistical Inference & Hypothesis Testing**

I am using a dataset called “U.S. Housing Market Factors”. It contains annual macroeconomic factors which can affect USA house prices. These factors are:

1. House Price Index (HPI): “House price change according to the index base period set.”
2. Stock Price Index: “Stock price change according to the index base period set.”
3. Consumer Price Index: “The Consumer Price Index measures the overall change in consumer prices based on a representative basket of goods and services over time.”
4. Population: “Population of the USA.”
5. Unemployment Rate: Unemployment rate of the USA in units of percentage
6. Real Gross Domestic Product (GDP): “GDP with adjusted inflation.”
7. Mortgage Rate: “Interest charged on mortgages” in units of percentage.
8. Real Disposable Income: “Money left from salary after all the taxes are paid.”

I hypothesize increase in unemployment rates correlates with a lower HPI score in the United States. In my process, I uploaded the dataset into RStudio. Since I am doing a correlation analysis, I knew t-test, z-scores, and chi-square testing would not be applicable to my hypothesis. Following the STHDA method to compute the correlation between HPI and Unemployment rate in R, I used the Pearson correlation method, as it is the most common used method.

After visualizing the graph as a scatterplot, I was able to see the correlation between HPI and Unemployment rates. The results show as Unemployment rates increased, HPI scores also decreased. The 95percentile confidence intervals are (rounded) [-0.571, -0.058], which indicates the average of HPI scores are lower than the Unemployment rates. Since multiple dots on the graph are far from the regression line, it indicates the correlation between the two variables is weak. I will need to agree that the correlation between the two variables is weak since the dots are too scattered apart to show a proper line of regression between the variables. Since the dots are too scattered, it is difficult to identify which is an outlier or not.

**References**

Memon, F. (2022). *U.S. Housing Market Factors*. Kaggle.com. <https://www.kaggle.com/datasets/faryarmemon/usa-housing-market-factors/data>

‌*Correlation Test Between Two Variables in R - Easy Guides - Wiki - STHDA*. (2024). Sthda.com. <https://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r>

‌