<u>**DSCI 5180: INTRODUCTION TO BUSINESS DECISION PROCESS**</u>

Final Project Report

Submitted by:

Name: Gokul

UNT ID: 11690939

This dataset is taken from Hawkes Learning –

https://www.hawkeslearning.com/Statistics/dbs/datasets.html

The dataset contains the comparison of stock prices and returns for Amazon, Starbucks, Coca-Cola, and the S&P 500 index. It includes information such as the date, closing price, price change, and return for each stock and the index. The data spans from January 3, 2000, to an unspecified date. The price change represents the difference between the closing prices on consecutive days, while the return indicates the percentage change in price. This data can be used to assess the performance and volatility of these stocks over time.

Module 1-

The price of Amazon stock is distributed normally. The Mean close of Amazon stock is $210.56 and the variance of the close of Amazon stock is $68135.13. Find the probability that the close of stock is more than $280.

Solution:

Given in the question,

Mean, μ = $210.56

Variance, σ^2 = $68135.13

- ⇨ Standard deviation = $\sqrt{\sigma^2}$
- ⇨ $\sqrt{68135.13} = 261.027$

Symbolic Representation of the Problem: $P(x > 280)$

Conversion to standard normal using, $z = (x- \mu / \sigma)$

- ⇨ $z = (280-210.56)/261.027$
- ⇨ 0.2660

$P(z > 0.2660) = 1 - P(z < 0.2660)$

- ⇨ $1 – 0.6049$
- ⇨ 0.3951

The probability that the closing price of a stock is more than 280 is 39.51%

Module 2 –

Let us take 4526 stock of S&P500 close samples. The average of close stock is 123.24. Using the sample mean, a confidence interval can be calculated within which the stock is relatively confident that the actual close mean is located. Suppose that the population standard deviation is known from Stock to be 49.10. Calculate the 95% confidence interval.

Solution:

Given in the question,

Sample size, $n = 4526$

Sample mean, $\bar{x} = 123.24$

Population Standard Deviation, $\sigma = 49.10$

Alpha, $\alpha = 1 – 0.95 = 0.05$

z-value, $z_{\alpha/2} = -1.96$

Confidence interval = $\bar{x} \pm z_{\alpha/2} \cdot (\sigma / \sqrt{n})$

- ⇨ $123.24 \pm (-1.96)(49.10/\sqrt{4526})$

Upon solving we get the confidence interval as (121.81,124.67)

Module 3-

As per a stock price analyzer, the mean close price of Starbucks stock is $25. You believe that the provided stats are wrong and claim that the average closing price is below the mentioned. In a random sample of 1112 stocks, you calculate the sample mean as $23.227 with a standard deviation of $8.475. Test your claim at a 0.03 level of significance assuming the distribution to be normal.

Solution:

Given in the question,

Sample size, $n = 1112$

Sample mean, $\bar{x} = 23.227$

Sample standard deviation, $s = 8.475$

Alpha, $\alpha = 0.03$

Degrees of freedom = $n - 1 = 1111$

Null Hypothesis, $H_0 : \mu = 25$

Alternate Hypothesis, $H_a : \mu < 25$

In this question, the population standard deviation is unknown hence we need to T-test. It is also to be noted that the test is left (one) tailed hypothesis test.

Critical value, $t_\alpha = -1.88$

T-test statistics, $Ts = (\bar{x} - \mu) / (s/\sqrt{n})$

- ⇨ $(23.227 - 25)/(8.475/\sqrt{1112})$
- ⇨ -6.684

We can observe that the Ts < critical value. As the test is left-tailed in this scenario, we reject the null hypothesis.

Hence, we can conclude that there is sufficient evidence at 0.03 level of significance to prove the claim.

Module 4 –

Professor claims that the average closing price of Coca-Cola stock is less than the S&P 500. You wonder if this is true, so you decide to compare the average closing price of the two companies. For a random sample of 1526 stocks of Coca-Cola, the mean is $15.58 with a standard deviation of $1.53. For 1243 randomly selected stocks of S&P 500, the mean is $83.97 with a standard deviation of $12.25. Test the professor's claim at 0.10 level of significance assuming the variances to be equal. Let Coca-Cola be Population 1 and S&P 500 be Population 2.

Solution:

Given in the question,

Coco-Cola

Sample mean, $\bar{x}_1 = 15.58$

Sample size, $n_1 = 1526$

Sample standard deviation, $s_1 = 1.53$

S&P 500

Sample mean, $\bar{x}_2 = 83.97$

Sample size, $n_2 = 1243$

Sample standard deviation, $s_2 = 12.25$

Alpha, $\alpha = 0.10$

Null Hypothesis, $H_0 : \mu_1 - \mu_2 = 0$

Alternate Hypothesis, $H_a : \mu_1 - \mu_2 < 0$

Here, the test is a left(one) tailed test and as σ is unknown, we need to do a T-test. It is mentioned that the population variances are equal.

Hence,
Degrees of Freedom, $df = n_1 + n_2 - 2$
$= 2767$

$$S_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

$= 68.65$

$$\text{T-Test statistics, } t_s = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 (1/ n_1 + 1/ n_2)}}$$

$= -216.04$

Critical value, $t_\alpha = -1.28$

We can observe that the Ts < critical value. As the test is left-tailed in this scenario, we reject the null hypothesis.

Hence, we can conclude that there is sufficient evidence at 0.03 level of significance to prove the professor's claim.

Module 5 –

Analyze the dataset and predict the Return as a function of Close and Price Change for Amazon using regression analysis. Also, find the solution for the following questions.

1. Identify the multiple regression equation if a significant linear relationship exists between the independent and dependent variables at the 0.05 level of significance.
2. What is the coefficient of determination for this model, R2? Round your answer to four decimal places.
3. What is the adjusted coefficient of determination for this model, Ra2? Round your answer to four decimal places.
4. Which statistic is most appropriate for the above scenario to determine the usefulness of the regression model and why?
5. What would be the value of Return if the values of Close and Price change are 187.54 and 1.38 respectively as per the regression equation?

Solution:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.499075115							
R Square	0.249075971							
Adjusted R Square	0.24874407							
Standard Error	2.949869478							
Observations	4528							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	13060.50631	6530.253157	750.4545883	3.4352E-282			
Residual	4525	39375.32796	8.701729937					
Total	4527	52435.83428						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.120432587	0.056332575	2.137885341	0.032579475	0.009993228	0.230871946	0.009993228	0.230871946
close	-0.000376651	0.00016849	-2.235457581	0.025435914	-0.000706973	-4.63294E-05	-0.000706973	-4.63294E-05
Price Change	0.303819405	0.007844001	38.73270833	1.0559E-283	0.288441332	0.319197478	0.288441332	0.319197478

1. We can observe that the p-value < α which is 0.05. Hence, we can conclude that the relationship between the dependent and the independent variables is statistically significant.

   We can also observe that each variable is significant as their corresponding p-value is also less than alpha (0.05). Therefore, the multiple linear regression equation is as follows:

   $\hat{y} = 0.1204 + (-0.0004)\text{Close} + (0.3038)\text{Price Change}$

2. $R2 = 1- (SSE/TSS)$
   $R2 = 1 – (39375.33/52435.83)$
   $= 0.2490$

3. $Ra2 = 1 – ((n-1/n-k-1)(SSE/TSS))$
   $= 0.2487$

4. Here, the adjusted R2 value is the appropriate measure scenario to determine the usefulness of the regression model.

   This is because two independent variables are making the model multiple linear regression. For these models, Ra2 is the appropriate measure as the computation of Ra2 values considers degrees of freedom.

5. Multiple linear regression equation is:
   $\hat{y} = 0.1204 + (-0.0004)Close + (0.3038)Price\ Change$

   As per question,
   Close = 187.54
   Price Change = 1.38
   Then,

   Return ($\hat{y}$) = -6.9584