**BRIAN NYAKWAMA**

**STATA ASSIGNMENT**

**HYPOTHESIS TESTING**

In this assignment, we make use of one of the example datasets given by STATA called auto.dta. I perform some hypothesis testing on the data so as to answer the below questions regarding the variables from the data. There are 74 observations and 12 variables. The variables include;

* Make – Make and model
* price – price
* mpg – mileage (mpg)
* rep78 – repair record 1978
* headroom – headroom (in.)
* trunk – trunk space (cu. Ft.)
* weight – weight (lbs.)
* length – length (in.)
* turn – turn circle (ft.)
* displacement – displacement (cu. In.)
* gear\_ratio – gear ratio
* foreign – Car type

1. **QUESTION 1**

Is Albert’s claim that the average price of a car is 7000US true?

**Statistical Test** : In this case, one sample t-test is applied. The variable **price** is continuous in nature and the test is of the mean of one continuous variable against an exact value.

**Hypothesis under investigation:**

Null hypothesis : The mean price of a car is equal to 7000USD.

H0 : μ = 7000

Alternative hypothesis : The mean price of a car is not equal to 7000USD

HA : μ ≠ 7000

**Results:**

The valid responses were 100% with no missing observations. The average price of a car is 6165.257 with a standard deviation of 2949.496. At 95% confidence level, we reject the null hypothesis implying that the average price of a car is not equal to 7000USD. This is because the p-value [0.0174] is less than the significance level 0.05.

1. **QUESTION 2**

If there is a difference in the prices between foreign and domestic cars?

**Statistical Test :** the two independent sample t-test is applied on the data. This is because we want to examine two groups of a categorical variable on one continuous variable.

**Hypothesis under investigation:**

Null hypothesis : There is no statistically significant difference in the price on average between domestic and foreign car types.

Diff : μ domestic - μ foreign   
H0 : diff = 0

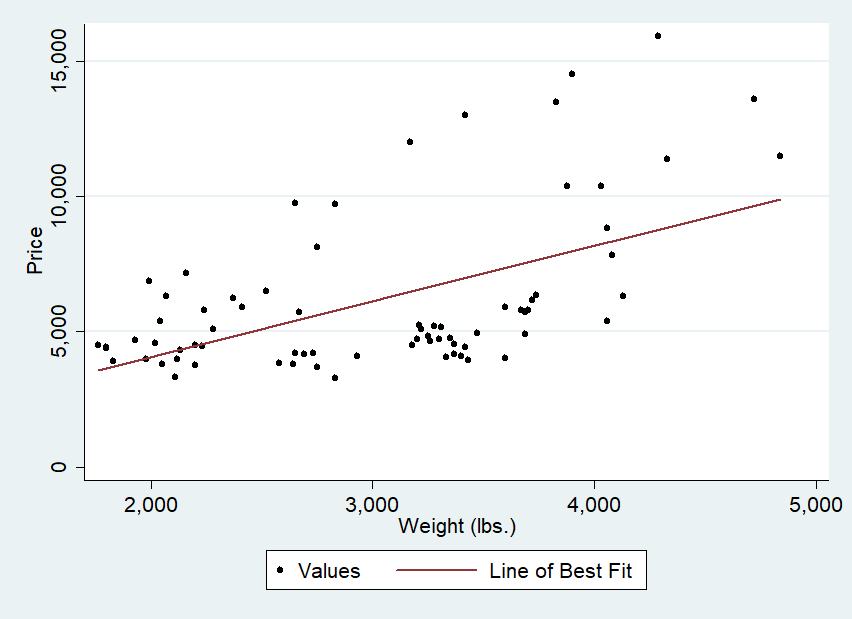
Alternative hypothesis : There is a statistically significant difference in the average price between the domestic and foreign car types.

HA : diff≠ 0

**Results:**

Valid responses are 100% (74), with domestic having 70.27% (52) and foreign 29.73% (22). The mean price of domestic car types is 6072.423 while that of foreign car types is 6384.682.

At 95% confidence level, we fail to reject the null hypothesis implying that there is no statistically significant difference between the mean prices between domestic and foreign car types. This is because the p-value [0.3802] is greater than the significance level 0.05.

1. **QUESTION 3**
2. Plot a scatter plot for the price (Y-axis) and weight (X-axis).

From the fitted values, we see that the variable price is directly related to weight.

1. Correlation analysis on price and weight.

Is there a linear relationship between price and weight variables?

**Statistical test :** Price and weight are continuous variables. We perform a correlation test between the two variables at 95% confidence level.

**Hypothesis under investigation:**

Null hypothesis : The linear relationship between price and weight is not statistically significant.

Alternative hypothesis : the linear relationship between the two variables is statistically significant.

**Results :**

Valid responses are 100% with no missing. The mean for price is 6165.257 while price for weight is 3019.459. At 95% confidence level, we reject the null hypothesis implying that there is a linear relationship between price and weight (lbs.). There is a correlation coefficient of 0.5386 which illustrates a moderate and positive relationship between the two variables.

1. **QUESTION 4**

Which factors are crucial in purchasing a car?

**Statistical test :**  We apply regression analysis on the continuous variables on price.

**Hypothesis under investigation :**

Null hypothesis : The model with the independent variables [mpg, rep78, weight, length, displacement] is not statistically significant in predicting the price of a car.

Alternative hypothesis: The independent variables are statistically significant predictors of price of a car.

**Results:**

The valid responses are 69 (93.24%) with 5 missing (6.77%). The p-value [0.000] is less than the significance level 0.05, and so we reject the null hypothesis implying the model is statistically significant in predicting the price of a car at 95% significance level. The R Squared (R2) is 0.4505, meaning that 45.05% of the variation in price is explained by mileage, repair record, weight, length and displacement, while the 54.95% is explained by other factors not in the model.

Mileage(mpg) and displacement(cu. In.) are not statistical significant predictors of price, as their p-values [0.190] and [0.334] respectively are greater than the 0.05 significance level. The other variables are statistical significant predictors as their p-values are less than 0.05 sf;

Repair record – p-value [0.03]  
Weight – p-value [0.014]  
Length – p-value [0.007]  
Constant – p-value [0.039]

The regression equation is as follows;