Assignment #1 KEY

1. Given this information

$n = 30$	$\sum_{i=1}^{n} X_i = 366$	$\sum_{i=1}^{n} Y_i = 631$	$\overline{X} = 12.20$	$\overline{Y} = 21.03$
$\sum_{i=1}^{n} (X_i)^2 = 5,564$	$\sum_{i=1}^{n} X_i Y_i = 7,524$	$\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = 1098.8$	$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = 882.97$	
$\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) = -174.20$		$\sum_{i=1}^{n} \hat{u}_i^2 = 873.14$		

Answer the following questions. Show your work.

a) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID\left(0, \sigma^2\right)$ (normally, identically and independently distributed), find the estimators of β_1 and β_2 with OLS method and explain the meaning of the model.

> $\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{-174.20}{1,098.8} = \mathbf{-0.1585}$

> $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 21.03 - (-0.1585)12.20 = \mathbf{22.9637}$

Then, our SRF is $\hat{Y}_i = 19.0963 + 0.1585 X_i$ meaning that the intercept is 19.0963 and the slope of this function is 0.1585.

b) Find r^2 and explain its meaning.

> $r^2 = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^{\,2}}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{873.14}{882.97} = 1 - 0.9889 = \mathbf{0.0111}$

The r^2 is a measure of 'goodness of fit' or the measure of the proportion of the total variation in Y explained by the regression model. In this case, 98.89% of Y is explained by the regression model.

c) If $X_i = 5$, estimate the value of $\hat{Y}_i$ and explain its meaning.

From the SRF, $\hat{Y}_i = 22.9637 - 0.1585 X_i$, replacing X_i with 5, we get,

> $\hat{Y}_i = 22.9637 - 0.1585(5) = \mathbf{22.1712}$

which means that when $X_i = 5$, the average of $\hat{Y}_i$ will be 22.1712.

d) Find the estimators of $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$

> $\text{var}(u_i) = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^{\,2}}{n-k} = \frac{873.14}{30-2} = \mathbf{31.1836}$

> $\text{var}(\hat{\beta}_1) = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (X_i)^2}{n\sum_{i=1}^{n}(X_i-\bar{X})^2}\hat{\sigma}^2 = \frac{5{,}564}{(30)1{,}098.8} 31.1836 = \mathbf{5.2635}$

> $\text{var}(\hat{\beta}_2) = \hat{\sigma}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{31.1836}{1{,}098.8} = \mathbf{0.0284}$

e) Test the hypothesis whether coefficients are different from zero at 0.05 level of significance.

First of all, we should find $\text{se}(\hat{\beta}_1)$ and $\text{se}(\hat{\beta}_2)$ by

> $\text{se}(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = 2.2942$

> $\text{se}(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = 0.1685$

Testing $\boldsymbol{\beta_1}$: set up the null and alternative hypothesis as (Note: the degrees of freedom is n-k or 30 – 2 = 28)

> $\text{H}_0: \beta_1 = 0$ and $\text{H}_a: \beta_1 \neq 0$

> Compute $t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{22.9637-0}{2.2942} = 10.0095$

> Find the critical values when $\alpha = 0.05$, then $t_{\frac{\alpha}{2}} = \pm 2.048$.

> Concluding the test: t_{cal} is beyond the critical value, therefore, we can **reject the null hypothesis**. In other words, we can make sure that β_1 is not zero 95 out of 100 times.

————————————————————

Testing $\boldsymbol{\beta_2}$: set up the null and alternative hypothesis as

> $\text{H}_0: \beta_2 = 0$ and $\text{H}_a: \beta_2 \neq 0$

> Compute $t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{-0.1585-0}{0.1685} = -0.9407$

> Find the critical values when $\alpha = 0.05$, then $t_{\frac{\alpha}{2}} = \pm 2.048$.

> Concluding the test: t_{cal} is within the critical value, therefore, we **cannot reject the null hypothesis**. In other words, we cannot make sure that β_2 is not zero 95 out of 100 times.

f) Test the hypothesis whether coefficients are less than zero at 0.01 level of significance.

Testing $\boldsymbol{\beta_1}$: set up the null and alternative hypothesis as (Note: the degrees of freedom is n-k or 30 – 2 = 28)

> $\mathrm{H_0}$: $\beta_1 \leq 0$ and $\mathrm{H_a}$: $\beta_1 > 0$

> Compute $t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{22.9637 - 0}{2.2942} = 10.0095$

> Find the critical values when $\alpha = 0.01$, then $t_\alpha = 2.467$.

> Concluding the test: t_{cal} is beyond the critical value, therefore, we **can reject the null hypothesis**. In other words, we can make sure that β_1 is more than or equal 0, 95 out of 100 times.

____

Testing $\boldsymbol{\beta_2}$: set up the null and alternative hypothesis as

> $\mathrm{H_0}$: $\beta_2 \leq 0$ and $\mathrm{H_a}$: $\beta_2 > 0$

> Compute $t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{-0.1585 - 0}{0.1685} = -0.9407$

> Find the critical values when $\alpha = 0.01$, then $t_\alpha = 2.467$.

> Concluding the test: t_{cal} is within the critical value, therefore, we **cannot reject the null hypothesis.** In other words, we cannot make sure that β_2 is more than or equal 0, 95 out of 100 times.

2. Given that Y is market price of a car (USD) while X is how long a car aged (years), results of the regression are as follows.

$$\hat{Y}_i = 7{,}836 - 502.4X_i$$
$$(52) \quad (411.8)$$

Given that u_i is normally, identically and independently distributed with zero mean and σ^2 variance, total number of observations is 11,

$\bar{X} = 7.45,$

$\hat{\sigma}^2 = 212{,}877,$

$\sum(X_i - \bar{X})^2 = 78.73,$

Answer the following questions. Show your work.

a) Does the sign of $\hat{\beta}_2$ make economic sense? Provide your explanation.

It does. The negative sign indicates that as a car ages, market price drops due to its depreciation.

b) If you are a car expert and someone asks you to estimate how much his car will be **averagely** priced at when his car is 5 years old, how much is the market price range that you would estimate that you can make sure that for 95% of the time, market price will be within the specific range?

For this question, we rely on the mean prediction. The X_0 of is 5, plugging this value into the SRF, we get,

> $\hat{Y}_0 = 7{,}836 - 502.4(5) = 5{,}324$; Then, find the $\mathrm{var}(\hat{Y}_0)$

> $\mathrm{var}(\hat{Y}_0) = \hat{\sigma}^2\left[\frac{1}{n} + \frac{(X_0-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right] = 212{,}877\left[\frac{1}{11} + \frac{(5-7.45)^2}{78.73}\right] = 35{,}582.5345$; Then, find the $\mathrm{se}(\hat{Y}_0)$

> $\mathrm{se}(\hat{Y}_0) = \sqrt{\mathrm{var}(\hat{Y}_0)} = \sqrt{35{,}582.5345} = 188.6333$

Construct a 95% confidence interval around the mean 5,324,

> $\mathrm{P}\left[\left(\hat{Y}_0 - t_{\frac{\alpha}{2}} \cdot \mathrm{se}(\hat{Y}_0)\right) \le Y_0 \le \left(\hat{Y}_0 + t_{\frac{\alpha}{2}} \cdot \mathrm{se}(\hat{Y}_0)\right)\right] = 0.95$; $t_{\frac{\alpha}{2}} = 2.262$ when the degrees of freedom is n-k or 11-2 = 9.

> $\mathrm{P}[(5{,}324 - 2.262 \cdot 188.6333) \le Y_0 \le (5{,}324 + 2.262 \cdot 188.6333)] = 0.95$

> $\mathrm{P}[4{,}897.3115 \le Y_0 \le 5{,}750.6885] = 0.95$

Therefore, we can estimate that when a car is 5 years old, we can make sure that, 95 out of 100 times, the market price will be between 4,897.3115 and 5,750.6885.

c) If you multiply all the X with 10, report the new SRF with the standard error resulted from the multiplication.

If all the X are multiplied with 10, scaling the X without scaling the Y will change the meaning of this model.

For example, if a car is 2 years old, market price is 7,000 USD. If we scale this X up to 10 times, it means that the car will be 20 years old and market price is the same at 7,000 USD. In other words, the depreciation is 10 times less compared to the original data.

As a result, the coefficient $\hat{\beta}_2$ and its standard error must be divided by 10. Hence, the new SRF will be

> $\boldsymbol{\hat{Y}_i = 7,836 - 50.24X_i}$

> **se** $\quad \mathbf{(52)} \qquad \mathbf{(41.18)}$

d) Calculate the elasticity of market price when a car is 10 years old

From the SRF

> $\hat{Y}_i = 7{,}836 - 502.4X_i$

Find the derivative of this function, we get

> $\frac{d\hat{Y}_i}{dX_i} = -502.4$; We know that the elasticity is $\frac{d\hat{Y}_i}{dX_i} \cdot \frac{X_i}{\hat{Y}_i}$.

We need to find $\hat{Y}_i$ when $X_i = 10$ which is $\hat{Y}_i = 7{,}836 - 502.4(10) = 2{,}812$.

Now we can plug all the values into $\frac{d\hat{Y}_i}{dX_i} \cdot \frac{X_i}{\hat{Y}_i} = -502.4\left(\frac{10}{2{,}812}\right) = \mathbf{-1.7866}$.

In other words, we can say that when a car is 10 years old, a decrease in car price of 1.7866 percent is a response for an increase in a year of car from this point.