ASSIGNMENT 1

6304641878
Ximan Yang

Question 1. (15 points) Given this information

$$n = 18 \qquad \sum_{i=1}^{n} X_i = 388.00 \qquad \sum_{i=1}^{n} Y_i = 50.90$$

$$\sum_{i=1}^{n} (X_i)^2 = 9{,}620.00 \qquad \sum_{i=1}^{n} X_i Y_i = 1{,}254.90$$

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 211.00 \qquad \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 2.5844$$

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = 20.58 \qquad \sum_{i=1}^{n} \hat{u}_i^2 = 0.5781$$

Use the above sample information to answer all the following questions. Show explicitly all formulas and calculations.

a) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i,\ u_i \sim NIID(0, \sigma^2)$, **find the estimators** of β_1 and β_2 with OLS method. Interpret the intercept and slope coefficients.

$$\hat{\beta}_2 = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma (X_i - \bar{X})^2}$$

$$\hat{\beta}_2 = \frac{20.58}{211}$$

$$\hat{\beta}_2 = 0.0975$$

(if x_i change by 1 unit, y_i will change 0.0975 in same direction)

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\hat{\beta}_1 = \frac{50.90}{18} - (0.0975)\left[\frac{388}{18}\right]$$

$$\hat{\beta}_1 = 2.8278 - 2.1017$$

$$\hat{\beta}_1 = 0.7261$$

(if x is 0, y equal to 0.7261 unit.)

$$\hat{Y}_i = 0.7261 + 0.0975 X_i$$

b) Compute the value of R^2 and explain its meaning.

$$r^2 = \frac{ESS}{TSS} = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2}$$

$$r^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2}$$

$$r^2 = 1 - 0.2237$$

$$r^2 = 0.7763$$

The independent variable (x) in model can explain dependent variable (y) by 77.63 %.

The propotion or percentage of the total variation in Y explained by regression model is 77.63 %.

c) If $X_i = 30$, estimate the value of $\hat{Y}_i$ and explain its meaning.

$$\hat{y}_i = 0.7261 + 0.0975\, X_i$$

if $X_i = 30$ $\rightsquigarrow$ $\hat{y}_i = 0.7261 + 0.0975\,(30)$

$\hat{y}_i = 3.6511$

if X_i equal to 30 unit, the prediction of Y_i ($\hat{y}_i$) equal to 3.6511 unit.

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

$$\text{var}(U_i) = \frac{\Sigma \hat{u}_i}{n-2} = \frac{RSS}{n-2} = \frac{0.5781}{18-2} = 0.0361$$

$\hat{\sigma}^2$

$$\text{var}(\hat{\beta}_1) = \frac{\Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2} \cdot \hat{\sigma}^2 = \frac{9620\,(0.0361)}{18\,(211)} = 0.0915$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\Sigma(X_i - \bar{X})^2} = \frac{0.0361}{211} = 0.00017$$

e) What are the 90-percent confident intervals for β_2 ? Interpret the meaning.

$$\hat{\beta}_i \pm t_{crit} \cdot se(\hat{\beta}_i)$$

$$\hat{\beta}_2 \pm t_{crit} \cdot se(\hat{\beta}_2)$$

$$\hat{\beta}_2 = 0.0975$$

$$se(\hat{\beta}_2) = \sqrt{var(\hat{\beta}_2)} = \sqrt{0.00017} = 0.0130$$

$$t_{crit} = t_{\frac{\alpha}{2}, n-2} \cdot t_{\frac{0.1}{2}, 18-2} = t_{0.05, 16} = 1.746$$

sub. $0.0975 \pm (1.746)(0.013)$

0.0975 ± 0.0227

$$0.0747 \leq \beta_2 \leq 0.1203$$

β_2 (true value / population term) have value between 0.0747 – 0.1203

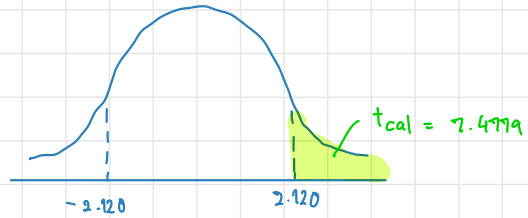
f) Test the hypothesis whether the slope coefficients are different from zero at 0.05 level of significance.

H_0 : $\beta_2 = 0$ [slope coefficients is zero]

H_1 : $\beta_2 \neq 0$ [slope coefficients is not zero]

$$t_{crit} = t_{\frac{\alpha}{2}, n-2} = t_{\frac{0.05}{2}, 18-2} = 2.120$$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0.0975 - 0}{0.0130} = 7.4779$$



∴ Reject H_0. It means, slope coefficient is not zero at 0.05 level of significance.

Question 2. Using the 2015 Health and Welfare Survey from the National Statistical Office, a simple linear regression is modeled as follows,

$$outp_i = \beta_1 + \beta_2 age_i + u_i$$

where $outp_i$ is how many times person i has visited hospital in 2015, from 0 to 7 times

age_i is how old is person i, from 0 to 97 years.

We assume that both $outp_i$ and age_i are continuous, the estimation results in the following table. Answer the following questions and show your work.

Source	SS	df	MS			
				Number of obs	=	27,886
				F(1, 27884)	=	186.96
Model	77.5444409	1	77.5444409	Prob > F	=	0.0000
Residual	11565.0627	27,884	.414756231	R-squared	=	0.0067
				Adj R-squared	=	0.0066
Total	11642.6072	27,885	.417522223	Root MSE	=	.64402

outp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
age	.0031338	.0002292	Omitted		.0026846	.003583
_cons	.4279898	.0140339			.4004828	.4554969

a) Test if both parameters are significantly different from zero or not. Use $\alpha = 0.05$.

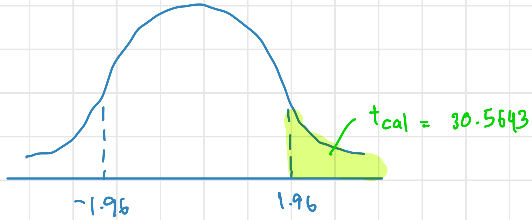
test β_1 (Intercept parameter)

$H_0 : \beta_1 = 0$ [Intercept parameter is zero]

$H_1 : \beta_1 \neq 0$ [Intercept parameter is not zero]

$t_{crit} = t_{\frac{\alpha}{2}, n-2} = t_{\frac{0.05}{2}, 27886} = 1.96$

$t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \dfrac{0.4279 - 0}{0.0140} = 30.5643$



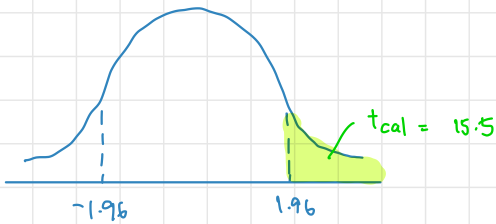
$\therefore$ Reject H_0. It means, intercept parameter is not zero at 0.05 level of significance.

<u>test β_2</u> (slope parameter)

$H_0: \beta_1 = 0$ [slope parameter is zero]
$H_1: \beta_1 \neq 0$ [slope parameter is not zero]

$t_{crit} = t_{\frac{\alpha}{2}, n-2} = t_{\frac{0.05}{2}, 27886} = 1.96$

$t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \dfrac{0.0031 - 0}{0.0002} = 15.5$



$\therefore$ Reject H_0. It means, slope parameter is not zero at 0.05 level of significance.

b) Interpret the meaning of $\hat{\beta}_2$. Does the sign of $\hat{\beta}_2$ make economic sense? Explain.

$\hat{\beta}_2 = 0.0031 \rightsquigarrow$ sig. at $\alpha = 0.05$ (from a)

Interpretion : If age change by 1 year,
the number of time visit hospital change by 0.0031 times,
in the same direction at 0.05 level percent of significant.

c) If $outp_i$ is turned into natural logarithmic scale (ln), how would you reinterpret the relationship between $\hat{\beta}_2$ and $\widehat{outp}_i$, assumed that the given coefficient given in the table above can be used to interpret this new functional form.

old $\widehat{outp} = 0.4279 + 0.0031$ age

New $\ln(outp) = 0.4279 + 0.0031$ age

old if age change by 1 year, number of time visit hospital change 0.0031 times,
in the same direction at 0.05 level percent of significance.

New if age change by 1 year, number of time visit hospital change 0.0031 %
in the same direction at 0.05 level percent of significance.

d) If age_i variable is divided by 10, how does it affect both the coefficients, standard errors, and confidence intervals? Answer the changes of both the constant and slope (if there is).

old $outp_i = \hat{\beta}_1 + \hat{\beta}_2 age_i + \hat{u}_i$

New $outp_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* age_i^* + u$

$w_1 = 1$, $w_2 = \frac{1}{10}$

$\hat{\beta}_2^* = \left[\frac{w_1}{w_2}\right] \hat{\beta}_2 \rightsquigarrow \hat{\beta}_2^* = \left[\frac{1}{1/10}\right] 0.0031 = 0.031$

$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \rightsquigarrow \hat{\beta}_1^* = [1]\, 0.4279 = 0.4279$

$se(\hat{\beta}_1^*) = w_1 \, se(\hat{\beta}_1) \rightsquigarrow se(\hat{\beta}_1^*) = [1]\, 0.0140 = 0.0140$

$se(\hat{\beta}_2^*) = \left[\frac{w_1}{w_2}\right] se(\hat{\beta}_2) \rightsquigarrow se(\hat{\beta}_2^*) = \left[\frac{1}{1/10}\right] 0.0002 = 0.002$

confidence interval of $\hat{\beta}_1$ at $\alpha = 0.05$

$$\hat{\beta}_1^* \pm t\,crit \cdot se(\hat{\beta}_1^*)$$

$$0.4279 \pm (1.96)(0.0140)$$

$$-0.4004 \leq \beta_1^* \leq 0.4553 \rightsquigarrow \text{not change}$$

confidence interval of $\hat{\beta}_2$ at $\alpha = 0.05$

$$\hat{\beta}_2^* \pm t\,crit \cdot se(\hat{\beta}_2^*)$$

$$0.031 \pm (1.96)(0.002)$$

$$-0.0271 \leq \beta_2^* \leq 0.0349 \rightsquigarrow \text{narrow down}$$

e) Find the confidence interval of mean prediction at the age of 50 years old, given that $var(\hat{Y}_0) = 0.00002$ and $\alpha = 0.01$.

$x_0 = age_0 = 50 \qquad \widehat{outp} = 0.4279 + 0.0031\, age$

sub. $age_0 = 50 \qquad \widehat{outp}_0 = 0.4279 + 0.0031(50) = 0.5829$

$var(\widehat{outp}_0) = 0.00002$

$$t\,cal = \frac{\widehat{outp}_0 - outp_0}{se(\widehat{outp}_0)} = \frac{0.5829 - 0}{\sqrt{0.00002}} = \frac{0.5829}{0.0044} = 130.3404$$

$$t\,crit = t_{\frac{\alpha}{2}, n-2} = t_{\frac{0.01}{2}, 27889-2} = 2.576$$

Interval : $\widehat{outp}_0 \pm t_{crit} \cdot se(\widehat{outp}_0) = 0.5829 \pm (2.576)(0.0044)$

$$0.5713 \leq outp_0 \leq 0.5944$$

Question 3. Discuss in a short paragraph why the confidence interval for both the mean prediction and individual prediction get larger as the X_0 is further away from $\bar{X}$.

The variance used for calculating confidence interval for the mean prediction is :

$$\mathrm{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma (X_i - \bar{X})^2} \right]$$

From above if $X_0 - \bar{X}$ gets larger, the variance gets larger leading to larger $se(\hat{Y}_0)$.