Solution to TUTORIAL 3

1. **(R code)**

   (a) [3.061384, 3.341033] on average, with 95% confidence, the mean freshman GPA is between 3.061384 and 3.341033 when their ACT test scores are 28

   (b) [1.959355, 4.443063], with 95% confidence her GPA will be between 1.959355 and 4.443063

   (c) Yes

2. **(R code)**

   (a) [16.67037, 19.72963], with 99% confidence, on average there are 16.67 to 19.73 broken ampuls after 2 times of transfer

   [22.77964, 29.62036], with 99% confidence, on average there are 22.78 to 29.62 broken ampuls after 4 times of transfer

   (b) [14.45319, 21.94681], with 99% confidence, there are 14.45 to 21.94 broken ampuls after 2 shipments

3. the α level used by the analyst was greater than 0.033, If the α level had been 0.01, he shod accept H_0

4. t-test can test $H_0 : \beta_1 = c$ for any constant c, but F-test can only test $H_0 : \beta_1 = 0$

5. **(R code)**

   (a) Set up the ANOVA table

   Response: y

source	Df	SS	MS	F-value	p-value
regression(x)	1	3.588	3.588	9.2402	0.002917
Residuals	118	45.818	0.388		
Total	119	49.406			

(b) conduct an F-test for $H_0 : \beta_1 = 0$ with $\alpha = 0.01$

Since p-value is smaller than $\alpha = 0.01$, we reject H_0, i.e. β_1 is significantly different from 0.

(c) what is the absolute magnitude of the reduction in the variation of Y when X is introduced into the model? what is the relative reduction? what is the name for the later measure?

the absolute magnitude of the reduction in the variation of Y when X is 3.588

the relative reduction 3.588/SST = 3.588/(3.588+45.818) = 7.262276%

called R^2

(d) obtain r_{XY} and attach the appropriate sign

$$r_{XY} = +\sqrt{R^2} = 0.2694818$$

(e) which measure R^2 or r has more clear-cut operational interpretation, explain.

r, because it give clear relationship between x and y.

6. **(R code)**

(a) set up the ANOVA table. which elements are additive?

Response: y

source	Df	SS	MS	F-value	p-value
regression(x)	1	160.0	160.0	72.727	2.749e-05
Residuals	8	17.6	2.2		
Total	9	177.6			

(b)

$$H_0 : \beta_1 = 0$$

$$F^* = 72.727 > F(1 - 0.05, 1, 8) = 5.32$$

reject H_0, there is significant linear association between X and Y

(c)

$$H_0 : \beta_1 = 0$$

$$|t^*| = 8.528 > t(1 - 0.05/2, 8) = 2.262$$

reject H_0, there is significant linear association between X and Y

(d)

$$R^2 = 90.09\%, \qquad r = 0.9491575$$

90.09% of the variation in Y is accounted for by introducing X into the model

7. **(R code)**

(a)

$$H_0 : \beta_1 \geq 0$$

$n = 60$

$$t^* = -13.19 < t(0.05, 58) = -1.671$$

we reject H_0, i.e there is a significant negative linear association between amount of muscle mass and age with $\alpha = 0.05$. The p-value for the test is $P(t(58) < -13.19)$

(b) Yes

(c) the output of the calculation is

Coefficients:

	Estimate	Std. Error	t value	P-value
(Intercept)	156.3466	5.5123	28.36	$< 2e-16$
x	-1.1900	0.0902	-13.19	$< 2e-16$

Residual standard error: 8.173 on 58 degrees of freedom

Multiple R-squared: 0.7501, Adjusted R-squared: 0.7458

F-statistic: 174.1 on 1 and 58 DF, $p-value :< 2.2e-16$

The difference is $\{b_0 + (x+1)b_1\} - (b_0 + xb_1) = b_1$, thus its 95% confidence interval

$$b_1 \pm s(b_1)t(1-0.05/2, 58) = -1.1900 \pm 0.0902 * t(1-0.05/2, 58) = -1.19 \pm 0.0902 * 2$$

(d) [82.83471, 87.05895], with 95% confidence the mean muscle mass for women of age 60 is from 82.83471 to 87.05895

(e) set up the ANOVA table

Response: y

source	Df	SS	MS	F-value	p-value
regression(x)	1	11627.5	11627.5	174.06	$< 2.2e-16$
Residuals	58	3874.4	66.8		
Total	59	15501.9			

(f)

$$1 - R^2 = 1 - 0.7501 = 0.2499$$

I think it is small

8. Find $Var(b_0 + 2b_1)$

Recall that

$$b_1 = \beta_1 + \sum_{i=1}^{n} k_i \varepsilon_i$$

where $k_i = (X_i - \bar{X}) / \sum_{j=1}^{n} (X_j - \bar{X})^2$.

$$\begin{aligned} b_0 + 2b_1 &= \bar{Y} - b_1\bar{X} + 2b_1 \\ &= \beta_0 + \beta_1\bar{X} + \bar{\varepsilon} + (2 - \bar{X})(\beta_1 + \sum_{i=1}^{n} k_i\varepsilon_i) \\ &= \beta_0 + 2\beta_1 + \bar{\varepsilon} + (2 - \bar{X}) \sum_{i=1}^{n} k_i\varepsilon_i \\ &= \beta_0 + 2\beta_1 + \sum_{i=1}^{n} \{\frac{1}{n} + (2 - \bar{X})k_i\}\varepsilon_i \end{aligned}$$

Thus

$$\begin{aligned} Var(b_0 + 2b_1) &= Var\Big\{ \sum_{i=1}^{n} \{\frac{1}{n} + (2 - \bar{X})k_i\}\varepsilon_i \Big\} \\ &= \sum_{i=1}^{n} \{\frac{1}{n} + (2 - \bar{X})k_i\}^2 Var(\varepsilon_i) \\ &= \{\frac{1}{n} + \frac{(2 - \bar{X})^2}{\sum_{j=1}^{n}(X_j - \bar{X})^2}\}\sigma^2. \end{aligned}$$

9. (a)

$$a_0 = \bar{Y} - a_1\bar{X}, \quad a_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

and

$$b_0 = \bar{X} - b_1\bar{Y}, \quad b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

(b) No

(c) $a_1 = b_1$