STAT 3004: Assignment 3

Cheng Li
1 Dec 2020

---

1 Problems 1

1.1 (11.13)

The least squares estimates are given by

$$b = \frac{L_{xy}}{L_{xx}},\ a = \frac{\sum y_i - b \sum x_i}{n},$$

where

$$\begin{aligned} L_{xx} &= \sum x_i^2 - \frac{(\sum x_i)^2}{17} = 1785 - \frac{153^2}{17} \\ &= 408, \\ L_{xx} &= \sum y_i^2 - \frac{(\sum y_i)^2}{17} = 226580 - \frac{1956^2}{17} \\ &= 1524.9, \\ L_{xy} &= \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{17} = 18387 - \frac{153 * 1956}{17} \\ &= 783, \end{aligned}$$

therefore, we can get $b = 1.92$ and $a = 97.8$. As a result, the least squares line is given by $y = 97.8 + 1.92x$

1.2 (11.14)

The standard errors of the regression parameters are given by

$$se(b) = \frac{s_{yx}}{\sqrt{L_{xx}}},\ se(a) = s_{yx}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}}},$$

where

$$\begin{aligned} s_{yx}^2 = ResMS &= \frac{L_{yy} - (L_{xy}^2/L_{xx}}{n-2} \\ &= \frac{1524.9 - (783^2/408)}{15} \\ &= 1.48, \end{aligned}$$

so, we can get $s_{yx} = \sqrt{1.48} = 1.22$. Thus, we have $se(b) = 0.0604$ and $se(a) = 0.619$.

1.3 (11.15)

Based on the linear model, the predicted systolic blood pressure for a 13-year-old boy is given by $\hat{y} = 97.8 + 1.92 * 13 = 122.8$ mm Hg. This is very close to the observed mean systolic blood pressure in this age group in these data (122 mm Hg).

1.4 (11.16)

We have

$$\begin{aligned} se &= s_{yx}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{L_{xx}}} \\ &= 1.22 * \sqrt{\frac{1}{17} + \frac{(13-9)^2}{408}} \\ &= 0.382\ mm\ Hg. \end{aligned}$$

2 Problem 2

2.1 (11.31)

We test the hypothesis $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$, where ρ = true correlation between reactivity as measured by the automated and manual monitors. We use the one-sample t test for correlation coefficients.

2.2 (11.32)

The test statistics is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.19 * \sqrt{77}}{\sqrt{1-0.19^2}} = 1.70 \sim t_{77}\ under\ H_0,$$

Since $t_{60,0.95} = 1.671 < 1.70 < t_{60,0.975} = 2.000$ and $t_{120,0.95} = 1.658 < 1.70 < t_{120,0.975} = 1.980$(You can also use R code to get the exact p-value of the test statistics.) Thus, there is a trend toward statistical significance; persons with greater changes in blood pressure as measured by the manual monitor tend to have higher changes as measured by the automated monitor. However, the relationship is weak and the significance level only borderline.($0.05 < p < 0.1$).

2.3 (11.33)

A 95% CI for z (the Fisher's z transform of ρ) is given by z_1, z_2, where

$$\begin{aligned} z_1 &= \hat{z} - 1.96\sqrt{\frac{1}{n-3}} \\ z_2 &= \hat{z} + 1.96\sqrt{\frac{1}{n-3}}. \end{aligned}$$

From the table 13, Appendix, text, we have that for $r = 0.19$, the z transform = 0.192. Therefore,

$$\begin{aligned} z_1 &= 0.192 - 1.96\sqrt{\frac{1}{76}} = -0.032 \\ z_2 &= 0.192 + 1.96\sqrt{\frac{1}{76}} = 0.417. \end{aligned}$$

The 95% CI for ρ is given by ρ_1, ρ_2, where

$$\rho_1 = \frac{exp(2z_1) - 1}{exp(2z_1) + 1} = -0.032$$
$$\rho_2 = \frac{exp(2z_2) - 1}{exp(2z_2) + 1} = 0.394.$$

Thus, the 95% CI for $\rho = (-0.032, 0.394)$.

3 Problem 3

3.1 (12.6)

We wish to test the hypothesis H_0 : all $\alpha_i = 0$ versus H_1 : at least one $\alpha_i \neq 0$. We will use the fixed effects one-way ANOVA. For this purpose, we compute the mean and standard deviation for each group as follows:

	$\bar{x}$	s	n
Group A	18.68	10.07	5
Group B	8.58	6.84	12
Group C	5.46	3.13	5

and the display of one-way ANOVA results is as follows:

Source of variation	SS	df	MS	F statistics	p-v
Between	$\sum_{i=1}^{3} n_i \bar{y_i}^2 - \frac{(\sum_{i=1}^{3} n_i \bar{y_i})^2}{\sum_{i=1}^{3} n_i} = 503.55$	2	$\frac{SSB}{df} = 251.77$	$\frac{MSB}{MSW} = 4.99 \sim F_{2,19}$ under H_0	$p <$
Within	$\sum_{i=1}^{3} (n_i - 1) s_i^2 = 958.80$	19	$\frac{SSW}{df} = 50.46$		
Total	1462.35				

Therefore there is a significant difference among the means.

3.2 (12.7)

We use the test statistic: $t = \frac{\bar{y_1} - \bar{y_2}}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim t_{n-k}$ under H_0. The results are given for each pair of groups as follows:

Groups	Test Statistic	p-value
A and B	$t = 2.67 \sim t_{19}$	$0.01 < p < 0.02$
A and C	$t = 2.94 \sim t_{19}$	$0.001 < p < 0.01$
B and C	$t = 0.82 \sim t_{19}$	$p > 0.05$

3.3 (12.8)

Under the Bonferroni method, the critical values are given $t_{19,\alpha^*/2}, t_{19,1-\alpha^*/2}$, where

$$\alpha = \frac{0.05}{\binom{3}{2}} = 0.0167$$

. Thus the critical values are $t_{19,0.0083}, t_{19,0.9917}$ and we can find that $t_{19,0.9917} = 2.63$. So we can get that there are significant differences between Group A and B and Group A and C while there is no significant difference between Group B and C.