Leyli Garryyeva
Professor deCamp

Assignment 4

CSCI 688
December 30, 2019

Chapter 4 - Simultaneous Inferences... : 4.3, 4.7, 4.16, 4.17, 4.22, 5.29, 5.31.

Problem 4.3

Refer to **Copier maintenance** Problem 1.20.
$X-$ number of copiers serviced
$Y-$ total number of minutes spent by service person

a. Will b_0 and b_1 tend to err in the same direction or in opposite directions here? Explain.

(4.5) $\sigma\{b_0, b_1\} = -\bar{X}\sigma^2\{b_1\}$
$\bar{X} = 5.111$

We know that if $\bar{X}$ is positive, then b_0 and b_2 are negatively correlated and have a positive covariance. Since our $\bar{X} > 0$ then b_0 and b_2 are negatively correlated and will tend to err in opposite directions.

b. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 95 percent family confidence coefficient.
$b_0 = -0.58$
$s\{b_0\} = 2.80$
$b_1 = 15.035$
$s\{b_1\} = 0.483$
$B = t(1-\alpha/4; n-2) = t(1-0.05/4; 45-2) = t(0.9875; 43) = 2.3226$

$$
\begin{aligned}
b_0 \pm Bs\{b_0\} &= -0.58 \pm 2.3226 \cdot 2.80 = -0.58 \pm 6.50328 = \\
&= \boxed{-0.58 \pm 6.503} \\
&= \boxed{-7.083 \leq \beta_0 \leq 5.923} \\
b_1 \pm Bs\{b_1\} &= 15.035 \pm 2.3226 \cdot 0.483 = 15.035 \pm 1.1218158 = \\
&= \boxed{15.035 \pm 1.122} \\
&= \boxed{13.913 \leq \beta_1 \leq 16.157}
\end{aligned} \tag{1}
$$

Then, β_0 is between -7.083 and 5.923 while β_1 is between 13.923 and 16.157. The family confidence coefficient is at least 0.95 that the procedure leads to correct pairs of interval estimates.

c. A consultant has suggested that β_0 should be 0 and β_1 should equal 14.0. Do your joint confidence intervals in part (b) support this view?

Yes, the suggested values of $\beta_0 = 0$ and $\beta_1 = 14$ are both within the joint confidence intervals.

Problem 4.7

Refer to **Copier maintenance** Problem 1.20.

a. Estimate the expected number of minutes spent when there are 3, 5, and 7 copiers to be serviced, respectively. Use interval estimates with a 90 percent family confidence coefficient based on the Working-Hotelling procedure.
$S_{xx} = 340.444$, $n = 45$, $\bar{X} = 5.111$, $MSE = 79.5$, $b_0 = -0.58$, $b_1 = 15.035$
$\hat{Y}_h \pm W \cdot s\{\hat{Y}_h\}$
$b_0 + b_1 X \pm W \cdot s\{\hat{Y}_h\}$
$W^2 = 2F(1-\alpha; 2, n-2) = 2F(0.9, 2, 43) = 2 * 2.430407155 = 4.86081431$
$W = \sqrt{W^2} = 2.2047$
For $X = 3$, $s\{\hat{Y}_h\} = \sqrt{MSE\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{xx}}\right]} = \sqrt{79.5 * \left[\frac{1}{45} + \frac{(3-5.111)^2}{340.444}\right]} = \sqrt{2.8073} = 1.6755$

For $X = 5$, $s\{\hat{Y}_h\} = \sqrt{79.5 * \left[\frac{1}{45} + \frac{(5-5.111)^2}{340.444}\right]} = \sqrt{1.76954} = 1.3302$

For $X = 7$, $s\{\hat{Y}_h\} = \sqrt{79.5 * \left[\frac{1}{45} + \frac{(7-5.111)^2}{340.444}\right]} = \sqrt{2.5999} = 1.6124$

$\boxed{\text{For X = 3}}$,

$$\begin{aligned} b_0 + b_1 X_h \pm W \cdot s\{\hat{Y}_h\} &= -0.58 + 15.035 * X_h \pm 2.2047 * s\{\hat{Y}_h\} \\ &= -0.58 + 15.035 * 3 \pm 2.2047 * 1.6755 = \\ &= \boxed{44.525 \pm 3.6940} \\ &= \boxed{40.831, 48.219} \end{aligned}$$

$\boxed{\text{For X = 5}}$,

$$\begin{aligned} b_0 + b_1 X_h \pm W \cdot s\{\hat{Y}_h\} &= -0.58 + 15.035 * X_h \pm 2.2047 * s\{\hat{Y}_h\} \\ &= -0.58 + 15.035 * 5 \pm 2.2047 * 1.3302 \\ &= \boxed{74.595 \pm 2.9327} \\ &= \boxed{71.662, 77.528} \end{aligned}$$

$\boxed{\text{For X = 7}}$,

$$\begin{aligned} b_0 + b_1 X_h \pm W \cdot s\{\hat{Y}_h\} &= -0.58 + 15.035 * X_h \pm 2.2047 * s\{\hat{Y}_h\} \\ &= -0.58 + 15.035 * 7 \pm 2.2047 * 1.6124 \\ &= \boxed{104.665 \pm 3.5549} \\ &= \boxed{101.110, 108.220} \end{aligned}$$

b. Two service calls for preventive maintenance are scheduled in which the numbers of copiers to be serviced are 4 and 7, respectively. A family of prediction intervals for the times to be spent on these calls is desired with a 90 percent family confidence coefficient. Which procedure, Scheffe or Bonferroni, will provide tighter prediction limits here?

Scheffe: $S^2 = gF(1-\alpha; g; n-2) = 2F(0.9; 2, 43) = 2 * 2.430407 = 4.8608$

$S = \sqrt{S^2} = \sqrt{4.8608} = 2.2047$

Bonferroni: $B = t(1-\alpha/4; n-2) = t(0.975, 43) = 2.01669$

Since $S = 2.2047 > B = 2.0170$, *Bonferroni procedure will provide tighter prediction limit.*

c. Obtain the family of prediction intervals required in part (b), using the more efficient procedure.

Bonferroni: $\hat{Y}_h \pm B \cdot s\{pred\}$

$\hat{Y}_h = b_0 + b_1 X_h$

$s\{pred\} = \sqrt{MSE\left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right]} = \sqrt{79.5\left[1 + \frac{1}{45} + \frac{(X_h - 5.111)^2}{340.444}\right]}$

For $X = 4$, $\hat{Y}_h = -0.58 + 15.035 * 4 = 59.56$

$s\{pred\} = \sqrt{79.5\left[\frac{46}{45} + \frac{(4-5.111)^2}{340.444}\right]} = \sqrt{81.5549} = 9.0308$

For $X = 7$, $\hat{Y}_h = -0.58 + 15.035 * 7 = 104.665$

$s\{pred\} = \sqrt{79.5\left[\frac{46}{45} + \frac{(7-5.111)^2}{340.444}\right]} = \sqrt{82.0999} = 9.0609$

$\boxed{\text{For X} = 4}$,

$$\begin{aligned} \hat{Y}_h \pm B \cdot s\{pred\} &= 59.56 \pm 2.0170 \cdot 9.0308 = \\ &= \boxed{59.56 \pm 18.215} = \\ &= \boxed{41.345, 77.775} \end{aligned} \tag{2}$$

$\boxed{\text{For X} = 7}$,

$$\begin{aligned} \hat{Y}_h \pm B \cdot s\{pred\} &= 104.665 \pm 2.0170 \cdot 9.0609 = \\ &= \boxed{104.665 \pm 18.2758} = \\ &= \boxed{86.389, 122.941} \end{aligned} \tag{3}$$

Problem 4.16

Refer to Copier maintenance Problem 1.20. Assume that linear regression through the origin model (4.10) is appropriate.

$Y_i = \beta_1 X_i + \epsilon_i$

$b_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$

a. Obtain the estimated regression function.

$\hat{Y} = b_1 X$. From the Minitab output in Figure 1:

$\boxed{\hat{Y} = 14.947X}$

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	338704	338704	4357.88	0.000
X	1	338704	338704	4357.88	0.000
Error	44	3420	78		
Lack-of-Fit	9	622	69	0.86	0.564
Pure Error	35	2798	80		
Total	45	342124			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8.81602	99.00%	98.98%	98.94%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
X	14.947	0.226	66.01	0.000	1.00

Regression Equation

Y = 14.947 X

Fits and Diagnostics for Unusual Observations

Obs	Y	Fit	Resid	Std Resid	
16	100.00	119.58	-19.58	-2.27	R
19	127.00	149.47	-22.47	-2.64	R

R Large residual

Figure 1: Minitab output for the Problem 4.16

b. Estimate β_1, with a 90 percent confidence interval. Interpret your interval estimate.
$b_1 \pm t(1-\alpha/2, n-1) \cdot s\{b_1\}$
$t(1-\alpha/2, n-1) = t(0.95, 44) = 1.68022992$
$s^2\{b_1\} = \frac{MSE}{\sum X_i^2}$. From Minitab output, $s\{b_1\} = 0.226$

$$\begin{aligned} b_1 \pm t(1-\alpha/2, n-1) \cdot s\{b_1\} &= 14.947 \pm 1.68022992 * 0.226 = \\ &= 14.947 \pm 0.3797319619 = \\ &= \boxed{14.947 \pm 0.3797} = \\ &= \boxed{14.567, 15.327} \end{aligned} \tag{4}$$

With 90 percent confidence we estimate that the mean variable total number of minutes spent by service person increases somewhere between 14.567 and 15.327 for each additional copier serviced.

c. Predict the service time on a new call in which six copiers are to be serviced. Use a 90 percent prediction interval.
$\hat{Y}_h \pm t(1-\alpha/2; n-1) \cdot s\{\hat{Y}_h\}$
$s\{\hat{Y}_h\} = 8.81602$ (from Minitab output)
$\hat{Y}_h = b_0 + b_1 X_h = b_1 X_h = 14.947 * 6 = 89.682$
$t(1-\alpha/2; n-1) = t(0.95, 44) = 1.68022992$

$$\begin{aligned}\hat{Y}_h \pm t(1-\alpha/2; n-1) \cdot s\{\hat{Y}_h\} &= 89.682 \pm 1.68022992 * 8.81602 = \\ &= 89.682 \pm 14.81294058 = \\ &= \boxed{89.682 \pm 14.813} = \\ &= \boxed{(74.869, 104.495)}\end{aligned} \tag{5}$$

Problem 4.17

Refer to **Copier maintenance** Problem 4.16.

a. Plot the fitted regression line and the data. Does the linear regression function through the origin appear to be a good fit here?

As we can see from the graph below, the fitted regression model is a good fit.

Figure 2: Plot of the fitted regression line and the data

b. Obtain the residuals e_i. Do they sum to zero? Plot the residuals against the fitted values $\hat{Y}_i$. What conclusions can be drawn from your plot?

Sum of the e_i's equals -5.86280. As we can see from the Figure 3, the residuals seem to have "constant variance."

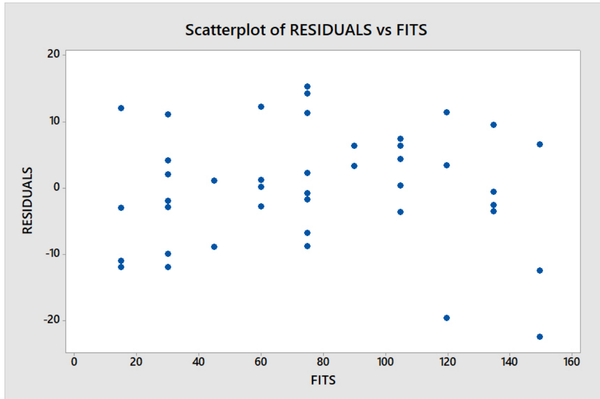

Figure 3: Plot of the fitted regression line and the data

c. Conduct a formal test for lack of fit of linear regression through the origin; use $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. What is the $P-$value of the test?

$$H_0 : E\{Y\} = \beta_1 X$$
$$H_a : E\{Y\} \neq \beta_1 X$$

Decision rule:

$$\text{If } F^* \leq F(1-\alpha; c-2, n-c), \text{ conclude } H_0$$
$$\text{If } F^* > F(1-\alpha; c-2, n-c), \text{ conclude } H_a.$$

From Figure 1:
Lack-of-fit $df = 9, SSLF = 622, MSLF = 69,$
Pure Error $df = 35, SSPE = 2798, MSPE = 80$
$P-$value $= 0.564$

$F(1-\alpha; c-2, n-c) = F(0.99, 9, 35) = 2.963011814$

$F^* = \frac{MSLF}{MSPE} = \frac{69}{80} = 0.8625$ (also given in the Figure 1)

$\boxed{F^* < 2.963}$ we conclude the H_0 that the regression function is linear ($P-$value $= 0.564$).

Problem 4.22

Derive an extension of the Bonferroni inequality (4.2a) for the case of three statements, each with statement confidence coefficient $1 - \alpha$.

$P(A_1) = P(A_2) = P(A_3) = \alpha$

$P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = 1 - P(A_1 \cup A_2 \cup A_3)$

$$
\begin{aligned}
P(A_1 \cup A_2 \cup A_3) &= P(A_1 \cup A_2) + P(A_3) - P\left((A_1 \cup A_2) \cap A_3\right) = \\
&= P(A_1) + P(A_2) - P(A_1 \cap A_2) + P(A_3) - P\left((A_1 \cap A_3) \cup (A_2 \cap A_3)\right) = \\
&= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - \\
&- \left[P(A_1 \cap A_3) + P(A_2 \cap A_3) - P\left((A_1 \cap A_3) \cap (A_2 \cap A_3)\right)\right] = \\
&= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + \\
&+ P(A_1 \cap A_2 \cap A_3)
\end{aligned}
$$

Then,

$$
\begin{aligned}
P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) &= 1 - P(A_1 \cup A_2 \cup A_3) = \\
&= 1 - P(A_1) - P(A_2) - P(A_3) + P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3) - \\
&- P(A_1 \cap A_2 \cap A_3)
\end{aligned}
$$

We know that $P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3) >= 0$ because $P(A_1 \cap A_2 \cap A_3) < P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)$. Thus,

$$
\begin{aligned}
P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) &>= 1 - P(A_1) - P(A_2) - P(A_3) \\
P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) &>= 1 - \alpha - \alpha - \alpha \\
P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) &>= 1 - 3\alpha
\end{aligned}
$$

$\boxed{\therefore P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) >= 1 - 3\alpha}$ is the extension of the Bonferroni inequality (4.2a) for the case of three statements, each with statement confidence coefficient $1 - \alpha$.

Problem 5.29

Consider the least squares estimator **b** given in (5.60). Using matrix methods, show that **b** is an unbiased estimator.

Bias $= E\{b\} - \beta = 0$

(5.60) $b = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y}$

$(\mathrm{X}'\mathrm{X})^{-1}(\mathrm{X}'\mathrm{X}) = 1$

$E\{\epsilon\} = 0$

$$
\begin{aligned}
E\{b\} &= E\{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'Y\} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'E\{Y\} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'E\{\mathrm{X}\beta + \epsilon\} = \\
&= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\left[E\{\mathrm{X}\beta\} + E\{\epsilon\}\right] = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{X}\beta = \beta
\end{aligned}
$$

Then, $E\{b\} - \beta = 0$ and **b** is an unbiased estimator of β.

Problem 5.31

Obtain an expression for the variance-covariance matrix of the fitted values $\hat{Y}_i, i = 1, ..., n$, in terms of the hat matrix.
$\hat{Y} = \mathrm{HY}$
$\sigma^2\{\mathrm{AY}\} = \mathrm{A}\sigma^2\{Y\}\mathrm{A}'$
$\sigma^2\{Y\} = \sigma^2\mathrm{I}$
$\mathrm{HH} = \mathrm{H}$

$$\sigma^2\{\hat{Y}\} = \sigma^2\{\mathrm{HY}\} = \mathrm{H}\sigma^2\{Y\}\mathrm{H}' = \mathrm{H}\sigma^2\mathrm{I}\mathrm{H}' = \sigma^2\mathrm{H}$$

$\boxed{\therefore \sigma^2\{\hat{Y}\} = \sigma^2\mathrm{H}}$