Chapter 4 Simultaneous Inferences and Other Topics in Regression Analysis

4.1 Joint Estimation of β_0 and β_1

- Assume that $x = 0$ is in the scope of the model, so inferences of β_0 is meaningful

- CI for β_0 and β_1

$$b_0 \pm t(1 - \frac{\alpha}{2})s(b_0)$$
$$b_1 \pm t(1 - \frac{\alpha}{2})s(b_1)$$

The coefficient for each confidence interval is $(1 - \alpha)$

- Question: what is the confidence coefficient for the collection of the above statements?

- Before the sample is taken and the confidence intervals computed, we know that

  (1) the statement that β_0 is in the interval

$$b_0 \pm t(1 - \frac{\alpha}{2})s(b_0)$$

  is correct with probability $1 - \alpha$.

  (2)the statement that β_1 is in the interval

$$b_1 \pm t(1 - \frac{\alpha}{2})s(b_1)$$

  is correct with probability $1 - \alpha$.

- What is the probability that both statements are simultaneously correct?

  (1) If the statements are independent, then the probability that both are correct is $(1-\alpha)(1-\alpha)$.

  (2) But they are not independent. The actual probability is difficult to compute.

- Want a family confidence coefficient for our family of statements

  (1) a joint rectangular region based on adjusted individual confidence interval for β_0 and β_1 using the Bonferroni method.

  (2) a joint elliptically shaped region for β_0 and β_1

Rectangular Joint Confidence Region

- Bonferroni Inequality

  Let $s_1, s_2 \cdots s_k$ be statements with

  $$p(s_i \text{ is true}) = 1 - \alpha_i$$

  then

  p(s_1 is true, s_2 is true $\cdots$ and s_k is true)

  =p(all s_i's are simultaneously true)

  $\geq$ 1-$\sum_{i=1}^{k} \alpha_i$

- Gives a lower bound on the probability that all statements are simultaneously true.

Example: Suppose $1 - \alpha_i = .90$, $k = 10$

$p(\text{All} \quad 10 \quad s_i's \quad \text{true}) \geq 1 - \sum_{i=1}^{10} .10 = 0$

The Bonferroni inequality works, but might not work very well.

Obtain a family confidence coefficient of $(1-\alpha)$ for confidence intervals for β_0 and β_1

- Example: If β_0 and β_1 both have 95% confidence intervals

$$b_0 \pm t(.975; n-2)s(b_0)$$

and

$$b_1 \pm t(.975; n-2)s(b_1)$$

The joint confidence coefficient using the Bonferroni inequality is greater than or equal to $1 - .05 - .05 = .90$

- To get a joint confidence coefficient of at least $(1-\alpha)$ for β_0 and β_1, we use the confidence intervals

$$b_0 \pm Bs\left\{b_0\right\}, \quad b_1 \pm Bs\left\{b_1\right\}$$

where $B = t(1-\alpha/4; n-2)$ The confidence coefficient is at least

$$1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

- To get a joint confidence coefficient of at least $(1-\alpha)$ for g parameters, we construct each interval estimate with statement confidence coefficient $1-\alpha/g$

The confidence coefficient is at least

$$1 - g * \frac{\alpha}{g} = 1 - \alpha.$$

Comments: For a given family confidence coefficient, the large the number of confidence intervals in the family, the greater becomes the multiple B, which may make some or all of the confidence intervals too wide to be helpful.

Example:

- Y outlier, examine the largest absolute standardized deleted residual, the appropriate α level test rejects if

$$\text{max}|(t_h)| \geq t(1-\frac{\alpha}{2n}, \text{df}_E - 1).$$

s_1: observation 1 is not an outlier

s_2: observation 2 is not an outlier

$\vdots$

s_n: observation n is not an outlier

To get a joint confidence coefficient of at least $(1-\alpha)$ for n parameters, we construct each interval estimate with statement confidence coefficient $1-\alpha/n$

Therefore, simultaneously, reject H_0, when

$$|(t_h)| \geq t(1-\frac{\alpha}{2n}, \text{df}_E - 1).$$

Mean response CI's

For all X_h with a confidence band: use Working-Hotelling

$$\hat{Y}_h \pm W s(\hat{Y}_h) \quad \text{where} \quad W^2 = 2F_{2;n-2}(1-\alpha)$$

For simultaneous estimation for a few X_h, say g different values, we may use Bonferroni approach

$$\hat{Y}_h \pm B s(\hat{Y}_h) \quad \text{where} \quad B = t_{n-2}(1-\alpha/(2g))$$

Examples (pages 158, 159)

Toluca company example, we require a family of estimates of the mean number of work hours at the following lot size level:

X_h	$\hat{Y}_h$	$s\left\{\hat{Y}_h\right\}$
30	169.5	16.97
65	294.4	9.918
100	419.4	14.27

For a family confidence coefficient of 0.90

- using Working-Hotelling procedure, we require $F(0.90; 2, 23) = 2.549$. Hence

$$W^2 = 2 * 2.549 = 5.098 \quad W = 2.258$$

- Using Bonferroni procedure,

$$B = t[1 - 0.10/2(3); 23] = t(0.9833; 23) = 2.263.$$

We can now obtain the confidence intervals for the mean number of work hours at $X_h = 30, 65,$ and 100:

$$131.2 = 169.5 - 2.258(16.97) \leq E\{Y_h\} \leq 169.5 + 2.258(16.97) = 207.8$$

$$272.0 = 294.4 - 2.258(9.918) \leq E\{Y_h\} \leq 294.4 + 2.258(9.918) = 316.8$$

$$387.2 = 419.4 - 2.258(14.27) \leq E\{Y_h\} \leq 419.4 + 2.258(14.27) = 451.6$$

With family confidence coefficient 0.90, we conclude that the mean number of work hours required is

- between 131.2 and 207.8 for lots of 30 parts
- between 272.0 and 316.8 for lots of 65 parts
- between 387.2 and 451.6 for lots of 100 parts.

- The family confidence coefficient 0.90 provides assurance that the procedure leads to all correct estimates in the family of estimates.

Using Bonferroni procedure,

$$B = t[1 - 0.10/2(3); 23] = t(0.9833; 23) = 2.263.$$

With 90 percent family confidence coefficient, we conclude that the mean number of work hours required is

- between 131.1 and 207.9 for lots of 30 parts
- between 272.0 and 316.8 for lots of 65 parts
- between 387.1 and 451.7 for lots of 100 parts.

Comments:

- In this instance the Working-Hotelling multiplier W is slightly smaller than the Bonferroni multiplier B. In other cases where the number of statements is small, the Bonferroni multiplier is usually smaller, so the confidence limits are tighter.
- For larger families, the Working-Hotelling confidence limits will always be the tighter, since W stays the same for any number of statements in the family whereas B becomes larger as the number of statements increase.
- In practice, once the family confidence coefficient has been decided upon, one can calculate the W and B to determine which procedure leads to tighter confidence limits.
- Both the Working-Hotelling and Bonferroni procedures provide lower bounds to the actual family confidence coefficient.

Simultaneous PIs

Simultaneous prediction intervals for g different X_h: use Bonferroni

$$\hat{Y}_h \pm Bs(\text{pred}) \quad \text{where} \quad B = t_{n-2}(1 - \alpha/(2g))$$

or Scheff$\acute{e}$

$$\hat{Y}_h \pm Ss(\text{pred}) \quad \text{where} \quad S^2 = gF_{g;n-2}(1 - \alpha)$$

Regression through the origin

$$Y_i = \beta_1 X_i + \varepsilon_i$$

where ε_i's are independent normal with mean 0 and variance σ^2.

The least square estimate of β_1 is

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum x_i^2}, \text{ so that } \hat{y}_i = b_1 x_i$$

$$MSE = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - b_1 x_i)^2$$

Comments:

- With regression through the origin, $\sum_{i=1}^{n} e_i \neq 0$. From the normal equation, the only constraints on the residuals is of the form $\sum X_i e_i = 0$. In a residual plot the residuals will usually not be balanced around the zero line.

- SSE may exceed the SSTO. This can occur when the data form a curvilinear pattern or a linear pattern with an intercept way from he origin.

- Care must be taken in using regression through the origin. If there is any doubt about $\beta_0 = 0$. A safer approach is to use the full model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ and test $H_0 : \beta_0 = 0$ v.s.$H_\alpha : \beta_0 \neq 0$