Chapter 4

Simultaneous Inferences and Other Topics in Regression Analysis

Instructor: Li, Caixia

Simultaneous Inference

- Start with ordinary confidence intervals for β_0 and β_1

$$b_0 \pm t(1-\alpha/2; n-2)s\{b_0\}$$
$$b_1 \pm t(1-\alpha/2; n-2)s\{b_1\}$$

- Each interval has $100(1-\alpha)\%$ confidence level.
- What about the overall/family confidence level?
  - Level of confidence that all constructed intervals contain their true parameter values
  - Often much lower than individual $100(1-\alpha)\%$ level

Bonferroni Inequality

- Let A_1 denote the event that the first confidence interval does not cover β_0 , i.e. $\mathrm{P}(A_1) = \alpha$
- Let A_2 denote the event that the second condence interval does not cover β_1 , i.e. $\mathrm{P}(A_2) = \alpha$
- We want to know the probability that both estimates fall in their respective confidence intervals, i.e. $P(\bar{A}_1 \cap \bar{A}_2)$

$$P\left(\bar{A}_1 \cap \bar{A}_2\right) = 1 - P\left(A_1 \cup A_2\right) = 1 - P\left(A_1\right) - P\left(A_2\right) + P\left(A_1 \cap A_2\right)$$

$$\Rightarrow 1 - 2\alpha \le 1 - P\left(A_1\right) - P\left(A_2\right) \le P\left(\bar{A}_1 \cap \bar{A}_2\right) \le 1 - P\left(A_1\right) = 1 - \alpha$$

$$\boxed{1 - 2\alpha \le \text{family confidence level} \le 1 - \alpha}$$

Bonferroni Correction for Simultaneous Inference

- Want to have family confidence level $100(1-\alpha)\%$ or family type I error rate α.
- Consider g CIs or hypothesis tests, each using $\alpha^* = \alpha/g$

$$\Pr\left(\bigcap_{i=1}^{g} \overline{A}_i\right) \geq 1 - g\alpha^* = 1-\alpha$$

where $\overline{A}_i$ is i-th acception region or CI

  - Use confidence level $1-\alpha/g$ for each CI
  - Use significance level α/g for each hypothesis test

- Increasingly conservative as g increases

Joint Confidence Intervals for β_0 and β_1

- Want to have family confidence level $100(1-\alpha)\%$ for joint estimation of β_0 and β_1
- Make each confidence interval at $100(1-\alpha/2)\%$ confidence level.

$$(1-\alpha/2)100\% \text{ CI for } \beta_0: \quad b_0 \pm t(1-\alpha/4; n-2)s\{b_0\}$$
$$(1-\alpha/2)100\% \text{ CI for } \beta_1: \quad b_1 \pm t(1-\alpha/4; n-2)s\{b_1\}$$

- E.g, if we want to find joint CIs with family confidence level 95%, we need set-up 97.5% confidence interval for each β_0 and β_1

Simultaneous Estimation of Mean Responses

- Working-Hotelling Method:
  - Confidence Band for Entire Regression Line.
  - Can be used for any number of Confidence Intervals for means, simultaneously
- Bonferroni Method:
  - Can be used for any g Confidence Intervals for means by creating $(1-\alpha/g)100\%$ CI[s] at each of g specified X levels

$$\text{Working-Hotelling: } \hat{Y}_h \pm W s\left\{\hat{Y}_h\right\} \qquad W = \sqrt{2F\left(1-\alpha; 2, n-2\right)}$$

$$\text{Bonferroni: } \hat{Y}_h \pm B s\left\{\hat{Y}_h\right\} \qquad B = t\left(1-\alpha/\left(2g\right); n-2\right)$$

Regression Through the Origin

- In some applications, it is believed that the regression line goes through the origin, this implies that $E\{Y|X\} = \beta_1 X$ (proportional relation)
- Should only be used if there is a strong theoretical reason
- Should use t-test for β_1
- Degrees of freedoms for some statistics are changed!

Model: $Y_i = \beta_1 X_i + \varepsilon_i \qquad \varepsilon_i \sim N\left(0, \sigma^2\right)$ independent

- LSE(least square estimator) or MLE of β_1

$$Q = \sum_{i=1}^{n}\left(Y_i - \beta_1 X_i\right)^2 \Rightarrow \frac{\partial Q}{\partial \beta_1} = 0 \Rightarrow b_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Regression Through the Origin

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \sum_{i=1}^{n} \left(\frac{X_i}{\sum_{i=1}^{n} X_i^2} \right) Y_i \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2} \right)$$

$$\hat{Y}_i = b_1 X_i \qquad e_i = Y_i - \hat{Y}_i \qquad SSE = \sum_{i=1}^{n} e_i^2 \sim \sigma^2 \chi^2 (n-1)$$

$$MSE = \frac{SSE}{n-1}, \quad s^2\{b_1\} = \frac{MSE}{\sum_{i=1}^{n} X_i^2}, \qquad \frac{b_1 - \beta_1}{s\{b_1\}} \sim t(n-1)$$

- $H_0: \quad \beta_1 = 0$
- Test statistic $\qquad t^* = \dfrac{b_1}{s\{b_1\}} \overset{H_0}{\sim} t(n-1)$

Regression Through the Origin

- Confidence intervals(CI) or prediction interval(PI)

$$(1-\alpha)100\%\ \text{CI for } \beta_1 : b_1 \pm t\left(1-(\alpha/2), n-1\right) s\{b_1\}$$

$$s^2\{b_1\} = \frac{MSE}{\sum_{i=1}^{n} X_i^2}; \quad s^2\left\{\hat{Y}_h\right\} = \frac{X_h^2}{\sum_{i=1}^{n} X_i^2} MSE; \quad s^2\{\text{pred}\} = MSE\left(1 + \frac{X_h^2}{\sum_{i=1}^{n} X_i^2}\right)$$

$$(1-\alpha)100\%\ \text{CI for } E(Y_h) = \beta_1 X_h : \hat{Y}_h \pm t\left(1-(\alpha/2), n-1\right) s\left\{\hat{Y}_h\right\}$$

$$(1-\alpha)100\%\ \text{PI for } Y_{h(new)} : \hat{Y}_h \pm t\left(1-(\alpha/2), n-1\right) s\{\text{pred}\}$$

Homework

- Page 172~(174):

  4.4 (b);   4.8(a),(b)

  4.16;   4.17(c)