STAT 525

Chapter 4
Miscellaneous Topics

Dr. Qifan Song

Simultaneous Inference

- Consider a collection of CIs / hypothesis tests
- Each interval has $(1-\alpha)\%$ confidence level
- What about the overall confidence level?
  - Level of confidence that all constructed intervals contain their true parameter values
  - Often much lower than individual $(1-\alpha)\%$ level
- Require to adjust individual confidence levels
- Recall confidence band $\longrightarrow$ widened individual intervals

Bonferroni Adjustment

- Want simultaneous CIs for θ_1 and θ_2
- Let A_1 denote event that CI excludes θ_1
- Let A_2 denote event that CI excludes θ_2
- We choose individual confidence level as $\text{Pr}(A_i) = \alpha_i$
- What is the probability that both events don't occur?

$$\begin{array}{rcl} \text{Pr}(\overline{A}_1 \cap \overline{A}_2) & = & 1 - \text{Pr}(A_1 \cup A_2) \\ \text{Pr}(A_1 \cup A_2) & = & \text{Pr}(A_1) + \text{Pr}(A_2) - \text{Pr}(A_1 \cap A_2) \\ & \leq & \text{Pr}(A_1) + \text{Pr}(A_2) \\ & \downarrow & \\ \text{Pr}(\overline{A}_1 \cap \overline{A}_2) & \geq & 1 - (\text{Pr}(A_1) + \text{Pr}(A_2)) \end{array}$$

- If $\alpha_1 + \alpha_2 = .05$, then $\text{Pr}(\overline{A}_1 \cap \overline{A}_2) \geq 0.95$

- Want to have family confidence level $1-\alpha$
- Consider g CIs each using $\alpha_i \equiv \alpha^*$

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

- Use level $1-\alpha/g$ for each test ($\alpha^* = \alpha/g$)
- Provides lower bound for confidence level
- Increasingly conservative as g increases
- True confidence level often higher than $1-\alpha$, so larger family-wise can be α used
- A universal method regardless of dependency

Joint Estimation of β_0 and β_1

- Joint confidence Intervals = rectangular confidence region of (β_0, β_1)
- If CIs were independent
  - Overall confidence level of rectangle is $(1 - \alpha_0)(1 - \alpha_1)$
  - Could set equal to 0.95 and solve for α
- Estimates (b_0,b_1) are not independent and MSE is shared by both CIs.
- Bonferroni Adjustment: $\alpha_0 = \alpha_1 = \alpha/2$
- Given normal error terms, can show (b_0, b_1) multivariate normal, thus can show natural (i.e., smallest) confidence region defined by an ellipse (STAT 524)

Mean Response CIs

- Could apply Bonferroni correction
  - Want to know $E(Y|X)$ for g X's
  - Construct CIs using $\alpha^* = \alpha/g$
  - Reasonable approach when g small

$$\widehat{Y}_h \pm B \times s(\widehat{Y}_h) \text{ where } B = t(1-\alpha/(2g), n-2)$$

- Previously discussed Working-Hotelling
  - Uses F distribution instead of t distribution
  - Coefficient W does not change as g increases

$$\widehat{Y}_h \pm W \times s(\widehat{Y}_h) \text{ where } W^2 = 2F(1-\alpha, 2, n-2)$$

Prediction Intervals

- Could apply Bonferroni correction
  - Want to know $Y_{h(new)}$ for g X's
  - Construct PIs using $\alpha^* = \alpha/g$
  - Reasonable approach when g small

$$\widehat{Y}_h \pm B \times s(\text{pred}) \text{ where } B = t(1 - \alpha/(2g), n - 2)$$

- Can also use Scheffé procedure
  - Uses F distribution instead of t distribution
  - Coefficient S increases as g increases

$$\widehat{Y}_h \pm S \times s(\text{pred}) \text{ where } S^2 = gF(1 - \alpha, g, n - 2)$$

Regression through the Origin

- Many instances where the line is known to go through the origin
- Statistical model is

$$Y_i = \beta_1 X_i + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, \sigma^2)$$

  - Can show $b_1 = \sum X_i Y_i / \sum X_i^2$, and $se(b_1) = MSE / \sum X_i^2$
  - Can fit using option `NOINT` for `MODEL` in `PROC REG`
  - Degrees of freedoms for some statistics are changed! $MSE = \sum (Y_i - \widehat{Y}_i)^2 / (n - 1)$
  - Problems with R^2
- If the line does go through the origin, little is lost fitting a line with both intercept and slope

Measurement Error

- Measurement Error in Y
  - Generally not a problem if the measurement error is random, independent with mean zero
  - Error term in model represents unexplained variation which is often a combination of many factors not considered

- Measurement Error in X
  - Does cause problems
  - Often results in biased estimators
  - Tends to reduce strength of association
  - $Y = \beta_0 + X\beta_1 + \epsilon = \beta_0 + (X^* - \delta)\beta_1 + \epsilon = \beta_0 + X^*\beta_1 + (\epsilon - \delta\beta_1) := \beta_0 + X^*\beta_1 + \epsilon^*$, where X^* and ϵ^* are correlated in general, unless X^* is fixed (Berkson model)

Inverse Predictions

- Given Y_h, predict X, such that $Y_h = X_h\beta_1 + \beta_0 + \epsilon$
- Given fitted equation this is

$$\hat{X}_h = \frac{Y_h - b_0}{b_1} \qquad\qquad b_1 \neq 0$$

- This is the MLE (i.e., function of b_0, b_1)
- Approximate CI can be constructed using inverse mapping of CI for new observation Y_h:
  Given a known X_h,

$$b_0+b_1X_h-t(1-\alpha/2,n-2)s\{pred\} \leq Y_h \leq b_0+b_1X_h+t(1-\alpha/2,n-2)s\{pred\}$$

$$Y_h - t(1-\alpha/2,n-2)s\{pred\} \leq b_0 + b_1X_h \leq Y_h + t(1-\alpha/2,n-2)s\{pred\}$$

$$\Longrightarrow X_h \in \hat{X}_h \pm t(1-\alpha/2,n-2)s\{pred\}/b_1$$

Note: $s\{pred\}$ requires true X_h, and we can plug in $\hat{X}_h$ for an approximation.

Chapter Review

- Simultaneous Inference / Multiplicity
- Regression through the origin
- Measurement Error
- Inverse predictions