Topic 6 - Miscellaneous Topics

STAT 525 - Fall 2013

Outline

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

Simultaneous Inference

- Consider a collection of confidence intervals
- Each interval has $(1-\alpha)\%$ confidence level
- What about the overall confidence level?
  - This is defined as the level of confidence that all constructed intervals contain their true parameter values
  - Often much lower than the individual $(1-\alpha)\%$ CI level
- We will adjust the individual confidence levels upward so the overall CI level is closer to $1-\alpha$
- Recall confidence band $\longrightarrow$ widened individual intervals

Joint Estimation of β_0 and β_1

- Will focus here on forming a joint rectangular region formed by the individual CIs
- If estimates were independent
  - Overall confidence level of rectangle is $(1-\alpha)^2$
  - Could set equal to 0.95 and solve for α
- Estimates (b_0,b_1) are not independent so how do we handle this?
- Given normal error terms, can show (b_0, b_1) multivariate normal
- Can show natural (i.e., smallest) confidence region defined by an ellipse (STAT 524)

Bonferroni Correction

- Let A_1 denote event that CI excludes β_0
- Let A_2 denote event that CI excludes β_1
- By construction $\Pr(A_1) = \Pr(A_2) = \alpha$
- What is prob that both events don't occur?

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

- If $\Pr(A_1) + \Pr(A_2) = .05$, then $\Pr(\overline{A}_1 \cap \overline{A}_2) \geq 0.95$

Bonferroni Correction

- Want to have family confidence level $1 - \alpha$
- Consider g tests or CIs each using α^*

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

- 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 much higher than $1 - \alpha$ so larger family-wise α used

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

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

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

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

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

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

- Can also use Scheffe' procedure
  - Uses F distribution instead of t distribution
  - Coefficient increases as g increases

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

Regression through the Origin

- Many instances where theory suggests true population line should go through origin
- Statistical model under this restriction is

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

- Can show $b_1 = \sum X_i Y_i / \sum X_i^2$ in this case
- In a sense you are forcing b_0 to be zero
- Model causes problems with R^2 and other statistics
- Little is lost fitting the intercept and slope in all cases
- **Note:** If no intercept, no adjustment necessary for family of tests

Measurement Error

- Measurement Error in Y
  - Generally not a problem provided error is random and unbiased
  - Error term in model represents unexplained variation which is often a combination of many factors not considered
- Measurement Error in X
  - Can cause problems
  - Often results in biased estimators (slope shrunk towards zero)
  - Reduces strength of association
  - Berkson error model: special case where predictor variable is set at a target level. This does not result in biased parameter estimates.

Inverse Predictions

- Given Y_h, predict corresponding X, $\hat{X}_h$
- Given fitted equation this is

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

- This is the MLE (i.e., function of b_0, b_1)
- Approximate CI can be constructed using inverse mapping of CI for $\hat{Y}_h$

$$\frac{Y_h \pm t(1-\alpha/2, n-2)s(\hat{Y}_h) - b_0}{b_1}$$
$$\hat{X}_h \pm t(1-\alpha/2, n-2)s(\hat{Y}_h)/b_1$$

Background Reading

- KNNL Chapter 4
- KNNL Chapter 5 : Matrix Algebra