Contents

Day 26

Final Exam Procedures

- If the grade for either lecture or lab is better than the worse score, that score will replace that worse score
- Lab portion is take home (will take a week to do it)
  - Four questions
    - probability (pbinom, qnorm, t distributions, probability stuff) <- question 1 exam 1 and 2
    - Midterm 1 inference problem (neyman pearson, NHST, fisher, chi squared)
    - Midterm 2 inference (t procedures, 1 sample, 2 sample, matched pairs, CI, Hypot test (NHST), power analysis (but not for two sample), one way ANOVA)
    - Inference for linear regression (ANOVA, rest of this current lecture can be on this problem)
- Lecture
  - longer version of the other midterms
  - Multiple cheat sheets allowed

t-Test for Slope

Almost exclusively, we use NHST:

- H_0: $\beta_1 = 0$
- H_a: $\beta \neq 0$

Recall:

$$y = \beta_0 + \beta_1 x + \epsilon$$

$$\epsilon \sim N(0, \sigma)$$

- $\mu_{\mathrm{y|x}} = \beta_0 + \beta_1 x$
- $y = \mu_{\mathrm{y|x}} + \epsilon$

$$t = \frac{\text{statistic - parameter}}{\text{standard error}} = \frac{b_1 - \beta_1}{SE_{\mathrm{b_1}}}$$

Where $\beta_1 = 0$

Under H_0, $t \sim t(n-2)$

We reject H_0:

- Slope is not 0, so a linear relationship between x and y exists.

Fail to reject H_0:

- It is reasonable to believe that slope is 0 and there is no linear relationship exists between x & y.

Confidence Interval for Slope

Recall: Find the two-sided critical region. Want to find values of β_1 for which the t-statistic is NOT in the critical region.

$$-t^{**} < \frac{b_1 - \beta_1}{SE_{\mathrm{b}_1}} < t^{**}$$

$$b_1 - t^{**} \times SE_{\mathrm{b}_1}(2.5\%) < \beta_1 < b_1 + t^{**} \times SE_{\mathrm{b}_1}(97.5\%)$$

$$t^{**} \sim t(n-2)$$

Interpretation

We are 95% confident in our estimate that when x-variable increases by 1 unit, the population mean of y-variable increases (1 decreases) by between lower bound and upper bound

Mean Response

Model: - $\mu_{\text{y|x}} = \beta_0 + \beta_1 x \leftarrow$ confidence interval for mean response

Confidence Interval for $\mu_{\mathbf{y|x}}$: requires a specific x^* value to plug in for x.

$$\hat{\mu} = b_0 + b_1 x^*$$

$\hat{y}$ (Estimate y at x^*) and $\hat{\mu}_{\text{y}}$ (estimate μ_{y} at x^*) have the same value.

- Bigger sample size (s) $\rightarrow$ wider confidence interval
- Bigger sample size (s) means weaker relationship (in general)
- Weaker relationship $\rightarrow$ wider confidence interval
- Confidence interval at x^* close to $\bar{x} \rightarrow$ narrower
- Confidence interval at x^* far from $\bar{x} \rightarrow$ wider

$$PI : \hat{y} \pm t^{**} \times SE_{\hat{y}}$$

include formula for SE hat y

At the same x^* value, prediction interval for y is always wider than the confidence interval for μ_{y}

Interpretation

Confidence interval for $\mu_{\text{y|x}}$:

We are 95% confidence in our estimate that when x-variable is value of x^*, the population mean of y-variable for a new observation is between lower bound and upper bound