MATH 900 Practicum - (STAT 231 - Statistics)

Ziye Lin

Mar. 21, 2025

Practicum Details

- We derived the sampling distribution of $\tilde{\mu}(x) = \tilde{\alpha} + \tilde{\beta}x$, shown $E(\tilde{\mu}) = \alpha + \beta x$, $\text{Var}(\tilde{\mu}) = \sigma^2 \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right]$ and derived the pivotal quantity $T = \frac{\tilde{\mu}(x) - \mu(x)}{S_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t(n-2)$

- We used the notation $b_i = \sum_{i=1}^{n} \frac{1}{n} + \frac{(x_i - \bar{x})(x-\bar{x})}{S_{xx}}$ and left the following exercises for students to take home and try to show:

$$(1)\ \sum_{i=1}^{n} b_i = 1\ ;$$

$$(2)\ \sum_{i=1}^{n} b_i x_i = x\ ;$$

$$(3)\ \sum_{i=1}^{n} b_i^2 = \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}$$

  I want students to try them at home to save some time.

- I added an extra page to show why $T = \frac{\tilde{\mu}(x) - \mu(x)}{S_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}}$ follows a t-distribution with $n-2$ degrees of freedom. Some students asked this question in the lecture and there is a theorem in Chapter 4 for this result in STAT 231 course notes.

- We constructed the $100p\%$ confidence interval for $\mu(x)$: $\hat{\mu}(x) \pm a \cdot S_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$

- For a future response Y, we shown that $E(Y - \tilde{\mu}(x)) = 0$, $\text{Var}(Y - \tilde{\mu}(x)) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right]$

- We shown that $\frac{Y - \tilde{\mu}(x)}{S_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t(n-2)$ and constructed the $100p\%$ prediction interval for Y: $\hat{\mu}(x) \pm a \cdot S_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$

- We discussed why the prediction interval for Y is wider than the confidence interval for $\mu(x)$, where the extra source of uncertainty is

- We started section 6.3, discussed the two main assumptions for Gaussian linear response models.

- Discussed why we need to check model assumptions

- We talked about the first method - scatterplot of data and fitted regression line superimposed

- Quickly mentioned the definition of residual plots, the three types of residuals plots. **But did not have time to discuss the expected pattern of residual plots when the model assumptions are correct**

- Suggestion to course instructor for the next lecture: start by talking about what do we expect to see in the residual plots if the model assumptions are correct

Observations

- There are 28 students attended the lecture
- Some students seem to be not familiar with the structure/assumptions of Gaussian linear response models. When I asked them what the expected value and the variance of Y_i are, some of them seem to have no idea how to find them, even when I write out the model $Y_i = \alpha + \beta x_i + R_i$.
- Some students seem to not understand the difference between the covariate x_i and the x in $\mu(x) = \alpha + \beta x$. Similarly, some students seem to not understand the difference between a future response Y and the Y_i in our sample data
- Some students asked why replacing σ by S_e in the denominator makes $\frac{\tilde{\mu}(x)-\mu(x)}{S_e\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}}$ to follow a t-distribution, also a similar question for $\frac{Y-\tilde{\mu}(x)}{S_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}}$