Gaussian Response Models

Selected Topic from STAT 221/231

Ziye Lin - MATH 900 Microteaching Session 3

Distributions of $\tilde{\beta}_0$ and $\tilde{\beta}_1$

Distributions of $\tilde{\beta}_0$ and $\tilde{\beta}_1$

Model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ , \ \epsilon_i \overset{i.i.d}{\sim} G(0, \sigma) \ , \ i = 1, \ldots, n \ .$$

We have derived the following estimators together:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} x_i (Y_i - \bar{Y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})} \quad \text{(for } \hat{\beta}_1\text{)}$$

$$\tilde{\beta}_0 = \bar{Y} - \tilde{\beta}_1 \bar{x} \qquad \text{(for } \hat{\beta}_0\text{)}$$

Distributions of $\tilde{\beta}_0$ and $\tilde{\beta}_1$

Model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ , \ \epsilon_i \overset{i.i.d}{\sim} G(0, \sigma) \ , \ i = 1, \ldots, n \ .$$

We have derived the following estimators together:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n x_i (Y_i - \bar{Y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} \quad \text{(for } \hat{\beta}_1\text{)}$$

$$\tilde{\beta}_0 = \bar{Y} - \tilde{\beta}_1 \bar{x} \qquad \text{(for } \hat{\beta}_0\text{)}$$

What are the distributions of $\tilde{\beta}_0$ and $\tilde{\beta}_1$?

Distribution of $\tilde{\beta}_1$

Let $a_i = \dfrac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $i = 1, \dots, n$

Distribution of $\tilde{\beta}_1$

$$\text{Let } a_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \ , \ i = 1, \dots, n$$

Identities (proofs are omitted for now):

$$\sum_{i=1}^{n} a_i = 0 \quad (1)$$

$$\sum_{i=1}^{n} a_i x_i = 1 \quad (2)$$

$$\sum_{i=1}^{n} a_i^2 = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (3)$$

Distribution of $\tilde{\beta}_1$

Notice that, with

$$a_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\ ,\ i = 1, \dots, n$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i(x_i - \bar{x})} = \dots = \sum_{i=1}^{n}\left\{\frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} y_i\right\} = \sum_{i=1}^{n} a_i y_i$$

(why?)

Distribution of $\tilde{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i(y_i - \bar{y})}{\sum_{i=1}^n x_i(x_i - \bar{x})}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \sum_{i=1}^n \left\{ \left[\frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] y_i \right\} = \sum_{i=1}^n a_i y_i \quad .$$

Distribution of $\tilde{\beta}_1$

$$\begin{aligned}
\mathbb{E}(\tilde{\beta}_1) &= \mathbb{E}\left[\sum_{i=1}^{n}\left\{\frac{(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}Y_i\right\}\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} a_i Y_i\right] \\
&= \sum_{i=1}^{n} a_i \mathbb{E}[Y_i] \\
&= \sum_{i=1}^{n} a_i(\beta_0 + \beta_1 x_i + \overbrace{\mathbb{E}(\epsilon_i)}^{=0}) \text{ (recall } \epsilon_i \overset{i.i.d}{\sim} G(0,\sigma)) \\
&= \beta_0 \sum_{i=1}^{n} a_i + \beta_1 \sum_{i=1}^{n} a_i x_i \\
&= \beta_1 \quad \text{using identities (1) and (2)}
\end{aligned}$$

Distribution of $\tilde{\beta}_1$

$$
\begin{aligned}
\mathrm{Var}(\tilde{\beta}_1) &= \mathrm{Var}\left[\sum_{i=1}^{n} a_i Y_i\right] \\
&= \sum_{i=1}^{n} a_i^2 \mathrm{Var}[Y_i] + \overbrace{\sum_{i \neq j} \mathrm{Cov}(Y_i, Y_j)}^{0} \\
&= \mathrm{Var}(\epsilon_1) \sum_{i=1}^{n} a_i^2 \quad (\text{recall } \epsilon_i \overset{i.i.d}{\sim} G(0, \sigma)) \\
&= \sigma^2 \sum_{i=1}^{n} a_i^2 \\
&= \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{using identity } (3)
\end{aligned}
$$

Distribution of $\tilde{\beta}_1$

Model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ , \ \epsilon_i \overset{i.i.d}{\sim} \ G(0, \sigma) \ , \ i = 1, \dots, n \ .$$

Estimator of β_1:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} x_i (Y_i - \bar{Y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})}$$

Distribution of $\tilde{\beta}_1$:

$$\mathbb{E}(\tilde{\beta}_1) = \beta_1$$

$$\mathrm{Var}(\tilde{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\Rightarrow \quad \tilde{\beta}_1 \sim G\left(\beta_1 \ , \ \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \ \right)$$

Exercises

Recall:

$$a_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \ , \ i = 1, \dots, n$$

$$\sum_{i=1}^{n} a_i = 0 \quad (1)$$

$$\sum_{i=1}^{n} a_i x_i = 1 \quad (2)$$

$$\sum_{i=1}^{n} a_i^2 = \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (3)$$

Exercises: prove these identities.

Distribution of $\tilde{\beta}_0$

How about the distribution of $\tilde{\beta}_0$?

Distribution of $\tilde{\beta}_0$

How about the distribution of $\tilde{\beta}_0$?

$$\tilde{\beta}_0 \;\sim\; G\left(\beta_0 \;,\; \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}\;\right)$$

Distribution of $\tilde{\beta}_0$

How about the distribution of $\tilde{\beta}_0$?

$$\tilde{\beta}_0 \ \sim \ G\left(\beta_0 \ , \ \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}\ \right)$$

(Assignment Question!)

Next step...

Why care about the distributions of $\tilde{\beta}_0$ and $\tilde{\beta}_1$?

$$\mathbb{E}(Y_i|x_i) = \beta_0 + \beta_1 x_i \;\; \text{(mean response given the covariate } x_i\text{)}$$

Next step...

Why care about the distributions of $\tilde{\beta}_0$ and $\tilde{\beta}_1$?

$$\mathbb{E}(Y_i|x_i) = \beta_0 + \beta_1 x_i \text{ (mean response given the covariate } x_i\text{)}$$

What is coming next:

- Confidence intervals of β_1 and β_0
- Test of hypothesis involving β_1 and β_0
- Inferences and predictions via simple linear regression
- Estimator of σ^2 (recall $\epsilon_i \overset{i.i.d}{\sim} G(0, \sigma)$)

References

Department of Statistics and Actuarial Science.
STAT 221/231 Course Notes Fall 2021 Edition.
University of Waterloo, Waterloo, ON, 2021.

Chris Springer, Jerry Lawless, Don McLeish, and Cyntha Struthers.
STAT 220/230 Course Notes Fall 2021 Edition.
University of Waterloo, Waterloo, ON, 2021.