

JOE SEIDEL

STAT 245

HOMEWORK 4

Sampling distribution: simple derivation

Consider independent observations $y_1, \dots, y_n \sim N(\beta_0, \sigma^2)$. The MLE of β_0 is $\hat{\beta}_0 = \bar{y}$. The residual is $\hat{e}_i = y_i - \bar{y}$ for $i = 1, \dots, n$. In class we learned $\bar{y} \sim N(\beta_0, \frac{\sigma^2}{n})$ and $\sum_{i=1}^n \hat{e}_i^2 / \sigma^2 \sim \chi_{n-1}^2$. In order to obtain a t-distribution, we need independence between \bar{y} and $\sum_{i=1}^n \hat{e}_i^2$. Here is a simple way to do it.

(a) Calculate $\text{Cov}(\bar{y}, \hat{e}_i)$

$$\begin{aligned} \text{Cov}(\bar{y}, \hat{e}_i) &= \text{Cov}(\bar{y}, y_i - \bar{y}) \\ &= \text{Cov}(\bar{y}, y_i) - \text{Cov}(\bar{y}, \bar{y}) \\ &= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0 \end{aligned}$$

(b) Can you claim the independence between \bar{y} and $\sum_{i=1}^n \hat{e}_i^2$?

Since we found $\text{Cov}(\bar{y}, \hat{e}_i) = 0$ by the Gaussian assumption we may assume independence.

Residual

Consider independent observations $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ for $i = 1, \dots, n$. For the LSE $\hat{\beta}_0$ and $\hat{\beta}_1$, define the residual $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ for $i = 1, \dots, n$.

(a) Calculate $\mathbb{E}(\frac{1}{n-2} \sum_{i=1}^n \hat{e}_i^2)$.

First calculate $\mathbb{E}(\hat{e}_i^2)$, where we can rewrite $\hat{e}_i = y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})$. Since $\mathbb{E}(\hat{e}_i) = 0$ we get

$$\begin{aligned} \mathbb{E}(\hat{e}_i^2) &= \text{Var}(\hat{e}_i) \\ &= \text{Var}(y_i - \bar{y}) + (x_i - \bar{x})^2 \text{Var}(\hat{\beta}_1) - 2(x_i - \bar{x}) \text{Cov}(y_i - \bar{y}, \hat{\beta}_1). \end{aligned}$$

Then,

$$\begin{aligned} \text{Var}(y_i - \bar{y}) &= \frac{n-1}{n} \sigma^2, \\ \text{Var}(\hat{\beta}_1) &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \end{aligned}$$

and

$$\begin{aligned}\text{Cov}(y_i - \bar{y}, \hat{\beta}_1) &= \text{Cov}\left(y_i - \bar{y}, \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}\right) \\ &= \frac{\sum_{j=1}^n (x_j - \bar{x}) \text{Cov}(y_i - \bar{y}, y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}.\end{aligned}$$

When $i = j$,

$$\text{Cov}(y_i - \bar{y}, y_j - \bar{y}) = \text{Var}(y_i - \bar{y}) = \frac{n-1}{n}\sigma^2$$

otherwise

$$\text{Cov}(y_i - \bar{y}, y_j - \bar{y}) = -\frac{\sigma^2}{n}.$$

Then

$$\frac{\sum_{j=1}^n (x_j - \bar{x}) \text{Cov}(y_i - \bar{y}, y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{(x_i - \bar{x}) \frac{n-1}{n}\sigma^2 - \sum_{j \neq i} (x_j - \bar{x}) \frac{\sigma^2}{n}}{\sum_{j=1}^n (x_j - \bar{x})^2},$$

which can also be written

$$\frac{(x_i - \bar{x}) \frac{n-1}{n}\sigma^2 - \sum_{j \neq i} (x_j - \bar{x}) \frac{\sigma^2}{n}}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{\sigma^2(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}.$$

Now we have everything we need to find

$$\begin{aligned}\mathbb{E}(\hat{e}_i^2) &= \text{Var}(y_i - \bar{y}) + (x_i - \bar{x})^2 \text{Var}(\hat{\beta}_1) - 2(x_i - \bar{x}) \text{Cov}(y_i - \bar{y}, \hat{\beta}_1) \\ &= \frac{n-1}{n}\sigma^2 + (x_i - \bar{x})^2 \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} - 2 \frac{\sigma^2(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}.\end{aligned}$$

Finally, we have

$$\sum_{i=1}^2 \mathbb{E}(\hat{e}_i^2) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2$$

which means

$$\mathbb{E}\left(\frac{1}{n-2} \sum_{i=1}^n \hat{e}_i^2\right) = \sigma^2.$$

(b) Calculate $\text{Cov}(\hat{\beta}_1, \hat{e}_i)$.

$$\begin{aligned}\text{Cov}(\hat{\beta}_1, \hat{e}_i) &= \text{Cov}(\hat{\beta}_1, y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})) \\ &= \text{Cov}(\hat{\beta}_1, y_i - \bar{y}) - (x_i - \bar{x}) \text{Cov}(\hat{\beta}_1, \hat{\beta}_1)\end{aligned}$$

Where we've found $\text{Cov}(\hat{\beta}_1, y_i - \bar{y})$ in the previous part and

$$\text{Cov}(\hat{\beta}_1, \hat{e}_i) = 0$$

(c) Can you claim the independence between $\hat{\beta}_1$ and $\sum_{i=1}^n \hat{e}_i^2$?

From the result in part (b) and the Gaussian assumption, they are independent.

(d) Are $\hat{\beta}_0$ and $\sum_{i=1}^n \hat{e}_i^2$ independent?

$$\begin{aligned}\text{Cov}(\hat{\beta}_0, \hat{e}_i) &= \text{Cov}(\bar{y} - \hat{\beta}_1 x_i, y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})) \\ &= -(x_i - \bar{x}) \text{Cov}(\bar{y}, \hat{\beta}_1) + \bar{x}(x_i - \bar{x}) \text{Var}(\hat{\beta}_1) - \bar{x} \text{Cov}(\hat{\beta}_1, y_i - \bar{y}).\end{aligned}$$

All items above have been found in previous part of this problem which are used to find

$$\text{Cov}(\hat{\beta}_0, \hat{e}_i) = 0$$

allowing us to claim independence.

Joint Distribution of $(\hat{\beta}_0, \hat{\beta}_1)$

Consider independent observations $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ for $i = 1, \dots, n$. For the LSE $\hat{\beta}_0$ and $\hat{\beta}_1$ we derived $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$ in the class.

(a) Calculate $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$. First observe

$$\begin{aligned}\text{Cov}(\bar{y}, \hat{\beta}_1) &= \text{Cov}\left(\bar{y}, \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x}) (\text{Cov}(\bar{y}, y_i - \bar{y}))}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= 0\end{aligned}$$

Then

$$\begin{aligned}\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= \text{Cov}(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1) \\ &= -\bar{x} \text{Cov}(\hat{\beta}_1, \hat{\beta}_1) \\ &= -\bar{x} \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.\end{aligned}$$

(b) What is the joint distribution of $(\hat{\beta}_0, \hat{\beta}_1)$

$$N \left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} & -\bar{x} \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ -\bar{x} \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix} \right)$$

Linear regression without slope

Consider independent observations $y_i \sim N(\beta_1 x_i, \sigma^2)$ for $i = 1, \dots, n$.

(a) Find the MLE for β_1 , denoted as $\hat{\beta}_1$.

Finding the MLE boils down to minimizing the following element of the joint likelihood function $f(y_1, y_2, \dots, y_n \mid \beta_1)$ where $f(x)$ is the normal density,

$$\sum_{i=1}^n (y_i - \beta_1 x_i)^2.$$

Minimize by taking the derivative and setting it equal to 0,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

(b) Find $\mathbb{E}(\hat{\beta}_1)$.

$$\mathbb{E}(\hat{\beta}_1) = \frac{\sum_{i=1}^n x_i \mathbb{E}(y_i)}{\sum_{i=1}^n x_i^2} = \beta_1$$

(c) Find $\text{Var}(\hat{\beta}_1)$.

$$\text{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^n x_i \text{Var}(y_i)}{\sum_{i=1}^n x_i^2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

(d) What is the distribution of $\hat{\beta}_1$?

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$$

Linear regression with centered covariates

Some people like to center their x_i before applying regression. This leads to the model $y_i \sim N(\beta_0 + \beta_1(x_i - \bar{x}), \sigma^2)$ independently for $i = 1, \dots, n$.

(a) Find the MLE for β_0, β_1 , denoted as $\hat{\beta}_0, \hat{\beta}_1$.

Minimize

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1(x_i - \bar{x}))^2$$

by setting the derivative equal to 0 to find the MLE

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

and

$$\hat{\beta}_0 = \bar{y}.$$

(b) Find the expectations of $\hat{\beta}_0, \hat{\beta}_1$.

Expectations are β_0 and β_1 , respectively.

(c) Find the variances of $\hat{\beta}_0, \hat{\beta}_1$.

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

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n}$$

Check the matrix formula

For the multivariate linear regression with model $y \sim N(X\beta, \sigma^2 I_n)$, we showed in class that the MLE is $\hat{\beta} = (X^T X)^{-1} X^T y$. It has distribution $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$. Now consider the simple case of $p = 2$ so that

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

1. (a) For $p = 2$, work out the formula $(X^T X)^{-1} X^T y$.
2. (b) For $p = 2$, work out the formula $\sigma^2 (X^T X)^{-1}$.
3. (c) Do these formulas give you the same answers that we learned for $p = 2$ in class?

(LSE=MLE)

For $y \sim N(X\beta, \sigma^2 I_n)$ write down the likelihood function of y . Show that maximizing the likelihood function is equivalent to minimizing $\|y - X\beta\|^2$.