JOE SEIDEL

STAT 245 HOMEWORK 4

1. *Sampling distribution: simple derivation*

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

(a) Calculate $\mathrm{Cov}(\overline{y}, \hat{e}_i)$

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

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

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

2. *Residual*

Consider independent observations $y_i \sim N(\beta_0 + \beta_1 x_i,\ \sigma^2)$ for $i = 1, ..., n$. 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 $i = 1, ..., n$.

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

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

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

Then,

$$\mathrm{Var}(y_i - \overline{y}) = \frac{n-1}{n}\sigma^2,$$

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

and

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

When $i = j$,

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

otherwise

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

Then

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

which can also be written

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

Now we have everything we need to find

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

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

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

Where we've found $\text{Cov}(\hat{\beta}_1, y_i - \overline{y})$ 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?

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

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.

3. *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 $i = 1, ..., n$. 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

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

Then

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

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

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

4. *Linear regression without slope*

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

(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 $f(y_1, y_2, ...y_n \mid \beta_1)$ 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 $\hat{\beta}_1$?

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

5. *Linear regression with centered covariates*

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

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

Minimize

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

by setting the derivative equal to 0 to find the MLE

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

and

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

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

Expectations are β_0 and β_1, respectively.

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

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

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

6. 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

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

1. (a) For $p = 2$, work out the formula $(X^T X)^{-1} X^T y$.

First

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

and

$$(X^T X)^{-1} = \frac{1}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix}.$$

Therefore

$$(X^T X)^{-1} X^T y =$$
$$\frac{1}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} (\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n x_i y_i) \\ -(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i) + n \sum_{i=1}^n x_i y_i \end{pmatrix}.$$

The value on the second row is

$$\begin{aligned}
&\frac{-(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i) + n\sum_{i=1}^n x_i y_i}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \overline{xy}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2}.
\end{aligned}$$

For the first row

$$\begin{aligned}
&\frac{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n x_i y_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= \frac{(\frac{1}{n}\sum_{i=1}^n x_i^2)\overline{y} - \overline{x}(\frac{1}{n}(\sum_{i=1}^n x_i y_i)}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2} \\
&= \frac{(\frac{1}{n}\sum_{i=1}^n x_i^2)\overline{y} - \overline{x}^2\overline{y} + \overline{x}^2\overline{y} - \overline{x}(\frac{1}{n}(\sum_{i=1}^n x_i y_i)}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2} \\
&= \overline{y} - \overline{x}\frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \overline{xy}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2} \\
&= \overline{y} - \hat{\beta}_1\overline{x}.
\end{aligned}$$

2. (b) For $p = 2$, work out the formula $\sigma^2(X^TX)^{-1}$. On the first row, first column

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_0) &= \sigma^2\frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= \frac{\sigma^2}{n}\frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 + (\sum_{i=1}^n x_i)^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= \frac{\sigma^2}{n}\Big(1 + \frac{(\sum_{i=1}^n x_i)^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}\Big) \\
&= \frac{\sigma^2}{n} + \frac{\overline{x}\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}.
\end{aligned}$$

The entry in the second row, second column

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_1) &= \frac{\sigma^2 n}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}.
\end{aligned}$$

3. (c) Do these formulas give you the same answers that we learned for $p = 2$ in class? yes.

7. (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 $\|y - X\beta\|^2$.

The likelihood function is the PDF of $N(X\beta,\ \sigma^2 I_n)$. Therefore

$$l(\beta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\|y - X\beta\|^2}$$

which is maximized when $\|y - X\beta\|^2$ is minimized.