Homework 4 (due May 16)

In your solution include your name and the homework number. Please staple your pages together. When solving the problems below, give detailed derivations in order to receive credit.

1. *(Sample distribution: simple derivation)* Consider independent observations $y_1, ..., y_n \sim N(\beta_0, \sigma^2)$. The MLE of β_0 is $\hat{\beta}_0 = \bar{y}$. The residue is $\hat{e}_i = y_i - \bar{y}$ for $i = 1, ..., n$. In class, we learned $\bar{y} \sim N\left(\beta_0, \frac{\sigma^2}{n}\right)$ 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 the independence between $\bar{y}$ and $\sum_{i=1}^n \hat{e}_i^2$. Here is a simple way to do it:

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

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

   Solution. $\mathsf{Cov}(\bar{y}, \hat{e}_i) = \mathsf{Cov}\left(\bar{y}, y_i - \bar{y}\right) = \mathsf{Cov}\left(\bar{y}, y_i\right) - \mathsf{Cov}(\bar{y}, \bar{y}) = \sigma^2/n - \sigma^2/n = 0$. This implies $\bar{y}$ and $\hat{e}_i$ are independent because they are jointly Gaussian. Therefore, $\bar{y}$ and $\sum_{i=1}^n \hat{e}_i^2$ are independent.

2. *(Residue)* 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 residue $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ for $i = 1, ..., n$.

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

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

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

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

   Solution. We first calculate $\mathbb{E}\hat{e}_i^2$. Observe that $\hat{e}_i = y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})$. Since $\mathbb{E}\hat{e}_i = 0$, we get

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

   Then,

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

   and

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

   When, $i = j$, $\mathsf{Cov}(y_i - \bar{y}, y_j - \bar{y}) = \frac{n-1}{n}\sigma^2$. When $i \neq j$, $\mathsf{Cov}(y_i - \bar{y}, y_j - \bar{y}) = -\frac{\sigma^2}{n}$. Therefore,

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

which can be organized as

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

Thus,

$$\mathsf{Cov}(y_i-\bar{y},\hat{\beta}_1)=\frac{\sigma^2(x_i-\bar{x})}{\sum_{j=1}^n(x_j-\bar{x})^2}.$$

With the calculation above, we get

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

This leads to

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

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

Now we calculate $\mathsf{Cov}(\hat{\beta}_1,\hat{e}_i)$. By its formula,

$$\mathsf{Cov}(\hat{\beta}_1,\hat{e}_i)=\mathsf{Cov}\left(\hat{\beta}_1,y_i-\bar{y}-\hat{\beta}_1(x_i-\bar{x})\right)=\mathsf{Cov}\left(\hat{\beta}_1,y_i-\bar{y}\right)-(x_i-\bar{x})\mathsf{Cov}\left(\hat{\beta}_1,\hat{\beta}_1\right).$$

We have calculated $\mathsf{Cov}\left(\hat{\beta}_1,y_i-\bar{y}\right)$ above. Therefore, $\mathsf{Cov}(\hat{\beta}_1,\hat{e}_i)=0$. This implies the independence between $\hat{\beta}_1$ and $\sum_{i=1}^n\hat{e}_i^2$ by Gaussianity. Finally,

$$\mathsf{Cov}(\hat{\beta}_0,\hat{e}_i)=\mathsf{Cov}(\bar{y}-\hat{\beta}_1\bar{x},y_i-\bar{y}-\hat{\beta}_1(x_i-\bar{x})),$$

which equals

$$-(x_i-\bar{x})\mathsf{Cov}(\bar{y},\hat{\beta}_1)+\bar{x}(x_i-\bar{x})\mathsf{Var}(\hat{\beta}_1)-\bar{x}\mathsf{Cov}(\hat{\beta}_1,y_i-\bar{y}).$$

where we have used the fact that $\mathsf{Cov}(\bar{y},y_i-\bar{y})=0$. We already have the formulas of $\mathsf{Var}(\hat{\beta}_1)$ and $\mathsf{Cov}(\hat{\beta}_1,y_i-\bar{y})$. Then, it is not hard to see that $\mathsf{Cov}(\bar{y},\hat{\beta}_1)=0$. This gives

$$\mathsf{Cov}(\hat{\beta}_0,\hat{e}_i)=\bar{x}(x_i-\bar{x})\mathsf{Var}(\hat{\beta}_1)-\bar{x}\mathsf{Cov}(\hat{\beta}_1,y_i-\bar{y})=0.$$

We can claim the independence between $\bar{y}$ and $\sum_{i=1}^n\hat{e}_i^2$.

3. *(Joint distribution of $(\hat{\beta}_0,\hat{\beta}_1)$)* Consider independent observations $y_i\sim N(\beta_0+\beta_1x_i,\sigma^2)$ for $i=1,...,n$. For the LSE $\hat{\beta}_0$ and $\hat{\beta}_1$, we derived $\mathsf{Var}(\hat{\beta}_0)$ and $\mathsf{Var}(\hat{\beta}_1)$ in the class.

   (a) Calculate $\mathsf{Cov}(\hat{\beta}_0,\hat{\beta}_1)$.

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

Solution. We first show $\mathsf{Cov}(\bar{y}, \hat{\beta}_1) = 0$. This is because

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

Then

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

The joint distribution of $(\hat{\beta}_0, \hat{\beta}_1)$ is

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

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 $\hat{\beta}_1$.

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

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

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

   Solution. Write down the joint likelihood function and maximize over β_1. Observe that this is equivalent to minimizing over

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

   Take derivative of the above objective and set it to zero. We obtain $\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$. The mean is

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

   The variance is

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

   The distribution is

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

5. *(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 $i = 1, ..., n$.

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

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

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

Solution. The MLE is defined by minimizing

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

Take derivative of the above objective and set it to zero. We obtain

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

and

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

The expectations are

$$\mathbb{E}\hat{\beta}_1 = \beta_1, \quad \mathbb{E}\hat{\beta}_0 = \beta_0.$$

The variance of $\hat{\beta}$ has the old formula

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

and the variance of $\hat{\beta}_0$ is

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

6. *(Check the matrix formula)* For 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

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

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

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

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

Solution. Note that

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

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

and

$$X^T y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.$$

Therefore,

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

The value on the second row of $(X^T X)^{-1} X^T y$ is

$$\begin{aligned} & \frac{-\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right) + n\sum_{i=1}^n x_i y_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ = & \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2}, \end{aligned}$$

which is exactly the formula of $\hat{\beta}_1$ we had in the class. The value on the first row of $(X^T X)^{-1} X^T y$ is

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

which is exactly the formula of $\hat{\beta}_0$ we had in the class.

Next, we try to organize $\sigma^2(X^T X)^{-1}$. The entry on the first row first column is $\mathsf{Var}(\hat{\beta}_0)$, and it has formula

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

and this is the formula we had in the class. The entry on the second row second column of $\sigma^2(X^TX)^{-1}$ is $\mathsf{Var}(\hat{\beta}_1)$, and it has formula

$$\frac{\sigma^2 n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

which is also the formula we had in the class.

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

Solution. The likelihood function is the probability density function of $N(X\beta, \sigma^2 I_n)$. Thus,

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

Since $l(\beta)$ increases as $\|y - X\beta\|^2$ decreases, maximizing $l(\beta)$ is equivalent to minimizing $\|y - X\beta\|^2$.