HW 1

Madeline Blasingame

205412082

10/2/2021

Exercise 1

1. Predictor Assumption: $\hat{Y}_0 = \sum_{i=1}^{n} a_i Y_i$

in order for $\hat{Y}_0$ to be unbiased $E(\hat{Y}_0) = \beta_1 X_0$ must be true

$$E(\hat{Y}_0) = E\left(\sum_{i=1}^{n} a_i Y_i\right)$$

$$\beta_1 X_0 = \sum_{i=1}^{n} a_i E(Y_i) = \sum_{i=1}^{n} a_i \beta_1 X_i$$

$$\beta_1 X_0 = \beta_1 \sum_{i=1}^{n} a_i X_i$$

$\sum_{i=1}^{n} a_i X_i = X_0$ must be true in order for $\hat{Y}_0$ to be unbiased.

2. We know that $E(a^2) = var(a) + [E(a)]^2$
using this rule we can show:
$$E(Y_0 - \hat{Y}_0)^2 = var(Y_0 - \hat{Y}_0) + [E(Y_0 - \hat{Y}_0)]^2$$

We also know: $E(Y_0 - \hat{Y}_0) = 0$ because $E(Y_0) = E(\hat{Y}_0) = \beta_1 X_0$ when $\hat{Y}_0$ is unbiased.

Therefore:
$$E(Y_0 - \hat{Y}_0)^2 = var(Y_0 - \hat{Y}_0)$$

3. We need to minimize $var(Y_0 - \hat{Y}_0)$ with the constraint: $\sum_{i=1}^{n} a_i X_i = X_0$

$$\min Q = var(Y_0 - \hat{Y}_0) - 2\lambda\left[\sum a_i X_i - X_0\right]$$

$$= var(Y_0) + var(\hat{Y}_0) + \underline{2\,cov(Y_0, \hat{Y}_0)} - 2\lambda\left[\sum a_i X_i - X_0\right]$$

$$cov(Y_0, \hat{Y}_0) = cov\left(Y_0, \sum a_i Y_i\right)$$

$$= \sum a_i\, cov(Y_0, Y_i) = 0$$

because $Y_i \perp\!\!\!\perp Y_j$ s.t. $i \neq j$

$$\min\ Q = var(Y_0) + var(\hat{Y}_0) - 2\lambda\left[\sum a_i x_i - x_0\right]$$

$$= \sigma^2 + \sigma^2 \sum a_i^2 - 2\lambda\left[\sum a_i x_i - x_0\right]$$

$$\frac{\partial Q}{\partial a_i} = 2\sigma^2 a_i - 2\lambda x_i = 0$$

$$a_i = \frac{2\lambda x_i}{2\sigma^2} = \frac{\lambda x_i}{\sigma^2}$$

$$\sum a_i x_i = \sum \frac{\lambda x_i}{\sigma^2} x_i = \sum \frac{\lambda x_i^2}{\sigma^2} = \frac{\lambda}{\sigma^2} \sum x_i^2$$

$$\sum a_i x_i = x_0$$

$$x_0 = \frac{\lambda \sum x_i^2}{\sigma^2} \implies \lambda = \frac{x_0 \sigma^2}{\sum x_i^2}$$

$$a_i = \frac{2\lambda x_i}{2\sigma^2} = \frac{2 x_0 \sigma^2 x_i}{2\sigma^2 \sum x_i^2} = \frac{x_0 x_i}{\sum x_i^2}$$

$$\hat{Y}_0 = \sum a_i Y_i$$

$$\hat{Y}_0 = \frac{\sum x_0 x_i Y_i}{\sum x_i^2} = x_0 \frac{\sum x_i Y_i}{\sum x_i^2}$$

Exercise 2

a1, b1, c1 $\rightarrow$ $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

a2, b2, c2 $\rightarrow$ $y_i = \beta_1 x_i + \epsilon_i$

a.1. Show: $\sum_{i=1}^{n} e_i = 0$

$$\sum_{i=1}^{n} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = \sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

$\boxed{\text{We know: } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$

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

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

$$= 0$$

b.1. Show: $\sum_{i=1}^{n} e_i x_i = 0$

By Definition: $e_i = y_i - \hat{y}_i$

We know: $\hat{y}_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x})$

Therefore: $e_i = y_i - (\bar{y} + \hat{\beta}_1 (x_i - \bar{x}))$

$$\sum \left[y_i - (\bar{y} + \hat{\beta}_1 (x_i - \bar{x})) \right] [x_i]$$

$$\sum (y_i - \bar{y})(x_i) - \hat{\beta}_1 \sum (x_i - \bar{x}) x_i$$

We know: $\hat{\beta}_1 = \dfrac{\sum (Y_i - \bar{Y}) x_i}{\sum (x_i - \bar{x})^2}$

$$\sum (y_i - \bar{y})(x_i) - \frac{\sum (Y_i - \bar{Y}) x_i}{\sum (x_i - \bar{x})^2} \sum (x_i - \bar{x}) x_i$$

$$\sum (y_i - \bar{y})(x_i) \left[1 - \underbrace{\frac{\sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x})^2}} \right]$$

If: $k_i = \dfrac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$

Then: $\sum k_i x_i = \dfrac{\sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x})^2}$

Property given in class: $\sum k_i x_i = 1$

$$\sum (y_i - y)(x_i) [1 - 1]$$

$$\sum (y_i - y)(x_i)(0) = 0$$

c.1. Show: $\sum_{i=1}^{n} e_i \hat{y}_i = 0$

$$\boxed{\text{We know: } \hat{y}_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x})}$$

$$\sum e_i (\bar{y} + \hat{\beta}_1 (x_i - \bar{x}))$$

$$\sum e_i \bar{y} + \hat{\beta}_1 e_i (x_i - \bar{x})$$

$$\bar{y} \sum e_i + \hat{\beta}_1 \sum e_i x_i - \hat{\beta}_1 \bar{x} \sum e_i$$

Shown in part a: $\sum e_i = 0$
Shown in part b: $\sum e_i x_i = 0$

$$\bar{y}(0) + \hat{\beta}_1 (0) - \hat{\beta}_1 \bar{x} (0) = 0$$

a2. show: $\sum e_i = 0$

$e_i = y_i - \hat{y}_i$
$\hat{y}_i = \hat{\beta}_1 x_i$

using MLE or least squares we find: $\hat{\beta}_1 = \frac{\sum x_i Y_i}{\sum x_i^2}$

$$\sum e_i = \sum y_i - \hat{y}_i = \sum y_i - \hat{\beta}_1 x_i =$$

$$\sum y_i - \hat{\beta}_1 \sum x_i = \sum y_i - \frac{\sum x_i Y_i}{\sum x_i^2} \sum x_i$$

This cannot be reduced to 0 therefore $\sum e_i \neq 0$.

b.2. show: $\sum_{i=1} e_i x_i = 0$

$e_i = y_i - \hat{y}_i$
$\hat{y}_i = \hat{\beta}_1 x_i$

$$\sum (y_i - \hat{\beta}_1 x_i) x_i$$

$$\sum y_i x_i - \hat{\beta}_1 \sum x_i^2$$

$\hat{\beta}_1 = \frac{\sum x_i Y_i}{\sum x_i^2}$ using MLE or least squares

$$\sum y_i x_i - \frac{\sum x_i y_i \sum x_i^2}{\sum x_i^2}$$

$$\sum y_i x_i - \sum y_i x_i = 0$$

C.2. show: $\sum e_i \hat{y}_i = 0$ $\qquad$ $\hat{y}_i = \hat{\beta}_1 x_i$

$$\sum e_i \hat{\beta}_1 x_i = \hat{\beta}_1 \sum e_i x_i = \hat{\beta}_1 \underbrace{(0)}_{\text{shown in part b}} = 0$$

The conditions $\sum e_i x_i = 0$ and $\sum e_i \hat{y}_i = 0$ still hold when $y_i = \beta_1 x_1 + \epsilon_i$ however $\sum e_i = 0$ does not.

Exercise 3

Given:

$Y_i = \beta_1 X_i + \epsilon_i$

$\epsilon_i \sim N(0, \sigma)$

Therefore:

$Y_i \sim N(\beta_1 X_i, \sigma)$

Because Y_i follows a normal distribution the likelihood functions are:

$$L = \prod \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Y_i - \beta_1 X_i)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} e^{-1/2\sigma^2 \sum (Y_i - \beta_1 X_i)^2}$$

$$\ln(L) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(y_i - \beta_1 x_i)^2$$

Find the MLE of β_1.

$$\frac{\partial \ln(L)}{\partial \beta_1} = 0$$

$$\ln(L) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum y_i^2 + \frac{2\beta_1}{2\sigma^2}\sum x_i y_i - \frac{\beta_1^2}{2\sigma^2}\sum x_i^2$$

$$\frac{\partial \ln(L)}{\partial \beta_1} = \frac{1}{\sigma^2}\sum x_i y_i - \frac{\beta_1}{\sigma^2}\sum x_i^2 = 0$$

$$\sum x_i y_i = \beta_1 \sum x_i^2$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

Show that the MLE of β_1 is the same as the estimator obtained using the method of least squares:

$$\text{Min } Q = \sum \epsilon_i^2$$

$$\text{Min } Q = \sum (Y_i - \beta_1 X_i)^2 = \sum Y_i^2 - 2\beta_1 \sum X_i Y_i + \beta_1^2 \sum X_i^2$$

$$\frac{\partial Q}{\partial \beta} = 0$$

$$\frac{\partial Q}{\partial \beta} = -2\sum X_i Y_i + 2\beta_1 \sum X_i^2$$

$$\sum X_i Y_i = \beta_1 \sum X_1^2$$

$$\hat{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$$

this is equivalent to the estimator found using MLE.

Find the MLE of σ^2:

$$\frac{\partial \ln(L)}{\partial \sigma^2} = 0 \qquad \ln(L) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum (y_i - \beta_1 x_i)^2$$

$$\frac{\partial \ln(L)}{\partial \sigma^2} = \frac{-n2\pi}{2\cdot 2\pi\sigma^2} + \frac{2}{(\sigma^2)^2}\sum (y_i - \beta_1 x_i)^2 = 0$$

$$\frac{2}{(\sigma^2)^2}\sum (y_i - \beta_1 x_i)^2 = \frac{n}{2\sigma^2}$$

$$\frac{2}{\sigma^2}\sum (y_i - \beta_1 x_i)^2 = \frac{n}{2}$$

$$\frac{4\sum (y_i - \beta_1 x_i)^2}{n} = \hat{\sigma}^2$$

Find the variance of $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum \overset{\downarrow \text{constant}}{x_i} y_i}{\underset{\nearrow \text{constant}}{\sum x_i^2}}, \quad y_i = \underset{\uparrow \text{constant}}{\beta_1 x_i} + \overset{\downarrow \text{random}}{\epsilon_i}$$

we know that y_i is independent of y_j when $i \neq j$ because ϵ_i and ϵ_j are independent when $i \neq j$ (given in the question)

$$\hat{\beta}_1 = \sum \frac{x_i}{\sum x_j^2} y_i \qquad a_i = \frac{x_i}{\sum x_j^2} \qquad \rightarrow \qquad \hat{\beta}_1 = \sum a_i y_i$$

$\Rightarrow \quad var(\hat{\beta}_1) = Var(\sum a_i y_i) = \sum var(a_i y_i)$ due to the independence of $y_1 \cdots y_i$

$$= \sum a_i^2 \, var(y_i) \qquad var(y_i) = var(\epsilon_i) = \sigma^2$$

$$= \sum \left[\frac{x_i}{\sum x_j^2}\right]^2 \sigma^2$$

Find the mean of $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\overset{\downarrow \text{constant}}{\sum_i} x_i y_i}{\underset{\nearrow \text{constant}}{\sum_i x_i^2}} , \quad y_i = \underset{\substack{\uparrow \\ \text{constant}}}{\beta_1 x_i} + \overset{\downarrow \text{random}}{\epsilon_i}$$

we know that y_i is independent of y_j when $i \neq j$ because ϵ_i and ϵ_j are independent when $i \neq j$ (given in the question)

$$\hat{\beta}_1 = \sum_i \frac{x_i}{\sum_j x_j^2} y_i \qquad a_i = \frac{x_i}{\sum_j x_j^2} \qquad \rightarrow \qquad \hat{\beta}_1 = \sum_i a_i y_i$$

$\Rightarrow E(\hat{\beta}_1) = E(\sum_i a_i y_i) = \sum_i E(a_i y_i)$ due to the independence of $y_1 \dots y_i$

$= \sum_i a_i E(y_i)$

$E(y_i) = \beta_1 x_i$

$$= \sum_i \frac{x_i}{\sum_j x_j^2} \beta_1 x_i$$

Exercise 4

$\hat{\beta}_0 = \sum l_i y_i \qquad l_i = \frac{1}{n} - \bar{x} k_i \qquad k_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$

consider: $b_0 = \sum a_i y_i$ another unbiased estimator of β_0

Then: $E(b_0) = \beta_0$

$E(\sum a_i Y_i) = \beta_0$

$\sum a_i E Y_i = \beta_0$

$\sum a_i (\beta_0 + \beta_1 X_i) = \beta_0$

$\beta_0 \sum a_i + \beta_1 \sum a_i X_i = \beta_0 \longrightarrow \begin{matrix} \sum a_i = 1 \\ \sum a_i X_i = 0 \end{matrix}$

$var(b_0) = var[\sum a_i y_i] = \sigma^2 \sum a_i^2$ due to the fact $y_i \perp\!\!\!\perp y_j$ s.t. $i \neq j$

and $var(\hat{\beta}_0) = \sigma^2 \sum l_i^2 = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\sum (x_i - \bar{x})^2}$

Now let $a_i = l_i + d_i$

Then: $var(b_0) = \sigma^2 \sum (l_i + d_i)^2$

$= \sigma^2 \sum (l_i + d_i)^2$

$= \sigma^2 \sum l_i^2 + \sigma^2 \sum d_i^2 + 2\sigma^2 \underbrace{\sum l_i d_i}$ we want this to be 0

Show that $\sum l_i d_i = 0 \qquad a_i = l_i + d_i \Rightarrow d_i = a_i - l_i$

$\sum l_i (a_i - l_i) = \sum l_i a_i - \sum l_i^2$

$\sum \frac{a_i}{n} - \frac{\sum \bar{x}(x_i - \bar{x}) a_i}{\sum (x_i - \bar{x})^2} - \frac{1}{n} - \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}$

$= \frac{1}{n} \sum a_i - \frac{\bar{x} \sum a_i x_i - \bar{x}^2 \sum a_i}{\sum (x_i - \bar{x})^2} - \frac{1}{n} - \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}$

$$= \frac{1}{n}(1) - \frac{\bar{x}(0) - \bar{x}^2(1)}{\sum(x_i - \bar{x})^2} - \frac{1}{n} - \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}$$

$$= \frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2} - \frac{1}{n} - \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}$$

$$= 0$$ Yay!

$$var(b_0) = \sigma^2 \sum l_i^2 + \sigma^2 \sum d_i^2 + 2\sigma^2 \sum l_i d_i$$

$$= var(\hat{\beta}_0) + \sigma^2 \sum d_i^2 + 0$$

$$= var(\hat{\beta}_0) + \sigma^2 \sum d_i^2$$

must be positive

Therefore: $var(b_0) \geq var(\hat{\beta}_0)$ therefore $\hat{\beta}_0$ is BLUE

Exercise 5

show $\underbrace{\sum (y_i - b_0 - b_1 x_i)^2} \geq \sum e_i^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

$$\hookrightarrow \sum (y_i - b_0 - b_1 x_i)^2 = \sum\Big([y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i] + [\hat{\beta}_0 + \hat{\beta}_1 x_i] - [b_0 + b_1 x_i]\Big)^2$$

$$= \sum\Big([y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i] + [\beta_0 + \beta_1 x_i - b_0 + b_1 x_i]\Big)^2$$

$$= \sum\Big([y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i]^2 + 2[y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i][\beta_0 + \beta_1 x_i - b_0 + b_1 x_i] +$$

$$+ [\beta_0 + \beta_1 x_i - b_0 + b_1 x_i]^2$$

$$\boxed{\text{We know: } e_i = y_i - \hat{\beta}_0 - \beta_1 x_i}$$

$$= \sum e_i^2 + 2\sum e_i[\beta_0 + \beta_1 x_i - b_0 + b_1 x_i] + \sum[\hat{\beta}_0 + \beta_1 x_i - b_0 - b_1 x_i]^2$$

$$\boxed{\text{Let: } \alpha_i = \hat{\beta}_0 + \hat{\beta}_1 x_i - b_0 - b_1 x_i}$$

$$= \sum e_i^2 + 2\sum e_i \beta_0 + 2\sum e_i \beta_1 x_i - 2\sum e_i b_0 - 2\sum e_i b_1 x_i + \sum \alpha_i^2$$

$$= \sum e_i^2 + 2\beta_0 \sum e_i + 2\beta_1 \sum e_i x_i - 2b_0 \sum e_i - 2b_1 \sum e_i x_i + \sum \alpha_i^2$$

using the properties we showed in exercise 2 we can simplify:

$\sum e_i = 0 \qquad \sum e_i x_i = 0$

$$= \sum e_i^2 + 0 + 0 - 0 + 0 + \sum \alpha_i^2$$

$$= \sum e_i^2 + \sum \alpha_i^2$$

We know that $\sum \alpha_i^2 \geq 0$ due to the fact α_i is squared

Therefore:

$$\sum e_i^2 + \sum \alpha_i^2 = \boxed{\sum (y_i - b_0 - b_1 x_i)^2 \geq \sum e_i^2} = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$