HW 2

Madeline Blasingame
205412002
10/10/2021

Exercise 1

a. $\hat{\sigma}^2 = \dfrac{\sum (y_i - \hat{\beta}_1 x_i)^2}{n} = \dfrac{\sum (e_i)^2}{n}$

$var(a) = E(a^2) - E(a)^2$
$E(a^2) = var(a) + E(a)^2$

$$E(\hat{\sigma}^2) = E\left[\frac{\sum (e_i)^2}{n}\right] = \frac{1}{n} E\left[\sum e_i^2\right]$$

$$= \frac{1}{n} \sum E(e_i^2) = \frac{1}{n} \sum \left[var(e_i) + \underbrace{E(e_i)^2}_{e_i \sim N(0,\sigma^2)}\right] \quad \therefore E(e_i) = 0$$

Note 1:

$$E(e_i) = E(Y_i - \hat{Y}) = E[Y_i] - E[x_i \hat{\beta}_1] = x_i \beta_1 - x_i E[\hat{\beta}_1] = x_i \beta_1 - x_i \beta_1 = 0$$

Note 2:

$$var(e_i) = var(Y_i - \hat{Y}) = var(Y_i) + var(x_i \hat{\beta}_1) - 2cov(Y_i, x_i \hat{\beta}_1)$$
$$= \sigma^2 + x_i^2 var(\hat{\beta}_1) - 2cov(Y_i, x_i \hat{\beta}_1)$$
$$= \sigma^2 + \frac{x_i^2 \sigma^2}{\sum x_i^2} - 2cov(Y_i, x_i \hat{\beta}_1)$$

Note 3:

$$cov(Y_i, \hat{Y}_i) = cov(Y_i, x_i \hat{\beta}_1) = x_i cov(Y_i, \hat{\beta}_1)$$

we know: $\hat{\beta}_1 = \sum \dfrac{x_i}{\sum x_j^2} y_i$ let $a_i = \dfrac{x_i}{\sum x_j^2}$

$$cov(y_i, \hat{\beta}_1) = cov(y_i, \sum a_i y_i) = \underbrace{cov(y_i, a_1 y_1)}_{0} + \dots + cov(y_i, a_i y_i) + \dots + \underbrace{cov(y_i, a_n y_n)}_{0}$$

we know that $y_i \perp\!\!\!\perp y_j$ when $i \neq j$ so nearly all the covariances are 0 and we are left with:

$$= cov(y_i, a_i y_i) = a_i cov(y_i, y_i) = a_i \sigma^2 = \frac{x_i}{\sum x_j^2} \sigma^2$$

therefore: $cov(Y_i, \hat{Y}_i) = \dfrac{x_i^2 \sigma^2}{\sum x_j^2}$

$$E(\sigma^2) = \frac{1}{n}\sum\left[\mathrm{var}(e_i) + E(e_i)^2\right]$$

$$= \frac{1}{n}\sum\left[\sigma^2 + \frac{x_i^2\,\sigma^2}{\sum x_i^2} - \frac{2x_i^2\,\sigma^2}{\sum x_j^2}\right) + 0^2\right]$$

$$= \frac{1}{n}\sum\left[\sigma^2 - \frac{x_i^2\,\sigma^2}{\sum x_i^2}\right] = \frac{n\sigma^2}{n} - \frac{\sum x_i^2\,\sigma^2}{n\sum x_i^2} = \sigma^2 - \frac{\sigma^2\sum x_i^2}{n\sum x_i^2}$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{(n-1)\sigma^2}{n} \quad \square$$

$$S_e^2 = \frac{\sum e_i^2}{n-1} \qquad E(S_e^2) = E\left[\frac{\sum e_i^2}{n-1}\right] = \frac{1}{n-1}E\left[\sum e_i^2\right]$$

As shown on the previous page:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n} \longrightarrow n\hat{\sigma}^2 = \sum e_i^2 \longrightarrow nE(\hat{\sigma}^2) = E\left[\sum e_i^2\right]$$

$$E\left[\sum e_i^2\right] = n\,\frac{(n-1)\sigma^2}{n} = (n-1)\sigma^2$$

$$E(S_e^2) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

Due to the fact that $E(S_e^2) = \sigma^2$ S_e^2 is an unbiased estimator for σ^2. $\square$

b. Show that $\hat{\beta}_1$ is BLUE

$b_1 = \sum c_i y_i$ such that $E(b_1) = \beta_1$

We want to show:

$$\mathrm{var}(\hat{\beta}_1) \leq \mathrm{var}(b_1)$$

Note: $E(b_1) = E(\sum c_i y_i) = \beta_1$

$$= \sum c_i E(y_i)$$
$$= \sum c_i \beta_1 x_i$$
$$= \beta_1 \sum c_i x_i \overset{\text{assume}}{=} \beta_1 \Rightarrow \sum c_i x_i = 1$$

$$\underbrace{var(\hat{\beta}_1) \le var(b_1)}$$
$$\hookrightarrow var(\hat{\beta}_1) + k \quad \text{where} \quad k \ge 0$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i y_i \quad \text{where} \quad k_i = \frac{x_i}{\sum x_i^2}$$

$$var(\hat{\beta}_1) = var(\sum k_i y_i) = \sum var(k_i y_i) \sum k_i^2 var(y_i) = \sigma^2 \sum k_i^2$$

Define d_i such that $d_i = c_i - k_i \Rightarrow c_i = d_i + k_i$

$$var(b_1) = \sum c_i^2 var(y_i) = \sigma^2 \sum c_i^2 = \sigma^2 \sum (d_i + k_i)^2$$

$$= \sigma^2 \left[\sum k_i^2\right] + \sigma^2 \left[\sum d_i^2\right] + 2\sigma^2 \left[\sum k_i d_i\right]$$

$$= var(\hat{\beta}_1) + \underbrace{\sigma^2 \left[\sum d_i^2\right] + 2\sigma^2 \left[\sum k_i d_i\right]}$$

in order for $\hat{\beta}_1$ to be BLUE this must be ≥ 0

we know $\sigma^2 \sum d_i^2 \ge 0$ because $\sigma^2 \ge 0$ and $d_i^2 \ge 0$, so now we will try to show $2\sigma^2 \sum k_i d_i = 0$

$$\sum k_i d_i = \sum k_i (c_i - k_i) \qquad \text{because } d_i = c_i - k_i$$
$$= \sum k_i c_i - k_i^2$$
$$= \sum k_i c_i - \sum k_i^2$$

Note: $\sum k_i^2 = \sum \left(\frac{x_i}{\sum x_i^2}\right)^2 = \frac{\sum x_i^2}{(\sum x_i^2)^2} = \frac{1}{\sum x_i^2}$

$$= \frac{\sum x_i c_i}{\sum x_i^2} - \frac{1}{\sum x_i^2}$$

Note: $\sum x_i c_i = 1$

$$= \frac{1}{\sum x_i^2} - \frac{1}{\sum x_i^2} = 0$$

Therefore: $var(b_1) = var(\hat{\beta}_1) + 0 + \sigma^2 \sum d_i^2$ where $\sigma^2 \sum d_i^2 \geq 0$

Therefore: $var(b_1) \geq var(\hat{\beta}_1) \Rightarrow \hat{\beta}_1$ is BLUE $\square$

c.

1. we want to find the distribution for $\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$

we know: $y_i \sim N(\beta_1 x_i, \sigma)$ because $\varepsilon_i \sim (0, \sigma)$

we also know that $\hat{\beta}_1$ can be written as a linear combination of $y_i \rightarrow \hat{\beta}_1 = \sum a_i y_i$ where $a_i = \frac{x_i}{\sum x_i^2}$

we also know that $y_i \perp\!\!\!\perp y_j$ when $i \neq j$ meaning that $\hat{\beta}_1$ is a linear combination of independent random variables and therefore also follows a normal distribution.

we also know:

$E(\hat{\beta}_1) = \beta_1$ and $var(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2}$

Therefore: $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum x_i^2}\right)$ $\square$

2. If $Z \sim N(0,1)$ we know that $Z^2 \sim \chi^2_1$

$\hat{\beta}_1$ follows a normal distribution, but we need to standardize it:

$$Z = \frac{(\hat{\beta}_1 - \beta_1)}{\sqrt{\sigma^2 / \sum x_i^2}} \sim N(0,1)$$

$$Z^2 = \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2 / \sum x_i^2} \sim \chi^2_1 \ \square$$

The mgf for $Z^2 \sim \chi^2_1$ is:

$$M_{Z^2}(t) = (1-2t)^{-1/2} \ \square$$

3. we know: $\epsilon_i \sim N(0, \sigma^2)$

therefore: $\frac{\epsilon_i}{\sigma} \sim N(0,1) \rightarrow \frac{\epsilon_i^2}{\sigma^2} \sim \chi_1^2$

Because $\frac{\epsilon_i^2}{\sigma^2} \sim \chi_1^2$: $M_{\epsilon_i^2/\sigma^2}(t) = (1-2t)^{-1/2}$ is the MGF for $\frac{\epsilon_i^2}{\sigma^2}$

$\epsilon_i \perp\!\!\!\perp \epsilon_j$ when $i \neq j \Rightarrow \frac{\epsilon_i^2}{\sigma^2} \perp\!\!\!\perp \frac{\epsilon_j^2}{\sigma^2}$ when $i \neq j$

This means that the MGF for $\sum_{i=1}^{n} \frac{\epsilon_i^2}{\sigma^2}$ will be equivalent to the MGF for χ_n^2 because we are summing n independent χ_1^2 random variables.

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

This means both $\frac{\sum \epsilon_i^2}{\sigma^2}$ and $\frac{\sum (y_i - \beta_1 x_i)^2}{\sigma^2}$ follow a χ_n^2 distribution and share the same MGF:

$$M_{\frac{\sum \epsilon_i^2}{\sigma^2}}(t) = M_{\frac{\sum (y_i - \beta_1 x_i)^2}{\sigma^2}}(t) = (1-2t)^{-n/2} \quad \square$$

4. $$\frac{1}{\sigma^2} \sum \epsilon_i^2 = \frac{1}{\sigma^2} \sum (y_i - \beta_1 x_i)^2 = \frac{1}{\sigma^2} \sum [\overset{e_i}{(y_i - \hat{\beta}_1 x_i)} + (\hat{\beta}_1 x_i - \beta_1 x_i)]^2$$

$$= \frac{1}{\sigma^2} \sum [e_i + (\hat{\beta}_1 x_i - \beta_1) x_i]^2 = \frac{1}{\sigma^2} \left[\sum e_i^2 + \sum x_i^2 (\hat{\beta}_1 - \beta_1)^2 + 2 \sum e_i x_i (\hat{\beta}_1 - \beta_1)\right]$$

we know: $\sum e_i x_i (\hat{\beta}_1 - \beta_1) = (\hat{\beta}_1 - \beta_1) \sum e_i x_i = 0$

Therefore:

$$\frac{1}{\sigma^2} \sum \epsilon_i^2 = \frac{1}{\sigma^2} \left[\sum e_i^2 + \sum x_i^2 (\hat{\beta}_1 - \beta_1)^2\right] = \sum \frac{e_i^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2 / \sum x_i^2}$$

Let $\sum \frac{e_i^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2 / \sum x_i^2} = Q$ with $Q_1 = \sum \frac{e_i^2}{\sigma^2}$

and $Q_2 = \sum \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2 / \sum x_i^2}$

Determine if Q_1 and Q_2 are independent

$$\text{cov}(e_i, \hat{\beta}_1) = \text{cov}(y_i - \hat{\beta}_1 x_i, \hat{\beta}_1) = \text{cov}(y_i, \hat{\beta}_1) - x_i \text{cov}(\hat{\beta}_1, \hat{\beta}_1)$$
$$= \text{cov}(y_i, \hat{\beta}_1) - x_i \text{var}(\hat{\beta}_1)$$

Note: $\text{cov}(y_i, \hat{\beta}_1) = \frac{\sigma^2 x_i}{\sum x_i^2}$ $\quad \text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2}$

$$= \frac{\sigma^2 x_i}{\sum x_i^2} - \frac{x_i \sigma^2}{\sum x_i^2} = 0$$

Note: $e_i = y_i - \hat{\beta}_1 x_i \sim N(0, \text{Var}(e_i))$
$y_i \sim N(\beta_1 x_i, \sigma^2)$
$\hat{\beta}_1 x_i \sim N(\beta_1 x_i, \frac{x_i^2 \sigma^2}{\sum x_i^2})$

Due to the fact that both e_i and $\hat{\beta}_1$ are normally distributed and their covariance is 0, we know that e_i and $\hat{\beta}_1$ must be independent.

Using this fact we know that e_i is independent of Q_2 since Q_2 is a function of $\hat{\beta}_1$. Additionally, Q_1 must be independent of Q_2 since Q_1 is a function of $e_1 \dots e_n$

We know (from part 2): $Q_2 = \sum \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2 / \sum x_i^2} \sim \chi_1^2 \Rightarrow M_{Q_2}(t) = (1-2t)^{-1/2}$

We know (from part 3): $Q_1 + Q_2 = \frac{\sum (y_i - \beta_1 x_i)^2}{\sigma^2} \sim \chi_n^2 \Rightarrow M_{Q_1+Q_2}(t) = (1-2t)^{-n/2}$

We need to find the MGF for Q_1

Due to the fact that Q_1 and Q_2 are independent, we can use the property: $M_{Q_1+Q_2}(t) = M_{Q_1}(t)\, M_{Q_2}(t)$

$$\text{or} \quad M_{Q_1}(t) = \frac{M_{Q_1+Q_2}(t)}{M_{Q_2}(t)} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}}$$

$$= (1-2t)^{-\frac{n-1}{2}}$$

Since MGFs are unique, this means that $Q_1 \sim \chi^2_{n-1}$

Reminder: $Q_1 = \Sigma \frac{e_i^2}{\sigma^2}$

Note: $S_e^2 = \frac{\Sigma e_i^2}{n-1}$ so $\frac{(n-1)S_e^2}{\sigma^2} = \frac{(n-1)\Sigma e_i^2}{(n-1)\sigma^2} = \frac{\Sigma e_i^2}{\sigma^2}$

Therefore $Q_1 = \frac{\Sigma e_i^2}{\sigma^2} = \frac{(n-1)S_e^2}{\sigma^2}$ meaning that $\frac{\Sigma e_i^2}{\sigma^2}$ and $\frac{(n-1)S_e^2}{\sigma^2}$ must follow the same distribution, implying:

$$\frac{(n-1)S_e^2}{\sigma^2} \sim \chi^2_{n-1} \quad \square$$

5. In part 4 we found $Q_1 = \frac{(n-1)S_e^2}{\sigma^2}$ with $Q_1 \sim \chi^2_{n-1}$

therefore: $\frac{Q_1 \sigma^2}{(n-1)} = S_e^2$

Using the property: $M_{cx}(t) = M_x(ct)$ we know:

$$M_{\frac{\sigma^2}{(n-1)}Q_1}(t) = M_{S_e^2}(t) = M_{Q_1}\left(\frac{\sigma^2}{n-1}t\right)$$

$$M_{Q_1}(t) = (1-2t)^{-\frac{n-1}{2}} \quad \text{so} \quad M_{Q_1}\left(\frac{\sigma^2}{n-1}t\right) = \left(1 - \frac{2\sigma^2}{n-1}t\right)^{-\frac{n-1}{2}} = M_{S_e^2}(t)$$

Also: $X \sim \Gamma(\alpha, \beta) \longrightarrow M_X(t) = (1 - \beta t)^{-\alpha}$

Due to the fact that every MGF is unique:

$$M_{S_e^2}(t) = \left(1 - \frac{2\sigma^2}{n-1} t\right)^{-\frac{n-1}{2}} \implies S_e^2 \sim \Gamma\left(\frac{n-1}{2}, \frac{2\sigma^2}{n-1}\right) \square$$

Exercise 2

1. $E\left[\frac{\sum Y_i}{\sum X_i}\right] = \frac{1}{\sum X_i} E\left[\sum Y_i\right] = \frac{\sum E[Y_i]}{\sum X_i} = \frac{\sum \beta_1 X_i}{\sum X_i} = \frac{\beta_1 \sum X_i}{\sum X_i} = \beta_1$

therefore $\frac{\sum Y_i}{\sum X_i}$ is an unbiased estimator for β_1 $\square$

$y_i = \overset{\nearrow \text{constant}}{\beta_1 X_i} + \overset{\nearrow \text{random}}{\epsilon_i}$

2. $var(\hat{\beta}_1) = \frac{\sigma^2}{\sum X_i^2}$ $\quad var\left(\frac{\sum y_i}{\sum X_i}\right) = \frac{1}{(\sum X_i)^2} var(\sum y_i) = \frac{1}{(\sum X_i)^2} \sum var(y_i) = \frac{n\sigma^2}{(\sum X_i)^2}$

Cauchy-Schwartz: $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots b_n \in \mathbb{R}$

$$\left(\sum a_i b_i\right)^2 \leq \left(\sum a_i^2\right)\left(\sum b_i^2\right)$$

Let $a_i = x_i$ for each $i = 1, 2, 3 \ldots, n$
$b_i = 1$

$$\left(\sum x_i\right)^2 \leq \left(\sum x_i^2\right)\left(\overset{=n}{\sum 1}\right)$$

$$\frac{1}{\sum x_i^2} \leq \frac{n}{(\sum x_i)^2}$$

$$\frac{\sigma^2}{\sum x_i^2} \leq \frac{n\sigma^2}{(\sum x_i)^2}$$

therefore $var(\hat{\beta}_1) \leq var\left(\frac{\sum y_i}{\sum X_i}\right)$ $\square$

3. $E\left[\frac{\sum \frac{Y_i}{x_i}}{n}\right] = \frac{1}{n} E\left[\sum \frac{Y_i}{x_i}\right] = \frac{1}{n} \sum E\left[\frac{Y_i}{x_i}\right] = \frac{1}{n} \sum \frac{1}{x_i} E[Y_i] =$

$\frac{1}{n} \sum \frac{\beta_1 X_i}{X_i} = \frac{1}{n} n\beta_1 = \beta_1$

therefore, $\frac{\sum \frac{Y_i}{x_i}}{n}$ is an unbiased estimator for β_1 $\square$

4. $var(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum x_1^2}$

$$var\left(\frac{\sum \frac{Y_i}{x_i}}{n}\right) = \frac{1}{n^2} var\left(\sum \frac{Y_i}{x_i}\right) = \frac{1}{n^2} \sum var\left(\frac{Y_i}{x_i}\right) = \frac{1}{n^2} \sum \frac{1}{x_i^2} var(Y_i)$$

$$= \frac{1}{n^2} \sum \frac{\sigma^2}{x_i^2} = \frac{\sigma^2}{n^2} \sum \frac{1}{x_i^2}$$

cauchy-schwartz: $a_1, a_2, \dots, a_n, b_1, b_2, \dots b_n \in \mathbb{R}$

$$\left(\sum a_i b_i\right)^2 \leq \left(\sum a_i^2\right)\left(\sum b_i^2\right)$$

Let $a_i = x_i$ for each $i = 1, 2, 3 \dots, n$
$b_i = x_i^{-1}$

$$\left(\sum \frac{x_i}{x_i}\right)^2 \leq \left(\sum x_i^2\right)\left(\sum x_i^{-2}\right)$$

$$\left(\sum 1\right)^2 \leq \left(\sum x_i^2\right)\left(\sum x_i^{-2}\right)$$

$$n^2 \leq \left(\sum x_i^2\right)\left(\sum x_i^{-2}\right)$$

$$\frac{1}{\sum x_i^2} \leq \frac{1}{n^2} \sum \frac{1}{x_i^2}$$

$$\frac{\sigma^2}{\sum x_i^2} \leq \frac{\sigma^2}{n^2} \sum \frac{1}{x_i^2}$$

therefore, $var(\hat{\beta}_1) \leq var\left(\frac{\sum \frac{Y_i}{x_i}}{n}\right)$ □

Exercise 3

a. Express $e_i = \sum q_\ell Y_\ell$ and $e_j = p_r Y_r$

We know:

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

$$\hat{\beta}_1 = \frac{\sum (x_\ell - \bar{x}) y_\ell}{\sum (x_k - \bar{x})^2}$$

Therefore:

$$e_i = \sum a_\ell y_\ell - \frac{1}{n} \sum y_\ell - \frac{(x_i - \bar{x})}{\sum (x_k - \bar{x})^2} \sum (x_\ell - \bar{x}) y_\ell$$

where: $a_\ell = \begin{cases} 1 & \text{when } \ell = i \\ 0 & \text{otherwise} \end{cases}$

$$= \sum \underbrace{\left[a_\ell - \frac{1}{n} - \frac{(x_i - \bar{x})(x_\ell - \bar{x})}{\sum (x_k - \bar{x})^2} \right]}_{= q_\ell} y_\ell$$

Similarly:

$$e_j = \sum b_r y_r - \frac{1}{n} \sum y_r - \frac{(x_j - \bar{x})}{\sum (x_r - \bar{x})^2} \sum (x_r - \bar{x}) y_r$$

where: $b_r = \begin{cases} 1 & \text{when } r = j \\ 0 & \text{otherwise} \end{cases}$

$$= \sum \underbrace{\left[b_r - \frac{1}{n} - \frac{(x_j - \bar{x})(x_r - \bar{x})}{\sum (x_k - \bar{x})^2} \right]}_{= p_r} y_\ell$$

$$\text{cov}(e_i, e_j) = \text{cov}\left(\sum q_\ell Y_\ell, \sum p_r Y_r\right) = \sum_\ell \sum_r q_\ell p_r \,\text{cov}(Y_\ell, Y_r)$$

Note: $\text{cov}(y_\ell, y_r) = 0$ when $\ell \neq r$ since $y_\ell \perp\!\!\!\perp y_r$

$$= \sum_{\ell, r : \ell = r} q_\ell p_r \,\text{cov}(y_\ell, y_r)$$

$$= \sum_\ell q_\ell p_\ell \,\text{cov}(y_\ell, y_\ell)$$

$$= \sum_\ell q_\ell p_\ell \,\text{var}(y_\ell)$$

$$= \sigma^2 \sum_\ell q_\ell p_\ell \quad \square$$

b. $cov(e_i, e_j) = cov(Y_i - \hat{Y}_i, Y_j - \hat{Y}_j)$

$$= cov(Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x}), Y_j - \bar{Y} - \hat{\beta}_1(x_j - \bar{x}))$$

covariance property:

$$cov(A+B+C, D+E+F) = cov(A, D+E+F) + cov(B, D+E+F) + cov(C, D+E+F)$$
$$= cov(A,D) + cov(A,E) + cov(A,F) + cov(B,D) + cov(B,E) + cov(B,F) + cov(C,D) + cov(C,E) + cov(C,F)$$

$$cov(\underbrace{Y_i}_{A} - \underbrace{\bar{Y}}_{B} - \underbrace{\hat{\beta}_1(x_i - \bar{x})}_{C}, \underbrace{Y_j}_{D} - \underbrace{\bar{Y}}_{E} - \underbrace{\hat{\beta}_1(x_j - \bar{x})}_{F})$$

$$= cov(Y_i, Y_j) - cov(Y_i, \bar{Y}) - cov(Y_i, \hat{\beta}_1(x_j - \bar{x})) - cov(\bar{Y}, Y_j) + cov(\bar{Y}, \bar{Y}) + cov(\bar{Y}, \hat{\beta}_1(x_j - \bar{x}))$$
$$- cov(\hat{\beta}_1(x_i - \bar{x}), Y_j) + cov(\hat{\beta}_1(x_i - \bar{x}), \bar{Y}) + cov(\hat{\beta}_1(x_i - \bar{x}), \hat{\beta}_1(x_j - \bar{x}))$$

Note: $cov(Y_i, Y_j) = 0$ because $Y_i \perp\!\!\!\perp Y_j$

Note: $cov(Y_i, \bar{Y}) = cov(\bar{Y}, Y_j) = \dfrac{\sigma^2}{n}$

Note: $cov(Y_i, \hat{\beta}_1(x_j - \bar{x})) = (x_j - \bar{x})\, cov(Y_i, \hat{\beta}_1) = \dfrac{\sigma^2 (x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2}$

Note: $cov(\hat{\beta}_1(x_i - \bar{x}), Y_j) = (x_i - \bar{x})\, cov(Y_j, \hat{\beta}_1) = \dfrac{\sigma^2 (x_j - \bar{x})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$

Note: $cov(\bar{Y}, \hat{\beta}_1(x_j - \bar{x})) = (x_j - \bar{x})\, cov(\bar{Y}, \hat{\beta}_1) = 0$

Note: $cov(\hat{\beta}_1(x_i - \bar{x}), \bar{Y}) = (x_i - \bar{x})\, cov(\hat{\beta}_1, \bar{Y}) = 0$

Note: $cov(\bar{Y}, \bar{Y}) = var(\bar{Y}) = \dfrac{\sigma^2}{n}$

Note: $cov(\hat{\beta}_1(x_i - \bar{x}), \hat{\beta}_1(x_j - \bar{x})) = (x_i - \bar{x})(x_j - \bar{x})\, cov(\hat{\beta}_1, \hat{\beta}_1)$

$$= (x_i - \bar{x})(x_j - \bar{x})\, var(\hat{\beta}_1) = (x_i - \bar{x})(x_j - \bar{x}) \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

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

$$+ \frac{\sigma^2 (x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2} = -\frac{\sigma^2}{n} - \frac{\sigma^2 (x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2} = -\sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2} \right]$$

$\square$

Exercise 4

a. $$\text{cov}(X,Y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{(n-1)} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)}$$

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1+n_2} = \frac{(5)(30.2)+(6)(52.17)}{11} = 42.1836$$

$$\bar{Y} = \frac{n_1\bar{Y}_1 + n_2\bar{Y}_2}{n_1+n_2} = \frac{(5)(39.0)+(6)(63.67)}{11} = 52.4563$$

$$\text{cov}(X_1, Y_1) = \frac{\sum X_{i1}Y_{i1} - n\overline{X_1Y_1}}{n-1} \rightarrow 88.5 = \frac{\sum X_{i1}Y_{i1} - (5)(30.2)(39.0)}{4}$$

$$\sum X_{i1}Y_{i1} = 6243$$

$$\text{cov}(X_2, Y_2) = \frac{\sum X_{i2}Y_{i2} - n\overline{X_2Y_2}}{n-1} \rightarrow 148.67 = \frac{\sum X_{i2}Y_{i2} - (6)(52.17)(63.67)}{5}$$

$$\sum X_{i2}Y_{i2} = 20673.3334$$

$$\sum X_iY_i = 6243 + 20673.3334 = 26916.3334$$

$$\text{cov}(X,Y) = \frac{26916.3334 - 11(42.1836)(52.4563)}{10} = 257.5582 \; \square$$

b. New values excluding the final pair:

$$\sum_{i=1}^{5} X_i = \sum_{i=1}^{6} X_i - X_6 = 9.545 - 1.565 = 7.98$$

$$\sum_{i=1}^{5} Y_i = \sum_{i=1}^{6} Y_i - Y_6 = 61.668 - 3.508 = 58.16$$

$$\sum_{i=1}^{5} X_i^2 = \sum_{i=1}^{6} X_i^2 - X_6^2 = 15.78468 - (1.565)^2 = 13.335$$

$$\sum_{i=1}^{5} Y_i^2 = \sum_{i=1}^{6} Y_i^2 - Y_6^2 = 719.9573 - (3.508)^2 = 707.651$$

$$\sum_{i=1}^{5} X_i y_i = \sum_{i=1}^{6} X_i y_i - X_6Y_6 = 96.43722 - (1.565)(3.508) = 90.9472$$

$$\bar{X} = \frac{1}{5}\sum_{i=1}^{5} X_i = \frac{7.98}{5} = 1.596 \qquad \bar{Y} = \frac{1}{5}\sum_{i=1}^{5} Y_i = \frac{58.16}{5} = 11.632$$

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{90.9472 - (5)(1.596)(11.632)}{13.335 - \frac{(7.98)^2}{5}} = -3.1325$$

$$\hat{\beta}_0 = 11.632 - (-3.1325)(1.596) = 16.6315$$

c. $$R^2 = \frac{SSR}{SST} = \frac{\hat{\beta}_1^2 \sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2} = \frac{\hat{\beta}_1^2 \left[\sum X_i^2 - 2\bar{X}\sum X_i + n\bar{X}^2\right]}{\sum Y_i^2 - 2\bar{Y}\sum Y_i + n\bar{Y}^2}$$

$$R^2 = \frac{(-3.1325)^2\left[13.335 - 2(1.596)(7.98) + (5)(1.596)^2\right]}{(707.651 - 2(11.632)(58.16) + (5)(11.632)^2} = \frac{5.8769}{31.1338}$$

$$= 0.1887$$

Exercise 5

a. show $\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2$

we know: $e_i = y_i - \hat{y}_i \rightarrow y_i = e_i + \hat{y}_i$

$$\sum y_i^2 = \sum (e_i + \hat{y}_i)^2 = \sum e_i^2 + 2e_i\hat{y}_i + \hat{y}_i^2$$

$$= \sum e_i^2 + 2\sum e_i\hat{y}_i + \sum \hat{y}_i^2$$

we know (from HW1): $\sum e_i \hat{y}_i = 0$

$$= \sum e_i^2 + \sum \hat{y}_i^2 \quad \square$$

b. Refer to Exercise 3 part b

c. Find $\text{cov}(e_i, \hat{y}_j)$ $\qquad e_j = Y_j - \hat{Y}_j$

$\text{cov}(e_i, e_j) = \text{cov}(e_i, Y_j) - \text{cov}(e_i, \hat{Y}_j)$

$\hookrightarrow \text{cov}(e_i, \hat{y}_j) = \text{cov}(e_i, Y_j) - \text{cov}(e_i, e_j)$

Note: $\text{cov}(e_i, Y_j) = \text{cov}(y_i - \bar{Y} - \hat{\beta}_1 (x_i - \bar{x}), y_j)$

$$= \text{cov}(y_i, y_j) - \text{cov}(\bar{Y}, y_j) - (x_i - \bar{x})\,\text{cov}(y_i, \hat{\beta}_1)$$

Note: $\text{cov}(y_i, y_j) = 0$

Note: $\text{cov}(\bar{Y}, y_j) = \frac{\sigma^2}{n}$

Note: $(x_i - \bar{x})\,\text{cov}(y_j, \hat{\beta}_1) = \frac{\sigma^2 (x_j - \bar{x})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$

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

Note: $\text{cov}(e_i, e_j) = -\frac{\sigma^2}{n} - \frac{\sigma^2 (x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2}$

$$\operatorname{cov}(e_i, \hat{y}_j) = \operatorname{cov}(e_i, Y_j) - \operatorname{cov}(e_i, e_j)$$

$$= \left[-\frac{\sigma^2}{n} - \frac{\sigma^2 (x_i - \bar{x})(x_j - \bar{x})}{\Sigma (x_i - \bar{x})^2} \right] - \left[-\frac{\sigma^2}{n} - \frac{\sigma^2 (x_j - \bar{x})(x_i - \bar{x})}{\Sigma (x_i - \bar{x})^2} \right]$$

$$= \frac{-\sigma^2}{n} - \frac{\sigma^2 (x_i - \bar{x})(x_j - \bar{x})}{\Sigma (x_i - \bar{x})^2} + \frac{\sigma^2}{n} + \frac{\sigma^2 (x_j - \bar{x})(x_i - \bar{x})}{\Sigma (x_i - \bar{x})^2} = 0 \; \square$$

d. Show that: $\Sigma (y_i - \bar{y})(\hat{y}_i - \bar{y}) = \Sigma (\hat{y}_i - \bar{y})^2$

We know: $e_i = Y_i - \hat{Y}_i$

$$\Sigma (y_i - \bar{y})(\hat{y}_i - \bar{y}) = \Sigma (e_i + \hat{Y}_i - \bar{y})(\hat{y}_i - \bar{y}) = \Sigma e_i \hat{y}_i - e_i \bar{y} + \hat{y}_i^2 - 2\hat{y}_i \bar{y} + \bar{y}^2$$

$$= \Sigma e_i \hat{y}_i - \bar{y} \Sigma e_i + \Sigma \hat{y}_i^2 - 2\bar{y} \Sigma \hat{y}_i + \Sigma \bar{y}^2$$

$$= 0 - 0 + \Sigma \hat{y}_i^2 - 2\bar{y} \Sigma \hat{y}_i + \Sigma \bar{y}^2$$

$$= \Sigma \hat{y}_i^2 - 2\bar{y} \Sigma \hat{y}_i + \Sigma \bar{y}^2 =$$

$$\Sigma \hat{y}_i^2 - 2\bar{y} \hat{y}_i + \bar{y}^2 =$$

$$\Sigma (\hat{y}_i - \bar{y})^2 \; \square$$