Ankit Jain

Linear Regression HW 1

1.) False, the distribution of Y being normal is not included in the Gauss-Markov assumptions of Best Linear Unbiased Estimator.

2.) $$\ell(\beta_0, \beta_1, \sigma^2) = \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

$$= \log(1) - \log\left((2\pi\sigma^2)^{n/2}\right) - \text{''}$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \text{''}$$

$$\frac{\partial}{\partial\beta_0} = \frac{-1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)(-1) = 0$$

$$\boxed{= \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)} \checkmark$$

$$\frac{\partial}{\partial\beta_1} = \frac{-1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)\, X_i = 0$$

$$\boxed{= \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)X_i} \checkmark$$

$$\frac{\partial}{\partial\sigma^2} = \frac{-2\pi n}{4\pi\sigma^2} - \left(-\frac{1}{(\sigma^2)^2}\cdot\frac{1}{2}\cdot\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2\right)$$

$$0 = \frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\frac{n}{2\sigma^2} = \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\boxed{\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2}$$

3.) $$\beta_1 = \frac{\sum (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}\bar{y})}{\sum (x_i^2 - \bar{x} x_i - \bar{x} x_i + (\bar{x})^2)}$$

$$= \frac{\sum x_i y_i - \bar{x}\sum y_i - \bar{y}\sum x_i + \sum \bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x}\sum x_i + \sum \bar{x}^2}$$

$$= \frac{\sum x_i y_i - \bar{x}\sum y_i - \bar{y}\sum x_i + n\bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2}$$

$$= \frac{\sum x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x}\cdot\bar{x}n + n\bar{x}^2}$$

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

4.) $y_i = B_0 + B_1 x_i + e_i \quad \hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i$

$$E(e_i) = B_0 + B_1 x_i + e_i - (\hat{B}_0 + \hat{B}_1 x_i)$$

$$= B_0 + B_1 x_i + e_i - \hat{B}_0 - \hat{B}_1 x_i$$

$$E(e_i) = E\left[(B_0 - \hat{B}_0) + (B_1 x_i - \hat{B}_1 x_i) + \varepsilon_i\right]$$

$$= E(B_0 - \hat{B}_0) + E(B_1 x_i - \hat{B}_1 x_i) + \underbrace{E(\varepsilon_i)}_{0}$$

$$E(e_i) = \underbrace{E(B_0 - \hat{B}_0)}_{0} + \underbrace{E(B_1 - \hat{B}_1)x_i}_{0}$$

The expectations of $(B_0 - \hat{B}_0)$ and $(B_1 - \hat{B}_1)x_i$ are both zero because the residuals are balanced above and below the line.

So, $E(e_i) = 0$ ✓

5a.) No, because ε_i should not be included. The expectation of Y_i is $E(\beta_0) + E(\beta_1 x_i) + E(\varepsilon_i)$

Since $E(\varepsilon_i) = 0$, that term should be left out.

b.) $\beta_0 = 2$, $\beta_1 = 4$, $\sigma^2 = 9$

For $x = 1$, $E(y_i) = 2 + 4(1) = 6$
$Y \sim N(6, 9)$

For $x = 2$,
$E(y_i) = 2 + 4(2) = 10$
$\boxed{Y \sim N(10, 9)}$

For $x = 4$,
$E(y_i) = 2 + 4(4) = 18$
$\boxed{Y \sim N(18, 9)}$

I found the mean for each dist. by plugging each x val into $y_i = \beta_0 + \beta_1 x_i$ where $\beta_0 = 2$ and $\beta_1 = 4$. This is the mean because it is the average value of Y for each x with an error of 0 since the distances balance out.

6 a.) The slope indicates that for every increase in the critic rating, we expect the audience score to increase by the slope times a one-unit increase in the predictor.

Intercept indicates that for an observation where the critic rating is 0, we expect the audience score to be the intercept on average.

b.) $\beta_0 = 34.5089$ $\beta_1 = 0.4461$

$\hat{y}_8 = 34.5089 + 0.4461(8) = \boxed{38.0777}$

c.) Critics are harsher since they give a lower rating suggested by b.

d.) $\hat{y}_{86} = 34.5089 + 0.4461(86) = \boxed{72.8735}$

e.) The audience since critics rate higher in d.

f.) The regression line is the average prediction result. Some will lie below and some above.

e.) $x = \beta_0 + \beta_1 x \longrightarrow x - \beta_1 x = \beta_0 \longrightarrow x(1-\beta_1) = \beta_0$

$x = \frac{34.5089}{1-0.4461} = \boxed{62.3} \longleftarrow \boxed{x = \frac{\beta_0}{1-\beta_1}}$