HW 1

1. Consider the following statement: "For the ordinary least squares method to be fully valid, it is required that the distribution of Y be normal." Is this statement true or false, and why?

False, as long as the errors average out to 0, the variance is constant for all errors, and the errors are independant in the model. (no covariance), then OLS can be utilized.

2. Read section 1.8 of the textbook. When there is a Normal distribution assumption on the error terms, we can also formulate Maximum Likelihood Estimators for β_0, β_1, & σ^2 . Use the likelihood function (1.26) to find the estimators $\hat{\beta}_0$, $\hat{\beta}_1$, & $\hat{\sigma}^2$, and show that the estimators β_0, β_1, are the same as the Least Square estimators. You do not need to check second derivatives to prove the maximum values.

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{-1}{2\sigma^2}(Y_i - \beta_0 - \beta_1 X_i)^2\right)}$$

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

Log likelihood

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

Set derivatives to 0

chain rule

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

$$\frac{d\ell}{dB_0} = \cancel{\frac{1}{\sigma^2}} \cdot \sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i) = 0$$

$$\frac{d\ell}{dB_1} = \cancel{2 \cdot \frac{1}{2\sigma^2}} \sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i) X_i = 0$$

replace σ^2 with x

$$\frac{d\ell}{d\sigma^2} = -\frac{n}{2}\frac{1}{x} + \frac{1}{2x^2} \sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i)^2 = 0$$

$$\frac{\cancel{n}^{\,\sigma^2}}{\cancel{2}\sigma^{\cancel{2}}} = \frac{1}{\cancel{2}\sigma^{\cancel{4}\,2}_{\;n}} (\qquad)$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i)^2}{n}$$

Normal eqn

$$\sum Y_i = nB_0 + B_1 \sum X_i$$

$$\sum X_i Y_i = B_0 \sum X_i + B_1 \sum X_i^2$$

$$B_0 = \frac{1}{n} \left(\sum Y_i - B_1 \sum X_i \right)$$

$$B_0 = \bar{Y} - B_1 \bar{X} \checkmark$$

log eqn

$$\sum (Y_i - \hat{B_0} - \hat{B_1} X_i) = 0$$

$$n\hat{B_0} = \sum Y_i - \hat{B_1} \sum X_i$$

$$\hat{B_0} = \bar{Y} - \hat{B_1} \bar{X} \checkmark$$

$$\sum (Y_i - \hat{\beta_0} - \hat{\beta_1} X_i) X_i = 0 \qquad B_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \checkmark$$

$$\sum (Y_i - \bar{Y} + \hat{\beta_1} \bar{X} - \hat{\beta_1} X_i) X_i = 0$$

$$\sum Y_i X_i - \bar{Y} \sum X_i + \hat{\beta_1} \bar{X} \sum X_i - \hat{\beta_1} \sum X_i^2 = 0$$

$$\sum Y_i X_i - \bar{Y} \sum X_i = \hat{\beta_1} \left(\sum X_i^2 - \bar{X} \sum X_i \right)$$

$$\hat{\beta_1} = \frac{\sum (Y_i - \bar{Y})(X_i - \bar{X})}{\sum (X_i - \bar{X})^2} \checkmark$$

3. The solution for the LS estimator of the slope in simple linear regression is:

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

Prove that this expression is equivalent to the alternate formulation:

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

$$SSXY = \Sigma (x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i y_i - x_i \bar{y} - y_i \bar{x} - \bar{x}\bar{y}$$

$$= \Sigma x_i y_i - \bar{y} \Sigma x_i - \bar{x} \Sigma y_i + n\bar{x}\bar{y}$$

$$= \Sigma x_i y_i - \bar{y} n \bar{x} \cancel{- \bar{x} n \bar{y}} \cancel{+ n\bar{x}\bar{y}}$$

$$= \Sigma x_i y_i - n\bar{x}\bar{y}$$

$$SSX = \Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - 2\bar{x} \Sigma x_i + n\bar{x}^2$$

$$= \Sigma x_i^2 - 2\bar{x} \cdot \bar{x} + n\bar{x}^2 = \Sigma x_i^2 - n\bar{x}^2$$

$$\frac{SSXY}{SSX} = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} \checkmark$$

4. Recall that the residual for the i^{th} observation is defined as $e_i = y_i - \hat{y}_i$. Prove that $E(e_i) = 0$.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$e_i = \beta_0 + \beta_1 x_i + \varepsilon_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$= (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) x_i + \varepsilon_i$$

$$E(e_i) = E[\qquad\qquad]$$

$$E(e_i) = E(\beta_0 - \hat{\beta}_0) + E(\beta_1 - \hat{\beta}_1) x_i + E(\varepsilon_i)$$

if unbiased $\hat{\beta}_0 = \beta_0$, $\hat{\beta}_1 = \beta_1$,

and $E(\varepsilon_i) = 0$ by property of OLS (Gauss-Markov thm.)

$$E(e_i) = E(0) + E(0) x_i + 0$$

$$E(e_i) = 0$$

5. (a) When asked to state the simple linear regression model, a student wrote:

$$E(Y_i) = \beta_0 + \beta_1 x_i + \epsilon_i.$$

Do you agree? Why or why not?

(b) Consider the classical simple linear regression model. Suppose that the true parameter values are $\beta_0 = 2$, $\beta_1 = 4$, and $\sigma^2 = 9$. State the distributions of Y at $x = 1, 2$, and 4, and explain how you found them.

a) No, the expected value would not include the error as that term and would just be

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

the actual model would just be

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

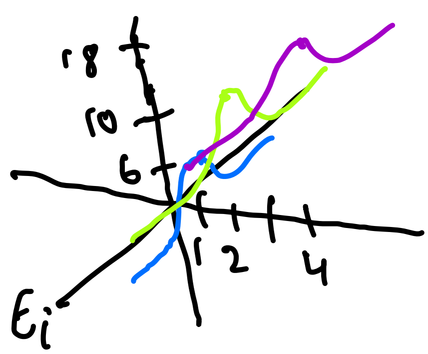
b)

$X = 1$ $\quad E(Y_i) = 2 + 4(1) = 6 : Y \sim N(6, 9)$

$X = 2$ $\quad E(Y_i) = 2 + 4(2) = 10 : Y \sim N(10, 9)$

$X = 4$ $\quad E(Y_i) = 2 + 4(4) = 18 : Y \sim N(18, 9)$

Expected value with normally distributed errors



6. Consider the Rotten Tomatoes movie rating example.

(a) Interpret the slope and the intercept in the real-life context of the problem.

(b) Suppose the Borderlands movie is about to be released and critics have given it a score of 8 (out of 100) on Rotten Tomatoes. Using the fitted simple linear regression line, what do we predict the audience rating will be?

(c) What does the SLR prediction in (b) suggest about who the regression line thinks is the harsher judge: audiences or critics?

(d) Suppose "The Quiet Place: Day One" movie is about to be released and critics have given it a score of 86 (out of 100) on Rotten Tomatoes. Using the fitted simple linear regression line, what do we predict the audience rating will be?

a) β_0: Hypothetical predicted audience rating based on critic rating of 0

β_1: Change in predicted audience rating at each individual point increase in critic rating (or decrease if negative)

b) $\hat{Y} = \beta_0 + \beta_1 X = \beta_0 + \beta_1 \cdot 8$

See python notebook, $\hat{\beta}_0 \approx 34.51$

$\hat{\beta}_1 \approx 0.446$

$\hat{Y}(8) \approx 38.08$

c) The regression model thinks Critics are much harsher
$\hat{Y}(8) >> 8$

d) $\hat{Y}(86) \approx 72.87$ (see python notebook

(e) What does the SLR prediction (d) suggest about who the regression line thinks is the harsher judge: audiences or critics?

(f) Consider your findings in (c) and (e). Provide a reasonable explanation as to how one can reconcile these two results.

(g) What value of critic ratings will the SLR model predict the exact same score for audience ratings? Derive a general formula for this value in terms of $\hat{\beta}_0$ & $\hat{\beta}_1$.

e) Since $\hat{Y}(86) << 86$ suggests audiences are harsher

f) It seems to suggest the ratings are not truly linearly related across the whole range.

g) $X = B_1 X + B_0$

$$X = \frac{B_0}{1 - B_1} \approx \frac{34.508}{1 - 0.446} \approx 62.29$$