Yuhan Zhu
A54482257

STT863 Homework 1

1.6 Consider the normal error regression model. Suppose that the parameter values are $\beta_0 = 200$, $\beta_1 = 5.0$, and $\sigma = 4$.

a. Plot this normal error regression model in the fashion of Figure 1.6. Show the distributions of Y for X = 10, 20, and 40.

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

$$Y_i = 200 + 5X_i + \varepsilon_i$$

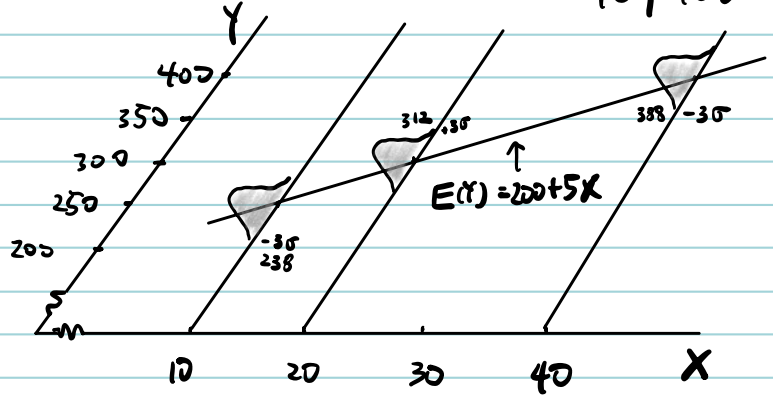
$$E\{Y\} = 200 + 5X$$

X	Y
10	250
20	300
40	400

$$\sigma^2(\varepsilon_i) = \sigma^2$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$\sigma = 4$
$2\sigma = 8$
$3\sigma = 12$



b. Explain the meaning of the parameters β_0 and β_1. Assume that the scope of the model includes X = 0.

β_0 is the Y intercept of the regression line. Which means when X = 0, the mean of the probability distribution is at (0, 200)

β_1 is the slope of the regression line. In this case it indicates mean of the Prob. dist of Y will increase 5 as X increase 1.

1.19. **Grade point average.** The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from the ACT test score (X). The results of the study follow. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	...	118	119	120
X_i:	21	14	28	...	28	16	28
Y_i:	3.897	3.885	3.778	...	3.914	1.860	2.948

a. Obtain the least squares estimates of β_0 and β_1, and state the estimated regression function.

b. Plot the estimated regression function and the data. Does the estimated regression function appear to fit the data well?

c. Obtain a point estimate of the mean freshman GPA for students with ACT test score $X = 30$.

d. What is the point estimate of the change in the mean response when the entrance test score increases by one point?

a. set least squares estimates b_0, b_1

$E(b_0) = \beta_0 \quad E(b_1) = \beta_1$

$\sum Y_i = nb_0 + b_1 \sum X_i$

$\sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2$

$\therefore b_1 = \dfrac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$

$b_0 = \frac{1}{n}(\sum Y_i - b_1 \sum X_i) = \bar{Y} - b_1 \bar{X}$

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

we try to minimize Q

$\therefore$ take partial derivaties

$\frac{\partial Q}{\partial \beta_0} = 0 = -2\sum (Y_i - \beta_0 - \beta_1 X_i) = -2\sum (Y_i - b_0 - b_1 X_i)$

$\frac{\partial Q}{\partial \beta_1} = 0 = -2\sum X_i (Y_i - \beta_0 - \beta_1 X_i) = -2\sum X_i (Y_i - b_0 - b_1 X_i)$

simplify and expanding we get

$$\sum Y_i - nb_0 - b_1 \sum X_i = 0$$

$$\sum X_i Y_i - b_0 \sum X_i - b_1 \sum X_i^2 = 0$$

we can get estimators b_0, b_1 by solving above equations.

I will use R to calculate b_0 and b_1

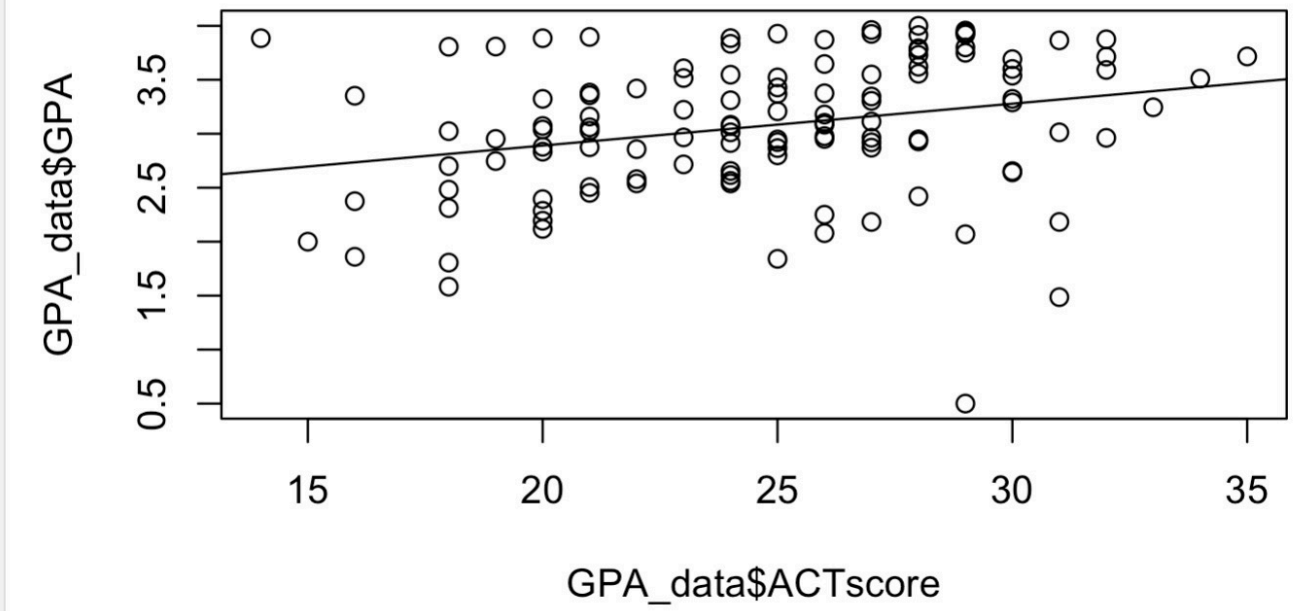
```
fit <- lm(GPA~ACTscore,data=GPA_data)
summary(fit)
```

```
Call:
lm(formula = GPA ~ ACTscore, data = GPA_data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.74004 -0.33827  0.04062  0.44064  1.22737

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.11405    0.32089   6.588  1.3e-09 ***
ACTscore     0.03883    0.01277   3.040  0.00292 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6231 on 118 degrees of freedom
Multiple R-squared:  0.07262,   Adjusted R-squared:  0.06476
F-statistic:  9.24 on 1 and 118 DF,  p-value: 0.002917
```

$$Y = 2.11405 + 0.03883X$$

b.

```r
plot(GPA_data$ACTscore, GPA_data$GPA)
abline(fit)
```



It kind of fit the data, but I think the correlation between ACT score and GPA are not very strong.

c.

$$E(Y) = \beta_0 + \beta_1 X + \varepsilon$$

$$= 2.11405 + 0.03883 \times 30$$

$$\approx 3.278803$$

Use R code:

```r
predict(fit, newdata = data.frame(ACTscore = 30))
```

Return:

```
       1
3.278863
```

d. When entrance test score increase by 1, the mean GPA score is estimated to increase by roughly 0.0388.

1.28. **Crime rate.** A criminologist studying the relationship between level of education and crime rate in medium-sized U.S. counties collected the following data for a random sample of 84 counties; X is the percentage of individuals in the county having at least a high-school diploma, and Y is the crime rate (crimes reported per 100,000 residents) last year. Assume that first-order regression model (1.1) is appropriate.

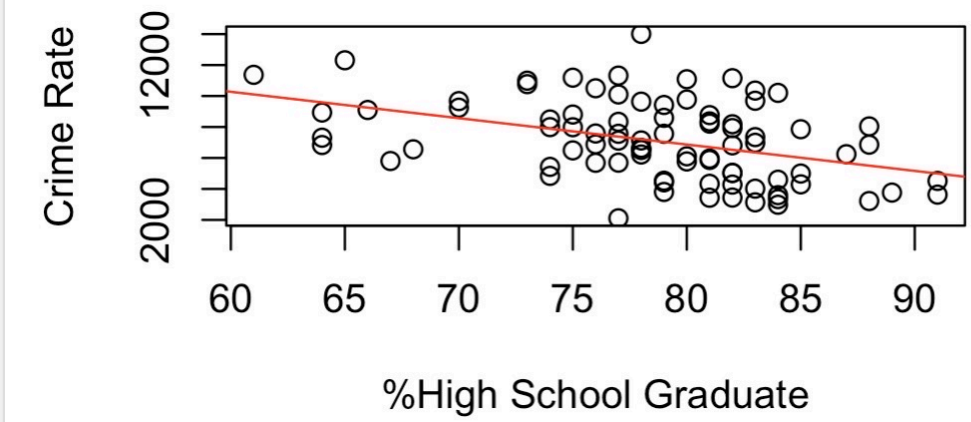
i:	1	2	3	...	82	83	84
X_i:	74	82	81	...	88	83	76
Y_i:	8,487	8,179	8,362	...	8,040	6,981	7,582

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does the linear regression function appear to give a good fit here? Discuss.

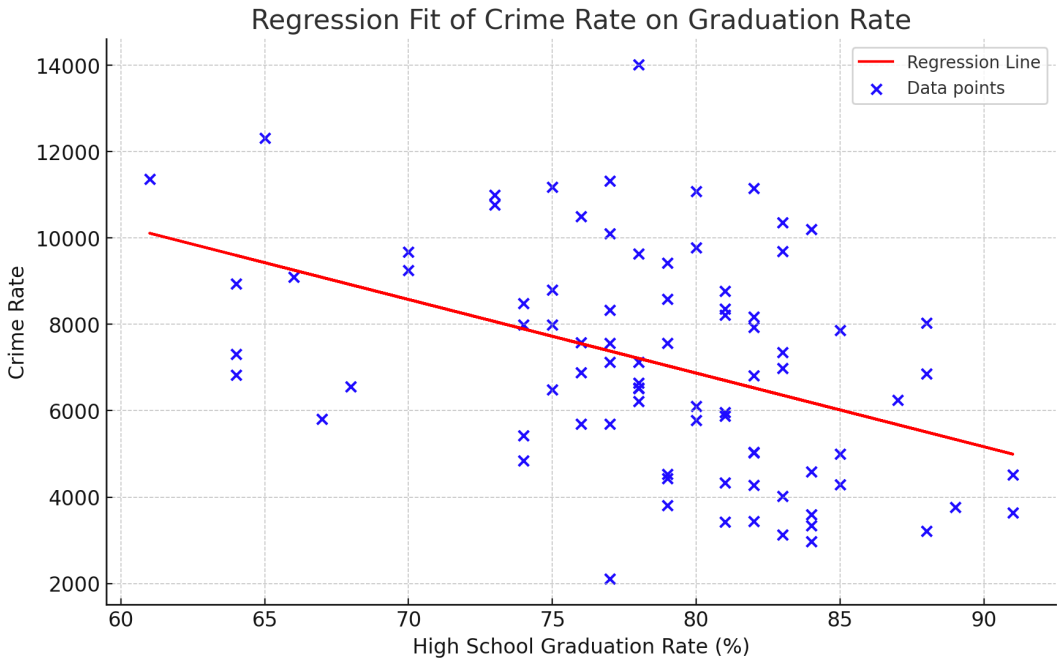
b. Obtain point estimates of the following: (1) the difference in the mean crime rate for two counties whose high-school graduation rates differ by one percentage point, (2) the mean crime rate last year in counties with high school graduation percentage $X = 80$, (3) ε_{10}, (4) σ^2.

a. $E(Y) = 20511.60 - 170.5732X$

```r
plot(data$Graduate,data$CrimeRate,xlab="%High School Graduate",ylab="Crime Rate")
abline(fit2, col="red")
```



A better plot



b. (1) For every one percentage point increase in graduation rate. the crime rate is estimate to decrease roughly 170.58 pre 10K residents.

(2) $20517.6 - 170.58 \times 80 \approx 6871.58$

(3) ε_{10}, using R code fit2$residuals[10]

return 10
1401.566

$\varepsilon_{10} = 1401.\bar{5}66$

(4) $\sigma^2 \approx 5485219$ using R code:

resi ← fit2$residuals
var(resi)
⇒ 5485219

1.30. Refer to regression model (1.1). What is the implication for the regression function if $\beta_1 = 0$ so that the model is $Y_i = \beta_0 + \varepsilon_i$? How would the regression function plot on a graph?

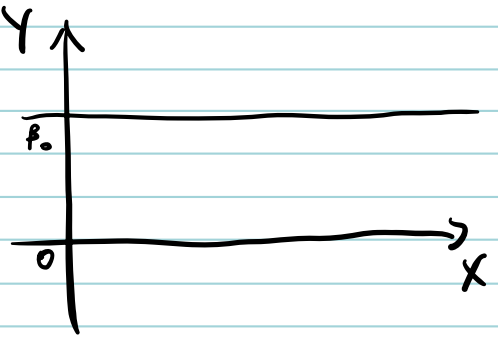
regression model (1.1) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

If $\beta_1 = 0$ then $\beta_1 X_i = 0$

means X_i and Y_i are independent from each other.

No matter how big or how small X_i is. Y will stay pretty constant around β_0 with an error ε_i.

The plot would look like this



The mean of probability distribution of Y, or $E(Y)$ is β_0.

1.33. (Calculus needed.) Refer to the regression model $Y_i = \beta_0 + \varepsilon_i$ in Exercise 1.30. Derive the least squares estimator of β_0 for this model.

1.34. Prove that the least squares estimator of β_0 obtained in Exercise 1.33 is unbiased.

1.33 Given $Y_i = \beta_0 + \varepsilon_i$ to minimize the sum of the squared residuals.

Residuals $= e_i = Y_i - b_0$

$$SSR = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - b_0)^2$$

Differentiate the SSR with respect to β_0 and result $= 0$

$$\frac{\partial}{\partial b_0} SSR = \frac{\partial}{\partial b_0} \sum_{i=1}^{n} (Y_i - b_0)^2$$

$$\Rightarrow -2 \sum_{i=1}^{n} (Y_i - b_0) = 0$$

$$\sum_{i=1}^{n}(Y_i - b_0) = 0$$

$$\sum_{i=1}^{n} Y_i = n b_0$$

$$\therefore b_0 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$\therefore b_0 = \bar{Y}$$

which means estimator b_0 is just the mean of probability distribution of Y.

1.34 To prove estimator of β_0 is unbiased

Let b_0 be the estimator of β_0

Show $E(b_0) = \beta_0$

from 1.33 we know $b_0 = \bar{Y}$

$$\therefore b_0 = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$E(b_0) = E(\bar{Y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$$

Now we put expectation in the equation $Y_i = \beta_0 + \varepsilon_i$

$$E(Y_i) = E(\beta_0 + \varepsilon_i) = E(\beta_0) + E(\varepsilon_i)$$

knowing β_0 is constant $E(\varepsilon_i) = 0$

we get $E(Y_i) = \beta_0$

Now $E(b_0) = \frac{1}{n}\sum_{i=1}^{n} \beta_0$

$\frac{1}{n}\sum_{i=1}^{n}$ cancel out each other

$$\Downarrow$$

$$E(b_0) = \frac{n}{n}\beta_0 = \beta_0$$

$$\Rightarrow E(b_0) = \beta_0$$

$\therefore$ estimator b_0 is unbiased estimator of β_0.

1.41. (Calculus needed.) Refer to the regression model $Y_i = \beta_1 X_i + \varepsilon_i, i = 1, \ldots, n$, in Exercise 1.29.

a. Find the least squares estimator of β_1.

b. Assume that the error terms ε_i are independent $N(0, \sigma^2)$ and that σ^2 is known. State the likelihood function for the n sample observations on Y and obtain the maximum likelihood estimator of β_1. Is it the same as the least squares estimator?

c. Show that the maximum likelihood estimator of β_1 is unbiased.

a. The residuals $e_i = Y_i - \beta_1 X_i$

$$\therefore SSR = \sum_{i=1}^{n}(Y_i - \beta_1 X_i)^2$$

Differentiate with respect to β_1 and set result to 0

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

$$\sum_{i=1}^{n} X_i(Y_i - \beta_1 X_i) = 0$$

$$\vdots$$

$$\Rightarrow \quad b_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

b. Assume ε_i are independent $N(0, \sigma^2)$ and know σ^2

The likelihood function can be written as

$$f(Y_i | X_i, \beta_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \beta_1 X_i)^2}{2\sigma^2}}$$

Joint likelihood: $L(\beta_1) = \prod_{i=1}^{n} f(Y_i | X_i, \beta_1)$

log likelihood $= l(\beta_1) = \ln(L(\beta_1))$

$$= \sum_{i=1}^{n} \ln(f(Y_i | X_i, \beta_1))$$

Substituting pdf into log-likelihood

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

Differentiate $l(\beta_1)$ with respect to β_1 and set result to 0

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

Same as part a $\rightarrow$ $\sum_{i=1}^{n} X_i(Y_i - \beta_1 X_i) = 0$

solving $b_1 \Rightarrow$ $b_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_1^2}$

$\therefore$ Maximum likelihood estimator b_1 for β_1 is the same as part a the least squares estimator.

c. Show $E(b_1) = \beta_1$

Given $b_1 = \dfrac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$

$$E(b_1) = E\left(\frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}\right) = E\left(\frac{\sum_{i=1}^{n} X_i(\beta_1 X_i + \varepsilon_i)}{\sum_{i=1}^{n} X_i^2}\right)$$

X_i^2 and X_i^2 canceled out $E(\varepsilon_i) = 0$

$\therefore E(b_1) \Rightarrow \beta_1$

$\therefore$ estimator b_1 is unbiased for β_1

1.42. **Typographical errors.** Shown below are the number of galleys for a manuscript (X) and the dollar cost of correcting typographical errors (Y) in a random sample of recent orders handled by a firm specializing in technical manuscripts. Assume that the regression model $Y_i = \beta_1 X_i + \varepsilon_i$ is appropriate, with normally distributed independent error terms whose variance is $\sigma^2 = 16$.

i:	1	2	3	4	5	6
X_i:	7	12	4	14	25	30
Y_i:	128	213	75	250	446	540

a. State the likelihood function for the six Y observations, for $\sigma^2 = 16$.

b. Evaluate the likelihood function for $\beta_1 = 17$, 18, and 19. For which of these β_1 values is the likelihood function largest?

c. The maximum likelihood estimator is $b_1 = \sum X_i Y_i / \sum X_i^2$. Find the maximum likelihood estimate. Are your results in part (b) consistent with this estimate?

d. Using a computer graphics or statistics package, evaluate the likelihood function for values of β_1 between $\beta_1 = 17$ and $\beta_1 = 19$ and plot the function. Does the point at which the likelihood function is maximized correspond to the maximum likelihood estimate found in part (c)?

a. $\varepsilon_i \sim N(0, \sigma^2)$ $\sigma^2 = 16$

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

$$L(\beta_1) = \frac{n}{i=1} \frac{1}{\sqrt{32\pi}} e^{-\frac{1}{32}(Y_i - \beta_1 X_i)^2}$$

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

plug in the value of $\beta_1 = 17$, 18, and 19

we get

$\log L(17) = -5,419,223$

$\log L(18) = -6,077,752$

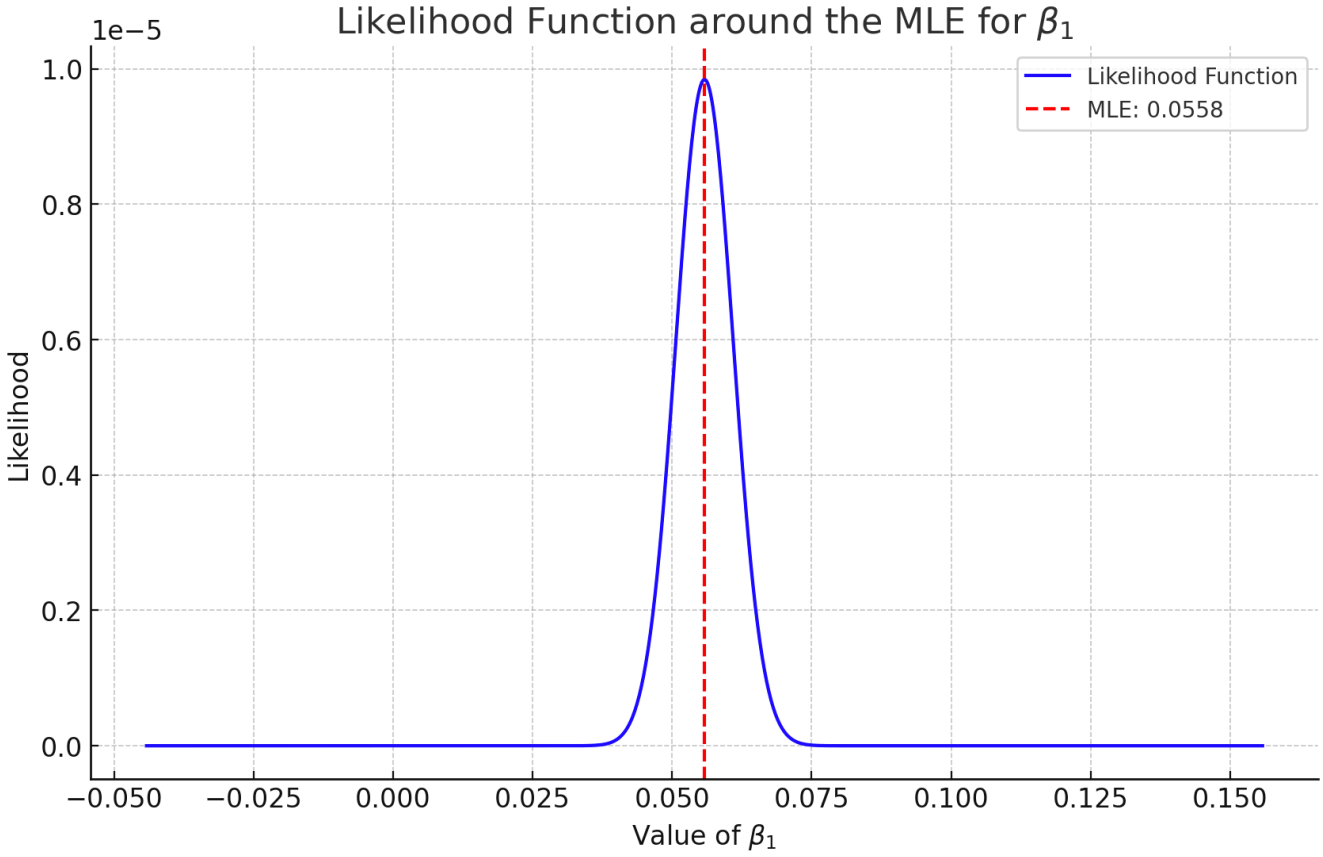
$\log L(19) = -6,774,032$

$\therefore$ $\beta_1 = 17$ has the largest likelihood value

c. Use computer we get MLE for β_1 is about 0.0658

which make sense the likelihood values for those β_1 were small.

d.


Likelihood Function around the MLE for β_1
1e−5
1.0
0.8
0.6
0.4
0.2
0.0
Likelihood
−0.050
−0.025
0.000
0.025
0.050
0.075
0.100
0.125
0.150
Value of β_1
Likelihood Function
MLE: 0.0558