S2208 MATH8050 Data Analysis - Section 001: Homework 3 Due on 09/14/22

Adithya Ravi, C09059838

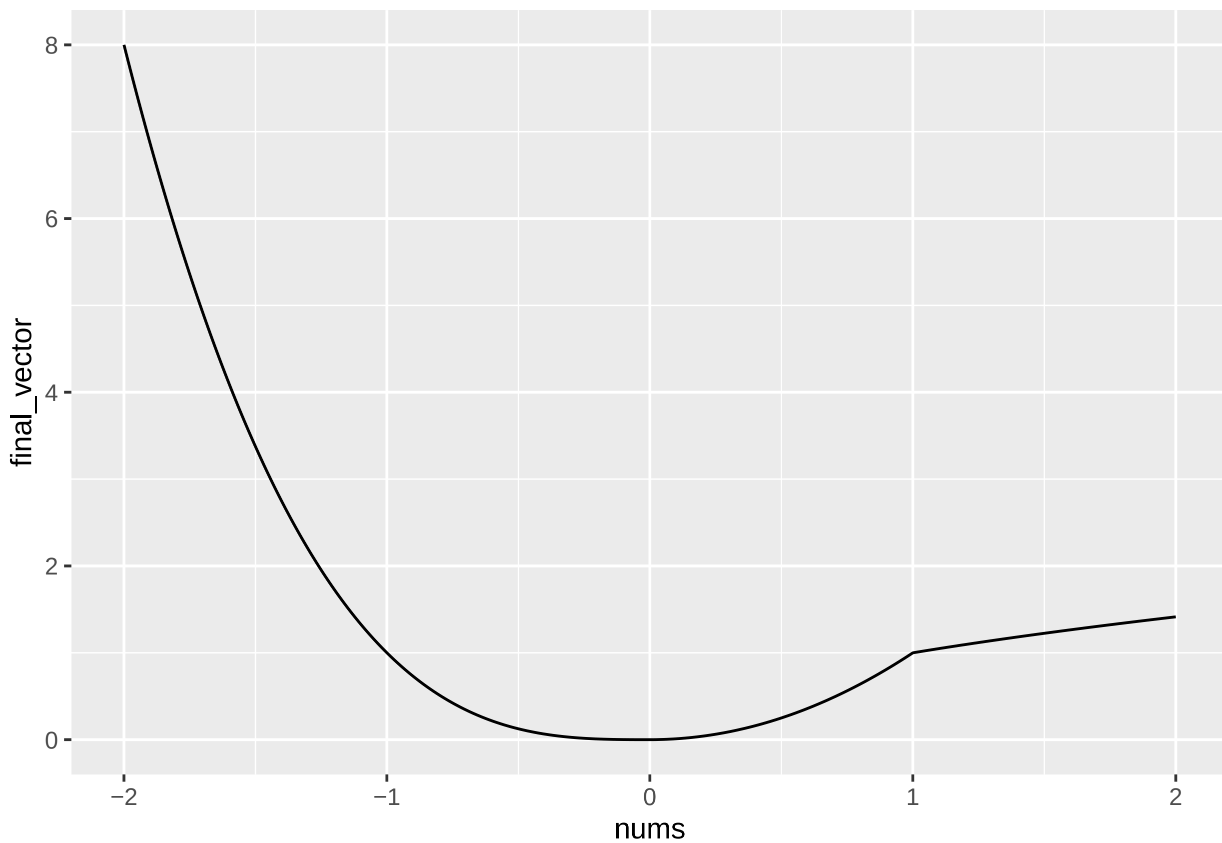
2022-09-21

Solutions

Question1

$$f(x) = \begin{cases} -x^3 & x \leq 0, \\ x^2 & x \in (0,1], \\ \sqrt{x} & x > 1. \end{cases}$$

```
library(ggplot2)
f <- function(input_vector) {

  myVector<-rep(0,length(input_vector))
  y<-0
  c<-1

  for (x in input_vector){
    if (x<=0){
      y<--(x)^3
    }
    else if(x<=1 & x>0){
      y<-x^2
    }
    else{
      y<-sqrt(x)
    }

    myVector<-replace(myVector,c,y)
    c<-c+1
  }
  return(myVector)
}

nums<-seq(2,-2,length=1000)
final_vector<-f(nums)


ggplot(mapping=aes(x=nums,y=final_vector))+geom_line()
```



Question2

2a

$$p(x|\theta) = \text{Poisson}(x|\theta) = \exp(-\theta)\frac{\theta^x}{x!}$$

Likelihood Function

$$p(\theta; x_1, ..., x_n) = \prod_{j=1}^{n} \frac{\theta^{x_j}}{x_j!} \exp(-\theta)$$

Log Likelihood Function

$$\ln p(\theta; x_1, ..., x_n) = \ln\left(\prod_{j=1}^{n} \frac{\theta^{x_j}}{x_j!} \exp(-\theta)\right)$$

$$\ln p(\theta; x_1, ..., x_n) = \sum_{j=1}^{n} \ln\left(\frac{\theta^{x_j}}{x_j!} \exp(-\theta)\right)$$

$$\ln p(\theta; x_1, ..., x_n) = \sum_{j=1}^{n} (\ln(\theta^{x_j}) + \ln(exp^{-\theta}) - \ln(x_j!))$$

$$\ln p(\theta; x_1, ..., x_n) = \sum_{j=1}^{n} (\ln(\theta^{x_j}) - \theta - \ln(x_j!))$$

$$\ln p(\theta; x_1, ..., x_n) = \sum_{j=1}^{n} (\ x_j \ln(\theta) - \theta - \ln(x_j!))$$

$$\ln p(\theta; x_1, ..., x_n) = -n\theta + \ln(\theta) \sum_{j=1}^{n} \ x_j - \sum_{j=1}^{n} \ln(x_j!)$$

The derivative of this function and equate it to zero

$$\frac{d}{d\theta} p(\theta; x_1, ..., x_n) = \frac{d}{d\theta}(-n\theta + \ln(\theta) \sum_{j=1}^{n} \ x_j - \sum_{j=1}^{n} \ln(x_j!))$$

$$\frac{d}{d\theta} p(\theta; x_1, ..., x_n) = -n + (\frac{1}{\theta}) \sum_{j=1}^{n} \ x_j$$

Set the derivative to zero

$$-n + \frac{1}{\theta} \sum_{j=1}^{n} x_j = 0$$

$$\theta = \frac{1}{n} \sum_{j=1}^{n} x_j$$

Therefore MLE is

$$\hat{\theta} = \frac{1}{n} \sum_{j=1}^{n} x_j$$

2b

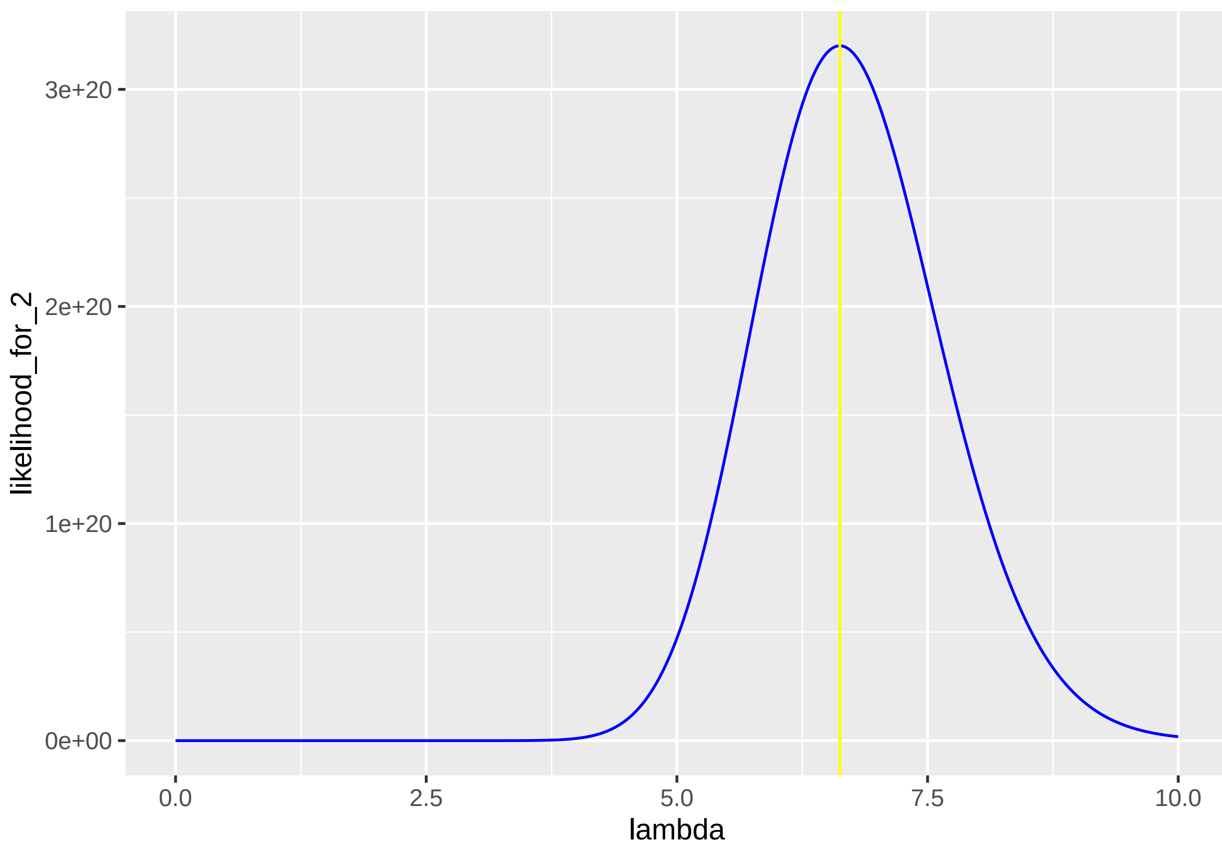
Now to prove it is the maximum value and unique take second derivative. and sub with value of $\theta = \frac{1}{n} \sum_{j=1}^{n} x_j$

$$\frac{d^2}{d^2\theta} p(\theta; x_1, ..., x_n) = \frac{-1}{\theta^2} \sum_{j=1}^{n} x_j = \frac{-n^2}{\sum_{j=1}^{n} x_j} < 0$$

we see that is the equation is less than zero, so it the maximum value and the unique value.

2c

```r
x = c(0, 2, 5, 6, 10, 15, 8, 7)

lambda = seq(0,10,length=5000)

likelihood_for_2=lambda^(sum(x)) * exp(-length(x)*lambda)

log.lklh.poisson <- function(x, lambda){
  -sum(x * log(lambda) - log(factorial(x)) - lambda)
}


theta_cap<-optim(par = 2, log.lklh.poisson, x = x)


MLE<-theta_cap$par
ggplot()+geom_line(aes(x=lambda,y=likelihood_for_2),color="blue")+
  geom_vline(xintercept =theta_cap$par,color="yellow")
```




```r
#mean(x)
```

Yes, the numerical solution is same as the true MLE.

2d

The prior and likelihood are given by

$$p(\theta) = Ga(\theta|a,b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta).$$

$$p_{x|\theta}(x1, ...x_{n|\theta}) = \prod_{i=1}^{n} \theta^{x_i} \frac{\exp^{-\theta}}{x_i!}$$

Dropping proportionality constants that do not depend on θ , the posterior distribution of θ given $X = x_1, ..., X_n = x_n$ is then

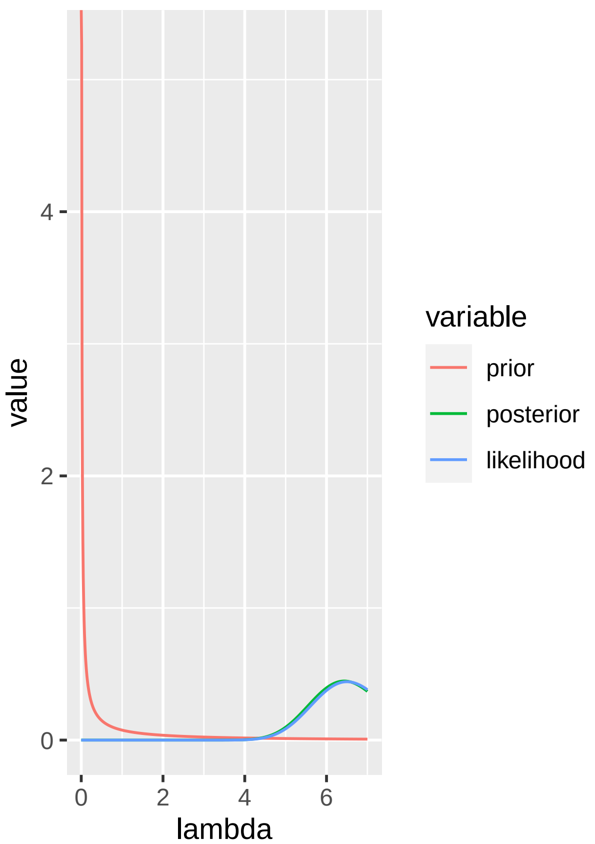
$p_{\theta|x}(x1, ...x_{n|\theta}) \propto \ p_{x|\theta}(x1, ...x_{n|\theta}) p_\theta(\theta) \propto \prod_{i=1}^{n} \theta^{x_i} \exp^{-\theta} \times \theta^{a-1} \exp^{-b\theta} = \theta^{s+a-1} \exp^{(n+b)\theta}$

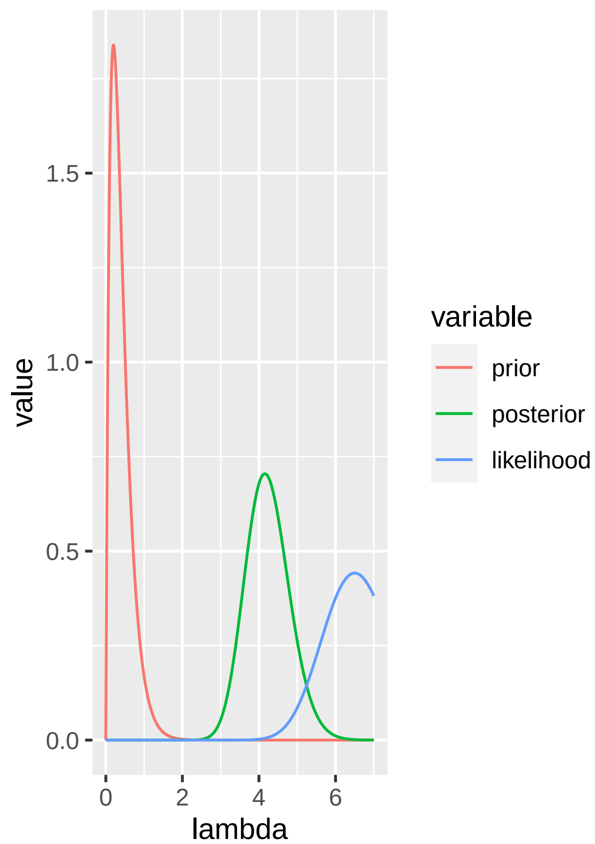
where $s = x1 + ... + x_n$.This is proportional to the PDF of the $\Theta(s + a, n + b)$distribution, so the posterior distribution $\Theta(s + a, n + b)$.

As the prior and posterior are both Gamma distributions, the Gamma distribution is a conjugate prior for Θ in the Poisson model.

2e

```
alpha <-2
beta <- 5

lambda <- seq(0,7, by = 0.01)
prior<-dgamma(lambda, alpha, beta)
posterior<-dgamma(lambda, alpha + sum(x), beta + length(x))
likelihood<-dgamma(lambda,sum(x),length(x))
df_of_values<-data.frame(lambda=lambda,prior=prior,posterior=posterior,
                         likelihood=likelihood)
d <- melt(df_of_values, id.vars="lambda")

two_5<-ggplot(d,aes(lambda,value,color=variable))+geom_line()

alpha2 <-0.1
beta2 <- 0.1

lambda2 <- seq(0,7, by = 0.01)
prior2<-dgamma(lambda2, alpha2, beta2)
posterior2<-dgamma(lambda2, alpha2 + sum(x), beta2 + length(x))
likelihood2<-dgamma(lambda,sum(x),length(x))
df_of_values2<-data.frame(lambda=lambda2,prior=prior2,posterior=posterior2,likelihood=likelihood2)
d2 <- melt(df_of_values2, id.vars="lambda")

plota <- ggplot(d2,aes(lambda,value,col=variable))+geom_line()
top_row<-plot_grid(plota,two_5,ncol=2)
top_row
```





2f

$$p(x_{1:n}) = \int_0^\infty p(x_{1:n}|\theta)p(\theta)d\theta$$
$$= \int_0^\infty \left[\frac{b^a}{\Gamma(a)}\theta^a - 1\exp^{-b\theta}\right]\left[\prod_{i=1}^n \theta^{x_i}\frac{\exp^{-\theta}}{x_i!}\right]$$
$$= \frac{b^a}{\Gamma(a)\prod_{i=1}^n x_i!}\int_0^\infty \theta^{(a-1+nx_i)}\exp^{-(n+b)\theta}d\theta$$
$$= \frac{\Gamma(a+\Sigma_{i=1}^n x_i)}{(n+b)^{a+\Sigma_{i=1}^n x_i}}\frac{b^a}{\Gamma(a)\prod_{i=1}^n x_i!}$$

2g

Posterior Predictive Distribution.

$$p(x_{new}|x) = \int_0^\infty p(x_{new}|\theta)p(\theta|x)d\theta$$
$$= \int_0^\infty \left[\frac{\theta^{x_{new}}\exp^{-\theta}}{x_{new}!}\right]\left[\frac{n+b^{\sum x_i+a}}{\Gamma(\sum x_i+a)}\theta^{\sum x_i+a-1}\exp^{-(n+b)\theta}\right]d\theta$$
$$p(x_{new}|x) = \frac{n+b^{\sum x_i+a}}{\Gamma(\sum x_i+a)\Gamma(x_{new}+1)}\int_0^\infty \theta^{x_{new}+\sum x_i+a-1}\exp^{-(n+b+1)\theta}d\theta$$

$$= \frac{n+b^{\sum x_i + a}}{\Gamma(\sum x_i + a)\Gamma(x_{new}+1)} \frac{\Gamma(x_{new} + \sum x_i + a)}{(n+b+1)^{x_{new}+\sum x_i + a}}$$

$$= \frac{\Gamma(x_{new} + \sum x_i + a)}{\Gamma(\sum x_i + a)\Gamma(x_{new}+1)} \left(\frac{n+b}{n+b+1}\right)^{\sum x_i + a} \left(\frac{1}{n+b+1}\right)^{x_{new}}$$

which is the a negative binomial

Question3

3a

Hat matrix is given as

$$H = X(X^T X)^{-1} X^T$$

Need to prove

$$H = H^T$$

$$H^T = (X(X^T X)^{-1} X^T)^T = X[(X^T X)^{-1}]^T X^T = X[(X^T X)^T]^{-1} X^T = X(X^T X)^{-1} X^T = H$$

$nbsp;

idempotent

$$H^2 = (X(X^T X)^{-1} X^T)(X(X^T X)^{-1} X^T) = (X(X^T X)^{-1})(X^T X)(X^T X)^{-1} X^T = X(X^T X)^{-1} X^T$$

idempotent

$$H^2 = H$$

$nbsp;

3b

$$E[\hat{Y}] = E[HY] = HE[Y] = HX\beta = X(X^T X)^{-1} X^T X\beta = X\beta$$

Next, the variance-covariance of the fitted values:

$$Var[\hat{Y}] = Var[HY] = Var[HY]$$

$$= Var[HY] = HVar[Y]H^T = \sigma^2 HIH = \sigma^2 H$$

Similarly, the expected residual vector is zero:

$$E[e] = (I - H)(X\beta) = X\beta - X\beta = 0$$

$$\begin{aligned} Var(e) &= Var[(I-H)(Y)] && (1)\\ &= Var[(I-H)Y] && (2)\\ &= (I-H)Var(Y)(I-H)^T && (3)\\ &= \sigma^2(I-H)(I-H) && (4)\\ &= \sigma^2(I-H) && (5) \end{aligned}$$

3c

$$E(\hat{Y}-Y)^2 = E(\hat{Y}^2)+E(Y^2)-2YE(\hat{Y}) = Var(\hat{Y})+[E(\hat{Y})]^2+Y^2-2YE(\hat{Y})$$
$$= Var(\hat{Y})+[E(\hat{Y})-Y]^2$$

where

$$[E(\hat{Y})-Y]^2$$

is bias. for an unbiased estimator it is variance. since the value is zero. There by

$$E(\hat{Y}-Y)^2 = Var(\hat{Y}) = \sigma^2$$

Alternative: To prove that $E(MSE)=\sigma^2$, we have to prove that $E(SSE)=(n-2)\sigma^2$.
Linear regression model:

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

Fitted regression line:

$$\hat{Y_i} = b_0 + b_1 x_i$$

By definition

$$e_i = Y_i - \hat{Y_i} = Y_i - (b_0 - b_1 x_i)$$

and since

$$b_0 = \overline{Y} - b_1\overline{X}$$

=>

$$e_i = Y_i - (\overline{Y} - b_1\overline{X} + b_1 x_1) = Y_i - \overline{Y} - b_1(x_i - \overline{X})$$

We know,

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

, and if we examine this at

$$(\overline{X},\overline{Y})$$

,

=>

$$\overline{Y} = \beta_0 + \beta_1\overline{X} + \overline{\epsilon}$$

=>

$$Y_i - \overline{Y} = (\beta_0 + \beta_1 x_i + \epsilon_i) - (\beta_0 + \beta_1\overline{X} + \overline{\epsilon})$$

=

$$\beta_1(x_i - \overline{X}) + (\epsilon_i - \overline{\epsilon})$$

we get,

$$e_i = Y_i - \overline{Y} - b_1(x_i - \overline{X}) = \beta_1(x_i - \overline{X}) + (\epsilon_i - \overline{\epsilon}) - b_1(x_i - \overline{X})$$

squaring both sides,

$$e_i^2 = (\epsilon_i - \overline{\epsilon})^2 - 2(\epsilon_i - \overline{\epsilon})(b_1 - \beta_1)(x_i - \overline{X}) + ((b_1 - \beta_1)(x_i - \overline{X}))^2$$

Sum of residuals squares = Sum of squared error

$$\sum e_i^2 = SSE = \sum(\epsilon_i - \overline{\epsilon})^2 - 2(\epsilon_i - \overline{\epsilon})(b_1 - \beta_1)(x_i - \overline{X}) + ((b_1 - \beta_1)(x_i - \overline{X}))^2$$

(i)

$$E(\sum \epsilon_i - \overline{\epsilon})^2 = E(\sum (\epsilon_i^2 - 2\epsilon_i\overline{\epsilon}) + \overline{\epsilon}^2)$$

$$E(\epsilon_i^2 - 2\overline{\epsilon} \sum \epsilon_i + n\bar{\epsilon^2})$$
$$E(\epsilon_i^2 - 2n\overline{\epsilon} \sum \frac{\epsilon_i}{n} + n\bar{\epsilon^2})$$
$$E(\epsilon_i^2 - 2n\overline{\epsilon} + n\bar{\epsilon^2})$$
$$E(\sum \epsilon_i^2 - n\bar{\epsilon^2})$$
$$\sum E(\epsilon_i^2) - nE(\overline{\epsilon^2})$$

To find

$$E(\bar{\epsilon_i^2})$$

, we recall the definition of variance of a random variable and $V(\epsilon_i) = E(\epsilon_i^2) - (E(\epsilon_i))^2$ and $E(\epsilon_i)$ is assumed to equal 0 and $V(\epsilon_i) = \sigma^2$

=>

$$E(\epsilon_i^2) = V(\epsilon_i) + (E(\epsilon_i))^2 = \sigma^2 + 0$$

likewise if

$$V(\epsilon_i) = \sigma^2$$

then

$$V(\overline{\epsilon}) = \frac{\sigma^2}{n}$$

$$E(\overline{\epsilon^2}) = V(\overline{\epsilon}) + (E(\epsilon_i))^2 = \frac{\sigma^2}{n} + 0$$

therefore,

$$E(\sum \epsilon_i - \overline{\epsilon})^2 = \sum E(\epsilon_i^2) - nE(\overline{\epsilon^2}) = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$

ii)

$$E[-2(b_1 - \beta_1) \sum (x_i - \overline{x})(\epsilon_i - \overline{\epsilon})]$$
$$b_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \sum K_i Y_i$$

where

$$K_i = \frac{\sum (X_i - \overline{X})}{\sum (X_i - \overline{X})^2} Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$
$$b_1 = \sum K_i(\beta_0 + \beta_1) = \beta_0 \sum K_i + \beta_1 \sum K_i x_i + \sum K_i \epsilon_i$$

We know $\sum K_i = 0$ and $\sum K_i x_i = 1$

$b_1 = \beta_0(0) + \beta_1(1) + \sum(K_i\epsilon_i) = \beta_1 + \sum(K_i\epsilon_i)$

$$E[-2(b_1 - \beta_1) \sum(x_i - \overline{X})(\epsilon_i - \overline{\epsilon})]$$
$$-2E[\sum K_i\epsilon_i \sum(x_i - \overline{X})(\epsilon_i - \overline{\epsilon})]$$

$$-2E[\sum \frac{x_i - \overline{X}}{\sum(x_i - \overline{X})^2}\epsilon_i \sum(x_i - \overline{X})(\epsilon_i - \bar{\epsilon})]$$

and

$$\bar{\epsilon} = 0$$

by assumption, so

$$= -2E[\frac{(\sum (x_i - \overline{X})\epsilon_i)^2}{\sum (x_i - \overline{X})^2}] = -2E(\epsilon_i)^2 = -2\sigma^2$$

Since

$$E(\epsilon_i^2) = V(\epsilon_i) + [E(\epsilon_i)]^2 = \sigma^2$$

(iii)

$E[(b_1 - \beta_1)^2 \sum(x_i - \overline{X})^2]$, Since the X values are not stochastic,

$$= \sum (x_i - \overline{X})^2 E(b_1 - \beta_1)^2$$

$E(b_1 - \beta_1)^2$ is the definition of variance of b , since $E(b_1) = \beta_1$ and $V(b_1) = \frac{\sigma^2}{\sum (x_i - X)^2)}$

we have , $E[(b_1 - \beta_1)^2 \sum(x_i - \overline{X})^2] = \frac{\sum (x_i - \overline{X})^2 \sigma^2}{(x_i - \overline{X})^2} = \sigma^2$

Question4

4a

$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$

$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix}$ $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

The above the matrix representation of y,X,β of a simple regression.

$X^T X = \begin{pmatrix} n & \Sigma_i x_i \\ \Sigma_i x_i & \Sigma_i x_i^2 \end{pmatrix}$

$(X^T X)^{-1} = \frac{1}{n\Sigma_i x_i^2 - (\Sigma_i x_i)^2} \begin{pmatrix} n & -\Sigma_i x_i \\ -\Sigma_i x_i & \Sigma_i x_i^2 \end{pmatrix}$

$$\begin{aligned} cov(\hat{\beta}) = cov[(X^T X)^{-1} X^T y] &= (X^T X)^{-1} X^T cov(y) [(X^T X)^{-1} X^T]^T \\ &= (X^T X)^{-1} X^T (\sigma^2 I) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\ cov(\hat{\beta}) &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

We know the values of the $(X^TX)^{-1}$ so we get the below equation.

$$cov(\hat{\beta}) = \sigma^2(X^TX)^{-1} = \begin{pmatrix} \frac{\sigma^2\Sigma_i x_i^2}{n\Sigma_i x_i^2-(\Sigma_i x_i)^2} & \frac{-\sigma^2\Sigma_i x_i}{n\Sigma_i x_i^2-(\Sigma_i x_i)^2} \\ \frac{-\sigma^2\Sigma_i x_i}{n\Sigma_i x_i^2-(\Sigma_i x_i)^2} & \frac{\sigma^2\Sigma_i x_i^2}{n\Sigma_i x_i^2-(\Sigma_i x_i)^2} \end{pmatrix}$$

$$cov(\hat{\beta}) = \sigma^2(X^TX)^{-1} = \begin{pmatrix} \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma_{i=1}^n(x_i-\bar{x})^2}\right) & \frac{-\sigma^2\bar{x}}{\Sigma_{i=1}^n(x_i-\bar{x})^2} \\ \frac{-\sigma^2\bar{x}}{\Sigma_{i=1}^n(x_i-\bar{x})^2} & \frac{\sigma^2\bar{x}}{\Sigma_{i=1}^n(x_i-\bar{x})^2} \end{pmatrix}$$

4b

$$cov(\hat{\beta}) = \begin{pmatrix} var(\hat{\beta}_0) & cov(\hat{\beta}_0,\hat{\beta}_1)) \\ cov(\hat{\beta}_0,\hat{\beta}_1) & var(\hat{\beta}_1) \end{pmatrix} = cov\begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{pmatrix} = \sigma^2(X^TX)^{-1}$$

From this equation we get the values of $var(\hat{\beta}_0)$,$var(\hat{\beta}_1)$ and $cov(\hat{\beta}_1,\hat{\beta}_1)$.

$$\begin{aligned}
Var(\hat{y_h}) = E(\hat{y_h} - y_0)^2 = E[(\beta_0 - \hat{\beta}_0) + (x_h - \bar{x})(\beta_1 - \hat{\beta}_1) + (\beta_1 - \hat{\beta}_1)(\bar{x})]^2 \\
= Var(\hat{\beta}_0) + (x_h - \bar{x})^2 Var(\hat{\beta}_1) + \bar{x}^2 Var(\hat{\beta}_1) + 2(x_h - \bar{x})Cov(\hat{\beta}_0,\hat{\beta}_1) + 2\bar{x}Cov(\hat{\beta}_0,\hat{\beta}_1) + 2(x_h - \bar{x})Var(\hat{\beta}_1) \\
= Var(\hat{\beta}_0) + [\bar{x}^2 + (x_h - \bar{x})^2 + 2(x_h - \bar{x})]Var(\hat{\beta}_1) + 2[(x_h - \bar{x}) + 2\bar{x}]Cov(\hat{\beta}_0,\hat{\beta}_1) \\
= Var(\hat{\beta}_0) + (x_h^2)Var(\hat{\beta}_1) + 2x_h Cov(\hat{\beta}_0,\hat{\beta}_1) \\
= \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right] + x_h^2\frac{\sigma^2}{s_{xx}} + \sigma^2 - 2x_h\frac{\bar{x}\sigma^2}{s_{xx}} \\
= \sigma^2\left[1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{s_{xx}}\right]
\end{aligned}$$