CAT 1

79546 - Stephen Ng'etich

Q1

$$\binom{10-1}{5-1}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{126}{1024} = \frac{63}{512}$$

Q2

a

Equating $E(X) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \overline{X}$

$$\overline{X} = \theta\sqrt{\frac{\pi}{2}}$$

Making θ the subject of th formula:

$$\widehat{\theta} = \overline{X}\sqrt{\frac{2}{\pi}}$$

Checking for Bias $E(\widehat{\theta}) = \theta$

$$E(\widehat{\theta}) = \overline{X}\sqrt{\frac{2}{\pi}} = \sqrt{\frac{2}{\pi}}E(\overline{X}) = \frac{1}{n}\sqrt{\frac{2}{\pi}}\sum_{i=1}^{n}E(x_i)$$

$$= \frac{1}{n}\sqrt{\frac{2}{\pi}}nE(x_i)$$

$$= \frac{1}{n}\sqrt{\frac{2}{\pi}}\theta\sqrt{\frac{\pi}{2}} = \theta$$

The estimator is unbiased

b

$$= \widehat{\sigma^2} = \frac{\widehat{\theta}^2}{2}(4-\pi) = \frac{2\overline{X^2}}{2\pi}(4-\pi) = \frac{\overline{X}^2}{\pi}(4-\pi)$$

$$E(\widehat{\sigma^2}) = \left(\frac{4-\pi}{\pi}\right)E(\overline{X^2})$$

$$= \left(\frac{4-\pi}{\pi}\right)\left(Var(\overline{X}) + (E(\overline{X}))^2\right)$$

$$= \left(\frac{4-\pi}{\pi}\right)\left(\frac{\theta^2(4-\pi)}{2n} + \frac{\theta^2\pi}{2}\right)$$

Q3

a

The population mean and variance of gamma distribution is $\mu = \alpha\beta$ and $sigma = \alpha\beta^2$ respectively The sample equivalent of the population sample mean and variance is $\mu = \overline{X}$ and $\sigma = S^2$

getting the value of β:

$$\alpha\beta = \overline{X}$$
$$\alpha = \frac{\overline{X}}{\beta}$$

Substituting α in S^2

$$S^2 = \alpha\beta^2$$
$$S^2 = \frac{\overline{X}}{\beta}\beta^2 = \overline{X}\beta$$

Get β estimate by making it the subject of the formula

$$\widehat{\beta} = \frac{S^2}{\overline{X}}$$

Getting the estimator of β

$$\alpha = \frac{\overline{X}}{\beta}$$

Submtiting β with the estimate $\widehat{b}$

$$\alpha = \frac{\overline{X}}{\frac{S^2}{\overline{X}}} = \frac{\overline{X}^2}{S^2}$$

b

Gamma distribution pdf is given by:

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$$

THe likelihood function is defined by:

$$L(f(x|\alpha,\beta)) = \ln\left(\prod_{i=1}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\right)$$

$$L(f(x|\alpha,\beta)) = \ln\left(\frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n}\prod_{i=1}^{n}x_i^{\alpha-1}e^{-\sum_{i=1}^{n}x_i\beta}\right)$$

$$\ln L = n\alpha\ln\beta - n\ln(\Gamma(\alpha)) + (\alpha-1)\sum_{x=i}^{n}\ln x - \sum_{x=i}^{n}x_i\beta$$

First partial derivative with respect to β

$$\frac{d\ln L}{d\beta} = \frac{n\alpha}{\beta} - \sum_{x=i}^{n}x_i = 0$$

Solve for β

$$\frac{\alpha n}{\beta} = \sum_{x=i}^{n} x_i$$

$$\frac{1}{\beta} = \frac{\sum_{x=i}^{n} x_i}{\alpha n}$$

$$\widehat{\beta} = \frac{\alpha}{\frac{1}{n}\sum_{x=i}^{n} x_i} = \frac{\alpha}{\overline{X}}$$

First partial derivative with respect to α

$$\frac{d \ln L}{d\alpha} = n \ln \beta - \frac{n\Gamma\alpha'}{\Gamma\alpha} + \sum_{i=1}^{n} \ln x_i = 0$$

$$= \frac{n\Gamma\alpha'}{\Gamma\alpha} = n \ln \beta + \sum_{i=1}^{n} \ln x_i$$

Dividing every term by n:

$$= \frac{\Gamma\alpha'}{\Gamma\alpha} = \ln \beta + \frac{\sum_{i=1}^{n} \ln x_i}{n}$$

c

$$\widehat{\beta} = \frac{S^2}{\overline{X}}$$

Calculate $\overline{X}$:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{168.5}{25} = 6.74$$

$$S^2 = 0.46166667$$

$$\widehat{\beta} = \frac{0.461666667}{6.74} = 0.06849$$

$$\alpha = \left(\frac{\overline{X}}{S}\right)^2$$

$$\alpha = (\frac{6.74}{0.461666667})^2 = 213.1389$$

Q4

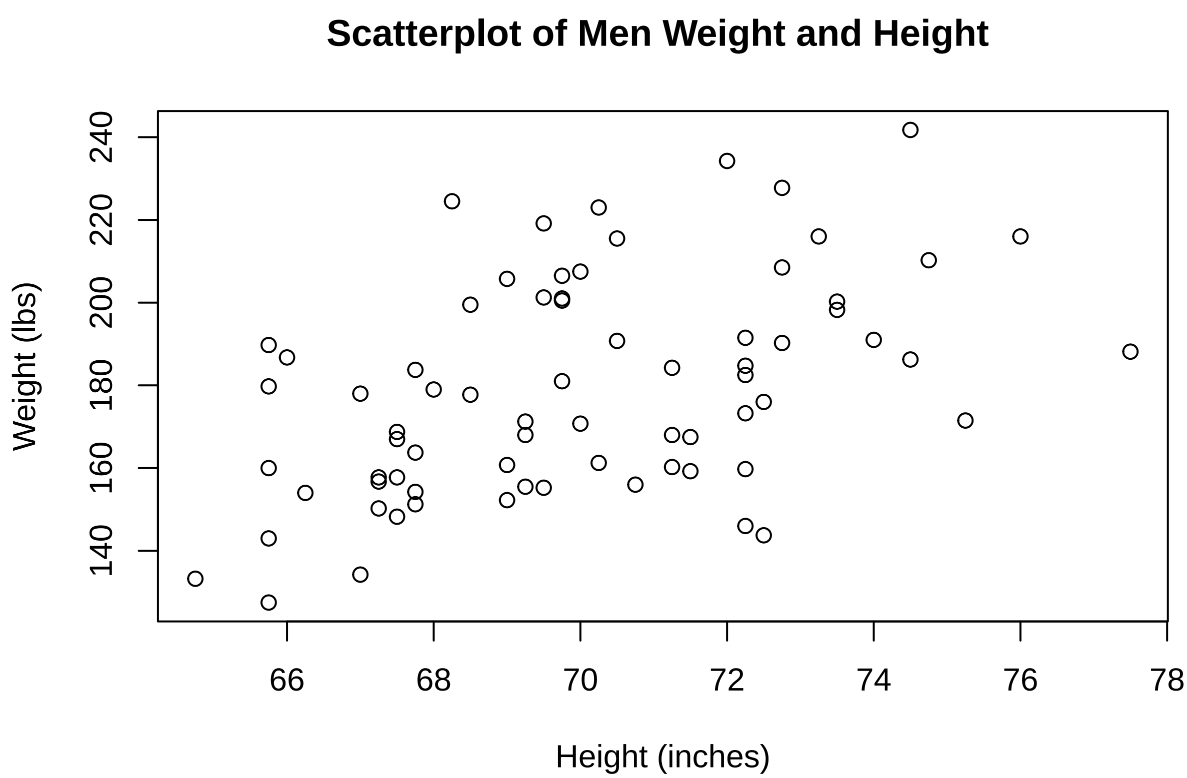
a

To predict weight,Weight is placed on the y-axis and height is place in the x-axis.The values of weight are varied as compared to the height of a person.

b

```
dataset = read.csv(file="dataset/body_composition.csv")
```

```
plot(dataset$height,dataset$weight,main ="Scatterplot of Men Weight and Height ",xlab = "Height (inches)
```



c

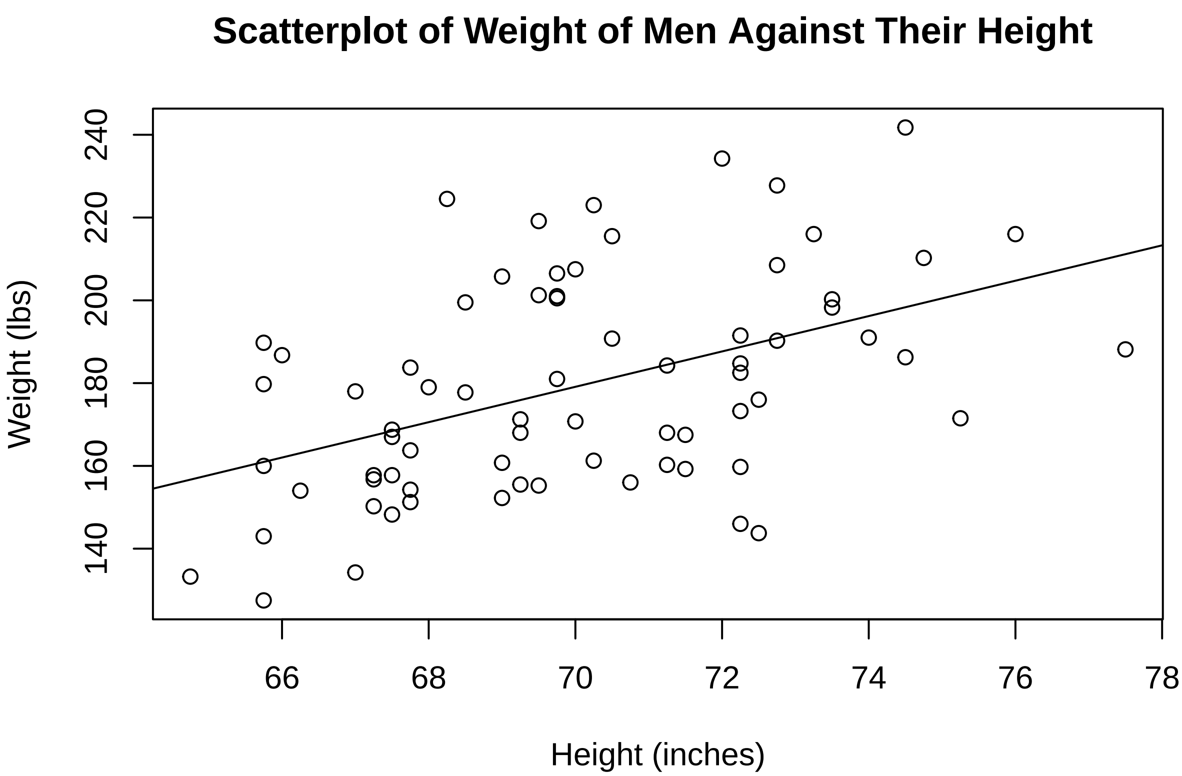
```
plot(
  dataset$height,
  dataset$weight,
  main ="Scatterplot of Weight of Men Against Their Height ",
  xlab = "Height (inches)",
  ylab = "Weight (lbs)"
  )

lm_model = lm(weight~ height , data=dataset)
summary(lm_model)
```

```
##
## Call:
## lm(formula = weight ~ height, data = dataset)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -46.04 -17.97  -4.31  17.93  52.87
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -119.9317    68.3942  -1.754   0.0839 .
## height         4.2720     0.9758   4.378  4.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.19 on 70 degrees of freedom
## Multiple R-squared:  0.2149, Adjusted R-squared:  0.2037
## F-statistic: 19.17 on 1 and 70 DF,  p-value: 4.105e-05
```

```r
abline(lm_model,v=0)
```



The Linear Regression equation for height and weight is:

$$Weight = -119.9317 \ + 4.272 \times Height + \epsilon$$

d

A slope of 4.272 represent the estimated change in weight for every one inch of height.The slope is positive hence there is a positive linear relationship between weight and height.
it is not appropriate to make a interpretation when the height is zero since the no person that has a height of 0.

e

```
predict(lm_model,newdata = data.frame(height=c(76)))
```

```
##        1
## 204.7376
```

Q5

$$MSE(\overline{Y_{str}}) = 0.36\frac{\sigma_A^2}{60} + 0.16\frac{\sigma_B^2}{40}$$
$$MSE(\overline{Y}) = 0.6\frac{\sigma_A^2}{100} + 0.4\frac{\sigma_b^2}{40} + 0.24\frac{(\mu_A - \mu_B)^2}{100}$$

Q6

Given

$$X_i \ N(\theta, \sigma^2)$$
$$E(X_i) = \theta$$
$$V(X_i) = \sigma^2$$

a

Y is an unbiased estimator of θ, the $E(Y) = \theta$

$$E(Y) = \left(\frac{E(X_1 + X_2)}{2}\right)$$
$$= \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2)$$
$$= \frac{1}{2}\theta + \frac{1}{2}\theta$$
$$= \theta$$

Y is an unbiased estimator of θ

b

$$Var(Y) \geq \frac{-1}{n\int_{-\infty}^{+\infty}\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right)f(x;\theta)dx}$$

T The normal p.d.f with $\mu = \theta$

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$$

Take the logarithm of the probability density function

$$\ln f(x;\theta) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)\right)$$

$$= -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\theta)^2}{2\sigma^2}$$

Determine the second derivative with respect to θ

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = \frac{-2(x-\theta)}{2\sigma^2}$$

$$= \frac{x-\theta}{\sigma^2}$$

$$\frac{\partial^2 \ln f(x;\theta)}{\partial^2\theta} = \frac{1}{\sigma^2}$$

Determining the Rao-Cramer lower bound:

$$\int_{-\infty}^{+\infty} \left(\frac{\partial^2 \ln f(x;\theta)}{\partial^2\theta}\right) f(x;\theta)dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma^2} f(x;\theta)dx$$

$$= -\frac{1}{\sigma^2}\int_{-\infty}^{+\infty} f(x;\theta)dx$$

$$= -\frac{1}{\sigma^2}.1 = -\frac{1}{\sigma^2}$$

Solving Rao-Creamer lower bound:

$$Var(Y) \geq \frac{-1}{n\int_{-\infty}^{+\infty}\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right) f(x;\theta)dx}$$

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

$$= \frac{\sigma^2}{n}$$

$$Var(Y) \geq \frac{\sigma^2}{n}$$

c

Determining the variance of Y

$$Var(Y) = Var\left(\frac{X_1 + X_2}{2}\right)$$

$$= \left(\frac{1}{2}\right)Var(X_1) + \left(\frac{1}{2}\right)Var(X_2)$$

$$= \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2$$

$$= \frac{\sigma^2}{2}$$

Efficiency is the Rao-Creamer lower bound to the actual variance of the random variable

$$Efficiency = \frac{Rao - Creamer\ lower\ bound\ of\ Y}{Var(Y)}$$

$$= \frac{\sigma^2/n}{\sigma^2/2}$$

$$= \frac{2}{n}$$