STAT504 Assignment1

Anuradha Ramachandran

2023-01-15

Problem 1

(a) Simulate 100 draws from $Y = 10 + 5X + \epsilon$ where $X \sim N(0, 1)$ and $\epsilon \sim N(0, 1)$

```
# Sample size n = 100

n = 100
# Generate 100 random X and e from No(0,1)
x = rnorm(n = n, mean = 0, sd = 1)
e = rnorm(n = n, mean = 0, sd = 1)
# y = 10+5x+e
y = 10 + 5 * x + e
summary(y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -4.884   6.510  10.610  10.270  13.445  23.634
```

(b) Fit OLS model by regressing Y on X

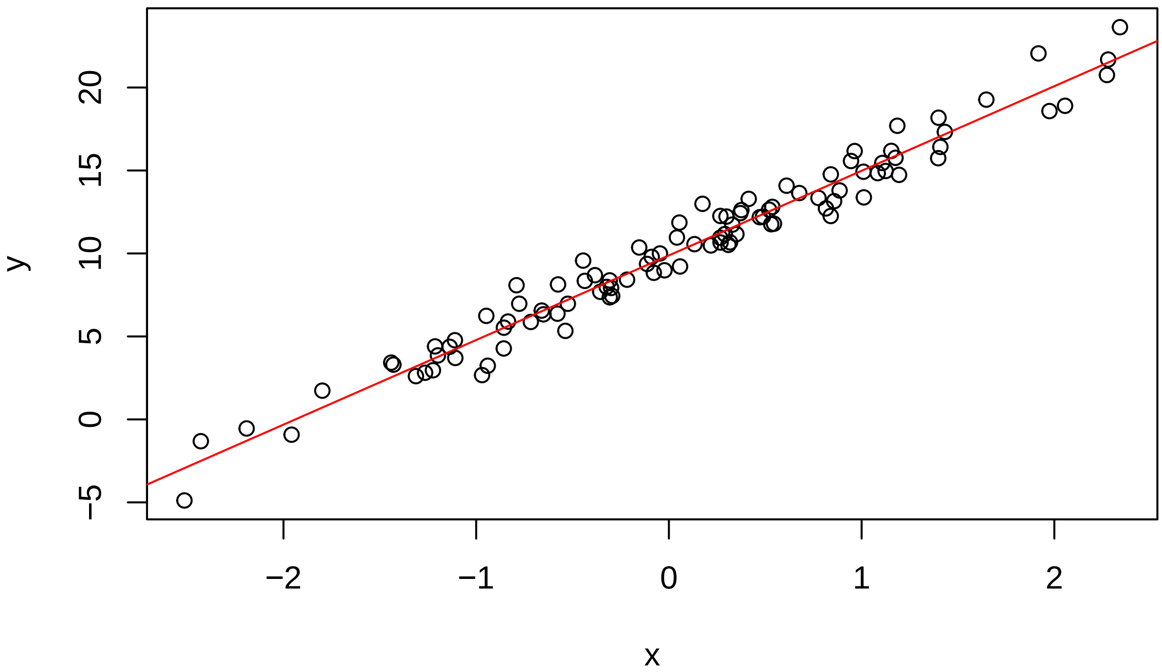
The coefficients from the regression is: Intercept = 10 Slope = 5

```
# OLS model for Y on X
ols = lm(y ~ x)
ols
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##       9.883        5.098
```

(c) Scatterplot of X & Y with the regression line

```
plot(y ~ x)
abline(ols, col = "red")
```



Problem 2

2(a)

Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$
Event space S = {all subsets of Ω} = 2^6
Probability measure of event $A \in S = P(A) = \frac{sizeofA}{sizeofomega} = \frac{|A|}{6}$
$P(\phi) = 0$
$P(\Omega) = 1$
$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$

2(b)

Let D=Democrat, R=Republican, I=Independent.
Sample space Ω = {R, I, D}, $|\Omega| = 1000$ since $|R| = 200, |D| = 400, |I| = 400$
Event space S = {all subsets of Ω} = {ϕ, R,I,D,{R,I},{R,D},{I,D},{R,I,D}}
Probability measure of event $A \in S = P(A) =$ Total people in event A/1000
$P(\phi) = 0$
$P(\Omega) = 1$
$P(D) = P(I) = \frac{400}{1000} = \frac{2}{5}$
$P(R) = \frac{200}{1000} = \frac{1}{5}$
$P(D, I) = \frac{800}{1000} = \frac{4}{5}$
$P(D, R) = \frac{600}{1000} = \frac{3}{5}$
$P(R, I) = \frac{600}{1000} = \frac{3}{5}$
$P(D, R, I) = \frac{1000}{1000} = 1$

Problem 3

3.1

(a) $E[X] = \int_{-\infty}^{\infty} xf(x)dx$ if x is continuous

(b) $Var(X) = E[(X - EX)^2]$

(c) From definition of variance, $var(X) = E[(X - EX)^2]$
$var(X) = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]^2 + E[X]^2$
$var(X) = E[X^2] - E[X]^2$

(d) $SD(X) = \sqrt{var(X)}$

(e) Assume X is discrete a random variable. Let $Y = g(x)$
$E[Y] = \sum\limits_{y \in Y} yP(Y = y)$
$E[Y] = \sum\limits_{y \in Y} yP(x = g^{-1}(y))$
Since, $P(x = g^{-1}(y)) = f_X(x)$
$E[Y] = \sum\limits_{y \in Y} \sum\limits_{x = g^{-1}(y)} yf_X(x)$
$E[Y] = \sum\limits_{x} g(x)f_X(x)$

(f) $E[a + bX] = \int_{-\infty}^{\infty} (a + bx)f(x)dx$
$E[a + bX] = a \int_{-\infty}^{\infty} f(x)dx + b \int_{-\infty}^{\infty} xf(x)dx$
Since $\int_{-\infty}^{\infty} f(x)dx = 1$
$E[a + bX] = a + b \int_{-\infty}^{\infty} xf(x)dx$
$E[a + bX] = a + bE[X]$

(g) By definition, $var(a + bX) = E[(a + bX - E[a + bX])^2]$
$var(a + bX) = E[(a + bX - a - bE[X])^2]$
$var(a + bX) = E[b^2(X - E[X])^2] = b^2E[(X - EX)^2] = b^2var(X)$

(h) $SD[a + bX] = \sqrt{var(a + bX)}$
From (g), $SD[a + bX] = \sqrt{b^2var(X)} = |b|SD(X)$

3.2

Markov's inequality: Let X be a random variable that takes only non negative values, then for any a >0, $P(X \geq a) \leq \frac{E[X]}{a}$, provided E[X] exists.

Proof: Let X be a continuous random variable. $E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$
$E[X] = \int_{-\infty}^{0} xf_X(x)dx + \int_{0}^{\infty} xf_X(x)dx$
$E[X] \geq \int_{a}^{\infty} xf_X(x)dx$, since $a \geq 0$
$E[X] \geq a \int_{a}^{\infty} f_X(x)dx = aP(X \geq a)$ Hence, $P(X \geq a) \leq \frac{E[X]}{a}$

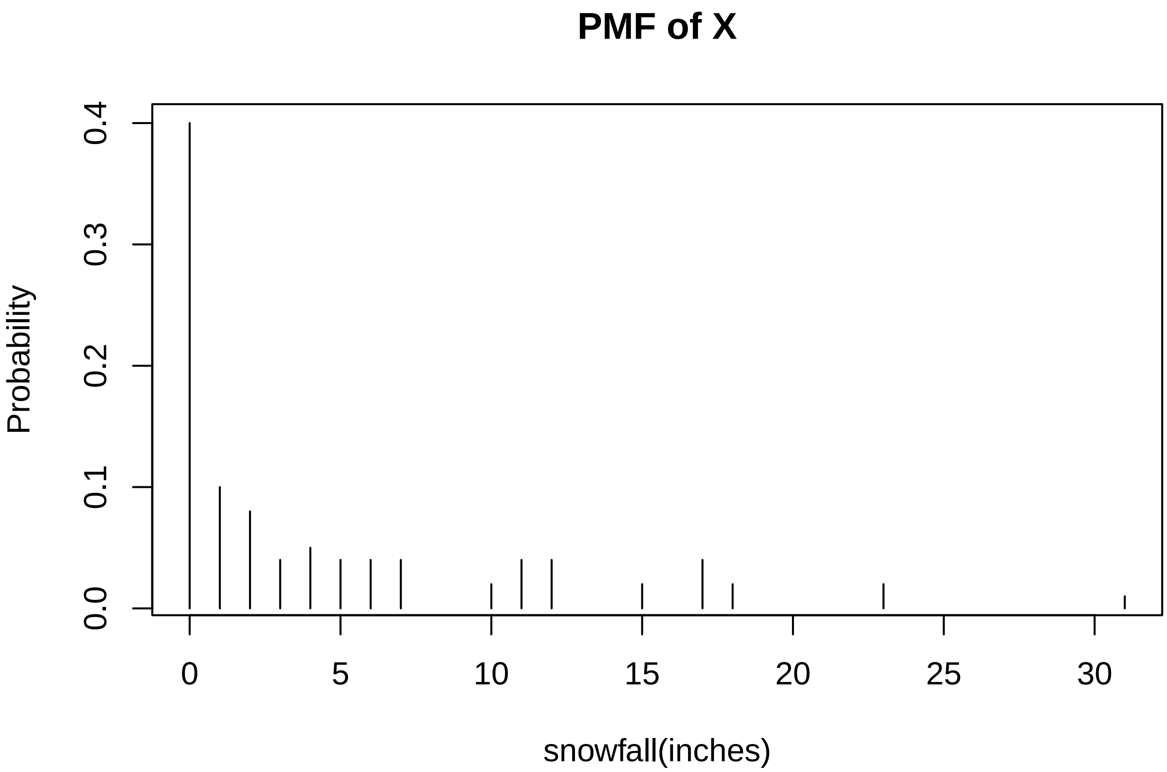
Chebychev's inequality: Let X be a random variable with finite variance then for any $\epsilon > 0$,
$P(|X - EX| \geq \epsilon) \leq \frac{var(X)}{\epsilon^2}$

Proof: Consider $(X - EX)^2$ to a random variable and it is strictly positive.
By applying Markov's inequality for any $\epsilon > 0$
$P((X - EX)^2 \geq \epsilon^2) \leq \frac{E[(X-EX)^2]}{\epsilon^2}$
Since $(X - EX)^2 \geq \epsilon^2 \implies |X - EX| \geq \epsilon$
By (b), $P(|X - EX| \geq \epsilon) \leq \frac{var(X)}{\epsilon^2}$

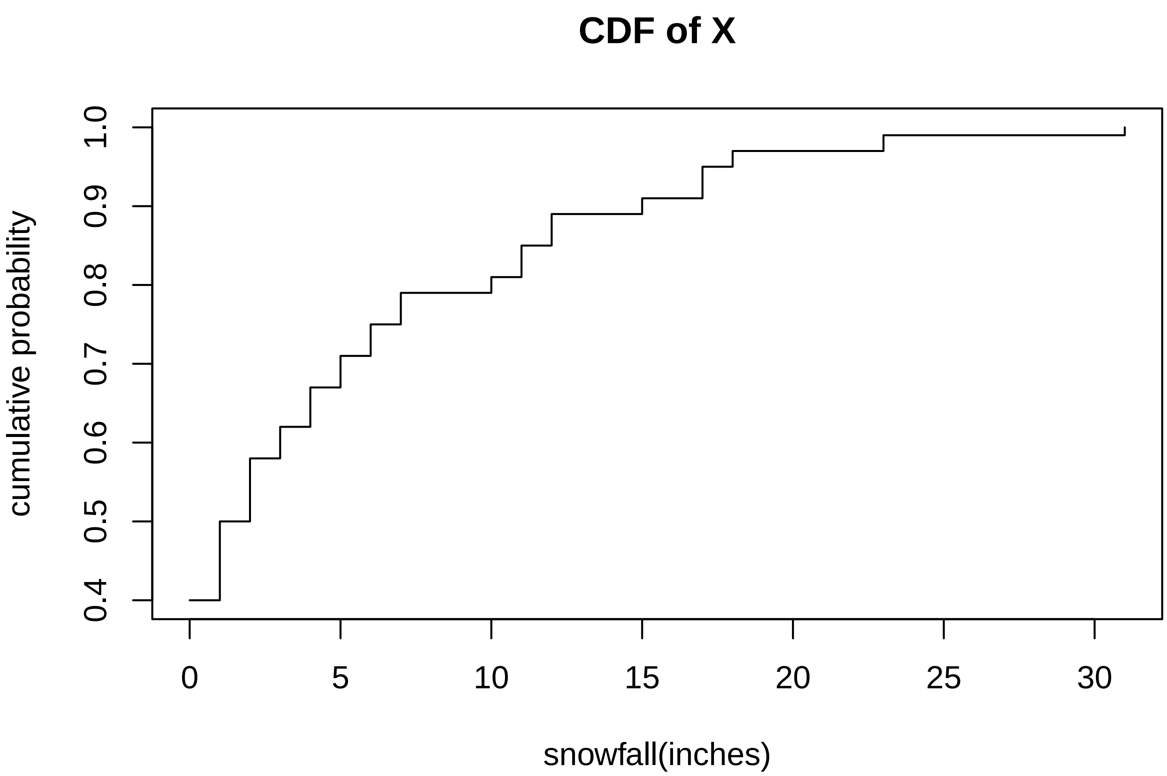
According to Chebychev's inequality, the probability that the absolute deviation of a random variable from its mean will exceed a threshold ϵ times standard deviation is less than or equal to $\frac{1}{\epsilon^2}$.

3.3(a) PMF and CDF of X

```
# Load CSV
snow = read.csv("/Users/anuram/Library/Mobile Documents/com~apple~CloudDocs/MS Stats/Winter 2023/STAT504
# PMF of X
pmf = plot(snow$prob ~ snow$snowfall, type = "h", xlab = "snowfall(inches)",
    ylab = "Probability", main = "PMF of X")
```




```
# cdf of X
snow$CDF = cumsum(snow$prob)
plot(snow$CDF ~ snow$snowfall, type = "s", xlab = "snowfall(inches)",
    ylab = "cumulative probability", main = "CDF of X")
```



3.3(b) Mean, median, mode, variance of X and 95% percentile of X

$E[X] = \Sigma_x xP(X = x) = 4.53$

Median(X) is $m for which P(X \leq m) = 0.5$.
Hence Median(X) = 1

Mode(X) is the X that has highest probability. Mode of this distribution is $X = 0$ inches

$Var(X) = E[X^2] - EX^2 = 40.79$

From the CDF, it can been that the 95% percentile of X is 17 inches.

```
# mean of X
mean = sum(snow$snowfall * snow$prob)
mean
```

```
## [1] 4.53
```

```
median = subset(snow$snowfall, snow$CDF == 0.5)
median
```

```
## [1] 1
```

```
# mode of X
mode = subset(snow$snowfall, snow$prob == max(snow$prob))
mode
```

```
## [1] 0
```

```
# variance of X
exp_sq = sum((snow$snowfall^2) * snow$prob)
variance = exp_sq - mean^2
variance
```

```
## [1] 40.7891
```

```
# 95% percentile of X
percentile = subset(snow, snow$CDF >= 0.95)
percentile
```

```
##     X snowfall prob  CDF
## 13 13       17 0.04 0.95
## 14 14       18 0.02 0.97
## 15 15       23 0.02 0.99
## 16 16       31 0.01 1.00
```

3.3(c) Odds of snowing

Odd of snowing = $\frac{P(snowfall>0)}{P(snowfall=0)} = 1.5$

```
# odds of snowing
nosnow = subset(snow$prob, snow$snowfall == 0)
odds = (1 - nosnow)/nosnow
odds
```

```
## [1] 1.5
```

3.3(d)

The best predictors could be any of the summary statistics like mean, median, mode calculated above depending on the definition of the loss function. Eg: If MSE is the loss function, then mean(X) would be the best predictor that minimizes MSE and gives the best prediction of snowfall.

3.3(e) MSE when E(X) is used

$MSE = E[(X - EX)^2] = var(X) = 40.7891$

3.3(f) 95% prediction interval

95% Prediction Interval = $[E[X] - 1.96 * SD[X], E[X] + 1.96 * SD[X]]$
Lower_limit = -7.987. Since X represents snowfall in inches, the lowest value it can take is 0. Hence the lower limit in the prediction interval is 0. Upper_limit = 17 inches

95% prediction interval of snowfall = [0,17]

```
Lower_limit = mean - 1.96 * sqrt(variance)
Lower_limit
```

```
## [1] -7.987804
```

```
Upper_limit = mean + 1.96 * sqrt(variance)
Upper_limit
```

```
## [1] 17.0478
```

Problem 4

4(a)

$E[(X - c)^2] = E[X^2 - 2cX + c^2] = E[X^2] - 2cE[X] + c^2$
$E[(X - c)^2] = E[X^2] - E[X]^2 + E[X]^2 - 2cE[X] + c^2$
$E[(X - c)^2] = E[X^2] - E[X]^2 + (E[X] - c)^2$
$E[(X - c)^2] = var(X) + (E[X] - c)^2$

4(b)

Let $Y = argmin_{c \in R} E[(X-c)^2]$
From 4(a), $Y = argmin_{c \in R} var(X) + (E[X]-c)^2$
Differentiating Y with respect to c and equating to 0, $c = E[X]$
Hence E[X] is the best predictor of X when MSE is the loss function.

4(c)

X is continuous random variable. Let $\phi = E[|X-c|] = \int_{-\infty}^{c}(c-x)f(x)dx + \int_{c}^{\infty}(x-c)f(x)dx$
Differentiating ϕ with respect to c and equating to 0 by using Leibniz's rule, $\frac{d\phi}{dc} = \int_{-\infty}^{c} \frac{\partial d}{\partial c}(c-x)f(x)dx + \int_{c}^{\infty} \frac{\partial d}{\partial c}(x-c)f(x)dx$
$\frac{d\phi}{dc} = \int_{-\infty}^{c} f(x)dx - \int_{c}^{\infty} f(x)dx = 0$
$\implies P(X \leq c) = P(X > c)$
But $P(X \leq c) + P(X > c) = 1$
$\implies P(X \leq c) = P(X > c) = 1/2$
Hence Median[X] is the best predictor when mean absolute error is the loss function.

4(d)

Mode(X) is the value $x \in X$ for which the marginal distribution of X (PDF if X is continuous or PMF if X is discrete) is maximum. $Mode(X) = argmax_{x \in X} P(X = x)$
$\implies$ the value of c that maximizes $P(X = c)$ is Mode(X). Hence $Mode[X] = argmax_{c \in R} P(X = c)$

Problem 5

5.1

(a) From definition of expectation, $E[a + bX + cY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(a + bX + cY)f(x,y)dxdy$
$E[a + bX + cY] = a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy + c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy$
since,$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$
$\int_{-\infty}^{\infty} f(x,y)dy = f(x)$
$\int_{-\infty}^{\infty} f(x,y)dx = f(y)$
$\implies E[a + bX + cY] = a + b\int_{-\infty}^{\infty} xf(x)dx + c\int_{-\infty}^{\infty} yf(y)dy$
$\implies E[a + bX + cY] = a + bE[X] + cE[Y]$

(b) $E[Y|X = x] = \int_{-\infty}^{\infty} yf(y|x)dy$
$E[Y|X = x]$ is the expected values of Y given that a certain set of $X = x$ is known to occur.

(c) $var[Y|X = x] = E[(Y - E[Y|X])^2|X]$
$var[Y|X = x]$ is the variance of Y given that a certain set of $X = x$ is known to occur.

(d) $cov(X,Y) = E[(X - EX)(Y - EY)]$

(e) From (d), $cov(X,Y) = E[XY - XEY - YEX + EXEY]$
$cov(X,Y) = E[XY] - EXEY - EYEX + EXEY$
$cov(X,Y) = E[XY] - EXEY$

From the above results, $cov(X,X) = E[X^2] - E[X]^2 = var(X)$

(f) $cov(bX, cY) = E[bXcY] - E[bX]E[cY]$
$cov(bX, cY) = bcE[XY] - bcE[X]E[Y] = bc(E[XY] - E[X]E[Y]) = bcCov(X, Y))$

(g) $var(a + bX + cY) = var(bX + cY) = E[(bX + cY)^2] - E[(bX + cY)]^2$
By expanding and grouping terms we get,
$var(a + bX + cY) = b^2[E[X^2] - E[X]^2] + c^2[E[Y^2] - E[Y]^2] + 2bc(EXY - EXEY)$
$var(a + bX + cY) = b^2var(X) + c^2var(Y) + 2bcCov(X, Y)$

(h) $cov(Y + X, Z) = E[(Y + X)Z] - E[Y + X]EZ$
$cov(Y + X, Z) = E[YZ + XZ] - E[Y]E[Z] - E[X]E[Z]$
$cov(Y + X, Z) = E[YZ] - E[Y]E[Z] + E[XZ] - E[X]E[Z]$
$cov(Y + X, Z) = cov(Y, Z) + cov(X, Z)$

(i) $cor(X, Y) = \frac{cov(X,Y)}{SD(X)SD(Y)}$

(j) $cor(a + bX, c + dY) = \frac{cov(a+bX,c+dY)}{SD(a+bX)SD(c+dY)}$
$cov(a + bX, c + dY) = E[b(X - EX)d(Y - EY)] = bdcov(X, Y)$
$SD(a + bX) = |b|SD(X)$
$SD(c + dY) = |d|SD(Y)$
$cor(a + bX, c + dY) = \frac{bdcov(X,Y)}{|bd|SD(X)SD(Y)}$

5.2(a)

Joint distribution is the probability of two events occurring together. $P(X, Y) = P(X \cap Y)$

5.2(b)

```
# load the data set
income = read.csv("/Users/anuram/Library/Mobile Documents/com~apple~CloudDocs/MS Stats/Winter 2023/STAT5
# setting the column name and row name of the dataframe
# names(income)[1] = 'y'
colnames(income) = c("y", "0.5", "1.5", "2.5", "3.5", "4.5",
    "5.5", "6.7", "8.8", "12.5", "17.5")
rownames(income) = income$y
income = income[, -1]

# PMF of Y
PMF_Y = rowSums(income)

# marginal distribution of Y table
dist_Y = data.frame(as.list(rowSums(income)))
colnames(dist_Y) = rownames(income)
rownames(dist_Y) = "P(Y=y)"
dist_Y
```

```
##          0.5   0.4  0.25  0.15  0.05     0 -0.05 -0.18 -0.25
## P(Y=y) 0.07 0.065 0.098 0.208 0.302 0.029 0.095  0.07 0.063
```

```r
# PMF of X
PMF_X = numeric(ncol(income))
for (i in (1:ncol(income))) {
    PMF_X[i] = sum(income[i])
}


# marginal distribution of X table
dist_X = cbind(X = as.numeric(colnames(income)), `P(X=x)` = as.numeric(PMF_X))
dist_X
```

```
##          X P(X=x)
##  [1,]  0.5  0.041
##  [2,]  1.5  0.093
##  [3,]  2.5  0.093
##  [4,]  3.5  0.082
##  [5,]  4.5  0.113
##  [6,]  5.5  0.103
##  [7,]  6.7  0.155
##  [8,]  8.8  0.155
##  [9,] 12.5  0.113
## [10,] 17.5  0.052
```

5.2(c) Conditional distribution of Y given X for all values of X=x

$P(Y|X=x) = \frac{P(X=x,Y=y)}{P(X=x)} \forall x \in X$

```r
# Conditional distribution of Y given X=x for all value of
# X=x
cond_dist = data.frame(matrix(ncol = ncol(income), nrow = nrow(income)))
for (i in (1:ncol(income))) {
    cond_dist[i] = income[i]/PMF_X[i]
}
rownames(cond_dist) = rownames(income)
colnames(cond_dist) = colnames(income)
# conditional distribution table for all Y=y given X=x
cond_dist
```

```
##              0.5        1.5        2.5        3.5         4.5        5.5
## 0.5   0.02439024 0.11827957 0.07526882 0.07317073 0.044247788 0.04854369
## 0.4   0.02439024 0.02150538 0.06451613 0.08536585 0.088495575 0.06796117
## 0.25  0.04878049 0.06451613 0.04301075 0.08536585 0.088495575 0.10679612
## 0.15  0.04878049 0.09677419 0.09677419 0.14634146 0.141592920 0.19417476
## 0.05  0.24390244 0.24731183 0.35483871 0.37804878 0.362831858 0.28155340
## 0     0.31707317 0.13978495 0.00000000 0.02439024 0.008849558 0.00000000
## -0.05 0.02439024 0.12903226 0.11827957 0.06097561 0.106194690 0.15533981
## -0.18 0.04878049 0.08602151 0.13978495 0.07317073 0.079646018 0.07766990
## -0.25 0.21951220 0.09677419 0.10752688 0.07317073 0.079646018 0.06796117
##              6.7        8.8       12.5       17.5
## 0.5   0.05161290 0.05806452 0.12389381 0.07692308
## 0.4   0.05161290 0.05806452 0.07079646 0.13461538
## 0.25  0.12903226 0.12258065 0.11504425 0.11538462
## 0.15  0.27096774 0.34838710 0.21238938 0.38461538
## 0.05  0.30322581 0.25161290 0.37168142 0.13461538
## 0     0.00000000 0.00000000 0.00000000 0.00000000
```

```
## -0.05 0.10967742 0.09032258 0.03539823 0.05769231
## -0.18 0.05161290 0.05161290 0.05309735 0.03846154
## -0.25 0.03225806 0.01935484 0.01769912 0.05769231
```

5.2(d) Conditional expectation of Y given X for all values of X=x

$E[Y|X=x] = \Sigma_y yP(Y|X=x) \forall x \in X$

```
y = as.numeric(rownames(cond_dist))
cond_exp = numeric(ncol(income))
for (i in (1:ncol(income))) {
    cond_exp[i] = sum(y * cond_dist[i])
}
x = as.numeric(colnames(cond_dist))

# conditional expectation of Y given X=x for all values of
# X=x
condition_exp = cbind(X = x, `P(Y|X=x)` = cond_exp)
condition_exp
```

```
##          X    P(Y|X=x)
##  [1,]  0.5 -0.01121951
##  [2,]  1.5  0.06462366
##  [3,]  2.5  0.04849462
##  [4,]  3.5  0.09841463
##  [5,]  4.5  0.07946903
##  [6,]  5.5  0.08262136
##  [7,]  6.7  0.11167742
##  [8,]  8.8  0.12909677
##  [9,] 12.5  0.15371681
## [10,] 17.5  0.16134615
```

Expectation of Y, $E[Y] = \Sigma yP(Y=y) = 0.0987$

```
# Expectation of Y, E[Y]
mean_y = sum(y * PMF_Y)
mean_y
```

```
## [1] 0.0987
```

$E[E[Y|X]] = \Sigma_x E[Y|X=x]P(X=x) = 0.0987$

```
# Expectation of Y, E[E[Y|X]]
exp_y = sum(cond_exp * PMF_X)
exp_y
```

```
## [1] 0.0987
```

Hence it's proved that $E[E[Y|X]] = E[Y] = 0.0987$

5.2(e) Best Linear predictor (BLP) of Y given X

BLP of Y given X is $Y = \alpha + \beta X$
$\alpha = E[Y] - \frac{Cov(X,Y)}{var(X)}E[X] = 0.0432$
$\beta = \frac{Cov(X,Y)}{var(X)} = 0.0086$
BLP of $Y = 0.0432 + 0.0086X$

```r
# Expectation of X
x = as.numeric(colnames(cond_dist))
mean_x = sum(x * PMF_X)
mean_x
```

```
## [1] 6.4795
```

```r
# variance of X
exp_xsq = sum(x^2 * PMF_X)
var_x = exp_xsq - (mean_x^2)
var_x
```

```
## [1] 17.76773
```

```r
# Cov(X,Y) = E[XY] - E[X]E[Y]
exp_xy = 0
for (i in (1:ncol(income))) {
    for (j in (1:nrow(income))) {
        exp_xy = exp_xy + x[i] * y[j] * income[j, i]
    }
}
cov_xy = exp_xy - (mean_x * mean_y)
cov_xy
```

```
## [1] 0.1520084
```

```r
# intercept alpha
alpha = mean_y - (cov_xy * mean_x/var_x)
alpha
```

```
## [1] 0.0432659
```

```r
# slope beta
beta = cov_xy/var_x
beta
```

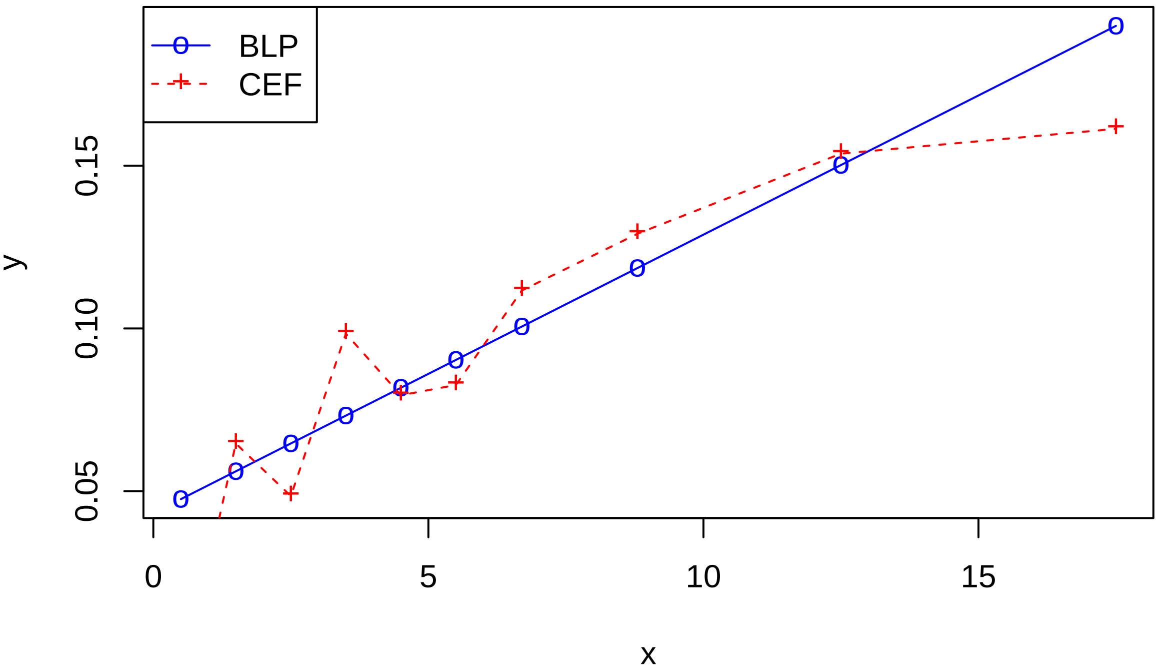
```
## [1] 0.008555305
```

```r
# Y from Best Linear Predictor
y_blp = alpha + beta * x
blp = cbind(X = x, Y_blp = y_blp)
blp
```

```
##          X      Y_blp
##  [1,]  0.5 0.04754355
##  [2,]  1.5 0.05609886
##  [3,]  2.5 0.06465416
##  [4,]  3.5 0.07320947
##  [5,]  4.5 0.08176477
##  [6,]  5.5 0.09032008
##  [7,]  6.7 0.10058644
##  [8,]  8.8 0.11855259
##  [9,] 12.5 0.15020721
## [10,] 17.5 0.19298374
```

5.2(f) Plot of BLP and CEF(=E[Y|X])

```
# Plotting BLP and CEF together plot x vs BLP in blue line
plot(x, y_blp, type = "o", col = "blue", pch = "o", ylab = "y",
    xlab = "x", lty = 1)
# adding x vs CEF in red line to the previous plot
points(x, cond_exp, col = "red", pch = "+")
lines(x, cond_exp, col = "red", lty = 2)
legend(x = "topleft", legend = c("BLP", "CEF"), col = c("blue",
    "red"), pch = c("o", "+"), lty = c(1, 2), ncol = 1)
```



Problem 6

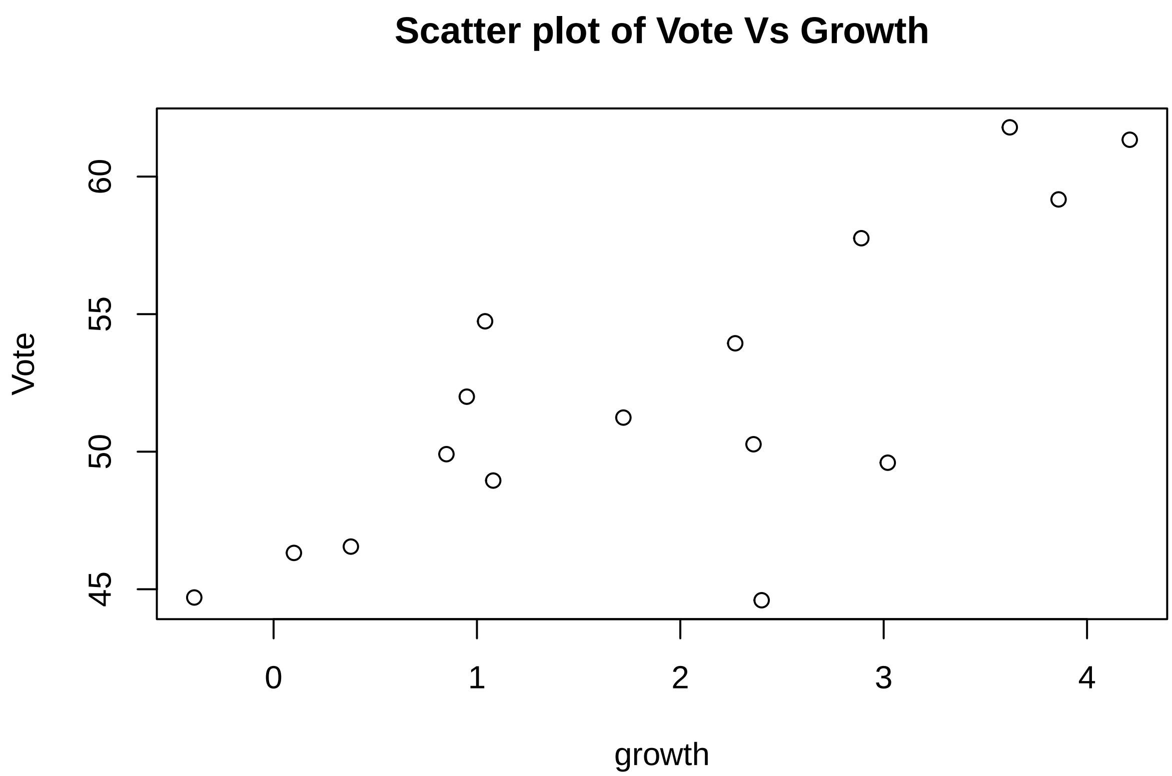
6(a)

Yes, there could be a positive correlation between income growth and incumbent vote%.As the income increases the confidence in the incumbent party's economic policy might get stronger and hence they might get higher share of vote.

6(b) Scatter plot of vote Vs growth

```
election = read.delim("/Users/anuram/Library/Mobile Documents/com~apple~CloudDocs/MS Stats/Winter 2023/S
    sep = " ")

# Scatter plot of vote vs growth
plot(election$growth, election$vote, main = "Scatter plot of Vote Vs Growth",
    xlab = "growth", ylab = "Vote")
```



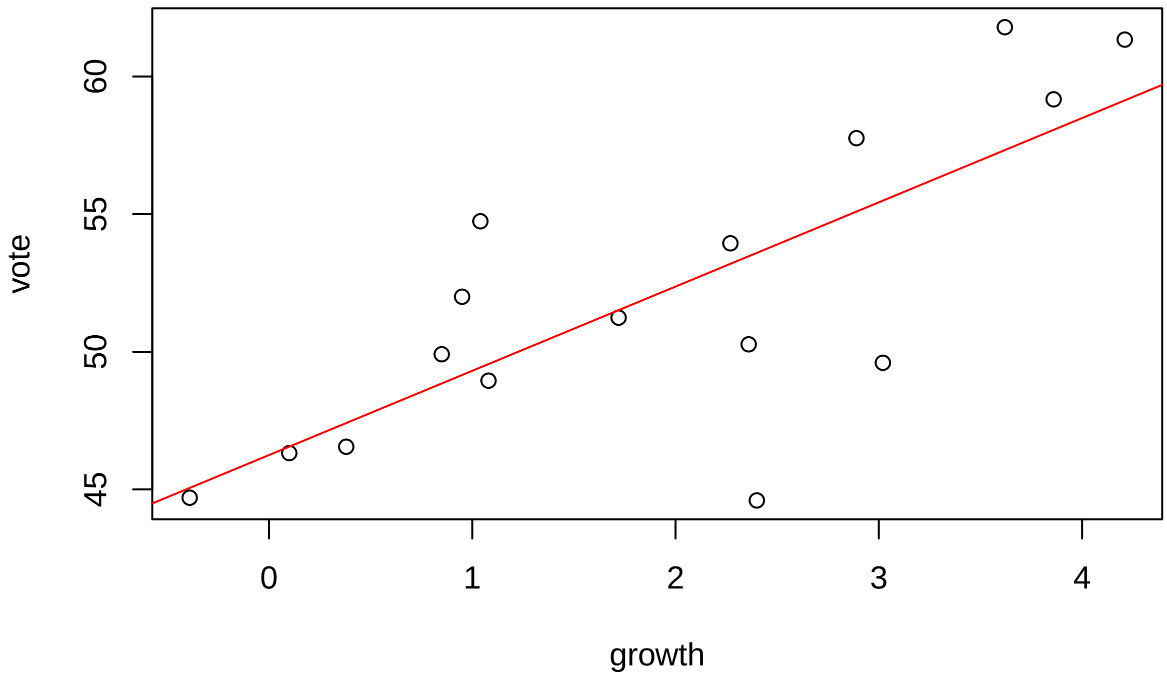
Simple linear regression model of Y on X Intercept: 46.25 Slope: 3.06

```
fit = lm(election$vote ~ election$growth)
summary(fit)
```

```
##
## Call:
## lm(formula = election$vote ~ election$growth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9929 -0.6674  0.2556  2.3225  5.3094
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      46.2476     1.6219  28.514 8.41e-14 ***
## election$growth   3.0605     0.6963   4.396  0.00061 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.763 on 14 degrees of freedom
## Multiple R-squared:  0.5798, Adjusted R-squared:  0.5498
## F-statistic: 19.32 on 1 and 14 DF,  p-value: 0.00061
```

The regression line seems to predict the trend of the data (ie, positive correlation between vote and growth), but adjusted R-squared is 0.55.

```
# plot with regression line
plot(election$vote ~ election$growth, ylab = "vote", xlab = "growth")
abline(fit, col = "red")
```



6(c) Regression model summary

The predicted regression model is $Vote = 46.25 + 3.06 * Growth$
From the regression line, it can interpreted that as growth in income increases by 1% the incumbent party's vote percentage increases by 3.06%.

The estimated regression coefficients are : Intercept: 46.25 This means that when there is no income growth in the previous years, the incumbent party's vote percentage on average would be 46.25%

Slope:3.06 Since the slope is positive, it indicates a positive linear relationship between income growth and incumbent party's vote %. Since slope is defined as $\frac{changeiny}{changeinx}$, if income grows by 1% in the previous years, it can predicted that the incumbent party's vote percentage would increase by 3.06%.

6(d) Prediction when average income growth is 2%

Given: if average income growth is 2%, from the regression model the incumbent party's vote percentage is 52.37%.
Vote% $= 46.25 + 3.06 * 2 = 52.37$
The actual vote% for the incumbent party was 51.1% in 2016 when the income growth was 2%. Hence the linear regression model is similar to the actual vote% observed.