Assignment 1 - Probability Review

Jeffrey Chang, JLC190004

1/20/2020

Question 4:

In the U.S. the North and South are quite different. Below is the joint probability distribution of political affiliation for a Northern city and a Southern city. (R = Republican, I = Independent and D = Democrat)

```
# Political Affiliation= PA, R= Republican, I= Independent, D= Democrat
PA = matrix( c(0.24, 0.04, 0.12, 0.18, 0.12, 0.3), nrow=2, ncol=3, byrow = TRUE)
colnames(PA) <- c("R","I","D")
rownames(PA) <- c("Southern","Northern")
PA <- as.table(PA)
PA
```

```
##             R    I    D
## Southern 0.24 0.04 0.12
## Northern 0.18 0.12 0.30
```

```
#Caculate the Marginal Probability
addmargins(PA)
```

```
##             R    I    D  Sum
## Southern 0.24 0.04 0.12 0.40
## Northern 0.18 0.12 0.30 0.60
## Sum      0.42 0.16 0.42 1.00
```

(a) What is the probability of selecting a Republican given that we sample from the Northern city? Show your calculation.

```
#P(Northern|Republican)
p1 <- 0.18/0.6
round(c(p1), digits = 2)
```

```
## [1] 0.3
```

(b) Are political affiliation and region of residence statistically independent random variables? Explain.

```
# Test whether  R, I, D statistically independent ? P(R)=0.42, P(I)=0.16, P(D)=0.42
# whether P(Southern|R)=P(R) ?
print(paste0("P(Southern|R)= ", round(0.24/0.42, 2)))
```

```
## [1] "P(Southern|R)= 0.57"
```

```
# whether P(Southern|I)=P(I) ?
print(paste0("P(Southern|I)= ", round(0.04/0.16, 2)))
```

```
## [1] "P(Southern|I)= 0.25"
```

```
# whether P(Southern|D)=P(D) ?
print(paste0("P(Southern|D)= ", round(0.12/0.42, 2)))
```

```
## [1] "P(Southern|D)= 0.29"
```

```r
# whether P(Northern|R)=P(R) ?
print(paste0("P(Northern|R)= ", round(0.18/0.42, 2)))
```

```
## [1] "P(Northern|R)= 0.43"
```

```r
# whether P(Northern|I)=P(I) ?
print(paste0("P(Northern|I)= ", round(0.12/0.16, 2)))
```

```
## [1] "P(Northern|I)= 0.75"
```

```r
# whether P(Northern|D)=P(D) ?
print(paste0("P(Northern|D)= ", round(0.30/0.42, 2)))
```

```
## [1] "P(Northern|D)= 0.71"
```

Explanation:

P(Southern|R) is not equaled to P(R), P(Southern|I) is not equaled to P(I) and P(Southern|D) is not equaled to P(D) also same like Northern, so political affiliation and region of residence are not statistically independent.

(c) Assign the values R = 0, I = 2 and D = 5 to political affiliation (PA). That is, if a citizen is selected at random, the variable PA can take the values 0, 2 and 5. Find the mathematical expectation of the random variable (PA).

```r
#E(PA)=0*P(R)+2*P(I)+5*P(D)
E <- 0*0.42+2*0.16+5*0.42
E
```

```
## [1] 2.42
```

(d) Find the expected value of X = 2(PA) + 2(PA)^2, where PA is the random variable political affiliation.

```r
#E(PA^2) =E2
E2 <- 0^2*0.42+2^2*0.16+5^2*0.42
X <- 2*E+2*E2
X
```

```
## [1] 27.12
```

Question 12:

Based on many years of experience, an instructor in econometrics has determined that the probability distribution of X, the number of students absent on Mondays, is as follows:

```
# f(x) = y
ab <- data.frame("x" = c(0, 1, 2, 3, 4, 5, 6,7),
                 "y" = c(0.02, 0.03, 0.26, 0.34, 0.22, 0.08, 0.04, 0.01))
ab1 <- t(ab)
ab1
```

```
##   [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## x 0.00 1.00 2.00 3.00 4.00 5.00 6.00 7.00
## y 0.02 0.03 0.26 0.34 0.22 0.08 0.04 0.01
```

(a) Sketch the probability function of X.

```
plot(ab, main = "The Probability Function of X")
```

(b) Find the probability that on a given Monday either 2, or 3 or 4 students will be absent.

```
sum(ab1[2,c(3:5)])
```

```
## [1] 0.82
```

(c) Find the probability that on a given Monday more than 3 students are absent.

```r
sum(ab1[2,c(5:8)])
```

```
## [1] 0.35
```

(d) Compute the expected value of the random variable X. Interpret this expected value.

```r
weighted.mean(ab1[1,],ab1[2,])
```

```
## [1] 3.16
```

The average number of students absent on Monday is 3.16.

(e) Compute the variance and standard deviation of the random variable X.

```r
# Variance = E(X^2)-u^2
# E(X^2) = 1*0.03+2^2*0.26+3^2*0.34+4^2*0.22+5^2*0.08+6^2*0.04+7^2*0.01
Ex2 <- 1*0.03+2^2*0.26+3^2*0.34+4^2*0.22+5^2*0.08+6^2*0.04+7^2*0.01
Var=Ex2-3.16^2
print(paste0("Variance is ", Var))
```

```
## [1] "Variance is 1.5944"
```

```r
print(paste0("Standard Deviation is ", round(c(Var^(1/2)), 4)))
```

```
## [1] "Standard Deviation is 1.2627"
```

(f) Compute the expected value and variance of $Y = 7X + 3$.

```r
# E(7X+3)=7E(X)+3
# Var(7X + 3)=7^2V(X)
print(paste0("Expected value of Y = (7X + 3) is ", 7*3.16+3))
```

```
## [1] "Expected value of Y = (7X + 3) is 25.12"
```

```r
print(paste0("The variance of Y = (7X + 3) is ", round(c(7^2*Var), 4)))
```

```
## [1] "The variance of Y = (7X + 3) is 78.1256"
```

Question 13:

Suppose a certain mutual fund has an annual rate of return that is approximately normally distributed with mean (expected value) 5% and standard deviation 4%. Use Table 1, the table of cumulative probabilities for the standard normal distribution, for parts (a)–(c).

(a) Find the probability that your 1-year return will be negative.

```
#P(X<0) = P(Z<(0-0.05)/0.04=-1.25) = P(Z<-1.25)
round(pnorm(0, mean=0.05, sd=0.04, lower.tail=TRUE), 4)
```

```
## [1] 0.1056
```

(b) Find the probability that your 1-year return will exceed 15%.

```
#P(X>0.15) = P(Z<(0.15-0.05)/0.04=2.5) = P(Z>2.5)
round(pnorm(0.15, mean=0.05, sd=0.04, lower.tail=FALSE), 4)
```

```
## [1] 0.0062
```

(c) If the mutual fund managers modify the composition of its portfolio, they can raise its mean annual return to 7%, but will also raise the standard deviation of returns to 7%. Answer parts (a) and (b) in light of these decisions. Would you advise the fund managers to make this portfolio change?

Modify the mean to 0.07 and standard deviation to 0.07

```
# mean = 0.07, sd = 0.07
#P(X<0)
print(paste0("The probability that your 1-year return will be negative is ",
             round(pnorm(0, mean=0.07, sd=0.07, lower.tail=TRUE), 4) ))
```

```
## [1] "The probability that your 1-year return will be negative is 0.1587"
```

```
#P(X>0.15)
print(paste0("The probability that your 1-year return will exceed 15% is ",
             round(pnorm(0.15, mean=0.07, sd=0.07, lower.tail=FALSE), 4) ))
```

```
## [1] "The probability that your 1-year return will exceed 15% is 0.1265"
```

After the calculation above,

1. The probability of a negative return has increased by 5.31%.
2. The probability of a return greater than 15% has increased by 12.03%.

So I think this situation depends on how much risk that managers can accept.

(d) Verify your computations in (a)–(c) using your computer software.

Please refer to the above calculation process

Question 14:

An investor holding a portfolio consisting of two stocks invests 25% of assets in Stock A and 75% into Stock B. The return R_a from Stock A has a mean of 4% and a standard deviation of sd_a = 8%. Stock B has an expected return E(R_b) = 8% with a standard deviation of sd_b = 12%. The portfolio return is P = 0.25*R_a + 0.75*R_b.

(a) Compute the expected return on the portfolio.

```
# E(P)= E(0.25*R_a + 0.75*R_b) = 0.25*E(R_A)+0.75*E(R_b)
print(paste0("The expected return on the portfolio is ",
             c(0.25*0.04+0.75*0.08)))
```

```
## [1] "The expected return on the portfolio is 0.07"
```

(b) Compute the standard deviation of the returns on the portfolio assuming that the two stocks' returns are perfectly positively correlated.

Var(P)= Var(0.25*R_a + 0.75*R_b)= (0.25^2)*var(R_a)+(0.75^2)*Var(R_b)+2*0.25*0.75*cov(R_a, R_b)

The question defined "perfectly positively correlated", so p=1

p = cov(R_a, R_b)/[var(R_a)^(1/2)]*[var(R_b)^(1/2)]=1

thus, cov(R_a, R_b) = $(0.08^{2)}(1/2)*(0.12^{2)}(1/2)$

```
cov <- (0.08^2)^(1/2)*(0.12^2)^(1/2)*1
Var_p <- (0.25^2)*(0.08^2)+(0.75^2)*(0.12^2)+2*0.25*0.75*cov
print(paste0("The standard deviation of the returns is ",
             Var_p^(1/2)))
```

```
## [1] "The standard deviation of the returns is 0.11"
```

(c) Compute the standard deviation of the returns on the portfolio assuming that the two stocks' returns have a correlation of 0.5.

The question defined "perfectly positively correlated", so p=0.5

p = cov(R_a, R_b)/[var(R_a)^(1/2)]*[var(R_b)^(1/2)]=0.5

```
cov2 <- (0.08^2)^(1/2)*(0.12^2)^(1/2)*0.5
Var_p2 <- (0.25^2)*(0.08^2)+(0.75^2)*(0.12^2)+2*0.25*0.75*cov2
print(paste0("The standard deviation of the returns is ",
             round(Var_p2^(1/2),4)))
```

```
## [1] "The standard deviation of the returns is 0.1015"
```

(d) Compute the standard deviation of the returns on the portfolio assuming that the two stocks' returns are uncorrelated.

The question defined "perfectly positively correlated", so p=0

p = cov(R_a, R_b)/[var(R_a)^(1/2)]*[var(R_b)^(1/2)]=0

```
cov3 <- (0.08^2)^(1/2)*(0.12^2)^(1/2)*0
Var_p3 <- (0.25^2)*(0.08^2)+(0.75^2)*(0.12^2)+2*0.25*0.75*cov3
print(paste0("The standard deviation of the returns is ",
             round(Var_p3^(1/2),4)))
```

```
## [1] "The standard deviation of the returns is 0.0922"
```

Question 18:

```r
library(jpeg)
library(ggplot2)
library(reshape)
readImage<-readJPEG('Q18_S.jpg')
longImage<-melt(readImage)
rgbImage<-reshape(longImage,timevar='X3',idvar=c('X1','X2'),direction='wide')
rgbImage$X1<- -rgbImage$X1
colorColumns<- rgbImage[, substr(colnames(rgbImage), 1, 5)== "value"]
with(rgbImage,plot(X2, X1, col = rgb(colorColumns), asp = 1, pch =".",axes=F,xlab='',ylab=''))
```

(a) $\frac{x_1+x_2+x_3+x_4}{4} = 3$

(b) $(1-3)+(3-3)+(5-3)+(3-3)$

$= 0$

(c) $(1-3)^2+(3-3)^2+(5-3)^2+(3-3)^2$

$= 8$

(d) $(1+9+25+9)-4\cdot 3^2 = 8$

(e) set $\sum_{i=1}^{n} = \sum$

$\Rightarrow \sum x_i^2 - 2\sum x_i\bar{x} + \sum \bar{x}^2$

$\Rightarrow \sum x_i^2 - 2n\bar{x}\left(\frac{\sum x_i}{n}\right) + n\bar{x}^2$

$\Rightarrow \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$

$\Rightarrow \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$ #

Question 19:

```
readImage<-readJPEG('Q19_S.jpg')
longImage<-melt(readImage)
rgbImage<-reshape(longImage,timevar='X3',idvar=c('X1','X2'),direction='wide')
rgbImage$X1<- -rgbImage$X1
colorColumns<- rgbImage[, substr(colnames(rgbImage), 1, 5)== "value"]
with(rgbImage,plot(X2, X1, col = rgb(colorColumns), asp = 1, pch =".",axes=F,xlab='',ylab=''))
```

$$\sum_{i=1}^{n}\left(x_i y_i - \bar{x} y_i - \bar{y} x_i + \bar{x}\bar{y}\right) \quad , \text{ set } \sum_{i=1}^{n} = \Sigma$$

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

$$= \Sigma x_i y_i - n\bar{x}\left(\frac{1}{n}\Sigma y_i\right) - n\bar{y}\left(\frac{1}{n}\Sigma x_i\right) + n\bar{x}\bar{y}$$

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

$$= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \quad \#$$