Sectional Homework 1

Group 3

1/28/2021

Group #: 3

Group Leader and ID: Brianna Johnson (01722540)

Member Names and IDs: Marco Sousa (01747512), Ben Pfeffer (01791790), Nikita Seleznev (01938236)

Sectional Written Homework #1: (25 points):

- (6 points; 2 points * 3) Consider two random variables X (0 = male; 1 = female) and Y (0= low risk; 1= medium risk; 2 = high risk) with a joint pmf given in the Table below.

Table Joint pmf of X and Y:

	Y=0	Y=1	Y=2
X=0	1/25	1/10	1/5
X= 1	2/5	4/25	1/10

- P(X=Female,Y=High Risk) = P(X=1,Y=2) = 1/10
- P(X=Female)=P(X=1)=2/5+4/25+1/10 = 22/50+8/50+5/50 = 33/50
- P(Y=High Risk | X=Female) = P(X=Female,Y=High Risk)/ P(X=Female) = (1/10)/(33/50) = 5/33
- (3 points) Suppose a success of a medical trial, X (yes/no), follows a true binomial population distribution. We randomly draw samples of size, n=30, from this binomial population distribution, with the probability of success =0.2. Answer the following questions:
  1. Given this sample size (n=30), will the CLT hold true? Show your steps to prove if the CLT holds true.
  2. If the CLT holds true, what is the mean of these sample means? What is the standard deviation of these sample means?

1. CLT holds true for Binomial Distribution sampling if min(np,n(1-p)) $\geq$ 5. Thus: min(30 $\times$ 0.2, 30(1-0.2)) = min(6,24) = 6 $\geq$ 5. This condition holds true. Thus, CLT will hold true.
2. μ = np = 30(0.2) = 6. $\sigma_x = \frac{\sigma_{pop}}{\sqrt{n}} = \sqrt{(30 \times .2 \times .8)}/\sqrt{30} = \sqrt{4.8}/\sqrt{30} = 0.4$

Note that $\sigma^2_{pop} = np(1-p)$.

- (4 points) Let's assume the number of spams follows the Poisson distribution and we randomly draw samples of size =40, with the mean of sample means of 6. Estimate the population mean and the population standard deviation based on CLT. Show you steps and no scores are assigned if only answers are provided.

$\mu = \lambda$, and $\sigma^2 = \lambda$, generally. $\mu_{sample} = 6$ and $\sigma^2_{sample} = 6$.

$\mu_{pop} = \mu_{sample} = \lambda = 6$. Thus $\mu_{pop} = 6$.

$\sigma_{sample} = \sigma_{pop}/\sqrt{n}$. $\sigma^2_{sample} = (\sigma_{pop}/\sqrt{n})^2 = \sigma^2_{pop}/n$

$6 = \sigma^2_{pop}/n$. $40 \times 6 = 240 = \sigma^2_{pop}$

But this is variance. We want standard deviation: $\sqrt{240} = \sigma_{pop} \approx 15.491$

Thus, final answers: $\mu_{pop} = 6, \sigma_{pop} \approx 15.4919$ are population estimates based on the sample.

- (12 points; 4 points*3) Use R to
  - Randomly generate 10 values (n=10) for each of three random variables (RVs), x1, x2, and x3, using a random seed = 490, from a multivariate Gaussian distribution, with a population mean vector of [2, 4, 6] (ie., mean of X1= 2, mean of X2= 4, mean of X3= 6) and a population covariance matrix of these RVs (shown below):

	X1	X2	X3
X1	4	3	2
X2	3	3	1
X3	2	1	2

Note: please use the exact random seed = 490, so we can replicate your values;

Reference: R: https://astrostatistics.psu.edu/su07/R/html/MASS/html/mvrnorm.html

Please past your R code and data output below:

```
#Construction of our covariance matrix and vector means. Setting seed
Sigma <- matrix(c(4,3,2,3,3,1,2,1,2),3,3)
means <- c(2, 4, 6)
set.seed(490)
#Sampling using mvrnorm.
samples = MASS::mvrnorm(n=10, means, Sigma)
samples
```

```
##             [,1]     [,2]     [,3]
##  [1,] -0.1524272 1.852485 5.791038
##  [2,]  2.4377223 6.095741 4.364562
##  [3,]  4.3934862 5.863520 8.713240
##  [4,]  1.5636755 4.975932 5.954129
##  [5,]  2.7477722 3.302398 7.658680
##  [6,]  4.0393360 7.089616 5.923624
##  [7,]  3.7774719 6.635451 7.535122
##  [8,]  0.4034914 3.233714 3.859881
##  [9,]  5.9677875 7.966577 6.710664
## [10,]  1.5275238 3.296755 4.920883
```

- Use relevant R functions to compute Expected Value (ie., mean), Variance, Standard Deviation, Mode, Median, Skewness and Kurtosis using the data generated for x1, x2 and x3 in Problem (a), respectively. Please past your R code for computing each statistic, and corresponding R output below:

```r
#Defining a mode function
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

library(moments);
```

```r
means <- apply(samples, 2, mean)
means
```

```
## [1] 2.670584 5.031219 6.143182
```

```r
vars <- apply(samples, 2, var)
vars
```

```
## [1] 3.626638 4.065381 2.360043
```

```r
sds <- apply(samples, 2, sd)
sds
```

```
## [1] 1.904373 2.016279 1.536243
```

```r
modes <- apply(samples, 2, Mode)
modes
```

```
## [1] -0.1524272  1.8524852  5.7910381
```

```r
medians <- apply(samples, 2, median)
medians
```

```
## [1] 2.592747 5.419726 5.938876
```

```r
skewness(samples)
```

```
## [1]  0.1334386 -0.1198388  0.1265788
```

```r
kurtosis(samples)
```

```
## [1] 2.113405 1.726293 2.043652
```

- Use relevant R functions to compute the correlation and covariance matrix of x1, x2 and x3, with the data generated from Problem (a). Please paste your R code and output below:

```r
cov(samples)
```

```
##          [,1]      [,2]      [,3]
## [1,] 3.626638 3.3796399 1.7608224
## [2,] 3.379640 4.0653812 0.9362867
## [3,] 1.760822 0.9362867 2.3600429
```

```r
cor(samples)
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.8801723 0.6018711
## [2,] 0.8801723 1.0000000 0.3022723
## [3,] 0.6018711 0.3022723 1.0000000
```