HW1

Minh Luc, Devin Pham, Kyle Moore

Friday of Week 1, 04/01/2022

This is a template `Rmarkdown` file. Please find your partners and form a group for the assignments and final project. It is better to not change after your group is fixed.

Please read `Math189_HW_template.Rmd` carefully when you write your reports, and use the template file `Math189_HW_template.Rmd` to generate your reports. If you generate your report as a html file, please save it as a pdf file. Please submit your report to **Gradescope** before the deadline.

We all contributed equally for this homework.

Question 0

Please write down you name and ID like below:

Member 1:

- Name: Minh Luc
- Student ID: A17209607

Member 2:

- Name: Kyle Moore
- Student ID: A14271413

Member 3:

- Name: Devin Pham
- Student ID: A17198936

Question 1.

*Let $Z \sim N(0, 1)$.

(a) $P(Z > 1)$;

Answer: The probability can be calculated by

```
p <- pnorm(q = 1,mean = 0,sd = 1,lower = FALSE)
print(p)
```

```
## [1] 0.1586553
```

```
p3 <- round(p, digits = 3) # x rounded to 3 decimal places
print(p3)
```

```
## [1] 0.159
```

Therefore,

$$P(Z > 1) \approx 0.159.$$

(b) $P(Z < -1.06)$;

Answer: The probability can be calculated by

```
p <- pnorm(q = -1.06,mean = 0,sd = 1,lower = TRUE)
print(p)
```

```
## [1] 0.1445723
```

```
p3 <- round(p, digits = 3) # x rounded to 3 decimal places
print(p3)
```

```
## [1] 0.145
```

Therefore,

$$P(Z < -1.06) \approx 0.145.$$

(c) $P(-2.33 < Z \leq 1.06)$.

Answer: The probability can be calculated by

```
p <- pnorm(q = 1.06,mean = 0,sd = 1,lower = TRUE) - pnorm(q = -2.33,mean = 0,sd = 1,lower = TRUE)
print(p)
```

```
## [1] 0.8455246
```

```
p3 <- round(p, digits = 3) # x rounded to 3 decimal places
print(p3)
```

```
## [1] 0.846
```

Therefore,

$$P(-2.33 < Z \leq 1.06) \approx 0.846.$$

Question 2.

Let (X, Y) be a random vector whose pdf is given by

$$f_{X,Y}(x, y) = \frac{2}{3}(x + 2y), \quad 0 \leq x \leq 1, 0 \leq y \leq 1.$$

(a) *Find the marginal distribution of X.*

Answer: The marginal distribution of X is given by

$$
\begin{aligned}
f_X(x) &= \int_0^1 f_{X,Y}(x,y)dy \\
&= \int_0^1 \left(\frac{2}{3}(x+2y)\right)dy \\
&= \frac{2}{3}x\int_0^1 dy + \frac{2}{3}\int_0^1 2ydy \\
&= \frac{2}{3}x \times 1 + \frac{2}{3} \times y^2\Big|_0^1 \\
&= \frac{2}{3}x + \frac{2}{3}, \quad 0 \le x \le 1.
\end{aligned}
$$

Like this, if you want to split a long equation into several lines, you can use `\begin{aligned} ... \end{aligned}` environment or `\begin{split} ... \end{split}` environment inside the `$$ ... $$` environment. Please note that blank lines are not allowed in the `$$ ... $$` environment, otherwise the compile will not be successful. The double backslash `\\` at the end of each line works as a newline character. Use the ampersand character &, to set the points where the equations are vertically aligned.

(b) *Find the marginal distribution of Y.*

Answer: The marginal distribution of y is given by

$$
\begin{aligned}
f_Y(y) &= \int_0^1 f_{X,Y}(x,y)dx \\
&= \int_0^1 \left(\frac{2}{3}(x+2y)\right)dx \\
&= \frac{2}{3}\int_0^1 xdx + \frac{4}{3}y\int_0^1 dx \\
&= \frac{2}{3} \times \frac{x^2}{2}\Bigg|_0^1 + \frac{4}{3}y \times 1 \\
&= \frac{2}{3} \times \frac{1}{2} + \frac{4}{3}y \\
&= \frac{1}{3} + \frac{4}{3}y, \quad 0 \le y \le 1.
\end{aligned}
$$

(c) *Find $E(X)$, $E(Y)$, $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$*
Answer: $E(X)$**:**

Since $E(X) = \int x f_X(x)dx$, using the above marginal distribution:

```r
fx = function(x){
  x*((2/3)*x +(2/3)) # x times marginal fx
}

mean_x = integrate(fx, lower=0, upper=1)
print(round(mean_x$value, digits=3))
```

```
## [1] 0.556
```

$$E(X) \approx .556$$

$\mathrm{Var}(X)$**:**

Since $\mathrm{Var}(X) = E(X^2) - E(X)^2$, using the above marginal distribution and $E(X)$:

```r
fx = function(x){
  x*((2/3)*x +(2/3)) # x times marginal fx
}

fx2 = function(x){
  (x^2)*((2/3)*x +(2/3)) # x^2 times marginal fx
}

var_x = (integrate(fx2, lower=0, upper=1)$value -
           (integrate(fx, lower=0, upper=1)$value)^2)
print(round(var_x, digits=3))
```

```
## [1] 0.08
```

$$\text{Var}(X) \approx .080$$

$E(Y)$:

Since $E(Y) = \int y f_Y(y) dy$, using the above marginal distribution:

```r
fy = function(y){
  y*((1/3) +(4/3)*y) # y times marginal fy
}

mean_y = integrate(fy, lower=0, upper=1)
print(round(mean_y$value, digits=3))
```

```
## [1] 0.611
```

$$E(Y) \approx .611$$

$\text{Var}(Y)$:

Since $\text{Var}(Y) = E(Y^2) - E(Y)^2$, using the above marginal distribution and $E(Y)$:

```r
fy = function(y){
  y*((1/3) +(4/3)*y) # y times marginal fy
}

fy2 = function(y){
  (y^2)*(((1/3) +(4/3)*y)) # y^2 times marginal fy
}

var_y = (integrate(fy2, lower=0, upper=1)$value -
           (integrate(fy, lower=0, upper=1)$value)^2)
print(round(var_y, digits=3))
```

```
## [1] 0.071
```

$$\text{Var}(Y) \approx .071$$

(d) *Find* $\text{Cov}(X, Y)$
Answer: We will use the equation $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

$$
\begin{aligned}
E(XY) &= \int_0^1 \int_0^1 xy f_{X,Y}(x,y)dxdy \\
&= \int_0^1 \int_0^1 xy\left(\frac{2}{3}x + \frac{4}{3}y\right)dxdy \\
&= \frac{1}{3}\int_0^1 \int_0^1 \left(2x^2y + 4xy^2\right)dxdy \\
&= \frac{1}{3}\int_0^1 2y\frac{x^3}{3}\bigg|_0^1 + 4y^2\frac{x^2}{2}\bigg|_0^1 dy \\
&= \frac{1}{3}\int_0^1 2y^2 + \frac{2}{3}ydy \\
&= \frac{1}{3}\left(\frac{2}{3}\frac{y^2}{2}\bigg|_0^1 + 2\frac{y^3}{3}\bigg|_0^1\right) \\
&= \frac{1}{3}\left(\frac{1}{3} + \frac{2}{3}\right) \\
&= \frac{1}{3} \approx .333
\end{aligned}
$$

Then plug into the equation:

$$
\begin{aligned}
\operatorname{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\
&= .333 - (.556)(.611) \\
&= .333 - .339716 \\
&= .006716 \approx .007
\end{aligned}
$$

(e) *Find* $\operatorname{Cor}(X,Y)$
Answer:

$$
\begin{aligned}
\operatorname{Cor}(X,Y) &= \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y} \\
&= \frac{.007}{\sqrt{Var(X)}\sqrt{Var(Y)}} \\
&= \frac{.007}{\sqrt{.080}\sqrt{.071}} \\
&= .0928804 \approx .093
\end{aligned}
$$

Question 3.

(a) *Let $n \geq 1$ and let $X_1, \ldots, X_n$ be a sample from $N(\mu, \sigma^2)$, where μ and σ^2 are both unknown. Provide an estimator $\hat{\mu}$ of the unknown parameter μ?*

Answer: An estimator of the population mean is given by

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + \cdots + X_n).$$

(b) *Following (a), provide an estimator $\hat{\sigma}^2$ of the unknown parameter σ^2.*

Let $n \geq 1$ and let $X_1, \ldots, X_n$ be a sample from $N(\mu, \sigma^2)$, where μ and σ^2 are both unknown.
Answer: An estimator of the population variance is given by

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

(c) *Read the R documentation webpage again, and use `rnorm` to generate `n = 10` normal random numbers with `mean = 1` and `sd = 3`.*

```
rnorm(n = 10, mean = 1, sd = 3)
```

```
##  [1]  2.1164778  1.4624571  2.0378571  3.1649453 -2.2011252  1.0000950
##  [7] -1.0079492  1.8294728  0.2100583  0.6715186
```

However, when we re-run the line above, the result is different:

```
rnorm(n = 10, mean = 1, sd = 3)
```

```
##  [1] -0.7028021  4.1274640  1.4627429  3.3318142 -1.1823396 -1.7314206
##  [7]  0.4362805 -4.5916137  0.3446164 -2.4476644
```

To keep the results unchanged, we can set a random seed every time before we use the `rnorm` function:

```
set.seed(1)
rnorm(n = 10, mean = 1, sd = 3)
```

```
##  [1] -0.87936143  1.55092997 -1.50688584  5.78584241  1.98852332 -1.46140515
##  [7]  2.46228716  3.21497412  2.72734405  0.08383484
```

```
set.seed(1)
rnorm(n = 10, mean = 1, sd = 3)
```

```
##  [1] -0.87936143  1.55092997 -1.50688584  5.78584241  1.98852332 -1.46140515
##  [7]  2.46228716  3.21497412  2.72734405  0.08383484
```

Now, choose an arbitrary integer as your own seed, and generate `n = 5` normal random numbers with `mean = 2` and `sd = 1`. Based on (a), find your value of $\hat{\mu}$ using your random numbers and find the difference between $\hat{\mu}$ and the true mean μ.

Answer:

```
mu <- 2
set.seed(2)
rnorm(n = 5, mean = mu, sd = 1)
```

```
## [1] 1.1030855 2.1848492 3.5878453 0.8696243 1.9197482
```

```
set.seed(2)
muHat <- mean(rnorm(n = 5, mean = mu, sd = 1))
diff <- muHat - mu
print(muHat)
```

```
## [1] 1.933031
```

```
print(diff)
```

```
## [1] -0.06696949
```

```
print(round(diff, digits=3))
```

```
## [1] -0.067
```

Value of

$$\hat{\mu} - \mu \approx -0.067$$

(d) *Discuss how can you estimate μ more accurately, and explain your idea.*

We can set n to a higher value and/or set sd to a smaller value. The higher the value of n, the closer we will get to μ, which is stated by the Law of Large Numbers. Setting the sd to a smaller value will reduce the spread of values from μ. Without altering the distribution itself, we could minimize the MSE by setting a to the sample mean in the mean squared error fucntion $MSE(a) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - a)^2$