Universidad Nacional de Ingeniería
Facultad de Ciencias
Escuela Profesional de Matemática — Ciclo 2019–I

SOLUCIONARIO DE LA PRÁCTICA CALIFICADA 1 - CM2H2A

Indicaciones

La siguiente práctica calificada consta de 3 partes, lea detalladamente que indica cada parte. La nota de la practica se calculará en base a la expresión: Nota $= \dfrac{\text{Puntaje obtenido}}{40} \times 20$.

Parte 1. (20 puntos)

Resuelva los siguientes ejercicios en el aula, tiempo de evaluación 2 horas.

1. Suponga que Y_1, Y_2, Y_3 denotan una muestra aleatoria de una distribución exponencial con función de densidad

$$f(y) = \begin{cases} \dfrac{1}{\theta}e^{-y/\theta} & , \quad y > 0 \\ 0 & , \quad \text{en cualquier otro punto.} \end{cases}$$

Considere los siguientes cinco estimadores de θ:

$$\hat{\theta}_1 = Y_1, \quad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \quad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \quad \hat{\theta}_4 = \min(Y_1, Y_2, Y_3), \quad \hat{\theta}_5 = \bar{Y}.$$

a. ¿Cuáles de estos estimadores son insesgados? (3 puntos)

b. Entre los estimadores insesgados. ¿Cuál tiene la varianza más pequeña? (2 puntos)

Solución 1 *a. Los estimadores $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ y $\hat{\theta}_5$ son combinaciones lineales simples de Y_1, Y_2 e Y_3. Por lo tanto, se muestra fácilmente que los cuatro estimadores son insesgados.* *(0.25 ptos c/u.)*

$$\begin{aligned} E[\hat{\theta}_1] &= E[Y_1] = \theta \\ E[\hat{\theta}_2] &= E[\frac{Y_1 + Y_2}{2}] = \frac{2\theta}{2} = \theta \\ E[\hat{\theta}_3] &= E[\frac{Y_1 + 2Y_2}{3}] = \frac{3\theta}{3} = \theta \\ E[\hat{\theta}_5] &= E[\bar{Y}] = E[\frac{Y_1 + Y_2 + Y_3}{3}] = \frac{3\theta}{3} = \theta \end{aligned}$$

En el caso de $\hat{\theta}_4 = \min(Y_1, Y_2, Y_3)$, tenemos que *(1 pto.)*

$$\begin{aligned} F_{\hat{\theta}_4}(y) &= P\{\hat{\theta}_4 \leq y\} \\ &= 1 - P\{\hat{\theta}_4 > y\} \\ &= 1 - P\{Y_1 > y, Y_2 > y, Y_3 > y\}, \quad Y_1, Y_2, Y_3 \text{ independientes} \\ &= 1 - P\{Y_1 > y\}P\{Y_2 > y\}P\{Y_3 > y\} \\ &= 1 - P\{Y_1 > y\}^3 \\ &= 1 - (1 - F_{Y_1}(y))^3 \end{aligned}$$

Derivando $F_{\hat{\theta}_4}(y)$ respecto a y, se tiene que *(0.5 ptos.)*

$$\begin{aligned} f_{\hat{\theta}_4}(y) &= \frac{dF_{\hat{\theta}_4}(y)}{dy} \\ &= 3(1 - F_{Y_1}(y))^2 f_{Y_1}(y), \quad F_{Y_1}(y) = (1 - e^{-y/\theta}) \\ &= \frac{1}{\theta/3}e^{-y/(\theta/3)}. \end{aligned}$$

Por lo tanto, $\hat{\theta}_4 \sim Exp(\theta/3)$, entonces $E[\hat{\theta}_4] = \frac{\theta}{3} \neq \theta$ (sesgado). *(0.5 ptos.)*

b. Calculando la varianza de los estimadores insesgados ($\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ y $\hat{\theta}_5$) son *(0.5 ptos c/u.)*

$$\begin{aligned} V(\hat{\theta}_1) &= \theta^2 \\ V(\hat{\theta}_2) &= \frac{\theta^2}{2} \\ V(\hat{\theta}_3) &= \frac{5\theta^2}{9} \\ V(\hat{\theta}_5) &= \frac{\theta^2}{3} \end{aligned}$$

Por lo tanto, el estimador $\hat{\theta}_5$ es el que posee la menor varianza.

2. Suponga que $\hat{\theta}$ es un estimador para un parámetro θ y $E(\hat{\theta}) = a\theta + b$ para algunas constantes diferentes de cero a y b.

   a. En términos de a, b, y θ, ¿cuál es $B(\hat{\theta})$? (1 punto)

   b. Encuentre una función de $\hat{\theta}$, es decir $\hat{\theta}^*_{(\hat{\theta})}$, que sea un estimador insesgado para θ. (1 puntos)

Solución 2 *a. Por definición de sesgo, $E(\hat{\theta}) = a\theta + b - \theta = (a-1)\theta + b$.* *(1 pto.)*

b. Sea $\hat{\theta}^* = \frac{\hat{\theta}-b}{a}$. *(1 pto.)*

3. Considere el estimador insesgado $\hat{\theta}^*$ que usted propuso en el Ejercicio 2.

   a. Exprese $MSE(\hat{\theta}^*)$ como función de $V(\hat{\theta})$. (2 puntos)

   b. Dé un ejemplo de un valor para a para el cual $MSE(\hat{\theta}^*) < MSE(\hat{\theta})$. (1 punto)

   c. Dé un ejemplo de valores para a y b para los cuales $MSE(\hat{\theta}^*) > MSE(\hat{\theta})$. (1 punto)

Solución 3 *a. Se tiene que $E[\hat{\theta}^*] = \theta$ y $V[\hat{\theta}^*] = V[\frac{\hat{\theta}-b}{a}] = \frac{V[\hat{\theta}]}{a^2}$.* *(1 pto)*
Entonces

$$MSE(\hat{\theta}^*) = V[\hat{\theta}^*] = \frac{V[\hat{\theta}]}{a^2} \qquad (1\ pto.)$$

b. Se sabe que $MSE(\hat{\theta}) = V[\hat{\theta}] + B(\hat{\theta}) = V[\hat{\theta}] + [(a-1)\theta + b]^2$. Un valor, lo suficientemente grande de a para obtener $MSE(\hat{\theta}^*) < MSE(\hat{\theta})$ es $a = 10$. *(1 punto)*

c. Un valor pequeño para a donde $MSE(\hat{\theta}^*) > MSE(\hat{\theta})$ es $a = 5$ y $b = 0$. *(1 punto)*

4. Si Y tiene una distribución binomial con parámetros n y p, entonces $\hat{p}_1 = \frac{Y}{n}$ es un estimador insesgado de p. Otro estimador de p es $\hat{p}_2 = \frac{Y+1}{n+2}$.

   a. Deduzca el sesgo de $\hat{p}_2$. (2 puntos)

   b. Deduzca $MSE(\hat{p}_1)$ y $MSE(\hat{p}_2)$. (2 puntos)

   c. ¿Para qué valores de p es $MSE(\hat{p}_1) < MSE(\hat{p}_2)$? (2 puntos)

Solución 4 *Se sabe que $\hat{p}_1$ es insesgado y que $E[Y] = np$. $E[\hat{p}_2] = \frac{np+1}{n+2}$.*

a. Entonces, $E[\hat{p}_2] = \frac{np+1}{n+2}$. *(1 pto.)*
Luego, $B(\hat{p}_2) = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}$ *(1 pto.).*

b. Se tiene que $MSE(\hat{p}_1) = V[\hat{p}_1] = \frac{p(1-p)}{n}$ ($\hat{p}_1$ es insesgado). *(1 pto.)*
Asimismo, $MSE(\hat{p}_2) = V[\hat{p}_2] + B(\hat{p}_2) = \frac{np(1-p)+(1-2p)^2}{(n+2)^2}$. *(1 pto.)*

c. Considerando la desigualdad

$$\frac{np(1-p) + (1-2p)^2}{(n+2)^2} < \frac{p(1-p)}{n}$$

Se tiene que

$$(8n+4)p^2 - (8n+4)p + n < 0 \qquad (0.5\ ptos.)$$

Igualando a 0 la desigualdad, se tiene que

$$p = \frac{8n + 4 \pm \sqrt{(8n+4)^2 - 4(8n+4)n}}{2(8n+4)} = \frac{1}{2} \pm \sqrt{\frac{n+1}{8n+4}} \qquad (1\ pto.)$$

En consecuencia, p se encuentra alrededor de 0.5, es decir, $p \in \langle \frac{1}{2} - \sqrt{\frac{n+1}{8n+4}}, \frac{1}{2} + \sqrt{\frac{n+1}{8n+4}} \rangle$.(0.5 ptos.)

5. Hemos visto, en clase, que si Y tiene una distribución binomial con parámetros n y p, entonces $\frac{Y}{n}$ es un estimador insesgado de p. Para calcular la varianza de Y, por lo general usamos $n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)$.

   a. Demuestre que el estimador sugerido es un estimador sesgado de $V(Y)$. (2 puntos)

   b. Modifique ligeramente $n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)$ para formar un estimador insesgado de $V(Y)$. (1 punto)

Solución 5 *a. Dada Y una variable aleatoria con distribución binomial, $E[Y]=np$ y $V[Y]=npq$, entonces $E[Y^2]=npq+(np)^2$.* *(1 pto.)*
Luego,

$$E[n(\frac{Y}{n})(1-\frac{Y}{n})]=E[Y]-\frac{1}{n}E[Y^2]=np-pq-np^2=(n-1)pq\neq npq. \qquad (1\ pto.)$$

Por lo tanto, $V[Y]$ es sesgado.

b. Un estimador insesgado podría ser $\hat{\theta}=\left(\frac{n}{n-1}\right)n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)$. *(1 pto.)*

Parte 2. Inferencia computacional. (10 puntos)

Resuelva y programe en R lo solicitado en el script "clase3.R"

Solución 6 *i.* `#Ejercicio1`

```
# El 10% de los articulos producidos por una maquina son defectuosos. Si elige una
# muestra aleatoria con reemplazo de 6 articulos y se define la variable X como el
# numero de articulos defectuosos elegidos. Determine la probabilidad que al menos
# un articulo sea defectuoso

# Solución: ######################################################################

prob=sum(dbinom(1:6,6,0.10))
prob

# o

prob=1-dbinom(0,6,0.10)
prob

> prob
[1] 0.468559
```

ii. `#Ejercicio2`

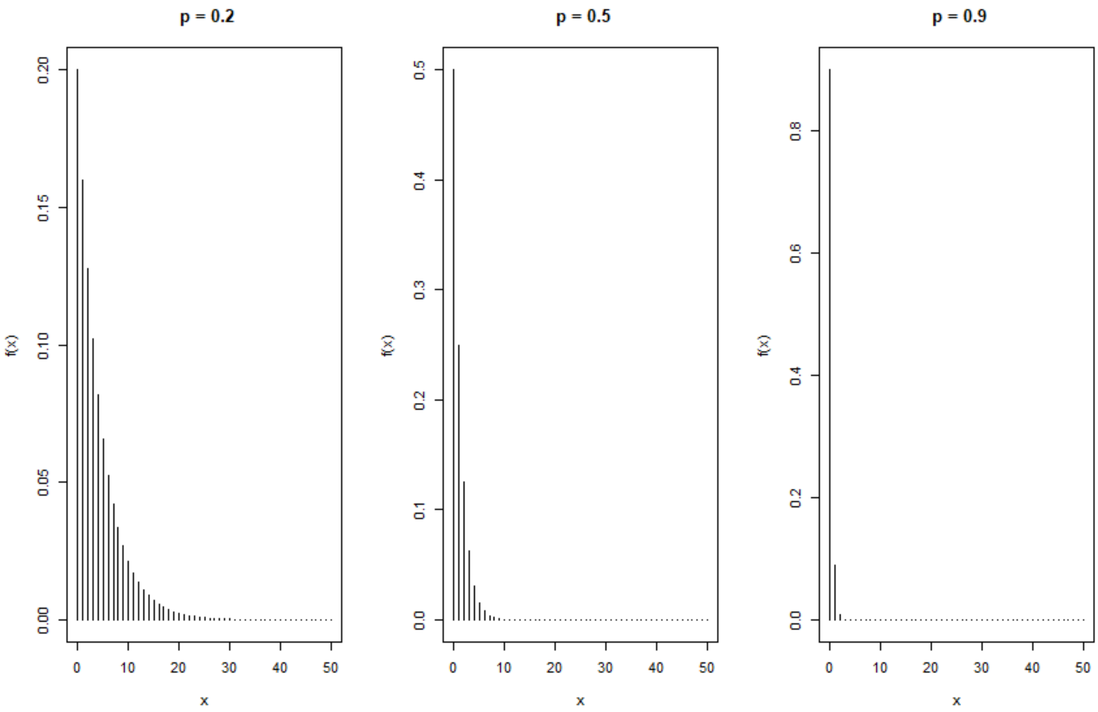
```
# Suponga que existe 20% de probabilidad de que un adulto en una cierta region sufra
# de una enfermedad. Elegimos aleatoriamente una muestra de 25 adultos de esta region.
# Se desea calcular la probabilidad de que a lo mas 3 de las personas seleccionadas
# presenten la enfermedad

# Solución: ######################################################################

prob=pbinom(3,25,0.2)
prob

# o

prob=sum(dbinom(0:3,25,0.2))
prob

> prob
[1] 0.2339933
```

iii. #Ejercicio3

```
# Suponga que un interruptor electrico de una maquina tiene una probabilidad de 0.04
# de fallar, y que cuando ello ocurre es necesario reemplazarlo por uno nuevo. Calcule
# la probabilidad que el interruptor pueda ser usado mas de 100 veces antes de ser
# reemplazado.

# Solución: ##########################################################################

prob=1-sum(dgeom(0:99,0.04))
prob

# o

prob=1-pgeom(99,0.04)
prob

> prob
[1] 0.01687032
```

iv. #Ejercicio4

```
# Graficar 3 distribuciones geometricas diferentes

# Solución: ##########################################################################

# Obtener los valores del rango de x
x=seq(0,50,1)
# Obtener los valores de la funcion de probabilidad
# primero para p=0.2
pdf1=dgeom(x,0.2)
# segundo para p=0.5
pdf2=dgeom(x,0.5)
# tercero para p=0.9
pdf3=dgeom(x,0.9)
# realizar los graficos
par(mfrow=c(1,3)) # pone los graficos en fila
plot(x,pdf1,type="h",xlab="x",ylab="f(x)")
title(" p = 0.2")
plot(x,pdf2,type="h",xlab="x",ylab="f(x)")
title(" p = 0.5")
plot(x,pdf3,type="h",xlab="x",ylab="f(x)")
title(" p = 0.9")
```



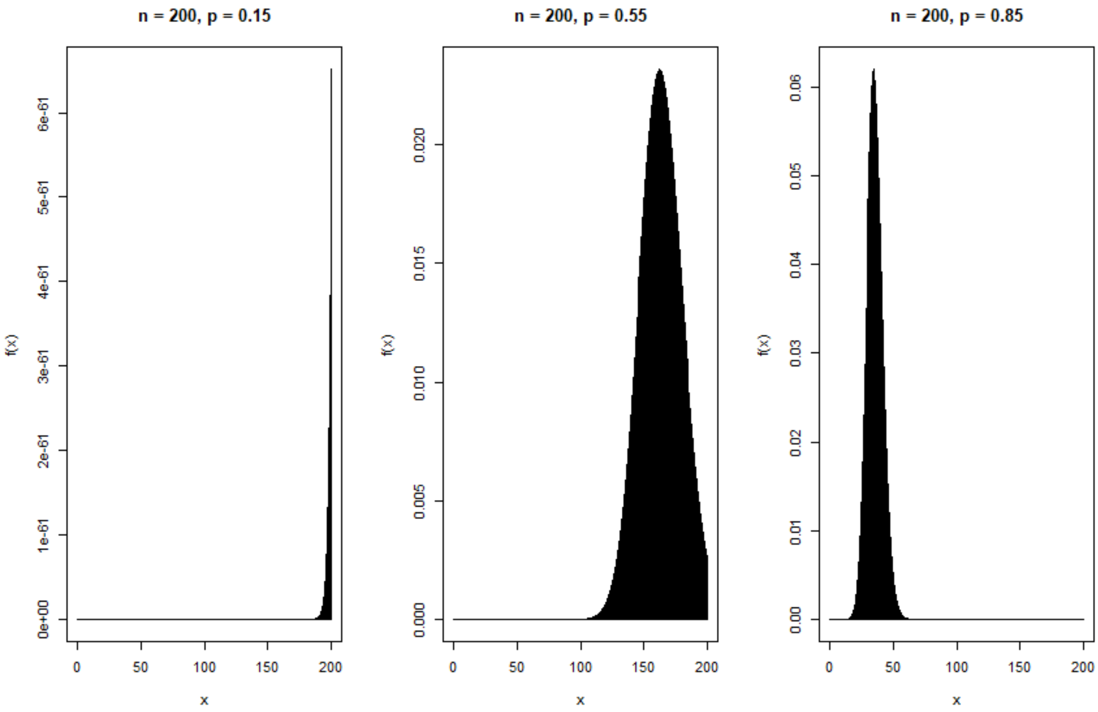
v. #Ejercicio5

```
# Un componente electronico tiene una probabilidad de 0.90 de pasar un control de
# calidad. Se puede asumir que existe independencia entre los resultados del control
# de calidad de diferentes componentes electronicos. Calcule la probabilidad de
# que sea necesario revisar 5 componenetes para obtener que 3 pasen el control de
# calidad.

# Solución: ########################################################################

prob=dnbinom(2,5,0.90)
prob

> prob
[1] 0.0885735
```

vi. #Ejercicio6

```
# Graficar 3 distribuciones binomial negativa diferentes

# Solución: #########################################################################

# Obtener los valores del rango de x
x=seq(0,200,1)
# Obtener los valores de la funcion de probabilidad
# primero para n=200 y p=0.15
pdf1=dnbinom(x,200,0.15)
# segundo para n=200 y p=0.55
pdf2=dnbinom(x,200,0.55)
# tercero para n=200 y p=0.85
pdf3=dnbinom(x,200,0.85)
# realizar los graficos
par(mfrow=c(1,3)) # pone los graficos en fila
plot(x,pdf1,type="h",xlab="x",ylab="f(x)")
title("n = 200, p = 0.15")
plot(x,pdf2,type="h",xlab="x",ylab="f(x)")
title("n = 200, p = 0.55")
```

```
plot(x,pdf3,type="h",xlab="x",ylab="f(x)")
title("n = 200, p = 0.85")
```



vii.
```
#Ejercicio7

# Los clientes de una tienda comercial ingresadn al estacionamiento a razon de 4
# clientes por cada 5 minutos. Si se elige al azar un intervalo de 2 minutos,
# calcule la probabilidad que ingresen al menos dos clientes al establecimiento
# comercial.

# Solución: ######################################################################

prob=1-sum(dpois(0:1,1.6))
prob

# o

prob=1-ppois(1,1.6)
prob

> prob
[1] 0.4750691
```

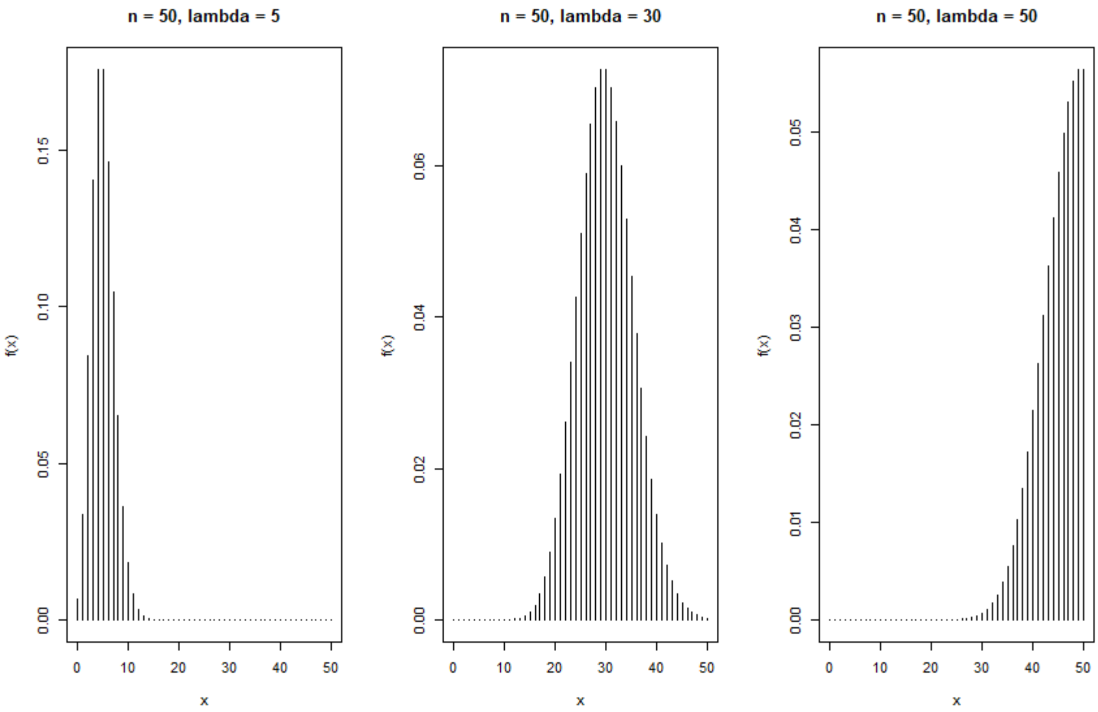
viii.
```
#Ejercicio8

# Suponga que los accidentes en una cierta interseccion siguen un proceso de Poisson
# con una tasa de 2 accidentes por semana. Calcule la probabilidad de que a lo mas
# ocurran 3 accidentes durante las proximas 2 semanas-

# Solución: #######################################################################

prob=ppois(3,2*2)
prob
> prob
[1] 0.4334701
```

xi.
```
#Ejercicio9
```

```
# Halle los cuantiles para la distribucion poisson del problema anterior,
# grafiquelos e interpreta que significan estos.

# Solución: ###################################################################

p=seq(0,1,0.1)
cuant=qpois(p,2*2)
cuant

> cuant
[1]   0   2   2   3   3   4   4   5   6   7 Inf

cuant=qpois(0.75,2*2)
cuant

> cuant
[1] 5
```

x. #Ejercicio10

```
# Graficar 3 distribuciones Poisson diferentes

# Solución: #################################################################

# Obtener los valores del rango de x
x=seq(0,50,1)
# Obtener los valores de la funcion de probabilidad
# primero para n=50 y lambda = 5
pdf1=dpois(x,5)
# segundo para n=50 y lambda = 30
pdf2=dpois(x,30)
# segundo para n=50 y lambda = 50
pdf3=dpois(x,50)
# realizar los graficos
par(mfrow=c(1,3)) # pone los graficos en fila
plot(x,pdf1,type="h",xlab="x",ylab="f(x)")
title("n = 50, lambda = 5")
plot(x,pdf2,type="h",xlab="x",ylab="f(x)")
title("n = 50, lambda = 30")
plot(x,pdf3,type="h",xlab="x",ylab="f(x)")
title("n = 50, lambda = 50")
```

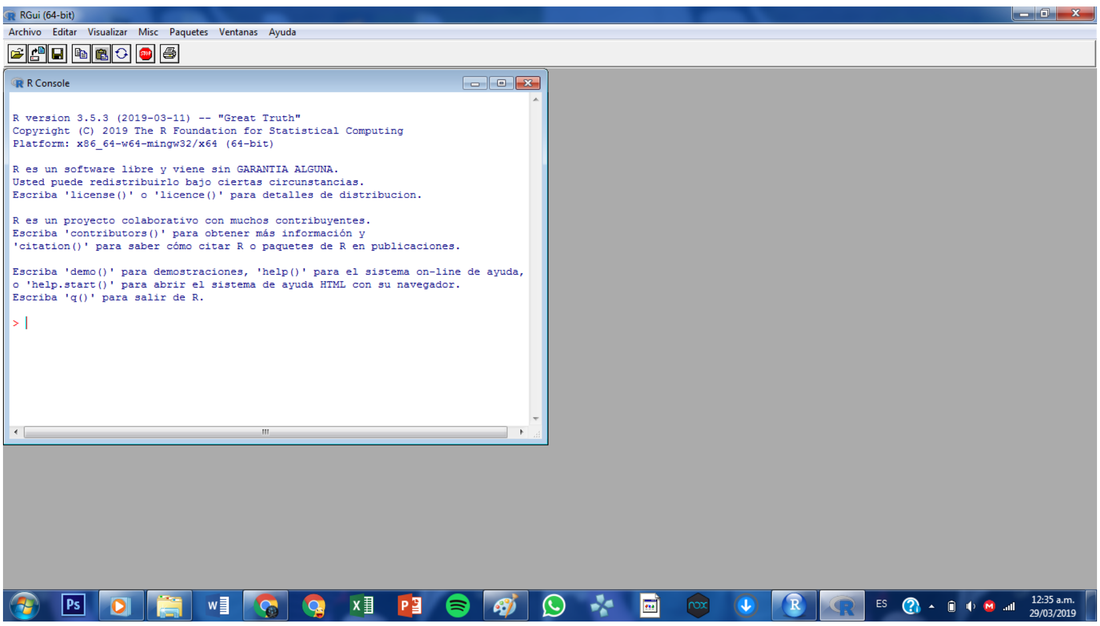


Parte 3. Manejo de herramientas de R. (10 puntos)

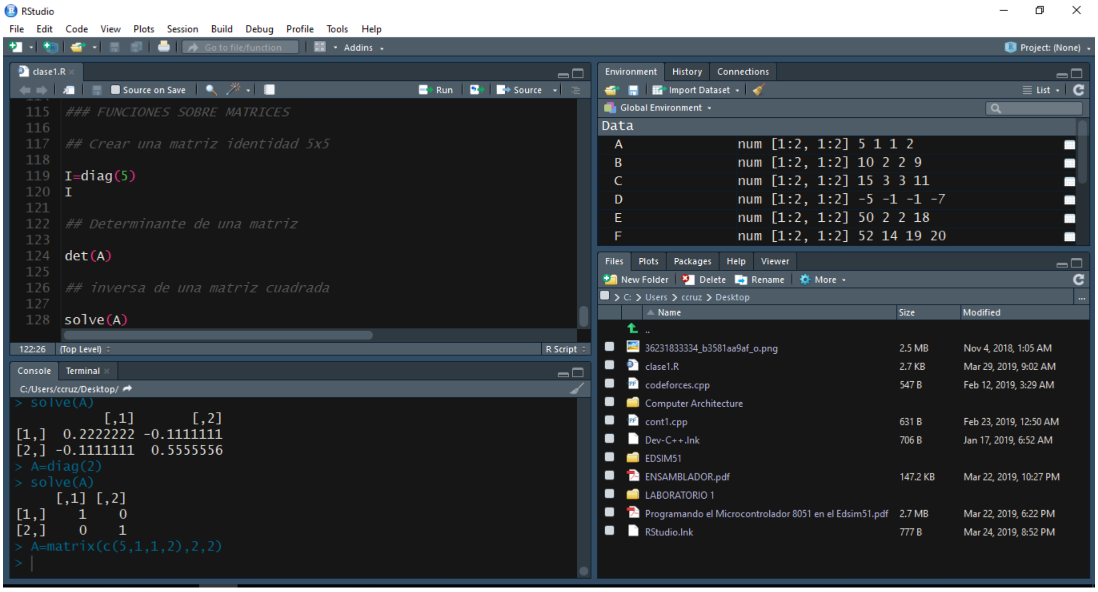
1. Evaluación de la primera tarea en R (presentada 29 de marzo). (2 puntos.)

   Debe mostrar que las instalaciones de R y R-Studio han sido exitosas, para eso deben enviar un screenshot de la pantalla donde se muestre

Solución 7 *i. R instalado o funcionando* *(1 punto.)*



ii. R studio instalado o funcionando *(1 punto.)*



2. Evaluación de la segunda tarea en R (presentada 30 de marzo). (4 puntos.)

Solución 8

i. Crear matrices de orden 8x8 en las cuales deben operar todas las operaciones conocidas hasta el momento. (2 puntos.)

```
####Parte 1
#########################################################
###Matrices de 8x8
#########################################################
P=matrix(15:78,nrow=8,ncol=8)
P
Q=matrix(6:69,nrow=8,ncol=8,T)
Q
###Operaciones con matrices
#########################################################

##Adicion
#########
A=P+Q
A
##Sustraccion
#############
B=P-Q
B

##Multiplicacion escalar
########################
C=P*5
C
D=Q*2
D

##Multiplicacion de matrices
############################
E=P%*%Q
E
```

ii. A esas matrices creadas deben multiplicar por derecha un vector columna y por izquierda un vector fila. (2 puntos.)

```
#########################################################
###Matrices de 8x8
#########################################################
P=matrix(15:78,nrow=8,ncol=8)
P
Q=matrix(6:69,nrow=8,ncol=8,T)
Q
###Creando Matrices
########################################################
M=matrix(5:12)            #Vector Columna
M
N=matrix(8:15,T)          #Vector Fila
N

###Operando las Matrices
########################
F=P%*%M                   #Multiplicando por un vector columna a P
F

G=N%*%F                   #Multiplicando por un vector fila al resultado anterior F
G



H=Q%*%M                   #Multiplicando por un vector columna a Q
H

I=N%*%H                   #Multiplicando por un vector fila al resultado anterior H
I
```

3. Evaluación de la tercera tarea en R (presentada 1 de abril). (4 puntos.)

Solución 9

i. En base al punto 1, debe crear una data frame sencillo de 5 columnas y 10 filas eligiendo solo una de las temáticas planteadas

- *Bodega*
- *Universidad*
- *Hospital*
- *Empresa de transporte*
- *Banco*

Luego, guarde su data frame en formato csv bajo el formato "nombre1_apellido1_1.csv" (1 pto.)

```
##Data Frame de universidades

Nombre = c("Pontificia Universidad Católica de Perú",
"Universidad Nacional Mayor de San Marcos",
"Universidad Nacional Agraria La Molina" ,
"Universidad Nacional de San Antonio de Abad del Cusco" ,
"Universidad Nacional de Ingeniería","Universidad ESAN" ,
 "Universidad de San Martín de Porres",
"Universidad Peruana de Ciencias Aplicadas" ,
"Universidad de Ingeniería y Tecnología",
"Universidad Peruana Cayetano Heredia")

Tipo = c("Privada","Pública","Pública","Pública","Pública", "Privada" ,
 "Privada" , "Privada" , "Privada" , "Privada")

Creación =c(1917,1551, 1960 , 1692 , 1955 , 2003, 1962 , 1994 , 2011 , 1961)
```

```
Rector = c("Efraín Gonzales de Olarte" , "Orestes Cachay Boza" ,
"Enrique Ricardo Flores Mariazza" ,"Baltazar Nicolás Cáceres Huambo" ,
"Jorge Elias Alva Hurtado" ,"Jorge Talavera Traverso" ,
"José Antonio Chang Escobedo" ,"Edward Roekaert Embrechts" ,
"Eduardo Hochschild" ,"Luis Varela Pinedo")

Postulantes.2015 = c(16534,77766,6068,35249,11479,3883,9398,25338,772,5091)

Datos = data.frame(Nombre,Tipo,Rector,Creación,Postulantes.2015)

Datos

write.table(Datos,file = "C:/Users/Jesus/Desktop/Universidad/Ciclo III/
Estadística Inferencial/Jesús_Ramirez_1.csv", sep = ",", row.names = F)
```

ii. *En base al punto 2, seleccione una base de datos (de las tantas que ofrece R), luego explique o describa que ofrece dicha data, finalmente guarde esa data en un archivo csv bajo el formarto "nombre1_apellido1_2.csv" (1 pto.)*

```
#La tabla elegida fue "HairEyeColor", en la tabla se describe el color de cabello,
#ojos y el género de una muestra de estudiantes de estadística.

View (HairEyeColor)

class(HairEyeColor) #tabla

Carac = data.frame(HairEyeColor)
Carac

class(Carac) #data.frame


write.table(Carac,file = "C:/Users/Jesus/Desktop/Universidad/Ciclo III/
Estadística Inferencial/Jesús_Ramirez_2.csv", sep = ",", row.names = F)
```

iii. *En base al punto 3, debe abrir el archivo "info_peru.csv", luego debe elegir solo 5 columnas (años) que a usted le parezcan las mas importantes, para luego crear un nuevo data frame con la informacion basica de las 4 primeras columnas y las 5 columnas que usted eligió, para crear y guardar esto ultimo como el nuevo data frame llamado "nombre1_apellido1_3.csv" (2 ptos.)*

```
##info_peru

setwd ("C:/Users/Jesus/Desktop/Universidad/Ciclo III/Estadística Inferencial")

getwd()                              #Para saber en que directorio estamos

peru = read.csv("info_peru.csv", sep = ",", header = TRUE)   #Crea data.frame

class(peru)                          #data.frame

View(peru)                           #Aparece la tabla de info_peru

#Creando la nueva data.frame

Country.Name = peru[ ,1]

Country.Code = peru[ ,2]

Indicator.Name = peru[,3]
```

```
Indicator.Code = peru[ , 4]

#Años elegidos:

X1960 = peru[ , 5]

X1961 = peru[ , 6]

X1962 = peru[ , 7]

X1963 = peru[ , 8]

X1985 = peru[ , 30]

X2000 = peru[ , 45]

X2001 = peru[ , 46]

X2010 = peru[ , 55]

X2016 = peru[ , 61]

Jesus_Ramirez_3 = data.frame (Country.Name , Country.Code , Indicator.Name ,
Indicator.Code, X1960, X1961 , X1962 , X1963 ,
X1985 , X2000 , X2001 , X2010 , X2016)
Jesus_Ramirez_3

write.table(Datos,file = "C:/Users/Jesus/Desktop/Universidad/Ciclo III/
Estadística Inferencial/Jesús_Ramirez_3.csv", sep = ",", row.names = F)
```

Clifford Torres.[1]
May 7, 2019

[1] Hecho en LaTeX