UNIVERSIDAD CATOLICA BOLIVIANA “SAN PABLO”

MAESTRIA EN CIENCIA DE DATOS, TERCERA VERSION

Materia: ANALISIS ESTADÍSTICO I

Examen

Maestrante: Ramón Wilder Serdán Cárdenas

Noviembre 2021

La Paz – Bolivia

PREGUNTAS

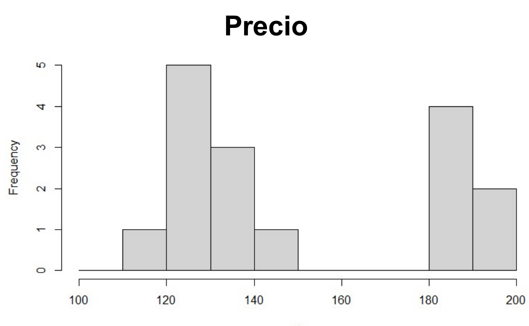
1. Con la base de datos: datos_ p_18.xlsx, se pide:
Realizar todos los gráficos estadísticos, e interprete. (se evalúa interpretación)

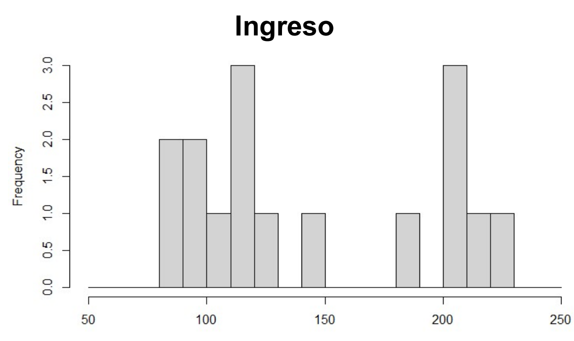
Supuesto: La descripción de los campos es la siguiente:

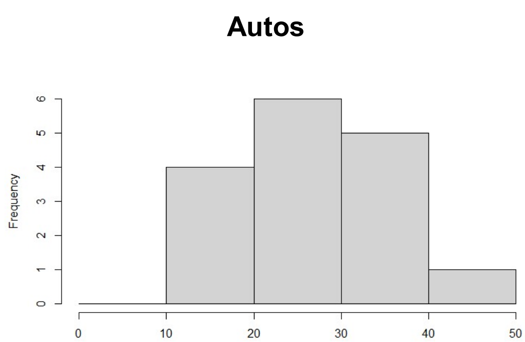
- **Precio** es el precio de autos en pesos.
- **Ingreso** es el ingreso expresado en pesos.
- **Autos** es el número de autos en una escala (miles).
- **Ventas** número de autos vendidos en una escala (miles).

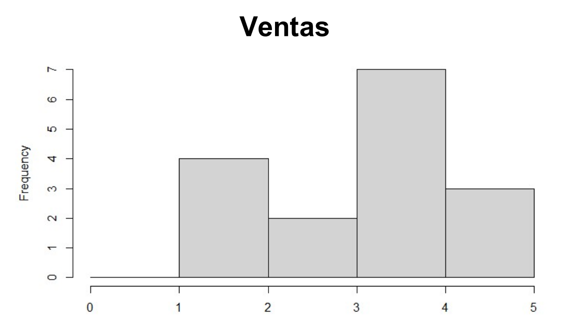
Histogramas de las variables cuantitativas

Los precios de los autos se encuentran en 2 rangos, el primero entre 110 pesos y 150 pesos y el segundo entre 180 pesos y 200 pesos. El ingreso tiene relación con los precios puesto que se encuentran también en dos rangos diferenciados. Las variables "Autos" y "Ventas" muestran una distribución aparentemente normal.





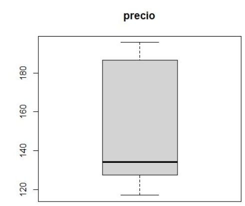




Boxplot

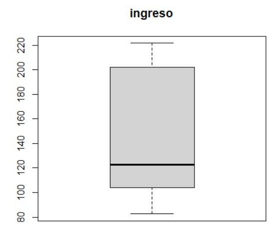
Ninguna de las variables presentan valores atípicos salvo las Ventas en miles de pesos que presenta un valor inferior que se podría tomar como atípico.

Precio



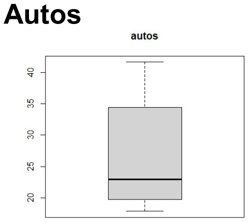

```
medidas(bd$`precio`)
           Medida       Valor
            Media 150.9250000
          Mediana 134.0500000
             Moda 128.5000000
       Cuartil 25 128.0000000
       Cuartil 75 186.6000000
              IQR  58.6000000
Coef. de variación   0.2023812
```

Ingreso



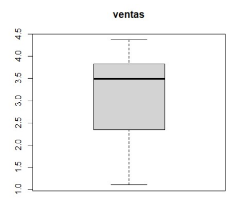

```
medidas(bd$`ingreso`)
           Medida       Valor
            Media 145.2750000
          Mediana 122.7500000
             Moda  82.6000000
       Cuartil 25 106.5750000
       Cuartil 75 201.3000000
              IQR  94.7250000
Coef. de variación   0.3579058
```

Autos




```
medidas(bd$`autos`)
           Medida      Valor
            Media 26.6937500
          Mediana 22.9500000
             Moda 17.9000000
       Cuartil 25 19.9250000
       Cuartil 75 33.7500000
              IQR 13.8250000
Coef. de variación  0.3082172
```

Ventas



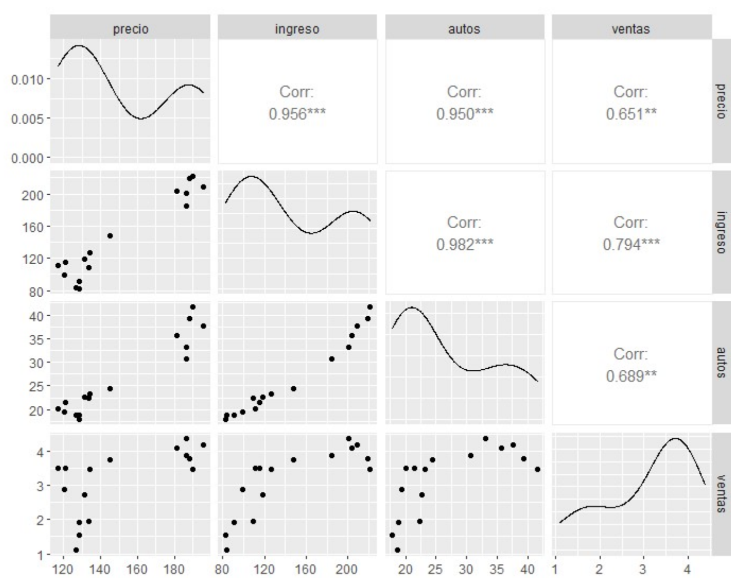

```
medidas(bd$`ventas`)
           Medida     Valor
            Media 3.1325000
          Mediana 3.4900000
             Moda 3.5100000
       Cuartil 25 2.5300000
       Cuartil 75 3.8025000
              IQR 1.2725000
Coef. de variación 0.3212334
```

Covarianza y el coeficiente de correlación

Como se aprecia en el siguiente cuadro, las variables “Ingreso” y “Precio” muestran un grado de asociación alto.

```
asociacion(bd$`ingreso`,bd$`precio`)
                Medida        Valor
            Covarianza 1518.3353333
Coef. de Correlación      0.9560415
asociacion(bd$`ventas`,bd$`precio`)
                Medida     Valor
            Covarianza 19.998400
Coef. de Correlación    0.650657
```

Asimismo, se corroboraron los resultados con el siguiente gráfico utilizando la librería GGally en el que también se muestran los gráficos de dispersión y correlaciones entre todas las variables de la base de datos.



2.- Determinar los estadísticos de las 4 variables e interpretar

Variable Precio

El precio expresado en pesos no se encuentra muy disperso en virtud al coeficiente de variación.

```
medidas(bd$`precio`)
              Medida       Valor
               Media 150.9250000
             Mediana 134.0500000
                Moda 128.5000000
          Cuartil 25 128.0000000
          Cuartil 75 186.6000000
                 IQR  58.6000000
Coef. de variación   0.2023812
```

Ingreso

El ingreso, que se encuentra en función del precio, tampoco muestra mucha dispersión.

```
medidas(bd$`ingreso`)
              Medida       Valor
               Media 145.2750000
             Mediana 122.7500000
                Moda  82.6000000
          Cuartil 25 106.5750000
          Cuartil 75 201.3000000
                 IQR  94.7250000
Coef. de variación   0.3579058
```

Autos

```
medidas(bd$`autos`)
              Medida       Valor
               Media  26.6937500
             Mediana  22.9500000
                Moda  17.9000000
          Cuartil 25  19.9250000
          Cuartil 75  33.7500000
                 IQR  13.8250000
Coef. de variación   0.3082172
```

Ventas

La empresa presenta un promedio de ventas de 3,13 miles de pesos con datos ligeramente dispersos.

```
medidas(bd$`ventas`)
              Medida     Valor
               Media 3.1325000
             Mediana 3.4900000
                Moda 3.5100000
          Cuartil 25 2.5300000
          Cuartil 75 3.8025000
                 IQR 1.2725000
Coef. de variación 0.3212334
```

3. Calcular la probabilidad:

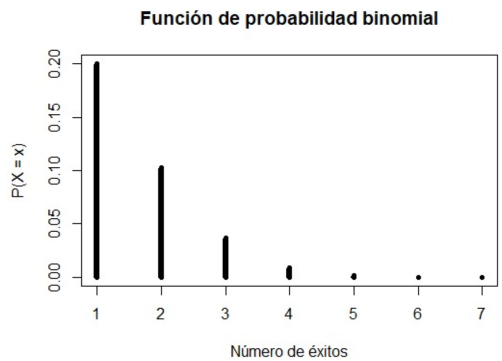
a) Cuando X~binomial(10;0.3)

1. P(X=5)

```
> n<-10
> p<-0.3
> x<-5 #igual a 5
> dbinom(x=x, size=n, prob=p)
[1] 0.1029193
```

2. P(4<=X<=10)

```
> x<-4
> y<-10
> sum(dbinom(x:y, size=n, prob=p))
[1] 0.3503893
> #metodo alternativo
> pbinom(y,size=n, prob=p,lower.tail =TRUE)- pbinom(x-1,size=n, prob=p,lower.tail =TRUE)
[1] 0.3503893
```



3. P(X<=5)

```
> pbinom(x,size=n, prob=p,lower.tail =TRUE)
[1] 0.952651
> sum(dbinom(0:x, size=n, prob=p))
[1] 0.952651
```

b) Cuando X~Poisson(lambda=3.52)

1. P(X=3)

```
> lambda<-3.52
> dpois(x=x, lambda=lambda)
[1] 0.2151593
```

2. P(3<=X<=7)

```
> x<-3
> y<-7
> sum(dpois(x:y, lambda=lambda))
[1] 0.6553185
> ppois(y,lambda=lambda,lower.tail =TRUE)-ppois(x-1,lambda=lambda,lower.tail =TRUE)
[1] 0.6553185
```

3. P(X=4)

```
> lambda<-3.52
> x<-4
> dpois(x=x, lambda=lambda)
[1] 0.1893402
>
```

4. Utilizando los datos de la pregunta 1, se pide determinar los intervalos de confianza para todas las variables con 99% de confianza(interpretar)

```
> intervalos<-function(nombre,bd,ene,media,desv){
+   media<-mean(bd)
+   ene<-length(bd)
+   desv<-sd(bd)
+   error.est <- desv/sqrt(ene) # error estándar
+   margen.error <- 2.576 * error.est # nivel de confianza de 99%
+   lim.inf <- media - margen.error # Límite inferior del intervalo
+   lim.sup <- media + margen.error # Límite superior del intervalo
+   cat("Variable: ", nombre, ", limite inferior: ", lim.inf, ", limite superi
+   #metodo alternativo
+   #zsum.test(mean.x=media,sigma.x=desv, n.x=ene,conf.level=0.99)
+ }
> intervalos("Precio",bd$`precio`)
Variable:  Precio , limite inferior:  131.2544 , limite superior:  170.5956
> intervalos("Ingreso",bd$`ingreso`)
Variable:  Ingreso , limite inferior:  111.7904 , limite superior:  178.7596
> intervalos("Autos",bd$`autos`)
Variable:  Autos , limite inferior:  21.39526 , limite superior:  31.99224
> intervalos("Ventas",bd$`ventas`)
Variable:  Ventas , limite inferior:  2.484466 , limite superior:  3.780534
```

Las medias poblacionales en todos los casos se encuentran dentro del intervalo de confianza, lo que permite deducir que la media poblacional se encuentra dentro de los intervalos descritos en el cuadro anterior con un nivel de confianza de 99%.
Asimismo, calculados los p-value de todas las variables nos permiten aceptar la Hipótesis nula de que las medias poblacionales sean iguales a las medias muestrales.

5. Utilizando la base de datos datos_p_18.xlsx, probar:

a) H0: mu_precio =150, con significancia=0.06
Se plantean la hipótesis:

Ho: mu_precio =150
H1: mu_precio <>150

```
> precio<-150
> media<-mean(bd$precio)
> desv<-sd(bd$precio)
>   probZ<-pnorm(precio,media,desv,lower.tail = FALSE)
>   probMayor<-2*probZ
>   probMayor
[1] 1.024159
```

Dado el nivel de significancia, el p-value alto permite **no rechazar** la Hipotesis nula de que la media poblacional sea 150 pesos.

b) H0: mu_ingreso =160, con significancia=0.03
Ho: mu_precio =160
H1: mu_precio <>160

```
> ingreso<-160
> media<-mean(bd$ingreso)
> desv<-sd(bd$ingreso)
> probZ<-pnorm(ingreso,media,desv,lower.tail = FALSE)
> prob<-2*probZ
> prob
[1] 0.7770223
```

Ya que el nivel de significancia es 0,03, el p-value superior a dicho valor permite que no rechacemos la Hipotesis nula de que la media poblacional del ingreso sea igual a 160 pesos.

6. Aplicar el método bisección o Newton Raphson para la función:

X^2-Cos(x)-1=0
Considerar X* en [1,2], X0=1.5

a	b	m	Error est.
1.0000000	1.5000000	1.5000000	0.2500000
1.0000000	1.2500000	1.2500000	0.1250000
1.1250000	1.2500000	1.1250000	0.0625000
1.1250000	1.1875000	1.1875000	0.0312500
1.1562500	1.1875000	1.1562500	0.0156250
1.1718750	1.1875000	1.1718750	0.0078125
1.1718750	1.1796875	1.1796875	0.0039062
1.1757812	1.1796875	1.1757812	0.0019531
1.1757812	1.1777344	1.1777344	0.0009766
1.1757812	1.1767578	1.1767578	0.0004883
1.1762695	1.1767578	1.1762695	0.0002441
1.1762695	1.1765137	1.1765137	0.0001221
1.1763916	1.1765137	1.1763916	0.0000610
1.1764526	1.1765137	1.1764526	0.0000305
1.1764832	1.1765137	1.1764832	0.0000153
1.1764984	1.1765137	1.1764984	0.0000076
1.1764984	1.1765060	1.1765060	0.0000038
1.1764984	1.1765022	1.1765022	0.0000019
1.1765003	1.1765022	1.1765003	0.0000010

Cero de f en [1 , 2] es approx: 1.1765 con error <= 9.536743e-07