

PROBABILIDAD Y ESTADÍSTICA

Damian Guillermo Morales Cruz

2-2020

Parcial 1

PROBLEMA 1 El Peso y la Presión arterial (PS) de 26 personas del género masculino seleccionadas al azar de un grupo con edades entre 25 a 30 años se muestra en la tabla.

Sujeto	Peso	PS sistólica	Sujeto	Peso	PS sistólica
1	165	130	14	172	153
2	167	133	15	159	128
3	180	150	16	168	132
4	155	128	17	174	149
5	212	151	18	183	158
6	175	146	19	215	150
7	190	150	20	195	163
8	210	140	21	180	156
9	200	148	22	143	124
10	149	125	23	240	170
11	158	133	24	235	165
12	169	135	25	192	160
13	170	150	26	187	159

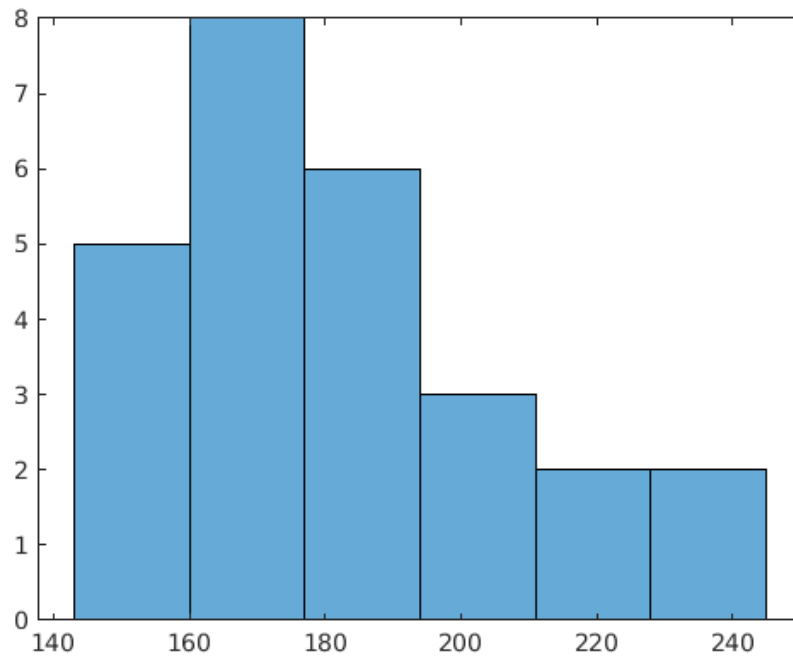
1. Determine para las variables (Peso y Presión arterial sistólica):

a. Tabla de frecuencia de los datos utilizando 6 clases.

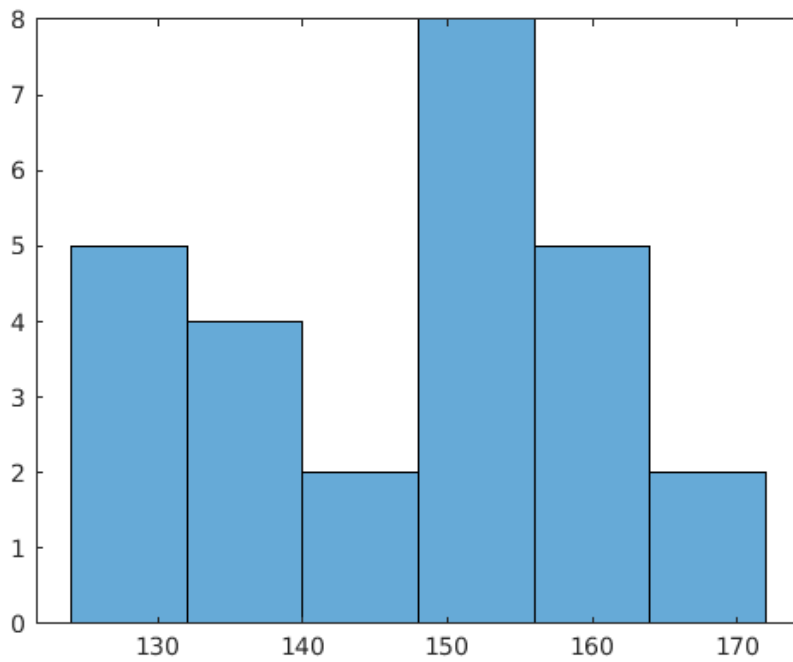
	marca_peso	fa_peso	fr_peso	F_peso	Fr_peso	marca_PS	fa_PS	fr_PS	F_PS	Fr_PS
1	151.5000	5	0.1923	5	0.1923	128	5	0.1923	5	0.1923
2	168.5000	8	0.3077	13	0.5000	136	4	0.1538	9	0.3462
3	185.5000	6	0.2308	19	0.7308	144	2	0.0769	11	0.4231
4	202.5000	3	0.1154	22	0.8462	152	8	0.3077	19	0.7308
5	219.5000	2	0.0769	24	0.9231	160	5	0.1923	24	0.9231
6	236.5000	2	0.0769	26	1.0000	168	2	0.0769	26	1.0000

- b. Histograma de cada variable para representar los datos.

Histograma peso



Histograma PS



c. Diagrama tipo pastel (con porcentajes) de cada variable e identificar en qué clase se encuentra el mayor porcentaje de los datos.

Diagrama de peso

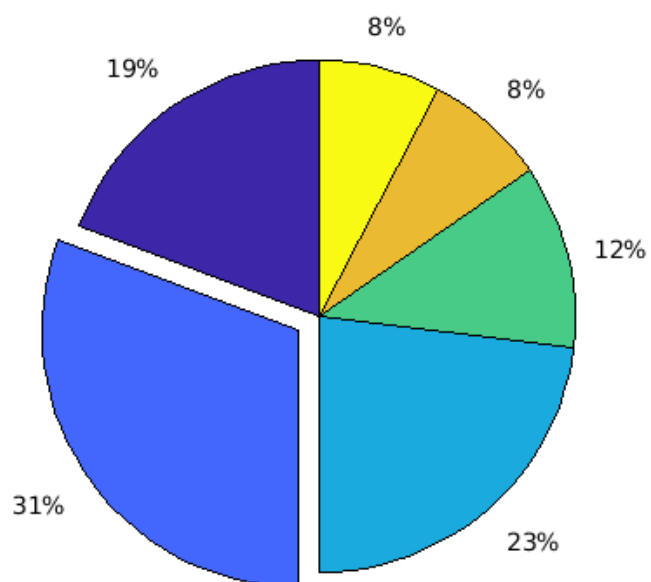
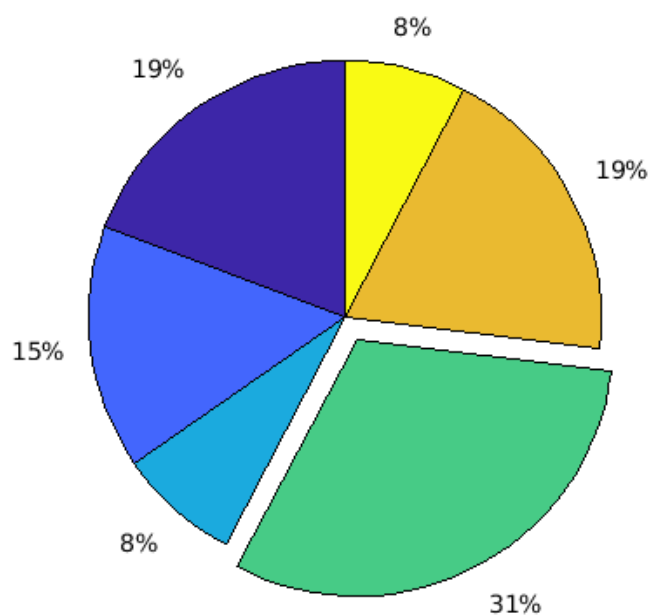
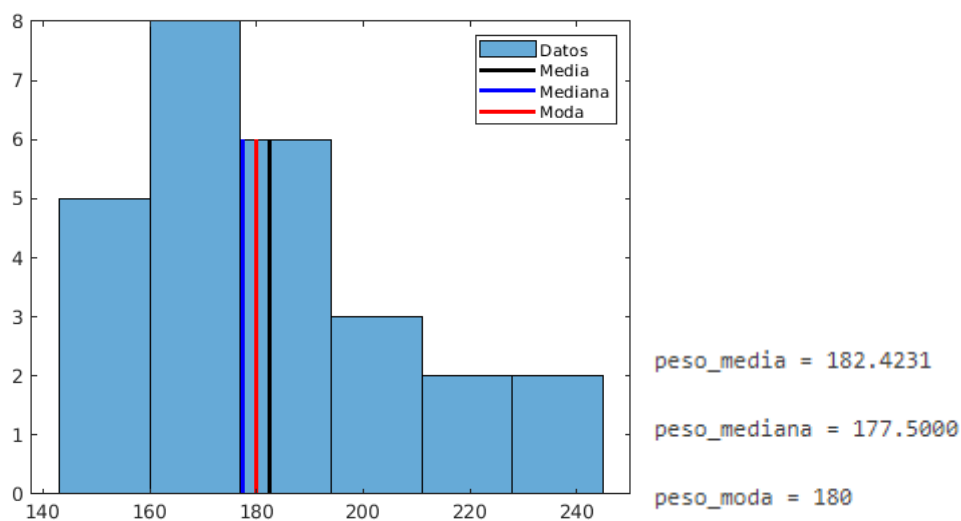


Diagrama de PS



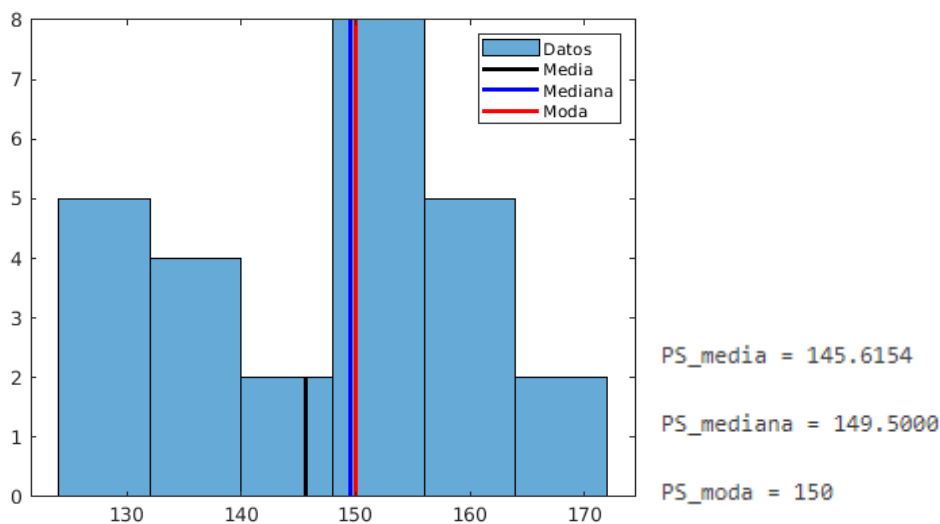
d. Media, mediana y moda de cada variable, graficarlos en el histograma e identificar el tipo de sesgo que presentan los datos, justificando su respuesta con los valores calculados.

Histograma de pesos



Sesgado a la derecha ya que se concentran los datos en la parte izquierda

Histograma de PS



Campana asimetrica ya que presenta una tendecia a la concentraacion de datos en la zona central, pero no es tan notoria

e. Varianza y desviación estándar de cada variable.

Peso

```
peso_varianza = 612.4938
```

```
peso_desviacion = 24.7486
```

PS

```
PS_varianza = 180.0862
```

```
PS_desviacion = 13.4196
```

f. Utilizar el teorema de Chebyshev para describir la distribución de los datos en el polígono de frecuencias.

Peso

El intervalo (132.9259 a 231.9203) contiene aproximadamente $3/4$ de las mediciones .

El intervalo (108.1773 a 256.6689) contiene aproximadamente $8/9$ de las mediciones .

PS

El intervalo (118.7762 a 172.4546) contiene aproximadamente $3/4$ de las mediciones .

El intervalo (105.3566 a 185.8742) contiene aproximadamente $8/9$ de las mediciones .

g. Coeficiente de variación de los datos, ¿cuál de las variables (Peso y Presión arterial sistólica) presenta mayor variación?

Peso

```
peso_cvariacion = 0.1357
```

Tiene un 13.57% de variación

PS

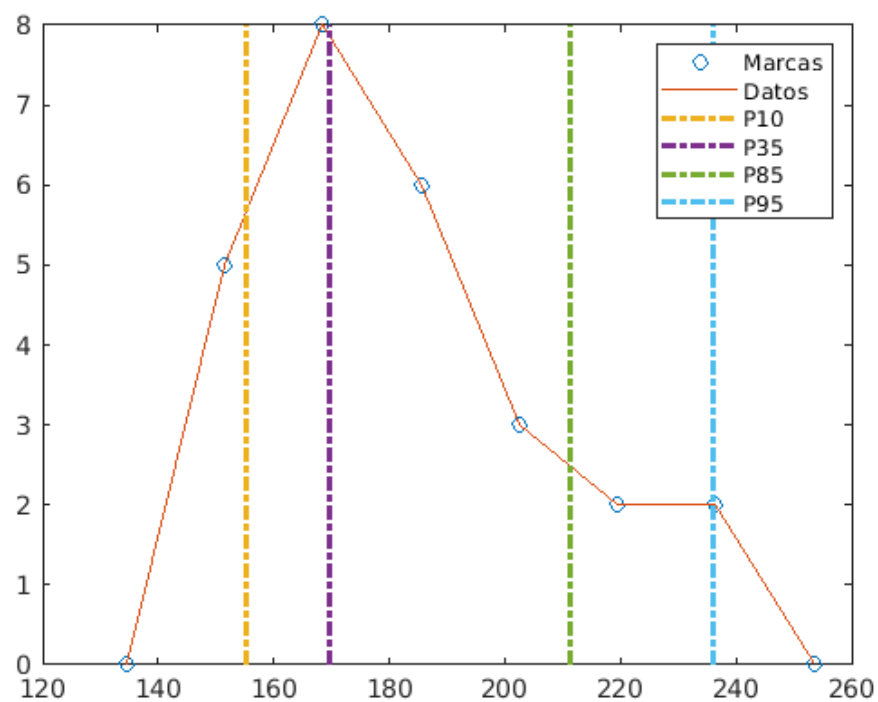
```
PS_cvariacion = 0.0922
```

Tiene un 9.22% de variación

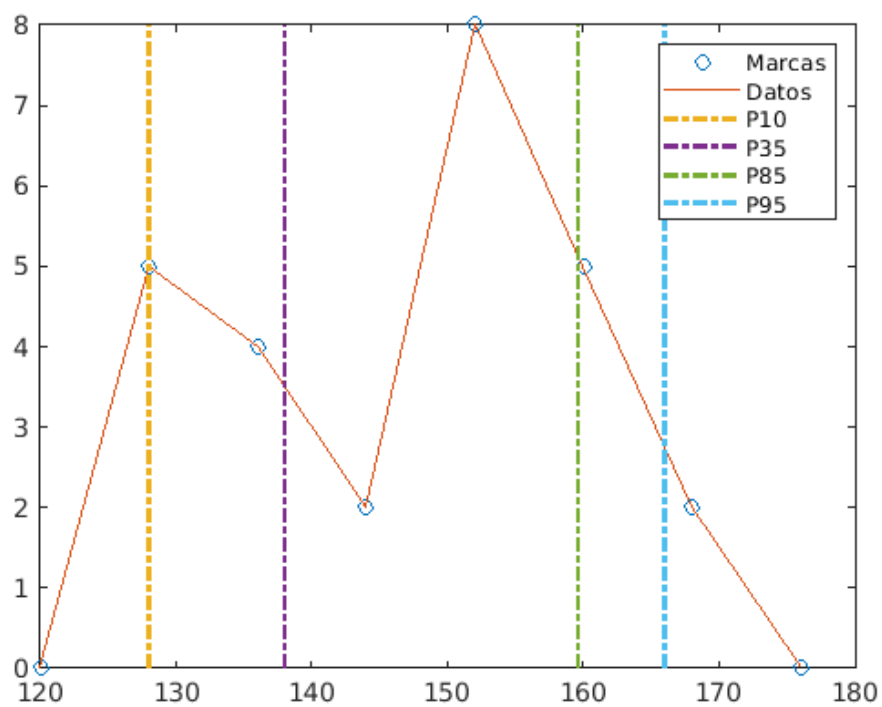
La variable peso presenta mayor variación

h. Percentiles P10, P35, P85, P95 y dibujarlos sobre el polígono de frecuencias de cada variable.

Peso

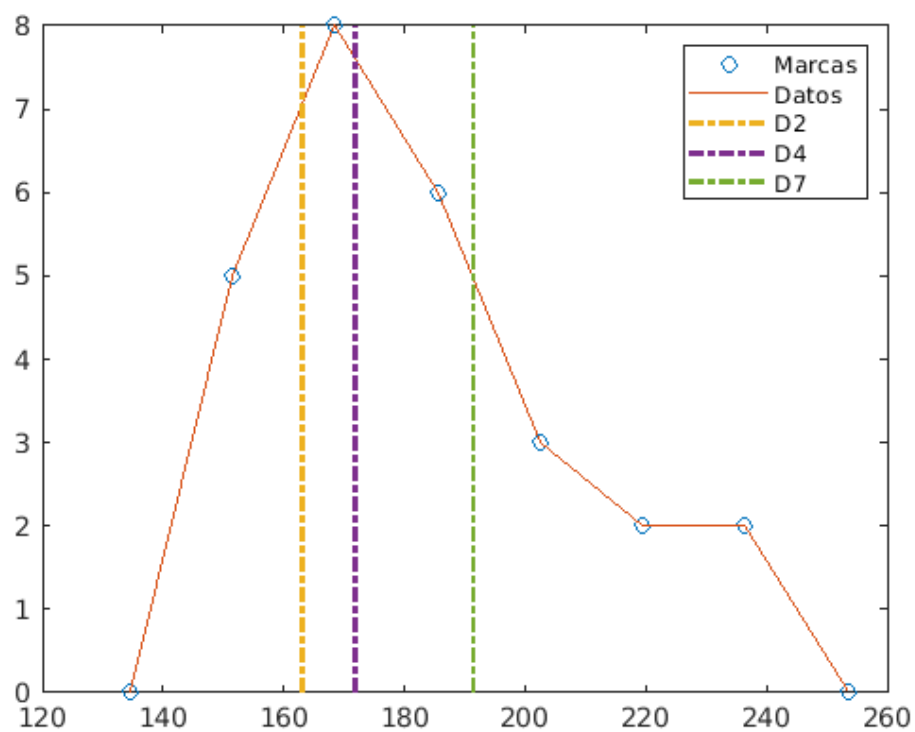


PS

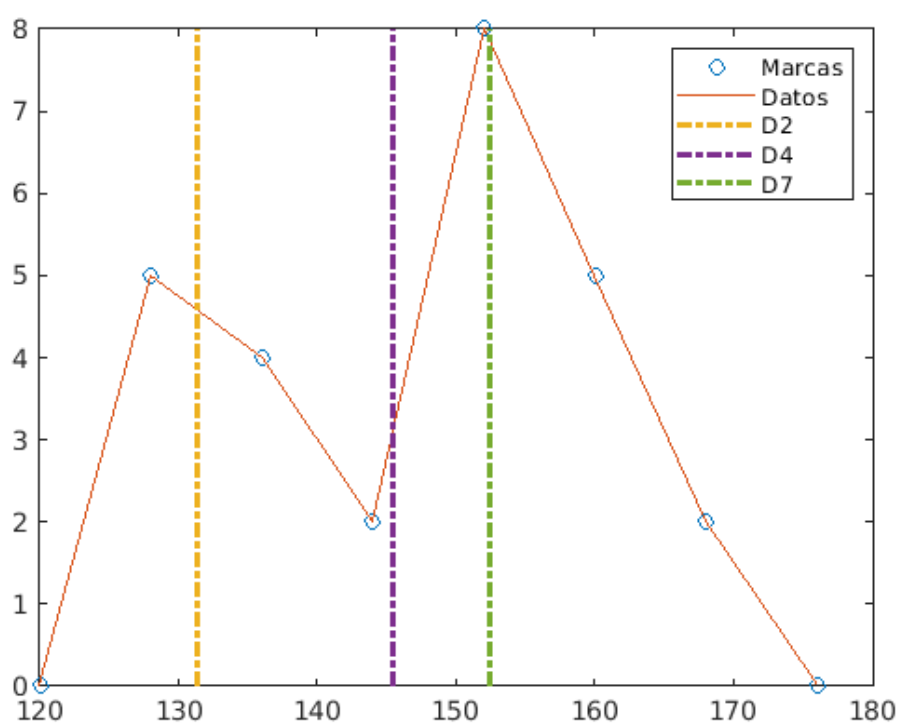


i. Deciles D2, D4, D7 y dibujarlos sobre el polígono de frecuencias de cada variable.

Peso

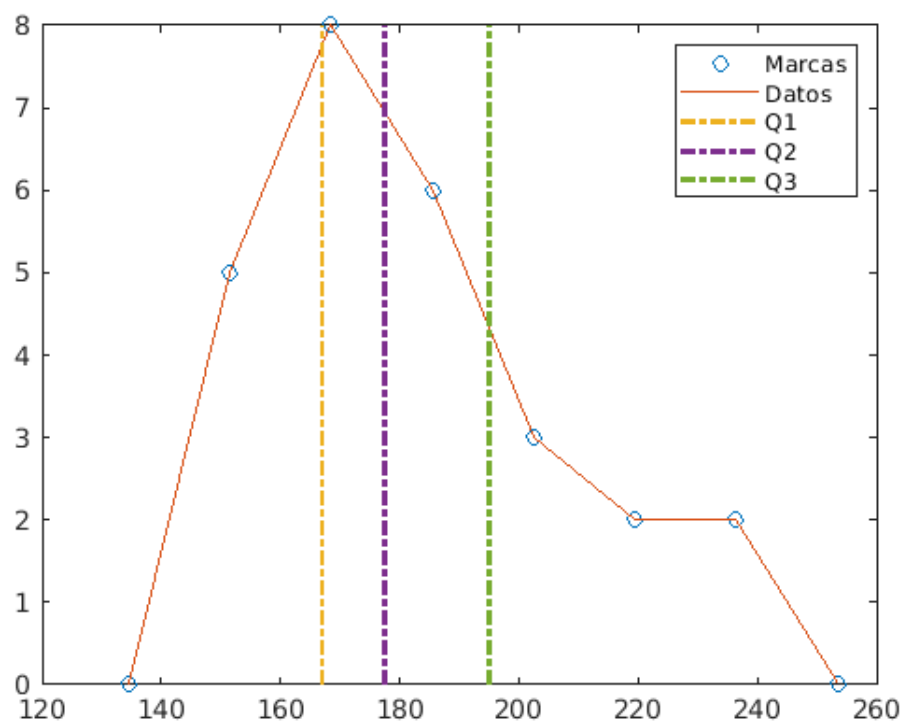


PS

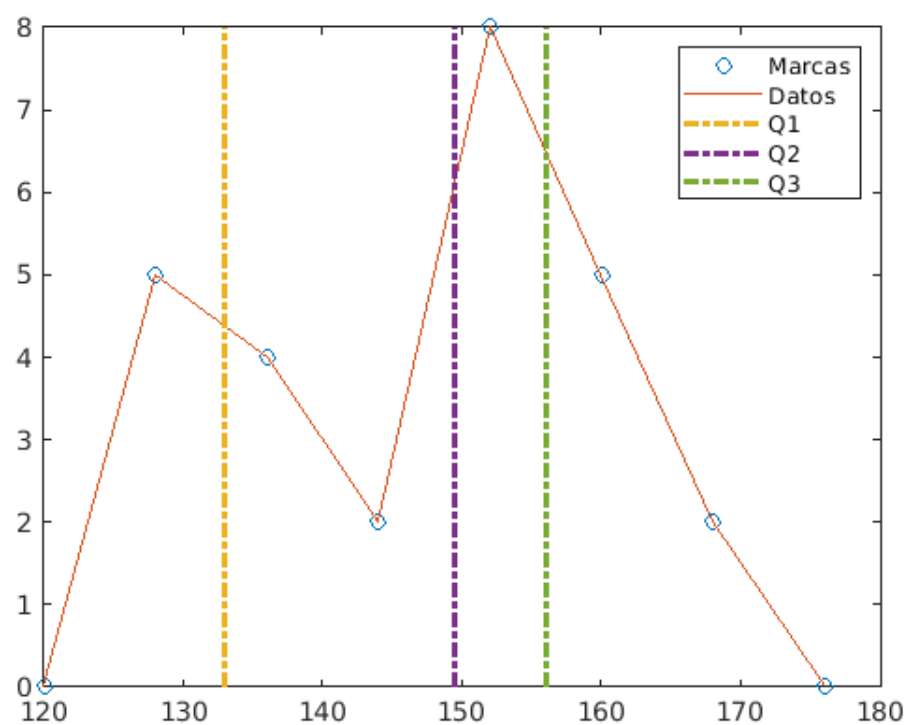


j. Cuartiles Q1, Q2, Q3 y dibujarlos sobre el polígono de frecuencias de cada variable.

Peso

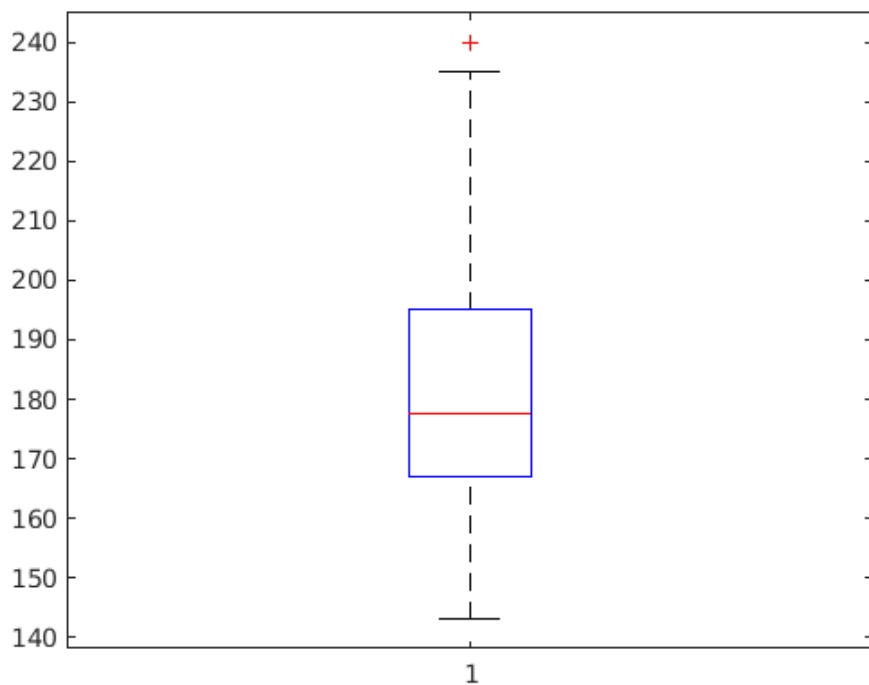


PS

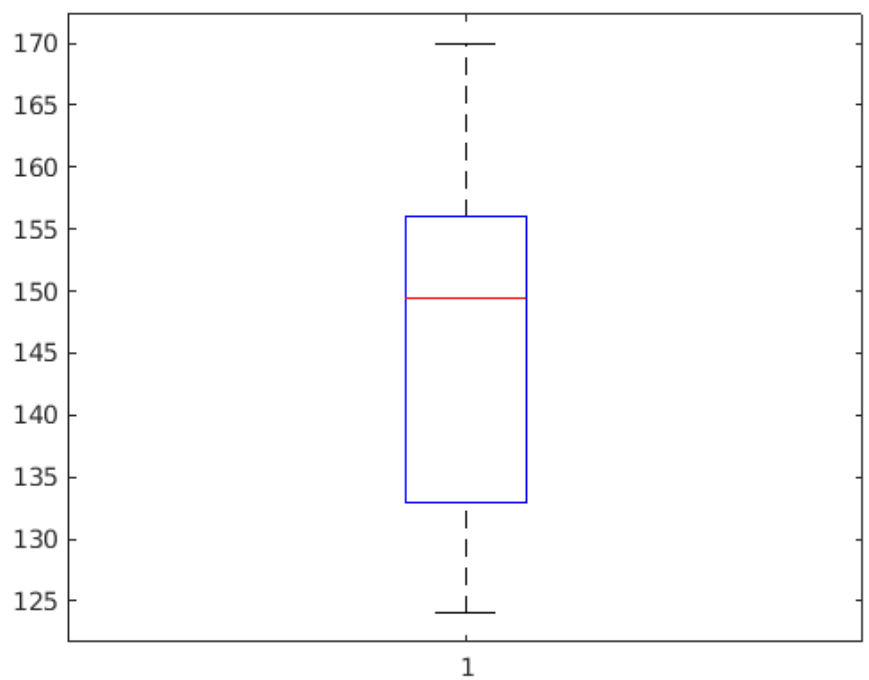


k. Dibujar el diagrama de caja y bigotes de cada variable.

Peso

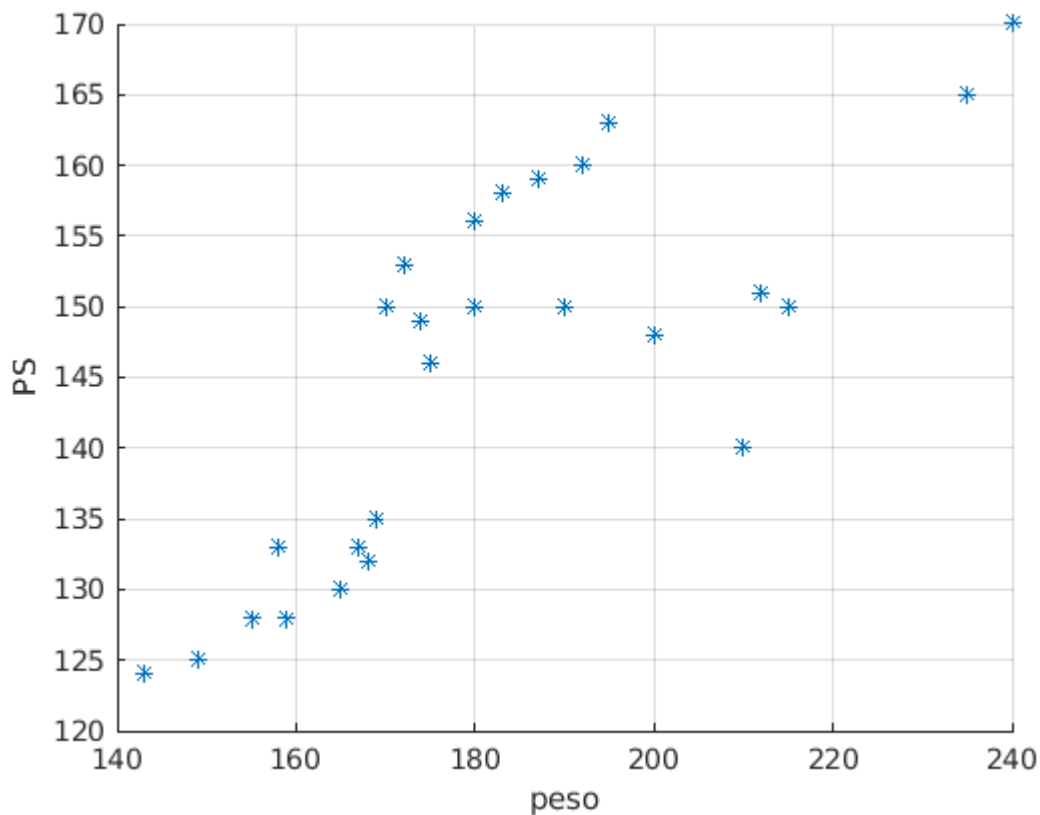


PS



2. Encontrar si existe una relación entre el Peso (variable independiente) y la Presión arterial (variable dependiente), para ello realice:

a. Diagrama de dispersión de los datos.



b. Calcular covarianza y coeficiente de correlación lineal, ¿Existe relación lineal entre las variables?

`covarianza = 256.8892`

`coeficiente_correlacion = 0.7735`

Si hay relación lineal, en esta caso positiva ya que su coeficiente de correlación esta cercano al valor de 1.

d. Realizar la regresión lineal, mostrando la ecuación de la línea recta y el coeficiente de determinación. ¿Es la recta aceptable para modelar los datos? Justifique su respuesta.

`recta =`

$$\frac{59027442236395}{140737488355328} x_a + \frac{37990530707039}{549755813888}$$

`r2 = 0.5983`

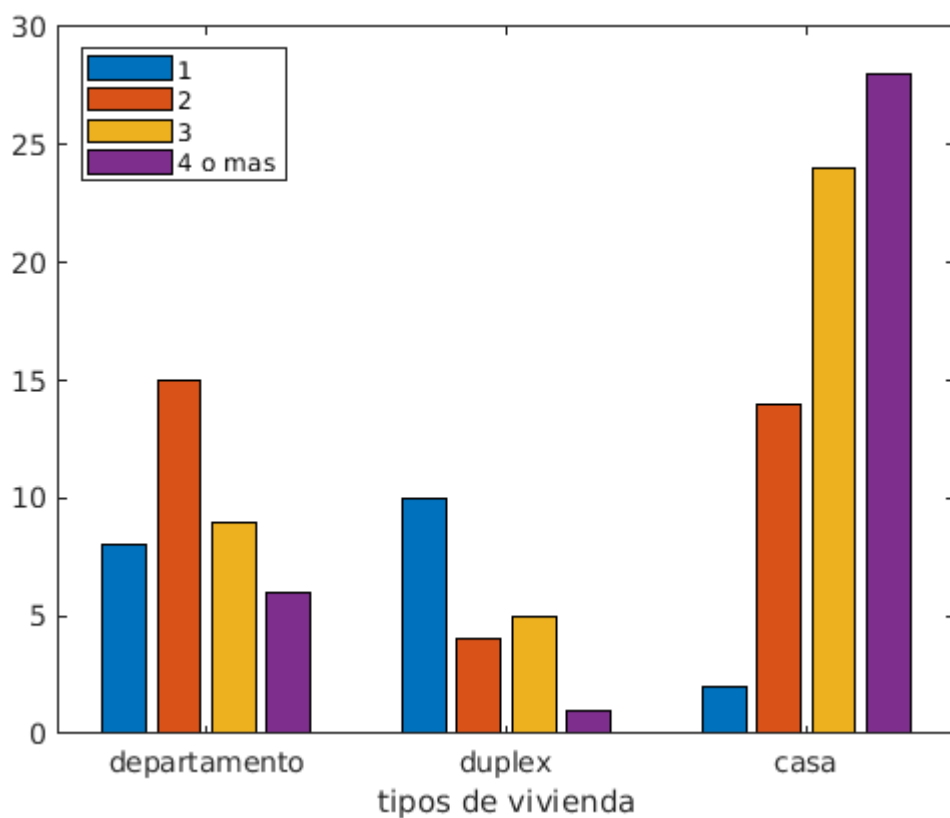
Recta= 0.42x + 69.10 La recta no es aceptable para modelar los datos ya que su índice de determinación es menor a 0.85 .

PROBLEMA 2

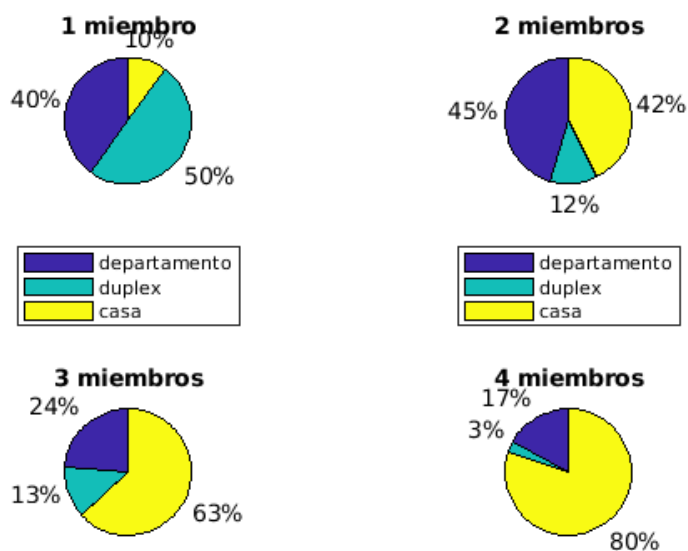
Se entrevistaron 126 familias dentro de una ciudad, registrando el tipo de residencia y el número de miembros de la familia en cada una de éstas. Como se muestra en la tabla.

Miembros en la familia	Tipo de residencia		
	Departamento	Dúplex	Casa
1	8	10	2
2	15	4	14
3	9	5	24
4 o más	6	1	28

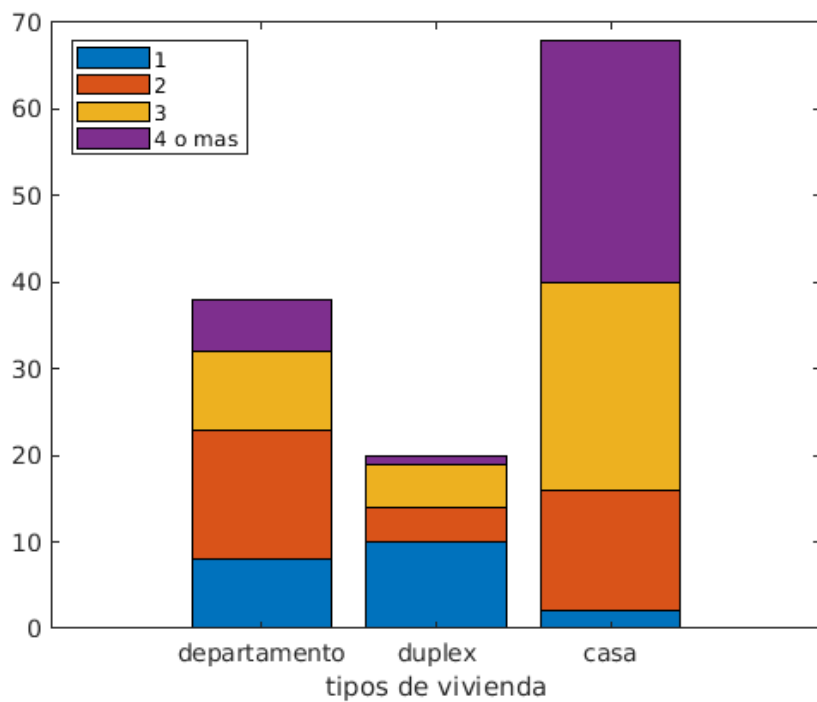
a. Realice una gráfica de barras lado a lado para comparar el número de miembros de una familia que viven en cada uno de los tres tipos de residencia.



b. Realice una gráfica tipo pastel para comparar en porcentaje el tipo de residencia que se prefiere según la cantidad de miembros en una familia.



c. Realice una gráfica de barras apiladas para comparar la preferencia de las familias al momento de escoger un tipo de residencia.

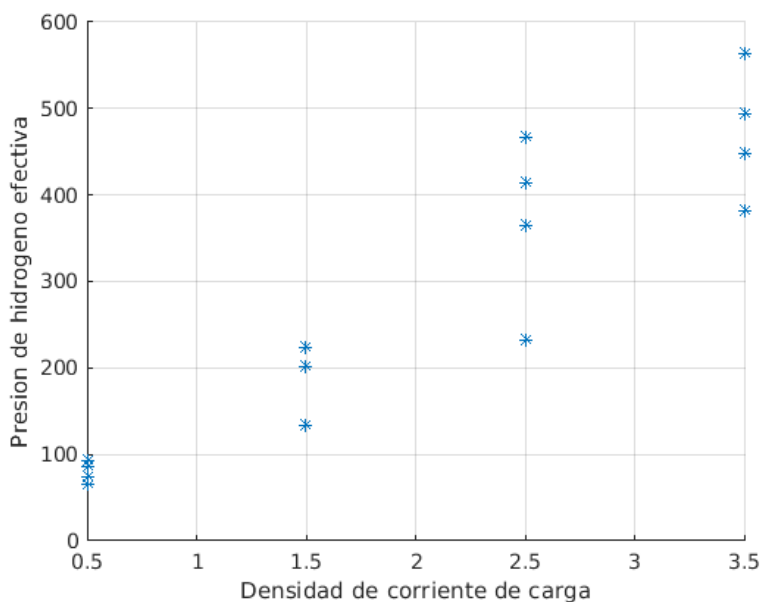


PROBLEMA 3

Se diseñó un experimento para el Departamento de Ingeniería de Materiales de una empresa, con el fin de estudiar las propiedades de deterioro del nitrógeno con base en las mediciones de la presión de hidrógeno electrolítico. La densidad de corriente de carga catódica fue controlada y variada en cuatro niveles. Se observó la presión de hidrógeno efectiva como la respuesta, como se muestra en los datos:

Ensayo	Densidad de corriente de carga, x (mA/cm ²)	Presión de hidrógeno efectiva, y (atm)
1	0.5	86.1
2	0.5	92.1
3	0.5	64.7
4	0.5	74.7
5	1.5	223.6
6	1.5	202.1
7	1.5	132.9
8	2.5	413.5
9	2.5	231.5
10	2.5	466.7
11	2.5	365.3
12	3.5	493.7
13	3.5	382.3
14	3.5	447.2
15	3.5	563.8

a. Diagrama de dispersión de los datos.



b. Transforme los datos para obtener una linealización del modelo exponencial y determine si existe una relación lineal entre las variables transformadas.

Cambiar la variable independiente por su logaritmo en este caso la presión:

```
p1 = 1×15
      4.4555    4.5229    4.1698    4.3135    5.4099    5.3088    4.8896

      6.0247    5.4446    6.1457    5.9007    6.2019    5.9462    6.1030    6.3347

coeficiente_correlacion_exp = 0.8975
```

Si hay una relación lineal entre las variables transformadas, ya que su coeficiente de correlación es cercano a uno, debido a esto se afirma que poseen relación lineal positiva.

c. En caso de existir una relación lineal entre las variables transformadas, determine la ecuación de la regresión de los datos linealizados y el coeficiente de determinación.

```
recta_exp =
      170230385409957 xa_exp + 1177044692006883
      281474976710656      281474976710656

r2_exp = 0.8877
```

Recta simplificada = $0.604x + 4.18$

d. Transforme los datos para obtener una linealización del modelo de potencias y determine si existe una relación lineal entre las variables transformadas.

Cambiar ambas variables por sus logaritmos: d2=densidad p2=presión

```
d2 = 1×15
      -0.3010   -0.3010   -0.3010   -0.3010    0.1761    0.1761    0.1761    0.3979

      0.3979    0.3979    0.3979    0.5441    0.5441    0.5441    0.5441

p2 = 1×15
      1.9350    1.9643    1.8109    1.8733    2.3495    2.3056    2.1235    2.6165

      2.3646    2.6690    2.5626    2.6935    2.5824    2.6505    2.7511

coeficiente_correlacion_pot = 0.8975
```

Si hay una relación lineal entre las variables transformadas, ya que su coeficiente de correlación es cercano a uno, debido a esto se afirma que poseen relación lineal positiva.

e. En caso de existir una relación lineal entre las variables transformadas, determine la ecuación de la regresión de los datos linealizados y el coeficiente de determinación.

recta_pot =

$$\frac{1044318493192881}{1125899906842624} x_{a_{pot}} + \frac{4861505083841293}{2251799813685248}$$

r2_pot = 0.9175

Recta simplificada = 0.927x + 2.15

f. ¿Cuál de los dos modelos representa mejor los datos? Justifique su elección.

En este caso ambos modelos los representan de manera satisfactoria ya que su coeficiente de determinación es mayor a 0.85 , pero si se comparan ambos modelos se aprecia que el modelo de potencias los representa mejor ya que su coeficiente de determinación es mayor.