**UNIVERSIDAD EIA**

**COMPUTACIÓN CIENTÍFICA**

**TALLER DE ESTADÍSTICA DESCRIPTIVA (10%)**

1. Identificar qué variables son cualitativas y qué variables son cuantitativas.
2. Comida favorita. **Cualitativa**
3. Profesión que gusta. **Cualitativa**
4. Número de goles marcado por un equipo de fútbol. **Cuantitativa**
5. Número de alumnos de la universidad. **Cuantitativa**
6. Color de ojos de una persona. **Cualitativa**
7. Coeficiente intelectual de un grupo de personas. **Cuantitativa, asumo como número**
8. Identifique qué variables son discretas y qué variables son continuas.
9. Número de acciones vendidas en la bolsa. **Discreta**
10. Temperatura registrada cada hora en un observatorio. **Continua**
11. Periodo de duración de un automóvil. **Discreta**
12. Diámetro de las ruedas de un auto. **Continua**
13. Número de hijos de 50 familias. **Discreta**
14. Censo anual de una población. **Continuo**
15. En una clínica infantil se ha ido anotando, durante un mes, el número de metros que cada niño anda, seguido y sin caerse, el primer día que comienza a caminar, obteniéndose la siguiente tabla de información:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Número de metros | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Número de niños | 2 | 6 | 10 | 5 | 10 | 3 | 2 | 2 |

Se pide:

1. Tabla de frecuencias.
2. Diagrama de barras para frecuencias absolutas, relativas y acumuladas.
3. Medidas de tendencia central (Mediana, promedio, moda y cuartiles).
4. Medidas de desviación (Varianza y desviación estándar).
5. Gráfico de dispersión.
6. Entregue conclusiones de los presentado en los datos.
7. Las cifras dadas en la siguiente tabla corresponden a miligramos de hidroxiprolina absorbidos por 1 miligramo de masa intestinal, analizados en distintos pacientes:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hidroxiprolina (mg) | 77.3 | 61.2 | 82.4 | 75.9 | 61 | 70.2 | 65 | 80 |
| Número de pacientes | 3 | 10 | 15 | 13 | 8 | 5 | 2 | 0 |

Se pide:

1. Generar la tabla de frecuencias.
2. Calcular medidas de tendencia central (Mediana, promedio, moda y cuartiles).
3. Medidas de desviación (Varianza y desviación estándar).
4. Gráfico de barras.
5. Gráfico de dispersión.
6. Entregue conclusiones de lo presentado en los datos.
7. Los valores del pH sanguíneo de 32 individuos son los siguientes:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 7.33 | 7.31 | 7.26 | 7.33 | 7.37 | 7.27 | 7.30 | 7.33 |
| 7.33 | 7.32 | 7.35 | 7.39 | 7.33 | 7.38 | 7.33 | 7.31 |
| 7.37 | 7.35 | 7.34 | 7.32 | 7.29 | 7.35 | 7.38 | 7.32 |
| 7.32 | 7.33 | 7.32 | 7.40 | 7.32 | 7.34 | 7.33 | 7.33 |

Se pide:

1. Generar la tabla de frecuencias por medio de la generación de 5 intervalos.
2. Calcular las medidas de tendencia central.
3. Calcular las medidas de variación.
4. Graficar el histograma de frecuencias.
5. Entregar conclusiones de los presentado en los datos.
6. En pacientes con distrofia muscular progresiva, la actividad de cretinquinasa sérica se eleva llamativamente sobre el valor normal de 50 unidades por litro. Los siguientes son los niveles séricos de creatiquinasa medidos en 24 pacientes con la enfermedad:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 3720 | 3795 | 3340 | 5600 | 3800 | 3580 |
| 5500 | 2000 | 1570 | 2360 | 1500 | 1840 |
| 3725 | 3790 | 3345 | 3805 | 5595 | 3575 |
| 1995 | 5505 | 2055 | 1575 | 1835 | 1505 |

Se pide:

1. Generar la tabla de frecuencias utilizando 5 intervalos.
2. Calcule las medidas de tendencia central.
3. Calcule las medidas de variación.
4. Genere el histograma de frecuencias.
5. Entregue conclusiones con respecto a los datos.
6. Las puntuaciones obtenidas en una prueba por un grupo de estudiantes son las siguientes:

3, 4, 3, 3.6, 4.4, 2.6, 2.6, 3.2, 3, 3.8, 3.6, 3, 3.2, 4, 3.2, 3, 3.6, 3.2, 2.8, 2.2

Se pide construir la tabla de frecuencias, calcular las medidas de tendencia central y generar el histograma de frecuencias.

1. Los retrasos en minutos de una ruta aérea están dados por los siguientes datos:

0,43 1,50 2,05 2,58 2,86 2,93 3,90 4,38 4,46 5,11 5,23 5,36 5,44 5,54 6,08 6,12 6,36 6,39 6,48 6,51 6,88 7,31 7,34 8,06 8,10 8,20 8,21 8,23 8,34 8,56 8,64 8,73 8,73 8,73 8,73 8,74 8,88 8,90 8,93 9,14 9,25 9,56 9,68 9,85 9,87 9,94 9,99 10,06 10,11 10,24 10,26 10,51 11,23 11,60 11,63 11,64 11,85 11,92 12,34 12,78 12,94 13,05 13,18 13,31 13,48 13,88 14,05 14,15 14,23 14,24 14,30 14,55 14,59 15,42 15,45 15,71 16,07 16,54 16,84 17,04 19,39

Obtenga la tabla de frecuencias, las medidas de tendencia central, las medidas de variabilidad, el histograma de frecuencias, el polígono de frecuencias y genere conclusiones alrededor de lo obtenido.

**Solución**

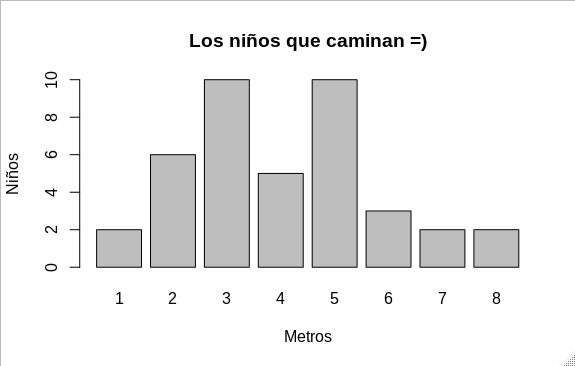
**3.**

* 1. Tabla de frecuencias.

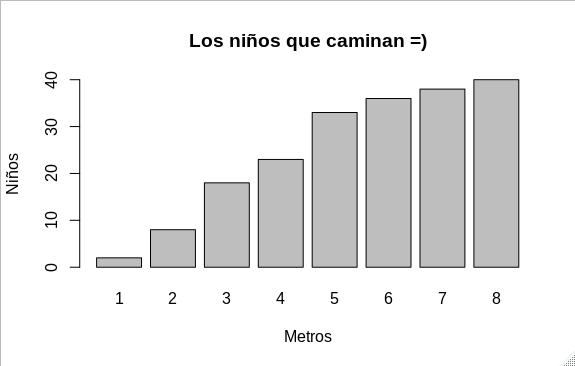
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | **fx** | **Fx** | **fax** | **Fax** |
| 1 | 2 | 2 | 0.050 | 0.050 |
| 2 | 6 | 8 | 0.150 | 0.200 |
| 3 | 10 | 18 | 0.250 | 0.450 |
| 4 | 5 | 23 | 0.125 | 0.575 |
| 5 | 10 | 33 | 0.250 | 0.825 |
| 6 | 3 | 36 | 0.075 | 0.900 |
| 7 | 2 | 38 | 0.050 | 0.950 |
| 8 | 2 | 40 | 0.050 | 1.000 |

* 1. Diagrama de barras para frecuencias absolutas, relativas y acumuladas.

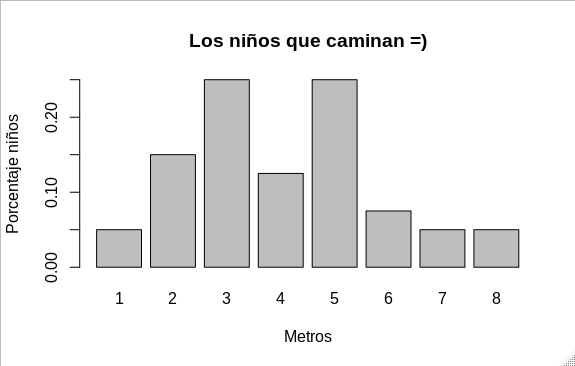
**fx**



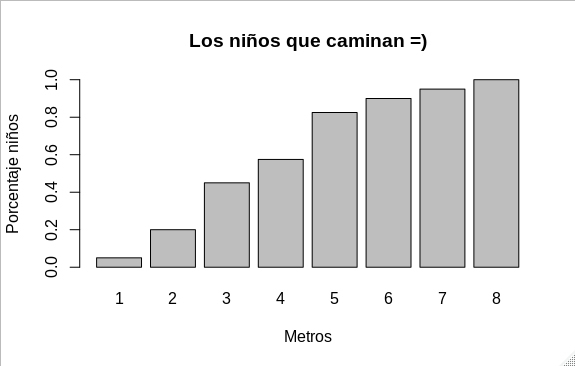
**Fx**



**fax**



**Fax**



* 1. Medidas de tendencia central (Mediana, promedio, moda y cuartiles).
     + Respecto a la cantidad de niños

Min. 1st Qu. Median Mean 3rd Qu. Max.

2 2 4 5 7 10

* + - Respecto a los metros

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.00 3.00 4.00 4.05 5.00 8.00

* 1. Medidas de desviación (Varianza y desviación estándar).
     + Respecto a la cantidad de niños

Sd: [1] 3.422614

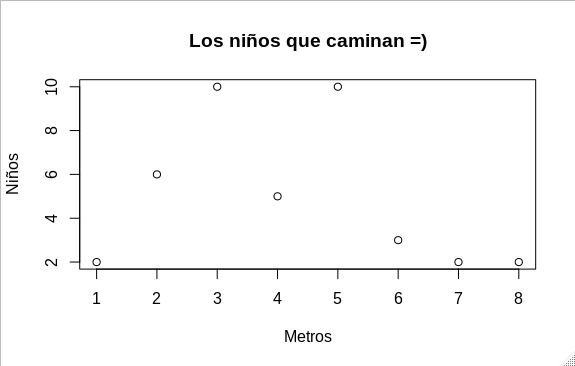
Var:[1] 11.71429

* + - Respecto a la cantidad de metros

Sd: [1] 1.782393

Var:[1] 3.176923

* 1. Gráfico de dispersión.



* 1. Entregue conclusiones de los presentado en los datos.

La mayor cantidad de niños está ubicada entre 4 y 5 metros. En promedio los niños caminaron un total de 4.05 metros con una desviación estándar de 1.78 m. Solo 25% de la muestra caminó más de 5 metros y también este mismo porcentaje caminó 1 metro o menos.

**4.**

* 1. Generar la tabla de frecuencias.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | **fx** | **Fx** | **fax** | **Fax** |
| 77.3 | 3 | 3 | 0.05357143 | 0.05357143 |
| 61.2 | 10 | 13 | 0.17857143 | 0.23214286 |
| 82.4 | 15 | 28 | 0.26785714 | 0.50000000 |
| 75.9 | 13 | 41 | 0.23214286 | 0.73214286 |
| 61.0 | 8 | 49 | 0.14285714 | 0.87500000 |
| 70.2 | 5 | 54 | 0.08928571 | 0.96428571 |
| 65.0 | 2 | 56 | 0.03571429 | 1.00000000 |
| 80.0 | 0 | 56 | 0.00000000 | 1.00000000 |

* 1. Calcular medidas de tendencia central (Mediana, promedio, moda y cuartiles).
     + Respecto a los pacientes

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 2.75 6.50 7.00 10.75 15.00

* + - Respecto a los mg

Min. 1st Qu. Median Mean 3rd Qu. Max.

61.00 61.20 75.90 72.06 82.40 82.40

* 1. Medidas de desviación (Varianza y desviación estándar).
     + Respecto a los pacientes

Sd: [1] 5.398412

Var:[1] 29.14286

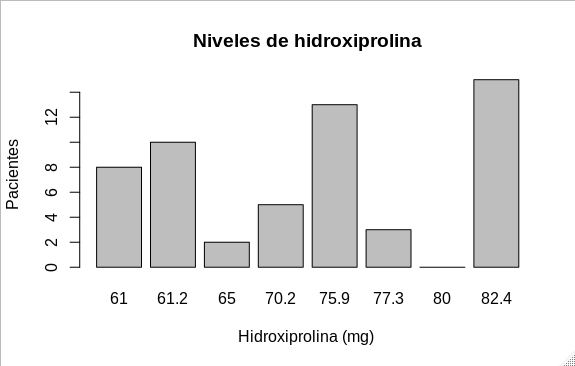
* + - Respecto a los mg

Sd: [1] 8.689355

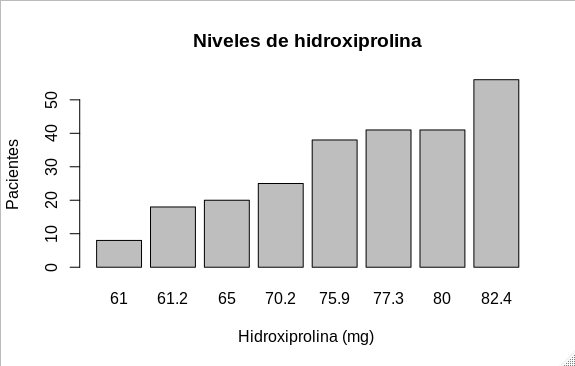
Var:[1] 75.50488

* 1. Gráfico de barras.

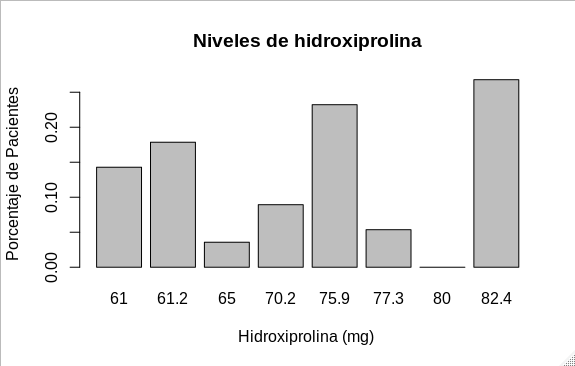
**fx:**



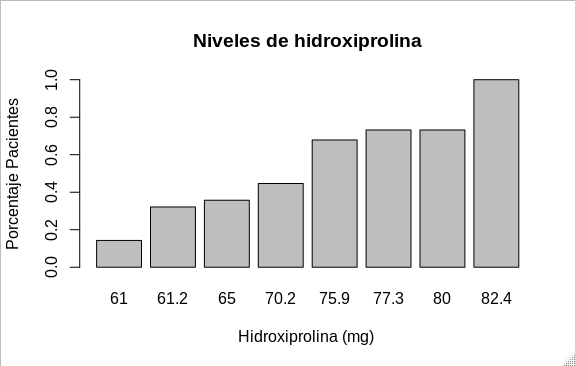
**Fx:**



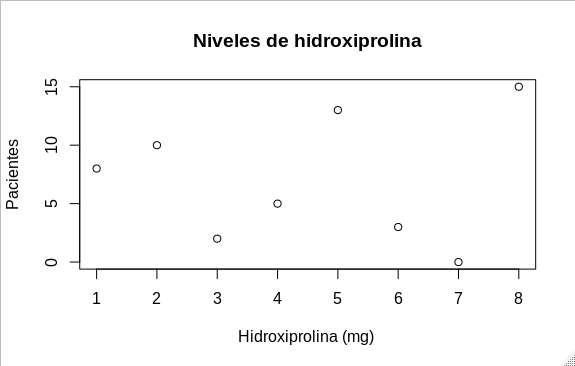
**fax:**



**Fax:**



* 1. Gráfico de dispersión.



* 1. Entregue conclusiones de lo presentado en los datos.

La mayor cantidad de Hidroxiprolina absorbida por los pacientes fue del 82.40%, frente a un promedio de 72.06%. En la gráfica podemos percibir un pequeño sesgo a la izquierda esto también es evidenciable ya que la mediana es mayor a la media.

**5.**

* 1. Generar la tabla de frecuencias por medio de la generación de 5 intervalos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | **fx** | **Fx** | **fax** | **Fax** |
| 7.260-7.288 | 2 | 2 | 0.06250 | 0.06250 |
| 7.288-7.316 | 4 | 6 | 0.12500 | 0.18750 |
| 7.316-7.344 | 17 | 23 | 0.53125 | 0.71875 |
| 7.344-7.372 | 5 | 28 | 0.15625 | 0.87500 |
| 7.372-7.400 | 4 | 32 | 0.12500 | 1.00000 |

* 1. Calcular las medidas de tendencia central.

Min. 1st Qu. Median Mean 3rd Qu. Max.

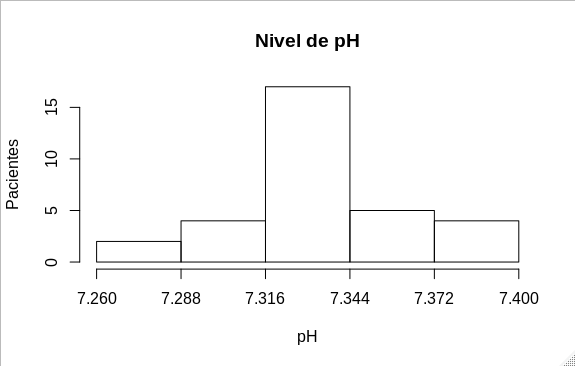
7.260 7.320 7.330 7.333 7.350 7.400

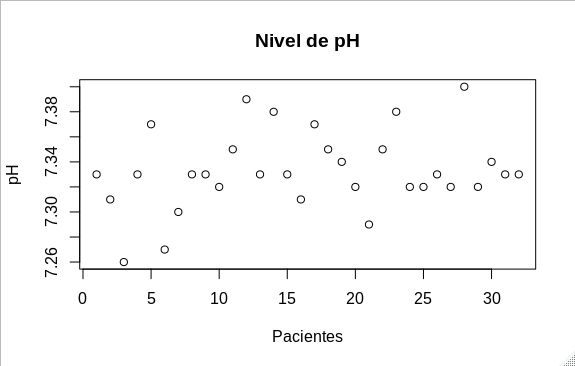
* 1. Calcular las medidas de variación.

Sd: [1] 0.03123629

Var:[1] 0.0009757056

* 1. Graficar el histograma de frecuencias.





* 1. Entregar conclusiones de los presentado en los datos.

Si analizamos la forma de la gráfica podemos pensar que los datos siguen una distribución normal, también esto se puede sustentar debido a que la media y la mediana son iguales. La mayor cantidad de individuos se encuentra en el rango de 7.316 - 7.344, en el cual se encuentra contenida su promedio.

**6.**

* 1. Generar la tabla de frecuencias utilizando 5 intervalos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | **fx** | **Fx** | **fax** | **Fax** |
| 2320 | 9 | 9 | 0.37500000 | 0.3750000 |
| 3140 | 1 | 10 | 0.04166667 | 0.4166667 |
| 3960 | 10 | 20 | 0.41666667 | 0.8333333 |
| 4780 | 0 | 20 | 0.00000000 | 0.8333333 |
| 5600 | 4 | 24 | 0.16666667 | 1.0000000 |

* 1. Calcule las medidas de tendencia central.

Min. 1st Qu. Median Mean 3rd Qu. Max.

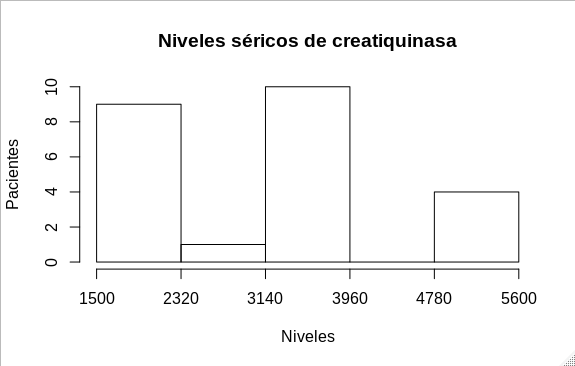
1500 1956 3460 3205 3796 5600

* 1. Calcule las medidas de variación.

Sd: [1] 1384.391

Var:[1] 1916539

* 1. Genere el histograma de frecuencias.



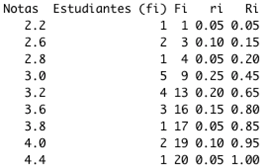
* 1. Entregue conclusiones con respecto a los datos.

El mayor número de pacientes está ubicado en el rango de 3140 a 3960, el 25% de la población tiene niveles séricos mayores a 3796. Los datos de esta muestra tienen una alta volatilidad y esto se evidencia con su varianza.

**7.**

* 1. Tabla de frecuencias.

Sin agrupar:



Grupos de 5:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | **fx** | **Fx** | **fax** | **Fax** |
| 2.64 | 3 | 3 | 0.15 | 0.15 |
| 3.08 | 6 | 9 | 0.30 | 0.45 |
| 3.52 | 4 | 13 | 0.20 | 0.65 |
| 3.96 | 4 | 17 | 0.20 | 0.85 |
| 4.40 | 3 | 20 | 0.15 | 1.00 |

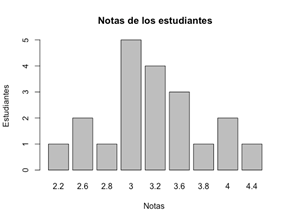
* 1. Calcular las medidas de tendencia central

Min. 1st Qu. Median Mean 3rd Qu. Max.

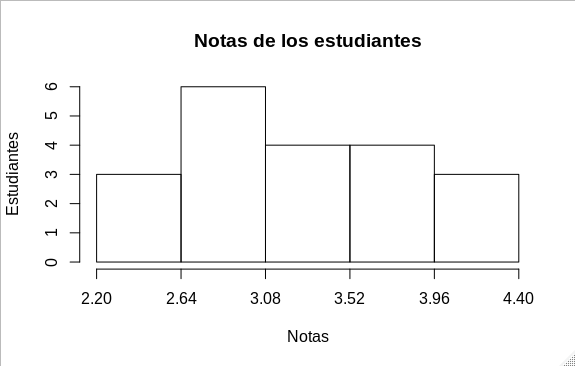
2.20 3.00 3.20 3.25 3.60 4.40

* 1. Generar el histograma de frecuencias.

Sin agrupar:



Grupos de 5:



* 1. Conclusiones

En promedio los estudiantes obtuvieron una nota del 3.25 con una desviación estándar del 0.5424. El 15% de los estudiantes perdieron el examen y un 25% tiene notas menores o iguales a 3. El desempeño del grupo no es el mejor ya que el 50% está ubicado.

**8.**

* 1. Tabla de frecuencias.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | **fx** | **Fx** | **fax** | **Fax** |
| 4.222 | 7 | 7 | 0.08641975 | 0.08641975 |
| 8.014 | 16 | 23 | 0.19753086 | 0.28395062 |
| 11.806 | 33 | 56 | 0.40740741 | 0.69135802 |
| 15.598 | 19 | 75 | 0.23456790 | 0.92592593 |
| 19.390 | 6 | 81 | 0.07407407 | 1.00000000 |

* 1. Las medidas de tendencia central.

Min. 1st Qu. Median Mean 3rd Qu. Max.

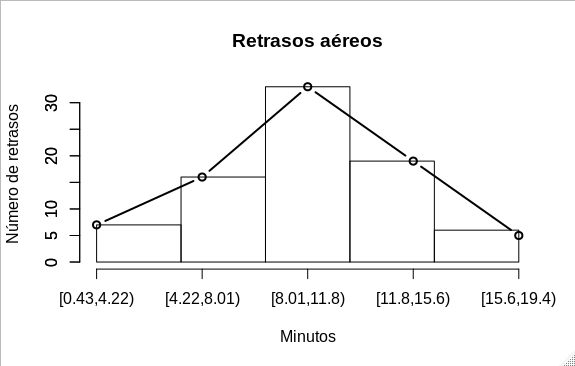
0.430 6.880 9.250 9.677 12.940 19.390

* 1. Las medidas de variabilidad.

Sd: [1] 4.042094

Var: [1] 16.33852

* 1. El histograma de frecuencias y el polígono de frecuencias.



* 1. Genere conclusiones alrededor de lo obtenido.

La gráfica aparenta que se trata de una distribución normal pero la diferencia entre la media y la mediana hace pensar que existe un sesgo a la derecha de la distribución ya que la media es mayor a la mediana. La mayor cantidad de retrasos están contenidos en el intervalo 8.01-11.8.