

|  |  |
| --- | --- |
| Colegio Universitario **IES** *Siglo 21* | |
| TRABAJO PRACTICO Nº 1 | |
| **Materia:** Probabilidad y Estadística | **Docente:** María Sofía Portillo Mongelli |
| **Modalidad:** Distancia | Condición: REGULAR |

**Reservado para el alumno**

|  |  |
| --- | --- |
| **Alumno: Vallejos Matias** | **Carrera: Simulaciones virtuales y videojuegos** |
| **DNI**: 42559030 | **Fecha límite de entrega: 17/04/20** |

|  |
| --- |
| NOTA |
|  |

###### Reservado para el docente

**CONSIDERACIONES GENERALES**

**Objetivos:**

* Utilizar la representación gráfica como herramienta para describir el comportamiento de los datos.

* Extraer conclusiones según los datos obtenidos.

**Criterios de evaluación:**

* Se evaluará:
  + La capacidad de analizar, plantear y resolver situaciones concretas empleando las herramientas de la estadística.
  + La fundamentación –en los casos que se requiera- de los pasos que ha utilizado para resolver una determinada situación.
  + La prolijidad, el orden, la ortografía y la presentación general del documento de evaluación.

**Modalidad:**

* Escrito e individual.

**Puntaje:**

1. El TP se aprueba con 4 (cuatro). Esta nota corresponde al 60% del trabajo práctico, realizado correctamente.

**Actividades:**

1. En un grupo de estudiantes se considera el número de ensayos que necesita cada uno para memorizar una lista de seis pares de palabras. Los resultados fueron:

5 8 3 9 6 7 10 6 7 4 6 9 5 6 7 9 4 6 8 7

a) Construya la tabla de frecuencias.

b) Calcule la moda, la media, y la mediana de las observaciones dadas. Obtenga la frecuencia del conjunto de los resultados superiores a 5.

c) Represente los datos a través de un histograma de frecuencias. Qué conclusiones puede obtener, a partir del gráfico?

a) *# Tabla de frecuencias 1.a*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Clases** | **Intervalos** | **Frecuencia** | **Xi . Fi** | **Frecuencia relativa** | **Porcentaje relativo** | **Frecuencia acumulada** |
| **C** | **X** | **Fi** | **Xi . Fi** | **Fr** | **%** | **F** |
| **1** | 3 | 1 | 3 | 0.05 | 5 | 1 |
| **2** | 4 | 2 | 8 | 0.1 | 10 | 3 |
| **3** | 5 | 2 | 10 | 0.1 | 10 | 5 |
| **4** | 6 | 5 | 30 | 0.25 | 25 | 10 |
| **5** | 7 | 4 | 28 | 0.2 | 20 | 14 |
| **6** | 8 | 2 | 16 | 0.1 | 10 | 16 |
| **7** | 9 | 3 | 27 | 0.15 | 15 | 19 |
| **8** | 10 | 1 | 10 | 0.05 | 5 | 20 |
|  |  | 20 | 132 | 1 | 100% |  |

b) *# Valores de posición 1.b*

[ 3 4 4 5 5 6 6 6 6 6 7 7 7 7 8 8 9 9 9 10]

* **Moda:**

*(5 repeticiones) (Unimodal)*

* **Media:**

* **Q1:**

= 5

* **Mediana (Q2):**

* **Q3:**

* **Frecuencia de el conjunto C (resultados > 5)**

y resulta que

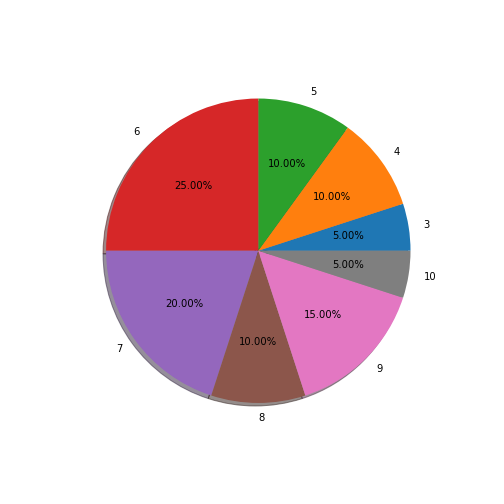
c) *# Gráficos punto 1.c*

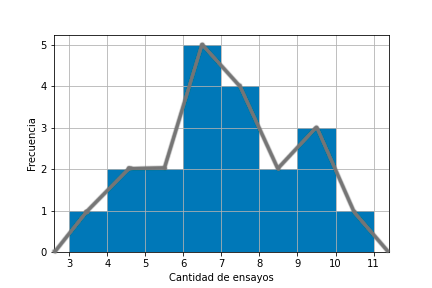
**Tabla (Referencia):**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Clases** | **N datos** | **Media** | **Min** | **Max** | **Q1** | **Mediana** | **Q3** | **Moda** |
| **Ensayos** | **20** | **6,6** | **3** | **10** | **5** | **6,5** | **8** | **6** |

**Grafico:**

*\* Frecuencias [0, .. ,5]*

*\* Aclaración se grafico con un intervalo 1 es decir [3, .. , 10 + 1] para una mejor visualización de los datos.*



*Gráficos con: Librería Matplotlib.pyplot https://matplotlib.org*

**Conclusión:**

Al observar el histograma de frecuencias se puede afirmar que la mayoría de los alumnos necesita 6,5 ensayos (moda), y que la media de los alumnos necesita 6,6 ensayos promedio para memorizar la información.

La cantidad mínima de ensayos es 3 y la máxima registrada es 10. El 50 % de los alumnos necesita + de 6,5 ensayos. Además se puede identificar un proceso centrado con una mucha variabilidad. No cuenta con excepciones o datos atípicos. Es un histograma Unimodal, es decir con una media. Por lo tanto el histograma confirma los datos y las predicciones marcadas.

2) Los niños, a diferencia de los adultos, tienden a recordar las películas, cuentos e historias como una sucesión de acciones más que el argumento en forma global y de conjunto. En el relato de una película, por ejemplo, utilizan con frecuencia las palabras "y entonces...". Una psicóloga con suprema paciencia pidió a 50 niños que le contaran una determinada película que ellos habían visto. Consideró la variable: cantidad de "y entonces..." utilizados en el relato y registró los siguientes datos:

8 15 22 19 15 17 18 20 17 12

16 16 17 21 23 18 20 21 20 20

15 18 17 19 20 23 22 10 17 19

19 21 20 18 18 24 11 19 31 16

17 18 19 20 18 18 40 18 19 16

Como parte del mismo estudio la experimentadora obtuvo de 50 adultos el mismo tipo de datos. Estos fueron:

10 12 5 8 13 10 12 8 7 9

11 10 9 9 11 15 12 17 14 10

9 8 15 16 10 14 7 16 9 1

4 11 12 7 9 10 3 11 14 8

12 5 10 9 7 11 14 10 15 9

Para ambas variables:

a) Construya la tabla de frecuencias.

b) Calcule la media, la mediana y la moda.

c) Grafique ambas distribuciones de manera que puedan ser comparadas.

d) Los puntos anteriores, ¿qué indican respecto de la conducta observada en niños y adultos?

a) *# Tabla de frecuencias 2.a*

*\* Aclaración: Para una mayor organización de las tablas se decidió trabajar con intervalos de clases calculados mediante la regla de sturges.*

* **Muestra 1 (Niños)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Clases** | **Li + Ls / 2** | **Frecuencia** | **Xi . Fi** | **Frecuencia relativa** | **Porcentaje relativo** | **Frecuencia acumulada** |
| **C** | **X** | **Fi** | **Xi . Fi** | **Fr** | **%** | **F** |
| [8 – 13) | 10,5 | 4 | 42 | 0.08 | 8 | 4 |
| [13 – 18) | 15,5 | 13 | 201.5 | 0.26 | 26 | 17 |
| [18 – 23) | 20,5 | 28 | 574 | 0.56 | 56 | 45 |
| [23 – 28) | 25,5 | 3 | 76.5 | 0.06 | 6 | 48 |
| [28 – 33) | 30,5 | 1 | 30.5 | 0.02 | 2 | 49 |
| [33 – 38) | 35,5 | 0 | 0 | 0 | 0 | 49 |
| [38 – 43] | 40,5 | 1 | 40.5 | 0.02 | 2 | **50** |
|  |  | **50** | 965 | 1,000002 | 100% |  |

* + **Rango ( Xmax – Xmin)**

* + **Intervalos (1 + 3,222\*log(x))**

7

* + **Amplitud:**

* **Muestra 2 (Adultos)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Clases** | **Li + Ls / 2** | **Frecuencia** | **Xi . Fi** | **Frecuencia relativa** | **Porcentaje relativo** | **Frecuencia acumulada** |
| **C** | **X** | **Fi** | **Xi . Fi** | **Fr** | **%** | **F** |
| [1 – 3) | 2 | 1 | 2 | 0.02 | 2 | 1 |
| [3 – 5) | 4 | 2 | 8 | 0.04 | 4 | 3 |
| [5 – 7) | 6 | 2 | 12 | 0.04 | 4 | 5 |
| [7 – 9) | 8 | 8 | 64 | 0.16 | 16 | 13 |
| [9 – 11) | 10 | 16 | 160 | 0.32 | 32 | 29 |
| [11 – 13) | 12 | 10 | 120 | 0.2 | 20 | 39 |
| [13 – 15) | 14 | 5 | 70 | 0.1 | 10 | 44 |
| [15 – 17] | 16 | 6 | 96 | 0.12 | 12 | **50** |
|  |  | **50** | 532 | 1 | 100% |  |

* + **Rango ( Xmax – Xmin)**

* + **Intervalos (1 + 3,222\*log(x))**

7

* + **Amplitud:**

b) *# Valores de posición 2.b*

* **Muestra 1 (Niños)**
* **Moda:**

*(9 repeticiones) (Unimodal)*

* **Media (Promedio):**

* **Mediana (Q2):**

* **Muestra 2 (Adultos)**
* **Moda:**

*(8 repeticiones) (Bimodal)*

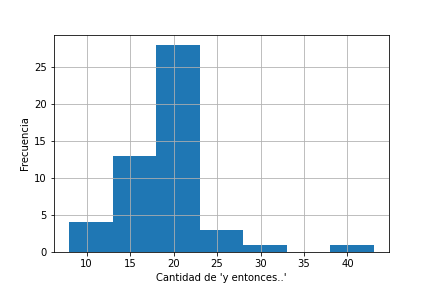
*(8 repeticiones) (Bimodal)*

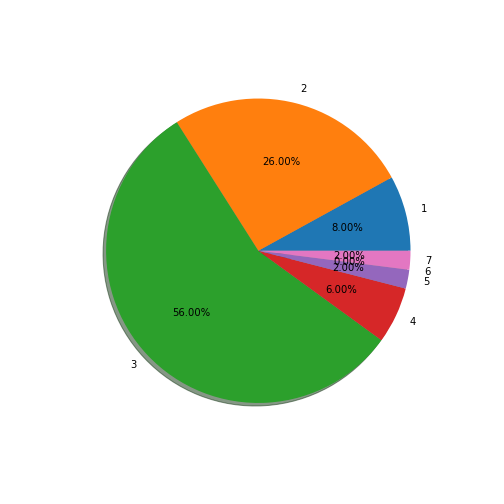
* **Media:**

* **Mediana (Q2):**

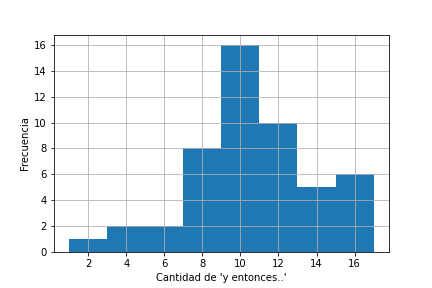
c) *# Gráficos punto 2.c*

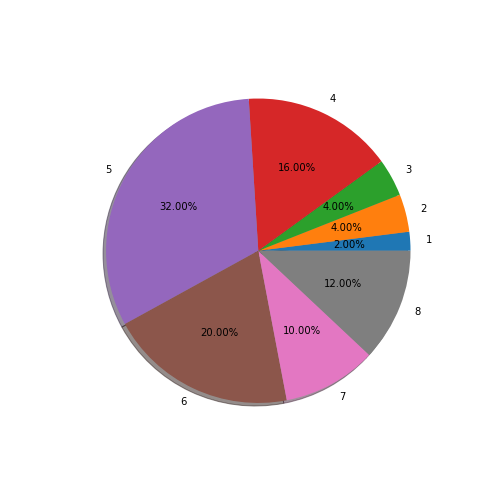
* **Muestra 1 (Niños)**





* **Muestra 2 (Adultos)**





*Gráficos con: Librería Matplotlib.pyplot https://matplotlib.org*

d) *# Conclusiones y observaciones*

Respecto a las conductas registradas referentes al estudio realizado por la psicóloga entre la distribución niños y adultos, se puede observar que los niños a diferencia de los adultos tienden a repetir con mayor frecuencia la palabra ‘y entonces..’.

En la media se encontró que los niños están en 19,3 repeticiones, mientras que los adultos se obtuvo un promedio de 10,64.

Además evaluando la moda nos indica que hay una tendencia mas exacta de 18 repeticiones en niños, mientras que en los adultos esa variable oscila entre 9 a 10.

Analizando los gráficos se puede confirmar estas tendencias diferentes entre ambas distribuciones. También se puede visualizar que en el grafico de los niños conseguimos un proceso centralizado a su izquierda y con poca variabilidad, mientras que en el grafico de los adultos tenemos un grafico centralizado con un índice de variabilidad mas alto que en el de los niños.

3) A fin de decidir cuántos mostradores de servicio se necesitarán en tiendas que serán construidas en el futuro, una cadena de supermercados quiso obtener información acerca del tiempo (en minutos) requerido para atender a los clientes. Se registró la duración de 60 casos:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3,6 | 1,9 | 2,1 | 0,3 | 0,8 | 0,2 | 1,0 | 1,4 | 1,8 | 1,6 |
| 1,1 | 1,8 | 0,3 | 1,1 | 0,5 | 1,2 | 0,6 | 1,1 | 0,8 | 1,7 |
| 1,4 | 0,2 | 1,3 | 3,1 | 0,4 | 2,3 | 1,8 | 4,5 | 0,9 | 0,7 |
| 0,6 | 2,8 | 2,5 | 1,1 | 0,4 | 1,2 | 0,4 | 1,3 | 0,8 | 1,3 |
| 1,1 | 1,2 | 0,8 | 1,0 | 0,9 | 0,7 | 3,1 | 1,7 | 1,1 | 2,2 |
| 1,6 | 1,9 | 5,2 | 0,5 | 1,8 | 0,3 | 1,1 | 0,6 | 0,7 | 0,6 |

1. ¿Cuál es la variable a analizar? ¿Es cuantitativa o cualitativa?
2. Organizarlos de la manera que le resulte más conveniente (de forma simple o en intervalos) y justificar la organización elegida.
3. Calcule los valores de posición y dispersión. Cuáles de estos valores tomará como representativos? justifique
4. Representar gráficamente las frecuencias anteriores mediante un histograma y un polígono de frecuencias.
5. Construir las frecuencias acumuladas crecientes, y graficarlas mediante una ojiva.
6. Calcule los cuartiles, e indique qué información le proporciona.
7. ¿Qué proporción de los tiempos de servicio es menor a un minuto?

a) *# Cuantitativa y continua*

La variable a analizar es la cantidad de mostradores que se necesitan para instalar. Es una variable cuantitativa ya que se puede expresar mediante un numero y es continua porque toma un numero infinito de valores ya que al medir el tiempo puede a ver muchas cantidades de posibles respuestas.

b) *# Tabla de frecuencias 3.b*

La organización elegida es por intervalos ya que al ser valores continuos se puede afirmar que la mejor forma de realizar la tabla es mediante este método.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Clases** | **Li + Ls / 2** | **Frecuencia** | **Xi . Fi** | **Frecuencia relativa** | **Porcentaje relativo** | **Frecuencia acumulada creciente** | **Desvió medio** | **Varianza** |
| **C** | **X** | **Fi** | **Xi . Fi** | **Fr** | **%** | **F** | **(Xi – X) fi** | **(Xi –X)2 fi** |
| [0,2 – 1) | 0,6 | 23 | 13,8 | 0,3833 | 38,33 | 23 | 19, 918 | 17,24 |
| [1 – 1,8) | 1,4 | 21 | 29,4 | 0,35 | 35 | 44 | 1,3986 | 0,091 |
| [1,8 – 2,6) | 2,2 | 10 | 22 | 0,1666 | 16,6666 | 54 | 7,334 | 5,387 |
| [2, 6 – 3,4) | 3 | 3 | 9 | 0,05 | 5 | 57 | 4,6002 | 7,059 |
| [3,4 – 4,2) | 3,8 | 1 | 3,8 | 0,01666 | 1,6666 | 58 | 2,334 | 5,447 |
| [4,2 – 5) | 4,6 | 1 | 4,6 | 0,01666 | 1,6666 | 59 | 3,134 | 9,821 |
| [5 – 5,8] | 5,4 | 1 | 5,4 | 0,01666 | 1,6666 | 60 | 3,934 | 15,476 |
|  |  | **60** | 88 | 1,0002 | 100% |  |  | 60,521 |

* + **Rango ( Xmax – Xmin)**

5

* + **Intervalos (1 + 3,222\*log(x))**

7

* + **Amplitud:**

c) *# Valores de posición y valores de dispersión 3.c*

* **Moda:**

*(7 repeticiones) (Unimodal)*

* **Media:**

* **Mediana (Q2):**

1. **Desvío medio**

1. **Varianza**

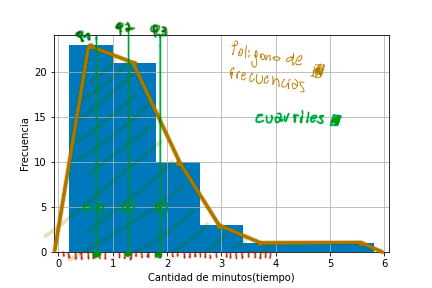
1. **Desvió estándar**

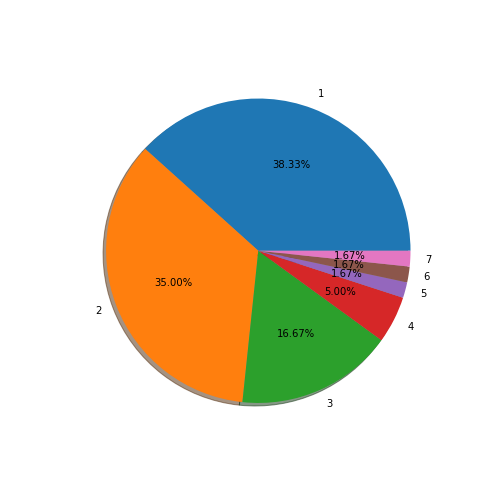
1. **Coeficiente variación**

**Conclusión:**

Los datos que tomaría y que a mi criterio son los mas representativos para analizar son la media que en este caso es 1,4666, la desviación estándar para ver que tan dispersos están los datos en este caso 1,0421 y lo tomaría en conjunto con el coeficiente de variación para poner esa desviación en términos relativos, en este caso 71% por lo que muestra una alta dispersión y variabilidad en los datos analizados.

d) *# Gráficos punto 3.d*

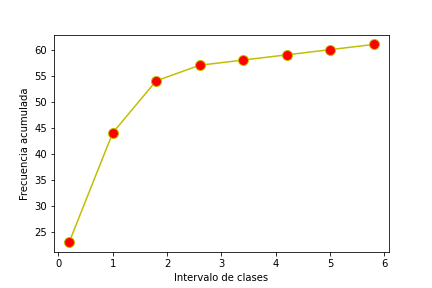




*Gráficos con: Librería Matplotlib.pyplot https://matplotlib.org*

En los gráficos vemos un proceso de alta variabilidad confirmando los datos anteriores que nos marcaban una alta dispersión.

e) *# Ojiva*



*Gráficos con: Librería Matplotlib.pyplot https://matplotlib.org*

f) *# Cuartiles 3.f*

* **Cuartil Q3:**

* **Cuartil Q2:**

* **Cuartil Q1:**

Los cuartiles son los tres valores —el 1er cuartil en 25% (Q1), el segundo cuartil en 50% (Q2 o mediana) y el tercer cuartil en 75% (Q3)— que dividen una muestra de datos ordenados en cuatro partes iguales. Estos nos ayudan a evaluar rápidamente la dispersión y la tendencia central de un conjunto de datos.

g) *# Análisis de proporción 3.g*

**Analizando los cuartiles podemos indicar que:**

El 25% de los tiempos de servicio es menor o igual que 0,721 minutos, que el 50% de los tiempos de servicio es menor o igual que 1,2666 minutos y que el 75% de los tiempos de servicio es menor o igual que 1,88 minutos.

Además al calcular los percentiles (kn/10) podemos determinar que el 40% de los tiempos de servicio son menor o igual que 1 minuto. Todo ello se puede confirmar mirando el grafico anterior del ejercicio d).

4) Una empresa consultora ha entrevistado un grupo de 50 personas a las cuales les han preguntado la edad. Se obtuvieron los siguientes datos:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 23 | 34 | 21 | 41 | 42 | 35 | 32 | 36 | 27 | 20 |
| 19 | 31 | 39 | 38 | 41 | 26 | 24 | 27 | 30 | 33 |
| 23 | 32 | 40 | 34 | 25 | 28 | 29 | 30 | 22 | 24 |
| 26 | 36 | 38 | 21 | 39 | 22 | 33 | 35 | 32 | 28 |
| 27 | 26 | 24 | 32 | 37 | 39 | 32 | 24 | 35 | 26 |

Se ha decidido organizar los datos en intervalos. A partir de estos datos, ya organizados, determine:

1. Valores de posición: media, mediana y moda. Cuál tomará como representativo de la distribución?
2. Valores de Dispersión. Cuál tomará como representativo?
3. Cuartiles. Qué información le proporcionan?
4. Representar gráficamente las frecuencias acumuladas.
5. Obtener conclusiones.

*# Tabla de frecuencias 4.a*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Clases** | **Li + Ls / 2** | **Frecuencia** | **Xi . Fi** | **Frecuencia relativa** | **Porcentaje relativo** | **Frecuencia acumulada creciente** | **Desvió medio** | **Varianza** |
| **C** | **X** | **Fi** | **Xi . Fi** | **Fr** | **%** | **F** | **(Xi – X)** | **(Xi –X)2 fi** |
| [19 – 23) | 21 | 6 | 126 | 0,12 | 12 | 6 | 59,76 | 595,20 |
| [23 – 27) | 25 | 11 | 275 | 0,22 | 22 | 17 | 65,56 | 390,73 |
| [27 – 31) | 29 | 8 | 232 | 0,16 | 16 | 25 | 15,68 | 30,73 |
| [31 – 35) | 33 | 10 | 330 | 0,2 | 20 | 35 | 20,4 | 41,61 |
| [35 – 39) | 37 | 8 | 296 | 0,16 | 16 | 43 | 48,32 | 291,85 |
| [39 – 43] | 41 | 7 | 287 | 0,14 | 14 | 50 | 63,28 | 705,61 |
|  |  | **50** | 1546 | 1 | 100% |  | 273 | 2055,73 |

* + **Rango ( Xmax – Xmin)**

23

* + **Intervalos (1 + 3,222\*log(x))**

6.643978354404254

6

* + **Amplitud:**

* **Nuevo rango:**

a) *# Valores de posición 4.a*

* **Moda:**

*(5 repeticiones) (Unimodal)*

* **Media:**

* **Mediana (Q2):**

(Fa+)

(Mediana es igual al limite superior de la Fa+ = 25)

**Conclusión valores de posición:**

El valor que tomaría como representativo de la distribución seria la media porque es siempre la mas útil siempre y cuando sea un distribución normal, es decir que los datos sean mas homogéneos y no heterogéneos. Por lo tanto estas medidas de tendencia central considero importante utilizarlas en termino relativo mirando también las tendencias y valores de dispersión es decir que tan dispersos están los datos en este caso en la muestra de las edades.

b) *# Valores de dispersión 4.b*

1. **Desvío medio**

1. **Varianza**

1. **Desvió estándar**

1. **Coeficiente variación**

**Conclusión valores de dispersión:**

Sabiendo que las medidas de dispersión tratan, a través del calculo arrojar un valor numérico que nos indique el grado de variabilidad lo que tomaría como representativo es el desvió estándar y el coeficiente de variación ya que me darían un panorama mas claro acerca de la distribución, su variabilidad y la dispersión que existe entre sus datos, permitiéndome sacar mejores conclusiones y decisiones al respecto.

En este caso puntual vemos que se tiene un conjunto de ‘baja variabilidad y de baja dispersión de los datos’ ya que el coeficiente de variación arroja un resultado de 17,63%, siendo el limite 25%, mientras que la desviación estándar arroja un resultado de 6,412 años. Por ello se puede establecer este parámetro para ver la variabilidad general del proceso de selección.

c) *# Cuartiles 4.c*

* **Cuartil Q3:**

* **Cuartil Q2:**

31

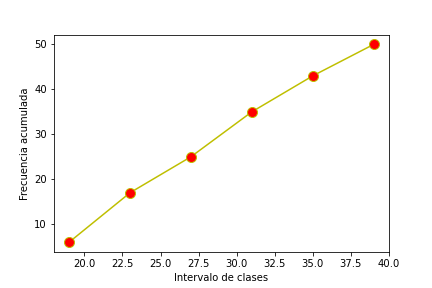
* **Cuartil Q1:**

**Conclusión cuartiles:**

Los cuartiles al dividirnos la distribución en este caso en 4 partes iguales nos permite analizar la muestra y administrar información ordenada.

En este caso puntual nos indican que el 25% de los entrevistados son menores o iguales que 25,36 años, que el 50% son menores o iguales que 31 años y que el 75% son menores o iguales que 36,25 años.

d) *# Grafico frecuencias acumuladas 4.d*

**

*Gráficos con: Librería Matplotlib.pyplot https://matplotlib.org*

e) *# Conclusiones 4.e*

Analizando el caso puntual de las edades podemos obtener que entre los entrevistados hay una media 30,92 años, mientras que en la mediana nos arroja un valor de 30,5 siendo así una distribución simétrica.

Además obteniendo las medidas de dispersión podemos ver que la distribución es homogénea con un coeficiente de variación cercano al 17,63% con una desviación estándar cercana a 6 años. Esto nos indica la baja variabilidad y dispersión de los datos.

También se puede analizar los cuartiles identificando como clave que el 75% de los entrevistados son menores o iguales que 36 años, mientras que el 25% restante son mayores que 36 siendo el máximo 41 y el mínimo 19, por lo que nos da un rango de 23 años de edad.

Por ultimo podemos ver como los cuartiles son confirmados mediante la grafica de frecuencias acumuladas en donde trazando una línea imaginaria podemos identificar esta separación entre las cuatro partes iguales siendo importante poder analizarlo para comparar esta muestra con alguna otra distribución mostrando así un papel importante donde 2 conjuntos cuyos números podrían ser diferentes.