Estadística y Probabilidad

**Tema: Estadística descriptiva**

Pantalla 1

**Conceptos básicos**

**Estadística**. Es la ciencia que estudia cómo aprender de los datos. Con ese objetivo se ocupa de desarrollar métodos para recolectar, organizar, resumir y analizar datos, así como para dar conclusiones válidas y tomar decisiones razonables. Los datos proporcionan un conocimiento más sólido y veraz que el conocimiento basado en anécdotas, creencias o intuiciones; por ello, la estadística se ha convertido en una herramienta valiosa en diversas áreas del quehacer humano y en la vida cotidiana.

**Variable**. Las características de interés en una población o una muestra se llaman variables. Como estas características no se mantienen constantes de un individuo a otro, pueden asumir más de un valor, (de ahí su nombre).

**Datos.** Son las observaciones, los valores que asumen las variables en cada individuo.

**Clasificación de las variables**:

**Variable**

Cualitativas

Cuantitativas

Nominales

Ordinales

Discretas

Continuas

No importa el orden

Si importa el orden

Toman valores enteros

Toman valores de []



**Ejemplos de los tipos de variables**:

Variable cualitativa nominal: El estado civil: soltero, casado, separado, divorciado y viudo.

Variable cualitativa ordinal: Medallas de una prueba deportiva: oro, plata y bronce.

Variable cuantitativa discreta: El número de hermanos de 5 amigos: 2, 1, 0, 1, 3.

Variable cuantitativa continua: La estatura de 5 amigos: 1.73, 1.82, 1.77, 1.69, 1.75.

Pantalla 2

**Tabla de frecuencias para datos no agrupados**

La **estadística descriptiva** se encarga de la organización, presentación y descripción de los datos recolectados, y de obtener información a partir de ellos. El objetivo de la organización de datos es acomodarlos en forma útil para revelar sus características esenciales y simplificar ciertos análisis.

La **distribución de frecuencias** es una ordenación en forma de tabla de los datos de un problema determinado, asignando a cada dato su frecuencia correspondiente. Los cuatro tipos de frecuencia son los siguientes:

1. **Frecuencia absoluta.** La frecuencia absoluta es el número de veces que aparece determinado valor en un estudio estadístico. Se representa con la letra .
2. **Frecuencia relativa.** La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se representa por y también se puede expresar en porcentaje.

Donde:

1. **Frecuencia acumulada.** La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Se representa con la letra .
2. **Frecuencia relativa acumulada.** La frecuencia relativa acumulada es el cociente entre la frecuencia acumulada de un determinado valor y el número total de datos. También se puede expresar en porcentaje.

Donde:

Utilicemos un ejemplo para identificar cada elemento de una distribución de frecuencias. En un grupo de Estadística I del CCH Oriente, los alumnos registraron las temperaturas máximas del mes de julio, los resultados fueron los siguientes:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 32 | 31 | 28 | 29 | 33 | 32 | 31 | 30 |
| 31 | 31 | 27 | 28 | 29 | 30 | 32 | 31 |
| 31 | 30 | 30 | 29 | 29 | 30 | 30 | 31 |
| 30 | 31 | 34 | 33 | 33 | 29 | 29 |  |

Ordenamos los datos de menor a mayor:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 27 | 28 | 28 | 29 | 29 | 29 | 29 | 29 |
| 29 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 |
| 32 | 32 | 32 | 33 | 33 | 33 | 34 |  |

Frecuencia absoluta:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Temperatura | Frecuencia | Frecuencia relativa | Frecuencia acumulada | Frecuencia acumulada relativa |
| 27 | 1 |  |  |  |
| 28 | 2 |  |  |  |
| 29 | 6 |  |  |  |
| 30 | 7 |  |  |  |
| 31 | 8 |  |  |  |
| 32 | 3 |  |  |  |
| 33 | 3 |  |  |  |
| 34 | 1 |  |  |  |
|  | **31** |  |  | |

**Frecuencia relativa.** La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se representa por y también se puede expresar en porcentaje.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Temperatura | Frecuencia | Frecuencia relativa | Frecuencia acumulada | Frecuencia acumulada relativa |
| 27 | 1 |  |  |  |
| 28 | 2 |  |  |  |
| 29 | 6 |  |  |  |
| 30 | 7 |  |  |  |
| 31 | 8 |  |  |  |
| 32 | 3 |  |  |  |
| 33 | 3 |  |  |  |
| 34 | 1 |  |  |  |
|  | **31** |  |  | |

Frecuencia relativa acumulada:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Temperatura | Frecuencia | Frecuencia relativa | Frecuencia acumulada | Frecuencia acumulada relativa |
| 27 | 1 |  | 1 | 0.032 |
| 28 | 2 |  | 3 | 0.097 |
| 29 | 6 |  | 9 | 0.290 |
| 30 | 7 |  | 16 | 0.516 |
| 31 | 8 |  | 24 | 0.774 |
| 32 | 3 |  | 27 | 0.871 |
| 33 | 3 |  | 30 | 0.968 |
| 34 | 1 |  | 31 | 1 |
|  | **31** |  |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Temperatura | Frecuencia | Frecuencia relativa | Frecuencia acumulada | Frecuencia acumulada relativa |
| 27 | 1 |  | 1 |  |
| 28 | 2 |  | 3 |  |
| 29 | 6 |  | 9 |  |
| 30 | 7 |  | 16 |  |
| 31 | 8 |  | 24 |  |
| 32 | 3 |  | 27 |  |
| 33 | 3 |  | 30 |  |
| 34 | 1 |  | 31 |  |
|  | **31** |  |  | |

Pantalla 3

**Tabla de frecuencias para datos agrupados**

Cuando el número de datos de estudio es grande, resulta conveniente agrupar los valores en intervalos consecutivos llamados **clases**. Estos intervalos son de la forma [Li, Ls], cuyo extremo Li es el límite inferior de la clase y el extremo Ls es el límite superior de la clase. Existen diferentes maneras de obtener el número de clases, pero la experiencia recomienda que sean entre 5 y 20 clases.

**Pasos para construir una tabla de frecuencias para datos agrupados**

**Paso 1.** Ordena todos los datos de menor a mayor.

**Paso 2.** Determinamos el **Rango** (R), que se calcula de la siguiente manera:

Donde: Xmax es el dato mayor

Xmin es el dato menor.

**Paso 3.** Determinamos el **número de** **intervalos** o **clases** (k). Hay varias formar de determinarlo, una alternativa es usando la raíz cuadrada del total de datos n (generalmente se redondea al siguiente valor). Por lo tanto:

**Paso 4.** Calculamos la **amplitud de clase** o **amplitud del intervalo** (a), que corresponde a la cantidad de datos que van en cada clase; esto se hace dividiendo el rango R entre el número de clases k (generalmente se redondea al siguiente valor).

**Paso 5.** Construimos los intervalos o clases:

Para obtener los limites inferiores Li:

* En la primera clase hacemos Li1 = Xmin (el dato más pequeño).
* Para las demás clases el límite inferior se obtiene sumando la Xmin y la amplitud.

Para obtener los limites superiores Ls:

* El límite superior de la primera clase será el valor anterior al límite inferior de la clase siguiente.
* Para las demás clases se va sumando la amplitud al límite superior de la clase anterior.

**Paso 6**. Calculamos los **límites reales de clase** o **fronteras**, son aquéllos que se obtienen restándole media unidad de medida al límite inferior de una clase y sumándole media unidad de medida al límite superior de las diferentes clases.

Ejemplo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Límite inferior | Límite superior | Frontera inferior | Frontera superior |
| 32-37 | 32 | 37 | 31.5 | 37.5 |

**Paso 7.** Calculamos las marcas de clase (M), que corresponden al punto medio cada intervalo. Se calcula sacando el promedio del límite inferior y el límite superior de cada clase o también, sacando el promedio de las fronteras de cada clase.

**Paso 8.** Para completar la tabla de frecuencias de datos agrupados se obtienen las siguientes frecuencias:

* **Frecuencia absoluta** (f), es el número de datos que caen en determinado intervalo.
* **Frecuencia relativa** (f’)
* **Frecuencia acumulada** (F)
* **Frecuencia acumulada relativa** (F’)

**Ejercicio:** Con los siguientes datos, que corresponden a los años de servicio de 60 empleados de una empresa, realiza una tabla de distribución de frecuencias para datos agrupados.

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

**Ejercicio**. Realiza la tabla de distribución de frecuencias para datos agrupados de las estaturas de los siguientes 16 alumnos.

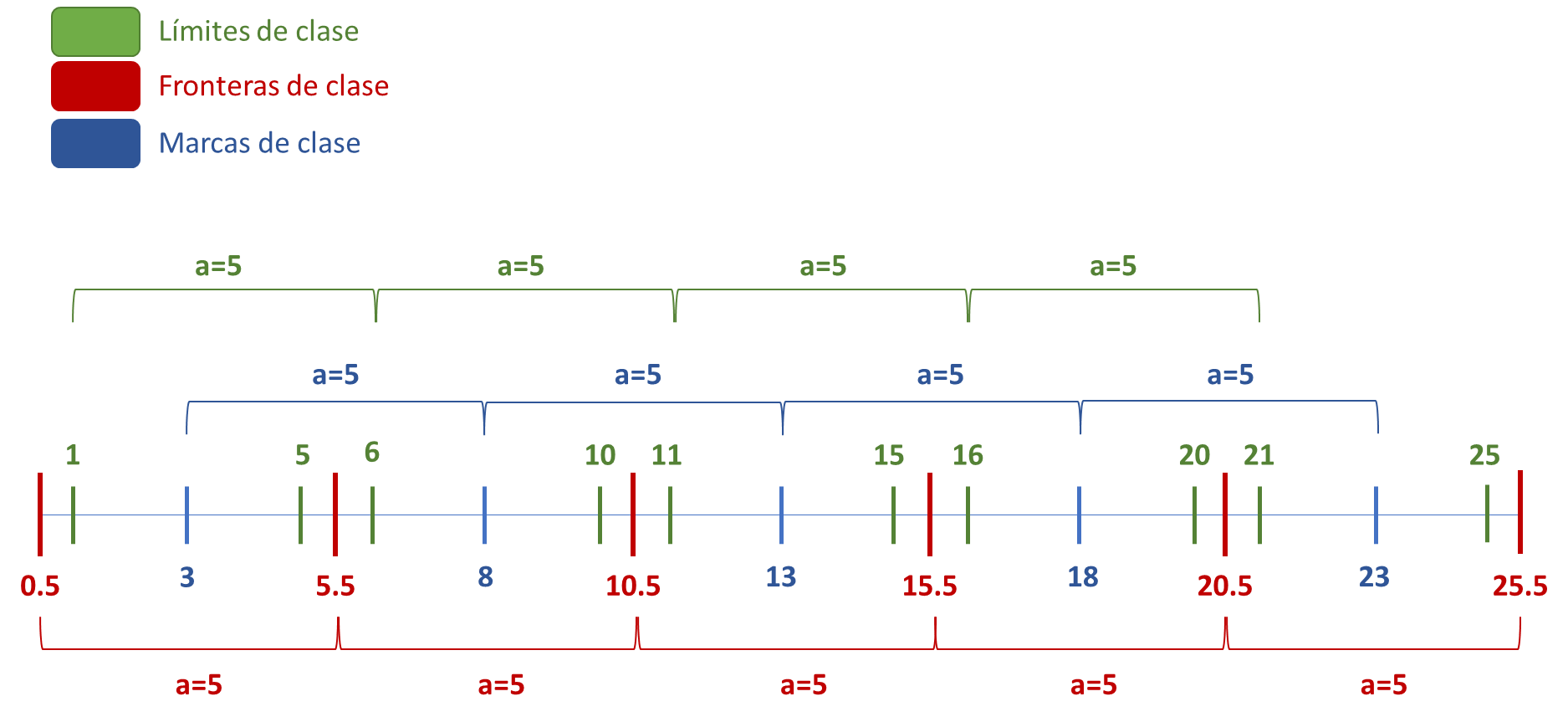
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1.58 | 1.64 | 1.79 | 1.58 | 1.64 | 1.53 | 1.64 | 1.66 |
| 1.53 | 1.52 | 1.76 | 1.57 | 1.70 | 1.74 | 1.66 | 1.52 |

**Marcas de clase**

Las marcas de clase son los puntos medios o los promedios de los límites de clase o de las fronteras de clase. La siguiente tabla muestra una distribución de frecuencias de datos agrupados, con diez clases y una amplitud de intervalo de 5 unidades.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **# de clases** | **Límites de clase** | | | **Fronteras de clase** | | | **Marcas de clase** | **f** | **f'** | **F** | **F'** |
| 1 | 1 | - | 5 | 0.5 | - | 5.5 | 3 | 1 | 0.025 | 1 | 0.025 |
| 2 | 6 | - | 10 | 5.5 | - | 10.5 | 8 | 1 | 0.025 | 2 | 0.05 |
| 3 | 11 | - | 15 | 10.5 | - | 15.5 | 13 | 3 | 0.075 | 5 | 0.125 |
| 4 | 16 | - | 20 | 15.5 | - | 20.5 | 18 | 3 | 0.075 | 8 | 0.2 |
| 5 | 21 | - | 25 | 20.5 | - | 25.5 | 23 | 3 | 0.075 | 11 | 0.275 |
| 6 | 26 | - | 30 | 25.5 | - | 30.5 | 28 | 6 | 0.15 | 17 | 0.425 |
| 7 | 31 | - | 35 | 30.5 | - | 35.5 | 33 | 7 | 0.175 | 24 | 0.6 |
| 8 | 36 | - | 40 | 35.5 | - | 40.5 | 38 | 10 | 0.25 | 34 | 0.85 |
| 9 | 41 | - | 45 | 40.5 | - | 45.5 | 43 | 4 | 0.1 | 38 | 0.95 |
| 10 | 46 | - | 50 | 45.5 | - | 50.5 | 48 | 2 | 0.05 | 40 | 1 |
|  | | | | | | | | **40** | **1** |  | |

Observa los límites de clase, las fronteras de clase, las marcas de clase y la amplitud en la recta numérica, tomemos como ejemplo las primeras cinco clases:



* Del límite inferior de una clase al límite inferior de la siguiente clase, la amplitud es la misma, en este caso es igual a 5.
* De la frontera inferior a la frontera superior de la misma clase, la amplitud es la misma, en este caso es igual a 5.
* El rango entre dos marcas de clase consecutivas debe ser igual a la amplitud, en este caso es igual a 5.

**Representaciones gráficas.**

Obtenidos los datos hay que mostrarlos de forma ordenada y comprensible. La forma más sencilla es colocarlos en una tabla, donde se muestran las variables, las categorías de cada variable y el número de eventos de cada categoría. En ciertas ocasiones, especialmente cuando trabajamos con un gran número de datos, las tablas no son prácticas y es necesario una mejor visión de los datos con una mirada rápida. Esto se consigue con los gráficos. La selección del gráfico dependerá del tipo de datos empleados. Comenzaremos con los gráficos para datos cuantitativos.

Los gráficos estadísticos nos transmiten información de modo más expresivo, nos van a permitir, con un sólo golpe de vista, entender de qué se nos habla, observar las características más importantes de un conjunto de datos e incluso sacar alguna conclusión sobre el comportamiento de la muestra donde se está realizando el estudio.

Pantalla 4

**Histograma**

Un histograma es una representación gráfica de una variable en forma de barras. Se utilizan para variables continuas o para variables discretas que están agrupadas en clases. En el eje abscisas (eje x) se construyen unos rectángulos que tienen por base la amplitud del intervalo, y por altura, la frecuencia de cada clase. Existen histogramas con frecuencias absolutas y también hay histogramas que se construyen con frecuencias acumuladas.

**Datos a utilizados para graficar:** Marcas de clase **vs** frecuencias absolutas o acumuladas.

**Polígono de Frecuencias**

Para construir el polígono de frecuencias se toma la marca de clase, que coincide con el punto medio de la parte superior de cada rectángulo del histograma y se unen los puntos mediante líneas rectas. Si deseas que el polígono de frecuencias toque el eje de las abscisas debes calcular una marca de clase anterior a la primera y una posterior a la última.

**Datos a utilizados para graficar:** Marcas de clase **vs** frecuencias absolutas o acumuladas.

**Ojiva**

La ojiva es la distribución de frecuencias de los datos, o sea, que en ella se permite ver cuántas observaciones están por encima o por debajo de ciertos valores, en vez de exhibir solo los números asignados a cada intervalo.

**Datos a utilizados para graficar:** Fronteras (frontera inferior de la primera clase y todas las fronteras superiores) **vs** frecuencias acumuladas.

**Identifica las gráficas:**

**Ejercicio:** treinta pilas AA se pusieron a prueba para determinar su tiempo de duración. Los resultados, que se muestran en minutos, se presentan continuación:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 423 | 369 | 387 | 411 | 393 | 394 | 371 | 377 | 389 | 409 |
| 392 | 408 | 431 | 401 | 363 | 391 | 405 | 382 | 400 | 381 |
| 399 | 415 | 428 | 422 | 396 | 372 | 410 | 419 | 386 | 390 |

En una hoja de tu cuaderno, realiza la tabla de distribución de frecuencias para datos agrupados y dibuja el histograma (de frecuencias absolutas y acumuladas), el polígono de frecuencias y la ojiva para dichos datos.

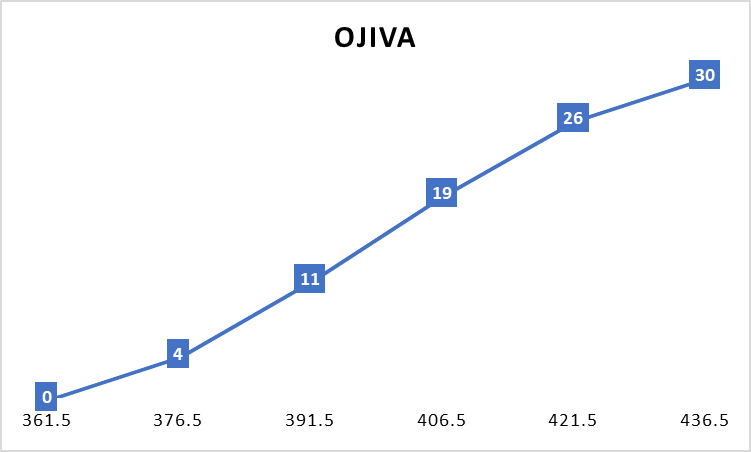
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| k | Límites | Fronteras | Marcas | f | f’ | F | F’ |
|  | - | - |  |  |  |  |  |
|  | - | - |  |  |  |  |  |
|  | - | - |  |  |  |  |  |
|  | - | - |  |  |  |  |  |
|  | - | - |  |  |  |  |  |
|  | - | - |  |  |  |  |  |

**Ejercicio:** Ahora completa la siguiente tabla de distribución de frecuencias para datos agrupados y realiza todas las gráficas vistas hasta el momento (en tu cuaderno).

**Tabla de frecuencias**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites de clase** | **Fronteras de clase** | **Marcas** | **f** | **F** | **f’** | **F’** |
| 1 | - | - **26.5** | **25** |  | **16** |  |  |
| 2 | - | - |  | **28** |  |  |  |
| 3 | - | - |  |  | **66** |  |  |
| 4 | - | - |  | **18** |  |  | **0.84** |
| 5 | - | - |  |  |  |  |  |
|  | | | |  |  |  |  |

**Ejercicio:** A partir de la siguiente gráfica, realiza la tabla de distribución de frecuencias para datos agrupados.



Pantalla 5

**Medidas de Tendencia Central.**

|  |  |
| --- | --- |
| Media aritmética | |
| **Definición**: La media aritmética es más conocida simplemente como media, y es el promedio de un conjunto de valores. Es sin duda la medida de tendencia central más utilizada, y por lo general es la más representativa. | |
| **Datos no agrupados**   |  |  |  | | --- | --- | --- | |  | = |  | |  | = |  |   Sumatoria de todos los valores del conjunto de datos divididos entre el número total de datos. | **Datos agrupados**   |  |  |  | | --- | --- | --- | |  | = |  | |  | = |  | |  | = |  |   Sumatoria del producto de las marcas de clase por si correspondiente frecuencia absoluta y dividido entre el número total de datos. |

**Ejemplo**. De igual manera, con papel y lápiz, realiza la tabla de distribución de frecuencias para datos agrupados y no agrupados de las estaturas de los siguientes 16 alumnos y calcula la media aritmética en ambos casos.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1.58 | 1.64 | 1.79 | 1.58 | 1.64 | 1.53 | 1.64 | 1.66 |
| 1.53 | 1.52 | 1.76 | 1.57 | 1.70 | 1.74 | 1.66 | 1.52 |

**Solución:**

## **Para datos no agrupados**

Primero ordenamos los datos de menor a mayor o de mayor a menor:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.52 | 1.52 | 1.53 | 1.53 | 1.57 | 1.58 | 1.58 | 1.64 | 1.64 | 1.64 | 1.66 | 1.66 | 1.70 | 1.74 | 1.76 | 1.79 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites** | | **Fronteras** | | **Marcas de clase** | **Frecuencia absoluta** | **Frecuencia relativa** | **Frecuencia acumulada** | **Frecuencia acumulada relativa** |
| 1 | 200 | 201 | 199.5 | 201.5 | 200.5 | 5 | 0.167 | 5 | 0.167 |
| 2 | 202 | 203 | 201.5 | 203.5 | 202.5 | 6 | 0.200 | 11 | 0.367 |
| 3 | 204 | 205 | 203.5 | 205.5 | 204.5 | 4 | 0.133 | 15 | 0.5 |
| 4 | 206 | 207 | 205.5 | 207.5 | 206.5 | 8 | 0.267 | 23 | 0.767 |
| 5 | 208 | 209 | 207.5 | 209.5 | 208.5 | 5 | 0.167 | 28 | 0.933 |
| 6 | 210 | 211 | 209.5 | 211.5 | 210.5 | 2 | 0.067 | 30 | 1 |
|  | | | | | | **30** | **1** |  | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites** | | **Fronteras** | | **Marcas de clase** | **F. absoluta** | **F. relativa** | **F. acumulada** | **F. acumulada relativa** |  |
|  | | | | |  |  |  |  |  |  |
| 1 | 200 | 201 | 199.5 | 201.5 | 200.5 | 5 | 0.1667 | 5 | 0.1667 | 1002.5 |
| 2 | 202 | 203 | 201.5 | 203.5 | 202.5 | 6 | 0.2000 | 11 | 0.3667 | 1215 |
| 3 | 204 | 205 | 203.5 | 205.5 | 204.5 | 4 | 0.1333 | 15 | 0.5000 | 818 |
| 4 | 206 | 207 | 205.5 | 207.5 | 206.5 | 8 | 0.2667 | 23 | 0.7667 | 1652 |
| 5 | 208 | 209 | 207.5 | 209.5 | 208.5 | 5 | 0.1667 | 28 | 0.9333 | 1042.5 |
| 6 | 210 | 211 | 209.5 | 211.5 | 210.5 | 2 | 0.0667 | 30 | 1 | 421 |
|  | | | | | | **30** | **1** |  | | **6151** |

La tabla de frecuencias para datos no agrupados es la siguiente:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dato | f | f' | F | F' | Dato x f |
| 1.52 | 2 | 0.125 | 2 | 0.125 | 3.04 |
| 1.53 | 2 | 0.125 | 4 | 0.25 | 3.06 |
| 1.57 | 1 | 0.0625 | 5 | 0.3125 | 1.57 |
| 1.58 | 2 | 0.125 | 7 | 0.4375 | 3.16 |
| 1.64 | 3 | 0.1875 | 10 | 0.625 | 4.92 |
| 1.66 | 2 | 0.125 | 12 | 0.75 | 3.32 |
| 1.7 | 1 | 0.0625 | 13 | 0.8125 | 1.7 |
| 1.74 | 1 | 0.0625 | 14 | 0.875 | 1.74 |
| 1.76 | 1 | 0.0625 | 15 | 0.9375 | 1.76 |
| 1.79 | 1 | 0.0625 | 16 | 1 | 1.79 |
|  | **16** | **1** |  |  | **26.06** |

Tenemos dos opciones para calcular la media en datos no agrupados, la primera es la tradicional, sumamos todos los datos y el resultado lo dividimos por el número total de datos:

La segunda es sumar todas las multiplicaciones de cada dato por su respectiva frecuencia absoluta y el resultado dividirlo por el número total de datos:

## **Para datos agrupados**

La tabla de frecuencias para datos agrupados es la siguiente:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Limites**  **[Li - Ls]** | | | **Fronteras** | | | | **Marcas de clase**  **M** | **Frecuencia**  **f** | **Frecuencia relativa**  **f’** | **Frecuencia Acumulada**  **F** | **Frecuencia Acumulada Relativa**  **F’** |
| 1 | 1.52 | - | 1.58 | 1.515 | | - | 1.585 | 1.55 | 7 | 0.4375 | 7 | 0.4375 |
| 2 | 1.59 | - | 1.65 | 1.585 | | - | 1.655 | 1.62 | 3 | 0.1875 | 10 | 0.625 |
| 3 | 1.66 | - | 1.72 | 1.655 | | - | 1.725 | 1.69 | 3 | 0.1875 | 13 | 0.8125 |
| 4 | 1.73 | - | 1.79 | 1.725 | | - | 1.795 | 1.76 | 3 | 0.1875 | 16 | 1 |
|  |  |  | |  |  | | |  | **16** | **1** |  |  |

Para datos agrupados, la forma de obtener la media es sumar todas las multiplicaciones de cada una de las marcas de clase por su respectiva frecuencia absoluta y el resultado lo dividimos por el número total de datos:

|  |  |
| --- | --- |
| Mediana | |
| **Definición:** Es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor. La mediana es el valor que divide el total de datos en estudio en dos conjuntos de igual tamaño, o bien, es el promedio de los dos valores centrales. | |
| **Datos no agrupados**  Primero hay que ordenar los dos, de menor a mayor o de mayor a menor. Debes encontrar la posición de la mediana de la siguiente forma:  à Cuando el número total de datos es par  à Cuando el número total de datos es impar   |  |  |  | | --- | --- | --- | |  | = |  | | **Datos agrupados**   |  |  |  | | --- | --- | --- | |  | = |  | |  | = |  | |  | = |  | |  | = |  | |  | = |  | |

**Datos no agrupados**

Primero hay que ordenar los dos, de menor a mayor o de mayor a menor. Debes encontrar la posición de la mediana de la siguiente forma:

à Cuando el número total de datos es par

à Cuando el número total de datos es impar

|  |  |  |
| --- | --- | --- |
|  | = |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | **205** |
| **206** | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

**Ejemplo**. Realiza la tabla de distribución de frecuencias para datos agrupados y no agrupados de las estaturas de los siguientes 16 alumnos y calcula la mediana en ambos casos.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1.58 | 1.64 | 1.79 | 1.58 | 1.64 | 1.53 | 1.64 | 1.66 |
| 1.53 | 1.52 | 1.76 | 1.57 | 1.70 | 1.74 | 1.66 | 1.52 |

**Solución:**

# **Para datos no agrupados**

* Tenemos 16 datos, por lo tanto, es un número de datos par y utilizamos la siguiente relación:
* Ubicamos los valores y hacemos el promedio de los datos:

**7 valores**

**7 valores**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.52 | 1.52 | 1.53 | 1.53 | 1.57 | 1.58 | 1.58 | 1.64 | 1.64 | 1.64 | 1.66 | 1.66 | 1.70 | 1.74 | 1.76 | 1.79 |

Por lo tanto:

## **Para datos agrupados**

La tabla de frecuencias para datos agrupados es la siguiente:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Limites**  **[Li - Ls]** | | | **Fronteras** | | | | **Marcas de clase**  **M** | **Frecuencia**  **f** | **Frecuencia relativa**  **f’** | **Frecuencia Acumulada**  **F** | **Frecuencia Acumulada Relativa**  **F’** |
| 1 | 1.52 | - | 1.58 | 1.515 | | - | 1.585 | 1.55 | 7 | 0.4375 | 7 | 0.4375 |
| 2 | 1.59 | - | 1.65 | 1.585 | | - | 1.655 | 1.62 | 3 | 0.1875 | 10 | 0.625 |
| 3 | 1.66 | - | 1.72 | 1.655 | | - | 1.725 | 1.69 | 3 | 0.1875 | 13 | 0.8125 |
| 4 | 1.73 | - | 1.79 | 1.725 | | - | 1.795 | 1.76 | 3 | 0.1875 | 16 | 1 |
|  |  |  | |  |  | | |  | **16** | **1** |  |  |

Para saber cuál es la **clase mediana**, debemos fijarnos en la columna de Frecuencia relativa acumulada y elegir la clase que cubra el 50% de los datos, en este caso 0.625 es la frecuencia que cubre el 0.5 o el 50% de los datos.

Identificamos los datos de la fórmula:

|  |  |  |
| --- | --- | --- |
|  | = |  |
|  | = |  |
|  | = |  |
|  | = |  |
|  | = |  |

Sustituimos los valores en la fórmula:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites** | | **Fronteras** | | **Marcas de clase** | **Frecuencia absoluta** | **Frecuencia relativa** | **Frecuencia acumulada** | **Frecuencia acumulada relativa** |
| 1 | 200 | 201 | 199.5 | 201.5 | 200.5 | 5 | 0.167 | 5 | 0.167 |
| 2 | 202 | 203 | 201.5 | 203.5 | 202.5 | 6 | 0.200 | 11 | 0.367 |
| 3 | 204 | 205 | 203.5 | 205.5 | 204.5 | 4 | 0.133 | **15** | **0.5** |
| 4 | 206 | 207 | 205.5 | 207.5 | 206.5 | 8 | 0.267 | 23 | 0.767 |
| 5 | 208 | 209 | 207.5 | 209.5 | 208.5 | 5 | 0.167 | 28 | 0.933 |
| 6 | 210 | 211 | 209.5 | 211.5 | 210.5 | 2 | 0.067 | 30 | 1 |
|  | | | | | | **30** | **1** |  | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites** | | **Fronteras** | | **Marcas de clase** | **Frecuencia absoluta** | **Frecuencia relativa** | **Frecuencia acumulada** | **Frecuencia acumulada relativa** |
| 1 | 200 | 201 | 199.5 | 201.5 | 200.5 | 5 | 0.167 | 5 | 0.167 |
| 2 | 202 | 203 | 201.5 | 203.5 | 202.5 | 6 | 0.200 | **11** | 0.367 |
| 3 | 204 | 205 | **203.5** | 205.5 | 204.5 | **4** | 0.133 | 15 | 0.5 |
| 4 | 206 | 207 | 205.5 | 207.5 | 206.5 | 8 | 0.267 | 23 | 0.767 |
| 5 | 208 | 209 | 207.5 | 209.5 | 208.5 | 5 | 0.167 | 28 | 0.933 |
| 6 | 210 | 211 | 209.5 | 211.5 | 210.5 | 2 | 0.067 | 30 | 1 |
|  | | | | | | **30** | **1** |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | = |  | = | **30** |
|  | = |  | = | **203.5** |
|  | = |  | = | **4** |
|  | = |  | = | **11** |
|  | = |  | = | **2** |

|  |  |
| --- | --- |
| Moda | |
| **Definición:** La moda de un conjunto de datos es el valor que se repite con mayor frecuencia y se denota por . Si existe más de una moda, entonces se dice que los datos tienen distribución bimodal. Si todos los datos aparecen el mismo número de veces, entonces se dice que no existe moda. | |
| **Datos no agrupados**  Contar las repeticiones que puedan existir y el que se repita mayor número de veces será la moda. | **Datos agrupados**  Se aproxima con la marca de clase del intervalo con mayor frecuencia. O se utiliza la siguiente formula:   |  |  |  | | --- | --- | --- | |  | = |  | |  | = |  | |  | = |  | |  | = |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | **207** | **207** | **207** |
| **207** | **207** | **207** | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites** | | **Fronteras** | | **Marcas de clase** | **Frecuencia absoluta** | **Frecuencia relativa** | **Frecuencia acumulada** | **Frecuencia acumulada relativa** |
| 1 | 200 | 201 | 199.5 | 201.5 | 200.5 | 5 | 0.167 | 5 | 0.167 |
| 2 | 202 | 203 | 201.5 | 203.5 | 202.5 | 6 | 0.200 | 11 | 0.367 |
| 3 | 204 | 205 | 203.5 | 205.5 | 204.5 | 4 | 0.133 | 15 | 0.5 |
| 4 | 206 | 207 | 205.5 | 207.5 | 206.5 | 8 | 0.267 | 23 | 0.767 |
| 5 | 208 | 209 | 207.5 | 209.5 | 208.5 | 5 | 0.167 | 28 | 0.933 |
| 6 | 210 | 211 | 209.5 | 211.5 | 210.5 | 2 | 0.067 | 30 | 1 |
|  | | | | | | **30** | **1** |  | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites** | | **Fronteras** | | **Marcas de clase** | **Frecuencia absoluta** | **Frecuencia relativa** | **Frecuencia acumulada** | **Frecuencia acumulada relativa** |
| 1 | 200 | 201 | 199.5 | 201.5 | 200.5 | 5 | 0.167 | 5 | 0.167 |
| 2 | 202 | 203 | 201.5 | 203.5 | 202.5 | 6 | 0.200 | 11 | 0.367 |
| 3 | 204 | 205 | 203.5 | 205.5 | 204.5 | 4 | 0.133 | 15 | 0.5 |
| 4 | 206 | 207 | 205.5 | 207.5 | 206.5 | 8 | 0.267 | 23 | 0.767 |
| 5 | 208 | 209 | 207.5 | 209.5 | 208.5 | 5 | 0.167 | 28 | 0.933 |
| 6 | 210 | 211 | 209.5 | 211.5 | 210.5 | 2 | 0.067 | 30 | 1 |
|  | | | | | | **30** | **1** |  | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Límites** | | **Fronteras** | | **Marcas de clase** | **Frecuencia absoluta** | **Frecuencia relativa** | **Frecuencia acumulada** | **Frecuencia acumulada relativa** |
| 1 | 200 | 201 | 199.5 | 201.5 | 200.5 | 5 | 0.167 | 5 | 0.167 |
| 2 | 202 | 203 | 201.5 | 203.5 | 202.5 | 6 | 0.200 | 11 | 0.367 |
| 3 | 204 | 205 | 203.5 | 205.5 | 204.5 | **4** | 0.133 | 15 | 0.5 |
| 4 | 206 | 207 | **205.5** | 207.5 | 206.5 | 8 | 0.267 | 23 | 0.767 |
| 5 | 208 | 209 | 207.5 | 209.5 | 208.5 | **5** | 0.167 | 28 | 0.933 |
| 6 | 210 | 211 | 209.5 | 211.5 | 210.5 | 2 | 0.067 | 30 | 1 |
|  | | | | | | **30** | **1** |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | = |  | = | **205.5** |
|  | = |  | = | **5** |
|  | = |  | = | **4** |
|  | = |  | = | **2** |

**Ejemplo**. Realiza la tabla de distribución de frecuencias para datos agrupados y no agrupados de las estaturas de los siguientes 16 alumnos y calcula la **moda** en ambos casos.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1.58 | 1.64 | 1.79 | 1.58 | 1.64 | 1.53 | 1.64 | 1.66 |
| 1.53 | 1.52 | 1.76 | 1.57 | 1.70 | 1.74 | 1.66 | 1.52 |

**Solución:**

## **Para datos no agrupados**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.52 | 1.52 | 1.53 | 1.53 | 1.57 | 1.58 | 1.58 | 1.64 | 1.64 | 1.64 | 1.66 | 1.66 | 1.70 | 1.74 | 1.76 | 1.79 |

## **Para datos agrupados**

La tabla de frecuencias para datos agrupados es la siguiente:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Limites**  **[Li - Ls]** | | | **Fronteras** | | | | **Marcas de clase**  **M** | **Frecuencia**  **f** | **Frecuencia relativa**  **f’** | **Frecuencia Acumulada**  **F** | **Frecuencia Acumulada Relativa**  **F’** |
| 1 | 1.52 | - | 1.58 | 1.515 | | - | 1.585 | 1.55 | 7 | 0.4375 | 7 | 0.4375 |
| 2 | 1.59 | - | 1.65 | 1.585 | | - | 1.655 | 1.62 | 3 | 0.1875 | 10 | 0.625 |
| 3 | 1.66 | - | 1.72 | 1.655 | | - | 1.725 | 1.69 | 3 | 0.1875 | 13 | 0.8125 |
| 4 | 1.73 | - | 1.79 | 1.725 | | - | 1.795 | 1.76 | 3 | 0.1875 | 16 | 1 |
|  |  |  | |  |  | | |  | **16** | **1** |  |  |

Para saber cuál es la **clase modal**, debemos fijarnos en la columna de Frecuencia absoluta y elegir la clase que tenga mayor frecuencia; si hay clases donde la mayor frecuencia se repita, se deberá sacar la moda para cada una de esas clases.

Identificamos los datos de la fórmula:

|  |  |  |
| --- | --- | --- |
|  | = |  |
|  | = |  |
|  | = |  |
|  | = |  |

Sustituimos los valores en la fórmula:

Pantalla 6

**Medidas de dispersión.**

|  |  |
| --- | --- |
| Rango | |
| **Definición:** el rango de un conjunto de datos es la diferencia entre el dato mayor y el dato menor. Generalmente se denota simplemente con la letra **R**. | |
| **Datos no agrupados**  El rango de un conjunto de datos es la diferencia entre el dato mayor y el dato menor. | **Datos agrupados**  Para datos agrupados se realiza la diferencia entre la frontera superior del último intervalo y la frontera inferior del primer intervalo. |

**Ejemplo:** La experiencia en años de 10 trabajadores se muestra a continuación, determina el **rango** para una distribución de frecuencias con datos agrupados y no agrupados.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 3 | 4 | 4 | 5 | 5 | 6 | 8 | 8 | 8 |

**Solución:**

## **Para datos no agrupados**

Para datos no agrupados, lo ideal es ordenarlos de menor a mayor y después hacer su diferencia.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 3 | 4 | 4 | 5 | 5 | 6 | 8 | 8 | 8 |

Dato mayor

Dato menor

## **Para datos agrupados**

La tabla de frecuencias para datos agrupados es la siguiente:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | Límites | | Fronteras | | M | f | f' | F | F' |
| 1 | 2 | 3.8 | 1.95 | 3.85 | 2.9 | 2 | 0.2 | 2 | 0.2 |
| 2 | 3.9 | 5.7 | 3.85 | 5.75 | 4.8 | 4 | 0.4 | 6 | 0.6 |
| 3 | 5.8 | 7.6 | 5.75 | 7.65 | 6.7 | 1 | 0.1 | 7 | 0.7 |
| 4 | 7.7 | 9.5 | 7.65 | 9.55 | 8.6 | 3 | 0.3 | 10 | 1 |
| 10 | 1 |

|  |  |
| --- | --- |
| Varianza | |
| **Definición:** la variancia de un conjunto de datos es el promedio de las distancias cuadradas de cada valor con respecto a su media. Se denota por o por dependiendo del valor que se utilice para promediar. Se divide entre cuando se considera que se tienen todos los datos posibles (población), y se divide entre cuando se tiene solo una fracción de los datos (muestra). | |
| **Datos no agrupados**   |  |  |  | | --- | --- | --- | |  | = |  | |  | = |  | |  | = |  | | **Datos agrupados**   |  |  |  | | --- | --- | --- | |  | = |  | |  | = |  | |  | = |  | |  | = |  | |  | = |  | |

**Ejemplo:** La experiencia en años de 10 trabajadores se muestra a continuación, determina el **rango** para una distribución de frecuencias con datos agrupados y no agrupados.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 3 | 4 | 4 | 5 | 5 | 6 | 8 | 8 | 8 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

**Solución:**

## **Para datos no agrupados**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 3 | 4 | 4 | 5 | 5 | 6 | 8 | 8 | 8 |

Utilizamos la fórmula:

Recordamos que la media para datos no agrupados es:

Sustituimos valores:

## **Para datos agrupados**

La tabla de frecuencias para datos agrupados es la siguiente:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | Límites | | Fronteras | |  | f | f' | F | F' |
| 1 | 2 | 3.8 | 1.95 | 3.85 | 2.9 | 2 | 0.2 | 2 | 0.2 |
| 2 | 3.9 | 5.7 | 3.85 | 5.75 | 4.8 | 4 | 0.4 | 6 | 0.6 |
| 3 | 5.8 | 7.6 | 5.75 | 7.65 | 6.7 | 1 | 0.1 | 7 | 0.7 |
| 4 | 7.7 | 9.5 | 7.65 | 9.55 | 8.6 | 3 | 0.3 | 10 | 1 |
| 10 | **1** |

Utilizamos la fórmula:

Recordado la media para datos agrupados:

Para hacer los cálculos más sencillos, podemos extender la tabla con las siguientes columnas:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | Límites | | Fronteras | |  | f | f' | F | F' |  |  |  |  |
| 1 | 2 | 3.8 | 1.95 | 3.85 | 2.9 | 2 | 0.2 | 2 | 0.2 | 5.75 | -2.85 | 8.1225 | 16.245 |
| 2 | 3.9 | 5.7 | 3.85 | 5.75 | 4.8 | 4 | 0.4 | 6 | 0.6 | 5.75 | -0.95 | 0.9025 | 3.61 |
| 3 | 5.8 | 7.6 | 5.75 | 7.65 | 6.7 | 1 | 0.1 | 7 | 0.7 | 5.75 | 0.95 | 0.9025 | 0.9025 |
| 4 | 7.7 | 9.5 | 7.65 | 9.55 | 8.6 | 3 | 0.3 | 10 | 1 | 5.75 | 2.85 | 8.1225 | 24.3675 |
| 10 | **1** |  | | | | | **45.125** |

|  |  |
| --- | --- |
| Desviación Estándar | Coeficiente de variación |
| **Definición:** es una medida de dispersión, que nos indica cuánto pueden alejarse los valores respecto al promedio (media). | **Definición:** mide la variación de los datos respecto a la media, sin tener en cuenta las unidades en la que están. El coeficiente de variación toma valores entre 0 y 1. Si el coeficiente es próximo al 0, significa que existe poca variabilidad en los datos y es una muestra muy compacta. En cambio, si tienden a 1 es una muestra muy dispersa. |
| **Para datos agrupados y no agrupados**  Raíz cuadrada de la varianza | **Para datos agrupados y no agrupados**  Dividir la desviación estándar entre la media |

**Ejemplo:** La experiencia en años de 10 trabajadores se muestra a continuación, determina la **desviación estándar** y el **coeficiente de variación** para una distribución de frecuencias con datos agrupados y no agrupados.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 3 | 4 | 4 | 5 | 5 | 6 | 8 | 8 | 8 |

**Solución:**

## **Para datos no agrupados**

La desviación estándar es la raíz cuadrada de la varianza, por lo tanto, únicamente hacemos el siguiente cálculo:

Para obtener el coeficiente de variación, dividimos la desviación estándar por la media.

Recordamos que la media para datos no agrupados es:

## **Para datos agrupados**

La desviación estándar es la raíz cuadrada de la varianza, por lo tanto, únicamente hacemos el siguiente cálculo:

Para obtener el coeficiente de variación, dividimos la desviación estándar por la media.

Recordado la media para datos agrupados:

**Ejercicio:** Un estudiante compra una torta diaria durante 20 días. Los siguientes datos muestran los precios en pesos de la torta que compró el alumno cada día con “El Tony” en su negocio afuera del CCH Oriente:

12 34 12 24 20 34 39 34 12 40

25 40 25 12 24 20 32 12 32 24

Determina las medidas de tendencia central y las medidas de dispersión para datos agrupados y no agrupados de los precios de las tortas.

Pantalla 7

**Medidas de posición.**

|  |  |
| --- | --- |
| Cuartiles | |
| **Definición:** Los cuartiles dividen al conjunto total de datos en cuatro partes iguales. | |
| **Datos no agrupados**  Se busca la posición del cuartil dentro de un conjunto de datos ordenado de menor a mayor de la siguiente manera:  En el caso en que la posición no sea un número entero, se realiza la siguiente interpolación: | **Datos agrupados**   |  |  |  | | --- | --- | --- | | *Qk* | = | Valor del cuartil | | *k* | = | Número del cuartil | | *n* | = | Número total de datos | | *Fi-1* | = | Frecuencia acumulada anterior a la clase del cuartil | | *fi* | = | Frecuencia absoluta donde se encuentra el cuartil | | *Li* | = | Frontera inferior de la clase del cuartil | | a | = | Amplitud del intervalo. | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

Se busca la posición del cuartil dentro de un conjunto de datos ordenado de menor a mayor de la siguiente manera:

**Para el cuartil 1:**

Como no es un valor entero, se está indicando que el valor del cuartil 1 está entre la posición 7 y la 8, entonces debemos hacer una interpolación:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

**Para el cuartil 2 que debe coincidir con la mediana:**

Como no es un valor entero, se está indicando que el valor del cuartil 1 está entre la posición 15 y la 16, entonces debemos hacer una interpolación:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

**Para el cuartil 3:**

Como no es un valor entero, se está indicando que el valor del cuartil 1 está entre la posición 23 y la 24, entonces debemos hacer una interpolación:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

|  |  |
| --- | --- |
| Deciles | |
| Definición: Los deciles dividen al conjunto total de datos en diez partes iguales. | |
| **Datos no agrupados**  Se busca la posición del decil dentro de un conjunto de datos ordenado de menor a mayor de la siguiente manera:  En el caso en que la posición no sea un número entero, se realiza la siguiente interpolación: | **Datos agrupados**   |  |  |  | | --- | --- | --- | | *Dk* | = | Valor del decil | | *k* | = | Número del decil | | *n* | = | Número total de datos | | *Fi-1* | = | Frecuencia acumulada anterior a la clase del decil | | *fi* | = | Frecuencia absoluta donde se encuentra el decil | | *Li* | = | Frontera inferior de la clase del decil | | a | = | Amplitud de la clase. | |

|  |  |
| --- | --- |
| Percentiles | |
| Definición: Los percentiles dividen al conjunto total de datos en cien partes iguales. | |
| **Datos no agrupados**  Se busca la posición del percentil dentro de un conjunto de datos ordenado de menor a mayor de la siguiente manera:  En el caso en que la posición no sea un número entero, se realiza la siguiente interpolación: | **Datos agrupados**   |  |  |  | | --- | --- | --- | | *Pk* | = | Percentil | | *k* | = | Número de percentil | | *n* | = | Número total de datos | | *Fi-1* | = | Frecuencia acumulada anterior a la clase del percentil | | *fi* | = | Frecuencia absoluta donde se encuentra el percentil | | *Li* | = | Frontera inferior de la clase del percentil | | a | = | Amplitud de la clase. | |

**Ejemplo:** calcularás las Medidas de Posición para datos agrupados y no agrupados del siguiente problema:

Estaturas de alumnos:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1.58 | 1.64 | 1.79 | 1.58 | 1.64 | 1.53 | 1.64 | 1.66 |
| 1.53 | 1.52 | 1.76 | 1.57 | 1.70 | 1.74 | 1.66 | 1.52 |

Tabla de frecuencias de datos agrupados

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Limites**  **[Li - Ls]** | | | **Fronteras** | | | | **Marcas de clase**  **M** | **Frecuencia**  **f** | **Frecuencia relativa**  **f’** | **Frecuencia Acumulada**  **F** | **Frecuencia Acumulada Relativa**  **F’** |
| 1 | 1.52 | - | 1.58 | 1.515 | | - | 1.585 | 1.55 | 7 | 0.4375 | 7 | 0.4375 |
| 2 | 1.59 | - | 1.65 | 1.585 | | - | 1.655 | 1.62 | 3 | 0.1875 | 10 | 0.625 |
| 3 | 1.66 | - | 1.72 | 1.655 | | - | 1.725 | 1.69 | 3 | 0.1875 | 13 | 0.8125 |
| 4 | 1.73 | - | 1.79 | 1.725 | | - | 1.795 | 1.76 | 3 | 0.1875 | 16 | 1 |
|  |  |  | |  |  | | |  | **16** | **1** |  |  |

**Solución:**

# **Cuartiles para datos no agrupados**

El primer paso es ordenar los datos de menor a mayor y localizamos la posición de los cuartiles 1, 2 y 3, de la siguiente manera:

**Primer cuartil:**

Lo cual indica que el primer cuartil está entre la posición 4 y 5, por lo tanto, definimos un valor anterior (Li) y un valor superior (Ls) para poder interpolar:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.52 | 1.52 | 1.53 | 1.53 | 1.57 | 1.58 | 1.58 | 1.64 | 1.64 | 1.64 | 1.66 | 1.66 | 1.70 | 1.74 | 1.76 | 1.79 |

Q1 se encuentra entre estos dos valores, por lo tanto, hay que interpolar.

Interpolamos de la siguiente manera:

Por lo tanto:

**Segundo cuartil:**

Lo cual indica que el primer cuartil está entre la posición 8 y 9, por lo tanto, definimos un valor anterior (Li) y un valor superior (Ls) para poder interpolar:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.52 | 1.52 | 1.53 | 1.53 | 1.57 | 1.58 | 1.58 | 1.64 | 1.64 | 1.64 | 1.66 | 1.66 | 1.70 | 1.74 | 1.76 | 1.79 |

Q2 se encuentra entre estos dos valores, por lo tanto, hay que interpolar.

Interpolamos de la siguiente manera:

Por lo tanto:

**Tercer cuartil:**

Lo cual indica que el primer cuartil está entre la posición 12 y 13, por lo tanto, definimos un valor anterior (Li) y un valor superior (Ls) para poder interpolar:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.52 | 1.52 | 1.53 | 1.53 | 1.57 | 1.58 | 1.58 | 1.64 | 1.64 | 1.64 | 1.66 | 1.66 | 1.70 | 1.74 | 1.76 | 1.79 |

Q2 se encuentra entre estos dos valores, por lo tanto, hay que interpolar.

Interpolamos de la siguiente manera:

Por lo tanto:

# **Cuartiles para datos agrupados**

**Primer cuartil:**

Identificamos la clase donde se encuentra el primer cuartil, que es la que tiene una frecuencia acumulada relativa igual al 25% o la inmediata superior al 25%:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Limites**  **[Li - Ls]** | | | **Fronteras** | | | | **Marcas de clase**  **M** | **Frecuencia**  **f** | **Frecuencia relativa**  **f’** | **Frecuencia Acumulada**  **F** | **Frecuencia Acumulada Relativa**  **F’** |
| 1 | 1.52 | - | 1.58 | 1.515 | | - | 1.585 | 1.55 | 7 | 0.4375 | 7 | 0.4375 |
| 2 | 1.59 | - | 1.65 | 1.585 | | - | 1.655 | 1.62 | 3 | 0.1875 | 10 | 0.625 |
| 3 | 1.66 | - | 1.72 | 1.655 | | - | 1.725 | 1.69 | 3 | 0.1875 | 13 | 0.8125 |
| 4 | 1.73 | - | 1.79 | 1.725 | | - | 1.795 | 1.76 | 3 | 0.1875 | 16 | 1 |
|  |  |  | |  |  | | |  | 16 | 1 |  |  |

Sustituimos en la fórmula:

**Segundo cuartil:**

Identificamos la clase donde se encuentra el segundo cuartil, que es la que tiene una frecuencia acumulada relativa igual 50% o la inmediata superior al 50%:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Limites**  **[Li - Ls]** | | | **Fronteras** | | | | **Marcas de clase**  **M** | **Frecuencia**  **f** | **Frecuencia relativa**  **f’** | **Frecuencia Acumulada**  **F** | **Frecuencia Acumulada Relativa**  **F’** |
| 1 | 1.52 | - | 1.58 | 1.515 | | - | 1.585 | 1.55 | 7 | 0.4375 | 7 | 0.4375 |
| 2 | 1.59 | - | 1.65 | 1.585 | | - | 1.655 | 1.62 | 3 | 0.1875 | 10 | 0.625 |
| 3 | 1.66 | - | 1.72 | 1.655 | | - | 1.725 | 1.69 | 3 | 0.1875 | 13 | 0.8125 |
| 4 | 1.73 | - | 1.79 | 1.725 | | - | 1.795 | 1.76 | 3 | 0.1875 | 16 | 1 |
|  |  |  | |  |  | | |  | 16 | 1 |  |  |

Sustituimos en la fórmula:

**Tercer cuartil:**

Identificamos la clase donde se encuentra el tercer cuartil, que es la que tiene una frecuencia acumulada relativa igual 75% o la inmediata superior al 75%:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **k** | **Limites**  **[Li - Ls]** | | | **Fronteras**  **[Li - Ls]** | | | | **Marcas de clase**  **M** | **Frecuencia**  **f** | **Frecuencia relativa**  **f’** | **Frecuencia Acumulada**  **F** | **Frecuencia Acumulada Relativa**  **F’** |
| 1 | 1.52 | - | 1.58 | 1.515 | | - | 1.585 | 1.55 | 7 | 0.4375 | 7 | 0.4375 |
| 2 | 1.59 | - | 1.65 | 1.585 | | - | 1.655 | 1.62 | 3 | 0.1875 | 10 | 0.625 |
| 3 | 1.66 | - | 1.72 | 1.655 | | - | 1.725 | 1.69 | 3 | 0.1875 | 13 | 0.8125 |
| 4 | 1.73 | - | 1.79 | 1.725 | | - | 1.795 | 1.76 | 3 | 0.1875 | 16 | 1 |
|  |  |  | |  |  | | |  | 16 | 1 |  |  |

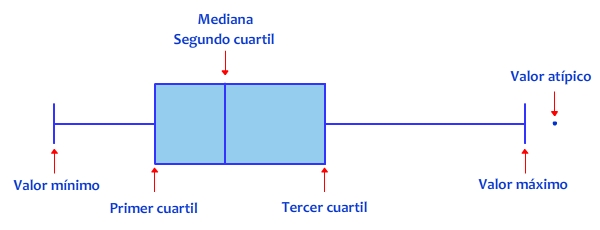
Sustituimos en la fórmula:

Pantalla 8

**Diagrama de caja y bigotes**

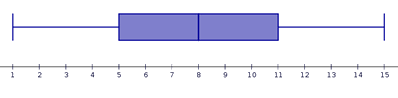
Los diagramas de caja y bigotes son una presentación visual que describe características importantes de un conjunto de datos, como la dispersión y simetría. Está formado por un rectángulo (**la caja**) y dos segmentos (**los bigotes**), uno a cada lado del rectángulo. La caja abarca el **rango** **intercuartílico**, que es el intervalo comprendido entre el primer cuartil **Q1** y el tercer cuartil **Q3**. Dentro de la caja se representa con un segmento la **mediana** de la distribución. El bigote izquierdo queda determinado por el **valor mínimo** y el primer cuartil y el derecho por el tercer cuartil y el **valor máximo**, excepto si existen **valores atípicos**.

Se consideran **valores atípicos** los menores que y los mayores que . Cuando hay valores atípicos, estos se representan fuera del diagrama de caja, mediante asteriscos o puntos.

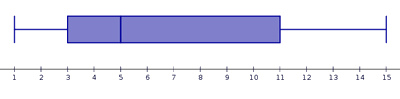


Los diagramas de caja permiten una fácil lectura e interpretación del tipo de simetría de una distribución:

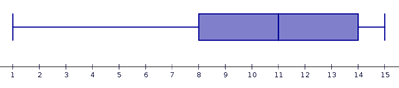
* Es simétrica cuando la media, mediana y moda de la distribución coinciden y los datos se distribuyen de igual forma a ambos lados de esas medidas.



* Es asimétrica positiva o sesgada a la derecha, cuando los datos tienden a concentrarse hacia la parte inferior de la distribución. La media se situaría a la derecha de la mediana.



* Es asimétrica negativa o sesgada hacia la izquierda cuando los datos tienden a concentrarse hacia la parte superior de la distribución. La media se situaría a la izquierda de la mediana.



**Ejemplos de interpretación de un diagrama de caja y bigotes:** Podemos obtener abundante información de una distribución a partir de estas representaciones. Supongamos que estamos hablando de edades:

* La parte izquierda de la caja es mayor que la de la derecha; ello quiere decir que las edades comprendidas entre el 25% y el 50% de la población están más dispersas que entre el 50% y el 75%.
* El bigote de la izquierda (Valor mínimo, Q1) es más corto que el de la derecha; por ello el 25% de los más jóvenes están más concentrados que el 25% de los mayores.
* El rango intercuartílico = Q3 - Q1 = 14.5; el 50 % de la población tiene 14.5 años.

**Ejercicio:** Dibuja el diagrama de caja y bigotes para los datos de las estaturas de los 16 estudiantes. Utiliza datos no agrupados.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1.58 | 1.64 | 1.79 | 1.58 | 1.64 | 1.53 | 1.64 | 1.66 |
| 1.53 | 1.52 | 1.76 | 1.57 | 1.70 | 1.74 | 1.66 | 1.52 |

|  |  |  |
| --- | --- | --- |
| Concepto | Fórmula | Cálculos |
| Valor mínimo | - | 200 |
| Valor máximo | - | 210 |
| Rango intercuartílico |  |  |
| Valores atípicos inferiores |  |  |
| Valores atípicos superiores |  |  |

Como no hay valores menores a **194** ni mayores a **216**, no existen valores atípicos. Por lo tanto, ya podemos dibujar el diagrama de la siguiente forma:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 200 | 201 | 201 | 201 | 201 |
| 202 | 202 | 203 | 203 | 203 |
| 203 | 204 | 205 | 205 | 205 |
| 206 | 206 | 207 | 207 | 207 |
| 207 | 207 | 207 | 208 | 208 |
| 208 | 209 | 209 | 210 | 210 |

Q3=207.75

Q1=202.25

210

200

Q2=205.5

210

209

208

207

206

205

204

203

202

201

200