* 1. Distribución de frecuencia

**Distribución de frecuencia para datos agrupados**

En estadística se denomina distribución de [frecuencias](https://drive.google.com/file/d/1dC3lpMMwd6BTt_ZKso6IqBfvt9AxJG0X/view?usp=sharing) a la agrupación de datos, generalmente representada en una tabla, que concentran el número de veces que tales datos se repiten, es decir, su frecuencia de aparición en cierto conjunto. La intención es observar de manera más sencilla el número de datos existentes en cada categoría de la distribución.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamenteImagínate que la FCA decide construir una nueva biblioteca e invita a participar en el grupo de diseño del proyecto a un grupo de especialistas en ergonomía que garanticen que los usuarios podrán disponer de espacios, mobiliario e instalaciones óptimos para realizar su trabajo. Entre las primeras actividades recopilan información acerca de las características y demandas de los [usuarios potenciales](https://drive.google.com/file/d/1qm1HuxZsSHgM-ANtC6TCZNARrl67frvx/view?usp=sharing). Uno de los recursos que emplean es aplicar una [encuesta](https://drive.google.com/file/d/1kaLb4ry9eU3d2GxlFkQXz8ae6PN1XshA/view?usp=sharing) entre los estudiantes, profesores y trabajadores universitarios que usarán la biblioteca. La encuesta incluye preguntas acerca de la estatura, edad y género de los encuestados y se aplica a un total de 1000 personas. Tu trabajo será organizar la información.

Hay algunas preguntas que pueden tener muchas respuestas posibles por ejemplo la estatura. Las respuestas podrían variar de centímetro en centímetro entre 1.50 y 2.00 metros por lo menos, es decir, más de ¡50 posibles respuestas! ¿Te imaginas hacer una tabla de frecuencias de 50 renglones?

Cuando la cantidad de posibles respuestas es grande -digamos más de 20- es necesario trabajar la información de una manera simplificada; para ello, lo que se acostumbra es dividir el espectro de valores posibles en intervalos, llamados clases que luego se registran en una tabla de frecuencias conocida como distribución de frecuencias agrupadas. Así, en vez de usar cada uno de los valores individuales como categorías, se trabaja con rangos de valores, de modo que se trabaja con una distribución más compacta, que permitirá que los comportamientos y tendencias puedan visualizarse mejor.

Para la construcción de una distribución de [frecuencias agrupadas](https://drive.google.com/file/d/1QZ0NdKbK7BpFSwYBpH8PBB0Y310MWQne/view?usp=sharing) primero aprenderemos a determinar las clases y con ellas construir la tabla de frecuencias hasta llegar a expresar los resultados gráficamente mediante un histograma.

Ejemplo: Revisemos las estaturas de 40 alumnos.

Aplicación, Calendario

Descripción generada automáticamente

**Determinemos los datos importantes para la construcción de la tabla de frecuencias.**

1. Rango de valores. Ordena los datos e identifica el valor máximo y mínimo. Con ellos vamos a determinar el rango. Como puedes ver, el valor mínimo es 1.51 m, el valor máximo es 1.99 m. El rango es: 1.99 - 1.51=0.48 m
2. Número de clases. Cada uno de los intervalos de valores que utilizaremos en la distribución de frecuencias agrupadas se llama clase. Es una costumbre muy frecuente dividir el rango de valores en 10 clases como máximo, pues con más clases, el trabajo puede ser muy laborioso y con menos clases se puede perder precisión (pero todo depende del tamaño de la muestra). Cuando el número de datos no es muy grande (digamos a lo más 150), una manera de determinar el número de clases es usando la siguiente fórmula:

número de clases =

Donde n es el número de datos. Así que aplicando la fórmula a nuestro ejercicio tenemos  = 6.32 ≈ 6, esto es 6 clases.

Nota importante: En algunos problemas se puede indicar el número de intervalos o clases con los que debes construir la tabla de frecuencias y en este caso lo que se debe hacer para conocer la amplitud del intervalo es dividir el rango entre el número de clases propuesta y se recomienda redondear a un número que resulte conveniente y que sea múltiplo de 5.

1. Amplitud o ancho de clase. Todos los intervalos o clases deben tener el mismo ancho y no deben superponerse, de manera que cada dato caiga dentro de solamente una clase. Para darnos una primera idea del ancho de clase dividiremos el rango entre el número de clases:

Este resultado indica que el ancho de cada clase deberá ser de 8 cm. Sin embargo, siempre que sea posible resulta práctico usar números que sean múltiplos de 5 por ello, en nuestro caso elegiremos un ancho de clase de 10 cm. Si esto no es posible, elige un número que resulte conveniente para lograr que todos los datos sean cubiertos por las clases.

1. Límites de clases. Una vez que hemos definido el número de clases y su ancho, podemos empezar a construir la tabla con la distribución de frecuencias. tendremos tantos renglones como clases y una que indique los valores que componen cada intervalo. Podemos también colocar dos columnas adicionales: una que indique específicamente cuál es el límite inferior y otra para el límite superior. Para establecer los límites de cada intervalo observamos el valor del dato mínimo que es 1.51 cm y damos un pequeño margen previo por lo que elegimos como inicio del primer intervalo 1.45 m, y como el ancho es de 10 cm, este intervalo abarcará de 1.45 a 1.54 m, el segundo abarcará de 1.55 m a 1.64 m y así sucesivamente.

Nota importante: No siempre es necesario dar un pequeño margen, regularmente esto se hace cuando los datos están dados con punto decimal.

Veamos cómo queda la tabla.

Tabla

Descripción generada automáticamente

1. Límites exactos. Existen muchas variables que son continuas, es decir, deben toman valores de entre todos los números reales todos. En nuestro caso, la estatura, es una variable continua, por lo que entre un intervalo y el siguiente no debe quedar fuera ningún número.

Si el límite superior del primer intervalo es 1.54 m y el límite inferior del segundo intervalo es 1.55 m, quedan fuera una infinidad de números, todos los que están entre estos dos números, para evitar esto, vamos a calcular intervalos exactos.

Primero hay que encontrar la diferencia entre 1.54 y 1.55

1.55 − 1.54 = 0.01

y la dividimos en 2 partes iguales, esto es

Restaremos a cada límite inferior este valor y en el límite superior lo sumaremos. De esta manera tenemos intervalos continuos y abarcaremos todos los números. Colocamos esta información en dos nuevas columnas.

Nota importante: Los intervalos o límites exactos regularmente no se incluyen en las tablas de frecuencias, solo si se piden.

1. Marcas de clase. Antes de proceder a obtener las frecuencias de cada intervalo calcularemos la marca de clase, que es el valor central de cada intervalo, y que se utiliza como valor representativo, al calcular la media y la desviación estándar en las distribuciones agrupadas.

Para calcular la marca de clase, sólo hay que sumar los límites superior e inferior del intervalo y dividir el resultado entre 2. Si se calcularon los límites exactos, estos son los que se utilizan y si no se calcularon, entonces solo se utilizan los intervalos de valores. Agregamos una columna con esta información.

Nuestra tabla quedará entonces de la manera siguiente. Observa que también hemos agregado un renglón final para la suma de valores.

Texto

Descripción generada automáticamente

Sigamos con la construcción de la tabla.

1. Obtención de las frecuencias absolutas. De la lista de datos, obtenemos cuántos de ellos hay dentro de cada clase, esta es la frecuencia absoluta de los datos. Colocamos esta información en una columna adicional en la tabla. Al terminar, es importante sumar todas las frecuencias para verificar que el total sea igual al número de datos.
2. Cálculo de la frecuencia relativa y de la frecuencia relativa acumulada. Agregamos dos nuevas columnas, una con la frecuencia relativa que se obtiene al dividir la frecuencia absoluta entre el total de los datos y redondear a dos dígitos.

Después, multiplicamos por 100 para obtener porcentajes y así la presentamos en la tabla 0.08 × 100 = 8%; 0.28 × 100 = 28%, etc. En cada renglón se calcula de la misma manera.

Para obtener la frecuencia relativa acumulada, vamos a ir sumando cada una de las celdas de frecuencia relativa con la anterior -salvo la primera- en este ejemplo tenemos para la primera celda 8%, para la segunda sumamos 8% + 28% = 36%, la siguiente celda será 36% + 23% = 59% y así sucesivamente, en la última clase debes asegurarte de tener 100%.

Nota importante: Si al sumar el total de la frecuencia relativa, no te da 100% es normal por el redondeo, de la misma en la columna de la frecuencia relativa acumulada.

Quedando nuestra tabla final de la siguiente manera:

Tabla

Descripción generada automáticamente

**Interpretación**

Frecuencia absoluta (fi): Corresponde a la cantidad de veces que se repite un dato. La letra f se refiere a la palabra frecuencia y la letra i se refiere a la realización i-ésima del experimento aleatorio.

Frecuencia absoluta acumulada (Fi): Es la suma de las frecuencias absolutas observadas hasta el intervalo i. También la puedes encontrar en algunos textos como fa.

Frecuencia relativa (hi): Corresponde a la probabilidad de pertenecer a cierta categoría. Normalmente se simboliza con la letra h, aunque algunos textos lo simbolizan como fr

Frecuencia relativa acumulada (Hi): Es la probabilidad de observar un valor menor o igual al valor que toma la variable en estudio en ese intervalo.

1.3 Representación de datos (gráficas y sus tipos; tablas)

Revisar la siguiente información [Histogramas - Unidad de Apoyo Para el Aprendizaje (unam.mx)](http://uapas1.bunam.unam.mx/matematicas/histogramas/)