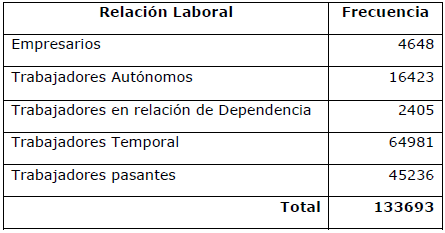
**PROBABILIDAD Y ESTADÍSTICA**

**1. Distribución de Frecuencia absoluta-relativa-relativa porcentual y acumulada.**

**1.1. Introducción de algunos conceptos**

Registraremos la cantidad de personas de una provincia de Argentina según su relación laboral.



**1.2. Frecuencias Absolutas**

Representar la característica a observar mediante la variable *X* y a la modalidad número *i* de dicha variable con la notación *X*i El número de individuos que presentan esa modalidad se llama frecuencia absoluta y se representará por *fi*.

**Propiedades de la frecuencia absoluta:**

1. Es un número entero mayor o igual a cero.

*fi ≥ 0*

1. La suma de las frecuencias absolutas es igual al tamaño de la muestra o la cantidad de datos u observaciones.

*Σ fi = n*

Ejemplo:

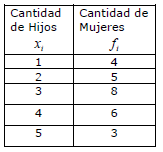
Supongamos que la variable en estudio es:

X: Cantidad de hijos por mujer.

Entonces debemos contar la cantidad de veces que se repite cada valor de la variable en estudio, frecuencia absoluta.

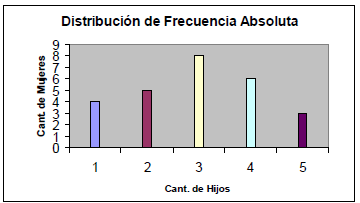
* 4 mujeres tienen 1 hijo cada una.
* 5 mujeres tienen 2 hijos cada una.
* 8 mujeres tienen 3 hijos cada una.
* 6 mujeres tienen 4 hijos cada una.
* 3 mujeres tienen 5 hijos cada una.

Construimos un cuadro de dos columnas. En la primera, colocamos los valores de la variable ordenados, generalmente de menor a mayor, y en la segunda columna, los valores de las frecuencias absolutas correspondientes.

Distribución de frecuencia absoluta

El total de mujeres con hijos observados es 26, por lo tanto, la suma de las frecuencias es:

*Σ fi = n* Como *n* = 26 entonces *Σ fi =* 26

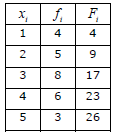
Graficamos:

**1.3. Frecuencias Absolutas Acumuladas**

Ejemplo:

Utilizando la distribución de frecuencia del primer Ejemplo, donde la variable es cantidad de hijos por mujer, construiremos una distribución de frecuencias absolutas acumuladas.

Distribución de Frecuencias Absolutas Acumuladas



Algunas interpretaciones referidas a este cuadro:

Sólo 4 mujeres tienen 1 hijo o menos, cada una, luego la frecuencia absoluta acumulada hasta el primer valor de la variable, x = 1, es:

*F*(x=1) = *F1* = 4

Sólo 9 mujeres tienen 2 hijos o menos cada una (4 que tienen 1 hijo, más 5 que tienen 2 hijos), luego la frecuencia absoluta acumulada hasta el segundo valor de la variable, x = 2, es:

*F*(x=2) = *F2* = 9

Sólo 17 mujeres tienen 3 hijos o menos cada una, luego la frecuencia absoluta acumulada hasta el tercer valor de la variable, x = 3, es:

*F*(x=3) = *F3* = 17

Sólo 23 mujeres tienen 4 hijos o menos cada una, luego la frecuencia absoluta acumulada hasta el cuarto valor de la variable, x = 4, es:

*F*(x=4) = *F4* = 23

Sólo 26 mujeres tienen 5 hijos o menos cada una, luego la frecuencia absoluta acumulada hasta el quinto valor de la variable, x = 5, es:

*F*(x=5) = *F5* = 26

**Propiedades de la frecuencia absoluta acumulada:**

* Indica el número de observaciones menores o iguales a un determinado valor de la variable.
* Es un número entero mayor o igual a cero
* Las frecuencias absolutas acumuladas forman una sucesión finita no decreciente comprendida entre 0 y n

Resumimos las tres propiedades anteriores y las escribimos en símbolos:

0 ≤ *F1* ≤ *F2* ≤ *…………… ≤* *Fs* = *n*

**1.4. Intervalos de Clases**

Para un adecuado análisis estadístico del comportamiento de una variable cuantitativa continua, es necesario agrupar a los valores individuales de ella en clases de equivalencia, llamadas Intervalos de Clases.

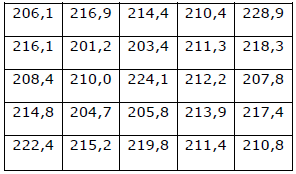
Detallaremos los pasos a seguir para determinar la cantidad de intervalos y la amplitud de cada uno de ellos:

* Primer paso: Hallamos el recorrido o rango de la variable que es la diferencia entre el Máximo (es el mayor valor que toma la variable en toda la serie estadística) y el Mínimo (es el menor valor que toma la variable en toda la serie estadística.). A esa diferencia la llamamos por ejemplo “A”.
* Segundo paso: Se determina la cantidad total de intervalos a utilizar con la siguiente fórmula:

donde *n* es el tamaño de la muestra o cantidad de datos u observaciones.

* Tercer paso: Se determina la amplitud a de cada intervalo utilizando la siguiente fórmula.
* Se simboliza con L*i* a los límites inferiores de los intervalos y L*s* a los límites superiores. A los intervalos se los considera semi-abiertos a la derecha, [L*i*; L*s*).

Ejemplo: En una empresa se realizaron 25 ventas en un día determinado, cuyos montos, en pesos, son los siguientes:



* Primer Paso: Se halla el recorrido, donde el menor valor observado es $201,2 y el mayor es $ 228,9. Entonces:

A = $230 - $200 = $30

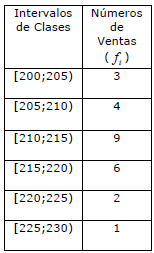
* Segundo Paso: Ahora hay que calcular la cantidad de intervalos, h



* Tercer Paso: Se determina la amplitud.



El límite inferior del primer intervalo es, entonces, 200. A partir de este número se forman los seis intervalos de amplitud 5 y se procede al conteo; como muestra el siguiente cuadro.

Distribución de frecuencias absolutas

Observamos que ambos cuadros tienen ambos el mismo título, lo que varía es el tipo de variable.

En general se definen algunas reglas para la formulación de intervalos:

* No establecer ni menos de 5 intervalos ni más de 16
* Los intervalos deben ser entre sí exhaustivos y excluyentes, es decir deben estar todos los datos incluidos en algún intervalo y cada dato debe pertenecer a un solo intervalo.
* Se deben evitar intervalos cuya frecuencia sea cero.
* No se debe perder información relevante, por ejemplo, máximos, mínimos, etc.
* Si algún valor de la variable coincide con los límites de clase de algún intervalo, se considerará en la clase inmediata superior.

**1.5. Frecuencias Relativas**

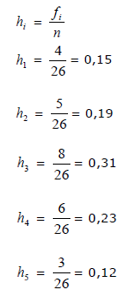
Para valorar la representatividad de cada categoría respecto al total de datos se calcula la frecuencia relativa (*hi*), dividiendo la frecuencia absoluta *fi* por el número total de observaciones (n), es decir:

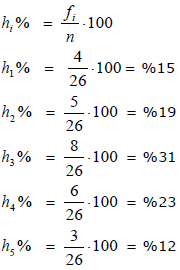
Usualmente en lugar de utilizar frecuencias relativas se utilizan porcentajes que muchos libros denominan frecuencia relativa porcentual y que no es otra cosa que la frecuencia relativa multiplicada por cien y la simbolizamos:

Ejemplo:

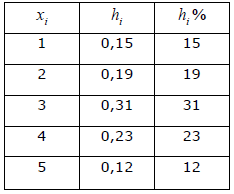
Con las distribuciones de frecuencias del primer Ejemplo, donde la variable es la cantidad de hijos por mujeres, vamos a construir las correspondientes distribuciones de frecuencia relativa y relativa porcentual.

Realizamos entonces el cociente entre la frecuencia absoluta correspondiente a cada valor de la variable y el número de observaciones.

Frecuencia relativa: Frecuencia relativa porcentual:



Distribución de frecuencias relativa y relativa porcentual



**Propiedades de la Frecuencia relativa:**

1. Es un número comprendido entre 0 y 1.

En símbolos: 0 < *hi* < 1

1. La suma de las frecuencias relativas es igual a 1.

En símbolos:

**Propiedades de la frecuencia relativa porcentual:** El concepto es el mismo de la frecuencia relativa con la diferencia de multiplicar por cien, entonces las propiedades son equivalentes.

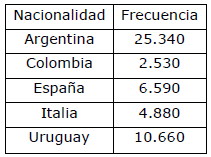
1. Es un número comprendido entre 0 y 100.

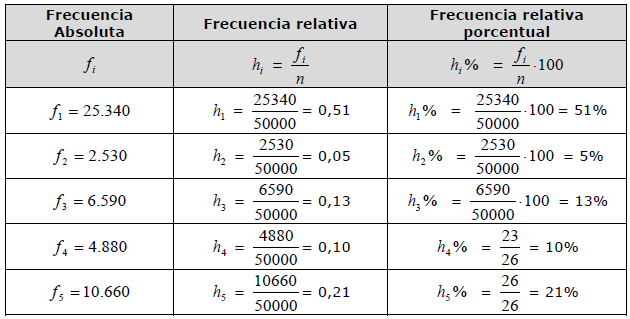
En símbolos: 0 < *hi* < 100

1. La suma de las frecuencias relativas porcentuales es igual al 100%.

En símbolos: *Σ hi* % = 100%

Ejemplo: En un estado hay 50.000 individuos de distintas nacionalidades.

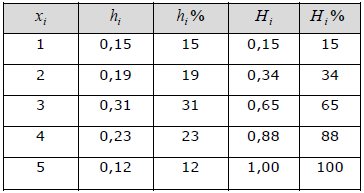
Distribución de frecuencia



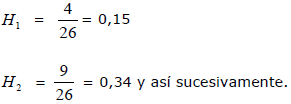
Interpretamos:

* *hi* 🡪 25. 340 individuos de un total de 50.000 son argentinos o 0,51 es la proporción de individuos que son argentinos.
* *hi* % 🡪 el 51% de individuos son argentinos.

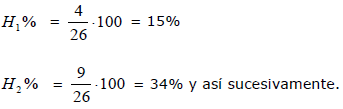
Dado el siguiente cuadro ya visto anteriormente recordando el concepto de frecuencia acumulada y razonando por analogía la frecuencia relativa acumulada será:



Frecuencia relativa acumulada:



Frecuencia relativa porcentual:



**Propiedades de la frecuencia relativa acumulada:** Como la frecuencia relativa acumulada indica la fracción de observaciones que son menores o iguales a un determinado valor de la variable, las interpretaciones son equivalentes a la de frecuencia absoluta acumulada.

**2. Medidas de Tendencia Central**

Para una variable en particular se cuenta con varios valores observados. Estos valores tienden a agruparse alrededor de algunos puntos centrales (de ahí su nombre) que fijan una posición. Estas medidas entonces resumen los datos en un solo valor que los representa. Este valor no tiene porqué corresponder a algún valor de la variable, pero *sí estar expresado en la misma magnitud.* Dentro de este tipo de medidas se incluye a la media, mediana y la moda o modo.

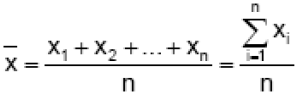
**Promedios Simples**

Los promedios pueden ser calculados sólo cuando las variables son cuantitativas, porque las variables cualitativas no admiten operaciones algebraicas.

Los promedios que se utilizan más frecuentemente son los siguientes:

* Promedio o Media Aritmética:

Es un número que resulta de sumar todos los valores observados de la variable y dividir esta suma por la cantidad de datos o tamaño de la muestra “n”.

Si x1, x2…, xn son los valores observados de la variable x, entonces

El ejemplo más sencillo sería el cálculo del promedio de notas obtenidas en un examen por un grupo de alumnos. Si los datos están organizados en una distribución de frecuencia, es decir que tengo una tabla de frecuencias entonces la fórmula será la misma, pero multiplicando cada valor de la variable por la frecuencia correspondiente. O sea:

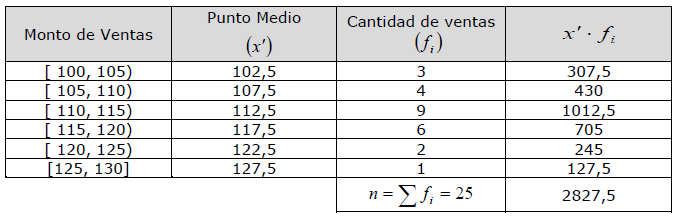
Si los datos están agrupados en intervalos, se toma el punto medio del intervalo (como el punto medio de un segmento) que se lo nota: xi'

Luego la fórmula resulta:



con

Ejemplo: Dada la siguiente distribución de frecuencia donde la variable es el monto, calcular la media aritmética.

Distribución de Frecuencia



La interpretación sería que el monto promedio por ventas es de $113,10. Es decir si todas las ventas tuviesen el mismo monto, el de cada una sería de ese precio.

El punto medio se obtiene haciendo, por ejemplo:

**Propiedades del Promedio o Media Aritmética:**

1. La media aritmética pertenece al campo de variación de la variable.



1. La media aritmética de una constante es una constante.



En símbolo:

Al realizar un cambio de unidad de medida a los datos (por ejemplo, si pasamos de gr. a Kg.), la medida estará afectada por dicho cambio de escala.

1. La media aritmética del producto de una constante por una variable es igual a la constante por la media aritmética de la variable.



En símbolo:

1. La media aritmética de la suma de una constante más una variable es igual a la constante más la media aritmética de la variable.



En símbolo:

Al efectuar un cambio en el origen desde el que se han medido todos los datos, la media quedará afectada por dicho cambio de origen.

1. La suma de desvíos dados por la diferencia entre cada valor observado de la variable y la media aritmética es igual a cero.



En símbolo:

 En una distribución de frecuencia esta propiedad se escribe:

En símbolo:

Las propiedades escritas en símbolos son las mismas para datos agrupados que para sin agrupar, solo se le agrega la *fi*.

1. La suma de los cuadrados de los desvíos con respecto a la media aritmética es un valor mínimo.



En símbolo:

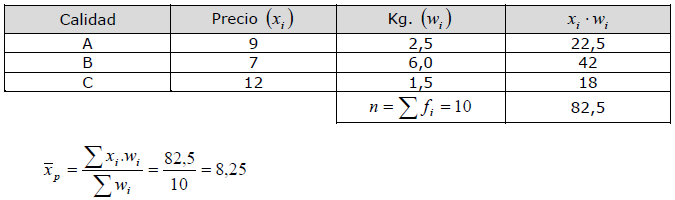
**Media aritmética ponderada:** No siempre los valores de una variable tienen la misma importancia o el mismo peso por este motivo es que recurrimos a “ponderar” es decir a darle un peso relativo a determinados valores de dicha variable.

El promedio ponderado surge de la suma del producto entre la ponderación y el valor de la variable, dividido por la suma de las ponderaciones.

Ejemplo:

Preparamos un envase con café de 3 calidades diferentes:

* 2,5kg de calidad A
* 6,0kg de calidad B
* 1,5kg de calidad C

Los precios respectivos son: $9, $7, y $12. Calcular el precio promedio por kilo del envase.

La media aritmética se ve afectada por valores extremos de la variable.