



ESTADISTICA

“La estadística es la rama de la matemática que, mediante distintas herramientas, nos permiten analizar datos cualitativos y cuantitativos, que describen el comportamiento de variables y sucesos”¹

Importancia de la Estadística

El término estadística tiene su raíz en la palabra Estado. Surge cuando se hace necesario para sus intereses cuantificar conceptos. En la mayoría de los casos esta cuantificación se hará en función de unos fines económicos o militares. El estado quiere conocer censo de personas, de infraestructura, de recursos en general, para poder obtener conclusiones de esta información.

Actualmente la estadística constituye una herramienta indispensable para la investigación científica, cualquiera que sea el área de investigación.

En su trabajo cotidiano el investigador se encuentra frente a situaciones en las cuales debe tomar decisiones en base a la observación de fenómenos. Así, por ejemplo, la estadística interviene cuando se quiere conocer el estado sanitario de un país, a través de ciertos parámetros como la tasa de mortalidad de la población.

En este caso la estadística describe la muestra en términos de datos organizados y resumidos, y luego infiere conclusiones respecto de la población.

Además, puede aplicarse a cualquier ámbito de la realidad, y por ello es utilizada en física, química, biología, medicina, astronomía, psicología, sociología, lingüística, etc.

Es así que la estadística se encuentra involucrada en todas las áreas científicas, principalmente en las ciencias sociales.

La estadística es la disciplina que estudia los modos más adecuados de recopilar, organizar, presentar, describir e interpretar datos.

Por esto, casi la totalidad de los alumnos que vayan a cursar distintas carreras (terciarias, universitarias, etc.) necesitarán estudiar y aplicar los métodos estadísticos en algún momento.

Análisis Estadísticos

Cuando se hace por ejemplo la misma pregunta a una determinada cantidad de personas, sus respuestas van a ser Datos Estadísticos. El hecho de hacer una misma pregunta a todas las personas es una “Recolección de datos estadísticos”.

Pasos del Análisis Estadístico: si bien podríamos tener distintos puntos de vista en la clasificación de estos pasos, básicamente los pasos a seguir en un análisis estadístico son los siguientes:

-  Recolección de datos: en planillas listas o simplemente en hojas en blanco, o bien por mail, teléfono, etc.
-  Organización de los datos: confección de tablas organizadas y completas.
-  Cálculo de variables: cálculo de promedios, totales y otras variables estadísticas que se verán más adelante.
-  Realización de gráficos: Gráficos circulares, de barras, cartesianos, de puntos, etc.
-  Análisis y Conclusiones: conclusiones finales del estudio.

¹ Ediciones “Logikamente”. Libros de matemática a medida. Estadística nivel II.



POBLACIÓN – MUESTRA – VARIABLE ESTADÍSTICA

Cuando se realiza un estudio estadístico, el conjunto de todas las personas o elementos que se está estudiando se denomina **POBLACIÓN**. A cada uno de los elementos de la población se lo denomina Unidad o Dato Estadístico.

Por ejemplo, en la Argentina, cada 10 años se realiza un censo para estudiar, entre otras cuestiones, determinadas características de todos los hogares del país.

Pero, como muchas veces resulta difícil estudiar una característica de todos los miembros de una población, se estudia solo una parte, a la cual se denomina **MUESTRA**. La finalidad es reducir la cantidad de información a obtener. Para que los resultados obtenidos con la muestra sean similares a los que se obtendrían estudiando toda la población, la muestra tiene que ser lo más representativa posible.

Aquellas características de una población (por ejemplo si poseen o no computadoras, si van a votar a un candidato o a otro, etc.) que se quiere analizar se llama **VARIABLE ESTADÍSTICA**.

Las variables que no se pueden medir con números se llaman **Variables Cualitativas**. Por ejemplo, si se tiene o no computadora.

Las que toman valores numéricos se denominan **Variables Cuantitativas**. Por ejemplo, la altura de las jugadoras de hockey

Las variables **cuantitativas** pueden ser **Continuas**, si sus resultados pueden tomar cualquier número (como la altura o el peso, el tiempo); o **Discretas**, si se pueden numerar, como la edad o la cantidad de hermanos.

DESCRIPCIÓN DE DATOS: DATOS AGRUPADOS Y NO AGRUPADOS

Los *datos no agrupados* son aquellos datos que pertenecen a muestras con menos de 20 elementos, por lo que para ser analizados no requieren ser agrupados

- Los datos se utilizan tal como se los recabaron, no se los clasifica.
- No generan tabla de frecuencias, dado que no tiene “mucho sentido”.

Los *datos agrupados* son aquellos que pertenecen a un tamaño de muestra mayor a 20 o más elementos, por lo que para ser analizados requieren ser agrupados en clases a partir de ciertas características.

- Su objetivo es resumir la información.
- Como se trata de muestras con más de 20 elementos, éstos requieren ser agrupados. Esto implica ordenar, clasificar de forma coherente y lógica, y expresarlos en una tabla de frecuencias.
- Esta agrupación de datos puede ser: “Simple” o mediante “Intervalos de clases”.

ORGANIZACIÓN Y RESUMEN DE DATOS: “TABLAS DE FRECUENCIAS”

Las técnicas para el resumen de datos permiten reducir la información relativa a un número muy grande de casos, a una forma muy simple que permitirá representar en qué forma están distribuidos los datos.

Se pueden organizar mediante Tablas de Distribución de Frecuencias. Dichas tablas presentan en forma ordenada a los distintos valores de una variable y sus correspondientes frecuencias.

- **FRECUENCIAS SIMPLES:** ellas son la Frecuencia Absoluta, Relativa y Porcentual.
 - **Frecuencia Absoluta:** es la cantidad de veces que se repite un determinado valor de una variable. La suma de las frecuencias es igual al número total de datos u observaciones (por ejemplo, total de personas encuestadas).}



- **Frecuencia Relativa:** es la parte que representa la frecuencia absoluta de cada valor respecto del total de datos. Se obtiene por medio del cociente entre la frecuencia absoluta y el número total de datos u observaciones.

$$\text{Frecuencia Relativa} = \frac{\text{Frecuencia absoluta}}{\text{Cantidad total de datos}}$$

La suma de las frecuencias relativas es igual a 1.

- **Frecuencia Porcentual:** es la **frecuencia** que se obtiene al multiplicar la frecuencia relativa por 100. La suma de las frecuencias porcentuales es 100.

$$\text{Frecuencia Porcentual} = \text{Frecuencia relativa} \times 100$$

- **FRECUENCIAS ACUMULADAS:** ellas son la Frecuencia Absoluta acumulada, Relativa acumulada y Porcentual acumulada.

- **Las Frecuencias Acumuladas** para un valor de la variable es la suma de las frecuencias (absolutas, relativas, porcentuales) de los valores menores o iguales al valor que se está considerando.

A continuación, vemos un ejemplo de recolección de datos y su posterior organización en una tabla de Frecuencias (simples y acumuladas). Este es un caso de DATOS EN AGRUPACIÓN SIMPLE.

Ejemplo: a los 30 alumnos de un curso se les preguntó cuántos hermanos hay en total en su casa. Y se obtuvieron los siguientes resultados:

1 - 4 - 2 - 3 - 3 - 1 - 5 - 4 - 2 - 2 - 3 - 2 - 2 - 1 - 3 - 2 - 2 - 1 - 3 - 3 - 4 - 4 - 5 - 1 - 1 - 3 - 2 - 2 - 5 - 4

Cantidad de hermanos	Frecuencia absoluta f_i	Frecuencia relativa r_i	Frecuencia porcentual p_i	Frecuencia absoluta acumulada F_i	Frecuencia relativa acumulada R_i	Frecuencia porcentual acumulada P_i
1	6	0,2	20	6	0,2	20
2	9	0,3	30	15	0,5	50
3	7	0,23	23	22	0,73	73
4	5	0,17	17	27	0,9	90
5	3	0,1	1	30	1	100
TOTAL	30	1	100			

Análisis de datos

Algunos análisis que se pueden extraer de la tabla son:

- hay 6 alumnos que tienen un hermano.
- hay 9 alumnos que tienen dos hermanos.
- El 17 % de los alumnos tienen cuatro hermanos.
- Hay 22 alumnos que tienen hasta tres hermanos.
- El 90% de los alumnos tienen como máximo 4 hermanos.



ESTADISTICA DESCRIPTIVA

TEMA

- ▶ ESTADÍSTICA DESCRIPTIVA: POBLACIÓN, MUESTRA, VARIABLES.
- ▶ AGRUPACIÓN DE DATOS: AGRUPACIÓN SIMPLE Y AGRUPACIÓN POR INTERVALOS DE CLASES.

OBJETIVOS

- ▶ DETERMINAR POBLACIÓN, MUESTRA, VARIABLE EN ESTUDIO Y TIPO DE VARIABLE.
- ▶ DIFERENCIAR LOS DISTINTOS TIPOS DE FRECUENCIAS: SIMPLES (ABSOLUTAS, RELATIVAS Y PORCENTUALES).
- ▶ ELABORACIÓN DE TABLAS DE FRECUENCIAS PARA ORGANIZAR LOS DATOS APLICADOS A LA ACTUALIDAD.

CRITERIOS

- ▶ IDENTIFICACION PERTINENTE DE LA POBLACIÓN, MUESTRAS Y VARIABLES EN LAS DISTINTAS SITUACIONES PLANTEADAS.
- ▶ COMPRENSIÓN DE LOS DISTINTOS TIPOS DE FRECUENCIAS.
- ▶ CONSTRUCCIÓN ADECUADA DE UNA TABLA DE FRECUENCIAS PARA DATOS EN AGRUPACIÓN SIMPLE.

BREVE INTRODUCCIÓN DEL TEMA

Este año la primera unidad del programa es Estadística Descriptiva.

Seguiremos trabajando los contenidos (Población, Muestra, qué es una Variable en estudio y los distintos tipos de variables) ahora aplicados a el tema de la actualidad que tanto nos preocupa: el CORONAVIRUS.

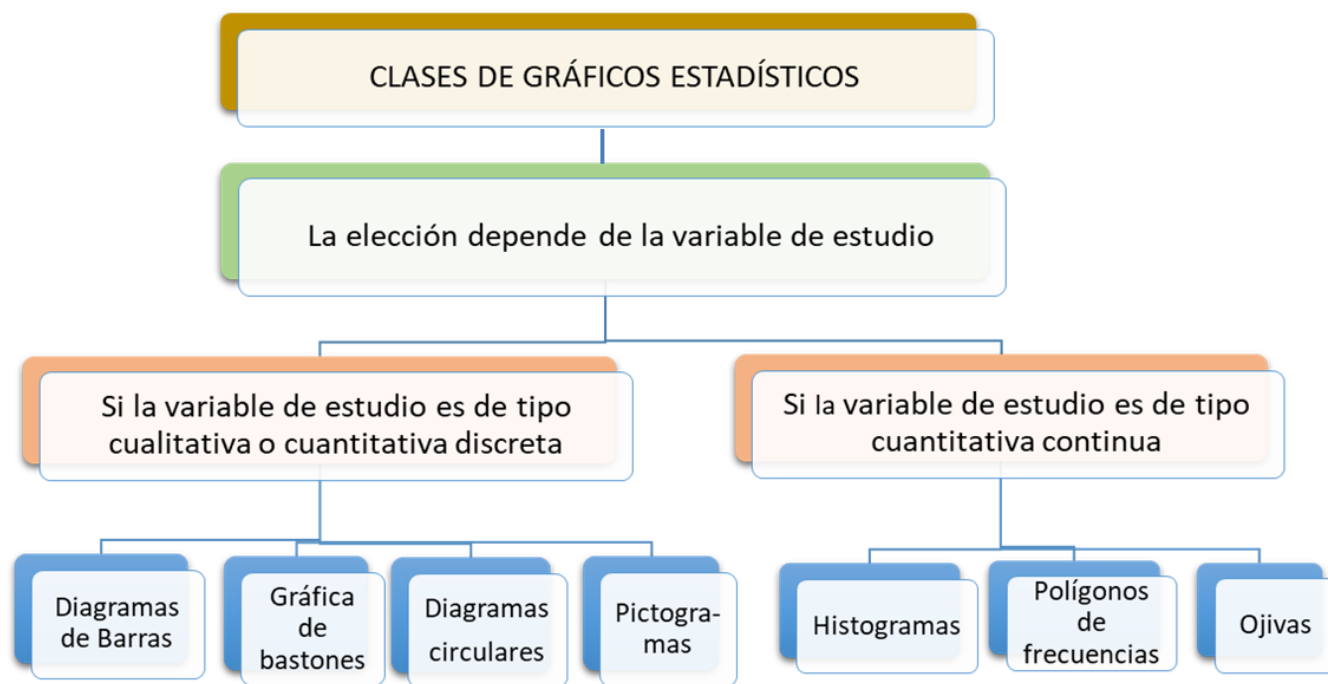
Para ello podrán seguir consultando al apunte enviado la semana pasada (en PDF) que tiene la teoría del tema como así también otras bibliografías que contengan el tema.

Así mismo, vuelvo a dejarles dos links en los cuales se explica con un ejemplo la elaboración de la tabla de Frecuencias para Datos agrupados en forma Simple:

<https://www.youtube.com/watch?v=xjTlzH6J8Vo>

<https://www.youtube.com/watch?v=cyXenZEbGz4>

ESTADÍSTICA DESCRIPTIVA – TEORIA



En general, la representación gráfica de una tabla de frecuencias permite percibir con mayor claridad algunas características de la masa de datos que se investiga. Por ello resulta más fácil transmitir conclusiones a personas no habituadas a la interpretación de distribuciones de frecuencias utilizando gráficos.

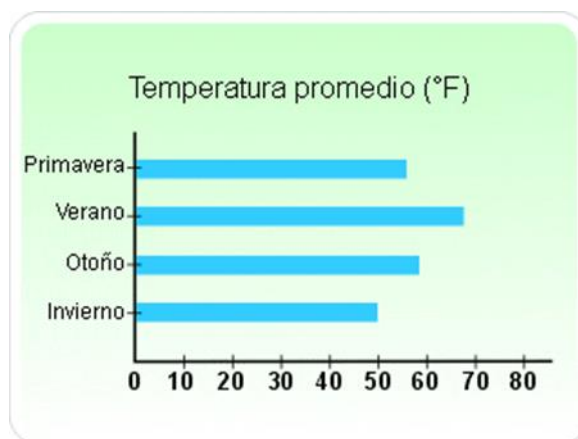
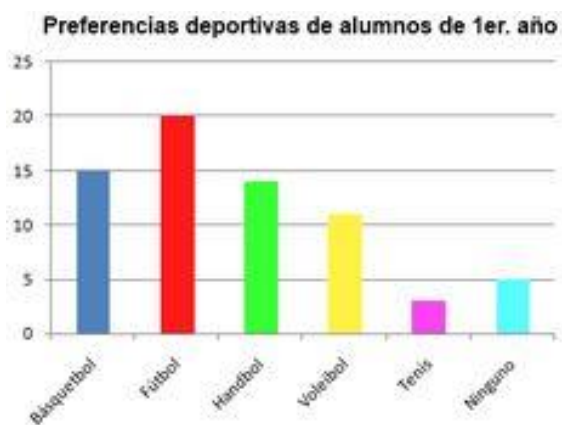
La utilidad de los gráficos es doble, ya que pueden servir no sólo como sustituto de las tablas, sino que también constituyen por sí mismos una poderosa herramienta para el análisis de los datos, siendo en ocasiones el medio más efectivo no sólo para describir y resumir la información, sino también para analizarla.

• Gráficos de Barras

Son los más utilizados para representar gráficamente. El **diagrama o gráfico de barras (verticales u horizontales)** se utiliza para variables cuantitativas discretas y cualitativas.

En el caso de las barras verticales, en el eje de las abscisas se representará la variable estudiada (en este caso, número de hermanos) y en el eje de las ordenadas las correspondientes frecuencias (las que se estén representando).

En el caso de las barras horizontales, en el eje de las abscisas se representarán las frecuencias, y en el eje de las ordenadas la variable estudiada. Las barras van de izquierda a derecha, y la longitud de cada barra es proporcional a la cantidad a representar.



• Gráficos de Bastones

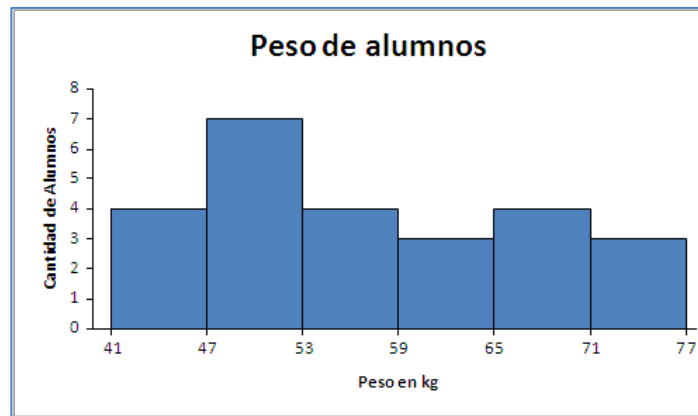
También es habitual representar por medio de **gráfico de bastones**: se utiliza generalmente para mostrar datos cuando la variable es cuantitativa discreta. Su construcción se hace levantando segmentos perpendiculares al eje de la variable (eje horizontal) y con una altura proporcional a su frecuencia absoluta o a su frecuencia relativa.

Representación gráfica de la distribución de frecuencias
Variable discreta

# hijos	F_i
0	7
1	11
2	8
3	3
4	1
Total	30



- **Histogramas:** Se utilizan para variables Cuantitativas continuas o para variables Cuantitativas discretas, con un gran número de datos, y que se han agrupado en intervalos de clases.
Formado por barras consecutivas cuyo ancho es proporcional a la amplitud de cada intervalo, y su altura es proporcional a la cantidad correspondiente al intervalo; es decir, la altura es la frecuencia absoluta de cada intervalo.

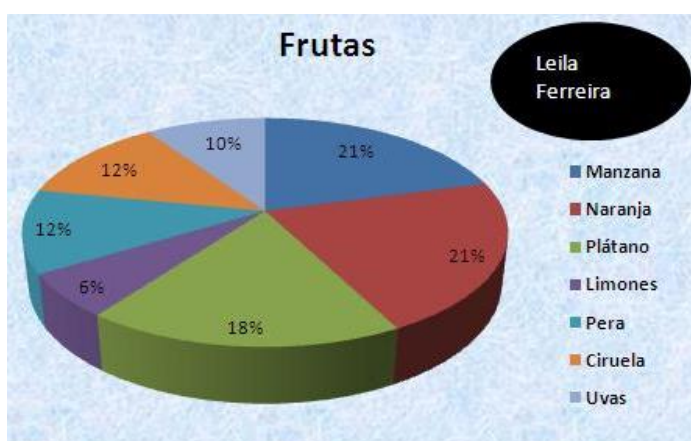


• GRÁFICOS CIRCULARES

Para representar porcentajes es muy útil el **gráfico circular o de sectores**: se utilizan para representar tanto variables Cualitativas como así también para variables Cuantitativas discretas.

Para construir un gráfico circular se divide un círculo en tantas partes como conceptos se están estudiando. Si tenemos en cuenta que el círculo completo representará la cantidad de la muestra y que el ángulo central del círculo es de 360° , obtendremos los valores de los ángulos si realizamos, para cada concepto, la siguiente operación:

$$\text{Frecuencia relativa simple} \times 360^\circ = \text{Ángulo}$$



Para realizar el gráfico circular trazamos una circunferencia y un radio y representamos, consecutivamente, los ángulos obtenidos.

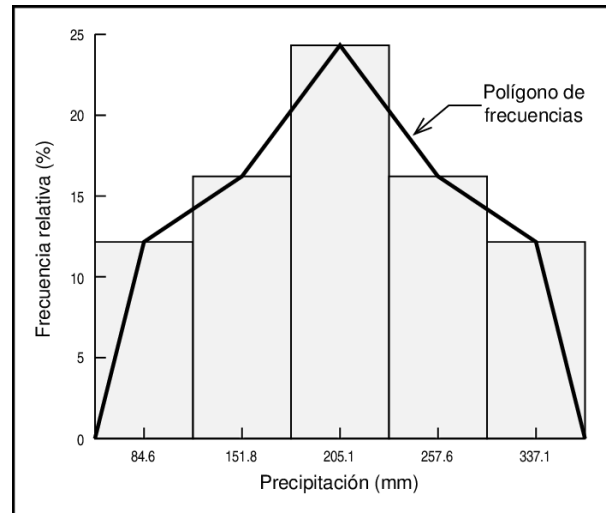
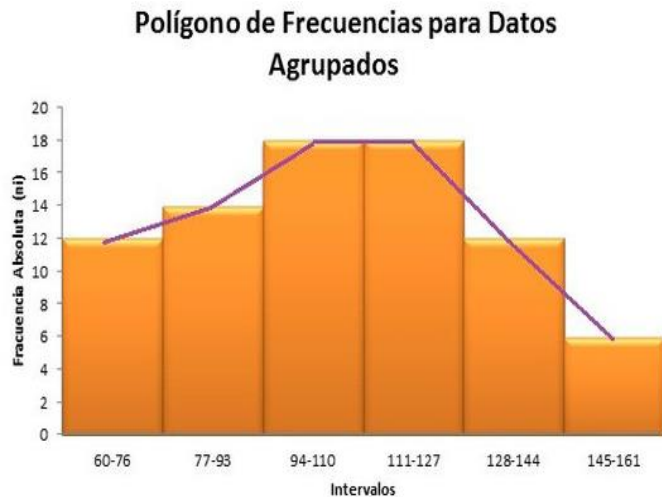
Cada uno de los sectores corresponde, en este ejemplo, a una fruta.

El gráfico circular debe presentar:

- los porcentajes correspondientes a cada sector.
- Referencias de cada sector (frutas en este caso).

• POLÍGONO DE FRECUENCIAS.

Anteriormente ya se definió lo que es un Histograma. El Polígono **de Frecuencias** es un gráfico de línea que se obtiene a partir de un Histograma, dado que se forma uniendo con segmentos las marcas de clase que coinciden con los puntos medios del extremo superior de las barras que forman el histograma. En los gráficos pueden identificarse los extremos o las **marcas de clases** de los intervalos.



• OJIVAS.

Las **Ojivas** son gráficos de líneas (ascendentes o descendentes) que se obtienen uniendo puntos cuyas coordenadas serán el límite superior de un intervalo y su correspondiente frecuencia acumulada. Estas frecuencias acumuladas pueden ser absolutas o relativas.

En el eje horizontal se marcan los límites superiores de los intervalos y en el eje vertical se indican las frecuencias Acumuladas. La curva puede comenzar en el total de observaciones y terminar con frecuencia cero, o bien comenzar con frecuencia cero y terminar con el total de las observaciones.





ESTADÍSTICA DESCRIPTIVA – TEORÍA

En los trabajos prácticos anteriores se ha enviado la teoría necesaria para ir trabajando el tema. En este trabajo se agregará algunas teorías para recordar a qué se refieren las Frecuencias Acumuladas, y la confección de Tablas de Frecuencias y los gráficos correspondientes cuando los datos se agrupan en INTERVALOS DE CLASES

DATOS AGRUPADOS EN INTERVALOS DE CLASES.

Puede ocurrir que al trabajar con variables cuantitativas continuas y en algunos casos con variables cuantitativas discretas, la cantidad de datos sea demasiado grande. Cuando trabajamos con variables cuantitativas continuas solemos tener muchos valores cercanos entre sí con poca frecuencia cada uno.

Ante estas situaciones, para intentar representar las características más importantes de los fenómenos que se están estudiando se divide el campo de variación de los datos en partes llamadas Intervalos; es decir, los datos se compactan y se tratan no como un valor individual, sino como pertenecientes a un grupo o conjunto de valores. Esta forma de presentar los datos es lo que se denomina una Distribución de DATOS AGRUPADOS EN INTERVALOS DE CLASE: ahora no interesa el valor particular de la variable, sino lo que ocurre dentro del intervalo. Ahora la frecuencia pertenece al intervalo.

Para ello, los intervalos tienen igual amplitud.

Aparecen además algunos parámetros importantes:

- **Límites de clase:** cada clase es un intervalo que va desde el límite inferior, hasta el límite superior.
- **Marca de clase:** es el punto medio de cada intervalo, y representa a la clase para el cálculo de algunos parámetros. Esto se utilizará más adelante.
- **Amplitud de clase:** es la diferencia entre el límite superior y el límite inferior.
- **En el límite inferior del Intervalo** se coloca **CORCHETE**, y en el **límite superior** se coloca **PARÉNTESIS**, excepto en el último intervalo donde en el **límite superior también se coloca corchete**.

A continuación sugiero ver un video que ayudará para la confección de la tabla de Frecuencias agrupados en intervalos de clases:

<https://www.youtube.com/watch?v=5XKf9vguOJE>

La primera parte del video por ahora no lo vamos a trabajar, pero es importante que ya lo vean. Los ejercicios ya estarán con las tablas y sus intervalos. Uds. tendrán que completar las frecuencias.

Si bien se pierde información de los datos con los intervalos de clase, se gana mucho:

- Con los datos formados en una distribución de frecuencias, las comparaciones entre dos o más series pueden efectuarse con mayor facilidad.
- Los cuadros de frecuencias son indispensables para acelerar los cálculos de muchas medidas descriptivas (media aritmética, mediana, moda entre otras).



Ejemplo:

Un grupo de atletas se está preparando para una maratón siguiendo una dieta muy estricta. A continuación, viene el peso en kilogramos que ha logrado bajar cada atleta gracias a la dieta y ejercicios:

0,2 8,4 14,3 6,5 3,4 4,6 9,1 4,3 3,5 1,5

6,4 15,2 16,1 19,8 5,4 12,1 9,6 8,7 12,1 3,2

Con los datos obtenidos se elabora una **tabla de frecuencias con 5 intervalos** de clases.

PESO (en Kg)	MARCA DE CLASE	Frecuencia absoluta simple	Frecuencia relativa simple	Frecuencia porcentual simple	Frecuencia absoluta Acumulada	Frecuencia relativa Acumulada	Frecuencia porcentual Acumulada
[0 – 4)	2	5	0,25	25	5	0,25	25
[4 – 8)	6	5	0,25	25	10	0,50	50
[8 – 12)	10	4	0,20	20	14	0,70	70
[12 – 16)	14	4	0,20	20	18	0,90	90
[16 – 20]	18	2	0,10	10	20	1	100
TOTAL		20	1	100			

FRECUENCIAS ACUMULADAS: ABSOLUTA, RELATIVA Y PORCENTUAL.

En el cuadro anterior pueden verse las frecuencias, tanto simples como las acumuladas.

→ Frecuencia absoluta acumulada

La **frecuencia absoluta acumulada** es la suma de las frecuencias absolutas que se va acumulando hasta ese dato, es decir, la frecuencia absoluta acumulada de un dato en concreto se obtiene sumando su frecuencia absoluta a las frecuencias absolutas de los datos que son menores que él.

Para diferenciarla de la frecuencia absoluta simple se representa con F_a (F mayúscula).

Se calcula **sumando la frecuencia absoluta de un dato más la frecuencia absoluta del dato anterior**. Por tanto, la **frecuencia absoluta acumulada del primer dato coincide con su frecuencia absoluta** y la **frecuencia absoluta acumulada del último dato coincide con el número total de datos**.

→ Frecuencia relativa acumulada

La **frecuencia relativa acumulada** es el mismo concepto que para la frecuencia absoluta acumulada. Se representa con F_r

Al igual que con la frecuencia absoluta, se **suma el dato de la frecuencia relativa de un dato más la frecuencia relativa del dato anterior**. Así que, la frecuencia relativa acumulada del primer dato coincide con su frecuencia relativa y la frecuencia relativa acumulada es igual a 1.

→ Frecuencia porcentual acumulada

Es el porcentaje de datos respecto al total que se han reportado hasta ese momento. Se puede calcular rápidamente multiplicando la frecuencia relativa acumulada por 100%.