Contenido

[1. Introducción 2](#_Toc174842891)

[1.1. Definición de estadística 2](#_Toc174842892)

[1.2. Ramas de la estadística y objetivos 2](#_Toc174842893)

[2. Población y muestra 2](#_Toc174842894)

[2.1. Población 2](#_Toc174842895)

[2.2. Muestra 2](#_Toc174842896)

[3. Clasificación de los parámetros o los estadísticos 3](#_Toc174842897)

[4. Distribuciones de frecuencias 3](#_Toc174842898)

[4.1. Frecuencia absoluta, ni 3](#_Toc174842899)

[4.2. Frecuencia absoluta acumulada, Ni 3](#_Toc174842900)

[4.3. Frecuencia relativa, fi 3](#_Toc174842901)

[4.4. Frecuencia relativa acumulada, Fi 3](#_Toc174842902)

[4.5. Porcentaje, pi 3](#_Toc174842903)

[4.6. Porcentaje acumulado, Pi 4](#_Toc174842904)

[4.7. Ejemplo de tabla con todas las frecuencias 4](#_Toc174842905)

[4.8. Marca de clase 4](#_Toc174842906)

[5. Representaciones gráficas 5](#_Toc174842907)

[5.1. Diagrama de barras 5](#_Toc174842908)

[5.2. Diagrama en escalera 5](#_Toc174842909)

[5.3. Histograma 6](#_Toc174842910)

[5.4. Polígono de frecuencias 7](#_Toc174842911)

# Introducción

## Definición de estadística

La estadística es la parte de las matemáticas que se encarga de estudiar sucesos no deterministas. Este carácter no determinista implica que no se pueden deducir el resultado de un experimento previamente a su realización.

Un ejemplo de una ciencia determinista es la física. Si un vehículo se desplaza a 50 km/h durante 5 minutos, puede determinarse con exactitud que distancia recorrerá. Sin embargo, en las ciencias sociales los resultados pueden variar incluso aunque se den las mismas condiciones iniciales.

A su vez la estadística trata con grupos grandes lo que genera la necesidad de organizar una cantidad de información más o menos grande.

## Ramas de la estadística y objetivos

La estadística suele dividirse en 2 ramas: estadística descriptiva y estadística inferencial

La estadística descriptiva es la parte de la estadística que analiza las características de un grupo que se esté estudiando, pero no intenta predecir esas características. Ejemplo, ordenar todas las notas de una clase y calcular la nota media de la clase.

La estadística inferencial sí intenta predecir algunas características de un grupo de personas a partir del resultado de una muestra. Por ejemplo, si durante el trascurso de los años, una asignatura A tiene más nota media que una asignatura B, se puede considerar que las y los estudiantes consideran la asignatura A es más fácil y es esperable que en años posteriores el número de aprobados de la asignatura A sea mayor que el número de aprobados de la asignatura B.

# Población y muestra

## Población

Población es el conjunto de individuos sobre los cuales se desea analizar una característica. Ejemplos:

* Ejemplo 1. Se desea conocer el número de ciudadanos de un concejo que aprueba la gestión de su alcalde.
* Ejemplo 2. En una fábrica de refrescos se desea realizar un análisis de calidad y conocer si las latas finales están en buenas condiciones.

La característica de la población que se desea estudiar se le llama parámetro. Si lo que está estudiando, esto es, con lo que se está trabajando es con el conjunto de la población, al estudio se le llama censo.

## Muestra

En algunas ocasiones no resulta práctico realizar el estudio a toda la población por lo que se coge una parte de ésta de manera que se considera que el resultado que se obtenga en la muestra representa de manera fiel al conjunto de la población. Determinar cuan fiel es esa consideración forma parte de la estadística inferencial.

A la característica de la muestra que se desea estudiar se le llama estadístico. Si se está trabajando con muestras, al estudio se le llama muestreo.

# Clasificación de los parámetros o los estadísticos

Los parámetros o los estadísticos se clasifican en atributos o variables.

* Atributos: son aquellos parámetros o estadísticos que representan características cualitativas, es decir, que no se pueden representar por números. En ejemplo de estos son por ejemplo el color de ojos de una población.
* Variables: son aquellos parámetros o estadísticos que representan una característica cuantitativa, es decir, se puede representar por números. Por ejemplo, el peso o la edad de una población. Las variables se suelen representar por letras mayúsculas: X, Y…. Estos a su vez se dividen en variables continuas o variables discretas. Las variables continuas son aquellas que pueden tomar infinitos valores, por ejemplo el peso; sin embargo las continuas solo pueden tomar unos valores concretos, por ejemplo con la edad solo pueden tomar los valores 0, 1, 2, 3…

# Distribuciones de frecuencias

Para cada valor de las variables se anotan lo siguientes valores:

## Frecuencia absoluta, ni

Representa el número de veces que se repite cada valor

## Frecuencia absoluta acumulada, Ni

La frecuencia absoluta acumulada de un valor representa el número de valores que son iguales o más pequeños que el valor indicado. Es decir, es como la frecuencia absoluta (ni), pero sumando todas las frecuencias absolutas de todos los valores más pequeños al dado. Por eso se llama acumulada, porque según va aumentando el valor va sumando o acumulando las frecuencias.

## Frecuencia relativa, fi

Representa el número de veces que se repite cada valor pero en tanto por uno respecto al tamaño muestral N. Es decir se calcula con la siguiente fórmula:

## Frecuencia relativa acumulada, Fi

La frecuencia relativa acumulada de un valor representa el número de valores que son iguales o más pequeños que el valor indicado expresado en tanto por uno respecto al tamaño muestran N. Es decir, es como la frecuencia relativa (fi), pero sumando todas las frecuencias relativas de todos los valores más pequeños pequeños al dado. Por eso se llama acumulada, porque según va aumentando el valor va sumando o acumulando las frecuencias.

## Porcentaje, pi

Es lo mismo que la frecuencia relativa (fi) de apartado 4.3 pero expresado en %. Es decir:

## Porcentaje acumulado, Pi

Es lo mismo que la frecuenta relativa acumulada (Fi) pero expresado en porcentaje. Es decir:

## Ejemplo de tabla con todas las frecuencias

Se tiene un bombo con 20 bolas que tienen los siguientes números:

* 4 bolas con el número 1.
* 1 bolas con el número 2.
* 2 bolas con el número 3.
* 1 bolas con el número 4.
* 3 bolas con el número 5.
* 1 bolas con el número 6.
* 5 bolas con el número 7.
* 1 bolas con el número 8.
* 2 bolas con el número 9.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Xi** | **ni** | **Ni** | **fi** | **Fi** | **pi** | **Pi** |
| 1 | 4 | 4 | 0,2 | 0,2 | 20 % | 20 % |
| 2 | 1 | 5 | 0,05 | 0,25 | 5 % | 25 % |
| 3 | 2 | 7 | 0,1 | 0,35 | 10 % | 35 % |
| 4 | 1 | 8 | 0,05 | 0,4 | 5 % | 40 % |
| 5 | 3 | 11 | 0,15 | 0,55 | 15 % | 55 % |
| 6 | 1 | 12 | 0,05 | 0,6 | 5 % | 60 % |
| 7 | 5 | 17 | 0,25 | 0,85 | 25 % | 85 % |
| 8 | 1 | 18 | 0,05 | 0,9 | 5 % | 90 % |
| 9 | 2 | 20 (1) | 0,1 | 1 (2) | 10 % | 100 % (3) |

1. Este valor debe coincidir con el número total de bolas, N=20.
2. Este valor debe coincidir siempre con 1.
3. Este valor debe coincidir siempre con 100 %.

## Marca de clase

En algunas situaciones se pueden agrupar varios valores Xi en intervalos. En esos casos se define la marca de clase como el valor medio del intervalo y se toma a este valor como representativo del intervalo pasando esta marca de clase a usar Xi y los 2 extremos del intervalo se denotarán como Li-1 y Li.

Para ejemplificar esto usaremos la como ejemplo la estura de 400 personas y se tabulan estas estaturas usando intervalos que van de 10 cm en 10 cm salvo en el primer intervalo que tiene una amplitud de 20 cm.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Li-1 - Li** | **Xi** | **ni** | **Ni** | **fi** | **Fi** | **pi** | **Pi** |
| 130 – 150 | 140 | 38 | 38 | 0 ,095 | 0,095 | 9,5 % | 9,5 % |
| 150 – 160 | 155 | 80 | 118 | 0,2 | 0,295 | 20 % | 29,5 % |
| 160 – 170 | 165 | 186 | 304 | 0,465 | 0,76 | 46,5 % | 76 % |
| 170 – 180 | 175 | 60 | 364 | 0,15 | 0,91 | 15 % | 91 % |
| 180 – 190 | 185 | 24 | 388 | 0,06 | 0,97 | 6 % | 97 % |
| 190 - 200 | 195 | 12 | 400 | 0,03 | 1 | 3 % | 100 % |

# Representaciones gráficas

Los diagramas suelen representarse horizontalmente, es decir, los valores de la variable se anotan en el eje x y las frecuencias en el eje y. Si bien se puede hacer al revés, todos los ejemplos de este documento estarán hecho de la primera forma mencionada.

## Diagrama de barras

En el diagrama de barras se representa un rectángulo asociado a cada valor de la variable. La altura del rectángulo representa la frecuencia para cada valor de la variable. En este caso particular el área del rectángulo no es relevante, de hecho, puede representarse como una simple línea.

Para el ejemplo de las bolas del apartado 4.7 el diagrama de barras sería:

Hay que darse cuenta de que este diagrama de barras está realizado con frecuencias absolutas. También podría plantearse un diagrama de barras con frecuencias relativas o porcentajes. En cualquier caso, no se deben usar nunca los datos acumulador.

## Diagrama en escalera

Cuando se usen los valores acumulador (Ni, Fi o Pi) se llamará diagrama en escalera, porque cada valor será mayor o igual que el anterior.

En este caso se ha usado la frecuencia relativa acumulada para el ejemplo.

## Histograma

El histograma se usa para valores agrupados como los vistos en el apartado 4.8. En estos casos importa tanto la altura de la barra, como la base del rectángulo.

* La base del rectángulo representa la longitud del intervalo.
* La altura del intervalo se llama densidad de frecuencia o densidad de porcentaje y se representa con hi.
* El área del rectángulo representa la frecuencia o el porcentaje.

El histograma puede ser de frecuencias absolutas, frecuencias absolutas acumuladas, frecuencias relativas, frecuencias relativas acumuladas, porcentajes y porcentajes acumulados.

Para ejemplificar esto, se hará un histograma de frecuencias relativas de la tabla del apartado 4.8. Previamente se escribirá la tabla nuevamente, pero añadiendo una columna nueva para indicar las densidades de frecuencia hi.

En la fórmula anterior, se puede despejar hi:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Li-1 - Li** | **Xi** | **ni** | **Ni** | **fi** | **hi** | **Fi** | **pi** | **Pi** |
| 130 – 150 | 140 | 38 | 38 | 0 ,095 | 4,75·10-3 | 0,095 | 9,5 % | 9,5 % |
| 150 – 160 | 155 | 80 | 118 | 0,2 | 0,02 | 0,295 | 20 % | 29,5 % |
| 160 – 170 | 165 | 186 | 304 | 0,465 | 0,0465 | 0,76 | 46,5 % | 76 % |
| 170 – 180 | 175 | 60 | 364 | 0,15 | 0,015 | 0,91 | 15 % | 91 % |
| 180 – 190 | 185 | 24 | 388 | 0,06 | 6·10-3 | 0,97 | 6 % | 97 % |
| 190 - 200 | 195 | 12 | 400 | 0,03 | 3·10-3 | 1 | 3 % | 100 % |

Gráfico, Gráfico de líneas

Descripción generada automáticamente

La línea roja, es el polígono de frecuencias, que se verá en el apartado 5.4

## Polígono de frecuencias

Si en el histograma se unen los puntos medio de la base superior (en la marca de clase) de cada rectángulo entre si, se obtiene el polígono de frecuencias. En el ejemplo del apartado 5.3 el polígono de frecuencias es el que está en rojo. El polígono de frecuencias podría usarse para cualesquiera de las gráficas anteriores.

El polígono de frecuencias es muy útil para visualizar tendencias, pues se le da mayor importancia a las variaciones de valores frente al valor de cada punto concreto. Puede por tanto entenderse la importancia de este grafo en áreas como la economía.