

Resumen sobre estadística y probabilidad.

Leandro Molina

24 de enero de 2023



Índice

1. Capítulo 1.	3
1.1. ¿Por qué se debe estudiar estadística?	3
1.2. ¿Que se entiende por estadística?	3
1.3. Tipos de estadística.	4
1.4. Tipos de variables.	5
1.5. Niveles de medición.	5
2. Capítulo 2.	7
2.1. Descripción de datos.	7
2.2. Construcción de una tabla de frecuencias.	7
2.3. Construcción de distribuciones de frecuencias: datos cuantitativos.	10
2.4. Distribución de frecuencias relativas	12
2.5. Representacion grafica de una distribucion de frecuencias	12

1. Capitulo 1.

1.1. ¿Por qué se debe estudiar estadística?

Hay 3 motivos para el estudio de la estadística estos son:

1. La primera razon, consiste en que la informacion numerica prolifera por todas partes. si revisas diarios o revistas contienen mucha cantidad de informacion numerica.
2. Una segunda razon, es que las tecnicas de la estadistica se emplean para tomar decisiones que afectan la vida diaria, es decir, que incluyen en su bienestar.
3. Una tercera razon, el conocimiento de sus metodos facilita la compresion de la forma en que se toman las decisiones y proporciona un entendimiento mas claro de como le afectan.

Al encarar la necesidad de tomar decisiones en las que tenes que saber hacer un analisis de datos resultara de utilidad. Con el fin de tomar una decision informada, sera necesario llevar a cabo lo siguiente para poder tomar una decision informada:

1. Determinar si existe informacion adecuada o si requiere informacion adicional.
2. Reunir informacion adicional, si se necesita, de manera que no se obtengan resultados erroneos.
3. Resumir los datos de manera util e informativa.
4. Analizar la informacion disponible.
5. Obtener conclusiones y hacer inferencias al mismo tiempo que se evalua el riesgo de tomar una decision incorrecta.

En resumen hay por lo menos tres razones para estudiar estadistica: 1) los datos proliferan por todas partes; 2) las tecnicas estadisticas se emplean en la toma de decisiones que influyen en su vida; 3) sin que importe la carrera que elija, tomara decisiones profesionales que incluyan datos.

1.2. ¿Que se entiende por estadística?

Posee dos significados: su aceptacion mas comun, la estadistica se refiere a informacion numerica. Una coleccion de informacion numerica recibe el nombre de **estadisticas**. La informacion estadistica se presenta en forma grafica, es util porque capta la atencion del lector e incluye una gran cantidad de informacion.

Estadistica: Ciencia que recoge, organiza, presenta, analiza e interpreta datos con el fin de propiciar una toma de decisiones mas eficaz.

El primer paso en el estudio de un problema consiste en recoger datos revelantes. Estos deben organizarse de alguna forma y, tal vez, representarse en una grafica.

1.3. Tipos de estadística.

El estudio de la estadística se divide en dos categorías: la estadística descriptiva y la estadística inferencial.

Estadística descriptiva.

Es la ciencia que recoge, organiza, presenta, analiza...datos". Esta parte de la estadística recibe el nombre de **estadística descriptiva**.

Estadística descriptiva: Métodos para organizar, resumir y presentar datos de manera informativa.

Se trata de estadística descriptiva si calcula el crecimiento porcentual de una década a otra. Sin embargo, no sería de naturaleza descriptiva si utiliza estos para el calcular con esos datos algo futuro. Una masa de datos desorganizados resulta de poca utilidad. Las técnicas de la estadística descriptiva permiten organizar esta clase de datos y darles significado. Los datos se ordenan en una **distribución de frecuencia** (más adelante lo veremos). Se emplean diversas clases de **gráficas** para describir datos.

Estadística inferencial.

La estadística inferencial, también denominada **inferencia estadística**. El principal interés que despierta esta disciplina se relaciona con encontrar algo relacionado con una población a partir de una muestra de ella. Ya que estas son inferencias relacionadas con una población, basadas en datos de la muestra, se trata de estadística inferencial. Se podría considerar a la estadística inferencial como la mejor conjetura que es posible obtener del valor de una población sobre la base de la información de una muestra.

Estadística inferencial: Métodos que se emplean para determinar una propiedad de una **población** con base en la información de una **muestra** de ella.

Atención a las palabras población y muestra en la definición de estadística inferencial. Una **población** puede constar de individuos, también puede consistir en objetos. Desde una perspectiva estadística, una población no siempre que tiene que ver con personas.

Población: Conjunto de individuos u objetos de interés o medidas que se obtienen a partir de todos los medios u objetos de interés.

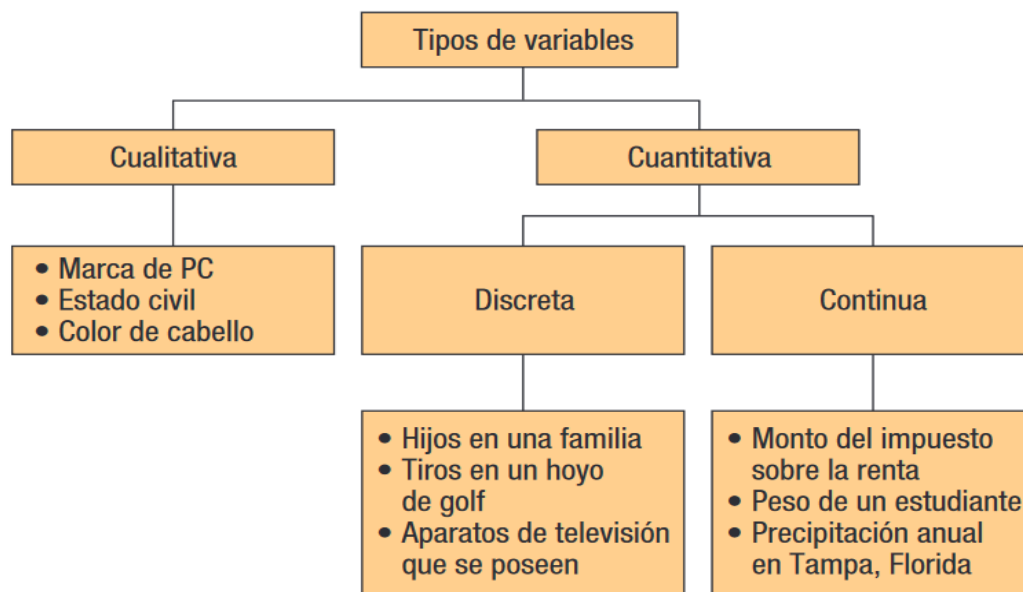
Con el objeto de inferir algo sobre una población, lo común es que se tome una muestra de ella.

Muestra: Porción o parte de la población de interés.

La toma de muestras para aprender algo sobre una población es de uso frecuente en administración, agricultura, política y acciones de gobierno.

1.4. Tipos de variables.

Dos tipos basicos de variables: 1)Cualitativas y 2)Cuantitativas, la caracteristica que se estudia es de naturaleza no numerica, recibe el nombre de **variable cualitativa** o **atributo**. Cuando los datos son de naturaleza cualitativa, importa la cantidad o proporcion que caen dentro de cada categoria. Los datos cualitativos se resumen en tablas o graficas de barras. Cuando la variable que se estudia aparece en forma numerica, se le denomina **variable cuantitativa**. Las variables cuantitativas pueden ser discretas o continuas. Las **variables discretas** adoptan solo ciertos valores y existen vacios entre ellos. Las variables discretas son el resultados de una relacion numerica, las observaciones de una **variable continua** toman cualquier valor dentro de un intervalo especifico. Por lo general las variables continuas son el resultado de mediciones. Resumen de los tipos de variables:



1.5. Niveles de medición.

Los datos se clasifican por niveles de medicion. El nivel de medicion de los datos rige los calculos que se llevan a cabo con el fin de resumir y presentar los datos. Tambien determina las pruebas estadisticas que se deben realizar. De hecho, existen cuatro niveles de medicion: nominal, ordinal, de intervalo y de razon. La medicion mas baja, o mas primaria, corresponde al nivel nominal. La mas alta, o el nivel que proporciona la mayor informacion relacionada con la observacion, es la medicion de razon.

Datos de nivel nominal.

Las observaciones acerca de una variable cualitativa solo se clasifican y se cuentan. No existe una forma particular para ordenar las etiquetas, no existe un orden natural. Para el nivel nominal, la medicion consiste en contar, a veces, para una mejor compresión de lectura, estos conteos se convierten en porcentajes. Es necesario hacer que el porcentaje sume un total de 100 %, no existe un orden natural para los resultados. Para procesar datos a menudo se codifica la informacion en forma numerica. El nivel nominal tiene las siguientes propiedades:

1. La variable de interes se divide en categorias o resultados.
2. No existe un orden natural de los resultados.

Datos de nivel ordinal.

El nivel inmediato superior de datos es el **nivel ordinal**. No es posible distinguir la magnitud de las diferencias entre los grupos, ¿la diferencia entre superior y bueno es la misma que entre lo malo e inferior? No es posible afirmarlo. Las propiedades del nivel ordinal de los datos son las siguientes:

1. Las clasificaciones de los datos se encuentran representadas por conjuntos de etiquetas o nombre (alto, medio, bajo), las cuales tienen valores relativos.
2. En consecuencia, los valores relativos de los datos se pueden clasificar u ordenar.

Datos de nivel de intervalo.

El **nivel de intervalo** de medicion es el nivel inmediato superior. Incluye todas las características de nivel ordinal, pero, además, la diferencia entre valores constituye una magnitud constante. Si las distancias entre los números tienen sentido, aunque las razones no, entonces tiene una escala de intervalo de medicion. Las propiedades de los datos de nivel intervalo son las siguientes:

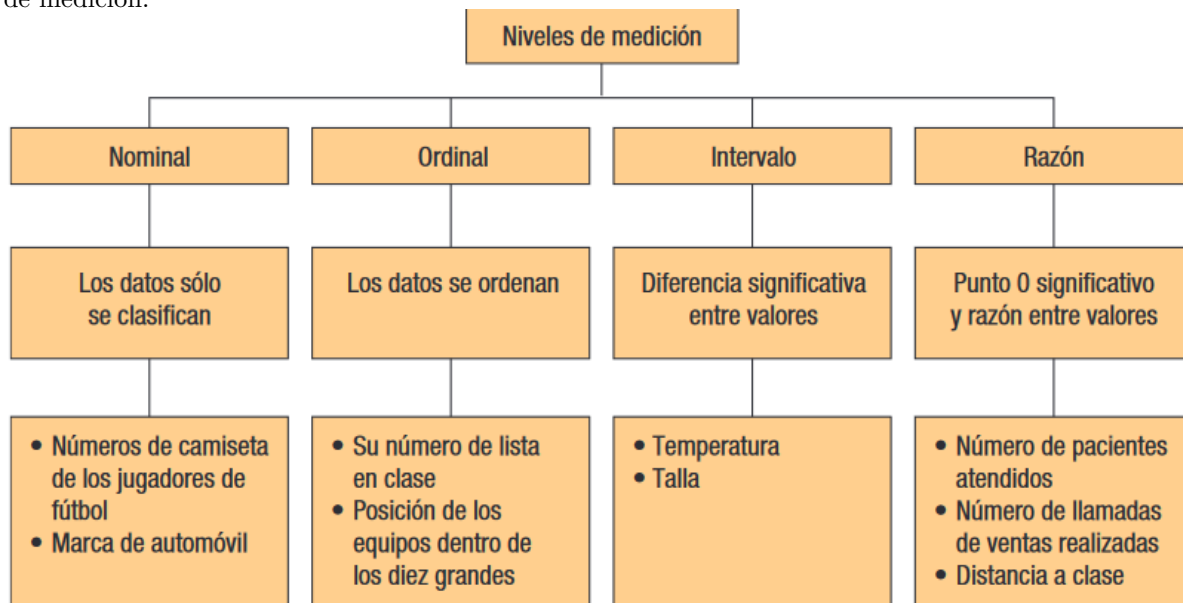
1. Las clasificaciones de datos se ordenan de acuerdo con el grado que posea de la característica en cuestion.
2. Diferencias iguales en la característica representan diferencias iguales en las mediciones.

Datos de nivel de razón.

Todos los datos cuantitativos son registrados en el nivel de razon de la medicion, el **nivel de razon** es el *mas alto*. Posee todas las características del nivel de intervalo, aunque, además, el punto 0 tiene sentido y la razon entre dos números significativa, si se encuentra en 0 significa la ausencia de algo (peso, dinero, etc). Las propiedades de los datos de nivel intervalo son las siguientes:

1. Las clasificaciones de datos se ordenan de acuerdo con la cantidad de características que poseen.
2. Diferencias iguales en la característica representan diferencias iguales en los números asignados a las clasificaciones.
3. El punto cero representa la ausencia de características y la razón entre dos números es significativa.

La siguiente grafica resume las principales características de los diversos niveles de medicion.



2. Capítulo 2.

2.1. Descripción de datos.

Tabla de frecuencias, distribuciones de frecuencias y su representación grafica.

2.2. Construcción de una tabla de frecuencias.

La estadística descriptiva se encarga de organizar datos con el fin de mostrar la distribución general de estos y el lugar en donde tienden concentrarse, además de señalar valores de datos pocos usuales o extremos. El primer procedimiento que se emplea para organizar y resumir un conjunto de datos es una **tabla de frecuencias**.

Tabla de frecuencias: Agrupacion de datos cualitativos en clases mutuamente excluyentes que muestra el numero de observaciones en cada clase.

Recordar que, una variable cualitativa es de naturaleza no numerica; es decir, que la informacion es clasificable en distintas categorias. No hay un orden particular en estas categorias. Por otro lado, estan las variables cuantitativas son de indole numerica.

Frecuencias relativas de clase.

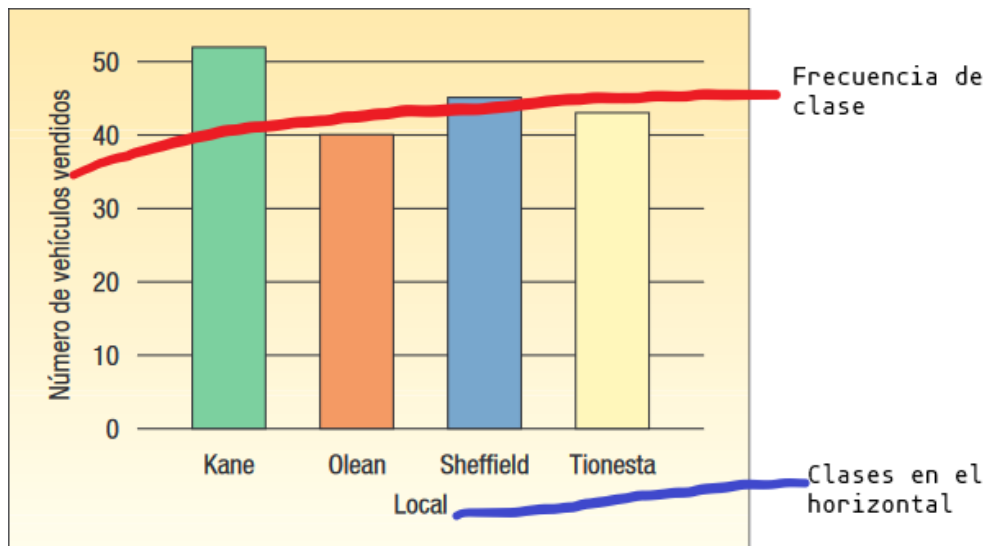
Es posible convertir las frecuencias de clase en frecuencias relativas de clase para mostrar la fraccion del numero total de observaciones en cada una de ellas. Una frecuencia relativa capta la relacion entre la totalidad de elementos de una clase y el numero total de observaciones. Para convertir una distribucion de frecuencias en una distribucion de frecuencias relativa, cada una de las frecuencias de clase se divide entre el total de observaciones.

Representación grafica de datos cualitativos.

El instrumento mas comun para representar una variable cualitativa en forma grafica es la **grafica de barras**. En la mayoria de los casos, el eje horizontal muestra la variable de interes y el eje vertical la frecuencia o fraccion de cada uno de los posibles resultados. Una caracteristica distinta de esta herramienta es que existe una distancia o espacio entre las barras. Una grafica de barras es una representacion grafica de una tabla de frecuencias mediante una serie de rectangulos de anchura uniforme, cuya altura corresponde a la frecuencia de clase.

Grafica de barras: En ella, las clases se representan en el eje horizontal y la frecuencia de clase en el eje vertical. Las frecuencias de clase son proporcionales a las alturas de las barras.

Se muestra ejemplo de una grafica de barras en la siguiente pagina:



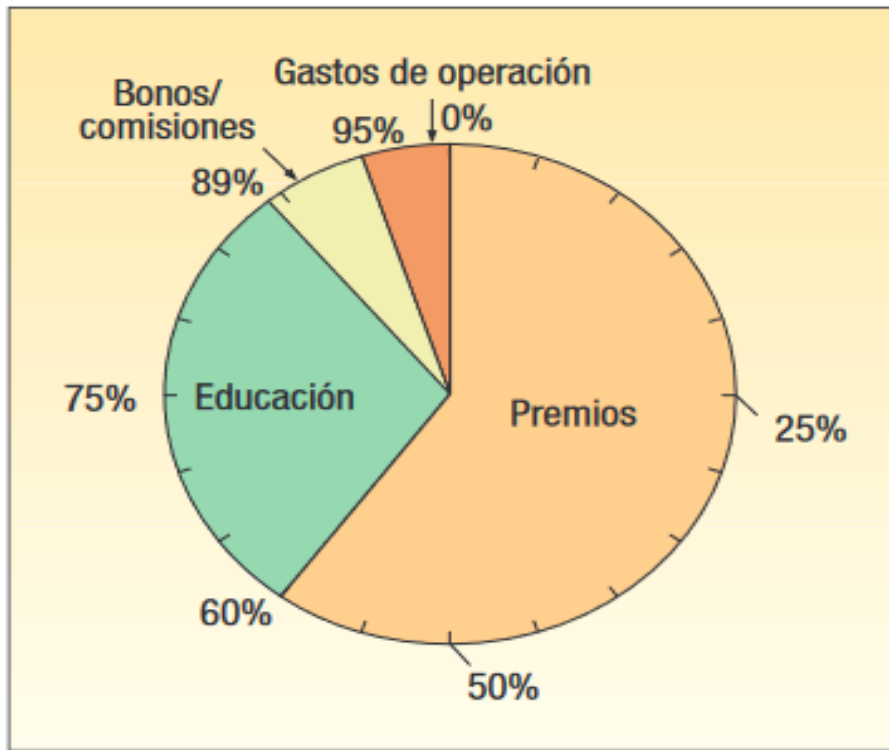
Otro tipo de grafica util para describir informacion cualitativa es la **grafica de pastel**.

Grafica de pastel: Grafica que muestra la parte o porcentaje que representa cada clase del total de numeros de frecuencia.

El primer paso para elaborar una grafica de pastel consiste en registrar los porcentajes 0, 5, 10, 15, etc, de manera uniforme alrededor de la circunferencia de un circulo. El area rebanada representa alguna clase, cada rebanada de pastel representa la porcion relativa de cada componente, es posible compararlas con facilidad.

Aca un ejemplo de grafica bizcochuelo.

Uso del dinero de las ventas	Cantidad (millones de dólares)	Porcentaje de ventas
Premios	1 460.0	60
Educación	702.3	29
Bonos	150.0	6
Gastos	124.3	5
Total	2 436.6	100



Las graficas de pastel y las de barras cumplen casi la misma funcion. ¿Cuales son los criterios para elegir una u otra? En la mayoria de los casos, las graficas de pastel son las mas informativas cuando se trata de comparar la diferencia relativa en el porcentaje de observaciones de cada uno de las variables de la escala nominal. Es preferible usar una grafica de barras cuando el objetivo es comparar el numero de observaciones en cada categoria.

2.3. Construcción de distribuciones de frecuencias: datos cuantitativos.

Distribucion de frecuencias puede ser util para describir la ganancias de ventas.

Distribucion de frecuencias: Agrupacion de datos en clases mutuamente excluyentes, que muestra el numero de observaciones que hay en cada clase.

¿Como crear una distribucion de frecuencias? El primer paso consiste en acomodar los datos en una tabla que muestre las clases y el numero de observaciones que hay en cada clase. Recorda que el **objetivo** es construir tablas, diagramas y graficas que revelen rapidamente la concentracion, los valores extremos y la distribucion de los datos. La informacion desorganizada como **datos en bruto** o **datos no agrupados**. Los datos en bruto se interpretan con mayor facilidad si se organizan como una distribucion de frecuencias.

Paso 1: Defina el numero de clases. El objetivo consiste en emplear suficientes agrupamientos o **clases**, de manera tal que se perciba la forma de la distribucion. Aca se necesita criterio. Una gran cantidad de clases o muy pocas podrian no permitir ver la conformacion fundamental del conjunto datos. Una receta util para determinar la cantidad de clases (k) es la regla de 2 a la k . Esta guia sugiere que se elija el menor numero (k) para el numero clases, de tal manera que 2^k sea mayor que el numero de observaciones (n). Ejemplo $n = 180$ obsecciones, si supone $k = 7$, lo cual significa que utilizara siete clases, entonces $2^7 = 128$, algo menos que las 180 observaciones. De ahi que 7 no represente suficientes clases. Si $k = 8$, entonces $2^8 = 256$, que es mayor 180. Por lo tanto, el numero de clases se recomienda es de 8.

Paso 2: Determine el intervalo o ancho de clase. El **intervalo** o **ancho de clase** deberia ser el mismo para todas las clases. Todas las clases juntas deben cubrir por lo menos la distancia del valor mas bajo al mas alto de los datos. Expresado esto en una formula seria:

$$i \geq \frac{H - L}{k}$$

Por lo general el tamaño de intervalo se redondea a una cifra conveniente, tal como un multiplo de 10 a 100. En las distribuciones de frecuencia son preferibles los intervalos de clase iguales. Sin embargo, en ciertos casos se necesita que no lo sean para evitar una gran cantidad de clases vacias, o casi vacias.

Paso 3: Establezca los limites de cada clase. Este paso es importante para que sea posible incluir cada observacion en una sola categoria. Esto significa que debe evitar la superposicion de limites de clase confusos. Por ejemplo, clases como \$1300 - \$1400 y \$1400 - \$1500 no deberian emplearse porque no resulta claro si el valor de \$1400 pertenece a la primera o a la segunda clases. Las clases como \$1300 - \$1400 y \$1500 - \$1600 se emplean con frecuencia, aunque tambien pueden resultar confusas si no se conviene en redondear todos los datos de \$1450 o por arriba de esta cantidad a la segunda clase y los datos por debajo de \$1400 a la primera clase. Al redondear el intervalo de clase hacia arriba con el fin de obtener un tamaño conveniente de clase, se cubre un rango mas amplio que el necesario. Una directriz consiste en convertir el limite inferior de la primera clase en un multiplo del intervalo de clase. A veces esto no es posible, pero el limite inferior por lo menos debe redondearse.

Paso 4: Anote las veces que encuentre las observaciones en el intervalo en las clases. PONER IMAGEN PASO 4

Paso 5: Cuente el numero de elementos de cada clase. El numero de elementos que hay en cada clase recibe el nombre de **frecuencia de clase**. Por ejemplo en numero de \$200 a \$600 hay 8 observaciones, y en la clase de \$600 a \$1000 hay 11 observaciones. Por lo tanto la frecuencia de clase de la primera es

de 8, mientras que en la segunda es de 11. Hay un total de 180 observaciones o frecuencias en todo el conjunto de datos. así que la suma de todas las frecuencias debe ser igual a 180. **PONER TABLA2-7**

Con frecuencia aparecerán otros dos términos: **punto medio de clase** e **intervalo de clase**. El punto medio, que se encuentra entre los límites inferiores de dos clases consecutivas, se calcula sumando los límites inferiores de clases consecutivas y dividiendo el resultado entre dos. En caso de la imagen del paso 5, el límite de clase inferior de la primera clase es de \$200 y el siguiente límite es de \$600. El punto medio de clase \$400, que se calcula mediante la operación

$$\frac{\$600 + \$200}{2}.$$

El punto medio de \$400 representa mejor. Para determinar el intervalo de clase, se resta el límite inferior de la clase del límite inferior de la siguiente clase, es decir

$$\$600 - \$200 = 400.$$

También se puede determinar el intervalo de clase calculando la diferencia entre puntos medios consecutivos. El punto medio de la primera clase es de \$400 y el punto medio de la segunda clase es de \$800. La diferencia es de \$400.

2.4. Distribución de frecuencias relativas