

La distribución de Frecuencia

INTRODUCCIÓN

Desde el inicio de la civilización han existido formas sencillas de estadística, puesto que en la antigüedad se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas que eran de importancia en aquellas civilizaciones. El término estadístico es ampliamente percibido y pronunciado a diario desde diversos sectores activos de la sociedad. No obstante, hay una gran diferencia entre el sentido del término cuando se utiliza en el lenguaje corriente, generalmente al anteceder una citación de carácter numérico, y lo que la estadística significa como ciencia.

La razón o razones que motivaron al hombre en un momento de su desarrollo a tomar en cuenta datos con propósitos estadísticos, posiblemente se encuentra si se toma en cuenta que es difícil suponer un organismo social, sea cual fuere la época, sin la necesidad, casi instintiva, de recoger aquellos hechos que aparecen como actos esenciales de la vida; y así, al ubicarnos en una etapa del desarrollo de la estadística podemos especular que se convirtió en una aritmética estatal para asistir al gobernante que necesitaba conocer la riqueza y el número de los súbditos entre otros, con el objeto de recaudar impuestos o presupuestar la guerra.

Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o cosas. Hacia el año 3000 a.C. los babilonios usaban pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y sobre las especies vendidas o cambiadas mediante trueque.

El uso de la Estadística es de gran importancia en la investigación científica y de finanzas. Casi todas las investigaciones aplicadas requieren algún tipo de análisis estadístico para que sea posible evaluar sus resultados.

En algunos casos, para resolver un problema de carácter empírico, es preciso llevar a cabo un análisis bastante complejo; otras veces, basta con efectuar un análisis muy simple y directo.

La elección de uno u otro tipo de análisis estadístico depende del problema que se plantee en el estudio así como de la naturaleza de los datos. Desde este punto de vista, la Estadística constituye un *instrumento de investigación* y no un *producto final* de esta última.

El presente trabajo, que las estadísticas deben ser dominadas y practicadas para obtener satisfactorios resultados, sobre todo siguiendo las reglas para evitar errores mayores del 5% o el 10%, en otras palabras, el verdadero diagnóstico de la realidad han de ser prácticas permanentes en el accionar del investigador y el estadístico aplicado.

La estadística es una ciencia o método científico que en la actualidad es considerada como un poderoso auxiliar en las investigaciones científicas, que le permite a ésta aprovechar el material

cuantitativo.

DISTRIBUCIÓN DE FRECUENCIA

A menudo en una investigación ya sea de finanzas o científica se recogen grandes cantidades de datos numéricos. Cuando esto ocurre es difícil visualizar un orden o estructura que ayude a analizarlos. Para lograrlo es necesario condensar los datos en grupos de acuerdo a ciertas divisiones de la recta numérica (intervalos o clases). Aunque con esta agrupación la información inicial sobre cada dato individual se pierde, es más fácil visualizar rápidamente las características principales del grupo total de datos.

La frecuencia de un intervalo es el número de datos que corresponden a ese intervalo.

Una distribución de frecuencia es una tabla en la que aparecen todos los intervalos y las frecuencias de datos correspondientes a cada intervalo. Esta agrupación de datos numéricos por intervalos o clases se llama una distribución de frecuencia porque en ella se indica cuan frecuentemente aparecen datos en cada intervalo.

-

2. Aspectos importantes que se deben tener en cuenta cuando se crea una distribución de frecuencia

2.1. Número de intervalos

El número de intervalos en una distribución de frecuencia depende del número total de observaciones.

Se recomienda que en una distribución de frecuencia no haya más de 15 ni menos de 5 intervalos. Si hay muy pocos no se pueden observar las características importantes de la distribución y si hay muchos no se obtiene un resumen adecuado de la distribución.

2.2. Límites de los intervalos

El límite inferior de un intervalo corresponde al valor mínimo que puede incluirse en el intervalo. El límite superior de un intervalo corresponde al valor máximo que puede incluirse en el intervalo.

Ejemplo:

Puntuaciones	frecuencia
200-299	2
300-399	8
400-499	6

En el ejemplo anterior 200 es el límite inferior y 299 el límite superior del primer intervalo.

Algunos autores agrupan los datos de variables continuas en clases o intervalos que incluyen todos los valores desde un número dado hasta (pero excluyendo) otro número.

Ejemplo:

Peso	frecuencia
100-bajo 120	5
120-bajo 130	3
130-bajo 140	10

2.3. Amplitud de los intervalos

La amplitud de un intervalo se obtiene al restar el límite inferior de una clase al límite inferior de la clase superior.

Todos los intervalos de una distribución de frecuencia deben tener la misma amplitud, aunque a veces se encuentran tablas en las que el último intervalo es diferente.

Para determinar la amplitud de los intervalos de una distribución se divide la amplitud o alcance de la distribución entre el número de intervalos deseados y se redondea.

El conjunto de intervalos debe incluir todos los datos.

No debe haber traslapo (overlapping) de intervalos.

2.4. Distribución de frecuencia absoluta

En la tabla de frecuencia absoluta se señala, para cada intervalo o clase, la cantidad de datos cuyos valores pertenecen al intervalo.

2.5. Distribución de frecuencia relativa

La frecuencia relativa es la razón que se obtiene al dividir la frecuencia absoluta de un intervalo entre el número total de datos en la distribución.

La frecuencia relativa se puede expresar como una proporción o como un porcentaje.

La distribución de frecuencia relativa es esencial para comparar datos de dos distribuciones diferentes.

Si la frecuencia relativa del intervalo se multiplica por 100 se obtiene el porcentaje correspondiente a dicho intervalo.

Por lo general, en las publicaciones no especializadas, se utiliza más la frecuencia relativa expresada en porcentajes que en proporciones.

2.6. Distribución de frecuencia absoluta acumulada

La frecuencia absoluta acumulada de una clase se obtiene al sumar las frecuencias de todas las clases que representan valores menores que los del límite superior de la clase dada.

En la frecuencia absoluta acumulada de un intervalo se incluyen los datos del intervalo más los datos de todos los intervalos inferiores.

Ejemplo:

Puntuaciones	frecuencia	frec.acum.
200-299	2	2
300-399	8	10
400-499	6	16

2.7. Distribución de frecuencia relativa acumulada

La frecuencia relativa acumulada de una clase se obtiene al dividir la frecuencia acumulada del intervalo entre el total de datos de la distribución. Se expresa como una proporción o un porcentaje.

Ejemplo:

Puntuaciones	frecuencia	frec.acum.	frec.relativa acum.
200-299	2	2	0.125
300-399	8	10	0.625
400-499	5	16	1

Nota: A menudo la frecuencia absoluta, la frecuencia relativa y la frecuencia acumulada aparecen en la misma tabla.

Distribución de Frecuencia:

En estadística existe una relación con cantidades, números agrupados o no, los cuales poseen entre sí características similares. Existen investigaciones relacionadas con los precios de los productos de la dieta diaria, la estatura y el peso de un grupo de individuos, los salarios de los empleados, los grados de temperatura del medio ambiente, las calificaciones de los estudiantes, etc., que pueden adquirir diferentes valores gracias a una unidad apropiada, que recibe el nombre de variable. La representación numérica de las variables se denomina dato estadístico.

La distribución de frecuencia es una disposición tabular de datos estadísticos, ordenados ascendente o descendientemente, con la frecuencia (f_i) de cada dato. Las distribuciones de frecuencias pueden ser para *datos no agrupados* y para *datos agrupados* o de intervalos de clase.

Distribución de frecuencia para datos no Agrupados:

Es aquella distribución que indica las frecuencias con que aparecen los datos estadísticos, desde el menor de ellos hasta el mayor de ese conjunto sin que se haya hecho ninguna

modificación al tamaño de las unidades originales. En estas distribuciones cada dato mantiene su propia identidad después que la distribución de frecuencia se ha elaborado. En estas distribuciones los valores de cada variable han sido solamente reagrupados, siguiendo un orden lógico con sus respectivas frecuencias.

Distribución de frecuencia de clase o de datos Agrupados:

Es aquella distribución en la que la disposición tabular de los datos estadísticos se encuentran ordenados en clases y con la frecuencia de cada clase; es decir, los datos originales de varios valores adyacentes del conjunto se combinan para formar un intervalo de clase. No existen normas establecidas para determinar cuándo es apropiado utilizar datos agrupados o datos no agrupados; sin embargo, se sugiere que cuando el número total de datos (N) es igual o superior 50 y además el rango o recorrido de la serie de datos es mayor de 20, entonces, se utilizará *la distribución de frecuencia para datos agrupados*, también se utilizará este tipo de distribución cuando se requiera elaborar gráficos lineales como el histograma, el polígono de frecuencia o la ojiva.

La razón fundamental para utilizar la distribución de frecuencia de clases es proporcionar mejor comunicación acerca del patrón establecido en los datos y facilitar la manipulación de los mismos. Los datos se agrupan en clases con el fin de sintetizar, resumir, condensar o hacer que la información obtenida de una investigación sea manejable con mayor facilidad.

Componentes de una distribución de frecuencia de clase

1.- Rango o Amplitud total (recorrido).- Es el límite dentro del cual están comprendidos todos los valores de la serie de datos, en otras palabras, es el número de diferentes valores que toma la variable en un estudio o investigación dada. Es la diferencia entre el valor máximo de una variable y el valor mínimo que ésta toma en una investigación cualquiera. El rango es el tamaño del intervalo en el cual se ubican todos los valores que pueden tomar los diferentes datos de la serie de valores, desde el menor de ellos hasta el valor mayor estando incluidos ambos extremos. El rango de una distribución de frecuencia se designa con la letra R .

2.- Clase o Intervalo de clase.- Son divisiones o categorías en las cuales se agrupan un conjunto de datos ordenados con características comunes. En otras palabras, son fraccionamientos del rango o recorrido de la serie de valores para reunir los datos que presentan valores comprendidos entre dos límites.

Para organizar los valores de la serie de datos hay que determinar un número de clases que sea conveniente. En otras palabras, que ese número de intervalos no origine un número pequeño de clases ni muy grande. Un número de clases pequeño puede ocultar la naturaleza natural de los valores y un número muy alto puede provocar demasiados detalles como para observar alguna información de gran utilidad en la investigación.

Tamaño de los Intervalos de Clase

Los intervalos de clase pueden ser de tres tipos, según el tamaño que estos presenten en una

distribución de frecuencia: a) *Clases de igual tamaño*, b) *clases desiguales de tamaño* y c) *clases abiertas*.

3.-Amplitud de Clase, Longitud o Ancho de una Clase

La amplitud o longitud de una clase es el número de valores o variables que concurren a una clase determinada. La amplitud de clase se designa con las letras *lc*. Existen diversos criterios para determinar la amplitud de clases, ante esa diversidad de criterios, se ha considerado que lo más importante es dar un ancho o longitud de clase a todos los intervalos de tal manera que respondan a la naturaleza de los datos y al objetivo que se persigue y esto se logra con la práctica.

4.-Punto medio o Marca de clase

El centro de la clase, es el valor de los datos que se ubica en la posición central de la clase y representa todos los demás valores de esa clase. Este valor se utiliza para el cálculo de la media aritmética.

5.-Frecuencia de clase

La frecuencia de clase se le denomina frecuencia absoluta y se le designa con las letras *fi*. Es el número total de valores de las variables que se encuentran presente en una clase determinada, de una distribución de frecuencia de clase.

6.- Frecuencia Relativa

La frecuencia relativa es aquella que resulta de dividir cada uno de los *fi* de las clases de una distribución de frecuencia de clase entre el número total de datos (*N*) de la serie de valores. Estas frecuencias se designan con las letras *fr*; si cada *fr* se multiplica por 100 se obtiene la frecuencia relativa porcentual (*fr %*).

7.-Frecuencias acumuladas

Las frecuencias acumuladas de una distribución de frecuencias son aquellas que se obtienen de las sumas sucesivas de las *fi* que integran cada una de las clases de una distribución de frecuencia de clase, esto se logra cuando la acumulación de las frecuencias se realiza tomando en cuenta la primera clase hasta alcanzar la última. Las frecuencias acumuladas se designan con las letras *fa*. Las frecuencias acumuladas pueden ser *menor que* (*fafa > que*).

8.- Frecuencia acumulada relativa

La frecuencia acumulada relativa es aquella que resulta de dividir cada una de las *fa* de las diferentes clases que integran una distribución de frecuencia de clase entre el número total de datos (*N*) de la serie de valores, estas frecuencias se designan con las letras *far*. Si las *far* se multiplican por 100 se obtienen las frecuencias acumuladas relativas porcentuales y las mismas

se designan así: *far* %.

Mediciones de tendencia central

Hay tres medidas matemáticas de la tendencia central, usadas frecuentemente. La primera es la conocida mediaaritmética, que es la suma aritmética de todos los valores de una distribución dividida por el numero de casos. En términos matemáticos es:

$$M = \frac{\text{SUMATORIA } X}{N}$$

N

M representa la media aritmética, la letra griega “suma de”, X los valores individuales y N el número de casos.

La segunda medida de tendencia central es la *moda*, es el puntaje más común, el puntaje obtenido por el mayor número de personas.

La tercera medida se llama *mediana* que es simplemente el puntaje intermedio de una distribución, o el número que representaría un punto entre las dos mitades.

Por consiguiente, el significado de hombre medio o de inteligencia media puede referirse a cualquiera de estas tres medias diferentes de tendencia central, cada una con distinto valor y las tres matemáticamente correctas. Cada media tiene ciertas ventajas e inconvenientes. La media aritmética está muy influida por los valores de los elementos extremos, en tanto que la moda y la mediana no. A veces los valores extremos son innaturales y no se puede dejar que ejerzan toda su influencia.

Para tener un promedio que refleje todas las puntuaciones, incluyendo las extremas, se usa la media aritmética; para que no esté influido por las puntuaciones extremas, se usa la moda o la mediana.

Si todo lo que se conoce acerca de un grupo es su promedio, las mejores conjeturas de sus características tendrían que expresarse basándose en el mismo; pero ello da lugar, a veces, a serios errores.

La mediana

La mediana (M_d) es una medida de posición que divide a la serie de valores en dos partes iguales, un cincuenta por ciento que es mayor o igual a esta y otro cincuenta por ciento que es menor o igual que ella. Es por lo tanto, un parámetro que esta en el medio del ordenamiento o arreglo de los datos organizados, entonces, la mediana divide la distribución en una forma tal que a cada lado de la misma queda un número igual de datos.

Para encontrar la mediana en una serie de datos no agrupados, lo primero que se hace es ordenar los datos en una forma creciente o decreciente y luego se ubica la posición que esta

ocupa en esa serie de datos; para ello hay que determinar si la serie de datos es par o impar, luego el número que se obtiene indica el lugar o posición que ocupa la mediana en la serie de valores, luego la mediana será el número que ocupe el lugar de la posición encontrada.

La moda

La moda es la medida de posición que indica la magnitud del valor que se presenta con más frecuencia en una serie de datos; es pues, el valor de la variable que más se repite en un conjunto de datos. De las medias de posición la moda es la que se determina con mayor facilidad, ya que se puede obtener por una simple observación de los datos en estudio, puesto que la moda es el dato que se observa con mayor frecuencia. La moda se designa con las letras *Mo*.

Desviación típica o estándar

Es la *medida de dispersión* más utilizada en las investigaciones por ser la más estable de todas, ya que para su cálculo se utilizan todos los desvíos con respecto a la media aritmética de las observaciones, y además, se toman en cuenta los signos de esos desvíos. Se le designa con la letra castellana **S** cuando se trabaja con una muestra y con la letra griega minúscula **s** (Sigma) cuando se trabaja con una población. Es importante destacar que cuando se hace referencia a la población el número de datos se expresa con **N** y cuando se refiere a la muestra el número de datos se expresa con **n**. La desviación típica se define como:

Interpretación de la desviación estándar

La desviación típica como medida absoluta de dispersión, es la que mejor nos proporciona la variación de los datos con respecto a la media aritmética, su valor se encuentra en relación directa con la dispersión de los datos, a mayor dispersión de ellos, mayor desviación típica, y a menor dispersión, menor desviación típica.

Varianza

Es otra de las variaciones absolutas y la misma se define como el cuadrado de la desviación típica; viene expresada con las mismas letras de la desviación típica pero elevadas al cuadrado, así S^2 y s^2 . Las fórmulas para calcular la varianza son las mismas utilizadas por la desviación típica, exceptuando las respectivas raíces, las cuales desaparecen al estar elevados el primer miembro al cuadrado.

Errores de medida

El error se debe a un instrumento de medida inexacto, a un método imperfecto de aplicar el instrumento, a nuestra manera inadecuada de leerlo o registrarlo o a cualquier otro factor.

En la ciencia, por depender en gran parte del raciocinio de las mediciones, se tiene mucho cuidado con los errores de medida y se ha aprendido mucho acerca de su naturaleza, origen y control. En los casos que se ha sido incapaz de eliminarlos, se han desarrollado técnicas que

permiten estimar el grado de error. Sabiendo la magnitud del error se puede enunciar el grado de confianza en las conclusiones basadas en las medidas. El estudio de los errores de medida es uno de los básicos de la estadística.

Fiabilidad

No existe un instrumento de medida absolutamente perfecto. Hasta el instrumento de medida más simple, la regla, no está libre de error. Algunos instrumentos de medida nos dan errores mayores que otros.

La fiabilidad de un aparato de medida(incluido su método de aplicación) puede definirse como *el grado en que medidas repetidas de la misma cantidad, con el mismo instrumento de medida, dan las mismas lecturas.*

La fiabilidad medida por correlación: el coeficiente de correlación nos da *un índice numérico que expresa el grado de fiabilidad de una prueba.* Cuando se usa con este fin, el coeficiente de correlación recibe el nombre de *coeficiente de fiabilidad.*

Veracidad de las formas comparables: La mayoría de las pruebas psicológicas constan de gran número de elementos, problemas y preguntas. La correlación de las dos formas comparables nos daría la fiabilidad de una y otra forma.

El método de las formas comparables evita el problema de la memoria y quizás el de fastidio, pero deja intacto el del tiempo. Las dos formas se aplican en tiempos diferentes, y durante el intervalo pueden suceder muchas cosas que dificultan la interpretación de la correlación entre las dos formas comparables.

Fiabilidad compartida: la base del método de fiabilidad bipartida es idéntica a la del de formas comparables. Este método suele llamarse del “ coeficiente de pares-impares” y cuenta con dos ventajas: primera, las dos subpruebas(pares y nones) se hacen a la vez, en las mismas condiciones de motivación, idénticas condiciones de examen y con el mismo grado de atención. Segunda, por haber dividido la prueba de pares-impares, hemos garantizado la comparabilidad de formas, no sólo en cuanto al contenido, sino también en cuanto al contenido, sino también en cuanto al modo de administración.

Estos y otros métodos pueden proporcionarnos una valiosa información sobre la utilidad de una prueba como instrumento de medida. Sin embargo, saber que una prueba es fiable no basta para permitirnos apreciar su valor como instrumento de medición Puede ser muy fiable y por el contrario, constituir un mal instrumento de medida, por carecer de validez.

Validez

Los términos de “fiabilidad” y “validez” se usan indistintamente en el lenguaje vulgar. No obstante, en la teoría de la medición, tienen un significado distinto. El estadístico preocupado por el problema de la fiabilidad e un instrumento con lo que mide. Cuando le interesa la cuestión de la validez, pregunta si el instrumento mide lo que él quiere medir. Un instrumento

puede hacer medidas acordes (puede tener fiabilidad), pero acaso no mide lo que se quiere medir (acaso tiene poca validez). Pero a la mayoría de los tests que tratan de medir fenómenos más complejos no se les adscribe la validez con tanta facilidad. En primer lugar, la validez, lo mismo que la fiabilidad, no es asunto de todo nada. Una prueba tiene grados de validez. El grado de validez de las preguntas de clase sólo estaría influido por la comprensión por parte del alumno de los principios psicológicos. En este caso diríamos que las preguntas tienen validez perfecta como medida de la comprensión de principios psicológicos; pero, más probablemente, la puntuación en las preguntas es la resultante de la comprensión psicológica, más la aptitud memorista. La prueba tiene alguna validez para la comprensión psicológica y alguna otra para la capacidad memorista, pero no es una prueba "pura" de ninguna de las dos. Como en la fiabilidad, necesitamos algún medio para expresar el *grado* de validez de un instrumento de medida de un instrumento de medida y, de nuevo como en aquella, el coeficiente de correlación nos facilita ese medio.

La validez medida por correlación: Es evidente que una prueba es válida en el grado en que sus medidas se correlacionan con lo que mide. Cuando se usa de este modo el coeficiente de correlación se llama *coeficiente de validez*.

El principio general para determinar la validez de una prueba es bastante simple, correlacionamos sus puntuaciones con su criterio. La dificultad consiste en que, frecuentemente, no podemos hallar un criterio con el que compararlas. Por ejemplo se quiere medir la validez de una prueba de inteligencia. Se puede obtener las puntuaciones del test con mucha facilidad, pero qué servirá de criterio de "inteligencia" ¿Las calificaciones escolares? ¿El dinero ganado en la vida real? ¿La originalidad y creatividad? ¿La primacía en cuestiones sociales? Personas diferentes sugerirían distintos criterios y algunos de ellos plantearían, por sí mismos, problemas e medida.

Se han hecho muchos intentos de resolver el problema del criterio. Entre las técnicas más corrientes está el llamado *método del "grupo conocido"*.

Grupos conocidos y validez: No hay puntuaciones-criterio de originalidad y creatividad fácilmente disponibles.

Una prueba puede tener gran fiabilidad y poca validez, en el sentido que no mida lo que intentábamos que midiese. En cambio, una prueba de mucha validez no puede tener poca fiabilidad. Las pruebas poco fiables no pueden compararse consecuentemente con un conjunto de puntuaciones-criterio, porque sus medidas son, en gran parte, erróneas y por consiguiente deben tener poca validez.

Fiabilidad y validez conciernen a errores de medida y de conceptualización; por eso surgen de la inadecuación del instrumento de medida, pero, completamente aparte de ello, aún hay otra fuente mayor de error en las investigaciones que utilizan medidas. Se puede cometer un *error de selección*.

Normalmente en la sección de resultados de una publicación científica debemos incluir algún tipo de gráfico que ilustre en forma clara nuestros principales hallazgos. Un gráfico no sólo

debe ser agradable a la vista, sino que también debe ser sencillo y autoexplicativo, y reflejar lo más fielmente posible los hechos. Es bien sabido que un gráfico dice más que muchas palabras y permite en forma clara establecer una relación o cuantificar el grado de diferencia entre dos grupos, o alguna otra relación de acuerdo a nuestros datos. Hoy en día, con la ayuda de algunos softwares computacionales tales como: Systat, Excel, Lotus, etc., es relativamente fácil realizar gráficos tales como de:

- torta ("pie"),
- barras,
- histogramas,
- cajas,
- tallo y hoja,

Siempre debe pensar si vale la pena colocar un gráfico o si será mejor colocar una frase que simplifique más las cosas. No será que a veces nos encontramos con gráficos para llenar espacio, o quizás para impresionar?

De acuerdo a los propósitos y a los datos será el tipo de gráfico que podamos realizar:

1. Para mostrar distribución de frecuencia de variables nominales, ordinales o discretas utilizamos gráficos de barras simples. Para graficar distribución de frecuencia de variables continuas o intervalares, histogramas o polígonos de frecuencia.
2. Para mostrar relación entre variables nominales, ordinales o discretas, usamos gráficos de barras pareadas o agrupadas. Con el fin de mostrar asociación entre dos variables continuas, gráficos de correlación.
3. Para graficar la relación de una parte de la información respecto al total podemos utilizar gráficos sectoriales o barras subdivididas, independientemente de las variables.
4. Para mostrar variación conjunta de dos variables continuas usamos plots o gráficos lineales.

Según las variables que tengamos nominales u ordinales, realizamos normalmente gráficos de barras separada y gráficos sectoriales, donde cada porción "de la torta" corresponde a una categoría. Para variables intervalares normalmente se utiliza el histograma y el gráfico de barras, que demuestran promedio y desviación estándar. En cuanto a asociaciones entre variables podemos utilizar los plots.

En resumen la representación gráfica de datos nos permite o sería deseable que tengan, según Gnanadesikan, 1980:

- Capacidad descriptiva
- Versatilidad
- Orientación de los datos
- Potencial para posibles comparaciones internas
- Ayuda en focalizar la atención
- Permiten autocrítica de posibles presunciones
- Adaptabilidad para grandes volúmenes de datos.

GRAFICOS SECTORIALES Y DE BARRAS

Este tipo de gráficos fundamentalmente se utilizan para variables nominales u ordinales, en las cuales tenemos categorías que difieren en la proporción o tamaño en que se han presentado en la muestra analizada. El gráfico de torta o "pie" debe tener identificado cada sector; es muy conveniente colocar, además, el número de observaciones y la frecuencia en porcentaje, adyacente a cada categoría.

En un gráfico de barras se despliega cada barra separada por categoría, donde el alto de cada barra puede representar:

1. Cuentas de cada categoría,
2. El promedio de los casos en cada categoría,
3. El porcentaje de casos en cada categoría, o
4. Alguna otra medida o estadístico ingresado por cada categoría.

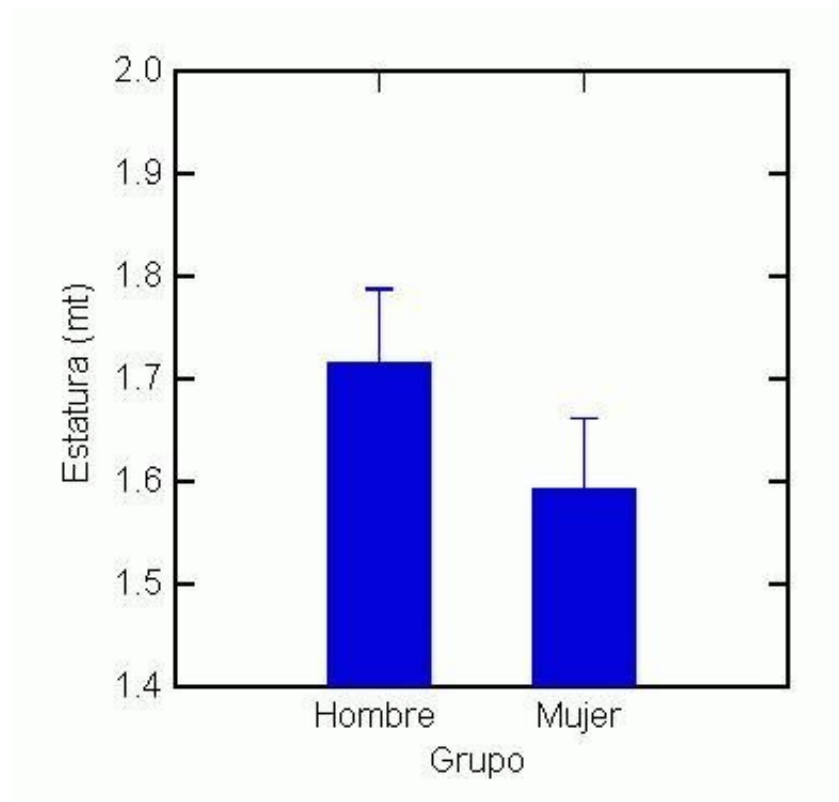


Fig. 1. Barras en que se presenta promedio de estatura de hombres y mujeres, y su desviación estándar (Dato de X. Haro).

HISTOGRAMA

Un histograma es una presentación de una distribución de frecuencia y se realiza por medio de

barras o rectángulos para cada intervalo de clase, de tal forma que cada barra es proporcional a la frecuencia de cada clase. La variable intervalar es la que normalmente debe tener un histograma, el cual va a servir para demostrar la normalidad en cuanto a la distribución de la muestra examinada. Al igual que en otros tipos de gráficos el eje X debe corresponder a la variable examinada, incluyendo el valor mínimo y máximo; en el eje Y se ubica la cantidad de individuos o muestras, o sea la frecuencia. Además como cualquier gráfico debe acompañarse de un título.

No debe llamar histograma a un gráfico de barras de variables nominales u ordinales. En muchas ocasiones vemos gráficos de barras que se les denomina histogramas; en estos casos se ha ubicado en el eje Y una variable nominal u ordinal (ver Rimm, pag. 46, fig. 2), y el largo de la columna representa la frecuencia observada de cada categoría. Estos son gráficos de barra y no histogramas.

Fig. 2. Histograma de la Edad. El símbolo indica el promedio de la edad, y los paréntesis () marcan una desviación estándar.

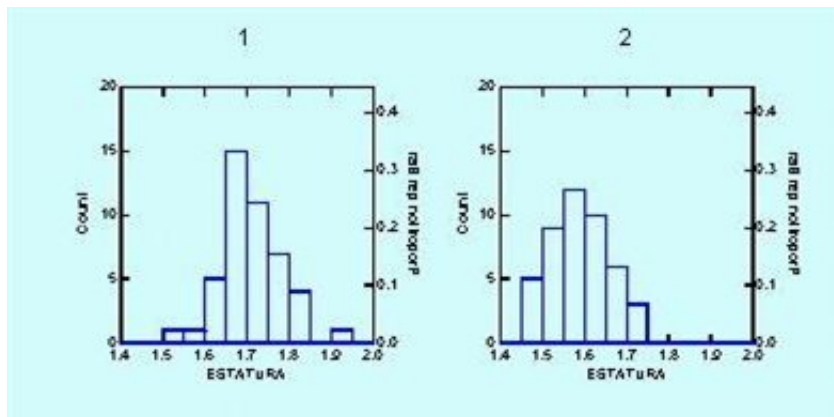


Fig. 3. Histograma de la Estatura, grupo 1= Hombres, 2= Mujeres. Datos de X. Haro, 45 hombres y 45 mujeres, chilenos. Observe la distribución para los hombres, entre 1.6 y 1.8 mt. está la mayoría, en cambio en las mujeres, entre 1.5 y 1.7 mt.

Polígono de frecuencia

En el polígono de frecuencia como en el histograma, el valor de la variable aparece en el eje horizontal y la frecuencia absoluta o relativa en el vertical.

La diferencia con respecto al histograma es que el polígono de frecuencia sólo toma en consideración el punto medio como representativo de cada clase o intervalo.

El punto medio de un intervalo se puede obtener de varias formas. Posiblemente la más fácil consiste en sumar los límites inferiores de dos intervalos consecutivos y dividir entre dos. Ejemplo:

Puntuaciones	frecuencia
200-299	2
300-399	8
400-499	6

En el ejemplo anterior 200 es el límite inferior del primer intervalo y 300 el límite inferior del segundo intervalo. Por lo tanto el punto medio del primer intervalo es $(200+300)/2 = 250$

De igual forma el punto medio del segundo intervalo es 350.

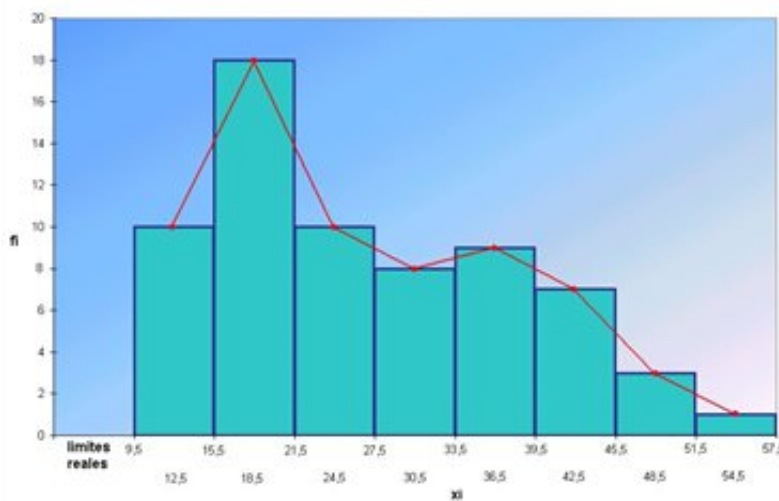
Nota:

Algunos autores definen el punto medio como el punto que se encuentra a mitad de camino entre los llamados límites exactos de cada intervalo. Por lo tanto calculan el punto medio sumando los límites exactos del intervalo y dividiendo entre 2. (Véase Hinkle capt.2)

Construcción

Para construir el polígono de frecuencia es necesario:

- a. Determinar los puntos medios de cada intervalo
- b. Determinar dos puntos medios adicionales correspondientes a un primero y último intervalo inexistentes (no siempre se añaden estos dos puntos).
- c. Graficar un punto sobre cada punto medio, a la altura de la frecuencia absoluta o relativa de cada intervalo.
- d. Conectar todos los puntos que aparecen a la altura de la frecuencia del intervalo.

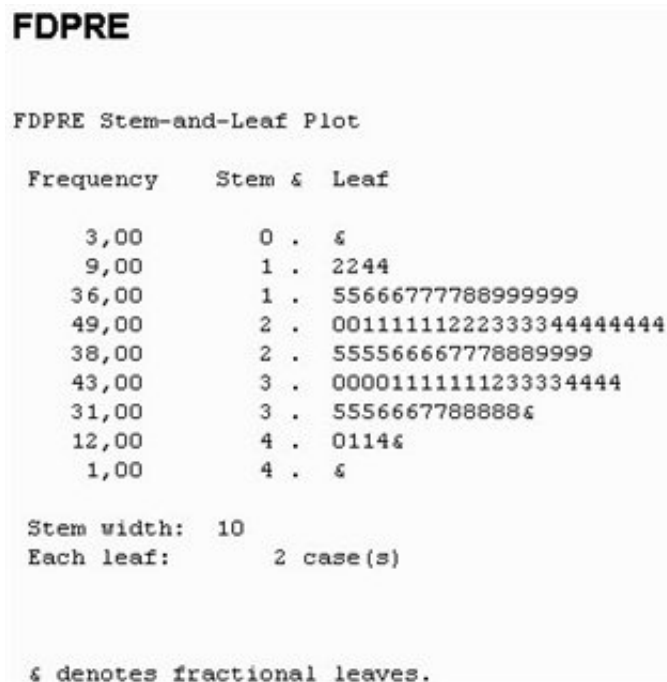


Nota: Cuando se construye un histograma o un polígono de frecuencia jamás se debe cortar el eje vertical de la frecuencia. Si se hace, la gráfica puede resultar engañosa. Sin embargo el eje

horizontal sí se puede cortar y hasta el cero se puede excluir, siempre que todas las clases o intervalos aparezcan.

Uno de los usos de los polígonos de frecuencia es poder comparar datos de dos distribuciones diferentes en un mismo eje.

Gráficos de tallo y hojas



Es una representación en la que los dígitos se dividen entre la parte troncal (izquierda denominada tallo) y la parte marginal (derecha denominada hoja). Los dígitos del tallo se ordenan de arriba abajo y las hojas de izquierda a derecha. Como en el histograma, la longitud de cada hoja corresponde al número de casos que contiene un determinado intervalo.

Gráficos de tallo y Hojas

Es una representación en la que los dígitos se dividen entre la parte troncal (izquierda denominada tallo) y la parte marginal (derecha denominada hoja). Los dígitos del tallo se ordenan de arriba abajo y las hojas de izquierda a derecha. Como en el histograma, la longitud de cada hoja corresponde al número de casos que contiene un determinado intervalo.

BIBLIOGRAFIA

<http://odontologia.uchile.cl/servicios/estadistica/graf.html>

http://www.cca.org.mx/dds/cursos/estadistica/html/m7/distribuciones_frecuencias_relativas.htm

<http://rrpac.upr.clu.edu:9090/~amenend/tabdistfrec.htm>

http://html.rincondelvago.com/metodologia-de-la-investigacion_1.html

<http://aldocgh.tripod.com/>

www.unav.es/.../practica_2/practica_2.html

Apuntes del Cuadernos Dictados en clase

Apuntes de año Anteriores

Libro de Estadística