



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
COLEGIO DE CIENCIAS Y HUMANIDADES

PLANTEL VALLEJO
ÁREA DE MATEMÁTICAS



ESTADÍSTICA Y PROBABILIDAD I

ESTADÍSTICA DESCRIPTIVA

Una vez recopilados los datos, es necesario en suministrar una presentación eficiente de los mismos, trabajo laborioso cuando la cantidad de datos es amplio.

La organización de los datos es necesaria, ya que es difícil obtener conclusiones o resultados de una gran cantidad de datos no organizados.

La organización se debe realizar de una forma tal que sea; sencilla, breve, que brinde una buena presentación y que sobre todo elimine lo irrelevante y resguarde la información de importancia de una manera precisa y sin ambigüedades.

Es decir los datos examinados deben clasificarse y comprimirse con el objeto de facilitar su análisis y el cálculo de medidas estadísticas.

Con la finalidad de lograr esto, los posibles valores de la variable a analizar deben de dividirse en clases o categorías que no se traslapen y que permitan clasificar a todos y cada uno de los datos recolectados, es decir que sean exhaustivas.

Por ejemplo sin nos interesara estudiar el "*Estado civil de un adulto residente del D. F*"

Podríamos dividir esta característica en las siguientes categorías: {soltero, casado, divorciado, viudo, unión libre}

Algunos otros ejemplos de variables y posibles categorías o intervalos son:

- *Color de ojos de un estudiante del C. C. H. Vallejo*, lo podemos dividir en las siguientes categorías: {negro, café, gris, verde, azul, otro}.
- *Relaciones entre compañeros de trabajo y colegas en un centro escolar*, podemos clasificarlas como {muy malas, malas, regulares, buenas, muy buenas}
- *Edad de los contrayentes en un matrimonio efectuado en el Estado de México*: se pueden establecer las siguientes categorías o intervalos de clasificación: {menores a 20 años, de 20 a 24, de 25 a 29, de 30 a 34, de 35 a 39, de 40 a 44, de 45 a 50 de 44 a 55 a 59, 60 o más}

Ejercicio: Proporciones posibles categorías o intervalos de clasificación para las siguientes variables:

- a) Lugar de residencia.
- b) Religión.
- c) Numero de mascotas en una familia.
- d) Salario de un empleado de la UNAM.
- e) Numero de materias adeudadas por un alumno del C. C. H.
- f) Mes de nacimiento.
- g) Complexión de una persona.

Una vez que se han establecido las categorías en las cuales podemos clasificar a los datos observados en un estudio, el siguiente paso es determinar la importancia o relevancia que tiene cada una de ellas. Es decir que tan frecuente es observada cada una de ellas.

Para nuestro ejemplo, “*Estado civil de un adulto residente del D. F.*”, nos interesaría saber ¿cuántos adultos son casados?, ¿qué porcentaje de adultos viven en unión libre? ¿Cuál es la categoría es mas frecuente? y algunas otras cuestiones relacionadas con esta variable.

REPRESENTACIÓN DE LA INFORMACIÓN

Con el objeto de que los datos obtenidos en una muestra o población sean más significativos y apropiados, es frecuente construir distribución de frecuencias y a partir de estas dibujar gráficas de diferentes tipos para representar dichos datos.

El objetivo de esta sección se centra en el análisis de la relación entre la variable en observación y la frecuencia de las observaciones.

Para poder iniciar el análisis de los datos es necesario conocer las el numero de observaciones o datos que corresponden a cada categoría de la variable. Una vez conocida la frecuencia de cada una de las categorías de la variable es posible realizar su presentación haciendo uso de algún sistema simbólico.

Distribuciones de frecuencias.

Es una presentación tabular que consiste en presentar en dos o más columnas, la información entre las categorías de la variable y el número de veces que observo dentro de los datos analizados. La primera columna habitualmente contiene las diferentes categorías o intervalos de la variable y las siguientes columnas algunas de las distintas frecuencias asociadas, de manera que tal que cada renglón contiene una categoría y sus frecuencias correspondientes.

Iniciaremos el análisis de datos con la variable nominal.

VARIABLES NOMINALES

La **distribución de frecuencias** de una variable de escala nominal, es una tabla que enlista todas sus categorías y proporciona el número y/o el porcentaje de elementos, de la muestra o población, que caen en cada una de las categorías.

El color del automóvil.

La siguiente tabla muestra los colores de automóviles particulares más populares adquiridos en Norte América durante el año 2001

Distribución de Frecuencias

Color	Frecuencia	Frecuencia relativa (%)
Plateado (P)	630	21.00
Blanco (B)	468	15.60
Negro (B)	336	11.20
Azul (A)	297	9.90
Verde (V)	228	7.60
Rojo (R)	207	6.90
Café (C)	168	5.60
Oro (O)	135	4.50
Rojo brillante (RB)	129	4.30
Gris (G)	60	2.00
Otros (OT)	342	11.40
Total	3000	100

Inspeccionando la información ya organizada, podemos concluir que de acuerdo a los 3000 consumidores contactados, el color de mayor preferencia es el color plateado ya que 630 prefirieron este, que representa al 21% de todos los compradores. El segundo color en preferencia fue el blanco ya que 468 (el 15.6%) de los 3000 automóviles seleccionaron ese color, que el color de menor demanda fue el gris ya que únicamente 60 de los interrogados indicaron que adquirieron un auto de este color.

Esta misma información puede también ser deducida o resaltada de forma grafica o visual, haciendo uso de algunos de las representaciones más comunes.

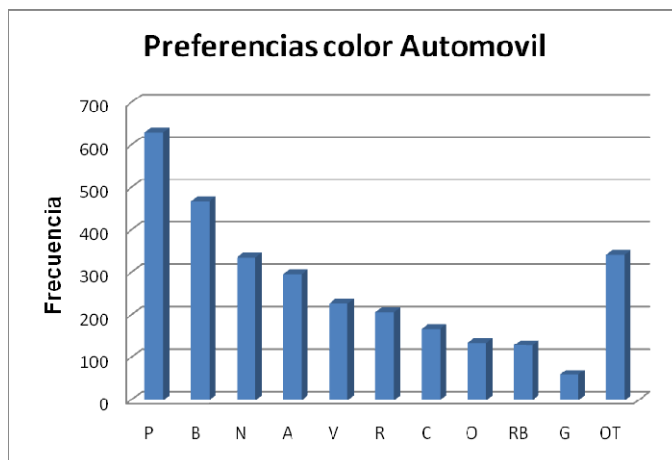
Comenzaremos con la representación de la información mediante el *diagrama de barras*.

Diagrama de Barras. Es un conjunto de barras o rectángulos cuyas alturas son proporcionales a la frecuencia o frecuencia relativa en cada categoría o nivel de la variable. Por lo general, las barras no son adyacentes, pero en ocasiones se dibujan tocándose entre sí.

Para la construcción de un diagrama de barras, se sitúan en el eje horizontal mediante un conjunto de puntos igualmente espaciados, cada una de las categorías de la variable nominal, sobre estos se traza una barra cuya altura es proporcional a la frecuencia o frecuencia relativa de la categoría considerada, debido a que las amplitudes de las bases de las barras no tienen relación con las categorías de la escala nominal, se considera práctico el tomar todas las bases iguales.

El diagrama de barras para la información del color de automóviles comprados en el año 2001, estaría expresado por el siguiente diagrama,

Diagrama de barras.



Otra grafica de gran utilidad para deducir información cualitativa acerca de los datos organizados es el diagrama de sectores o de pastel, que hace uso de la asignación de un sector de un círculo a cada uno de las categorías de la variable:

Diagrama circular, es la representación de las categorías de una variable mediante un círculo segmentado en sectores, de tal forma que los sectores son proporcionales a las frecuencias o frecuencias relativas de cada una de las categorías de la variable.

Para poder realizar esta grafica utilizamos una circunferencia de un radio cualquiera. Y dado que la circunferencia tiene un ángulo total de 360° y luego en forma proporcional se obtiene el ángulo central que corresponde a cada categoría de la variable nominal., de acuerdo a su frecuencia o frecuencia relativa.

El valor del ángulo central correspondiente a una categoría se determina mediante la siguiente expresión:

$$g = \frac{f \cdot 360}{n}$$

Donde n es en número total de datos en la distribución de frecuencias.

g es valor del ángulo central correspondiente a la categoría, cuya frecuencia es f.

Conocidos todos los ángulos centrales correspondientes a las frecuencias respectivas de cada categoría y con la ayuda de un transportador, se puede construir la división en sectores angulares de la circunferencia.

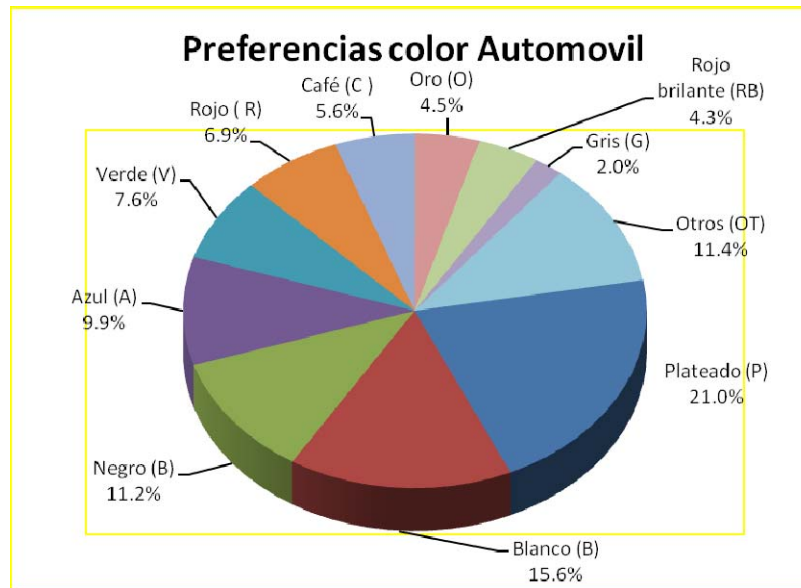
Son útiles para mostrar rápidamente diferencias de frecuencias entre las categorías, siempre que la cantidad de categorías en la variable nominal, no sea excesiva.

Para nuestro conjunto de datos de los automóviles adquiridos en 2001, los ángulos calculados para cada una de las categorías se expresan en la siguiente tabla.

Color	Frecuencia	f. r.	Grados
Plateado (P)	630	21.00	75.60
Blanco (B)	468	15.60	56.16
Negro (B)	336	11.20	40.32
Azul (A)	297	9.90	35.64
Verde (V)	228	7.60	27.36
Rojo (R)	207	6.90	24.84
Café (C)	168	5.60	20.16
Oro (O)	135	4.50	16.20
Rojo brillante (RB)	129	4.30	15.48
Gris (G)	60	2.00	7.20
Otros (OT)	342	11.40	41.04
Total	3000	100	360.00

Con la ayuda de un compás y un transportador podemos construir el diagrama de sectores correspondiente al conjunto de datos, el cual se muestra en la siguiente figura.

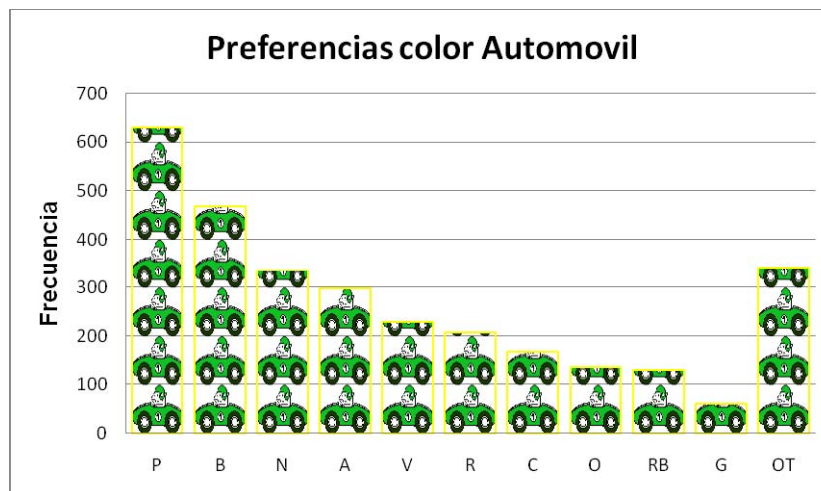
Diagrama de Sectores o circular



En algunas otras ocasiones con la finalidad de resaltar el aspecto visual, en lugar de utilizar un diagrama de barras, se hace uso de una representación grafica que utiliza una figura alusiva a la variable en estudio, generando lo que se conoce como *pictograma*.

Pictograma: Conjunto de símbolos alusivos a la variable que nos permiten representar la información entre una categoría y su frecuencia o frecuencia relativa.

Para construir un pictograma es posible utilizar cualquier clase de imagen, como pueden ser estrellas, columnas de monedas o billetes, de caras, de gente, etc. Cada símbolo representa un valor definido y uniforme, indicando la proporción que representa cada una de las figuras. Para la información relativa a los automóviles un posible pictograma es el siguiente:



Ejercicio: Elabore los diagramas de barras, circular y un pictograma para la variable superficie de áreas naturales protegidas (miles de hectáreas).

ÁREAS NATURALES PROTEGIDAS*	Superficie (miles de ha)
Parques Nacionales	1457
Reserva de la Biosfera	10957
Área de protección de recursos	40
Área de protección de flora y fauna	6073
Monumentos naturales	14
Santuarios	1
Otras categorías	187

* Agenda estadística INEGI 2004 pág. 9.

Variables Ordinales.

La **distribución de frecuencias** de una variable de escala ordinal es una tabla que enlista todas sus categorías y proporciona el número y/o el porcentaje de elementos, de la muestra o población, que caen en cada una de las categorías. Debido a que las categorías de la variable pueden ordenarse, es necesario representar las categorías o niveles en un orden determinado por ciertas reglas o criterios.

Por lo que los gráficos más usuales para una variable ordinal son los antes señalados con la salvedad de que las categorías o niveles deben de colocarse en un determinado orden.

Ejercicio: Los siguientes datos representan la intensidad de los Huracanes ocurridos en las costas del Atlántico de Estados Unidos, entre 1999 y 2007. El grado de intensidad se proporciona de acuerdo a la escala Saffir–Simpson, el cual toma diversos factores para su clasificación.

T 4 T 5 1 2 T 3 4 4 T T T T 1 T
 4 S T T T T T T 1 T T 1 T T T 1
 4 T 1 5 T 5 1 4 T T T 4 T T 1 2
 2 T 1 T T T 5 2 S 5 T 1 T T 1 2
 T 5 T 1 1 3 T T T 2 T T 3 3 T 1
 4 T 1 T 4 1 3 4 T 3 T 4 3 T 4 5
 4 5 T T 5 1 T 1 T T 2 1 1 2 1 1
 T T S 1 T 1 T T 1 T T 1 2 2 4
 2 1 T 1 T 5 1 5 1 T 3 3 4 T T
 2 T T T T 1 T 3 1 T T T T 4 2

*Orden de grados de intensidad S: tormenta subtropical, T: tormenta tropical, 1, 2, 3, 4, 5.

- a) obtenga la distribución de frecuencias del grado de intensidad de los huracanes.
- b) Calcule las frecuencias relativas, frecuencias acumuladas, frecuencias acumuladas relativas, frecuencias complementarias y frecuencias complementarias relativas.
- c) Construya el diagrama de barras, el diagrama circular y el diagrama de barras acumulativas correspondientes.

VARIABLES CUANTITATIVAS

Recordemos que las variables cuantitativas se dividen en dos tipos discretas y continuas.

Una variable es discreta cuando existan valores sucesivos para los que no hay valores intermedios, como es la característica: número de materias adeudadas en el C:C:H: en la cual podemos tener 0 materias, 1 materia, 2 materias,... etc., pero entre estos valores no hay valores intermedios posibles de observar. Es decir no podemos adeudar 1.5 materias, 1.7 materias, ya que adeudamos una o adeudamos dos.

Es decir es discreta es aquélla en la existen cortes, huecos o separaciones entre un valor posible y el siguiente, algunos ejemplos de esta variable son: numero de hijos, numero de habitaciones en una casa, numero de habitantes en la ciudad, numero de alumnos por grupo, etc.

Ejemplo: La siguiente tabla muestra el número de errores cometidos por el equipo de béisbol "Los cachorros" durante la temporada 2001.

Numero de errores	0	1	2	3	4	5	6	7	8	9	10
juegos	18	31	55	54	51	27	25	7	5	3	1

Una vez que la información está organizada podemos obtener la siguiente información:

Distribución de frecuencias:

Numero de errores	juegos	<i>f.r</i>	<i>f.a</i>	<i>f.a.r.</i>	<i>f.c.</i>	<i>f.c.r</i>
0	18	6.50	18	6.50	277	100.00
1	31	11.19	49	17.69	259	93.50
2	55	19.86	104	37.55	228	82.31
3	54	19.49	158	57.04	173	62.45
4	51	18.41	209	75.45	119	42.96
5	27	9.75	236	85.20	68	24.55
6	25	9.03	261	94.22	41	14.80
7	7	2.53	268	96.75	16	5.78
8	5	1.81	273	98.56	9	3.25
9	3	1.08	276	99.64	4	1.44
10	1	0.36	277	100.00	1	0.36

Analicemos el renglón de 2 errores cometidos en el juego, la frecuencia nos indica que se cometieron dos errores en 55 juegos (que es el numero de errores que mas se repitió), que representa al 19.46% de los todos los juegos, es decir en el 19.46% de los juegos se cometen dos errores. También podemos decir que en 104 juegos se cometieron dos o menos errores, que representan el 37.55%, es decir en el 37.55% de la temporada se cometieron dos o menos errores en los juegos. Aún más podemos decir que en 228 juegos se cometieron dos o más errores, que representa el 82.31%, con lo que podemos concluir que el 82.31% de las ocasiones se tuvieron juegos con al menos dos errores.

De una forma similar es posible analizar cada uno de los renglones, obviamente describirlo de esta forma nos daría bastantes líneas de información, se acostumbra solo escribir aquella información que se relevante en determinado momento.

Las representaciones graficas más comunes para este tipo de variable son:

Diagrama de Barras

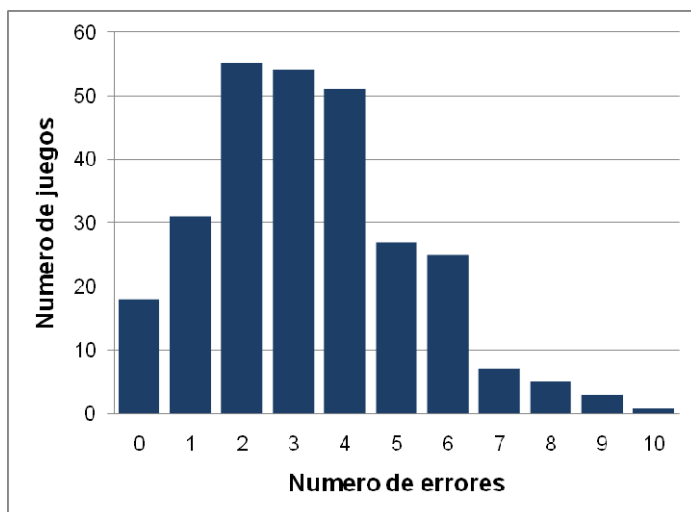
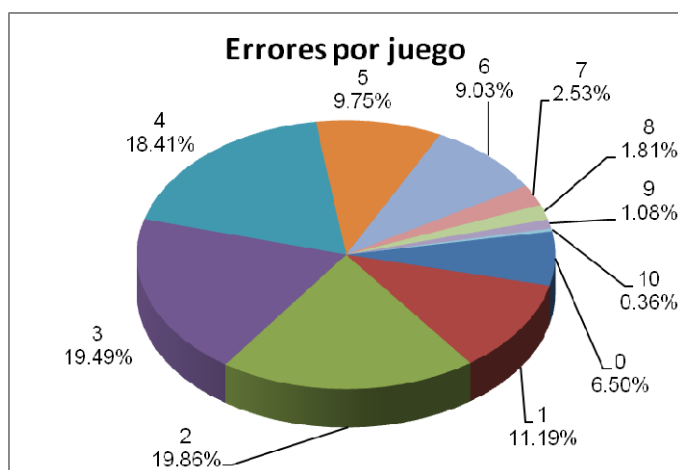
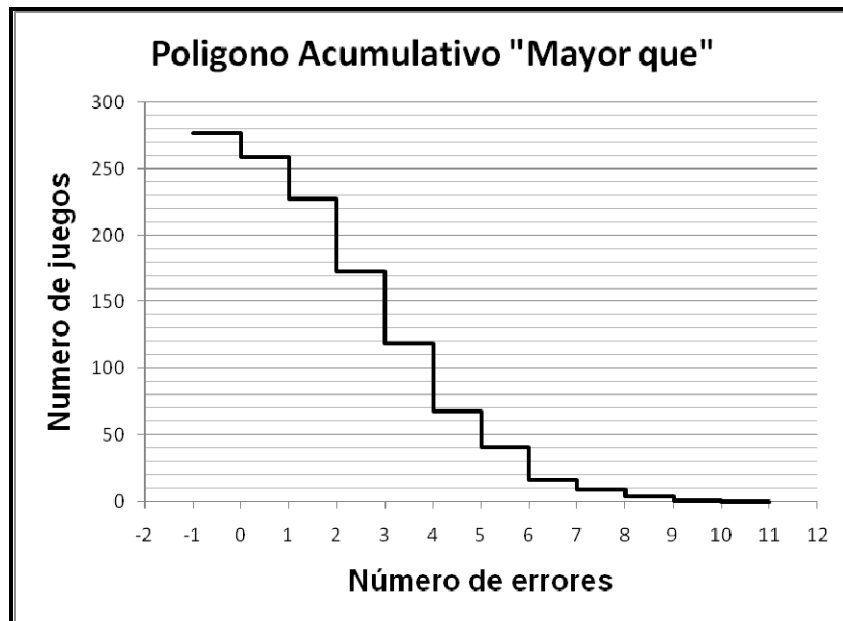
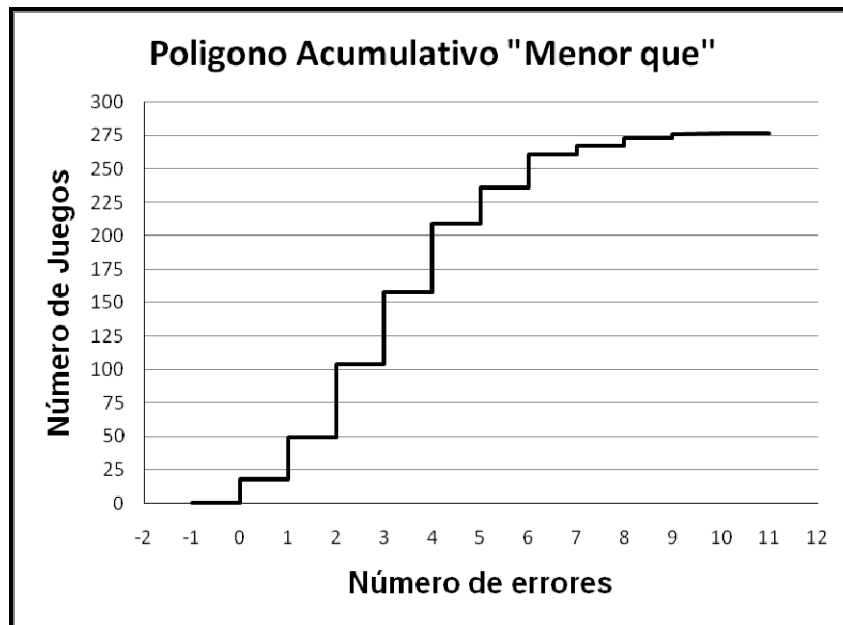


Diagrama Circular



POLÍGONOS ACUMULATIVOS

Se construye a partir del mismo eje horizontal del diagrama de barras, llevando sobre cada valor discreto una vertical de longitud proporcional a la frecuencia acumulada, absoluta o relativa, o la frecuencia complementaria de dicho valor. Se suele completar el gráfico dándole forma de una escalera de peldaños horizontales.



Variable continua es aquella en la cual es posible observar cualquier valor dentro de un intervalo de números reales, es decir que permite cualquier grado de subdivisión, algunas características humanas que caen dentro de esta situación son: estatura, peso, temperatura, edad etc.

Cuando en la población o muestra que estudiamos existen muchos valores diferentes, es conveniente, aún a costa de perder algo de información, dividir el intervalo de variación en una serie de subintervalos que cubran el total; a cada uno de ellos se le llama una clase, a sus extremos, extremos de clase, al punto medio de cada clase, marca de clase y a la diferencia entre sus extremos, ancho de la clase.

En el caso de la variable continua se complica un poco, ya que no está establecido ningún criterio para separar dos valores sucesivos, por la simple razón de que entre dos valores cualesquiera, tan cercanos como se desee, existe la posibilidad de que se presente otro valor.

El establecimiento de intervalos, para una variable continua sigue el criterio muy general, que dice que el número de intervalos de una distribución de frecuencias, no debe ser menor a 10 ni mayor de 20. Por lo que es necesario el establecer un número de intervalos que no sea muy grande, para que se destaquen las variaciones de frecuencia, ni tampoco muy pequeño, para que no se pierdan las diferencias que deseamos destacar.

Distribución de frecuencias.

A cada categoría de una variable le llamamos, cuando se trabaja con los datos agrupados, intervalo de clase y la distancia entre sus límites se llama ancho de de clase o amplitud.

El “tamaño o amplitud de intervalo” de clase depende de:

- El rango que abarcan los valores de la variable,
- La cantidad de datos u observaciones.

Con pocos intervalos desperdiciamos información; con muchos intervalos ahorramos poco trabajo. Conviene utilizar el mismo tamaño a lo largo de la distribución, pero no es obligatorio.

Por el “agrupamiento” se pierde información, porque una vez tabuladas las observaciones no se puede saber exactamente en qué valor dentro del intervalo de clase cae cada observación. Se supone que en el punto medio. Este es el “error por agrupamiento” que se compensa cuando el número de observaciones es grande y cuando el módulo (ancho de intervalo) es pequeño.

CONSTRUCCIÓN DE UNA DISTRIBUCIÓN DE FRECUENCIAS

CLASIFICACIÓN POR INTERVALOS

- A) Determine el rango del conjunto de datos. (valor máximo – valor mínimo)
- B) Seleccione el número apropiado de intervalos o clases. (de 6 a 20)
- C) Determine el ancho de cada clase. $c = \frac{\text{Rango}}{\text{N}^\circ \text{ de intervalos}}$ Redondeando a la siguiente unidad de acuerdo al formato de los datos.
- D) Indique el límite del primer intervalo, un valor igual o menor al valor mínimo.
- E) Sume el ancho de clase al punto inicial para obtener el límite inferior de la segunda clase. Suma la anchura de clase al límite inferior de la segunda clase para obtener el de la tercera, y así sucesivamente.
- F) Los límites inferiores sugieren los límites superiores, estableciéndolos de forma que no se superpongan y todos los datos puedan ser clasificados.
- G) Determinar mediante un conteo la frecuencia de cada intervalo.
- H) Calcular las frecuencias adecuadas para cada intervalo.

Apliquemos el procedimiento descrito anteriormente al siguiente conjunto de datos que proporcionan las estaturas en cm, (redondeado al cm más cercano), de una muestra de 300 estudiantes de la Universidad de Georgia.

180	180	177	162	162	177	182	172	180	180	200	182	195	170	170
193	175	180	182	170	159	188	193	185	175	182	177	179	176	159
172	175	162	175	161	193	200	164	166	162	170	188	180	162	172
172	172	167	180	170	164	180	185	180	170	170	185	170	164	180
188	182	182	154	170	190	167	170	170	157	180	164	177	154	177
167	170	190	167	182	193	172	170	170	180	167	168	166	188	185
188	177	172	159	175	177	149	167	164	190	167	157	170	180	164
180	159	182	172	159	173	170	180	185	157	144	167	159	164	167
167	159	185	180	167	188	170	171	182	162	177	175	162	164	175
188	175	177	177	162	190	177	154	170	185	172	172	159	185	164
182	234	170	180	173	180	164	168	175	167	172	157	195	172	177
168	170	164	162	180	180	180	175	159	159	162	175	189	182	164
190	164	164	164	159	180	180	167	167	164	159	172	154	170	188
177	198	164	166	170	182	173	177	162	170	167	164	184	159	159
164	195	175	170	182	175	164	190	167	177	185	162	159	164	175
185	172	157	162	159	170	171	167	164	167	157	177	166	175	172
177	164	170	182	159	172	162	167	191	164	177	172	172	172	177
168	181	170	157	167	172	164	167	162	182	172	159	164	164	162
180	180	200	167	159	177	182	167	175	172	159	172	164	167	162
170	168	154	177	190	182	182	146	188	164	172	154	172	172	172

A) *Determine el rango del conjunto de datos. (valor máximo – valor mínimo).* Tenemos que el valor mínimo de este conjunto de datos es de 144 y el máximo es de 234. Por lo que el rango de los datos es: $Rango = 234 - 144 = 90$

B) *Seleccione el número apropiado de intervalos o clases. (de 6 a 20).* Consideremos para nuestro caso consideremos que el número apropiado es: 11

C) *Determine el ancho de cada clase.* $c = \frac{Rango}{N^\circ \text{ de intervalos}}$ Redondeando a la siguiente unidad de acuerdo al formato de los datos. $c = \frac{90}{11} = 8.18 \approx 9$

D) *Indique el límite del primer intervalo, un valor igual o menor al valor mínimo, por ejemplo para nuestro primer límite podemos tomar un valor por debajo de 144, por ejemplo tomemos el 143.*

E) *Sume el ancho de clase al punto inicial para obtener el límite inferior de la segunda clase. Sume la anchura de clase al límite inferior de la segunda clase para obtener el de la tercera, y así sucesivamente. El segundo límite inferior es $143 + 9 = 152$, el tercero será $152 + 9 = 161$ y así sucesivamente hasta obtener los límites inferiores de los 11 intervalos de clasificación.*

F) *Los límites inferiores sugieren los límites superiores, estableciéndolos de forma que no se superpongan y todos los datos puedan ser clasificados. Para nuestro ejemplo podemos observar que el límite superior del primer intervalo será 151 (no debemos formar intervalos que se traslapen), el segundo sería 160, el tercero sería 169, y así sucesivamente.*

Intervalo	Límites	
	inferior	superior
1	143	151
2	152	160
3	161	169
4	170	
5	179	
6	188	
7	197	
8	206	
9	215	
10	224	
11	233	

G) *Determinar mediante un conteo la frecuencia de cada intervalo*

Intervalo	Límites		frecuencia	F
	inferior	superior		
1	143	151	///	3
2	152	160	////////////////////	33
3	161	169	////////////////////////////////////	80
4	170	178	////////////////////////////////////	98
5	179	187	////////////////////////////////////	56
6	188	196	////////////////////	25
7	197	205	////	4
8	206	214		0
9	215	223		0
10	224	232		0
11	233	241	/	1

H) *Calcular las frecuencias adecuadas para cada intervalo.*

Intervalo	Limites		f	fr	fa	far	fc	fcr
	inferior	superior						
1	143	151	3	3	3	1.00	300	100.00
2	152	160	33	36	36	12.00	297	99.00
3	161	169	80	116	116	38.67	264	88.00
4	170	178	98	214	214	71.33	184	61.33
5	179	187	56	270	270	90.00	86	28.67
6	188	196	25	295	295	98.33	30	10.00
7	197	205	4	299	299	99.67	5	1.67
8	206	214	0	299	299	99.67	1	0.33
9	215	223	0	299	299	99.67	1	0.33
10	224	232	0	299	299	99.67	1	0.33
11	233	241	1	300	300	100.00	1	0.33

REPRESENTACIONES GRAFICAS PARA DATOS AGRUPADOS

La diferencia significativa surge al considerar que lo que debemos representar es la frecuencia de cada intervalo y no la frecuencia de cada observación individual, ya que se trata de datos agrupados. Se pueden seguir dos procedimientos:

- 1) Representar la frecuencia del intervalo por una superficie, en cuyo caso la figura que resulta se denomina "histograma".
- 2) Representar la frecuencia del intervalo por una ordenada, en cuyo caso la figura de denomina "polígono de frecuencias".

Por supuesto, la información de ambos gráficos es exactamente igual y son reemplazables entre sí.

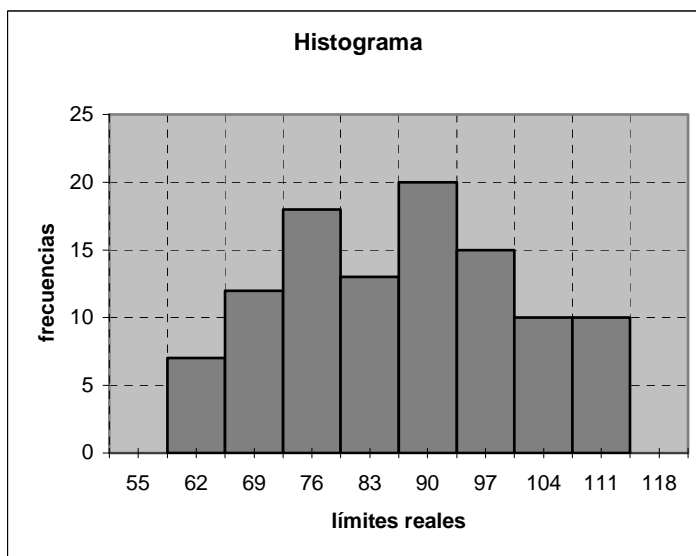
Para construir un histograma (del griego columna) se considera que cada rectángulo del mismo tiene como base la amplitud del intervalo real y como altura la frecuencia del intervalo. Es decir, que el histograma es un conjunto de rectángulos.

CONSTRUCCIÓN DEL HISTOGRAMA DE FRECUENCIAS

- A) Obtenga Los límites reales para cada intervalo
- B) En el eje horizontal de un sistema cartesiano grafique cada uno de los límites reales.
- C) Para cada intervalo construya una barra vertical, de tal forma que la base de dicha barra se encuentre entre los límites del intervalo y su altura sea igual a la frecuencia del intervalo. (Esto nos indica que no debe de haber espacio entre las barras, a menos que la frecuencia del intervalo sea cero)

Ejemplo: Construir el Histograma de frecuencias para el siguiente conjunto de datos.

Intervalo		f	límites reales	
59	65	7	58.5	65.5
66	72	12	65.5	72.5
73	79	18	72.5	79.5
80	86	13	79.5	86.5
87	93	20	86.5	93.5
94	100	15	93.5	100.5
101	107	10	100.5	107.5
108	114	10	107.5	114.5



POLÍGONO DE FRECUENCIAS.

Para representar las observaciones según un polígono de frecuencias, es válido lo dicho anteriormente, con la diferencia de que la frecuencia correspondiente a cada intervalo se representa por una ordenada que está ubicada en el punto medio de intervalo respectivo. El punto medio o marca de clase de cada intervalo se calcula por la fórmula:

$$m = \frac{L_i + L_s}{2}$$

Donde:

L_i Límite inferior del intervalo

L_s Límite superior del intervalo

En general se debe terminar el polígono de frecuencias sobre la intersección del eje de las abscisas, a la izquierda y la derecha de la primera y de la última marca de clase, respectivamente, a una distancia igual a una amplitud del intervalo.

Para dibujar un polígono de frecuencias se toma como base la marca de clase, localizando puntos cuyas coordenadas son la marca de clase y la frecuencia o la frecuencia relativa de cada intervalo, luego uniendo los puntos mediante segmentos de rectas,

Para completar el polígono, también se dibujan líneas rectas desde los puntos por encima de la primera y última marca de clase, respectivamente, a puntos en el eje horizontal, que

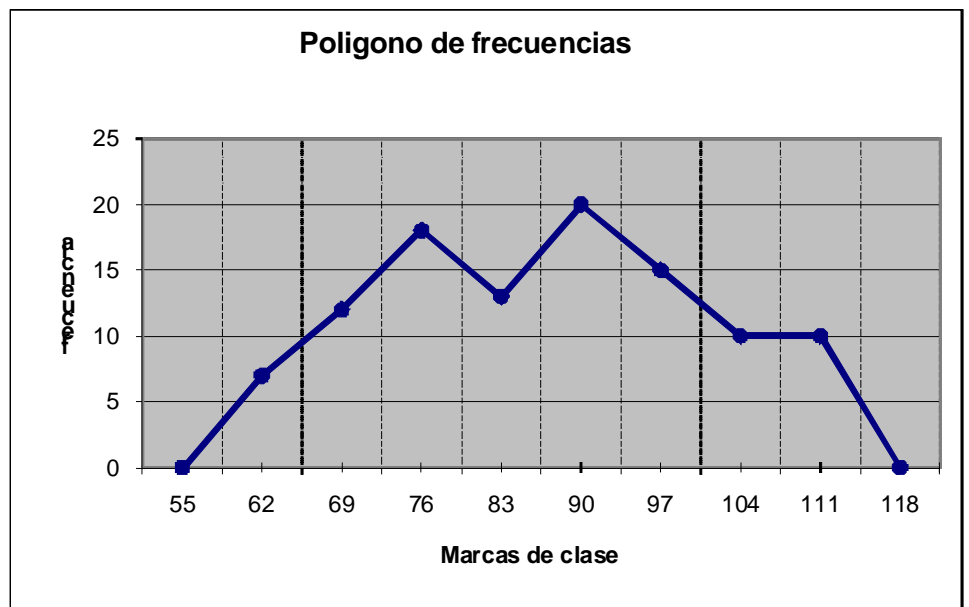
se encuentren a la mitad de la longitud de la unidad estándar de medida abajo del más bajo y del más alto límite de clase.

CONSTRUCCIÓN DEL POLÍGONO DE FRECUENCIAS

- A) Obtenga la marca de clase para cada intervalo
- B) Agregue dos marcas de clase adicionales, una antes de la primera marca de la clase y otra después de la última marca de clase, cada una de ellas con frecuencia cero y al agregarlas se debe tomar en cuenta el ancho de intervalo.
- C) En el plano cartesiano grafique cada uno de los puntos formado por x: marca de clase del intervalo; y: frecuencia correspondiente.
- D) Unir los puntos graficados en C) con segmento de recta. La figura así construida se le llama "Polígono de frecuencias".

Ejemplo: Construir el polígono de frecuencias de la siguiente distribución de datos:

intervalos	F	m
Auxiliar	0	55
58.5 - 65.5	7	62
65.5 - 72.5	12	69
72.5 - 79.5	18	76
79.5 - 86.5	13	83
86.5 - 93.5	20	90
93.5 - 100.5	15	97
100.5 - 107.5	10	104
107.5 - 114.5	10	111
Auxiliar	0	118



La forma de un histograma también puede aproximarse por una curva de frecuencias suavizadas. Esta aproximación puede lograrse mediante técnicas matemáticas o graficas y por lo general, sirve al propósito de eliminar irregularidades como resultado del error de muestreo de un histograma que describa información reunida en una encuesta muestral. La curva de frecuencias puede visualizarse como una estimación del histograma que resultaría

si la información se graficara la información del censo con muchas clases de ancho reducido.

Formas de las curvas de frecuencias.

La forma de la distribución puede indicarnos aproximadamente qué modelo de matemático o estadístico podremos utilizar para estudiarla, por ejemplo si es unimodal, simétrica o en forma de campana y si es posible utilizar el modelo de la curva normal de Gauss, que es muy común en muchas áreas científicas, y que es de gran utilidad para obtener conclusiones de mayor confiabilidad acerca del conjunto de datos.

POLÍGONOS ACUMULATIVOS (OJIVAS)

Cuando representamos una distribución de frecuencias utilizando algunas de la frecuencia acumulada (acumulada o complementaria) a la figura obtenida se le llama **ojiva**.

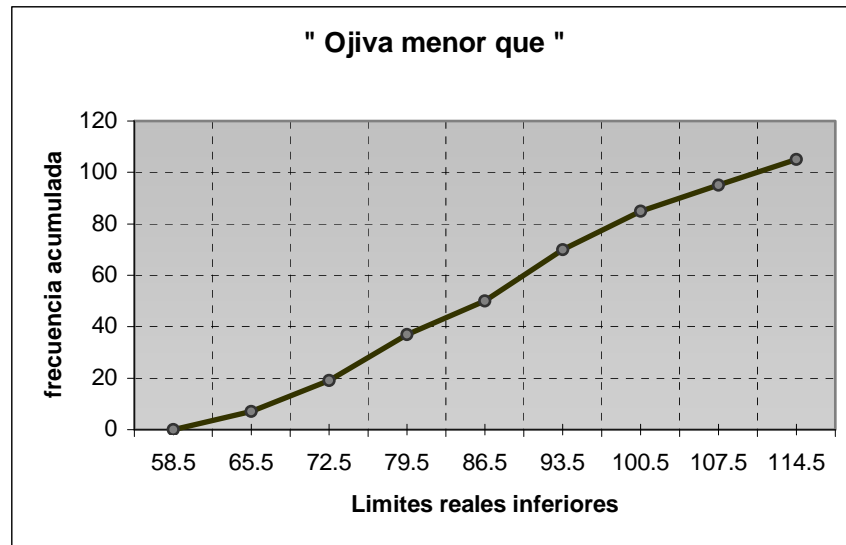
Si en una distribución de frecuencias acumuladas transformamos a éstas en frecuencias acumuladas relativas obtenemos una curva S estilizada que se llama ojiva. La famosa ojiva de Galton es la curva de distribución de frecuencias acumuladas en términos de porcentajes y facilita la comparación entre distintas distribuciones por ser la altura común.

Para una distribución de frecuencias acumulativas del tipo menor o igual, cada frecuencia acumulada de clase se graficará mediante un punto cuyas coordenadas serán el límite real superior del intervalo y la frecuencia acumulada correspondiente.

CONSTRUCCIÓN DE “LA OJIVA MENOR QUE”

- A) Obtenga la frecuencia acumulada para cada intervalo
- B) Considere los límites reales superiores de cada clase
- C) En el plano cartesiano grafique cada uno de los puntos formado por x: límite real superior del intervalo; y: la frecuencia acumulada correspondiente.
- D) Unir el primer límite real inferior y los puntos graficados en C) mediante segmentos de recta. La figura así construida se le llama “Ojiva menor que”, o “Polígono de frecuencias acumuladas”.

Intervalos		F	Fa
58.5	65.5	7	7
65.5	72.5	12	19
72.5	79.5	18	37
79.5	86.5	13	50
86.5	93.5	20	70
93.5	100.5	15	85
100.5	107.5	10	95
107.5	114.5	10	105

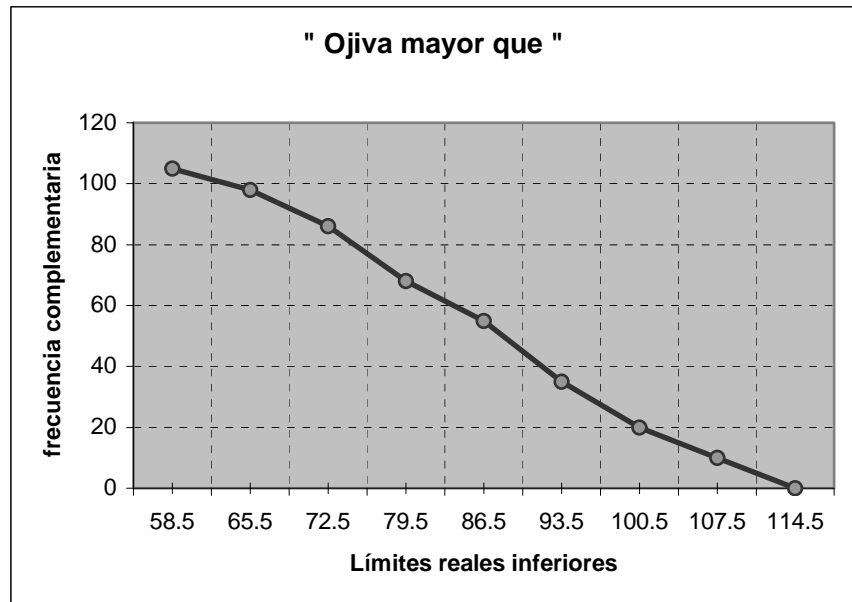


En el caso de una distribución de frecuencias acumulativa del tipo “mayor o igual que, cada frecuencia complementaria de clase se representa mediante un punto formado cuyas coordenadas son el límite real inferior de la clase y la frecuencia complementaria correspondiente.

CONSTRUCCIÓN DE “LA OJIVA MAYOR QUE”

- A) Obtenga la frecuencia complementaria para cada intervalo
- B) Considere los límites reales inferiores de cada clase
- C) En el plano cartesiano grafique cada uno de los puntos formado por x: límite real inferior del intervalo; y: la frecuencia complementaria correspondiente.
- D) Unir los puntos graficados en C) y el último límite real superior mediante segmentos de recta. La figura así construida se le llama “Ojiva mayor que”, o “Polígono de frecuencias complementarias”.

Intervalos		f	Fc
58.5	65.5	7	105
65.5	72.5	12	98
72.5	79.5	18	86
79.5	86.5	13	68
86.5	93.5	20	55
93.5	100.5	15	35
100.5	107.5	10	20
107.5	114.5	10	10



Los siguientes datos representan los pesos de una muestra de 140 estudiantes de la Universidad de Georgia. Clasifique estos datos haciendo uso de 9 intervalos.

Construya la distribución de frecuencias correspondiente.

Elabore las graficas: histograma, polígono de frecuencia, y las ojivas correspondientes a los datos clasificados por intervalos.

49	44	91	68	54	57	51	66	58	64
54	59	85	63	70	59	84	73	64	71
69	49	63	64	78	61	59	70	59	55
72	58	64	54	52	57	75	63	88	60
70	56	76	66	77	50	63	66	59	74
58	54	73	52	48	58	70	65	59	
77	52	60	58	56	53	59	49	63	
59	54	58	59	58	51	60	94	63	
64	54	54	62	54	77	68	62	64	
79	51	59	72	69	69	58	51	70	
45	49	73	62	56	64	73	49	68	
53	50	53	56	75	68	63	58	62	
60	53	77	75	61	65	55	58	55	
58	59	63	62	61	84	60	49	67	