

Unidad 2

Dato: Información concreta sobre hechos, elementos, etc., que permite estudiarlos, analizarlos o conocerlos, es una medida de cualquier variable.

2.1 Orden y clasificación de los datos

Una masa de datos es el grupo de *datos crudos*, es decir, tal como se obtuvieron en el proceso de captura de la información, sin llevar ningún tipo de corrección o modificación. El reunir o agrupar los datos implica nociones básicas de orden para establecer los posibles niveles de agrupamiento, dispersión, sesgo que se entre ellos, además de permitir posteriormente análisis de mayor complejidad como la comparación, la correlación y la proyección.

2.1.1 Tamaño de la masa de datos

- a) Masas de datos pequeña: Con un número máximo de 25 elementos.
- b) Masa de datos grande: Con un número mayor a 25 elementos.

2.1.2 Tipos de masas de datos por el nivel de medición

El conocer el *nivel de medición* de la variable permitirá determinar los procesamientos a los que se podrá someter dicha información, por lo que existen dos tipos de masas de datos:

Cualitativas: Cuando manejan datos nominales y no se le puede dar más allá de un tratamiento comparativo elemental.

Cuantitativas: cuando se presentan datos a nivel ordinal, intervalar o de razón (este último no aplicable a las ciencias sociales) y que por tal motivo es posible aplicar un tratamiento aritmético de mayor complejidad.

La descripción de una masa de datos inicia a través de las medidas que la caracterizan, en este sentido las medidas nominales solo permiten su identificación de acuerdo a una categoría y señalar el número de veces que se repiten los datos en las categorías que correspondan.

Mientras que las masas de datos a nivel ordinal o superior permite identificar las diferencias numéricas que se dan entre los integrantes de cada categoría ($X_{MIN}...X_{MAX}$).

2.1.3 Tipo de masas de datos de acuerdo al número de variables involucradas

Univariadas: Involucra una sola variable: (X_1, X_2, \dots, X_N)

Bivariadas: Involucra dos variables: $((X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N))$, cada pareja entre paréntesis indica los datos que corresponden a cada elemento.

Multivariadas: Involucra k variables

$((X_{1,1}, X_{2,1}, \dots, X_{k,1}), ((X_{1,2}, X_{2,2}, \dots, X_{k,2}), \dots, ((X_{1,N}, X_{2,N}, \dots, X_{k,N}))$

2.2 Distribución de Frecuencias y Distribución de Frecuencias Agrupadas

La frecuencia es la cantidad de veces que aparece un dato en un grupo.

La distribución se da al organizar la masa de datos y se determina la frecuencia con que aparece cada uno de ellos.

Por tal motivo, se puede entender que la distribución de frecuencias de una muestra de observaciones en un conjunto de clases, junto con el conjunto de observaciones de cada clase. Actúa como un útil resumen de las principales características de los datos, como son la localización, la forma y la dispersión. Para llevarlo de manera sistematizada, es necesario considerar tres aspectos importantes: la clasificación, la ordenación y la distribución de las frecuencias.

2.2.1 Clasificación: Es un arreglo que indica la frecuencia con que aparece cada dato y los agrupa en clases

$C_i \equiv$ Categoría (o clase) i -ésima	
$f \equiv$ Frecuencia con que aparece el dato	
$k \equiv$ Número de categorías	
Forma General de una Clasificación	
X	f
C_1	$f(C_1)$
...	...
C_k	$f(C_k)$

2.2.2 Rango: Conocido también como oscilación, recorrido o amplitud total, y es la distancia exacta entre el dato mayor y menor de una masa de datos.

$R \equiv$ Rango de una masa de datos
$R = X_{MAX} - X_{MIN} + 1$

2.2.3 Ordenación: Es el arreglo de los datos en los que se organizan de menor a mayor (ascendente) o de mayor a menor (descendente), aunque regularmente siempre se asume el orden ascendente.

Forma General de una Ordenación
X_{MIN}, \dots, X_{MAX}

La distribución de frecuencias adopta la siguiente forma:

$X \equiv$ Dato	
$f \equiv$ Frecuencia con que aparece el dato	
Forma General de una distribución de Frecuencias	
Dato	Frecuencias
X_{MIN}	$f(X)$

	
	X_{MAX}	$f(X)$	

La distribución de frecuencias permite calcular otros valores necesarios para otras etapas del procesamiento estadístico:

$X \equiv$ Dato $f_a \equiv$ Frecuencia con que aparece el dato Forma General de una distribución de Frecuencias					
Dato	Frecuencia	Frecuencia acumulada	Frecuencia del dato	Dato al cuadrado	Frecuencia del dato al cuadrado
X	f	Fa	fX	X^2	fX^2

2.3 Dato Puntual

Es aquel que constituye un punto en la recta numérica conformada por X_{MIN} y X_{MAX}

2.4 Dato Intervalar

Aquella que se construye con base a la cantidad de datos disponibles para reducir la proporción de renglones (o clases) y facilitar el proceso de análisis de la información.

El tamaño del intervalo depende del dividir el rango entre el número de datos para cada intervalo; el cociente redondeado al entero sucesor indica el número de renglones que tendrá la distribución de frecuencias. Aunque no hay criterios definidos, por conveniencia se recomienda elaborar una distribución de frecuencias cuando se tenga más de 15 renglones.

$\Sigma \equiv$ Suma total
$TAB =$ Tabulación
Construcción de una distribución de frecuencia puntual
1. Construir la columna X comenzando en X_{MIN} y termina en X_{MAX}
2. Construir la columna TAB , tabulando cada dato que aparezca en la masa de datos
3. Construir la columna f contando las tabulaciones de cada renglón
4. Construir la columna fa , sumando cada frecuencia con las anteriores
5. Obtener Σf .
6. Si $(\Sigma f \neq N)$ o $(fa_{MAX} \neq N)$ entonces existe error en la tabulación
7. Construir la columna fX
8. Obtener ΣfK
9. Construir la columna X^2
10. Construir la columna fX^2
11. Obtener ΣfX^2

2.5 Distribución de Frecuencia Intervalar

Para decidir el tamaño del intervalo, se divide el rango entre el número de datos que tendrá cada intervalo, el cociente redondeado al entero sucesor dará el número de renglones que tendrá la distribución de frecuencias.

$i \equiv$ Tamaño del Intervalo
A considerar: X_{MIN} , X_{MAX} , y R
$R = X_{MAX} - X_{MIN}$
$di =$ Datos por intervalo (a conveniencia)
$I = R/di$

2.6 Construcción Grafica de los Datos

2.6.1 Normas para la elaboración de las diferentes gráficas

Origen: Si se usan espacios coordenados, el cruce de los ejes X (abscisa) y Y (ordenada) constituye el origen y corresponde a cero. Si se contraen cualquiera de los dos o los dos se debe señalar y los ejes siempre cruzarán en cero. En las gráficas circulares no se debe desplazar el centro o rotar la gráfica en su conjunto.

Ubicación de las variables en los ejes: En distribuciones univariadas el eje X representa los datos y el Y la frecuencia.

Graduación de los ejes: Se debe indicar con claridad la graduación de cada eje de acuerdo a la escala que representa.

Proporción entre ejes: Se recomienda que la proporción del eje Y sea entre el 70% y 80% de la que tenga el eje X .

Rotulación: Los ejes, barras o sectores deben rotularse claramente para facilitar su identificación, así como también se indicarán proporciones o frecuencias.

2.6.2 Datos Generales: Deberán de contener:

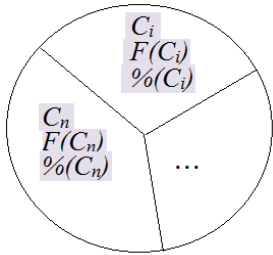
- Título
- Fecha a que corresponden los datos
- Fuente de donde se obtuvieron los datos
- Responsable de la elaboración particular o institucional
- Fecha en que se elaboró la gráfica

2.7 Tipos de gráficos

<i>Tipo de variable</i>	<i>Nivel de medición</i>	<i>Frecuencia simple</i>	<i>Frecuencia acumulada</i>
Discreta	Nominal	Circulograma	
	Ordinal	Barras	

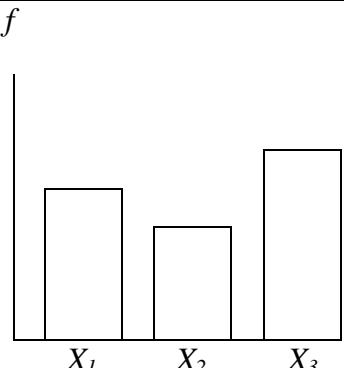
Continua	Intervalar	Histograma	Escalera
	De razón	Polígono de frecuencias	Ojiva

2.7.1 Circulograma (gráfica de pastel o gráfico de sector): Se diseña para enfatizar la naturaleza categórica de la variable y sus frecuencias (o porcentajes).

Características	Construcción																				
$C_i \equiv$ categoría o clase i -ésima	$ANG \equiv$ Ángulo de la categoría i -ésima																				
$f \equiv$ Frecuencia con que aparece el dato	1. Construir la nota de cálculo																				
$\% \equiv$ Porcentaje respecto al total	<table border="1"> <thead> <tr> <th>Categoría</th> <th>f</th> <th>$ANG \equiv (fN) \times 360$</th> <th>$\% = (fN) \times 100$</th> </tr> </thead> <tbody> <tr> <td>C_1</td> <td></td> <td></td> <td></td> </tr> <tr> <td>...</td> <td></td> <td></td> <td></td> </tr> <tr> <td>C_n</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Totales</td> <td>N</td> <td>360^0</td> <td>100%</td> </tr> </tbody> </table>	Categoría	f	$ANG \equiv (fN) \times 360$	$\% = (fN) \times 100$	C_1				...				C_n				Totales	N	360^0	100%
Categoría	f	$ANG \equiv (fN) \times 360$	$\% = (fN) \times 100$																		
C_1																					
...																					
C_n																					
Totales	N	360^0	100%																		
	2. Trazar un círculo																				
	3. Medir los ángulos correspondientes a cada categoría																				
	4. Trazar los radios de cada sector																				
	5. Rotular cada sector con el nombre de la categoría, su frecuencia y su porcentaje y completar datos generales																				

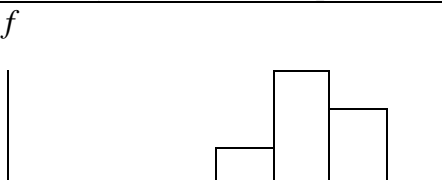
2.7.2 Gráfica de Barras

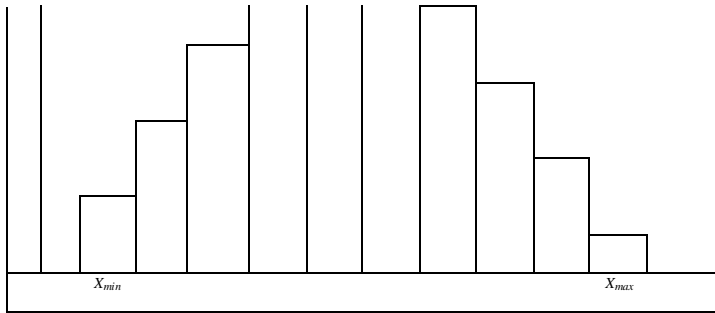
Las barras se colocan en orden jerárquico de la escala y se separan entre sí.

	1. Trazar un plano coordenado en que la longitud del eje Y sea entre 70% y 80% de la X
	2. Marcar en X las barras necesarias y rotular bajo el eje, cuidan el orden, ancho de las barras separadas por igual distancia.
	3. Graduar Y desde 0 hasta la frecuencia más alta de la distribución.
	4. Trazar las barras hasta la altura de su frecuencia.
	5. Rotular los ejes y agregar los datos generales

2.7.3 Histograma

Cada graduación del eje X corresponde a una barra y representa un intervalo (o categoría). Las barras se construyen juntas entre sí y las líneas verticales marcan los límites exactos del intervalo (la izquierda el inferior y la derecha el superior)

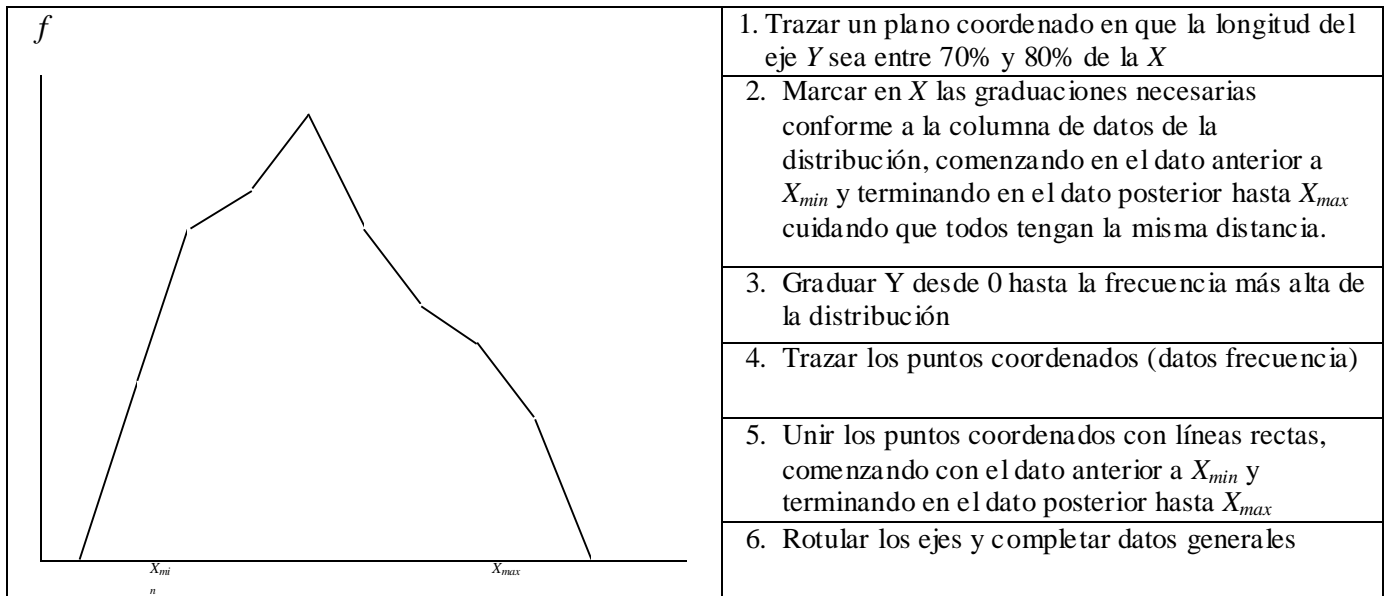
	1. Trazar un plano coordenado en que la longitud del eje Y sea entre 70% y 80% de la X
	2. Marcar en X las graduaciones desde X_{min} hasta X_{max} y rotularlas cuidando el mismo ancho en las barras, rotulando cada marca con el intervalo



correspondiente
3. Graduar Y desde 0 hasta la frecuencia más alta de la distribución
4. Trazar las barras hasta la altura de las frecuencias
5. Rotular los ejes y completar los datos generales

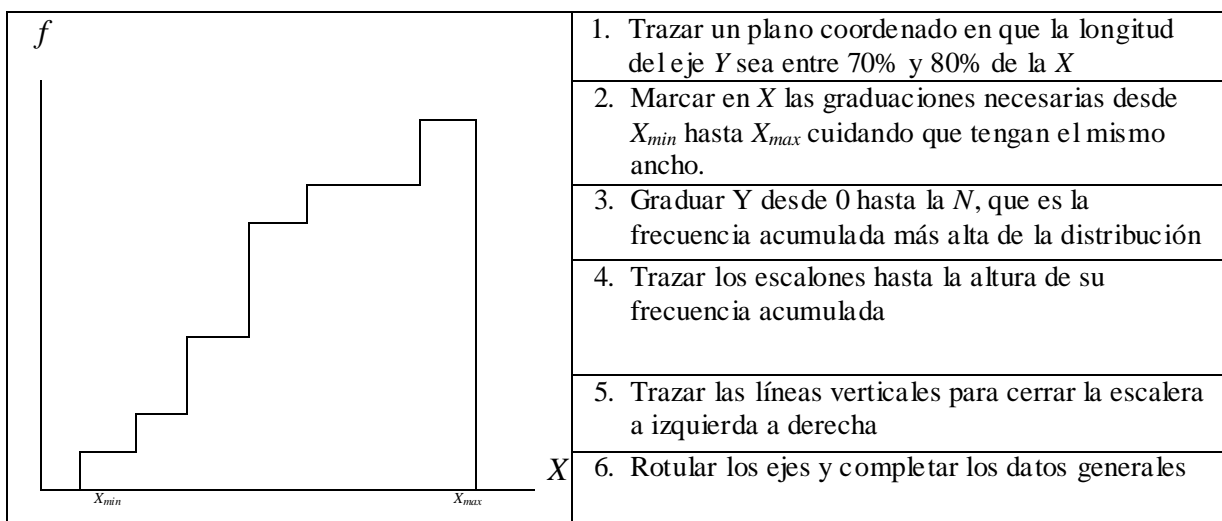
2.7.4 Polígono de Frecuencias

Representa distribuciones de frecuencias puntuales en el que los puntos coordenados se unen entre sí con líneas rectas que representan el carácter continuo de los datos. Se conoce también como *curva de frecuencias*, si no se tiene certeza en relación a la presentación continua de los datos es mejor realizar un histograma.



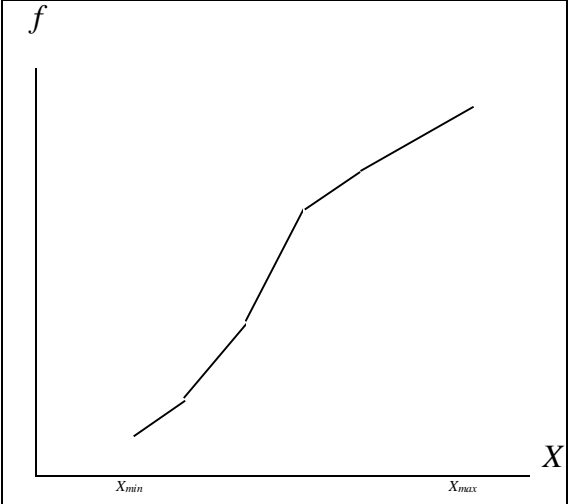
2.7.5 Gráfico de Escalera

O conocida también como *ojiva discreta*, representa distribuciones intervalares de frecuencias acumuladas, utiliza la misma distribución que el histograma, construyéndose a partir de la columna de frecuencias acumuladas (*fa*). La altura de cada escalón representa la frecuencia simple del intervalo.



2.7.6 Ojiva

Llamada también *gráfica acumulativa*, es equivalente al gráfico de escalera. Representa distribuciones puntuales de frecuencias acumuladas, en el que los cruces de coordenadas se unen entre sí con líneas rectas para enfatizar su carácter continuo.

	1. Trazar un plano coordenado en que la longitud del eje Y sea entre 70% y 80% de la X
	2. Marcar en X las graduaciones necesarias comenzando en X_{min} y terminando en X_{max} cuidando que tengan la misma distancia.
	3. Graduar Y desde 0 hasta la N, que es la frecuencia acumulada más alta de la distribución
	4. Trazar los puntos coordenados (dato, frecuencia acumulada)
	5. Unir los puntos coordenados con líneas rectas, comenzando en X_{MIN} y terminando en X_{MAX} .
	6. Rotular los ejes y completar los datos generales

Unidad 3

Curva Normal.

En las ciencias del comportamiento las distribuciones tienden a adoptar una forma característica conocida como *curva normal* o *distribución gaussiana*. La *curva normal* es una curva simétrica y asintótica, definida por la media, desviación estándar o típica y por el valor de la ordenada máxima. Los fenómenos conductuales tienden a distribuirse normalmente. Dentro de la curva normal, existen una serie de elementos importante para identificar y describir los fenómenos psicológicos, sí como la forma en que estos se observan en el mundo real. Para esto existe la clasificación de las diversas medidas que arroja la estadística descriptiva: Medidas de proporción, medidas de ubicación, medidas de tendencia central, medidas de variabilidad, medidas de forma y medidas de correlación.

La curva normal comprende para su entendimiento de una serie de elementos.

3.1 Medidas de proporción

3.1.1 Redondeo: Permite ajustar la fracción decimal a un número de dígitos preestablecido. Para el redondeo aunque no existen criterios claramente establecidos se considera que si la cifra representa una proporción, la fracción decimal se redondeará a cuatro dígitos, así al transformarla en porcentaje la fracción decimal quedará con cuatro dígitos. De no ser así solo se redondeará a dos dígitos.

3.1.2 Proporción: Es el resultado de dividir la cardinalidad (frecuencia) de un conjunto a entre la cardinalidad de un conjunto b . Al multiplicarla por 100, se obtiene un porcentaje.

3.1.3 Distribución de proporciones: Se da a partir de la clasificación o distribución de frecuencias. En la distribución de proporciones la suma de sus partes siempre es 1.0000

3.1.4 Ajuste de sumas: Al calcular la distribución de proporciones el redondeo de cifras puede provocar imprecisiones por lo que es necesario ajustar los datos para que su resultado final sea igual a 1.000. Considerando la siguiente regla: si la suma es mayor que 1.000, restar un diezmilésimo a la proporción más grande, otro a la siguiente y así sucesivamente hasta ajustar la suma a 1.000. Si la suma es menor que 1.0000, aumentar un diezmilésimo a la proporción más pequeña, otro a la siguiente y así sucesivamente hasta ajustar la suma a 1.000.

3.2 Medidas de ubicación

3.2.1 Límites exactos: En variables continuas cualquier puntaje empírico corresponde al punto medio de un rango. Es decir, un valor entero por ejemplo 110, señala que este valor se encuentra entre 109.50 y 110.50. Por otra parte, en distribuciones intervalares se consideran los límites del intervalo, por ejemplo un punto que se ubique entre el intervalo 91-95 señalará que sus límites exactos serán 90.50 y 95.50.

3.2.2 Fractil: Es el dato o puntaje X que limita por la derecha una proporción cualquiera de una distribución. El cálculo de fractiles permite calcular cualquier medida de ubicación al determinar la proporción de la distribución que limita.

3.2.3 Rango percentil: Es la proporción de la distribución que queda limitada por un dato determinado y permite evaluar la posición del resultado de una prueba en particular entre los resultados generales de la misma.

(Escotet, 1979) (Everitt & Wykes, 2001) (López A. , 2007) (López A. , 2011) (Ritchey, 2002)

Bibliografía

- Escotet, M. A. (1979). *Estadística Psicoeducativa* (3a reimpresión ed.). México: Trillas.
- Everitt, B., & Wykes, T. (2001). *Diccionario de Estadística para Psicólogos*. España: Ariel.
- López, A. (2007). *EL Método en Ciencias del Comportamiento*. México: Universidad Autónoma del Estado de México.
- López, A. (2011). *Estadística Descriptiva en Ciencias del Comportamiento*. México: Universidad Autónoma del Estado de México.
- Ritchey, F. J. (2002). *Estadística para las Ciencias Sociales*. México: McGraw-Hill.