

CAPITULO I

REPRESENTACION DE DATOS

A menudo, las mediciones son hechas con una muestra de ' n ' artículos tomados de un gran conjunto. Este gran conjunto se denomina '*población*'. Por ejemplo, se puede haber tomado una muestra de 100 lapiceros de un lote de 10,000 producidos con el objeto de efectuar mediciones de la calidad de lapiceros. Generalmente se desea "*generalizar*" los resultados obtenidos de la muestra hacia la población de la cual fue extraída. Esta generalización se denomina **inferencia de la muestra a la población**. La inferencia solo es posible si la muestra es totalmente representativa de la población. Por ejemplo, de la muestra de 100 lapiceros se concluye que el 1% de esa muestra (1 lapicero) presenta defectos al escribir. De esto se *infiere* que en el lote de 10,000 lapiceros, 100 de ellos presentaran defectos al escribir. El conjunto de datos obtenidos debe ser organizado, resumido y mostrado antes de que se intente alguna interpretación. Gráficos de ploteo y diagramas son muy útiles para observar adecuadamente la información obtenida.

En lo referente a notación, n , representa el número de observaciones en un conjunto de datos; las observaciones están representadas por una variable con subíndice; p. ej., X_1, X_2, \dots, X_n . Así, la representación de cinco valores, $n = 5$, de la resistencia a la fractura de una aleación medidos por un ingeniero será:

$$X_1 = 2310; X_2 = 2325, X_3 = 2315, X_4 = 2340, X_5 = 2335 \text{ (en unidades psi).}$$

1.1 Organización y reporte de datos: Tablas y Gráficas

Los datos tal como se obtienen de una investigación están en forma desordenada, por lo que es difícil su interpretación y análisis. Debido a esto se deben organizar en forma de Tablas y Gráficas para permitir una visualización clara y rápida de todo el conjunto. En la tabulación y descripción de los datos se debe seguir ciertos principios generales.

1.1.1 Principios generales sobre la construcción de Tablas.

1. Las Tablas se explicarán por sí mismas enteramente, se ha de dar suficiente información en el título y en los encabezados de las columnas para permitir que el lector identifique fácilmente su contenido.
2. Cada variable numérica debe contener sus unidades.
3. La función del rayado debe dar claridad de interpretación. Debe evitarse el rayado excesivo e innecesario.
4. No se debe incluir demasiada información en una sola Tabla.
5. Las anotaciones numéricas de cero se deben escribir explícitamente en vez de usar guión, ya que este se usará para indicar datos que faltan o no se han observado.
6. Una anotación numérica no debe comenzar con punto decimal.
7. Los números que indican valores de la misma característica se han de dar con el mismo número de cifras decimales.

Ejemplo:

Tabla Nro.: Cantidades de mineral en el proceso de lixiviación.

Tanque Nro.	Alimentación					Desechos	
	TMH ⁽¹⁾	% H ₂ O	TMS ⁽²⁾	% Cu oxidado	Contenido de Cu, kg.	% Cu oxidado	Contenido de Cu, kg.
4	7.160	2.43	6.986	8.35	583	0.35	24
6	3.580	2.50	3.491	8.44	295	0.40	14
10	7.000	2.58	6.819	8.50	580	0.45	31
12	3.500	2.40	3.416	9.10	311	0.54	18
14	3.600	2.45	3.152	9.00	316	0.60	21
16	3.800	2.50	3.705	8.40	311	0.50	19
18	3.600	2.40	3.514	8.70	306	0.60	21
TOTAL	32.240	2.47	31.443	8.59	2702	0.49	148

(1) Toneladas métricas húmedas. (2) Toneladas métricas secas.

1.1.2. Principios generales sobre construcción de gráficas

1. Las gráficas se han de explicar enteramente por si mismas.
2. Las escalas vertical (ordenada) y horizontal (abscisas) estarán rotuladas con claridad dando las unidades y con los mismos intervalos para las ordenadas y abscisas respectivamente.
3. No se debe abarcar demasiada información en un solo gráfico. Es mejor hacer varios gráficos que comprimir demasiada información en uno solo.
4. La finalidad de las gráficas es dar una visión general y no una imagen detallada del conjunto de datos.
5. Debe evitarse la inclusión de números dentro del cuerpo de la gráfica.

Ejemplos:

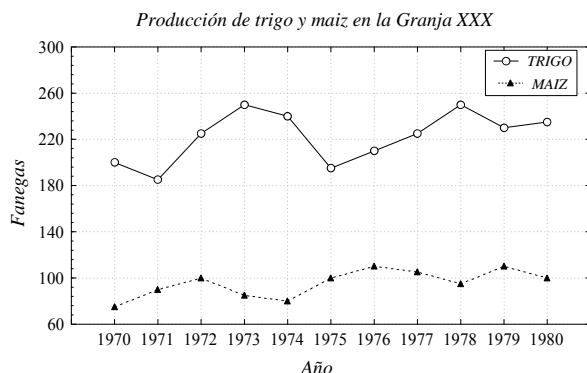


Figura Nro..... Gráfica de series de tiempo de producción de la granja

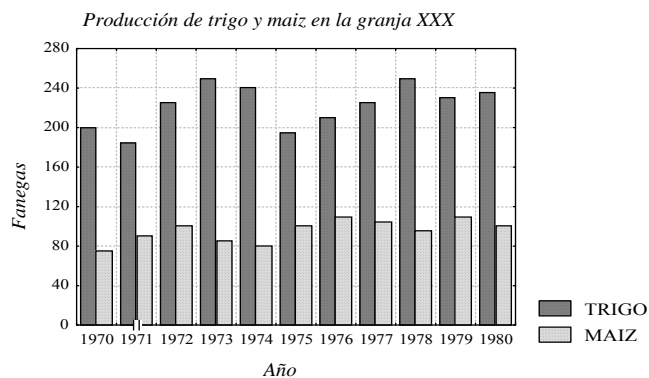


Figura Nro.....: Gráfico de barras de producción de la granja

Superficies de zonas del mundo

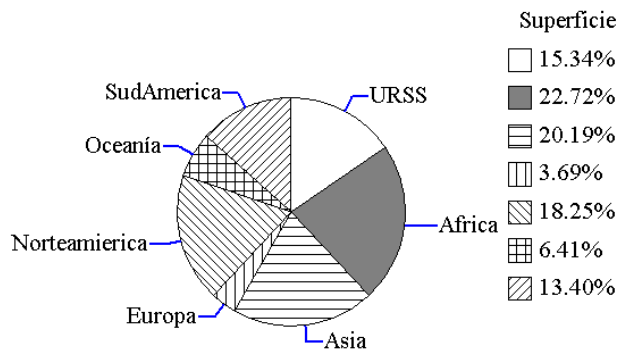


Figura Nro.: Gráfico de pastel de distribución de áreas en el mundo.

1.2. Tipos de datos.

Para seleccionar un procedimiento estadístico a utilizar es necesario conocer que tipo de datos tenemos; estos pueden ser: continuos, discretos, nominales o categóricos y ordinales o jerarquizados.

Los que provienen de variables continuas, tales como temperatura, viscosidad, peso, longitud, etc., están dentro de esta categoría. En general, los datos continuos son los que se obtienen al medir una variable con algún instrumento.

Una variable discreta es la que puede asumir solo ciertos valores, por lo general enteros. Los datos discretos surgen al contar el número de conceptos que posee cierta característica, como por ejemplo los defectos en un lote de tela, los accidentes de trabajo en una fábrica, el número de alumnos dentro de un salón de clase, etc.

Tanto los datos continuos como los discretos se conocen como datos cuantitativos. Por otra parte, los ordinales y los nominales, son datos cualitativos y se les convierte a números antes de trabajar con ellos.

Los datos nominales comprenden categorías como sexo, color de ojos o de piel, animales con o sin síntoma de determinada enfermedad, con defecto o sin defecto, turno de producción, lote producido, etc.

Los datos ordinales o jerarquizados se refieren a evaluaciones subjetivas según preferencia o logro. Por ejemplo, si un investigador desea analizar el efecto de cierta lesión cerebral sobre la conducta materna de ratones, puede establecer como criterio de evaluación la calidad de nido que construye: excelente, bueno, regular o malo. Estos términos son reducidos luego a números, (1, 2, 3,...) calificándose así las características que se evalúa.

Es interesante observar que algunas poblaciones pueden proporcionar los cuatro tipos de datos, por ejemplo:

POBLACIÓN	CONTINUOS	DISCRETOS	NOMINALES	ORDINALES
Automóviles	peso, longitud	número de defectos por auto.	colores	tamaño
Ratas de laboratorio	Peso, edad	Número de crías por rata	Color: negra, blanca, gris	

A pesar de la clasificación señalada, para la mayor parte de las aplicaciones estadísticas, basta con clasificar los datos como cuantitativos y cualitativos.

Es importante destacar en este punto que según el tipo de datos que se este analizando, la evaluación estadística difiere. Es así que existe tratamiento estadístico de datos si estos son discretos; si son datos continuos, el tratamiento difiere en cierta medida. Estos tipos de datos se tratan respectivamente por la Distribución Binomial y la Distribución Normal.

1.3. Diagramas circulares y diagramas de barras

Los datos se pueden agrupar según la frecuencia y la proporción o porcentaje de cada categoría y representar gráficamente mediante diagramas circulares (gráficas de pastel) y diagramas de barras. Por ejemplo, se ha recolectado 80 datos correspondientes a los colores de animales según se especifica en la Tabla:1.1

Tabla 1.1: Frecuencia y % según casos para ejemplificar diagramas.

CASOS	FRECUENCIA	%
Blancos	48	60
Manchados	16	20
Negros	12	15
Pardos	4	5

Para construir el diagrama circular se procede a dividir el círculo de acuerdo a las proporciones dadas, así 360° equivale al 100%; 216° equivale al 60 %, 72° al 20%, etc. Para distinguir las distintas regiones se utilizan diferentes colores o puntos, rayas, cruces, etc. El diagrama de barras representa los datos estableciendo barras que se levantan desde la abscisa hasta una altura correspondiente al dato a representar, como se muestra en la Figura 1.1.

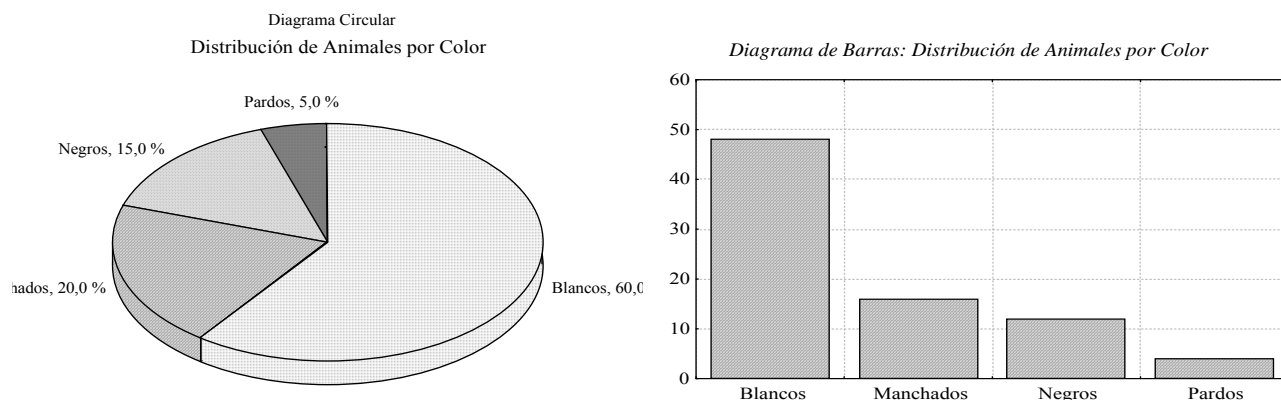


Figura 1.1: Diagrama de pastel y de barras para datos de colores de animales.

1.4 Diagramas de Pareto

Alfredo Pareto (1848 – 1923) llevó a cabo muy completos estudios sobre distribución de la riqueza en Europa. Descubrió que unos cuantos concentraban la mayor parte de la riqueza, en tanto que era muy grande el número de pobres. Esta desigual distribución de la riqueza se convirtió en parte fundamental de la teoría económica y se convirtió en un concepto universal. De esto se acuñó las frases *minoría vital* y *mayoría útil*.

Un diagrama de Pareto es una gráfica en donde se organizan diversas clasificaciones de datos por orden descendente, de izquierda a derecha, como puede observarse en la Figura 1.2. En este caso, las clasificaciones de datos corresponden a tipos de fallas producidas en el campo. La minoría vital aparece a la izquierda de la gráfica y la mayoría útil a la derecha. Hay veces que hay que combinar elementos de la mayoría útil en una sola clasificación denominada *otro*, y que en la Figura se designa con 'o'. Siempre que se utilice la categoría *otros*, esta deberá colocarse en el extremo derecho. La escala vertical es para el uso de valores, en tanto que la escala horizontal se coloca categorías.

Mediante el diagrama de Pareto se puede detectar los problemas que tienen más relevancia. Por lo general, alrededor del 80% de los resultados generales se origina en el 20% de los elementos. La gráfica es muy útil al permitir identificar visualmente de una sola revisión las minorías vitales a las que es importante prestar atención.

Conviene aclarar que una mejora en la calidad de la minoría vital, por ejemplo, del 50%, producirá mucho mayor rendimiento que una mejora del 50% de la mayoría útil. Asimismo, la experiencia ha mostrado que es más fácil lograr una mejora del 50% de la minoría vital.

Supóngase por ejemplo que la *falla tipo F* es el objetivo de los esfuerzos correctivos del programa de mejora de la calidad, Figura 1.2. Se comisiona a un equipo de ingenieros para que investigue y efectúe mejoras sobre ese tipo de fallas. La siguiente vez que se efectúe otro análisis de Pareto, ya debe ser el tipo de *falla C* el objetivo de la corrección, y de esta forma el proceso de mejora continua, hasta lograr que las fallas se conviertan en un problema insignificante en un contexto generalizado de lograr mejorar los atributos de calidad de un proceso.

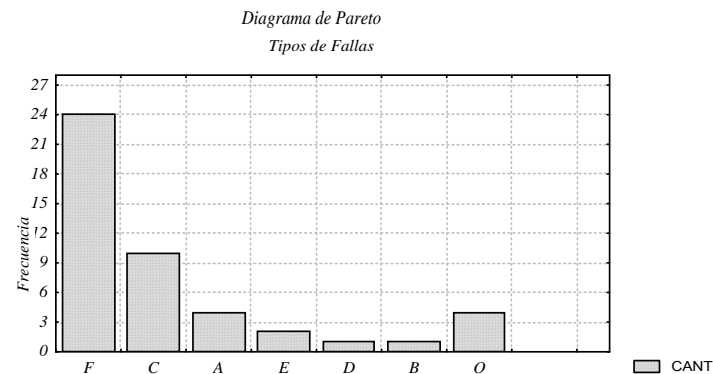


Figura 1.2: Diagrama de Pareto para tipos de fallas

1.5 Gráficas de Series de Tiempo

Las gráficas consideradas hasta el momento son métodos visuales muy útiles para mostrar la variabilidad presente en los datos.; sin embargo, con frecuencia el tiempo es un factor importante que contribuye a la variabilidad observada en los datos y los métodos anteriores no lo toman en cuenta.

Una serie de tiempo, o secuencia de tiempo, es un conjunto de datos en los que las observaciones se registran en el orden en que ocurren. La gráfica de una serie de tiempo es un diagrama en el que el eje vertical denota el valor observado (por ejemplo, $X = \text{Ventas}$), mientras que el eje horizontal denota tiempo (que puede ser en años, minutos, días, etc.). Cuando se grafican las mediciones como una serie de tiempo, a menudo se observan tendencias, ciclos u otras características importantes de los datos que, de otra forma, pasarían inadvertidas.

Por ejemplo, la Figura 1.3 presenta la gráfica de una serie de tiempo de las ventas anuales de una compañía durante los últimos diez años. La impresión general que ofrece esta gráfica es que las ventas tienen una tendencia a crecer. Existe cierta variabilidad en esta tendencia, donde, las ventas en algunos años aumentaron con respecto a las del año anterior, mientras que las ventas de otros años disminuyeron.

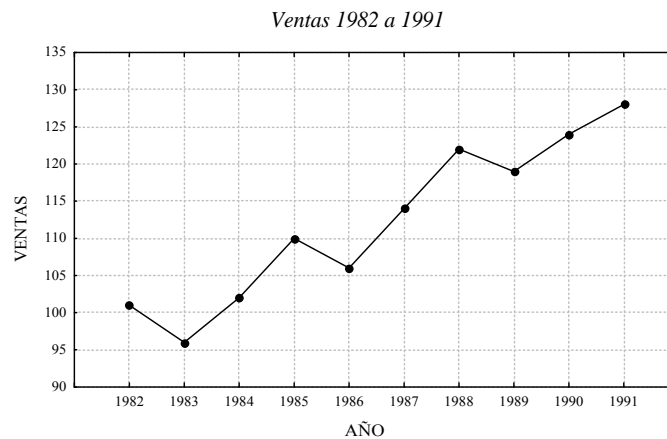


Figura 1.3: Ventas de una compañía por año.

1.6 Cartas de Control

Estos gráficos representan otra manera útil de examinar la variabilidad en datos que dependen del tiempo. La Figura 1.4 representa una carta de control para los datos de concentración de un proceso químico tomadas cada hora. La línea central de la carta de control representa el promedio de las mediciones de concentración para las 20 primeras muestras (91.45g/l). El límite superior de control y el límite inferior de control son un par de límites obtenidos estadísticamente que reflejan la variabilidad inherente o natural del proceso. Si el proceso está trabajando como se debe, sin ninguna fuente externa de variabilidad presente en el sistema, las mediciones de concentración deberán fluctuar de manera aleatoria alrededor de la línea central, y casi todas ellas deben caer dentro de los límites de control.

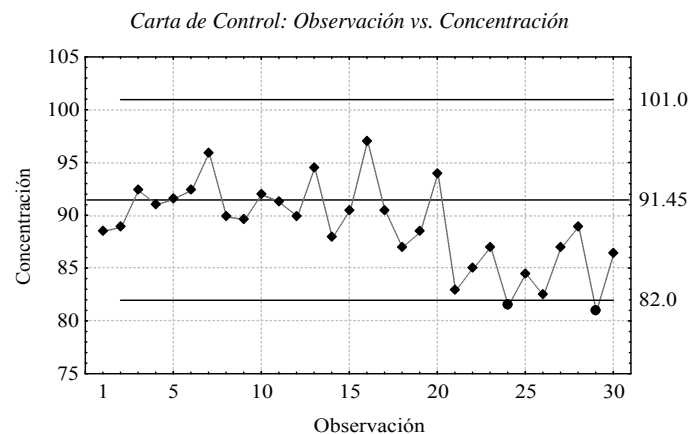


Figura 1.4: Carta de Control para los datos de un proceso químico.

En la carta de control de la Figura 1.4 el marco visual de referencia proporcionado por la línea central y los límites de control, indica que algún cambio o perturbación alrededor de la muestra 20 ha tenido efecto sobre el proceso, ya que todas las observaciones que siguen están por debajo de la línea central, y mas aún, dos de ellas se encuentran por debajo del límite inferior de control. Esto constituye un indicador muy fuerte de que el proceso requiere de una acción correctiva. Si se puede encontrar y eliminar la causa que originó el cambio, podrá mejorarse de manera considerable el rendimiento del proceso.

Las cartas de control son una aplicación muy importante de la estadística para la vigilancia, control y perfeccionamiento de un proceso. La rama de la estadística que hace uso de las cartas de control se conoce como *Control Estadístico de Procesos, CEP*. Estos métodos son procedimientos para vigilar la variación de un proceso y para generar información de la estabilidad del proceso, ya que un proceso inestable resultará en la producción de artículos defectuosos, baja calidad y en general, pérdida de la confianza del consumidor.

Si bien es cierto las gráficas de control son recursos excelentes para resolver problemas al facilitar el mejoramiento de la calidad, también es cierto que su utilidad es limitada cuando se trata de monitorear y mantener un proceso.

1.7. Diagramas de dispersión.

La manera más sencilla de definir si existe una relación causa y efecto entre dos variables, es dibujando un diagrama de dispersión. En la Figura 1.5 muestra la relación que existe entre presión y volumen de un gas. A medida que aumenta la presión, el volumen del gas disminuye. Aquí se establece una situación de dependencia e independencia, en donde la variable presión es independiente y la variable volumen depende de los efectos de la presión. La variable independiente se colocara siempre en el eje X y la variable dependiente o respuesta, en el eje Y.

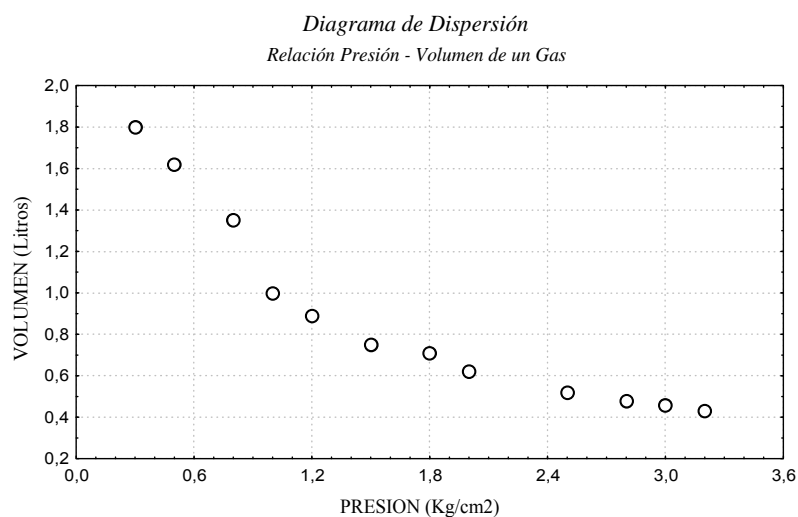


Figura 1.5: Diagrama de dispersión para la relación P – V de un gas.

1.8. Diagramas de puntos, distribución de frecuencias e histograma.

El siguiente ejemplo, Tabla 1.2, permitirá ilustrar la forma de graficar los resultados de mediciones. Se anota en esta tabla la resistencia a la compresión (unidades de 100 libras por pulgada cuadrada) de 90 bloques de concreto. Estas mediciones fueron efectuadas para medir la variabilidad entre los bloques comprados a un proveedor.

Esta información sin un tratamiento previo y sin gráficos apropiados no es muy útil. Para obtener información concreta de los datos obtenidos se puede elaborar un diagrama de puntos. En este diagrama se marca en un eje horizontal las observaciones realizadas. En la Figura 1.6 se muestra un diagrama de puntos para las 10 primeras observaciones. Este diagrama muestra la variabilidad existente en 10 observaciones. Esta gráfica permite ver con rapidez y facilidad la ubicación o tendencia central de los datos; así como su dispersión o **variabilidad**

Tabla 1.2 (a): Resistencia a la compresión de bloques de concreto x 100 psi; tal como se obtuvieron los datos.

49.2	53.9	50.0	44.5	42.2	42.3	32.3	31.3	60.9	47.5
43.5	37.9	41.1	57.6	40.2	45.3	51.7	52.3	45.7	53.4
51.0	45.7	45.9	50.0	32.5	67.2	55.1	59.6	48.6	50.3
45.1	46.8	47.4	38.3	41.5	44.0	62.2	62.9	56.3	35.8
38.3	33.5	48.5	47.4	49.6	41.3	55.2	52.1	34.3	31.6
38.2	46.0	47.0	41.2	39.8	48.4	49.2	32.8	47.9	43.3
49.3	54.5	54.1	44.5	46.2	44.4	45.1	41.5	43.4	39.1
39.1	41.6	43.1	43.7	48.8	37.2	33.6	28.7	33.8	37.4
43.5	44.2	53.0	45.1	51.9	56.6	48.5	39.0	47.3	48.8

Tabla 1.2 (b): Resistencia a la compresión de bloques de concreto (100 libras por pulgada cuadrada); tabla ordenada.

28.7	34.3	39.1	41.6	44.0	45.7	47.4	48.8	51.9	55.2
31.3	35.8	39.1	42.2	44.2	45.7	47.4	49.2	52.1	56.3
31.6	37.2	39.8	42.3	44.4	45.9	47.5	49.2	52.3	56.6
32.3	37.4	40.2	43.1	44.5	46.0	47.9	49.3	53.0	57.6
32.5	37.9	41.1	43.3	44.5	46.2	48.4	50.0	53.4	59.6
32.8	38.2	41.2	43.4	45.1	46.8	48.5	50.0	53.9	60.9
33.5	38.3	41.3	43.5	45.1	46.9	48.5	50.3	54.1	62.2
33.6	38.3	41.5	43.5	45.1	47.0	48.6	51.0	54.5	62.9
33.8	39.0	41.5	43.7	45.3	47.3	48.8	51.7	55.1	67.2

El diagrama de puntos es una gráfica muy útil para visualizar un conjunto pequeño de datos. A menudo estos diagramas son útiles al comparar dos ó más conjuntos de datos. Si el número de observaciones es pequeño, a menudo es difícil identificar algún patrón específico; sin embargo, el diagrama de puntos es útil y puede proporcionar información sobre características poco usuales de datos. Cuando el número de observaciones es moderadamente grande, pueden ser más útiles otros tipos de gráficas.

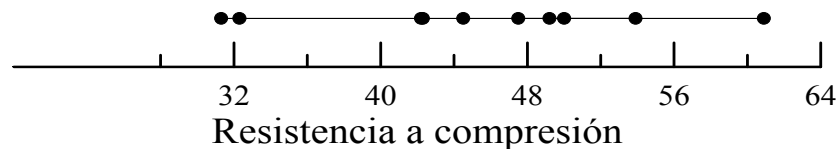


Figura 1.6: Diagrama de puntos para 10 mediciones de resistencia a la compresión.

Para las 90 observaciones este diagrama no sería muy útil por la gran cantidad de puntos que tendrían que agruparse en una escala corta. Para gran cantidad de datos es mejor construir una Tabla de distribución de frecuencias, como se muestra en la Tabla 1.3 y graficar los resultados en forma de un histograma como se muestra en la Figura 1.7.

Para determinar la frecuencia de distribución, primero se ordena las observaciones de menor a mayor, X_{min} , X_{max} . El **rango**, una medida de variabilidad de los datos, se establece según:

$$R = X_{max} - X_{min} = 67.2 - 28.7 = 38.5$$

El rango de los datos se divide en intervalos, los cuales se conoce como *intervalos de clase o celdas*. Las clases deben tener el mismo ancho con la finalidad de mejorar la información visual en la distribución de frecuencias. El número de intervalos depende del número de observaciones y de la *dispersión* de los datos. También dependerá del número de observaciones que se tenga, por lo que se podrá tener más intervalos con menores longitudes con mayor cantidad de datos. En muchos casos resulta satisfactorio usar entre 5 y 20 clases, y que el número de clases debe aumentar en función de n . En la práctica se obtiene buenos resultados si se hace la selección del número de clases aproximadamente igual a:

$$Nro. de clases = \sqrt{n}$$

En el ejemplo, se determina 10 intervalos, cada uno de los cuales con un ancho de 4. La primera clase se inicia en 28.0 a 32 (sin incluir 32), la segunda de 32 a 36 (sin incluir 36), etc. La clase final corresponde 64.0 - 68.0. Los puntos finales de cada celda se denominan *límites de celda* y el valor central de cada celda el *punto medio*. Luego se cuenta el número de valores que pertenecen a cada clase y se elabora una tabla de distribución de frecuencia. Por convención, si el valor de una observación corresponde exactamente a un límite, se respeta la designación del intervalo cerrado al comienzo y abierto al final. Así cada observación caerá en alguna de las clases establecidas. En la tabla se anotará la frecuencia, f_i , es decir, el número de ocurrencia de observaciones en cada clase. La frecuencia relativa, f_i/n resulta de dividir la frecuencia respectiva entre el total de observaciones. La frecuencia acumulada se obtiene de:

$$f_{cum} = \sum_{j=1}^i f_j = f_1 + f_2 + \dots + f_i \quad (i=1, 2, \dots, k)$$

y la frecuencia relativa acumulada:

$$\frac{f_{cum}}{n} = \frac{1}{n} \sum_{j=1}^i f_j$$

Tabla 1.3: Distribución de frecuencias.

Intervalo	Punto Medio		Frecuencia f_j	Frecuencia Relativa f_j/n	Frecuencia Acumulada f_{cum}	Frecuencia Acumulada Relativa f_{cum}/n
[28.0 - 32.0[30.0		3	0.033	3	0.033
[32.0 - 36.0[34.0		8	0.089	11	0.122
[36.0 - 40.0[38.0		10	0.111	21	0.233
[40.0 - 44.0[42.0		15	0.167	36	0.400
[44.0 - 48.0[46.0		22	0.244	58	0.644
[48.0 - 52.0[50.0		15	0.167	73	0.811
[52.0 - 56.0[54.0		9	0.100	82	0.911
[56.0 - 60.0[58.0		4	0.044	86	0.955
[60.0 - 64.0[62.0		3	0.033	89	0.988
[64.0 - 68.0[66.0		1	0.011	90	1.000
			90	1.000		

Un histograma está formado por un conjunto de rectángulos que representan la frecuencia de una categoría; representa gráficamente las frecuencias correspondientes a los valores observados. En el histograma, según la Figura 1.7, se gráfica en la abscisa las clases y en la ordenada o bien la frecuencia o la frecuencia relativa. El histograma muestra como están distribuidos los valores de la variable de interés; condensa la información en forma tal que la comprensión visual se hace muy fácil mostrando los valores típicos, dispersión y forma de la distribución.

En la Figura 1.8 se gráfica la frecuencia acumulada vs. los puntos medios de cada intervalo. Una *ojiva* es la línea que conecta los puntos. Las zonas de mayor pendiente en la curva respectiva, corresponden a las clases más frecuentes. La frecuencia acumulada muestra por ejemplo que 36 de 90 observaciones (o 40 % de las observaciones) son menores que 44. Similarmemente, que 89 de 90 (98.8 %de las observaciones) son menores que 64. La definición de frecuencia acumulada implica que la ojiva no es decreciente y siempre se encuentra entre 0 y 1.

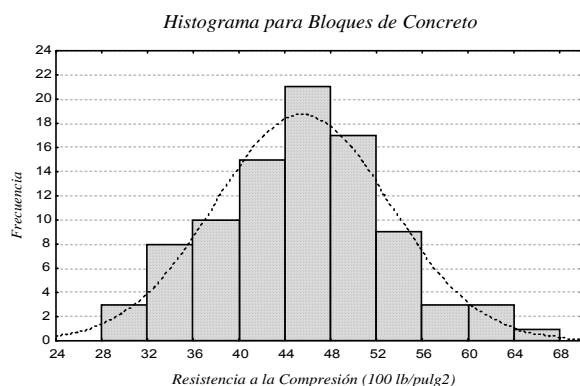


Figura 1.7: Histograma para las mediciones de resistencia a la compresión

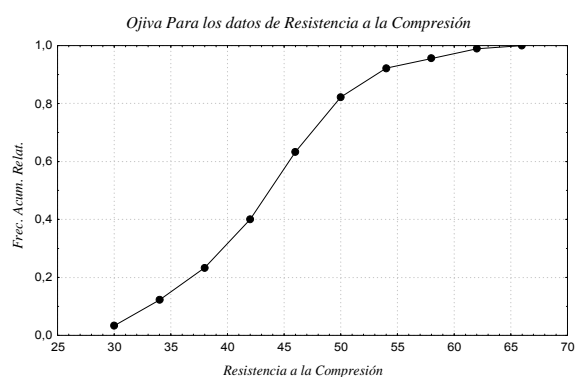


Figura 1.8: Gráfico de los datos de resistencia a la compresión, frecuencia acumulada.

El histograma de la Figura 1.7 es aproximadamente simétrico, con un solo pico o 'joroba' en la parte media. Este pico representa a la clase de mayor frecuencia, que en el ejemplo corresponde a la clase del 44 al 48.

El punto medio de la clase es el 46, el cual se denomina 'moda' y representa el valor más frecuente en el conjunto de datos. Este tipo de distribuciones con un solo pico en el medio son muy comunes. Lo son tanto que muchas de ellas se les denomina *distribuciones normales*.

Los histogramas de la Figura 1.9(a) y 1.9(b) tienen un solo pico pero no son simétricas. Se dicen que son distribuciones sesgadas a la izquierda o la derecha. Pueden corresponder a índices de mortalidad humana, la que es mayor a ciertas edades. El histograma de la Figura 1.9(c) presenta dos picos o dos modas. A este tipo de histogramas se les denomina '*bimodales*'. Pueden explicarse, por ejemplo, por la presencia de dos diferentes tipos de procesos, dos máquinas diferentes, o que las condiciones de una fábrica varía.

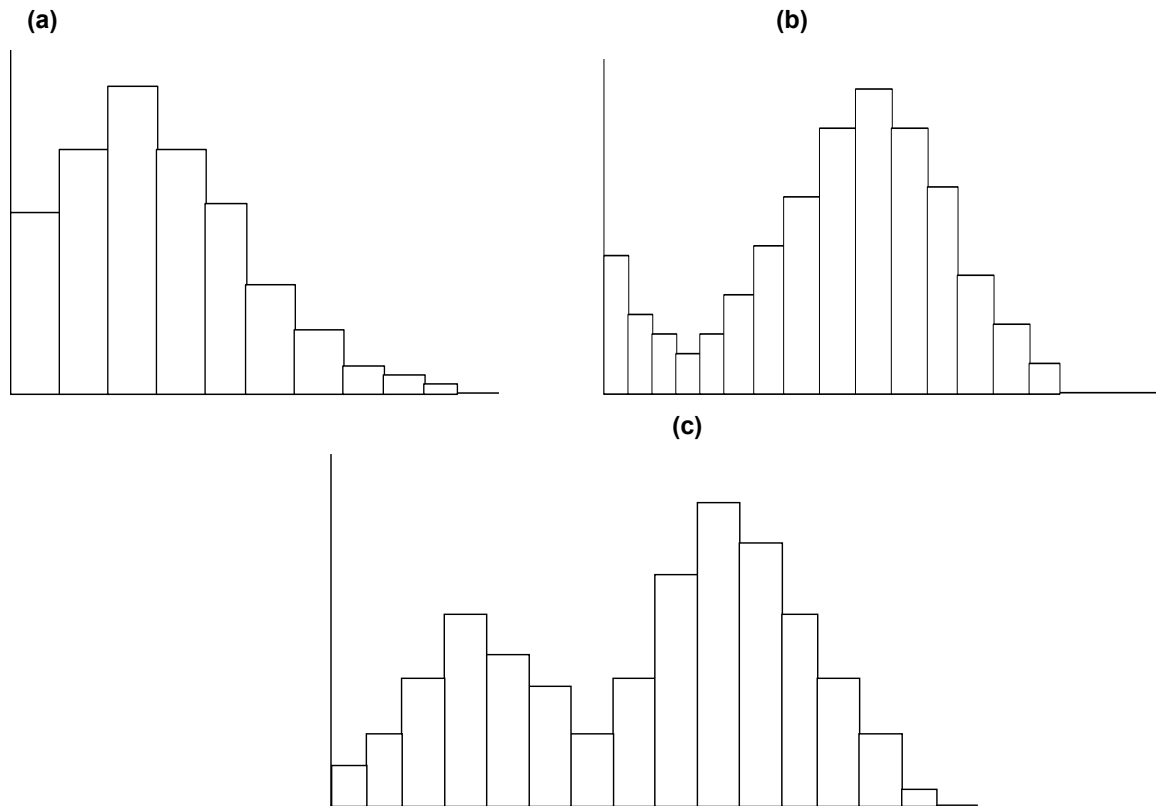


Figura 1.9: Tipos de histogramas.

Una distribución de frecuencias dice también cual es la variación de los datos. Contesta a la pregunta ¿cuánta variación existe? Si una distribución es pequeña y la otra grande, estando ambas dibujadas en la misma escala horizontal, ello significa que la primera tiene menos variación que la segunda, Figura 1.10.



Figura 1.10: Distribuciones de variabilidad menor y mayor respectivamente.

Una característica adicional de una distribución por frecuencias es la simetría de sus variaciones. La distribución puede ser simétrica o estar cargada a uno de los lados. Si se trata del segundo caso, se dice que la distribución es *asimétrica*. Una distribución asimétrica tiene un solo pico, pero está situado en el extremo superior o inferior de ella. Cuando la cola más larga de la curva se encuentra a la derecha, se dice que es asimétrica a la derecha o asimétrica positiva. Cuando la cola más larga de la curva está a la izquierda, la curva es asimétrica a la izquierda o asimétrica negativa.

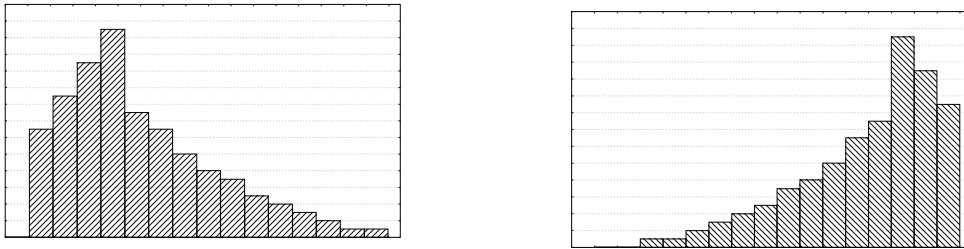


Figura 1.11: Distribución positivamente asimétrica y negativamente asimétrica respectivamente.

El modelo positivamente asimétrico es muy común en los datos económicos y comerciales; por ejemplo, la distribución del ingreso familiar; o también el número de tiendas que venden al por menor es mayor que las tiendas que venden al por mayor, sin embargo estas últimas registran mayores cantidades de ventas.

Un histograma condensa la información de modo que sea fácil su comprensión visual. Los histogramas pueden exhibir cierta sensibilidad con respecto al número de clases y ancho. Para conjuntos pequeños de datos, los histogramas pueden cambiar claramente de apariencia si el número de clases o el ancho de estas cambia. Los histogramas son más estables si el conjunto de datos es grande, de preferencia de 75 a 100 o más datos. Sin embargo, frecuentemente es necesario sintetizar la información de modo 'numérico' para caracterizar los datos de manera estadística. Se requiere entonces de medidas de centralización y de dispersión.

Ejemplo de aplicación

Con el objeto de caracterizar el comportamiento metalúrgico del material en determinadas operaciones individuales, es necesario cuantificar la cantidad de partículas de un tamaño dado con respecto al total de partículas en la muestra, lo que se denomina Distribución Granulométrica o Granulometría de Material.

Un sistema de partículas con un amplio rango de tamaños solo puede describirse en detalle mediante el uso de distribuciones estadísticas. De estas distribuciones puede obtenerse un tamaño característico que represente al sistema particulado.

La metodología normalmente utilizada en diversas plantas metalúrgicas consiste en hacer pasar la muestra a ensayar por una serie de tamices ordenados secuencialmente de mayor a menor aberturas, en forma descendente, haciendo pasar posteriormente el bajo tamaño del último tamiz, por ejemplo el material –325 mallas Tyler, con el objeto de tener un análisis granulométrico completo de la muestra. Así por ejemplo, se ha realizado el siguiente análisis granulométrico:

Malla	Abertura	Peso Retenido
20	850	18.0
28	600	30.0
35	425	41.5
48	300	43.5
65	212	40.5
100	150	42.0
150	106	38.0
200	75	40.5
325	45	44.0
400	38	37.5
-400		124.5

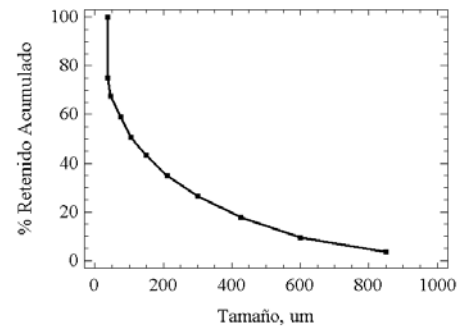
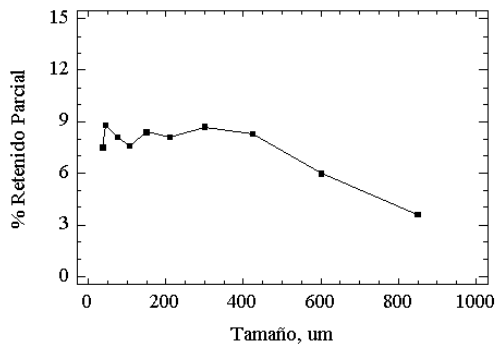
Con ello:

- Construir la Tabla de Análisis Granulométrico.

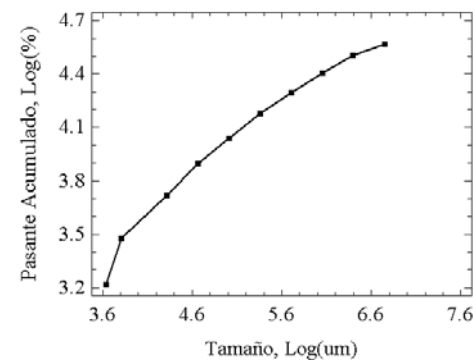
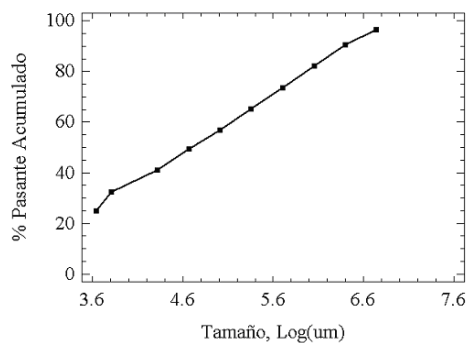
- b) Construir gráficos en escala aritmética de: % retenido parcial vs. tamaño, y % retenido acumulado vs tamaño.
- c) Construir el gráfico de % pasante acumulado vs. $\log(\text{tamaño})$
- d) Construir el gráfico de $\log(\% \text{ pasante acumulado})$ vs $\log(\text{tamaño})$

Tabla de análisis granulométrico

Malla	Abertura μm	Peso Retenido; g.	% Reten. Parcial	% Reten acum..	% Pasante acum.
20	850	18.0	3.6	3.6	96.4
28	600	30.0	6.0	9.6	90.4
35	425	41.5	8.3	17.9	82.1
48	300	43.5	8.7	26.6	73.4
65	212	40.5	8.1	34.7	65.3
100	150	42.0	8.4	43.1	56.9
150	106	38.0	7.6	50.7	49.3
200	75	40.5	8.1	58.8	41.2
325	45	44.0	8.8	67.6	32.4
400	38	37.5	7.5	75.1	24.9
-400	-	124.5	24.9	100.0	-



Generalmente los puntos experimentales se agrupan en el comienzo de la abcisa, cuando se utiliza un gráfico con escala aritmética para el tamaño; por eso es conveniente usar una escala logarítmica para el tamaño de partículas y de esta forma el espaciado de las mallas se hace constante.



De los cálculos y gráficos desarrollados es posible determinar tamaños característicos de partículas, los cuales son muy útiles en los procesos metalúrgicos; estos pueden ser el tamaño de 80% o d_{80} , o el tamaño de 50% ó d_{50} .

EJERCICIOS GRUPO 1

1. La siguiente tabla muestra la superficie en millones de millas cuadradas de los océanos del mundo. Represente los datos utilizando un diagrama de barras y un diagrama circular.

Océano	Superficie Millones de millas ²
Pacífico	70.8
Atlántico	41.2
Indico	28.5
Antártico	7.6
Artico	4.8

2. Los siguientes datos muestran el número de toneladas de materiales transportados por vía acuática a un país en 1972:

PRODUCTO	CANTIDAD, MILES DE TON
Petróleo y derivados	63,150
Carbón	20,100
Acero	13,800
Arena y grava	11,700
Fibras	7,650
Madera	4,650
Químicos	8,700
Conchas marinas	1,650
Otros	18,600

- a) Construya una tabla de porcentajes para los datos anteriores y calcule la porción a asignar a cada categoría.
- b) Construya un diagrama de barras y una gráfica de pastel para los datos anteriores.
3. La distribución en la siguiente Tabla muestra el número de accionistas de una empresa, clasificados por el número de acciones que poseen:

ACCIONES	NÚMERO DE ACCIONISTAS
menos de 1000	486
1000 - 1999	372
2000 - 2999	210
3000 - 3999	117
4000 - 4999	43
5000 - 5999	27
6000 - ó más	13

- a) Construya una tabla de porcentajes para los datos anteriores y calcule la porción a asignar a cada categoría.
- b) Construya un diagrama de barras y una gráfica de pastel para los datos anteriores.
4. Aproximadamente dos tercios de los accidentes automovilísticos se deben a una inadecuada forma de conducir. Construya un diagrama de Pareto para los siguientes datos: (a) Dar vuelta en sentido incorrecto: 3.6%; (b) Velocidad de manejo excesiva para las condiciones imperantes: 28.1 %; (c) Distancia insuficiente entre dos vehículos: 8.1 %; (d) Violaciones a la indicaciones de preferencia: 30.1 %; (e) Ocupar carriles de la izquierda o permanecer al centro: 3.3 %; (f) Rebasar mal : 3.2 %; (g) Otras causas: 23.6 %.

5. Un equipo de proyecto estudia el costo que implicaría las descomposturas en una línea para embotellar bebidas refrescantes. El análisis en miles de dólares correspondientes a tres meses son: (a) Regulador de presión: 30; (b) Ajuste de gusano de alimentación: 15; (c) Atoramiento de la cabeza de cobre: 6; (d) Pérdidas de enfriamiento: 52; (e) Reemplazo de válvulas: 8; (f) Otras: 5. Construya un diagrama de Pareto.
6. Los siguientes datos corresponden a la resistencia a la tensión (100 psi) y dureza (Rockwell E) de aluminio moldeado en matriz. Dibuje un diagrama de dispersión y defina la relación respectiva:
(0.5; 1.62), (1.5; 0.75), (2.0; 0.62), (3.0; 0.46), (2.5; 0.52), (1.0; 1.00), (0.8; 1.35), (1.2; 0.89), (2.8; 0.48), (3.2; 0.43), (1.8; 0.71), (0.3; 1.80). Dibuje un diagrama de dispersión y defina la relación correspondiente.

7. Los datos de la cantidad de agua utilizada, (en pulgadas) y el rendimiento de alfalfa en toneladas por acre, son las siguientes:

Agua	12	18	24	30	36	42	48	60
Rendimiento	5.3	5.7	6.3	7.2	8.2	8.7	8.4	8.2

Dibuje un diagrama de dispersión y analice los datos.

8. Los datos siguientes representan el número de ciclos transcurridos hasta que se presenta una falla en una prueba de piezas de aluminio sujetas a un esfuerzo alternante repetido de 21,000 psi a 18 ciclos por segundo:

1115	1567	1223	1782	1055	798	1016	2100	910	1501
1310	1883	375	1522	1764	1020	1102	1594	1730	1238
1540	1203	2265	1792	1330	865	1605	2023	1102	990
1502	1270	1910	1000	1608	2130	706	1315	1578	1468
1258	1015	1018	1820	1535	1421	2215	1269	758	1512
1315	845	1452	1940	1781	1109	785	1260	1416	1750
1085	1674	1890	1120	1750	1481	885	1888	1560	1642

Construya una tabla de frecuencias y un histograma para los datos anotados.

9. Los siguientes datos corresponden al octanaje de varias mezclas de gasolina:

88.5	87.7	83.4	86.7	87.5	91.5	88.6	100.3	96.5	93.3
94.7	91.1	91.0	94.2	87.8	89.9	88.3	87.6	84.3	86.7
84.3	86.7	88.2	90.8	88.3	98.8	94.2	92.7	93.2	91.0
90.1	93.4	88.5	90.1	89.2	88.3	85.3	87.9	88.6	90.9
89.0	96.1	93.3	91.8	92.3	90.4	90.1	93.0	88.7	89.9
89.8	89.6	87.4	88.4	88.9	91.2	89.3	94.4	92.7	91.8
91.6	90.4	91.1	92.6	89.8	90.6	91.1	90.4	89.3	89.7
90.3	91.6	90.5	93.7	92.7	92.2	92.2	91.2	91.0	92.2
90.0	90.7								

Construya una tabla de distribución de frecuencias y un histograma utilizando 8 clases y 16 clases.

Compare los histogramas obtenidos en cada caso, ¿los dos histogramas presentan información similar?.

10. Los siguientes datos son mediciones de viscosidad de un producto químico tomadas cada hora (de arriba abajo y de izquierda a derecha):

47.9	48.8	48.6	43.2	43.0
47.9	48.1	48.0	43.0	42.8
48.6	48.3	47.9	43.5	43.1
48.0	47.2	48.3	43.1	43.2
48.4	48.9	48.5	43.0	43.6
48.1	48.6	48.1	42.9	43.2

48.0	48.0	48.0	43.6	43.5
48.6	47.5	48.3	43.3	43.0

Construya una gráfica de series de tiempo para estos datos.

11. En una prueba de laboratorio se mide la fuerza de tirantez de un conector. Los siguientes son los datos obtenidos (y registrados en orden) para 40 muestras bajo prueba (de arriba abajo y de izquierda a derecha):

241	220	249	209
258	194	251	212
237	245	238	185
210	209	210	187
194	201	198	218
225	195	199	190
248	255	183	175
203	245	213	178
195	235	236	175
249	220	245	190

Elabore una gráfica de series de tiempo.

12. La siguiente tabla muestra las velocidades orbitales de los planetas de nuestro sistema solar. Representar estos datos con el gráfico apropiado.

Planeta	Velocidad, Millas/seg.
Mercurio	29.7
Venus	21.8
Tierra	18.5
Marte	15.0
Júpiter	8.1
Saturno	6.0
Urano	4.2
Neptuno	3.4
Plutón	3.0

13. Los siguientes son los números de imperfecciones observadas en muestras tomadas de rollos de tela:

2	0	4	4	1	4	0	3	2	0
0	1	1	1	0	1	2	4	1	1
1	5	2	2	5	3	4	0	4	0
0	0	3	0	1	4	2	1	2	0
3	1	3	4	2	0	5	6	3	2

- Agrupar los datos en una Tabla de Frecuencias.
 - Construir una ojiva para los datos.
14. Remítase a la bibliografía y describa, con un ejemplo, los procedimientos de elaboración de los diagramas de tallo y hoja.