

Datos y Estadísticas

Profesor: alberto alvaradejo

Estadística

- La **estadística** se define como el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos
 - Especialmente en los negocios y en la economía

Aplicaciones en los negocios y en la economía

- Contabilidad
 - Las empresas de contabilidad al realizar auditorías para sus clientes emplean procedimientos de muestreo estadístico.
- Finanzas
 - Los analistas financieros emplean una diversidad de información estadística como guía para sus recomendaciones de inversión.
- Marketing
 - Escáneres electrónicos en las cajas de los comercios recogen datos para diversas aplicaciones en la investigación de mercado.
- Producción
 - La importancia que se le da actualmente a la calidad hace del control de calidad una aplicación importante de la estadística a la producción.
- Economía
 - Los economistas suelen hacer pronósticos acerca del futuro de la economía o sobre algunos aspectos de la misma.

Datos

- Datos son hechos/informaciones y cifras que se recogen, analizan y resumen para su presentación e interpretación.
- A todos los datos reunidos para un determinado estudio se les llama conjunto de datos para el estudio.

Empresa	Bolsa de valores	Denominación abreviada Ticker	Posición en <i>BusinessWeek</i>	Precio por acción (\$)	Ganancia por acción (\$)
Abbott Laboratories	N	ABT	90	46	2.02
Altria Group	N	MO	148	66	4.57
Apollo Group	NQ	APOL	174	74	0.90
Bank of New York	N	BK	305	30	1.85
Bristol-Myers Squibb	N	BMY	346	26	1.21
Cincinnati Financial	NQ	CINF	161	45	2.73
Comcast	NQ	CMCSA	296	32	0.43
Deere	N	DE	36	71	5.77
eBay	NQ	EBAY	19	43	0.57
Federated Dept. Stores	N	FD	353	56	3.86
Hasbro	N	HAS	373	21	0.96
IBM	N	IBM	216	93	4.94
International Paper	N	IP	370	37	0.98
Knight-Ridder	N	KRI	397	66	4.13
Manor Care	N	HCR	285	34	1.90
Medtronic	N	MDT	53	52	1.79
National Semiconductor	N	NSM	155	20	1.03
Novellus Systems	NQ	NVLS	386	30	1.06
Pitney Bowes	N	PBI	339	46	2.05
Pulte Homes	N	PHM	12	78	7.67
SBC Communications	N	SBC	371	24	1.52
St. Paul Travelers	N	STA	264	38	1.53
Teradyne	N	TER	412	15	0.84
UnitedHealth Group	N	UNH	5	91	3.94
Wells Fargo	N	WFC	159	59	4.09

Fuente: *Business Week* (4 de abril de 2005).

Elementos, variables y observaciones

- **Elementos** son las entidades de las que se obtienen los datos.
- **Variable** es una característica de los elementos que es de interés.
- **Observación:** conjunto de mediciones obtenidas para un determinado elemento.
- Datos: los valores encontrados para cada variable en cada uno de los elementos.

Escalas de medición

- La escala de medición determina la cantidad de información contenida en el dato e indica la manera más apropiada de resumir y de analizar estadísticamente los datos.
 - **Escala nominal:** cuando el dato de una variable es una etiqueta o un nombre que identifica un atributo de un elemento.
 - **Escala ordinal:** los datos muestran las propiedades de los datos nominales y además tiene sentido el orden o jerarquía de los datos (etiquetas excelente, bueno o malo).
 - **Escala de intervalo:** los datos tienen las características de los datos ordinales y el intervalo entre valores se expresa en términos de una unidad de medición fija. Los datos de intervalo siempre son numéricos.

Datos cualitativos y cuantitativos

- **Datos cualitativos:** comprenden etiquetas o nombres que se usan para identificar un atributo de cada elemento.
 - emplean la escala nominal o la ordinal y pueden ser numéricos o no
- **Datos cuantitativos** requieren valores numéricos que indiquen cuánto o cuántos.
 - Los datos cuantitativos se obtienen usando las escalas de medición de intervalo o de razón.

Fuentes de datos

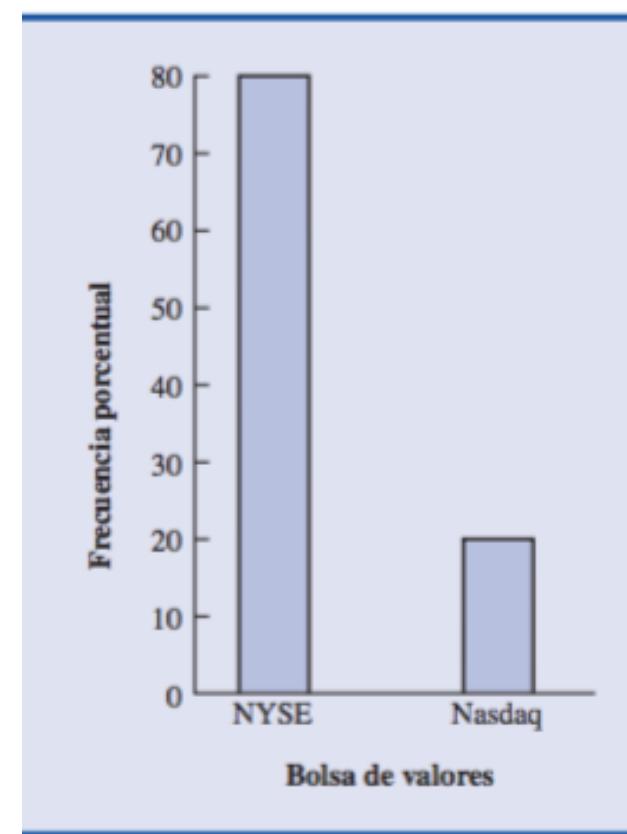
- **Fuentes existentes**
 - En algunos casos los datos que se necesitan para una determinada aplicación ya existen.
- **Estudios estadísticos**
 - Los datos suelen conseguirse realizando un estudio estadístico. Dichos estudios se clasifican como *experimentales* u *observacionales*.

Estadística Descriptiva

- Resúmenes de datos, que pueden ser tabulados, graficados o numerados.

Bolsa de valores	Frecuencia	Frecuencia porcentual
Bolsa de Nueva York	20	80
Mercado Nacional Nasdaq	5	20
Totales	25	100

La estadística es el arte y la ciencia de recolectar, analizar, presentar e interpretar datos.



Inferencia estadística

- **POBLACIÓN**
 - La población es el conjunto de todos los elementos de interés en un estudio determinado.
- **MUESTRA**
 - La muestra es un subconjunto de la población.
- **inferencia estadística**
 - Emplea datos de una muestra para hacer estimaciones y probar hipótesis acerca de las características de una población mediante un proceso.

Glosario

- **Estadística** El arte y la ciencia de recolectar, analizar, presentar e interpretar datos.
- Datos** Los hechos y las cifras que se recolectan, analizan y resumen para su presentación e interpretación.
- Conjunto de datos** Todos los datos recolectados en un estudio determinado.
- Elementos** Entidades sobre las que se recolectan los datos.
- Variable** Una característica que interesa de un elemento.
- Observación** El conjunto de mediciones obtenidas de un elemento determinado.
- Escala nominal** Escala de medición de una variable cuando los datos son etiquetas o nombres que se emplean para identificar un atributo de un elemento. Los datos nominales pueden ser no numéricos o numéricos.
- Escala ordinal** Escala de medición de una variable cuando los datos presentan las propiedades de los datos nominales y el orden o jerarquía de los datos tiene sentido. Los datos ordinales pueden ser no numéricos o numéricos.
- Escala de intervalo** Escala de medición de una variable cuando los datos presentan las propiedades de los datos ordinales y los intervalos entre valores se expresan en términos de una unidad o medida fija. Los datos de intervalo siempre son numéricos.

- **Datos cualitativos** Etiquetas o nombres utilizados para identificar un atributo de cada elemento. Los datos cualitativos usan las escalas de medición nominal y ordinal y pueden ser no numéricos o numéricos.
- Datos cuantitativos** Valores numéricos que indican cuánto o cuántos de algo. Los datos cuantitativos se obtienen mediante la escala de intervalo o de razón.
- Variable cualitativa** Una variable con datos cualitativos.
- Variable cuantitativa** Una variable con datos cuantitativos.
- Datos de sección transversal** Datos recolectados en el mismo o aproximadamente en el mismo momento.
- Datos de series de tiempo** Datos recolectados a lo largo de varios períodos de tiempo. **Estadística descriptiva** Resúmenes tabulares, gráficos o numéricos de datos.
- Población** Conjunto de todos los elementos que interesan en un estudio determinado.
- Muestra** Un subconjunto de la población.
- Censo** Un estudio para recolectar los datos de toda la población.
- Encuesta muestral** Un estudio para recolectar los datos de una muestra.
- Inferencia estadística** El proceso de emplear los datos obtenidos de una muestra para hacer estimaciones o probar hipótesis acerca de las características de la población.

Tablas y gráficos

- Distribución de frecuencia: es un resumen tabulado de datos que muestra el número (frecuencia absoluta) de elementos en cada una de las diferentes clases.

Coke Classic	Sprite	Pepsi
Diet Coke	Coke Classic	Coke Classic
Pepsi	Diet Coke	Coke Classic
Diet Coke	Coke Classic	Coke Classic
Coke Classic	Diet Coke	Pepsi
Coke Classic	Coke Classic	Dr. Pepper
Dr. Pepper	Sprite	Coke Classic
Diet Coke	Pepsi	Diet Coke
Pepsi	Coke Classic	Pepsi
Pepsi	Coke Classic	Pepsi
Coke Classic	Coke Classic	Pepsi
Dr. Pepper	Pepsi	Pepsi
Sprite	Coke Classic	Coke Classic
Coke Classic	Sprite	Dr. Pepper
Diet Coke	Dr. Pepper	Pepsi
Coke Classic	Pepsi	Sprite
Coke Classic	Diet Coke	

DISTRIBUCIÓN DE FRECUENCIA DE LAS VENTAS DE REFRESCO

Refresco	Frecuencia
Coke Classic	19
Diet Coke	8
Dr. Pepper	5
Pepsi	13
Sprite	5
Total	50

Distribuciones de frecuencia relativa y de frecuencia porcentual

En un conjunto de datos, en el que hay n observaciones, la frecuencia relativa de cada clase se determina como sigue:

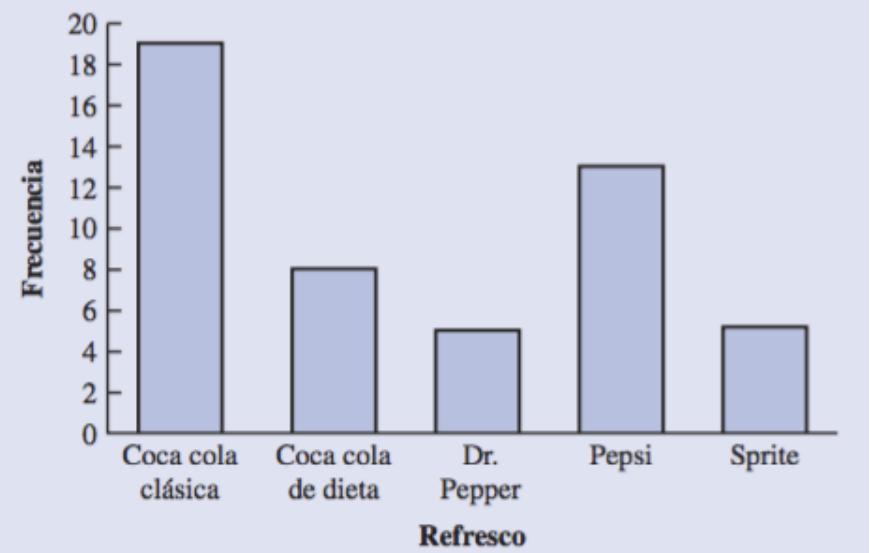
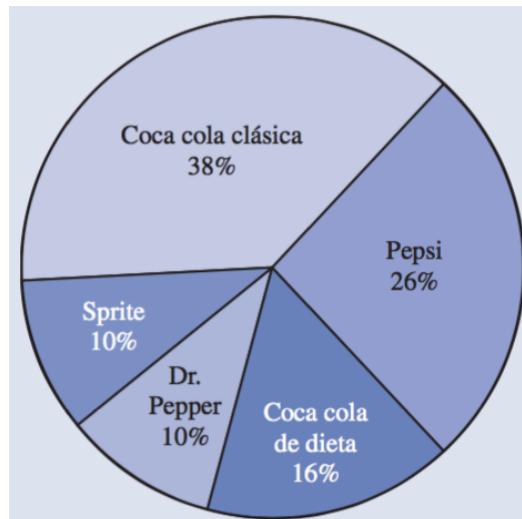
$$\text{Frecuencia relativa} = \frac{\text{Frecuencia absoluta}}{n}$$

La *frecuencia porcentual* de una clase es la frecuencia relativa multiplicada por 100.

Gráficos de barra y de pastel

Un **gráfico de barras** representa los datos cualitativos de una distribución de frecuencia absoluta, de frecuencia relativa o de frecuencia porcentual.

Refresco	Frecuencia relativa	Frecuencia porcentual
Coke Classic	0.38	38
Diet Coke	0.16	16
Dr. Pepper	0.10	10
Pepsi	0.26	26
Sprite	0.10	10
Total	1.00	100



Resumen de datos cuantitativos

Los siguientes datos se recopilaron con el fin de determinar la edad de 50 estudiantes de IIº Medio del Colegio Especial Montaña

15	16	17	18	19	15	20	18	20	17
15	16	15	19	25	15	30	42	15	20
15	16	19	20	16	15	16	20	20	42
16	17	17	20	19	18	19	60	42	22
19	19	25	17	25	31	20	25	30	42

Clase	Intervalo	Marca de Clase	Frecuencia absoluta	Frec. abs. acumulada	Frecuencia relativa	Frec. rel. acumulada
Nro.	Nro.	x_i	f_i	F_i	h_i	H_i
1	15 - 22	18,5	37	37	0,74	0,74
2	22 - 29	25,5	5	42	0,10	0,84
3	29 - 36	32,5	3	45	0,06	0,90
4	36 - 43	39,5	4	49	0,08	0,98
5	43 - 50	46,5	0	49	0,00	0,98
6	50 - 57	53,5	0	49	0,00	100
7	57 - 64	60,5	1	50	0,02	
	Total		50		1,00	

Gráficos

- **Histograma**

Se usa para representar las frecuencias de una variable cuantitativa continua. En uno de los ejes se posicionan las clases de la variable continua (los intervalos o las marcas de clase que son los puntos medios de cada intervalo) y en el otro eje las frecuencias. No existe separación entre las barras.

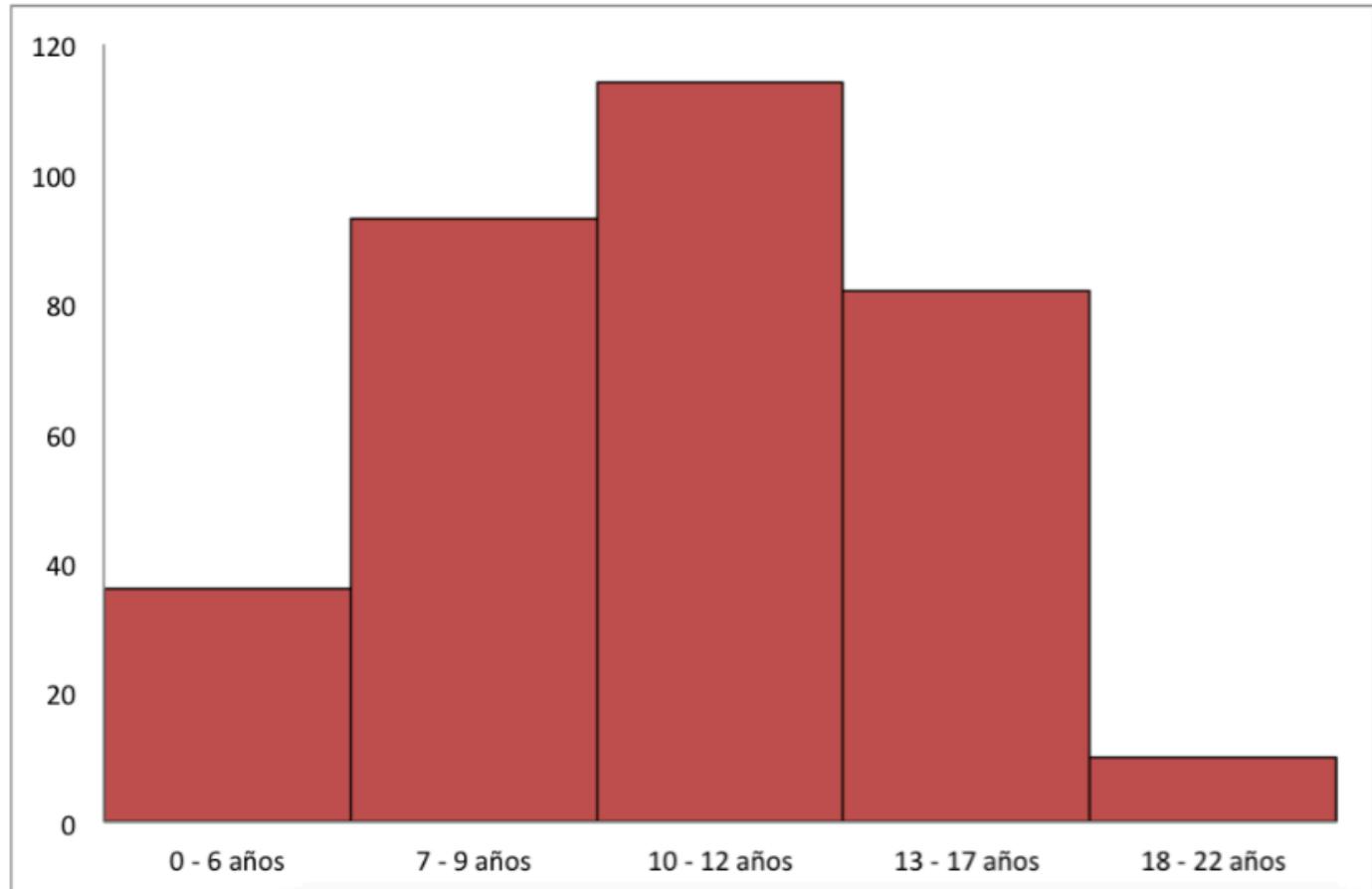


Gráfico de líneas

- Es una representación gráfica en un eje cartesiano de la relación que existe entre dos variables reflejando con claridad los cambios producidos.

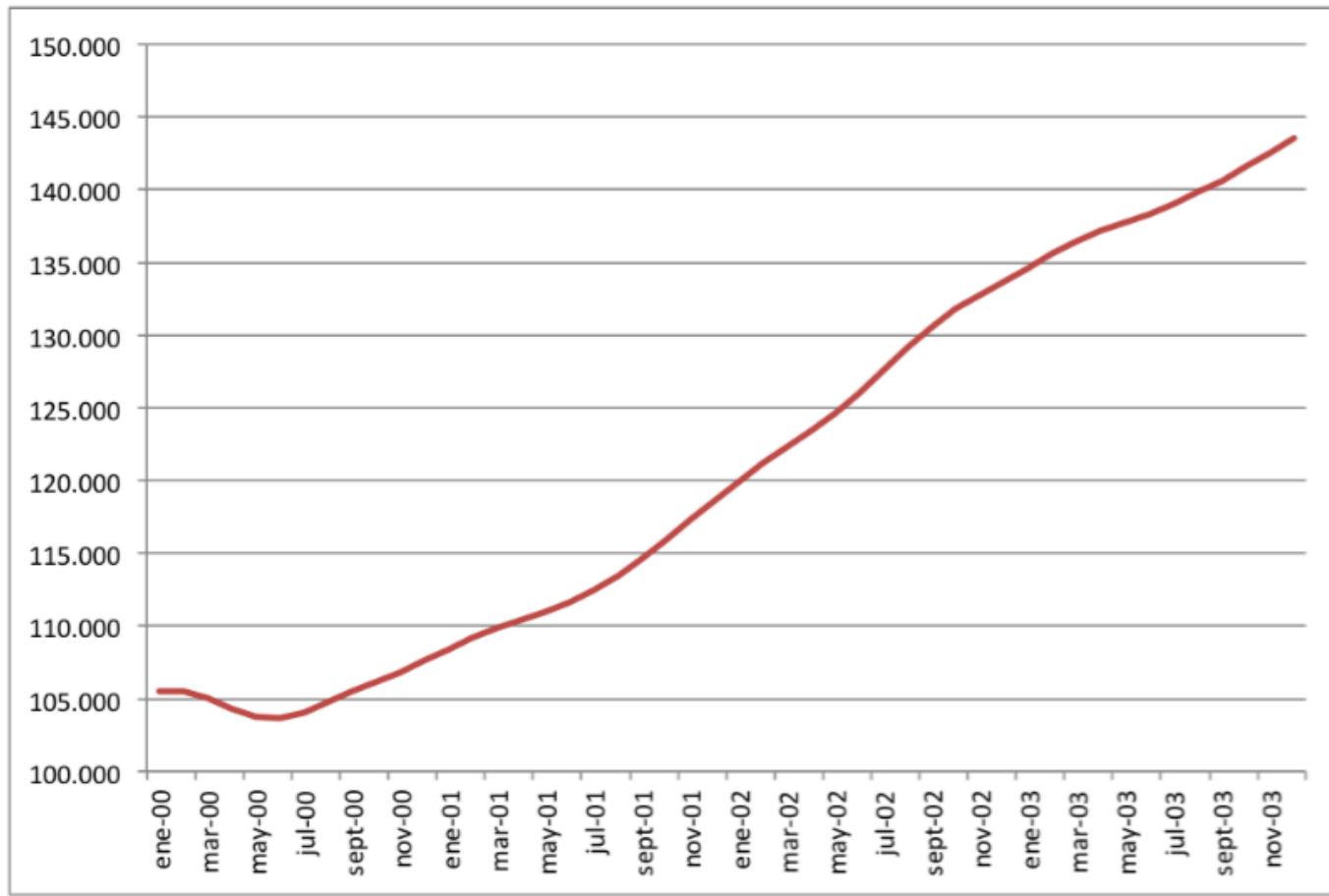


Diagrama de tallo y hojas

Las técnicas del **análisis exploratorio de datos** emplean aritmética sencilla y graficas fáciles de dibujar útiles para resumir datos. La técnica conocida como **diagrama de tallo y hojas** muestra en forma simultánea el orden jerárquico y la forma de un conjunto de datos.

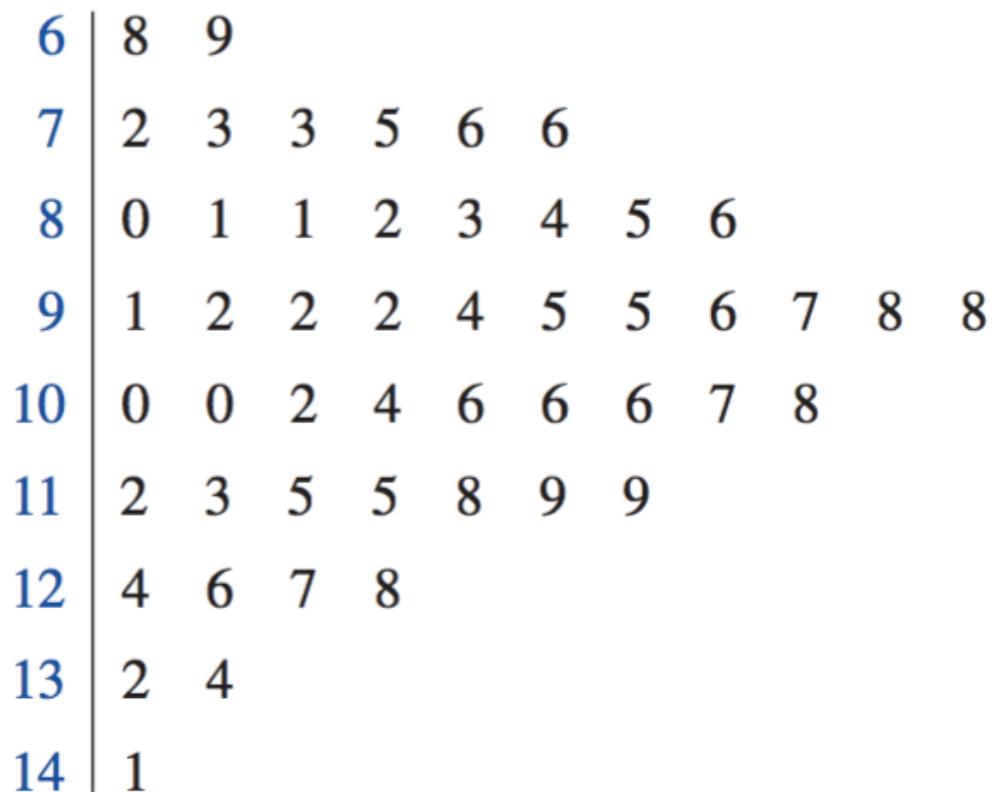
Ejemplo

NÚMERO DE PREGUNTAS CONTESTADAS CORRECTAMENTE EN UN EXAMEN

112	72	69	97	107
73	92	76	86	73
126	128	118	127	124
82	104	132	134	83
92	108	96	100	92
115	76	91	102	81
95	141	81	80	106
84	119	113	98	75
68	98	115	106	95
100	85	94	106	119

Para elaborar un diagrama de tallo y hoja se acomodan los primeros dígitos de cada uno de los datos a la izquierda de una línea vertical. A la derecha de la línea vertical se anota el último dígito de cada dato.

Una vez organizados los datos de esta manera, se ordenan los datos de cada renglón de menor a mayor. Entonces se obtiene el diagrama de tallo y hojas



Al rotar el gráfico sobre su costado en contra de las manecillas del reloj se obtiene una imagen de los datos que es parecida a un histograma y en el que las clases son 60–69, 70–79, 80–89, etcétera.

Aunque el diagrama de tallo y hojas parece proporcionar la misma información que un histograma, tiene dos ventajas fundamentales.

1. El diagrama de tallo y hojas es más fácil de construir a mano.
2. En cada intervalo de clase proporciona más información que un histograma debido a que el tallo y la hoja proporcionan el dato.

