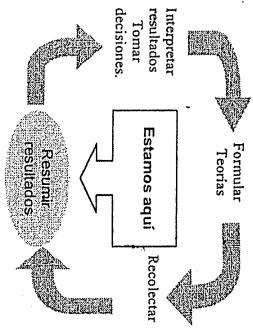


## CAPÍTULO 3

### 3.1. INTRODUCCIÓN

Tenemos una teoría que deseamos ensayar. A tal fin recolectamos datos, los miramos y resumimos los resultados.

En capítulos anteriores aprendimos cómo tomar nuestros datos, nuestras mediciones, ahora las resumiremos de una manera útil.



Una de las maneras más útiles de mirar o presentar los datos es a través de un cuadro, un gráfico o un diagrama. Luego mejoraremos la presentación **gráfica** con el cálculo de algunas **medidas resúmenes** en el capítulo 4, por ejemplo una medida numérica de resumen es la media aritmética o promedio.

Más adelante, tomaremos la representación gráfica de nuestros datos los cuales están basados en una muestra y trataremos de encontrar un **modelo** adecuado para representar los resultados de la población.

Comenzaremos este capítulo destacando que hay distintos tipos de datos; podríamos medir por ejemplo el género de un sujeto: masculino y altura por ejemplo: 1,80 cm. A veces una medida puede ser un rótulo, otras veces un valor numérico. Diferentes tipos de datos y escalas de medición conducen a diferentes tipos de presentaciones gráficas. Aprendremos a hacerlos y también incluiremos algunos ejemplos de los erro-



**conjunto (desordenado) de datos 2**  
**contiene la LONGITUD EN CM DE 20 PARTES**  
**CONSECUTIVAS SACADAS DE UNA LÍNEA DE MONTAJE**

Pieza #1	20.011	2	3	4	5	6	7	8	9	10
Pieza #20	20.007	12	13	14	15	16	17	18	19	20



**conjunto de datos 3**  
 Contiene varias mediciones de 20 individuos. Estos individuos fueron parte de un estudio médico, evaluando la bondad de una nueva droga para reducir la presión sanguínea y la relación entre el número de tabletas tomadas (droga) y la reducción de presión sanguínea. Las características (variables) que pudieron influir en la presión sanguínea son: género, edad, y diferencia de presión sanguínea antes y después de la ingesta de tabletas.

Sujeto número	Género	Edad	Presión sanguínea diastólica	
			Al comienzo de la prueba	Al final del estudio
1001	M	45	2	100,2
1002	M	41	1	98,5
1003	F	51	2	100,8
1004	F	46	2	101,1
1005	F	47	3	100,0
1006	M	42	2	99,0
1007	M	43	4	100,7
1008	F	50	2	100,3
1009	M	39	1	100,6
1010	M	32	1	99,9
1011	M	41	2	101,0
1012	M	44	2	100,9
1013	F	47	2	97,4
1014	F	49	3	98,8
1015	M	45	3	100,9
1016	F	42	1	101,1
1017	M	41	2	100,7
1018	F	40	1	97,8
1019	M	45	2	100,0
1020	M	37	3	101,5

¿Qué nota en estos tres conjuntos de datos? Las características

res comunes que se cometen en la presentación gráfica, los cuales pueden conducir a una mala interpretación de los resultados. El objetivo de este capítulo es mostrar la utilidad de un gráfico o diagrama cuando el mismo está bien hecho.



**conjunto de datos 1**  
**PLAN FINANCIERO A LARGO PLAZO**

En los últimos años, muchos individuos han invertido su dinero en una gran variedad de fondos mutuos. El motivo de estas inversiones incluye objetivos a corto y largo plazo.

Suponga que una consultora que evalúa fondos comunes generales de inversión nacional, lo contrata para que realice recomendaciones de compras a potenciales inversionistas.

¿Cómo podría detectar fondos para inversiones posibles como parte de un plan financiero a largo plazo?

Recuerde que el estudio de las medidas de desempeño financiero de fondos comunes está basado en las siguientes características:

a) objetivo de los fondos (fondos de crecimiento contra fondos mixtos),

b) estructura de los cargos (pagos de cargos; sin cargo);

c) monto de capitalización (grande, mediano y pequeño).

Una forma de comparar el desempeño financiero, medido en los rendimientos al año, se basa en los objetivos del fondo (crecimiento contra mixto).

A continuación se presenta un arreglo ordenado de los rendimientos totales que alcanzaron al año, los 59 fondos de crecimiento, en base a los datos obtenidos de una muestra de 194 fondos comunes de inversión nacional, [135 corresponden a fondos de inversión mixtos]

Rendimientos totales al año que alcanzaron los 59 Fondos de crecimiento										
20,4	23,8	25,6	26,2	27,6	27,7	28,3	28,6	28,8	28,9	
28,9	29,3	29,3	29,5	29,9	30,1	31,5	31,6	31,6	31,8	
31,9	32,1	32,3	32,3	32,4	32,8	32,9	33,0	33,3		
33,4	33,7	33,8	34,0	34,0	34,3	34,7	34,7	34,8		
38,2	39,0	39,4	40,7	41,1	42,8	42,9	43,3	43,4	43,5	
43,6	43,7	44,6	44,7	45,4	45,7	46,6	48,0	48,6		

FUENTE: archivo MUTUAL, B.L.K

sobre los individuos o unidades son variables, y en algunos casos medidas repetidas sobre el mismo individuo son variables.

## ¡Los datos varían!!

Como los datos varían, las conclusiones basadas sobre los mismos son inciertas.

La estadística nos ayuda a producir datos útiles los cuales pueden ser analizados de manera tal que podamos extraer conclusiones con un pequeño grado de incertidumbre. El razonamiento estadístico nos permite cuantificar cuán inciertas son nuestras conclusiones.

### Definiciones

- Una **UNIDAD** es el elemento u objeto que observamos. Cuando el objeto es una persona, la designamos **sujeto**.
- Una **OBSERVACIÓN** es la información o característica registrada para cada unidad.
- Una **característica** que varía de unidad a unidad se denomina **variable**.
- Un **conjunto de observaciones** sobre una o más **variables** se denomina **conjunto de datos**.

El conjunto de datos 2, consiste de solo una variable: **longitud**, la cual fue medida en 20 unidades.

En el conjunto de datos 1, se presenta solo una de las 7 variables consideradas de interés en la muestra de tamaño 194, como parte de un plan financiero a largo plazo

¿Cómo clasificaremos a estas variables?

### 3.2.1. Tipos de variables

Las variables pueden ser **cualitativas** o **cuantitativas**.

Las variables **cualitativas** asumen valores que no son necesariamente numéricos pero pueden ser categorizados.

La variable género tiene dos posibles valores: **femenino**, **masculino**.

Estos dos valores pueden ser codificados como:

1. femenino
2. masculino

Necesitamos recordar el esquema de codificación. Tales valores carecen de significado si se los suma, resta o promedia.

**Tipo de trabajo, Estado civil, Profesión, Situación económica,** (cargos de los activos del fondo, cargos diferidos, cargos por venta directa, cargos múltiples, fondos sin cargo): son ejemplos de **variables cualitativas**.

Las observaciones hechas sobre variables cualitativas a menudo se denominan **datos categóricos**.

Las variables cuantitativas toman valores numéricos, para los cuales sumar, restar, o promediar dichos valores tienen significado.

Ejemplo de variables cuantitativas: peso, altura, número de niños. A veces, una variable cuantitativa puede categorizarse; por ejemplo la variable ingreso, registrada en pesos es cuantitativa, sin embargo podríamos registrar la variable **ingreso** como "menos de \$300"; "de \$300 a \$600" y "\$601 y más" y tratarla como variable cualitativa. Idem con variable **Monto de capitalización**.

Existen dos tipos principales de variables **cuantitativas**

**Discreta**  
**Continua**

Una variable cuantitativa se dice que es discreta si su conjunto de resultados posibles es finito o infinito numerable (o contable).

El número de pisos de los edificios de una determinada zona donde las ordenanzas sobre construcción restringe el número de pisos a 20 será una variable discreta.

El número de valores posibles es finito: 1, 2, 3, ..., 20.

El número de pisos no podría ser 2, 1 o 7/6.

El número de aciertos en su último "quiz" el cual tenía 5 preguntas es discreto.

Los valores posibles son 0, 1, 2, 3, 4, 5. El número de llamadas telefónicas hechas por una vendedora hasta antes de concretar una venta es también discreta, con un número posible de valores: 1, 2, 3, ...

Una variable cuantitativa se dice que es continua si su conjunto de valores posibles es un intervalo o conjunto de intervalos de números reales.

La altura de un recién nacido, puede ser cualquier valor entre una altura mínima,  $h_{\min}$ , y una altura máxima posible,  $h_{\max}$ .

 **PARA PENSAR!!!**

**Variable discreta:** sólo puede tomar un número finito o infinito numerable de posibles valores.

El número de respuestas correctas sobre 5 preguntas en el test de elección múltiple.

- •
- 0 1 2 3 4 5

**Variable continua:** puede tomar cualquier valor en un intervalo o conjuntos de intervalos.

La cantidad de agua vertida en una probeta de 50 ml de capacidad.

0 \_\_\_\_\_ 50



A veces una variable puede ser tratada como discreta o continua. Consideremos la proporción de mujeres en una población. Si la población consiste sólo de 10 personas, luego las proporciones son: 0; 0,1; 0,2; ...; 0,9 ó 1,0.

Hay sólo un número finito de resultados posibles, así, tendremos una variable discreta. Sin embargo, si la población es muy grande, el número de posibles valores para la proporción es también muy grande (cualquier valor comprendido entre 0 y 1) y para los fines prácticos trataríamos la proporción de mujeres como una variable continua.

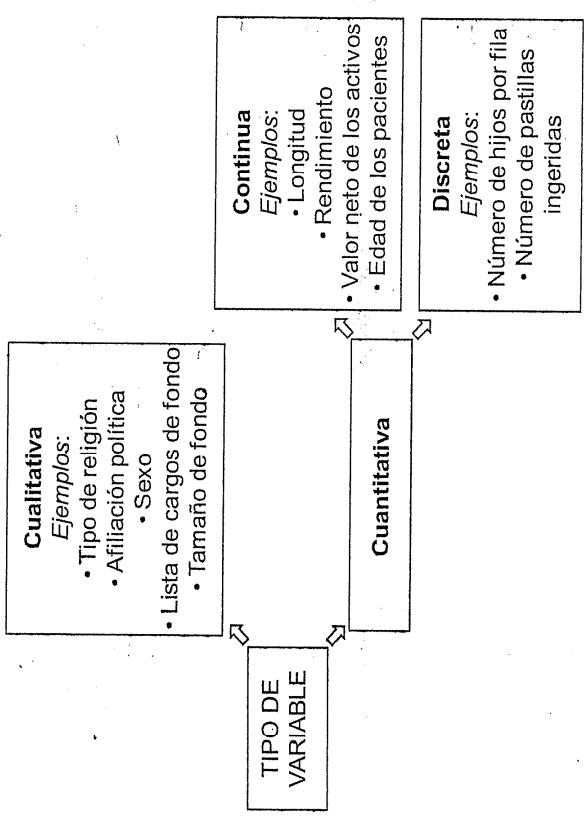
A veces una variable continua puede parecer discreta por la forma en que se efectúa su medición (a la unidad más cercana). La edad es de hecho una variable continua (estamos permanentemente volviéndonos más viejos con el tiempo). Sin embargo, a menudo medimos la edad discretamente en años.

#### Observaciones

El conjunto de **datos 2** consiste de una **variable continua** longitud, medida sobre 20 unidades.

Una unidad es una parte de una línea de producción.

En el conjunto de **datos 1**: el rendimiento alcanzado por los fondos de crecimiento, es una **variable cuantitativa continua**, mientras que por ejemplo si observo la base de datos, hallaremos información sobre la variable "objetivo del fondo", "lista de cargos", y "tamaño del fondo" que son **variables categóricas o cualitativas**.



#### Definiciones (\*)

Las **variables cualitativas**: son aquellas que clasifican a las unidades en categorías. Las categorías pueden tener o no, un orden. Las variables cualitativas son también denominadas **categóricas**.

Las **variables quantitativas** son aquellas cuyas observaciones provienen de procesos de medición o conteo (finito, infinito numerable o infinito no numerable).

Las operaciones aritméticas definidas sobre tales variables tienen significado. Distingúendose:

(\*)Estos conceptos se irán ampliando en el desarrollo del curso

**variable discreta** aquella en la cual puedo contar el número de posibles valores y **variable continua** que puede asumir cualquier valor sobre un intervalo dado.

### EJEMPLO 3.1. ¿Qué tipo de variable es?

a) El número de personas que ingresan al banco K el día viernes entre las 12.00 y las 13.00hs.

[**Quantitativa discreta**]

b) Tiempo empleado para correr 100 mts.

[**Quantitativa continua**]

c) Arrojar dos dados y registrar si las caras resultantes coinciden o no.

c1) Defina la variable

c2) Clasifíquela

Tipo de coincidencia: con modalidades: coinciden; no coinciden.

[**Cualitativa**]

d) Un restaurante de "comidas rápidas" tiene una promoción para los niños. La misma consiste en un premio entre cinco disponibles.

Una promoción es comprada y registra el premio recibido.

Tipo de premio [**Cualitativa**]

e) Número impreso en la camiseta de los jugadores de fútbol.

[**Cualitativa**. Dado que no podemos sumar o restar los número. No tiene sentido]

f) Una mujer es seleccionada aleatoriamente de una ciudad y se registra si la misma padece de cáncer de mama.

[**Cualitativa**]

### Para Resolver!!!

3.1. ¿Qué tipo de variable es?

a) El peso de varias cartas postales

b) Marcas de automotores alquilados en el mes de enero.

c) El tiempo de duración de un CD.  
d) La cantidad de lluvia caída durante una estación en una región determinada.

- e) La religión de una persona.
- f) La longitud de una cuerda.
- g) El número de fechas distintas de nacimientos de los estudiantes de una clase.

### PARA PENSAR!!!

Un número de bultos es llevado a una oficina de encomiendas por correo. Los paquetes son pesados y se registran los pesos: 9kg, 5kg, 4kg, 12kg, 20kg, etc. Estos valores son todos números enteros. ¿Implica esto que la variable es discreta?

La variable **peso** **es continua**.

Hemos medida el peso lo más cercano al kilo. Un paquete teniendo un peso de 12 kg, podría realmente pesar 12,2kg ó 11,997kg. O cualquier otro valor del intervalo 11,5 a 12,5kg.

**Punto clave:** no permitamos que el aspecto del dato luego de ser registrado sea engañoso!

Consideremos nuevamente la variable peso: Los paquetes que pesan menos de 5kg son clasificados como **liviano**, los que pesan entre 5 y 20kg son clasificados como **mediano** y los que pesan más de 20kg como **pesado**.

Registraremos a la variable peso con modalidades **liviano**, **mediano** y **pesado**. Ahora la variable peso es cualitativa (se ha categorizado).

**Punto clave:** el tipo de variable depende principalmente del proceso de medición, no sobre la propiedad a ser medida. Es importante preguntar algunas cuestiones sobre los datos: ¿cómo fueron obtenidos?

### 3.2.2. ¿Qué? ¿Cómo? ¿Quién? ¿Cuándo?

Un conjunto de datos es un conjunto de observaciones o mediciones. Cuando nos confrontamos con un conjunto de datos, es importante hacernos algunas preguntas. Por supuesto que necesitamos saber **¿Qué se está midiendo?** Pero el **QUÉ** no puede estar separado del

**CÓMO** de la medición. Necesitamos conocer cómo las medidas (o mediciones) fueron tomadas, como así también qué instrumentos fueron usados para hacer la medición.

También es importante registrar otro tipo de información sobre las variables reales de interés. Por ejemplo **¿Quién realizó las mediciones? ¿Qué día y a qué hora fueron realizadas? ¿Cómo eran las condiciones ambientales cuando se realizaron las mediciones?** Tal información es útil. Cuando miramos los datos, si una observación se aleja demasiado del resto (un potencial externo o extremo outlier), luego registrar una información extra puede ayudarnos a entender por qué esta observación fue diferente y puede garantizarse su exclusión para un análisis posterior.

Consideremos los siguientes datos sobre tasas de graduaciones, presentadas en Julio de 1993 editadas del New York Times:

Graduation rates at selected colleges and universities as reported to the National Collegiate Athletic Association Figures are for freshmen who entered in 1986 and had graduated by summer 1992. Athletes are those who find athletic scholarships. (Source: Chronicle of Higher Education)

	All students	Athletes	Students	Athletes
Yale	96	*	Univ. of So. California	66
Princeton	95	*	Indiana	65
Notre Dame	94	84	Syracuse	64
Virginia	92	88	Texas (Austin)	63
Georgetown	92	95	Alabama (Tuscaloosa)	55
Stanford	92	86	Miami (Fla.)	57
Cornell	90	*	Ohio State	55
Rice	87	78	Brigham Young	48
Penn State	77	78	Howard	45
California (Berkeley)	77	61	Houston	31
U.S. Military Academy	76	*	Morgan State	16

\*Institution does not award athletic scholarships.

- ¿Qué colegio o institución tuvo la tasa más alta de graduación?
- ¿Quién la más baja?

Las tasas de graduación son a menudo usadas como una herramienta de comercialización para atraer alumnos a los colegios y Universidades. Este artículo nos invita a preguntar ¿Qué es una tasa de graduación? ¿Es el porcentaje de estudiantes que se gradúa en 4 años? ¿Hay algunos programas que normalmente toman más de 4 años? ¿Cómo son manejados los alumnos transferidos? El primer problema en el reporte de estos datos, es que no hay estándares, ni guías sobre qué

constituye una "tasa de graduación", es decir, el **qué** y el **cómo** de la medición. Por lo tanto, los datos pueden no ser comparables y no muy útiles.

Las conclusiones posiblemente cambien según sean los caminos utilizados para producir una tasa de graduación. Por ejemplo Stanford tiene una tasa de graduación más alta que Penn State, en ambas categorías.

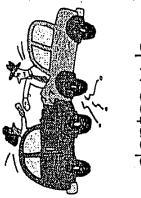
Por quizás si clasificamos los resultados en "grado de ciencia" y "no ciencia" que tienen los estudiantes, Penn State puede tener tasas más altas de graduación. Registrar información extra, tal como "tipo de grado", puede proveernos potenciales subgrupos, que examinaremos separadamente y permitirán ver si los resultados totales generalizan a través de los subgrupos.

Este artículo reconoce los problemas al reportar sobre los datos.

No todos los artículos que se refieren a datos ocupan nuestra atención en tales preguntas. Pero preguntar tales cuestiones nos ayuda a dirigir intelligentemente la información que recibimos, y así poder usarla para buenas y claras conclusiones decisiones.

### PARA RESOLVER !!!

### 3.2. ¿Información equivocada o faltante?



Los registros de una ciudad indicaron que en el pasado mes, doce accidentes de autos ocurrieron en una misma esquina muy transitada. El hombre fue el conductor en 10 de los accidentes y la mujer fue la conductora en los otros 2.

Ud. es interrogado para comparar los resultados de accidentes de autos cometidos por los varones versus las mujeres.

- ¿Cuál sería su conclusión si aproximadamente 1000 conductores masculinos y 20 femeninos han pasado por dicha intersección o esquina el mes pasado?
- ¿Cuál sería su conclusión si aproximadamente 100 conductores y 200 conductoras han pasado por dicha esquina el mes pasado?
- ¿Liste varios Qué? ¿Quién? ¿Cómo? ¿Cuándo? Antes de comparar los resultados entre femeninos y masculinos.

### 3.2.3. Distribución de una variable

Si elegimos una variable y medimos el valor de esta variable para una unidad, luego a otra unidad, y luego a otra y continuamos midiendo para un gran número de unidades, encontraremos que dichos valores varían de una unidad a otra.

La variación está presente en los datos, es decir, esperamos que los valores de una variable "varíen".

La distribución de una variable está dada por el conjunto de los valores posibles de esa variable y la frecuencia con que ocurre cada uno de dichos valores.

La distribución de una variable puede ser presentada gráficamente, numéricamente y con un modelo.

Hay muchas formas gráficas de presentar una distribución.

El método usado dependerá del tipo de variable y de la idea a ser presentada.

En la sección 3.3. examinaremos gráfico de sectores y de barras para presentar datos cualitativos. En sección 3.4. veremos cómo usan las distribuciones de frecuencias, diagramas de Tallo y Hojas e histogramas para presentar variables cuantitativas.

#### Definición

La distribución de una variable provee los posibles valores que una variable puede tomar y cuan frecuentes ocurren estos posibles valores.

La distribución de una variable muestra el modelo de variación de la variable.<sup>(1)</sup>



3.1. Establezca el tipo de variable. Si la variable es cuantitativa, decir si es discreta o continua.

- a) Preferencia religiosa.
- b) Cantidad de leche en un vaso.
- c) Número de tarjetas de créditos.
- d) Número de estudiantes de una clase de 35 que entregan sus trabajos en término.

<sup>(1)</sup>Estos conceptos se irán ampliando en el desarrollo del curso.

- e) Marcas de PC compradas por un mayorista.
- f) Cantidad de líquido dispensada por una máquina usada para llevar pocillos de café.

3.2. La ocupación de una persona clasificada por el Bureau Of Statistics está codificada por números.

Por ejemplo a la clase ocupacional "gerencial y profesional" se le asigna el número 1, a la clase "técnica" se le asigna el número 2, a la clase "empleados" el número 3, etc.

Estos datos son descriptos mejor como variable:

- a) discreta
- b) cualitativa
- c) continua
- d) cuantitativa
- e) ninguna de las de arriba
- f) no se da suficiente información.

3.3. Un amigo nos cuenta que en su clase de Inglés sólo 5 fallaron en sus exámenes parciales, mientras que en sus clases de Química un total de 20 estudiantes fallaron en el examen parcial. ¿Es el número total de alumnos que fallaron en los parciales una medida válida de la dificultad de la clase? Si es una medida válida, explique porqué. Si no lo es, sugiera una mejor.

3.4. Dos proyectos gubernamentales se están considerando. Con el proyecto A se estima una tasa de desempleo del 10%, mientras que con el proyecto B se estima una tasa desempleo del 5%. ¿Puede Ud. concluir que la adopción de ambos proyectos resultaría una tasa de desempleo del 15%? Explique.

### 3.3. PRESENTACIÓN DE DISTRIBUCIONES. VARIABLES CUALITATIVAS

Generalmente las variables cualitativas no son medidas en escala numérica, pueden ser categorizadas.

Se las presenta en tablas de distribución de frecuencias. La misma muestra el número (frecuencia:  $f_j$ ) de elementos correspondientes a cada una de las varias clases que no se traslapan. Sin embargo, muchas veces nos interesa conocer la proporción o porcentaje, de los artículos en cada clase. La frecuencia relativa ( $h_j$ ) de una clase es la fracción o proporción de elementos que pertenecen a esa clase. Para un conjunto de datos con  $n$  observaciones, la frecuencia relativa de cada clase ( $h_j$ ) resulta ser el cociente:

$$\text{frecuencia absoluta de la clase}$$

La frecuencia relativa porcentual ( $h_i \%$ ) de una clase es la frecuencia relativa multiplicada por 100. Los dos tipos de gráficos más comunes para variables cualitativas son los gráficos de sectores y de barras.

#### EJEMPLO 3.1.

**Beverage Digest**, manifestó que Coke Classic, Diet Coke, Dr. Pepper, Pepsi-Cola y Sprite son las cinco bebidas carbonatadas más vendidas.



A continuación se presenta una tabla que muestra la distribución de las marcas de gaseosas más vendidas.

Esta tabla proporciona un resumen de cómo se distribuyen las 50 compras en las cinco bebidas carbonatadas. Con ella se tiene una

#### Distribución de las bebidas gaseosas compradas según marcas.

1998

MARCA	Nº (f)	PROPORCIÓN (h)	PORCENTAJE (h %)
Coke classic	19	0,38	38
Diet Coke	8	0,16	16
Dr. Pepper	5	0,10	10
Pepsi-Cola	13	0,26	26
Sprite	5	0,10	10
<b>TOTAL</b>	<b>50</b>	<b>1,00</b>	<b>100</b>

FUENTE: Datos obtenidos en base a una muestra aleatoria de 50 compras.  
Beverage Digest, 1998

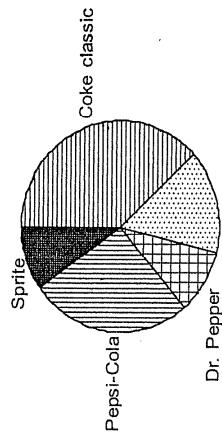
visión más clara que con los datos originales mostrados en la tabla 3.1 del Anexo.

Se aprecia que la marca líder es Coke Classic, y que Pepsi-Cola está en segundo lugar, Diet Coke en tercero y Sprite y Dr. Pepper compiten por el cuarto lugar. Esto nos indica la popularidad relativa de las cinco bebidas.

**NOTA:** En clases aprenderemos a interpretar las columnas de la tabla.

A continuación presentamos el diagrama de sectores circulares para representar las distribuciones de frecuencias relativas o relativos porcentuales.

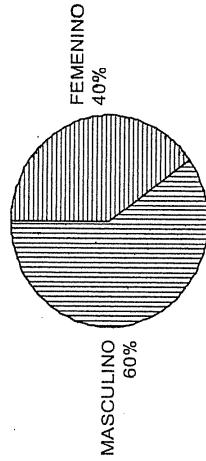
#### VENTAS DE GASEOSAS SEGÚN MARCAS



FUENTE: Datos obtenidos en base a una muestra aleatoria de 50 compras.  
Beverage Digest, 1998.

#### EJEMPLO 3.2 Distribución de los pacientes según su sexo

El gráfico de sectores para el Conjunto de Datos N°3 resulta:



FUENTE: conjunto de datos N°1

#### 3.4. Listas de cargos de fondos de acciones generales nacionales-período: 2001-2002

La siguiente tabla resume los tipos de cargos que realizan los 194 fondos de inversión:

Fondos de acciones generales nacionales según listas de cargos 2001-2002

LISTA DE CARGOS	NÚMERO DE FONDOS
Cargos de activos del fondo	17
Cargos diferidos	5
Cargos por venta directa	19
Cargos múltiples	46
Fondos sin cargo	107
<b>TOTAL</b>	<b>194</b>

Fuente: Archivo Mutual, Berenson-Levine-Krehbiel

a) ¿Qué porcentaje de fondos cubren los costos de venta que pagan los activos de los fondos?

b) ¿Cuántos fondos presentan cargos de venta diferidos o amortizados?

c) ¿Qué porcentaje de fondos están en la categoría "sin cargo"? (es decir, no requieren pagos)

d) Construya un gráfico de sectores.

Dicha información luego será también presentada en un gráfico de barras horizontales

### DIAGRAMA DE PARETO

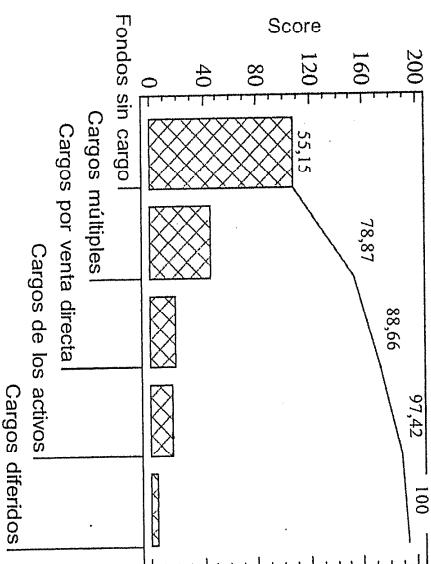
Es un buen recurso para representar datos categóricos. A menudo proporciona más información visual que los diagramas de barras y el de sectores circulares.

Es un tipo especial de gráfico de barras verticales, donde las respuestas categorizadas se grafican en orden descendente de frecuencias y se combinan con un polígono acumulado en la misma escala.

El diagrama de Pareto resulta más ventajoso cuando la variable categórica contiene muchas categorías. Se usa ampliamente en control estadístico de procesos y en el control estadístico de la calidad del producto.

### Ilustración

Diagrama de Pareto para Lista de cargos



### COMENTARIOS

En la construcción de un diagrama de Pareto, el eje vertical de la izquierda contiene las frecuencias (o porcentajes), el eje vertical de la derecha contiene los porcentajes acumulados de (de 0 a 100) y el eje horizontal contiene las categorías de interés. Las barras con separación uniforme tienen el mismo ancho.

Generalmente (según software), el punto en el polígono de porcentajes acumulados para cada categoría se centra en el punto medio de cada barra.

Al estudiar un diagrama de Pareto se buscan dos cosas: las magnitudes de las diferencias en las longitudes de las barras que corresponden a las categorías adyacentes decrecientes y los porcentajes acumulados de estas categorías adyacentes.

Class label	Diagrama de Pareto con frecuencias acumulativas					
	Rank	Count	Weight	Weight Score	Cum. Score	Percent
Fondos sin cargo	1	1	107	107	107	55,15
Cargos múltiples	2	1	46	46	153	78,87
Cargos p/ vía dir.	3	1	19	19	172	9,79
Cargos de los act.	4	1	17	17	189	8,76
Cargos diferidos	5	1	5	194	194	97,42
<b>TOTAL</b>		5	194			100,00

### The StatAdvisor

This table shows the frequency of occurrence of the 5 unique values of fondos. The classes are sorted according to the weighted scores, which are equal to the counts times the weights. The highest class is Fondos sin cargo with a score of 107, which represents 55,1546% of the total.

Salida de Statgraphics

### PARA RESOLVER!!

### 3.5. Ingresantes a la F.C.C.YE

La siguiente tabla muestra el número de ingresantes por carrera de la facultad de Ciencias Económicas y Estadística. Año 1997 de la UNR.

#### INSCRIPTOS SEGUN CARRERA

#### FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. AÑO 1997

##### CARRERA

Contador Público

Nº

2168

Lic. En Administración

Nº

632

Lic. Economía

Nº

123

Lic. Estadística

Nº

79

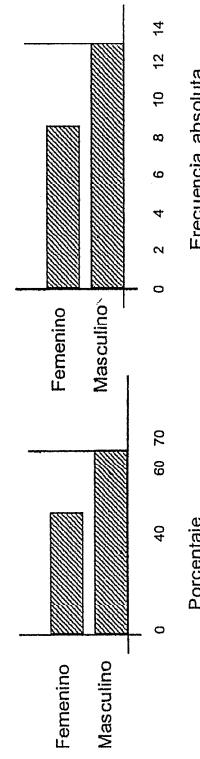
FUENTE: Sección alumnado Fac. Cs. Ecs. y Est. UNR

- a) ¿Qué porcentaje de alumnos se inscribieron en la carrera de Lic. en Estadística?  
 b) ¿Qué porcentaje de alumnos lo hicieron en la carrera de Contador Público?  
 c) Construya el gráfico de sectores.

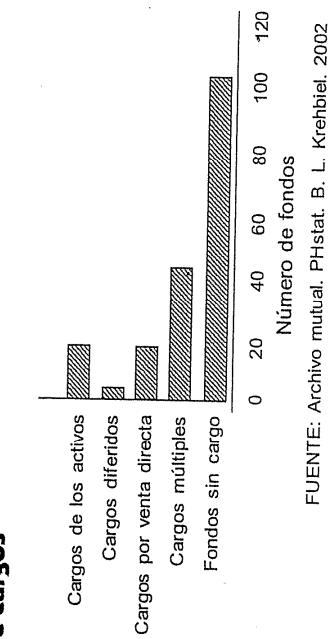
### 3.3.2. Gráfico de barras

Los gráficos de barra también muestran el porcentaje de ítems que caen en cada categoría. Un gráfico de barras muestra una barra para cada categoría. El ancho de las barras no tienen significado pero debería ser el mismo para todas las categorías. La Fig. 3.2(a) es un gráfico de barras según género de los 20 sujetos del estudio médico sobre la reducción de la presión arterial usando la frecuencia relativa porcentual. La Fig. 3.2(b) es el gráfico de barra de los mismos datos pero usando en número o frecuencia.

**Fig. 3.2 (a)**  
Distribución de los pacientes según género



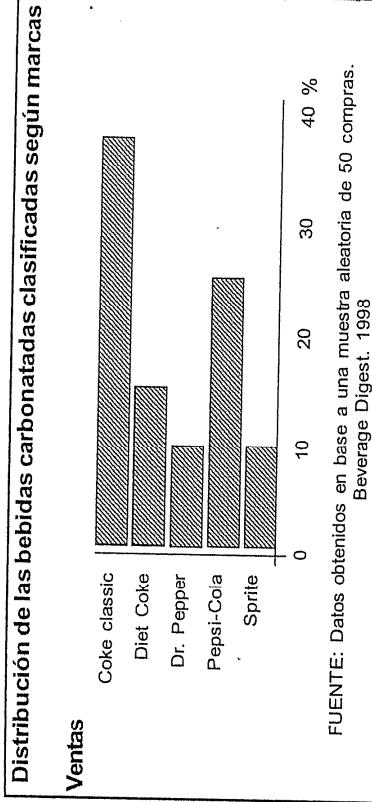
**Fig. 3.2 (b)**  
Distribución de los pacientes según género



### EJEMPLO 3. 3. Gráfico de Barras

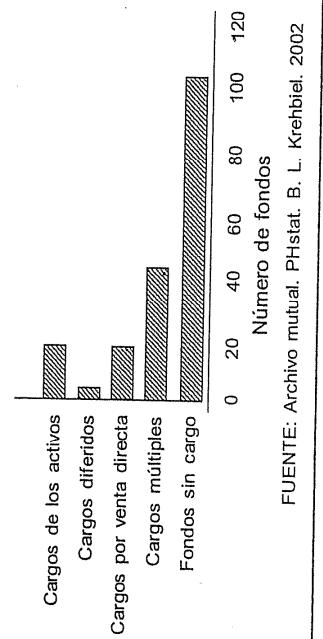
**Compras de bebidas carbonatadas**  
Este gráfico de barras muestra la distribución de las ventas de bebidas gaseosas, según marcas.

En uno de los ejes de la gráfica (por lo general el eje vertical), especificamos las etiquetas que se utilizan para las clases (categorías) de los datos. Para el otro eje de gráfica (por lo general el horizontal), se puede usar una escala de frecuencias absolutas, frecuencias relativas, o de frecuencias relativas porcentuales.



No hay un orden natural para las categorías según marcas. Este gráfico muestra que el mayor porcentaje de ventas es el correspondiente a Coke Classic (representada por la barra más alta). Para este gráfico de barras los porcentajes suman 100%.

**Otro gráfico de barras: Distribución de los fondos según lista de cargos**



### Observaciones:

- Las barras pueden ser verticales u horizontales.
  - Luego se analizarán según distintos software y recomendaciones del INDEC.
- Los gráficos de barras pueden ser usados para representar dos categorías cualitativas al mismo tiempo ⇒ **GRÁFICO DE BARRAS COMPUESTO**

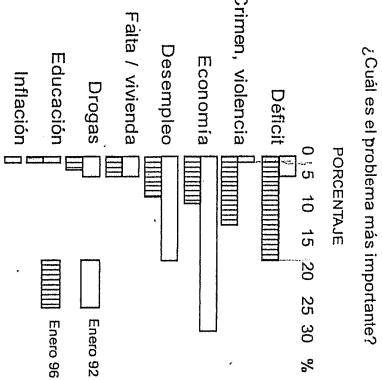


### PARA RESOLVER!!!

#### 3.6 ¿Cuál es el problema más importante?

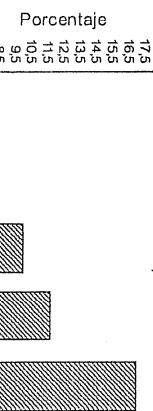
El siguiente gráfico de barras muestra el porcentaje de encuestados sobre cuál es, a su criterio, el problema más importante por el que está atravesando América. Se consideraron dos años enero 1992 y enero 1996.

- En enero de 1992, ¿qué problema o categoría tuvo el porcentaje más alto de respuestas?
- Este mismo problema obtuvo el mayor porcentaje de respuestas en enero de 1996?
- En enero de 1992, ¿qué porcentaje de encuestados respondieron "crimen y violencia" como el más importante problema de América? ¿y en enero 1996?



### PARA PENSAR!!!

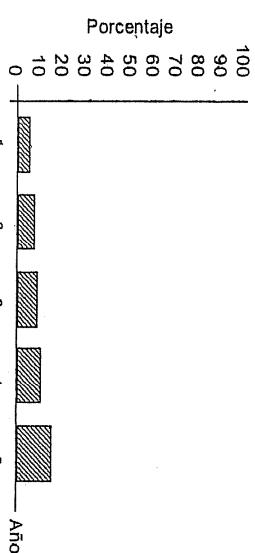
El gráfico de barras #1 muestra el % de alumnos inscriptos en los distintos cursos de la carrera de Antropología en la Universidad y han recibido cursos de computación en el Polimodal.



El gráfico de barras #1 da la apariencia de que el porcentaje ha (indique).

- Aumentado drásticamente
- Aumentado moderadamente
- Aumentado levemente

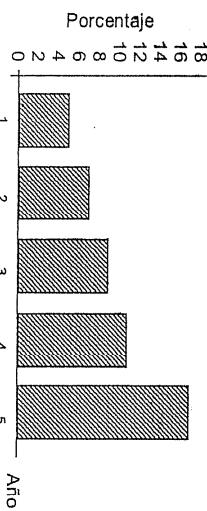
El gráfico de barras #2 muestra el porcentaje de alumnos inscriptos a los distintos cursos de Antropología de la Universidad y han recibido cursos de PC en la Polimodal.



Este gráfico muestra que el porcentaje ha ..... ( rodee uno)

- Aumentado drásticamente
- Aumentado moderadamente
- Aumentado levemente

Observe cuidadosamente los dos gráficos. Ellos muestran los mismos datos. Note que en el gráfico #1, el eje vertical comienza en 4.5 en vez de 0. Esta supresión del cero muestra un incremento más drástico en el porcentaje de los estudiantes que han recibido clases de PC en su ciclo anterior. En cambio en el gráfico #2 se observa una tendencia de leve incremento (observe la escala vertical del eje %). La forma adecuada de presentar estos datos es la Figura 3.4.



Un gráfico de barras nos permite comparar algunos grupos o categorías mediante las alturas de las barras correspondientes. Sin embargo, nuestros ojos tienden a enfocar no justamente la

## EJERCICIOS

altura de la barra sino su área, de ahí la importancia de que el ancho de las barras sea el mismo para todas.

### EJEMPLO 3. 4. Un gráfico de barras "enganoso"

El gráfico de barras siguiente muestra el total de ventas para tres inmobiliarias.

Cuando las barras son reemplazadas por diagramas relacionados con algún tópico, el gráfico recibe el nombre de **Pictograma**.

Es fácil confundir el gráfico de barras por un pictograma.



Las alturas de las casas reflejan los niveles correctos.

La inmobiliaria #1 presenta casi el doble de ventas que la inmobiliaria #3. Sin embargo el área de la casa correspondiente a la inmobiliaria #1 es más de 4 veces el área de la casa de la inmobiliaria #3.

#### Observación

Cuando observe un Pictograma sea cuidadoso con la interpretación, no permita que el área de los diagramas lo distraigan o engañen.

#### NOTA

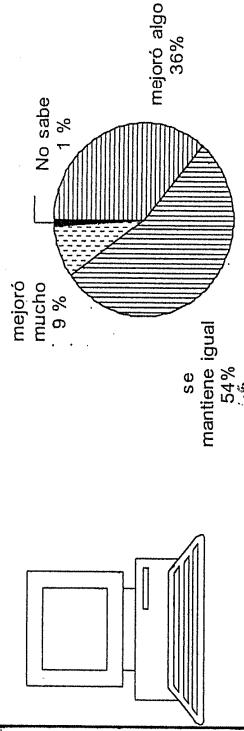
- Un GRÁFICO DE SECTORES CIRCULARES (comúnmente llamado de "torta") muestra la distribución de una variable cuantitativa, dividiendo al círculo en porciones o sectores circulares correspondientes a las categorías o modalidades de la variable, de modo que el tamaño de cada porción (o el ángulo) es proporcional al porcentaje de ítems en esa categoría.

- Un GRÁFICO DE BARRAS muestra la distribución de una variable cualitativa, listando las categorías (o modalidades) de la variable a lo largo de un eje y dibujando una barra según cada categoría con una altura (o longitud) igual al porcentaje de ítems en esa categoría. Las barras deben ser de igual ancho.

3.5. El gráfico de sectores circulares presenta información sobre como los estudiantes han cambiado y mejorado su rendimiento desde que comenzaron a usar computadoras.

#### COMPUTERS MAKING GRADE?

How students age 12-18 who do at least some of their homework assignments on a computer say their grades have changed since they started using a computer?



FUENTE: The USA Today, Marzo 4, 1997, pág. 11.

a) ¿Suman los porcentajes el 100%?  
 b) ¿Qué porcentaje de estudiantes han experimentado una gran mejora o alguna mejoría?

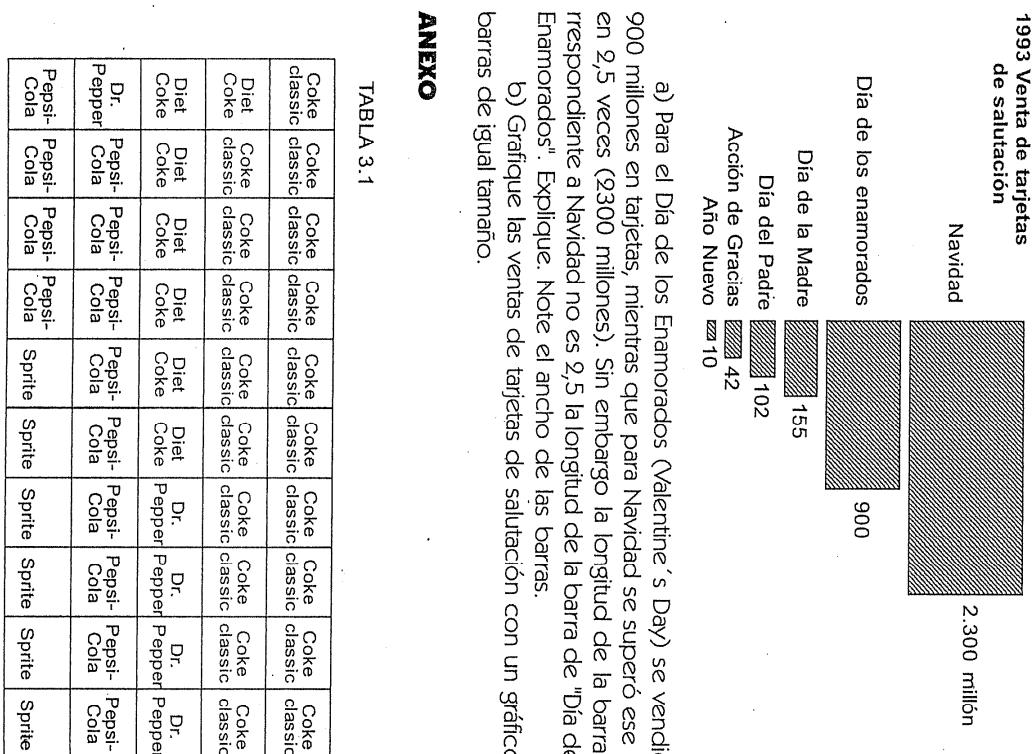
c) Presente esta misma información en un Gráfico de barras.  
 d) Si Ud. tuviera un familiar de algún estudiante de 12 a 18 años que no tiene PC. ¿Qué preguntas haría antes de aconsejar la compra? ¿Qué información no está presentada y qué le gustaría saber?

3.6. De acuerdo al Centro Nacional de Estadísticas de la Salud en el año 1940, el porcentaje de personas de 65 años que esperaban sobrevivir a los 90 fue del 7%. Para 1980 aquél porcentaje fue del 25%. Basado en proyecciones; los porcentajes para los años 2000 y 2050 fueron 26% y 42% respectivamente.

FUENTE "El viejo llegará a más viejo" The Ann Arbor News, febrero 1997, página A1.

Presente un gráfico de barras para mostrar el porcentaje de personas de 65 años de edad quienes esperan vivir hasta los 90 para los años 1940, 1960, 1980 y las proyecciones para los años 2000 y 2050.

**3.7** Pensando en Ud. Un gran número de tarjetas de salutación son compradas para celebrar "Navidad" y el "Día de los enamorados". El gráfico muestra el monto de ventas de tarjetas intercambiadas en varias celebraciones.



a) Para el Día de los Enamorados (Valentine's Day) se vendieron 900 millones en tarjetas, mientras que para Navidad se superó ese nivel en 2,5 veces (2300 millones). Sin embargo la longitud de la barra correspondiente a Navidad no es 2,5 la longitud de la barra de "Día de los Enamorados". Explique. Note el ancho de las barras.

b) Grafique las ventas de tarjetas de salutación con un gráfico de barras de igual tamaño.

## ANEXO

TABLA 3.1

| Coke classic |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Diet Coke    | Coke classic |
| Diet Coke    |
| Dr. Pepper   | Pepsi-Cola   |
| Pepsi-Cola   |

## PRESENTACIÓN DE RELACIONES ENTRE DOS VARIABLES CUALITATIVAS

Muchas investigaciones son conducidas para intentar estudiar la relación entre dos variables.

Para poder estudiar las respuestas simultáneas de dos variables categóricas, primero se construye una tabla de clasificación cruzada conocida como **tabla de contingencia o de clasificación cruzada**.

Tener una mascota incrementa el tiempo de sobrevida de pacientes con enfermedades coronarias? ¿Hay una relación entre el estado nutricional de los niños en edad escolar y su rendimiento académico? Las tablas de frecuencia o las tablas de contingencia son utilizadas para presentar los datos de conteo sobre dos variables cualitativas. Como podemos ver, las relaciones entre dos variables cualitativas se verán mejor si calculamos los porcentajes apropiados y usamos gráficos de barras para presentar esos porcentajes.

¿Hay una relación entre el estado nutricional de los niños en edad escolar y su nivel académico? El estado nutricional de 1.000 niños en edad escolar fue evaluado y codificado como "pobre", "adecuada" o "excelente". El nivel académico para estos niños codificadas como "por debajo del promedio", "promedio" o "por encima del promedio". Los datos son los siguientes:

Nivel Académico	Estado Nutricional			Totales
	Pobre	Adecuado	Excelente	
Por debajo del Promedio	70	95	35	200
Promedio	130	450	30	610
Por encima del Promedio	90	30	70	190
<b>Total</b>	<b>290</b>	<b>575</b>	<b>135</b>	<b>1000</b>

Esta **tabla de frecuencia** muestra cuántas observaciones caen dentro de las diversas categorías. Cada intersección de una fila y una columna es llamada celda. El valor 130 en la primera columna y segunda fila de la tabla, nos dice que 130 de los 1.000 niños en edad escolar, o el 13%, tienen un estado nutricional pobre y nivel académico promedio. En la anterior tabla hubo 90 niños con un estado nutricional que tenían un nivel académico por encima del promedio, en tanto que hubo sólo 70 estudiantes con un estado nutricional excelente y que tenía un nivel académico por encima del promedio. Sin embargo, los 90 niños eran de un total de 290 niños con un estado nutricional pobre (solo el 31%), mientras que los 70 niños pertenecían a un total de 135 niños con estado

nutricional excelente (52%). La relación entre las dos variables se verán mejor si calculamos los porcentajes apropiados.

### ¿Qué información podemos obtener de esta tabla?

De la tabla de doble entrada del estado nutricional y el nivel académico, podemos obtener la **distribución marginal** de cada variable. Encontrar la distribución marginal para el estado nutricional, la variable columna, tomamos los totales columna y los dividimos por 1,000, que es el total de la tabla. Multiplicando estos resultados por 100 encontramos los porcentajes para cada nivel nutricional. Vemos que un poco más de la mitad de los niños en edad escolar tienen un adecuado nivel estado nutricional y sólo una tercera parte tienen un estado nutricional pobre.

		Estado Nutricional		
		Pobre	Adecuada	Excelente
Porcentaje	Porcentaje	29%	57.5%	13.5%

La distribución marginal del Nivel Académico, la variable fila, se puede encontrar de forma similar. Los valores se hallan tomando los totales fila, dividiendo por 1000, y luego multiplicando por 100 para presentar los porcentajes.

### PARA RESOLVER!!

#### 3.7

De la distribución marginal del nivel académico.

		Nivel Académico		
		Por debajo del promedio	Por encima del promedio	
Porcentaje	Porcentaje	%	%	%

¿Qué porcentaje de los niños en edad escolar tuvieron un nivel académico promedio? ¿Por debajo del promedio? ¿Por encima del promedio?

De la tabla de doble entrada para el estado nutricional y el nivel académico, también podemos encontrar la **distribución condicional** de una variable dada la otra variable. La distribución condicional de una variable fila, dada la variable columna, se encuentra expresando los valores de la

tabla original como porcentajes de los totales columna. La distribución condicional de una variable columna, dada la variable fila, se encuentra expresando los valores de la tabla original como porcentajes de los totales fila.

Para encontrar la distribución condicional del nivel académico, la variable fila, dado el estado nutricional, variable columna, vamos a la tabla de doble entrada original y escribimos cada columna de valores como un porcentaje del total de esa columna. Comenzando con la primera columna, estado nutricional pobre, tomariamos el primer valor de 70, dividimos por 290 (que es el total de la primera columna), y multiplicamos por 100 para dar el porcentaje:

$$(70 / 290) \times 100\% = 24.1\%$$

El segundo valor en la primera columna, 130, también será dividido por 290 y expresado como un porcentaje de la siguiente manera:

$$(130 / 290) \times 100\% = 44.8\%$$

Para el tercer y último valor de la primera columna, 90, podemos dividirlo por 290 y expresarlo como un porcentaje:

$$(90 / 290) \times 100\% = 31.0\%$$

Podríamos también obtener este último valor tomando la suma de los dos porcentajes anteriores y restarlo del 100%:

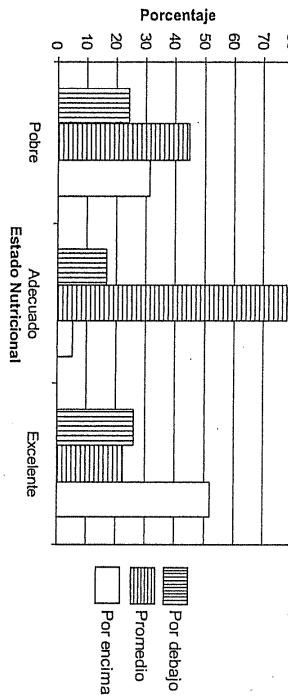
$$100\% - (24.1\% + 44.8\%) = 100\% - 68.9\% = 31.1\%$$

Estos porcentajes, 24.1%, 44.8% y 31.1%, dan la distribución condicional del nivel académico dado un estado nutricional **pobre**. La distribución condicional del nivel académico dado un estado nutricional **ade-cuado** y dado un estado nutricional **excelente** se pueden encontrar en forma similar. Los divisores serían 575 y 135, respectivamente, en lugar de 290. La tabla completa de la distribución condicional del nivel académico dado el estado nutricional está dada por:

		Estado Nutricional		
		Pobre	Adecuado	Excelente
Nivel Académico	Por debajo del Promedio	24.1	16.5	25.9
	Promedio	44.8	78.3	22.2
Por encima del Promedio		31.1	5.2	51.9
Totales		100%	100%	100%

Primero, notemos que cuando presentamos la distribución condicional de la variable fila dada la variable columna, los totales columna son iguales al 100%? ¿Qué vemos cuando examinamos esta tabla? En general vemos un cambio, según el estado nutricional mejora, la mayor proporción de niños se va situando hacia la clase de nivel académico más alto.

El siguiente gráfico de barras muestra la distribución condicional del nivel académico según el estado nutricional.



#### Definiciones

Los datos de dos variables cualitativas a menudo son presentados en forma de tabla de frecuencias a doble entrada. La **distribución marginal de una variable fila** se obtiene por el cálculo de porcentajes de cada total fila, en relación al tamaño muestral completo. La **distribución marginal de una variable columna** se obtiene por el cálculo de porcentajes de cada total columna en relación al tamaño muestral completo.

La asociación entre dos variables cualitativas se obtiene calculando algunos porcentajes adecuadamente elegidos. La **distribución condicional de la variable fila**, dada la variable columna, se encuentra expresando los totales columna. La **distribución condicional de la variable columna**, dada la variable fila, se encuentra expresando los datos de la tabla original como porcentajes de los totales fila. Si una variable es considerada explicatoria y la otra es la variable respuesta, entonces la distribución condicional de la variable respuesta, dada la variable explicatoria debería ser estudiada.

#### PARA RESOLVER!!!

#### 3.8 Poor women & abortion

El 19 de marzo de 1995, el New York Times publicó un artículo, "Los Enemigos del Aborto se Preocupan por Cortes de Operaciones de Bienestar Público", el cual contenía la información siguiente:

Poor Women & Abortion	Women aged 15-19 whose unplanned pregnancies ended in abortion in 1988 by income level...
Women aged 14-44 whose unplanned pregnancies ended in abortion in 1988 by income level...	43%      51%      51%
Women aged 14-44 whose unplanned pregnancies ended in abortion in 1988 by income level...	Poor      Low Income      Higher Income
Women aged 15-19 whose unplanned pregnancies ended in abortion in 1988 by income level...	39%      54%      70%
Women aged 15-19 whose unplanned pregnancies ended in abortion in 1988 by income level...	Poor      Low Income      Higher Income
Women aged 15-19 whose unplanned pregnancies ended in abortion in 1988 by income level...	Poor refers to those earning below the federal poverty level of \$14,800 for a family of four in 1994.
Women aged 15-19 whose unplanned pregnancies ended in abortion in 1988 by income level...	Low income earning 100% to 199% of that level. Higher income earning more than 200% of the Federal poverty level.

La parte de arriba de la presentación da información de dos variables: ingreso (pobre, baja y alta) y estado del embarazo no planeado (terminó con un aborto y no terminó en un aborto).

- a) De las mujeres pobres (de edades entre 15 y 44 años), ¿qué porcentaje no terminó su embarazo no planeado en un aborto en 1988?

Respuesta: ..... %

- b) Nombre la distribución condicional que es mostrada en la primera parte de la presentación:
- La distribución condicional de ..... dada

#### PARA RESOLVER!!!

#### 3.9 La Nueva Pócima

El pub "La Nueva Pócima" fabrica y distribuye tres tipos de cerveza: una cerveza suave baja en calorías, una cerveza común y una cerveza negra. Para investigar la relación entre el género y la cerveza que prefieren los clientes, se seleccionó una muestra de 450 clientes que son bebedores de cerveza. Después de probar cada una de las tres cervezas, los

Individuos fueron indagados para establecer su preferencia, definida según su primera elección. La siguiente tabla resume las respuestas.

Género	Cerveza que prefieren		
	Suave	Común	Negra
Masculino	60	120	.60
Femenino	90	90	30

a) Describa brevemente en palabras qué significa el valor 30 de la tabla.

b) Usando la siguiente tabla vacía, de la distribución condicional de la cerveza que prefieren dado el género.

Género	Cerveza que prefieren		
	Suave	Común	Negra
Masculino	45 %	52, 14%	6, 5%
Femenino	60 %	42, 86%	33, 3 %

c) Basados en los cálculos de (b), ¿el pub "la Nueva Póxima" debería iniciar una campaña de publicidad para todas sus cervezas, o dirigir sus promociones hacia los diferentes tipos de consumidores, masculino versus femenino? Explique.

d) Presente la distribución condicional de (b) en forma de un gráfico de barras.

Los investigadores a menudo están interesados en evaluar si las relaciones entre dos variables cualitativas son o no significativas. Esto se puede lograr probando la hipótesis de no asociación. La técnica para probar esta hipótesis involucra el cálculo de la estadística chi-cuadrado y la presentación del correspondiente p-value. La estadística chi-cuadrado se puede pensar como una medida de la distancia entre los datos observados presentes en la tabla de doble entrada y los valores que se presentarían si realmente no hubiese asociación en la población.

### 3.4. MOSTRANDO DISTRIBUCIONES. VARIABLES CUANTITATIVAS

Las variables cuantitativas están medidas sobre una escala numérica. (recordar escalas de medición).

Los gráficos más comunes para variables cuantitativas son:

- ⇒ De frecuencia
  - ⇒ De tallo y hoja
  - ⇒ Histogramas
  - ⇒ De caja y bigotes
- (comúnmente llamado box plot en próximo capítulo)

Si los datos de una variable cuantitativa han sido juntados a lo largo del tiempo, podríamos examinar los valores a través de un gráfico de tiempo.

### 3.4.1. Gráficos de frecuencia

Un gráfico de frecuencia constituye una manera rápida de mostrar la distribución de los datos sobre una recta. En el capítulo 1, hicimos uso de los gráficos de frecuencia en los ejemplos de nuestra bolsa A y B. Cada punto o valor observado de la variable está representado por una **X**, posicionada a lo largo de una escala. La escala generalmente es la escala horizontal, pero podría ser también una escala vertical. Como su nombre lo indica, La frecuencia o el número de los valores que se repiten será representado en la otra escala.

#### Pasos básicos para la construcción de un Gráfico de Frecuencia

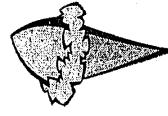
- Dibujar una recta
- Marcar los valores Mínimo y Máximo sobre el eje real
- Completar la escala para los números con incrementos igualmente espaciados
- Marcar cada valor observado de la variable con **X**, sobre la escala adecuada
- Si hay dos o más ítems con el mismo valor, debemos apilarlos verticalmente

**Observación:** Algunos textos usan puntos (•), en lugar de cruces (X)

### EJEMPLO 3.5. Precios de los helados de crema

Los precios (en cts. de dólares) de los helados conocidos de una sola bocha correspondientes a 17 negocios de Los Angeles, captados en setiembre del 2001 fueron los siguientes:

25, 51, 70, 75, 90, 90, 91, 95, 95, 95, 96,  
100, 105, 110, 115, 120



**Gráfico de frecuencias**

X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120				

El valor mínimo es 25 y el valor máximo es 120. En el gráfico de frecuencia correspondiente

El precio más bajo fue de 25 centavos y el más alto fue de \$1.20. Podemos observar cómo los datos están agrupados o concentrados cerca de los noventa. También, vemos unos pocos valores alejados del resto de los datos (25 y 51)

Se observa una brecha o claro entre los 25 y 51 y entre 51 y 70, también hay un ligero claro entre los 75 y 90.

**NOTA**

Si una o más observaciones están alejadas del resto, éstas se denominan valores extremos (outliers).

Un conjunto de observaciones separadas del resto de los datos forman una concentración, un conglomerado o racimo.

Un claro o brecha (gap) está dado por la distancia entre las observaciones.

**PARA RESOLVER!!!**

**3.10**

¿Cuántas llaves tienen en este momento?

Variable en estudio:.....

Respuestas:

Nº de llaves:

Nº de estudiantes:



Construya un gráfico de frecuencia. Escribe un breve resumen de la información contenida en su gráfico. Analice si hay valores extremos, claros (gaps o conglomerados).

**PARA RESOLVER!!**

**3.11 Generador de números aleatorios**

Las computadoras y calculadoras tienen la habilidad de producir listados de números enteros.

a) Use su TI con semilla 22, y genere 50 enteros aleatorios entre 1 y 10. Si no la tiene, use la TNA comenzando por la fila 6, columna 1. Construya un gráfico de frecuencias. Note que los enteros podrán ocurrir más de una vez, no los excluya.

b) Describa la distribución.

**EJERCICIOS**

**3.8.** Considere el número de goles registrados en cada partido de hockey jugados en la última temporada:

3 - 5 - 2 - 2 - 1 - 4 - 0 - 0 - 3 - 3 - 2 - 1 - 8 - 2 - 4 - 1 - 1 - 3 - 2 - 5

Construya una tabla de distribución de frecuencias y un gráfico de frecuencia.

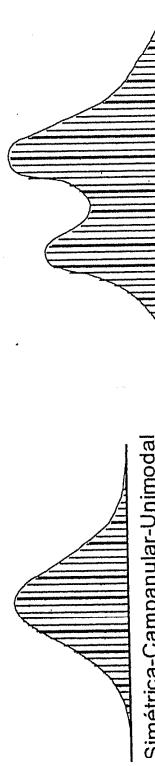
**3.9.** Describa gráficamente el conjunto de edades de los 20 sujetos sometidos al estudio médico para la reducción de la presión sanguínea. Los datos están presentados en CONJUNTO DE DATOS 3.

**3.10.** Obtenga los resultados del último "quiz" de Métodos Estadísticos 1 y presente la distribución de estos resultados con un gráfico de frecuencias.

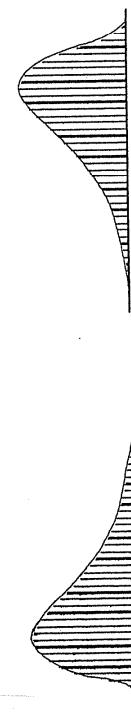
**3.4.2 Forma de las distribuciones**

Al observar un gráfico, podemos comentar acerca de la forma general, el centro aproximado de la distribución, y cualquier desviación de la forma general.

Ejemplos de las formas de distribuciones de algunas características:

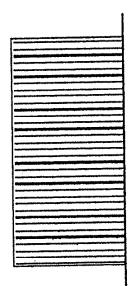


Simétrica-Campanular-Unimodal



Bimodal

Asimétrica a la izquierda  
(asimetría negativa)



Uniforme

Recordar los términos más usados para describir la forma de una distribución.

### 3.4.3 Diagrama de tallo y hoja

Un diagrama de tallo y hoja, también llamado diagrama de tallo, es una manera rápida de mostrar la distribución de un conjunto de datos con un número relativamente pequeño de unidades. Tiene el beneficio de retener los valores reales de las variables.

#### PASOS BÁSICOS PARA UN DIAGRAMA DE TALLO Y HOJA

- ✓ Separe cada medición en un tallo y una hoja, generalmente la hoja consiste en exactamente un dígito (el último) y el tallo consiste de 1 o más dígitos.

Por ej.: 734 ► tallo = 73, hoja = 4

Para 2.345 ► tallo = 234 hoja = 5

A veces el decimal es sacado del tallo pero una nota se agregará sobre como leer cada valor.

Para el ej. 2.345, podríamos decir que 2.345 debería leerse como 2.345. A veces, cuando los valores observados tienen muchos dígitos, sería beneficioso, ya sea: redondear los números (2.345 redondearlo a 2.35, con tallo = 2.3, hoja = 5); o bien truncar dígitos (truncar 2.345 en 2.34).

✓ Escriba los tallos en orden creciente verticalmente (de arriba hacia abajo) y trace una línea a la derecha de los tallos.

✓ Adjunte cada hoja al tallo apropiado.

✓ Ordene las hojas en orden creciente (de izquierda a derecha).

### EJEMPLO 3.6 [Conjunto de datos Nº 3]

Consideraremos las edades de los 20 sujetos del conjunto de datos número 3. Ellos varían de 32 a 51, de allí que para construir un diagrama básico de tallo y hoja, usaremos el último dígito como hoja y el primer dígito como tallo.

3	2	7	9
4	0	1	1
5	0	1	

Nota: 3 | 2 representan 32 años.

**Observa** que el más pequeño de los valores representados por 3 | 2, debería ser leído como 32 años.

De este diagrama básico podemos ver que la mayoría de los sujetos tratados estaban en los "cuarenta". Sin embargo, con sólo tres tallos y un gran número de hojas sobre un tallo, no están muy bien exhibidas la variación y la forma de la distribución.

Es útil otro tipo de arreglo. Dividiremos los tallos. En lugar de tener todos los dígitos de 0 a 9 sobre un mismo tallo, tendremos dos. Uno de los tallos es para los dígitos de 0 a 4 y el segundo es para los dígitos de 5 a 9.

3	2	7	9
4	0	1	1
4	5	5	6
5	0	1	

Nota: 3 | 2 representa 32 años.

Ahora podemos ver mejor que la distribución de las edades de los sujetos es aproximadamente simétrica, centrada en 43-44 aproximadamente, sin apariencia de valores extremos (observaciones que caen fuera del modelo general de variación).

### EJEMPLO 3.7. [ Conjunto de datos Nº 1 ]

Consideremos los rendimientos que alcanzaron los 59 fondos de crecimiento.

Un examen al diagrama de tallo y hoja permite obtener las primeras conclusiones sobre el rendimiento a un año de los fondos de inversión. Entre ellas:

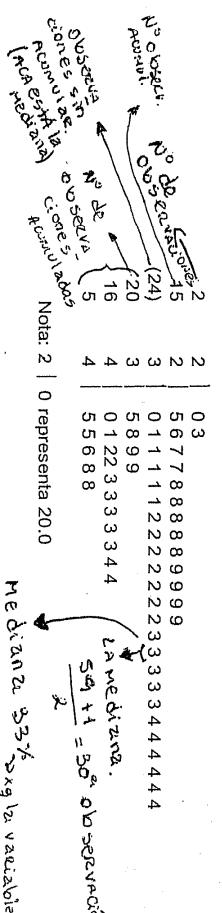
- El rendimiento más bajo al año es 20.4
- Los rendimientos de los 59 fondos de crecimiento se distribuyen entre las tasas más baja y más alta con cierta concentración entre 28 y 34
- Algunos fondos mutuos tienen los mismos rendimientos por ejemplo, hay dos fondos para cada porcentaje de rendimiento de 28.9, 31.6, 32.3, 32.9, 34.0 y 34.7.

Stem-and-leaf of 1Yr%Grow

Nº observación	Nº de observación
1	20.4
1	21.1
1	22.1
2	23.8
2	24.2
3	25.6
4	26.2
6	27.6
11	28.3
15	29.3
16	29.5
21	31.5
23	32.1
(5)	33.0
26	34.0
20	35.0
19	36.3
19	37.3
19	38.2
18	39.0
16	40.7
15	41.1
14	42.8
12	43.3
7	44.6
5	45.4
3	46.7
2	47.2
2	48.0
6	49.6

Diagrama de tallo y hoja obtenido con Minitab.

Otra presentación, truncando dígitos: Según salida de Statgraphs.



### EJEMPLO 3.8 [ Conjunto de datos Nº 2 ]

Recordemos el CONJUNTO DE DATOS Nº 2.

Las partes o piezas fueron programadas para tener 20 cm. de longitud. Un importante criterio en producciones de partes es lograr las mismas con medidas lo más cercanas posibles a la programada o especificada. Estas medidas muestran las desviaciones de 20 cm. ▶ con el valor más chico: 19.969 y el más grande: 20.011.

Usando este último dígito como hoja y los dígitos remanentes para el tallo, obtenemos el siguiente diagrama de tallo y hoja:

1996	9
1997	5
1998	4 4 5
1999	2 4 7 8
2000	0 1 1 2 4 4 7 8 8
2001	1

Nota: 1996 | 9 representa 19.969 cm.

Aquí tenemos una distribución no simétrica, con más variación por debajo de 20 cm especificado.

Si giramos el diagrama de tallo y hoja, la cola tira hacia la izquierda. Tal distribución es asimétrica a la izquierda, presenta asimetría negativa. Podemos identificar el centro de la distribución, o sea, podemos hallar el valor que divide a la distribución en dos partes iguales.

Con  $n = 20$  observaciones, y si las observaciones están ordenadas, el promedio de la 10<sup>a</sup> y 11<sup>a</sup> observación: 20,0005 es el valor central en que exactamente la mitad de las observaciones son menores que ella y la otra mitad cae arriba de este valor central.

Luego examinaremos este conjunto de datos 2, usando otra técnica gráfica llamada **Diagrama de caja** (Box-Plot).

Supongamos que queremos comparar este conjunto de datos con un segundo proceso de producción.

Las longitudes de las 20 partes producidas por el 2º proceso son presentadas en un **Diagrama de tallo y hoja comparativo**.

Proceso 2

Proceso 1	
1996	9
1997	5
	4 4 5
9	1998
6 4 4	1999 2 4 7 8
9 8 7 7 5 3 1 1 1 0 0 0	2000 0 1 1 2 4 4 7 8 8
1 0 0	2001 1

Nota: 9 | 1998 representa 19.98 cm.

Un diagrama de tallo y hoja de este tipo es útil para comparar dos distribuciones.

### Observación

- ❖ Las hojas para el 2º proceso están arregladas en orden creciente (de derecha a izquierda).
- ❖ En este caso, parece que el 2º proceso tiene una distribución más simétrica y menos variable comparada con la del 1º proceso.
- ❖ 20.001 es el valor central que divide a la distribución en dos partes, de manera tal que el 50% de las observaciones son menores que él y el otro 50%.

### PARA RESOLVER!!

#### 3.12 ¿Quién tuvo menor actuación?

Se registraron los resultados de las tareas de dos alumnos:

- Estudiante A: 80 - 52 - 86 - 94 - 76 - 48 - 92 - 69 - 79 - 45.
- Estudiante B: 73 - 87 - 81 - 75 - 78 - 82 - 84 - 74 - 80 - 76.

Construya un diagrama de tallo y hoja comparativo de datos. ¿Puede determinar qué estudiante realizó mejor las tareas?

### PARA PENSAR!!!

#### ¿Qué está mal?

Explique por qué cada uno de los siguientes diagramas de Tallos y Hojas no nos proveen una muy buena exhibición de la distribución

Diagrama de tallos y hojas N° 1

Diagrama de tallos y hojas N° 2	
27	9
32	0 1 17 8
33	1 2 25 9
34	0 3 4
35	1 1
41	0
58	6

Nota: 27 / 9 representa 279

Nota: 2 / 1 representa 21

Nota: 18 / 1 representa 181

**AYUDA:** Asegúrese que sus incrementos estén igualmente espaciados. Use partición de tallos, redondeos o truncamientos si tiene pocos tallos con muchas hojas o si tiene pocos tallos con pocas hojas.

Diagrama de tallos y hojas N° 3

Diagrama de tallos y hojas N° 3	
2	1 1 2 2 3 4 4 5 5 6 7 7 8 9
3	0 2 2 3 3 4 6 7 8
4	0 1 1
18	1
19	0
20	2
21	1 2 8
22	0
23	7
24	5 8
25	26 2 3
27	0 5
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35

Nota: 18 / 1 representa 181



**3.11. Aumento de ingresos.** Los siguientes datos son los porcentajes ganados por cápita del ingreso de personas por estados (de 1993 a 1994) registrado en el artículo ¿Cuánto ganan las personas?

Parade Magazine, Junio, 18 1995. Fuente: Bureau of Economic Analysis, U.S. Department of Commerce.

- a) Construya un diagrama de tallo y hoja. Asegúrese de incluir una nota, como:

----- representa -----

AL 5.1	HI 2.4	MA 4.9	NM 4.6	SD 9.5
AK 3.1	ID 4.1	MI 8.5	NY 4.7	TN 5.7
AZ 5.1	IL 5.4	MN 7.0	NC 5.4	TX 3.7
AR 5.6	IN 6.1	MS 7.4	ND 8.6	UT 5.6
CA 2.7	IA 10.9	MO 5.9	OH 6.3	VT 4.0
CO 3.9	KS 5.3	MT 2.8	OK 4.2	VA 4.3
CT 4.4	KY 5.4	NE 4.1	OR 5.1	WA 3.8
DE 4.5	LA 6.3	NV 4.9	PA 4.9	WV 6.4
FL 5.0	ME 4.7	NH 4.8	RJ 4.7	WI 6.1
GA 5.2	MD 4.3	NJ 4.3	SC 4.9	WY 3.6

- b)** Describa la forma de la distribución. ¿Es aproximadamente simétrica o asimétrica? ¿Cuál es el centro aproximado de la distribución? ¿Hay valores extremos?

- c)** ¿Qué estado tuvo el más alto porcentaje?

**3.12** Los siguientes datos son los octanajes de varias mezclas de gasolineras registradas por una organización de consumidores:

90.5	83.4	88.8	88.5
88.0	87.7	88.4	
86.7	89.1	87.5	

Construya un diagrama de tallo y hoja de manera tal que las hojas correspondan al primer dígito luego del punto decimal.

- 3.13** El pescado de mar *únicamente* tiene la línea o engorda?

Creáse o no, algunos tipos de pescados de mar, tienen un alto contenido graso.

La tabla adjunta, provee el contenido de grasa, grasa saturada, colesterol, y calorías para 3 onzas (1 onza = 28.35 grs) de consumo de pescado.

- a)** ¿Cuál de las 4 propiedades es más importante para Ud?  
**b)** Construya un diagrama de tallo y hoja para la propiedad elegida en a)

- c)** ¿Qué tipo de pescado actúa mejor según la propiedad elegida en a)? ¡el peor?

- d)** Elija otra propiedad y realice el diagrama de tallo y hoja. Describa la distribución.

- e)** El tipo de pescado s/c sigue comportándose de la misma manera para esta propiedad s/c) (por tratarse de intervalo de igual amplitud)

Tipo (3oz)	Grasa (g)	Grasa saturada (g)	Colesterol (mg)	Calorías
Salmon (chinook)	11.4	2.7	72	196
Salmon (atlantic)	6.9	1.1	60	155Tuna (bluefin)
5.3	1.4	42	157	
Swordfish	4.4	1.2	43	132
Oysters (eastern)				
(12 medium)	4.2	1.1	93	117
Salmon (pink)	3.8	0.6	57	127
Clams (19 small)	2.5	0.4	35	119
Snapper	1.7	0.2	57	126
Crab (alaska king)	1.3	0.3	40	109
Shrimp	1.5	0.1	45	82
Orange Roughy	0.9	0.2	166	84
Lobster	0.8	0.0	22	75
	0.5	0.1	61	83

### 3.4.3. Histogramas

Un histograma es otra opción para mostrar la distribución de una variable cuantitativa. Cuando la cantidad de datos es grande, un histograma es mejor que un gráfico de tallos y hojas, aunque no retiene los valores numéricos actuales. Muestra la distribución de una variable a través de la frecuencia o porcentajes del total de valores que hay en todo el rango de variación.

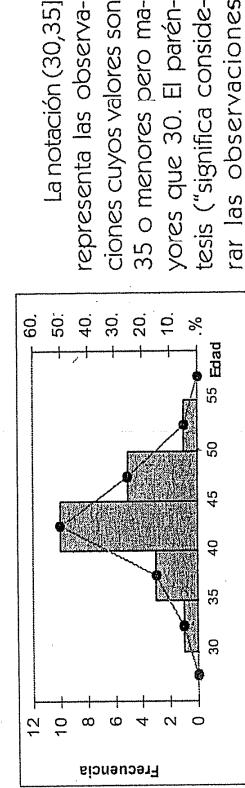
#### Pasos básicos para construir un histograma

- Identificar el menor y el mayor valor observado de la variable. Calcular el Alcance, Recorrido o Rango ( $X_{(n)} - X_{(1)}$ )
- Dividir el rango de los datos en clases o intervalos **de igual amplitud**. Las clases deben cubrir el total del rango de los valores, pero no deben superponerse. Si los datos van de 0 a 29, podría hacer que sus clases comiencen en 0 y vayan hasta 30 con una amplitud de 5.
- Contar el número de observaciones que caen en cada clase, a este número se lo llama frecuencia absoluta.
- Dibujar un eje horizontal y marcar las clases sobre él.
- En el eje vertical se puede representar la frecuencia absoluta, la proporción o el porcentaje.
- Dibujar un rectángulo (una barra vertical) sobre cada clase con la altura igual a la frecuencia, la proporción o el porcentaje.

### EJEMPLO 3.9 Histograma de las edades (Conjunto Nº 3)

Tomamos las edades de los 20 sujetos del Conjunto de Datos 3. Sus rangos van de 32 a 51, para un histograma podríamos crear clases que comiencen en 30 con incrementos de 5 hasta llegar a 55. Probaremos con diferente cantidad de clases y distintas amplitudes hasta seguir un buen gráfico.

Clase	Recuento	Observaciones Frecuencia absoluta	Proporción	Porcentaje %
(30,35]	1	1	$1/20 = 0.05$	5
(35,40]	11	3	$3/20 = 0.15$	15
(40,45]	HF-HF	10	$10/20 = 0.50$	50
(45,50]	HF-HF	5	$5/20 = 0.25$	25
(50,55]	1	1	$1/20 = 0.05$	5
<b>TOTAL</b>		<b>20</b>		<b>100</b>



La notación (30,35] representa las observaciones cuyos valores son 35 o menores pero mayores que 30. El parentesis "(" significa considerar las observaciones desde el 30 pero sin incluir este extremo, el corchete "] a la derecha significa considerar las observaciones hasta el 35 inclusive.

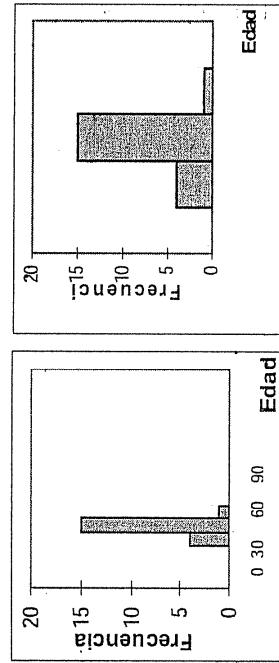
Otra vez se puede ver que la distribución de las edades de los sujetos es más o menos simétrica sin aparentes outliers. Muchas de las edades se encuentran entre los 40 y 45 años.

En el histograma anterior, note el impacto visual del área. En particular, el área de cada rectángulo es proporcional a la frecuencia o el conteo para cada clase. El concepto de área de área luego cuando hablamos de los modelos para variables continuas o curvas de densidad en el próximo capítulo.

La percepción visual de la distribución podría cambiar drásticamente si cambia la amplitud de las clases. Aquí se puede ver el histograma de los datos de las edades en donde la amplitud de las clases es 2 años en lugar de 5.

**¿Ha cambiado su percepción? ¿Hay alguna nueva información representada en este actual histograma?**

La percepción visual de la distribución podría cambiar al cambiar la escala en uno o en ambos ejes. Los siguientes 2 histogramas representan los mismos datos. El histograma en la derecha muestra una distribución muy concentrada con poca dispersión, mientras que el otro histograma representa una distribución con grandes variantes.



Finalmente si quisieramos comparar dos distribuciones basados en dos conjuntos de datos, podríamos usar porcentajes (frecuencias relativas porcentuales) en lugar de frecuencias absolutas al construir los histogramas, especialmente si los conjuntos de datos tienen diferente número de observaciones.

Los diagramas de tallos y hojas lado a lado sirven para comparar 2 distribuciones, al igual que los histogramas dibujados lado

a lado. Se usa el mismo eje horizontal (eje x) y las barras para un grupo ascendente (como en un histograma regular), mientras que las barras para el otro grupo descenden.

Los histogramas de lado a lado para los datos del tiempo de vida de las lamparitas del ejercicio 3.14 se muestran aquí.

### PARA RESOLVER!!

#### 3.13 Asocie las formas con las características

Considera las siguientes 4 distribuciones y las 4 características que aparecen a continuación. Determina cuál característica corresponde a cada distribución.

Distribución 1



Característica =

Distribución 2



Característica =

Distribución 3



Característica =

Distribución 4



Característica =

Edades de la población de los Estados Unidos en el año 1980. Describir y explicar la forma de la distribución.

Longitud (en millas) de la costa de los 50 Estados Unidos. Describir y explicar la forma de la distribución. ¿Qué Estados cree usted que podrían estar en la clase más extrema a la derecha?

Millas recorridas para ir a trabajar (distancia recorrida por adultos empleados en una ciudad). Describir y explicar la forma de la distribución.

Edad de muerte de la población de los Estados Unidos en el año 1980. Describir y explicar la forma de la distribución.

### Características

Edades de la población de los Estados Unidos en el año 1980. Describir y explicar la forma de la distribución.

Longitud (en millas) de la costa de los 50 Estados Unidos. Describir y explicar la forma de la distribución. ¿Qué Estados cree usted que podrían estar en la clase más extrema a la derecha?

Millas recorridas para ir a trabajar (distancia recorrida por adultos empleados en una ciudad). Describir y explicar la forma de la distribución.

Edad de muerte de la población de los Estados Unidos en el año 1980. Describir y explicar la forma de la distribución.

### PARA RESOLVER!!

#### 3.14 iDinero, dinero, dinero!

Pregunta: En este momento **cuánto dinero (en centavos) suman las monedas que tiene?**

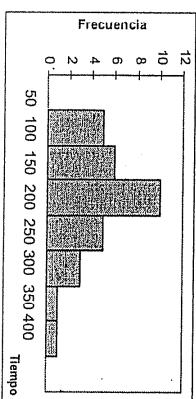
#### Datos:

Realice una tabla de distribución de frecuencias (absolutas, relativas, relativas porcentuales) e histograma(s) con los resultados de su grupo, con diferentes opciones para la amplitud de cada clase. Escriba un resumen de la información incluida en su histograma.



3.14 Dé un ejemplo real de una variable cuantitativa cuya distribución sea asimétrica a izquierda. Explique brevemente la respuesta.

3.15 El histograma de la izquierda muestra el comportamiento del tiempo (en segundos) que 31 clientes elegidos al azar del First Bank debieron aguardar antes de ser atendidos telefónicamente



- Describa la distribución.
- Aproximadamente qué porcentaje de clientes esperó más de 100 segundos?
- ¿Cuántos de los clientes esperaron más de 5 minutos?
- ¿Cuál es aproximadamente la máxima cantidad de tiempo que alguno de los 31 clientes debió esperar en línea? Asegúrese de incluir las unidades de medida en la respuesta.

### EJEMPLO 3.10 [conjunto datos N° 1]



#### Tabla de distribución de frecuencias Histogramas. Polígonos acumulativos

A medida que el número de observaciones crece, es necesario condensar más los datos en tablas apropiadas, a fin de presentar, analizar e interpretar los resultados en la forma correcta. Las observaciones se pueden agrupar en categorías o clases. Tal arreglo de datos en forma tabular se llama Distribución de Frecuencias.

Tabla de frecuencias para el rendimiento total a un año

Clase	Límite inferior	Límite superior	Punto medio	Frecuencia	Frecuencia relativa	Frecuencia acumulada	Frecuencia rel. acumulada
en o sobre	20.0	25.0	22.5	0	0.0000	0	0.0000
1	20.0	25.0	22.5	2	0.0339	2	0.0339
2	25.0	30.0	27.5	13	0.2203	15	0.2542
3	30.0	35.0	32.5	25	0.4237	40	0.6780
4	35.0	40.0	37.5	3	0.0508	43	0.7288
5	40.0	45.0	42.5	11	0.1864	54	0.9153
6	45.0	50.0	47.5	5	0.0847	59	1.0000
7	50.0	55.0	52.5	0	0.0000	59	1.0000

Menor: 34.9644 Desviación estándar: 6.65346

La distribución de frecuencias es una tabla de resumen en la cual los datos se colocan en categorías establecidas en forma conveniente de clases ordenadas numéricamente.

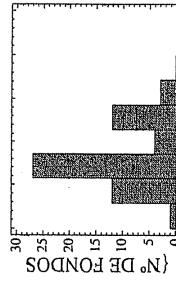
Al construir una tabla de distribución de frecuencias, se debe tener cuidado en la selección del número de clases y en el establecimiento de las fronteras o límites de cada clase (sin que se traslape). • **Selección del número de clases:** Por lo general la distribución de frecuencias debe tener como mínimo 5 clases, pero no más de 15. Una forma de determinar el número m apropiado de intervalos es:

• **Determinación del ancho de clase:** Es conveniente que cada intervalo de clase tenga la misma medida. Para determinar el ancho de intervalos de clase, se divide el Rango  $[X_{\max} - X_{\min}]$  por el número de intervalos (m)

• **Establecimiento de las fronteras:** Debe evitarse que las clases se traslapeen. Siempre que sea posible, estas fronteras se deben elegir en una forma que facilite la lectura e interpretación de los datos. Por ejemplo:

$$l_{\text{inferior}} < X_j \leq l_{\text{superior}}$$

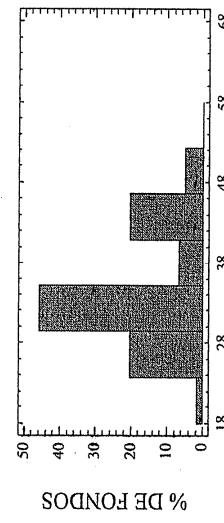
HISTOGRAMA



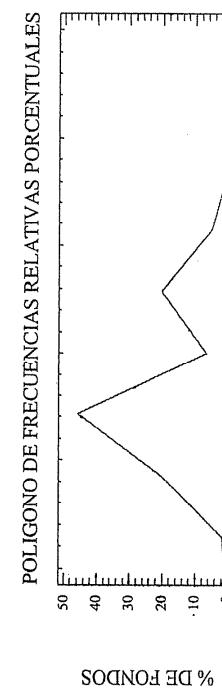
Rendimiento total a un año

(Recomendaciones y observaciones sobre escalas y cortes de ejes en clase)

HISTOGRAMA DE FRECUENCIAS RELATIVAS PORCENTUALES

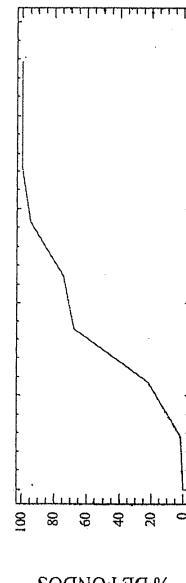


POLIGONO DE FRECUENCIAS RELATIVAS PORCENTUALES



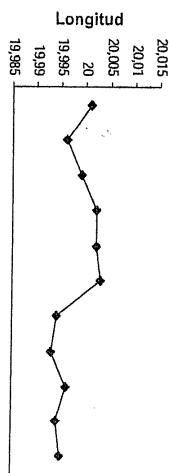
Rendimiento total a un año

POLÍGONO ACUMULATIVO



Rendimiento total a un año

### 3.4.4 Gráficos de series de tiempo



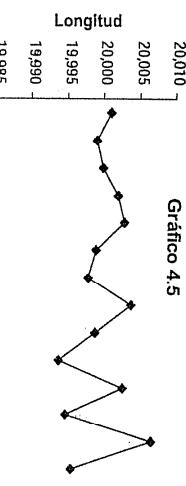
Cuando los datos de una variable se recogen a lo largo del tiempo, es útil graficar los datos contra el tiempo o en el orden en el cual se obtuvieron. Un gráfico de serie de tiempo de las observaciones sucesivas se conectan con líneas para ayudarnos a apreciar si la distribución parece cambiar con el tiempo. Esta gráfica nos muestra la distribución de longitudes de piezas (materia prima en una línea de producción), parece cambiar con respecto a la localización del centro. Tal vez, el cambio en el centro se deba a cambios en el material de origen.

La siguiente gráfica muestra una distribución que parece cambiar con respecto a la variación. El incremento de la variación puede deberse, por ejemplo, a las herramientas o la maquinaria.

Los patrones que podemos encontrar en un gráfico de serie de tiempo son los siguientes:

- **Tendencia:** creciente o decreciente, cambios en la ubicación del centro, cambios en la variación o dispersión.

- **Componente estacional o ciclo:** patrones de comportamiento que se repiten con regularidad.



Recuerde que al observar un conjunto de datos esperamos ver alguna variación. Pero si aparece un patrón de comportamiento es importante preguntarse ¿cuál será la causa de ese patrón? En la industria se usan versiones mejoradas de los gráficos de series de tiempo para controlar distintos procesos. Estas versiones mejoradas se llaman gráficos de control. Se usan para estudiar la estabilidad de un proceso. La idea general es que usando probabilidades y técnicas estadísticas, se pueda identificar un intervalo dentro del cual debe caer la mayor parte de la información obtenida. Los puntos que caen fuera de los límites son altamente improbables y sugieren que el proceso puede haberse desestabilizado (fuera de control).

Estos puntos nos conducen a buscar una causa assignable de variación, e implementar la acción correctiva correspondiente tal como cambiar la herramienta o ajustar algún parámetro, para que el proceso vuelva a su anterior estado bajo control.



### 3.15 Un Gráfico Simple de Control Estadístico de Procesos

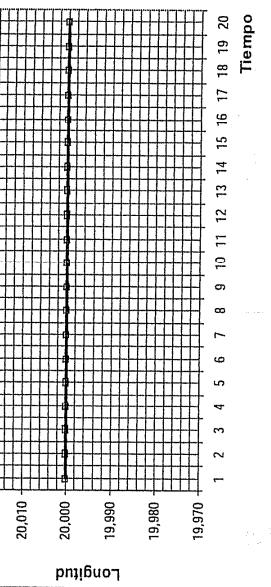


Volvamos a las veinte piezas consecutivas del Conjunto de datos 2.

El objetivo fue obtener piezas de 20 cm. Para estos datos ya hemos visto un gráfico de tallos y hojas y un histograma que mostró una ligera asimetría a la izquierda y más variación por debajo que por encima de los 20 cm. Grafique las observaciones en el orden en que se obtuvieron para ver qué nos dicen estos datos.

Parte #	1	2	3	4	5	6	7	8	9	10
Long. (cm)	20.01	19.98	19.99	19.99	20.00	20.00	19.99	20.00	20.00	20.00
Parte #	11	12	13	14	15	16	17	18	19	20
Long. (cm)	20.00	20.00	20.00	19.99	19.98	19.97	19.96	19.98	20.00	20.00
	7	4	1	7	4	5	9	4	4	2

## Dos maneras de "ver" los datos



a) ¿Qué le dicen los datos?

b) ¿Qué preguntas haría respecto de lo que observa?

### PARA RESOLVER!!

#### 3.16. ¿Patrones de Tardanza?



Los siguientes datos se refieren a la cantidad de estudiantes que han llegado tarde a la escuela durante un período de tres semanas.

	Lunes	Martes	Miércoles	Jueves	Viernes
Semana 1	10	7	6	8	11
Semana 2	14	5	10	8	7
Semana 3	9	3	6	4	6

Represente los datos con un gráfico de series de tiempo y comente lo que observa. ¿Hay alguna relación entre el día de la semana y el número de llegados tarde? Explique.

Los gráficos de frecuencias, los diagramas de tallos y hojas y los histogramas muestran la distribución de una variable cuantitativa. Pero si los datos se recogieron a lo largo del tiempo, un gráfico de series de tiempo también puede proveer información útil sobre la serie estudiada. Ambos modos de presentar la información pueden realizarse en un mismo gráfico. El gráfico de la izquierda muestra, por un lado, la distribución de los datos (representando las "x" en el eje "y" de la izquierda) y, por el otro, los datos en el orden en que fueron recolectados.

### 3.5. GUÍA PARA HACER GRÁFICOS

Hay una cantidad de errores comunes que aparecen en los gráficos. A veces estos errores están dados por la mala interpretación de la información. Hay algunos puntos a tener en cuenta antes de interpretar o construir un gráfico.

- Poner al gráfico un título apropiado.
- Incluir de dónde provienen los datos (fuente), detalles relevantes de cómo fueron obtenidos y una muestra de la magnitud.
- Estar seguro de marcar los ejes apropiadamente.
- Controlar si el eje de la frecuencia, proporción, o porcentaje comienza en cero.
- Incluir la unidad de medida de las variables analizadas.

### EJEMPLO 3.11. ¿Adónde han ido los ejes?



Un departamento de transporte publicó su reporte anual, que incluye el gráfico con el número de ciudadanos que utilizaron el transporte público en los últimos

5 años. En el gráfico se ve el aumento del número de usuarios. Sin embargo el gráfico no incluye un eje horizontal, que es necesario para interpretar los datos.

## EJERCICIOS

**3.16** Considere los siguientes gráficos de series de tiempo y las dos descripciones que acompañan. Determine qué series de tiempo corresponde a cada descripción. De una explicación del porqué de su elección.

### Descripciones

1. Gráfico de la proporción de chicas nacidas en cada mes de los últimos años en tres hospitales de la ciudad.
2. Gráfico de la cantidad acumulada de baterías falladas por día en los últimos 12 días.

**3.17** A veces un diagrama de tallos y hojas y un gráfico de series de tiempo dan información diferente y muy útil sobre una serie de datos. Los datos que aparecen a continuación son medidas de tensión tomadas en la rejilla de cables detrás de la pantalla de una computadora. Si la tensión es muy alta o muy baja, la pantalla no trabaja correctamente. Un operario mide la tensión de un pequeño grupo de pantallas muestradas cada hora de la línea de producción. Aquí aparecen los promedios para 20 horas consecutivas (de izquierda a derecha).

269.5	297.0	269.6	283.3	304.8	280.4	283.5	257.4	317.5	327.4
264.7	307.7	310.0	343.3	328.1	342.6	338.8	340.1	374.6	336.1

- Haga un diagrama de tallos y hojas.
- Describa la forma de la distribución. ¿Qué conclusiones pueden sacarse? (Tenga en cuenta que se espera que los datos muestren cierta variación).
- Haga un gráfico de serie de tiempo.
- Describa el patrón de comportamiento del gráfico anterior. ¿Qué recomendaciones haría al fabricante?

- 3.18** El artículo «Aumento de los accidentes de tránsito fatales protagonizados por mujeres» (The Ann Arbor News, 12 de febrero de

1995) presenta información sobre la tendencia creciente que se registra en los accidentes de tránsito fatales protagonizados por mujeres. Se presenta una serie de tiempo de los accidentes de tránsito fatales que involucran conductoras femeninas.

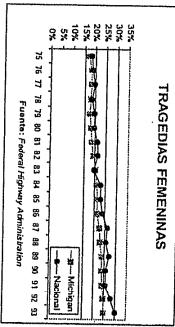
- ¿Qué porcentaje de accidentes de tránsito fatales con mujeres al volante ocurrió en 1993?
- ¿Qué porcentaje de accidentes de tránsito fatales con mujeres al volante ocurrió en 1975?
- ¿Qué porcentaje de accidentes de tránsito fatales involucró a conductores masculinos en 1993?
- Comente los resultados del año 1983.
- ¿Cuáles podrían ser factores de confusión que ocultan el por qué la cantidad de mujeres involucradas en accidentes de tránsito fatales va en aumento?

## RESUMEN DEL CAPÍTULO

En este capítulo hemos visto algunas de las técnicas más comunes para resumir datos gráficamente. Gráficoar los datos nos ayuda a "familiarizarnos" con ellos y los gráficos nos dan un resumen visual de los datos. Muchos resúmenes gráficos muestran la distribución de una variable de interés, de patrones de variación. Ver los datos en un resumen visual es a menudo el primer paso para obtener conclusiones. Si se tiene una idea de la distribución de una variable, se puede comparar con nuevas observaciones para evaluar si son consistentes con la distribución.

La distribución de una variable cualitativa se puede representar con un gráfico de torta o un gráfico de barras. Para ver la distribución de una variable cuantitativa, tendríamos que usar un gráfico de frecuencia, un diagrama de tallos y hojas o un histograma. Si los datos cuantitativos fueron recolectados a través de tiempo, sería bueno examinarlos en este sentido construyendo un gráfico de series de tiempo.

Cuando vemos un gráfico, debemos enfocarnos en el patrón general sin perdernos en los detalles. Una vez que hemos capturado el patrón general, podemos enfocarnos luego en los desvíos del patrón general. Los gráficos son una manera muy buena de mostrar los datos. Una impresión visual es mucho mejor que una presentación de datos en forma nu-



mérica. Pero también debemos recordar que hay buenos gráficos y gráficos que pueden dar una falsa impresión. Los gráficos se pueden usar para distorsionar la apariencia. Si cambiamos la escala o los intervalos podemos cambiar el resultado visual.

## EJERCICIOS

- a) Explique si el tiempo de espera es discreto o continuo.
- b) Describa esta distribución.
- c) ¿Cuál es el centro aproximado de esta distribución?
- d) ¿Cómo cree que se tomó la muestra? ¿Fue una simple muestra al azar? Explique.

- 3.19** Para cada una de las siguientes variables, determine si es cualitativa o cuantitativa. Si la variable es cuantitativa, explique si es discreta o continua.
- a) Estado civil.
  - b) El diámetro de los guijarros de un riachuelo.
  - c) El número de accidentes automovilísticos por año en Virginia.
  - d) El tiempo necesario para jugar 18 hoyos de golf.
  - e) El número mensual de permisos de construcción otorgados en una ciudad.

- 3.20** Se llevó a cabo un estudio para examinar las diferencias de performance laboral entre trabajadores de saco y corbata con tipo de comportamiento A (activo) y B (pasivo). Los datos fueron recopilados para varios trabajadores en las variables listadas. Para cada una, determine si es cualitativa o cuantitativa. Si es cuantitativa, explique si es discreta o continua.
- a) Tipo de comportamiento (A o B).
  - b) Edad (en años).
  - c) Número de empleados supervisados.

- 3.21** Uno de los problemas que enfrentan los administradores de los servicios públicos de salud es la queja de los pacientes. Una de esas es el tiempo que deben esperar para ser atendidos. En una encuesta para investigar los tiempos de espera, las secretarías debían anotar esos tiempos (desde la llegada hasta ser atendidos) de algunos pacientes. Los datos de un día (30 pacientes) son vistos en un gráfico de tallos y hojas lado a lado.
- |    |        |
|----|--------|
| 0  | 5      |
| 1  | 0555   |
| 2  | 000555 |
| 3  | 000555 |
| 4  | 0555   |
| 5  | 05     |
| 6  | 05     |
| 7  | 0      |
| 8  | 55     |
| 9  |        |
| 10 | 5      |
- Nota: 0/5 representa 5 minutos. El tiempo de espera fue tomado con 5 minutos de intervalo

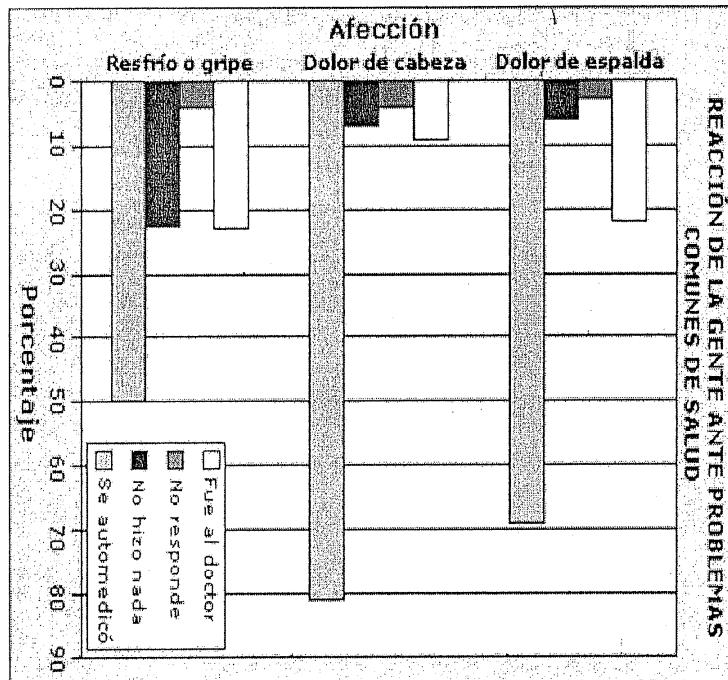
**RESUMEN DE DATOS EN FORMA GRÁFICA**  
**PRACTICA ADICIONAL**

1. Para cada una de las variables que siguen, determina si es cualitativa o cuantitativa (continua o discreta).
    - a) Estado civil.
    - b) Diámetro de piedras de un arroyo.
    - c) Número de accidentes automovilísticos en Rosario por año.
    - d) Tiempo que se tarda en jugar 18 hoyos en golf.
    - e) Número de permisos de construcción otorgados en una ciudad por mes.
    - f) Altura de niños de tercer grado.
    - g) Temperatura al mediodía durante el mes de junio en las cataratas del Iguazú.
    - h) Número de países que un adulto ha visitado durante su vida.
    - i) Si una persona a visitado o no Australia.
    - j) Número de visitantes por día al Parque de la Costa durante el mes de enero.
    - k) Contenido de gas en una garrafa.
  
  2. Médicos que asistieron a un seminario de medicina deben completar un formulario donde, entre otras cosas, se les pregunta: «Comparedo a otros seminarios que has tomado recientemente, ¿qué calificación que le pondrás a este? Mediocre – Aceptable – Bueno – Muy Bueno – Excelente». Una muestra de médicos seleccionados al azar produjo las siguientes respuestas. Para facilitar el procesamiento informático de los resultados de los cuestionarios, se utilizó una escala numérica con 1 = Mediocre, 2 = Aceptable, 3 = Bueno, 4 = Muy Bueno y 5 = Excelente.
- |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 4 | 4 | 5 | 1 | 5 | 3 | 4 | 5 | 2 |
| 4 | 5 | 3 | 4 | 4 | 4 | 5 | 5 | 4 | 1 |
| 4 | 5 | 4 | 2 | 5 | 4 | 2 | 4 | 4 | 4 |
| 5 | 5 | 3 | 4 | 5 | 5 | 2 | 4 | 3 | 4 |
| 5 | 4 | 3 | 5 | 4 | 4 | 3 | 5 | 4 | 5 |
| 4 | 3 | 5 | 3 | 4 | 4 | 3 | 5 | 3 | 3 |

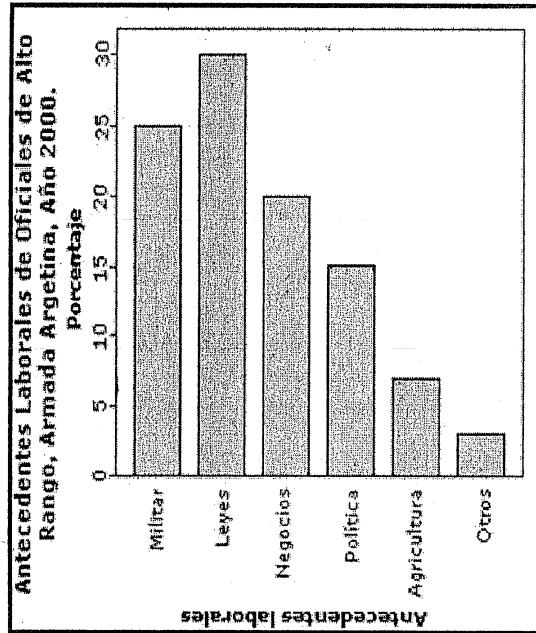
3. Para conocer la reacción de la gente ante problemas comunes de salud, una revista encuestó a una muestra representativa de la población nacional de 1752 hombres y mujeres mayores de 19 años. El gráfico enseña las respuestas. Completar las siguientes afirmaciones.
 

**REACCIÓN DE LA GENTE ANTE PROBLEMAS COMUNES DE SALUD**

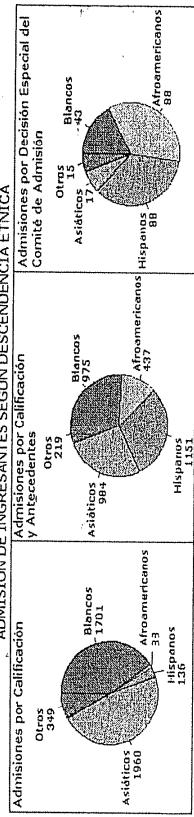
Categoría	Fue al doctor (%)	No responde (%)	No hizo nada (%)	Se automedicó (%)
Resfriado o gripe	10	15	10	65
Dolor de cabeza	15	10	10	65
Dolor de espalda	10	10	10	70



- a) Aproximadamente el \_\_\_\_\_ % de los entrevistados dijeron que se automedican si están resfriados o con gripe.
- b) Aproximadamente \_\_\_\_\_ encuestados no hacen nada si les duele la espalda.
- c) El porcentaje de encuestados que recurre al médico cuando tienen dolor de espalda es \_\_\_\_\_, mientras que el \_\_\_\_\_ % lo hace cuando tienen dolor de cabeza.
4. Una gran encuesta fue realizada entre oficiales de alto rango de la armada. Una de las variables medidas fue el antecedente laboral antes de incorporarse a la armada. La información recolectada se muestra en el siguiente gráfico:



- a) Aproximadamente el \_\_\_\_\_ % de los entrevistados dijeron que se automedican si están resfriados o con gripe.
- b) Aproximadamente \_\_\_\_\_ encuestados no hacen nada si les duele la espalda.
- c) El porcentaje de encuestados que recurre al médico cuando tienen dolor de espalda es \_\_\_\_\_, mientras que el \_\_\_\_\_ % lo hace cuando tienen dolor de cabeza.
4. Una gran encuesta fue realizada entre oficiales de alto rango de la armada. Una de las variables medidas fue el antecedente laboral antes de incorporarse a la armada. La información recolectada se muestra en el siguiente gráfico:
5. Los gráficos que siguen son diagramas de sectores circulares sobre descendencia étnica, uno para cada una de las tres posibles maneras en las que un estudiante puede ser admitido en una universidad. Los datos corresponden al tipo de admisión de la totalidad de los ingresantes a dicha universidad en un año particular, y a su descendencia.
- a) Resume en una tabla toda la información brindada por los gráficos.
- b) ¿Qué proporción de ingresantes admitidos por decisión especial del comité son hispanos?
- c) ¿Qué proporción del total de ingresantes son de descendencia asiática y fueron admitidos sólo por sus calificaciones?
- d) ¿Qué proporción de ingresantes admitidos por sus calificaciones y antecedentes son afroamericanos?
- e) ¿Qué proporción del total de ingresantes son blancos y su admisión fue una decisión especial del comité?
- f) ¿Qué porcentaje de los ingresantes afroamericanos fueron admitidos sólo por sus calificaciones?
- g) Realiza un diagrama de sectores para los ingresantes hispanos, que presente qué porcentaje de ellos fue admitido según cada modalidad.
- h) Ídem anterior con los ingresantes asiáticos, afroamericanos y blancos.

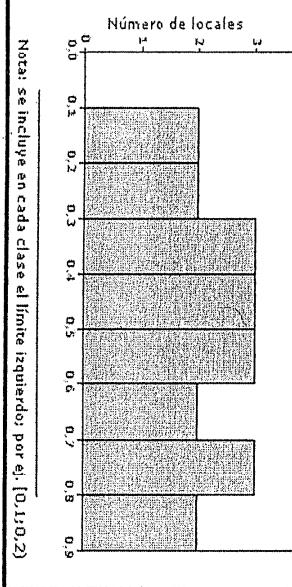


- a) Si 620 oficiales de alto mando respondieron la encuesta, ¿cuántos de ellos reportaron antecedentes laborales en negocios?
- b) ¿Cuántos oficiales tienen antecedentes laborales en el área de leyes o en política?
- c) Al observar este gráfico alguien dijo: «La distribución de los antecedentes laborales de los oficiales de alto rango es

Estado	Ingreso	Estado	Ingreso	Estado	Ingreso
AK	51074	LA	29518	OH	35022
AL	28530	MA	39604	OK	27263
AR	26850	MD	43123	OR	36470
AZ	31706	ME	34777	PA	35221
CA	38457	MI	38364	RI	36695
CN	41775	MN	40022	SC	32977
CO	41429	MO	35059	SD	29899
DE	37634	MS	27000	TN	30331
FL	30632	MT	28631	TX	33029
GA	33801	NC	34862	UT	37298
HI	42944	ND	30709	VA	38252
IA	34888	NE	33958	VT	33591
ID	34175	NH	39868	WA	36647
IL	39375	NJ	46345	Wash, DC	31811
IN	34759	NM	29592	WI	41082
KS	31911	NV	37845	WV	25431
KY	31552	NY	34707	WY	31707

7. En los locales de venta de electrodomésticos, la radiación emitida por un conjunto televisores encendidos en un área relativamente pequeña puede producir problemas de salud. El siguiente histograma resume medidas de radiación (en miliroentgenes por hora, mR/hr) que fueron tomadas en las 20 principales tiendas de electrodomésticos de Rosario, cada una de las cuales posee por lo menos cinco televisores encendidos en su área de exhibición:

Radiación Medida en las 20 Principales Tiendas de Electrodomésticos de Rosario, Marzo 2009.



- a) Escribe un nombre para el eje X.  
 b) ¿Puedes dar el rango de variación exacto de las medidas? Si puedes, cálcúlalo. Si no puedes, explica por qué.  
 c) El límite de seguridad para la exposición a radiación es de 0.5 mR/hr. ¿Qué porcentaje de locales tienen valores a ese nivel o mayor?  
 d) Si un televisor emite 0.08 mR/hr, ¿cuál es el máximo número de aparatos que puedes poner en tu habitación (y encenderlos) sin exceder el límite de seguridad?

EVALUACIÓN DEL CURSO SEGÚN LA CONDICIÓN DEL MISMO EN EL PLAN ESTUDIOS

Condición	Porcentaje
Bien	16
Regular	38
Mal	15
Total	100

20  
40  
10  
30  
100

- a) Explica qué representa el valor 4 en la tabla anterior.  
 b) Calcula las distribuciones condicionales de la evaluación del curso dado cada tipo de condición del mismo en el plan de estudio.  
 c) Brevemente describe tus resultados de b) – esto es, compara la distribución de la evaluación del curso por estudiantes que lo tomaron como una asignatura requerida, con la distribución de la evaluación del curso por estudiantes que lo tomaron como una asignatura optativa.  
 9. Una compañía de seguros ha examinado un gran número de reclamos y los ha clasificado de acuerdo al tipo de cobertura y de acuerdo a si fue o no un reclamo fraudulento. Los datos se muestran a continuación.

**CLASIFICACIÓN DE RECLAMOS SEGÚN TIPO DE COBERTURA Y CATEGORÍA**

Categoría	Tipo de cobertura	Total
Reclamos	Primeras	16
	Segundas	4
	Terceras	38
	Quartiles	10
	Quintiles	16

a) ¿Qué proporción de reclamos fue fraudulento?

- b) A la compañía le gustaría saber más acerca de la relación entre la categoría del reclamo y el tipo poliza, es decir, desea comparar los porcentajes para cada tipo de poliza. ¿Qué distribuciones condicionales debería calcular la compañía? Calculálas.

10. Un folleto de la Cruz Roja provee la siguiente información acerca de la sangre donada en un período determinado.

**SANGRE DONADA SEGÚN TIPO Y FACTOR**

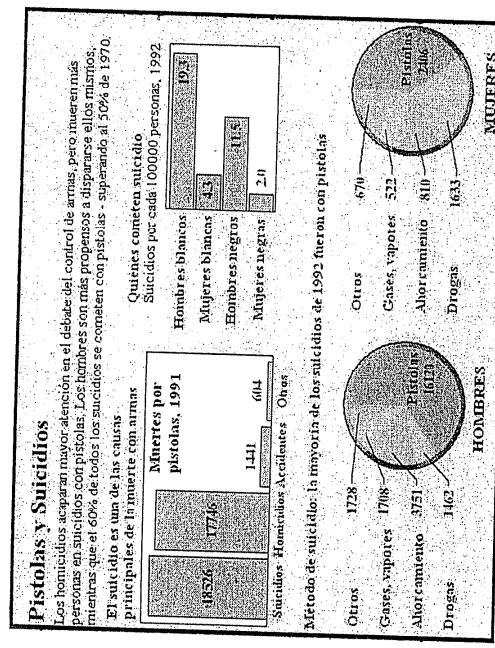
Factor	Rh+	Rh-	Total
O+ (marginal)	45,1%	37,8%	19,2%
O- (marginal)	44,4%	38,9%	11,1%
Total			5,6%

Fuente: Cruz Roja.

Esta tabla presenta (selecciona una):

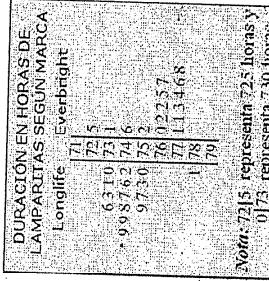
- La distribución marginal del tipo de sangre.
- La distribución marginal de factor Rh.
- La distribución general del tipo de sangre y factor Rh.
- La distribución condicional del tipo de sangre dado el factor Rh.
- La distribución condicional del factor Rh dado el tipo de sangre.

11. En 1991, 18526 de las 38317 muertes ocasionadas por disparos de pistolas en Estados Unidos fueron suicidios, superando los 17746 homicidios. Se dispone de la siguiente información adicional.  
(Fuente: «Silent but Deadly Statistic in Gun-Control Debate: Suicide», USA Today, 4 de abril, 1994)



- a) ¿Qué porcentaje de hombres que cometieron suicidio utilizaron una pistola? ¿Qué porcentaje de mujeres que cometieron suicidio utilizaron una pistola?
- ↳ b) Usa los números que proveen los gráficos para construir la distribución condicional de método de suicidio dado el género.
- c) Realiza el correspondiente gráfico de barras para mostrar la distribución condicional de b).
- d) Comenta dicha distribución.

12. El siguiente diagrama de tallo y hojas muestra la duración en horas de 15 lámparas «Longlife» y 15 lámparas «Everbright».



- a) ¿Cuál es la mayor cantidad de horas que una lámpara haya durado?
- b) Da una buena razón de por qué alguien podría preferir una lámpara Longlife.
- c) Da una buena razón de por qué alguien podría preferir una lámpara Everbright.

13. La tabla que sigue presenta los porcentajes de aterrizajes y despegues en hora en los 22 mayores aeropuertos de Estados Unidos para el cuarto trimestre de 2008.

PORCENTAJE DE ATERRIZAJES Y DESPEGUES EN HORA  
EN LOS 22 MAYORES AEROPUERTOS DE  
ESTADOS UNIDOS CUARTO TRIMESTRE, AÑO 2008

AEROPUERTO	ATERRIZAJES (%)	DESPEGUES (%)
Boston	75.4	83.3
Seattle-Tacoma	75.6	86.6
Los Angeles	76.0	83.9
Chicago	77.0	80.8
Denver	77.0	83.1
Dallas	77.1	84.3
Miami	77.8	88.6
San Francisco	79.6	85.7
Newark	80.1	87.0
Atlanta	81.1	88.1
Minneapolis	81.1	83.8
St. Louis	81.2	85.7
Philadelphia	81.5	86.1
San Diego	82.5	86.5
New York	83.1	88.9
Orlando	84.1	91.3
Washington	84.9	89.9
Dulles Int.	85.8	88.8
Las Vegas	86.0	89.0
Charlotte	86.4	87.6
Detroit	87.9	89.8
Raleigh	90.3	82.6

- a) Realiza un diagrama de tallo y hoja comparativo para estos datos.
- b) Compara las distribuciones.
- c) ¿Los aeropuertos en conjunto cumplen mejor el horario en los despegues o en los aterrizajes? Explica.
- d) Realiza histogramas lado a lado para estos datos.

14. Los siguientes datos corresponden a las ventas (en miles de pesos) del supermercado Red Line en doce meses consecutivos, correspondientes al período Abril 2007-Marzo 2008: 140-131-136-139-141-145-149-148-151-152-151-154.

a) Construye un gráfico de series de tiempo para este conjunto de datos.

- b) Dicho gráfico presenta (selecciona todas las respuestas correctas):
- Una tendencia creciente.
  - Una tendencia decreciente.
  - Ciclos crecientes.
  - Ciclos decrecientes.

15. Los siguientes datos corresponden al tiempo (en segundos) que le lleva a una persona sujetar un componente a una máquina en una línea de montaje en 15 oportunidades consecutivas: 150-152-148-140-141-136-134-130-125-128-125-118-120-110-112.

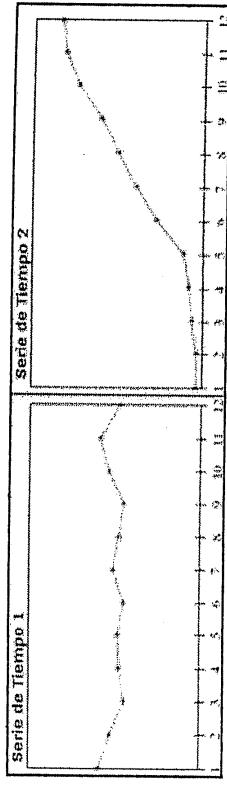
a) Construye un gráfico de series de tiempo para este conjunto de datos.

- b) Dicho gráfico presenta (selecciona todas las respuestas correctas):
- Una tendencia creciente.
  - Una tendencia decreciente.
  - Ciclos crecientes.
  - Ciclos decrecientes.

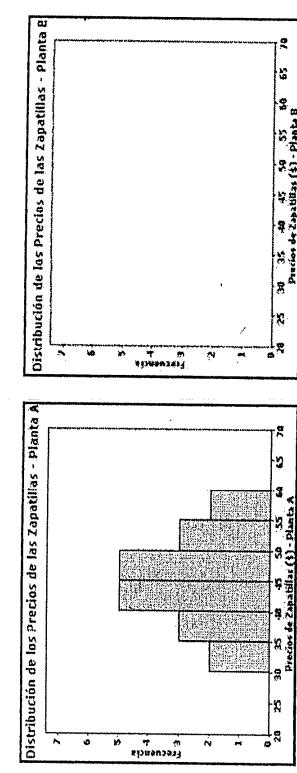
16. Considera los siguientes gráficos de series de tiempo y las dos descripciones. Determina a cuál gráfico hace referencia cada descripción.

**Descripción 1:** Gráfico del número de piezas reparadas por un nuevo operario en sus primeros doce días de trabajo.

**Descripción 2:** Gráfico de la proporción de niñas nacidas cada mes a lo largo del año pasado en los tres hospitales de una ciudad.



17. Las plantas manufactureras A y B producen zapatillas deportivas. El histograma de los 20 precios para los 20 distintos modelos fabricados en la Planta A es el siguiente:



a) ¿Cuántos modelos de zapatillas de la Planta A cuestan \$40 o más?

- b) Los siguientes son los precios para los 20 modelos producidos por la Planta B: 27-37-29-44-31-33-34-46-58-59-32-32-56-57-63-33-56-58-62-51. Presenta Gráficamente los precios.
- c) Selecciona la descripción apropiada para la distribución de los precios de las zapatillas de la planta A. Puedes marcar más de una respuesta:

- i. Simétrica
- ii. Unimodal
- iii. Bimodal
- iv. Asimétrica a la izquierda
- v. Asimétrica a la derecha
- vi. Uniforme

d) Selecciona la descripción apropiada para la distribución de los precios de las zapatillas de la planta B. Puedes marcar más de una respuesta:

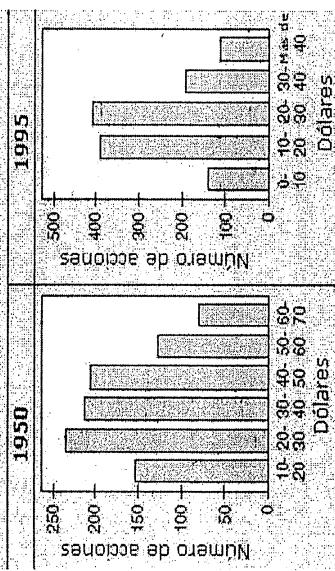
- i. Simétrica
- ii. Unimodal
- iii. Bimodal
- iv. Asimétrica a la izquierda
- v. Asimétrica a la derecha
- vi. Uniforme

e) Un modelo de zapatillas que no se sabe si proviene de la Planta A o de la Planta B tiene un precio de \$55. Queremos probar las siguientes hipótesis:

- $H_0$ : El modelo proviene de la Planta A.  
 $H_1$ : El modelo proviene de la Planta B.
- ¿Cuál es la dirección del test?
  - Encuentra el p-value correspondiente al precio de \$55 observado.
  - Con un  $\alpha=0.15$ , ¿cuál es tu decisión?

18. En el ámbito de la economía, se dice que en las últimas décadas se ha intensificado el fraccionamiento de acciones, lo cual implica que se incrementa el número de acciones como resultado de división de otras en varias partes de menor valor. Por ejemplo, en lugar de poseer una acción por \$80, se poseen dos acciones por \$40. A continuación se presentan dos gráficos que muestran cómo ha cambiado el Mercado de Acciones de Nueva York, en relación al número total de acciones y su valor monetario, entre los años 1950 y 1995.

**DISTRIBUCIÓN DE LOS PRECIOS DE LAS ACCIONES**



Compara las dos distribuciones: ¿lucen muy distintas? ¿parece haber algún cambio notorio en la distribución? Es difícil comparar estas dos distribuciones con exactitud porque los histogramas no tienen la misma escala, y presentan, además otros detalles incorrectos. Produce dos nuevos histogramas de estas distribuciones, realizando todas las correcciones necesarias. Resume cómo ha cambiado la distribución.

19. La siguiente tabla provee el número de estudiantes de una facultad en particular en varias categorías.

ESTUDIANTES SEGÚN SEXO Y AÑO DE CARRERA					
	Sexo	Facultad	Primer año	Segundo año	Tercer año
	Mujeres		1520	1035	876
	Hombres		600	610	950
			1465	1328	1210
			2120		

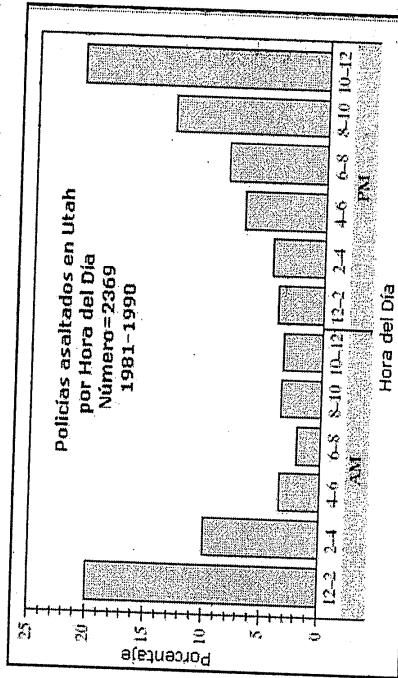
- a) ¿Qué porcentaje de los estudiantes en esta facultad son no graduados?
- b) ¿Qué porcentaje de los estudiantes son hombres no graduados?
- c) Considerando sólo las estudiantes mujeres, ¿qué porcentaje de ellas están en primer o segundo año?
- d) Considerando sólo los estudiantes graduados, ¿qué porcentaje de ellos son hombres?
- e) En general, en esta facultad hay más estudiantes no graduados que graduados. Si cada estudiante pertenece a un y sólo un departamento, entonces podemos concluir que (selecciona una respuesta):
- i. En cada departamento hay más estudiantes no graduados que estudiantes graduados.
  - ii. En cada departamento hay más estudiantes graduados que estudiantes no graduados.
  - iii. En algunos departamentos hay más estudiantes no graduados que estudiantes graduados.
  - iv. Ninguna de las anteriores se puede concluir.
- f) La siguiente tabla nos da (elige una):
- i. El porcentaje de estudiantes mujeres.
  - ii. El porcentaje de estudiantes no graduados y el porcentaje de estudiantes graduados.
  - iii. El porcentaje de estudiantes graduados y no graduados entre todas las estudiantes mujeres.
- |  |              |           |
|--|--------------|-----------|
|  | No graduados | Graduados |
|  | 65,4 %       | 34,6 %    |

20. Una empresa de guarderías infantiles tiene seis centros de cuidado de niños en una ciudad. El dueño le ha pedido a un estadístico que lleve a cabo una encuesta para obtener información sobre la opinión de los padres sobre varios aspectos del cuidado que los chicos reciben. Veinticuatro padres seleccionados al azar pusieron una calificación (0-100 puntos) y estos son los resultados: 81-86-84-41-87-69-76-61-77-79-73-77-80-52-78-67-71-64-56-72-65-71-82
- a) Confecciona un histograma para estas puntuaciones. Usa un límite inferior igual a 40, un límite superior igual a 90 y una amplitud de 5 puntos.
- b) ¿Cuál o cuáles de las siguientes descripciones son apropiadas para la distribución de las calificaciones? *Simétrica, unimodal, bimodal, asimétrica a la izquierda, asimétrica a la derecha, uniforme.*
- c) El estadístico también realizó el siguiente gráfico de series de tiempo del costo mensual del negocio de cuidado de niños en los doce meses del año 2008. ¿Cuál o cuáles de las siguientes descripciones son apropiadas para este gráfico? *Asimétrico a la izquierda, asimétrico a la derecha, tendencia creciente,*

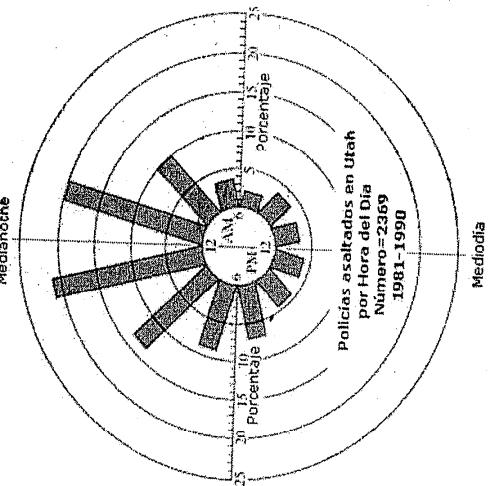
- ii. La distribución condicional de estudiantes de 3º y 4º año dado el sexo.
- iii. La distribución general de sexo y año de facultad.

- c) Describe la forma de la distribución de la edad en 1965, 1995 y la proyectada para 2025.
- d) ¿Cuál es tu opinión acerca de la altura de la barra correspondiente al último grupo etario?

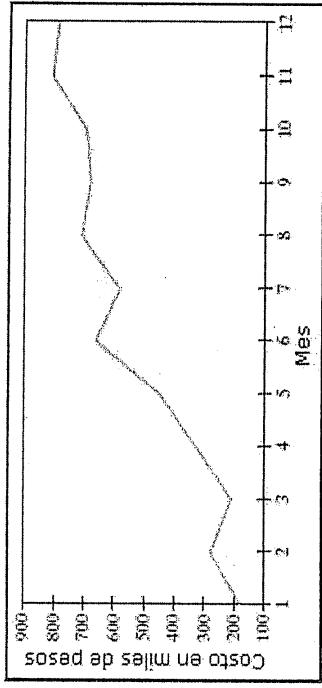
**22. ¿En qué momento del día ocurren más crímenes? Los datos de crímenes reportados**



en una gran ciudad se usaron para construir los siguientes gráficos

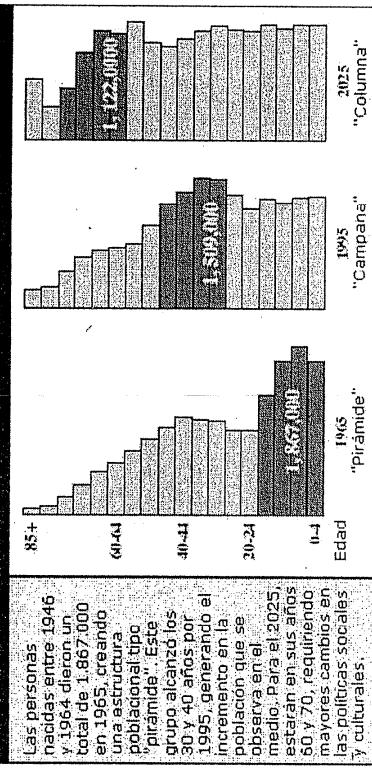


#### COSTO MENSUAL DE LA EMPRESA DE GUARDERÍAS INFANTILES EN EL AÑO 2008



21. Un artículo del 9 de noviembre de 2007 de The Ann Arbor News incluyó los siguientes gráficos para mostrar la distribución de la edad de la población en Michigan para los años 1965, 1995 y 2025 (proyección):

#### GRUPOS ETARIOS - POBLACIÓN DE MICHIGAN



a) El resumen del gráfico dice que «las personas nacidas entre 1946 y 1964 dieron un total de 1.867.000 en 1965». ¿Por qué el primer gráfico para 1965 muestra a estas 1.867.000 personas cubriendo las edades de 0 a 19 años?

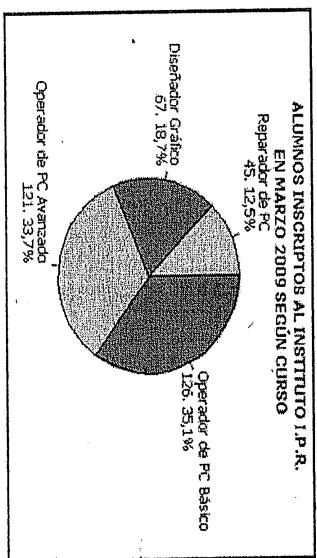
b) El gráfico reporta que los grupos etarios resaltados sólo totalizan 1.509.000 en 1995. ¿Por qué el total es menor que en 1965, pero representa al mismo sector de la población?

a) Escribe un breve resumen de lo que los gráficos muestran sobre los horarios del día en los que los policías de la ciudad fueron asaltados.

b) Discute la eficiencia de ambos gráficos en la exposición de estos datos. Incluye los pros y contras para cada uno. ¿Prefieres alguno de ellos? ¿Por qué?

23. El gráfico de sectores circulares que se expone a continuación representa el total de alumnos ingresantes al instituto I.P.R. en el mes de marzo de 2009, y presenta el porcentaje de los mismos que iniciaron cada uno de los distintos cursos dictados.

- Los distintos sectores del gráfico, representan variables o categorías de una variable?
- Identifica qué significa cada uno de los dos números en cada sector.
- Observando sólo el gráfico sin mirar los números, ¿sería más fácil identificar la moda con este gráfico o con un gráfico de barras? ¿Por qué?



24. ¿Qué países son los más visitados por los turistas? La tabla muestra los resultados de acuerdo a un estudio de la revista «Viajes y Recreación».

- País visitado, ¿qué tipo de variable es?
- Al crear un gráfico de barras para estos datos, ¿sería mejor listar los países alfabéticamente o como un diagrama de Pareto (ordenando las clases desde mayor frecuencia a menor)?

c) ¿Sería adecuado realizar un gráfico de frecuencia o un diagrama de tallo y hoja para estos datos?

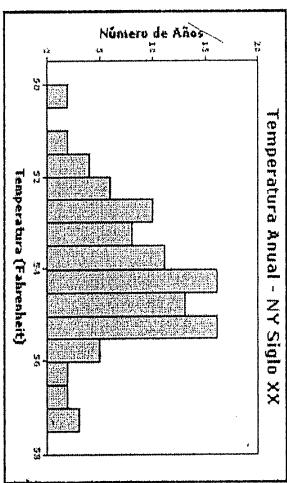
Francia	77.0
China	53.4
España	51.8
Estados Unidos	41.9
Italia	39.8
Reino Unido	24.2
Canadá	20.1
Méjico	19.7

25. E-Bay es una compañía de ventas por Internet muy popular. Considera estos datos de las ventas de la Palm M15 PDA, una computadora de mano, durante mayo de 2003. Los precios finales (en dólares) en los que estos productos fueron vendidos son: 235-225-225-

240-250-250-210-250-249-255-200-199-240-228-255-232-246-210-178-246-240-245-225-246-225

- Construye un diagrama de tallo y hoja.
- Resume qué te dice este diagrama sobre los datos.

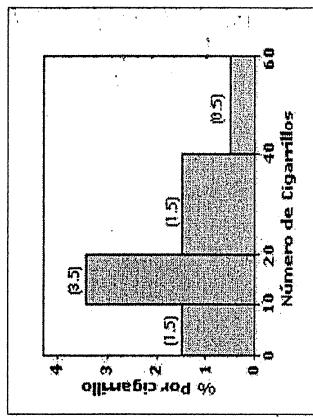
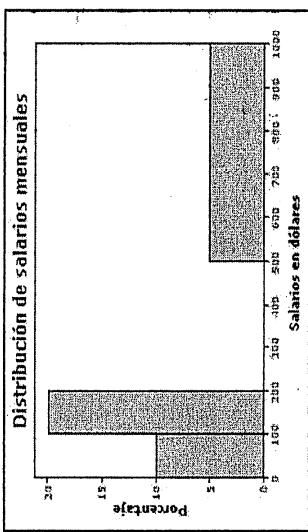
26. La figura muestra un histograma del promedio anual de temperaturas en Nueva York para el siglo veinte.



- a) Describe la forma de la distribución.  
 b) ¿Qué información nos podría dar un gráfico en series de tiempo que este histograma no provee?  
 a) ¿Qué información nos provee el histograma que el gráfico en series de tiempo no lo haría?
27. En cierta institución se tomó una muestra de 59 alumnos y se indagó sobre el tiempo que le dedicaban por día al estudio (en horas) y el sexo, identificado con 0 para los varones y 1 para las mujeres. Los datos se presentan a continuación. Construye un diagrama de tallos y hojas para todos los datos, y luego un histograma para cada sexo, utilizando en el eje vertical la frecuencia relativa porcentual. Describe los gráficos.

Sexo	Horas	Sexo	Horas	Sexo	Horas	Sexo	Horas
1	7	0	2	1	12	1	3
0	2	0	6	1	5	0	3
1	4	1	6	0	4	0	3,5
1	3,5	0	4	1	3	1	4,5
1	4,5	1	3	1	4	0	3,5
0	3	0	4	1	4	1	2
0	3	0	10	1	4	1	3
1	8	1	4,5	1	5	1	6
1	5,5	1	7	1	15	1	10
1	5,5	1	5	1	4,5	1	2
1	5	1	3	0	4	1	3
1	7	1	8	0	4	1	3
1	9	1	5	1	5	0	2
0	4	1	8	1	5	1	2,3
1	10	1	4,5	0	6		

28. El siguiente histograma muestra la distribución de salarios mensuales de un conjunto de empleados a tiempo parcial. Ninguno de estos trabajadores gana más de 1000 dólares por mes. En el dibujo hemos omitido el rectángulo que cubre el intervalo 200-500 dólares. ¿Cuál debería ser aproximadamente la altura de dicho rectángulo?
29. Durante la realización de un estudio por parte del Servicio de Salud Pública, se elaboró un histograma que mostraba el número de cigarrillos que fumaban diariamente un grupo de fumadores (varones) habituales. La densidad viene indicada entre paréntesis. Los intervalos de clase incluyen el límite derecho y excluyen el izquierdo. Contesta estas preguntas:



- a) El porcentaje que fumaba 10 cigarrillos o menos al día se aproxima al:
- 1,5%
  - 15%
  - 30%
  - 50%
- b) Sabiendo que en un paquete hay 20 cigarrillos, el porcentaje que fumaba más de un paquete pero no más de dos al día se aproxima al:
- 1,5%
  - 15%
  - 30%
  - 50%

c) El porcentaje que fumaba más de un paquete al día se sitúa

alrededor del:

- i. 1.5%
- ii. 15%

iii. 30%  
iv. 50%

d) El porcentaje que fumaba más de tres paquetes al día se

aproxima al:

- i. 0.25 del 1%
- ii. 0.5 del 1%
- iii. 10%

e) El porcentaje que fumaba 15 cigarrillos al día se acerca al:

- i. 0.35 del 1%
- ii. 0.5 del 5%
- iii. 1.5%

iv. 3.5%  
v. 10%

### Quiz 1

### Quiz – AUTOEVALUACIONES

Para las preguntas 1 a 10, determinar de qué tipo de variable se trata.

1. El sexo de varios estudiantes en una clase particular.
  - a) Cualitativa
  - b) Cuantitativa discreta
  - c) Cuantitativa continua
  - d) Ninguna de las anteriores
2. El tiempo que duermes cada noche, medido en horas y minutos.
  - a) Cualitativa
  - b) Cuantitativa discreta
  - c) Cuantitativa continua
  - d) Ninguna de las anteriores
3. Lugar de nacimiento de cada persona en un club?
  - a) Cualitativa
  - b) Cuantitativa discreta
  - c) Cuantitativa continua
  - d) Ninguna de las anteriores
4. El número de monedas en las billeteras de los estudiantes de tu clase de estadística.
  - a) Cualitativa
  - b) Cuantitativa discreta
  - c) Cuantitativa continua
  - d) Ninguna de las anteriores
5. La suma de los dos números que salen cuando se arrojan dos dados.
  - a) Cualitativa
  - b) Cuantitativa discreta
  - c) Cuantitativa continua
  - d) Ninguna de las anteriores
6. La vida útil de tu auto, medida en años y meses.
  - a) Cualitativa
  - b) Cuantitativa discreta
  - c) Cuantitativa continua
  - d) Ninguna de las anteriores

7. El peso de un dado legal:
- Cualitativa
  - Cuantitativa discreta
  - Cuantitativa continua
  - Ninguna de las anteriores

8. El color de un dado legal.
- Cualitativa
  - Cuantitativa discreta
  - Cuantitativa continua
  - Ninguna de las anteriores

9. El tiempo de espera entre dos clientes en cierta tienda.
- Cualitativa
  - Cuantitativa discreta
  - Cuantitativa continua
  - Ninguna de las anteriores

10. La calificación de los estudiantes en una clase redondeada al número entero más cercano.
- Cualitativa
  - Cuantitativa discreta
  - Cuantitativa continua
  - Ninguna de las anteriores

11. La calificación de los estudiantes en una clase redondeada al número entero más cercano.

- c) Ambas a) y b)

- d) Ninguna de las anteriores

12. El gráfico anterior es un:

- a) Gráfico de frecuencias

- b) Histograma

- c) Ambas anteriores

- d) Un gráfico de barras

13. En un gráfico de barras horizontales, es importante que las barras tengan:

- a) Igual longitud

- b) Igual altura

- c) Ambas a) y b)

- d) Ninguna de las anteriores

14. En un histograma, la frecuencia para cada clase representada en el eje y es:

- a) Constante

- b) Proporcional al ancho de cada clase

- c) Proporcional al área de cada rectángulo

- d) Ninguna de las anteriores

15. El término estadístico para una observación que se aleja demasiado del resto de los datos es:

- a) Brecha

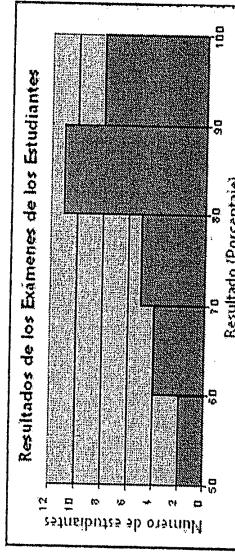
- b) Outlier

- c) Extremo

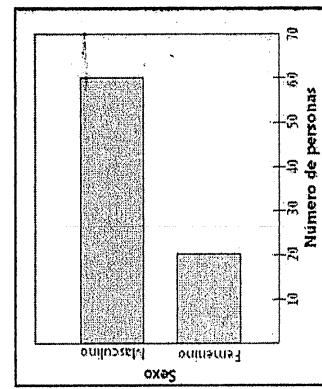
- d) Cluster

## Quiz 2

El histograma de los resultados de todos los estudiantes que tomaron cierto examen se presenta a continuación. Responde las preguntas 1 a 6 en relación al mismo. Nota: las clases incluyen el límite inferior y excluyen el límite superior.



Responde las preguntas 11 y 12 basándose en el siguiente gráfico:

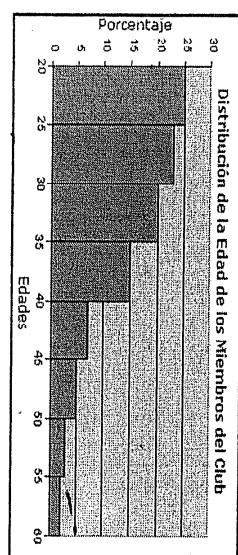


11. El gráfico representa la distribución de una variable:

- a) cualitativa  
b) cuantitativa

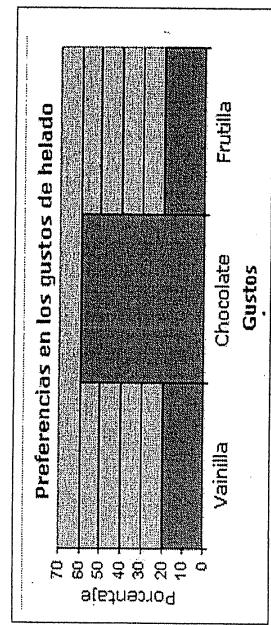
1. ¿Cuántos estudiantes tomaron el examen?
- No se puede determinar a partir del histograma
  - 100
  - 60
  - 30
2. ¿Cuántos estudiantes obtuvieron al menos el 90%?
- No se puede determinar a partir del histograma
  - 8
  - 10
  - 2
3. ¿Cuántos estudiantes obtuvieron exactamente el 90%?
- No se puede determinar a partir del histograma
  - 8
  - 10
  - 2
4. ¿Cuántos estudiantes obtuvieron estrictamente menos del 70%?
- No se puede determinar a partir del histograma
  - 4
  - 6
  - 9
5. ¿Cuál fue el resultado más alto obtenido?
- No se puede determinar a partir del histograma
  - 100
  - 95
  - 90
6. ¿Aproximadamente qué porcentaje de estudiantes obtuvieron al menos 80%?
- Más del 50%
  - Alrededor del 20%
  - Alrededor del 40%
  - Ninguna de las anteriores

A continuación se presenta un histograma con la distribución de la edad de todos los 300 miembros en un club. Responde las preguntas 7 a 12 acerca del mismo. Nota: las clases incluyen el límite inferior y excluyen el límite superior.



7. El ancho de las clases del histograma es:
- No se puede determinar a partir del histograma
  - 40
  - 5
  - 300
8. La distribución es:
- Asimétrica a la derecha.
  - Simétrica
  - Asimétrica a la izquierda.
  - Ninguna de las anteriores
9. ¿Cuántos miembros son estrictamente menores de 25?
- 50
  - 25
  - 75
  - Ninguna de las anteriores
10. ¿Cuál es la edad del mayor miembro del club?
- 60
  - 55
  - 57
  - No se puede determinar a partir del histograma
11. ¿Qué porcentaje de los miembros del club tienen por lo menos 50 años?
- Aprox. 0.67%
  - Aprox. 1.67%
  - Aprox. 1%
  - Ninguna de las anteriores
12. ¿Cuántos miembros tienen por lo menos 50 años?
- 12
  - 15
  - 5
  - Ninguna de las anteriores

Responder las preguntas 13 a 15 en base a la información sobre las preferencias de los clientes de una heladería en cuanto a los gustos de helado, que se muestra en el gráfico siguiente:



## RESUMEN DE DATOS EN FORMA GRÁFICA

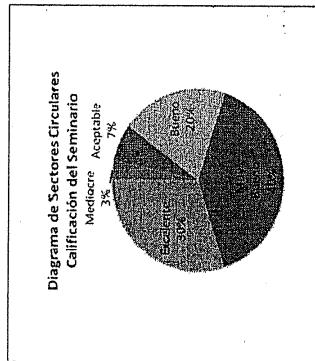
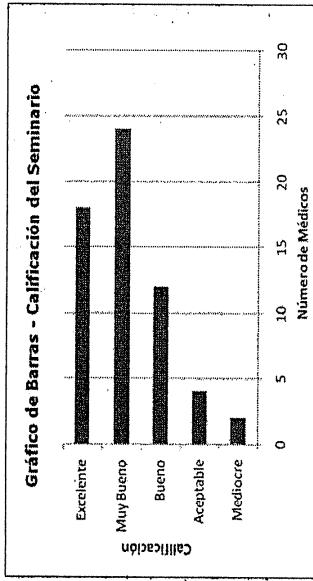
### Resolución de Ejercicios ADICIONALES

1.      a) Cualitativa  
       b) Cuantitativa continua  
       c) Cuantitativa discreta  
       d) Cuantitativa continua  
       e) Cuantitativa discreta  
       f) Cuantitativa continua  
       g) Cuantitativa continua  
       h) Cuantitativa discreta  
       i) Cualitativa  
       j) Cuantitativa continua  
       k) Cuantitativa continua

2.      a) Cualitativa  
       b)

13. El gráfico anterior es un:

- a) Gráfico de barras  
    b) Histograma  
    c) Diagrama de sectores circulares  
    d) Diagrama de tallo y hoja
14. El gráfico es:
- a) Asimétrico a la derecha.  
    b) Simétrico  
    c) Asimétrico a la izquierda.  
    d) Ninguna de las anteriores
15. La variable en el eje X es:
- a) Cualitativa  
    b) Cuantitativa discreta  
    c) Cuantitativa continua  
    d) Ninguna de las anteriores



3.

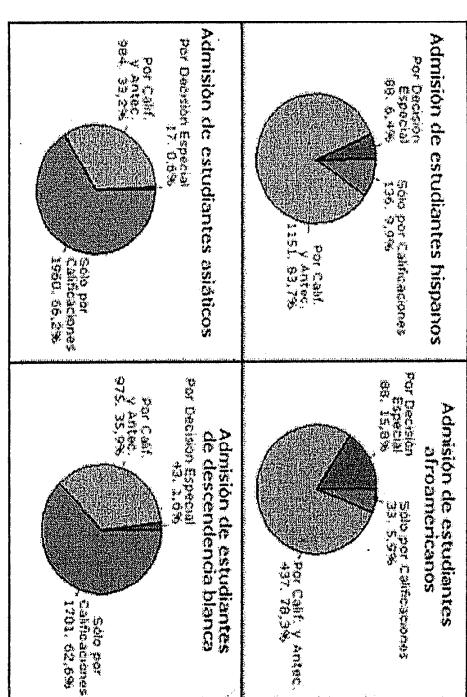
- a) Aproximadamente el 69% de los entrevistados dijeron que se automedian si están resfriados o con gripe.  
 b) Aproximadamente 385 (22%) encuestados no hacen nada si les duele la espalda.  
 c) El porcentaje de encuestados que recurre al médico cuando tienen dolor de espalda es 22%, mientras que el 9% lo hace cuando tienen dolor de cabeza.

4.

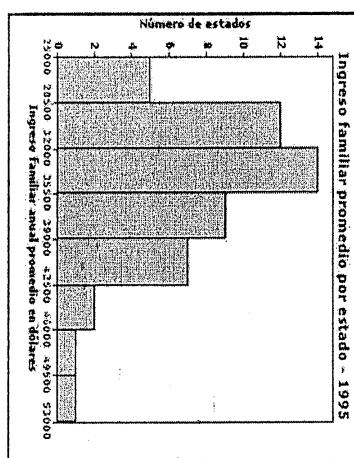
- a) 124 oficiales  
 b) 279 oficiales  
 c) No podemos hablar de simetría o simetría en el caso de la distribución de los antecedentes laborales porque ésta es una variable cualitativa, y el orden de sus categorías carece de sentido

	Solo Calif.	Calif. Y Antec.	Dec. Especial	Total
Blancos	1701	975	43	2719
Africanos	33	437	88	558
Hispanos	136	1151	88	1375
Asiáticos	1960	984	17	2961
Otros	349	219	15	583
Total	4179	3766	251	8196

5)



6)



7)

- a) Eje X: Radiación en mR/hr  
 b) No podemos determinarlo porque un histograma no nos muestra los valores reales observados sino sólo una representación de los valores dentro de cada clase o los porcentajes dentro de cada clase.  
 c) 50%  
 d) Se puede tener hasta 6 televisores encendidos.

8. a) El valor 4 significa que 4 de todos los estudiantes que tomaron el curso lo hicieron de forma optativa y lo calificaron como Mediocre.

b)

Clasificación		Bueno	Excelente
Razón	Requerida	22.86%	22.86%
	Optativa	13.33%	33.33%

c) Aproximadamente la mitad de los estudiantes que tomaron el curso porque así lo requería su plan de estudio lo calificaron como bueno, mientras que la mitad de los estudiantes que lo hicieron de forma optativa lo calificaron como excelente. De estos alumnos, la tercera parte lo calificó como bueno, y sólo el 13.33% dijo que fue mediocre, en contraste para quienes la materia era requerida, de los cuales el 22.86% lo hizo así. También el 22.86% de los que tenían el requisito de cursar la materia la consideró excelente.

9.

- a) 0.16  
b) La distribución de la categoría del reclamo dado el tipo de póliza

Categoría	Tipo de reclama			Total
	Fuego	Auto	Otros	
Fracturado	30%	3.33%	6%	
No fracturado	70%	96.67%	94%	

10. iv. La distribución condicional del tipo de sangre dado el factor Rh.

11.

Sexo	Métodos de Suicidio				Total
	Pistola	Dioses	Ahorc.	Gases	
Hombres	16190	1462	3751	1708	24769
Mujeres	9406	1633	810	522	6041
Total	18596	3095	4561	2330	30810

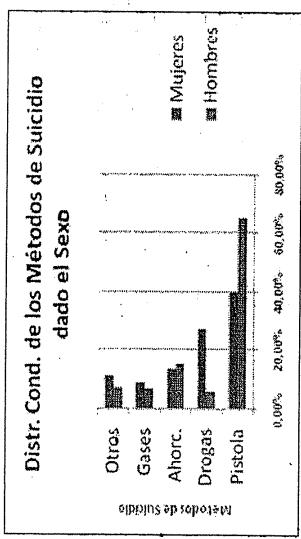
- a) Porcentaje de hombres que se suicidaron con una pistola: 65.08%

- Porcentaje de mujeres que se suicidaron con una pistola: 39.83%

b)

Sexo	Distr. Cond. Dados Métodos de Suicidio dado el Sexo				Otros
	Pistola	Dioses	Ahorc.	Gases	
Hombres	65,08%	5,90%	15,14%	6,90%	6,98%
Mujeres	39,83%	27,03%	13,41%	8,64%	11,09%

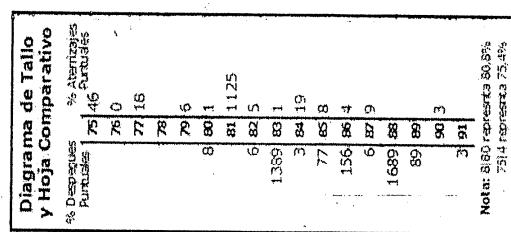
c)



- d) Para ambos sexos, el mayor porcentaje de los suicidios se cometió con una pistola, aunque este porcentaje es bastante mayor en el caso de los hombres que en el de las mujeres. En el caso de ella, las diferencias entre los porcentajes para cada método de suicidio no es demasiado grande, mientras que para los hombres, los porcentajes de suicidios cometidos con gases, ahorcamientos o drogas parecen ser mucho menores en comparación con el porcentaje de suicidios con pistola.

12.

- a) 781 horas.  
b) Porque la duración de estas lamparitas presenta menor variabilidad.  
c) Porque hay un conjunto de lamparitas Everbright que duraron más que muchas de las Longlife.

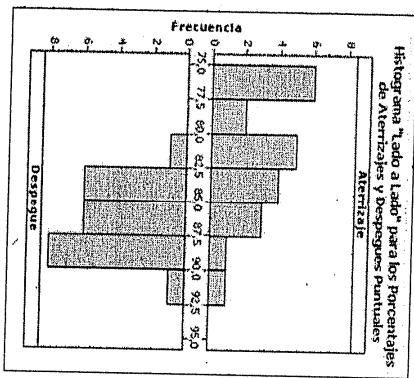


13.

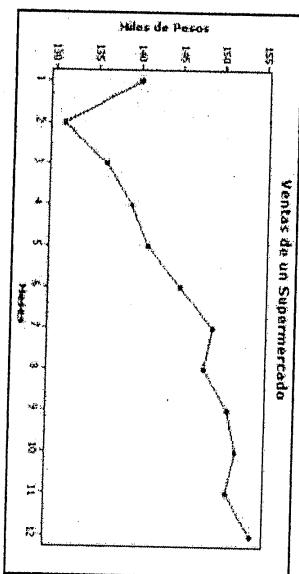
- b) La distribución del porcentaje de aterrizajes puntuales presenta mayor dispersión en los datos que la del porcentaje de despegues puntuales, la cual presenta mayor concentración en porcentajes más altos, principalmente entre el 83 y 88%.

C) Aparentemente, los aeropuertos en general cumplen mejor los horarios de despegue

a)



14.

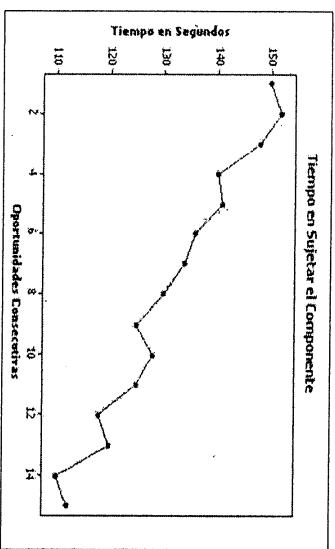


- b) Respuesta correcta: i - Tendencia creciente

212

15.

a)

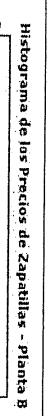


- b) Respuesta correcta: ii - Tendencia decreciente

16. La Descripción 1 corresponde al Gráfico 2; y la 2 al 1.

17.

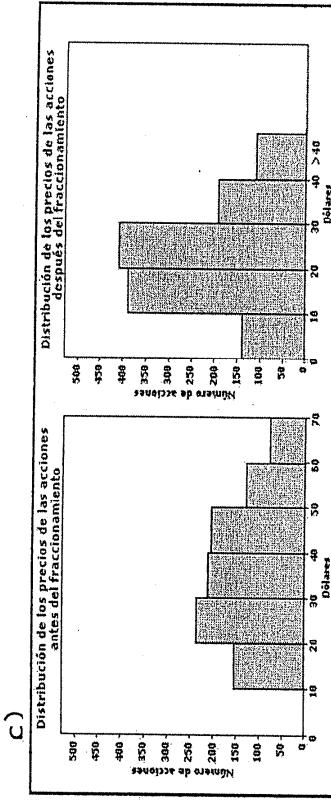
- a) 15 modelos



- a) i - Simétrica  
 b) i - Simétrica; iii - Bimodal  
 c) i. Es un test bilateral. ii.  $P\text{-value}=0.2$ . iii. No se rechaza  $H_0$ .

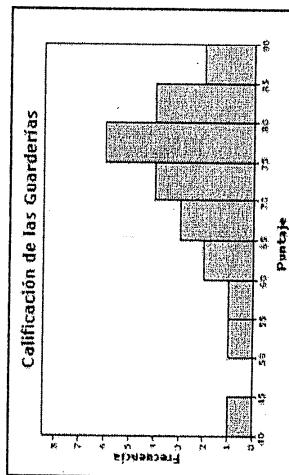
213

- 18.
- a) En 1925, los porcentajes en las distintas clases no difieren mucho unos de otros, exceptuando la clase +100, la cual presenta el mayor porcentaje de acciones. En 1960 se observa que la distribución es asimétrica a la derecha, como en 1994, pero en este año se observa una asimetría más pronunciada.  
 b) A lo largo de los años se observa que cada vez hubo más acciones con menores precios. Esto sustenta la idea expresada en el artículo.



- 19.
- a) 71.57%    b) 37.4%    c) 36.34%    d) iii    e) f) iii

- 20.
- a)



- b) Unimodal, asimétrica a la izquierda.  
 c) Tendencia creciente.

Notas: 17.8 representa 178 dólares

Precios de Venta de las Palms	
17	8
18	9
19	0
20	00
21	00
22	598
23	25
24	00669
25	00055

- 21.
- a) Porque las personas nacidas entre 1946 y 1964 son las que en 1965 tenían entre 0 y 19 años.  
 b) El total es menor porque integrantes de esos grupos estaban en 1965 pudiendo haber muerto para 1995; o quizás también el saldo migratorio para esas edades en esos años fue negativo en aquella ciudad.  
 c) Para 1965, la distribución de la población es asimétrica a la derecha. Esta asimetría es menos notoria en 1995 y desaparecería en la distribución de 2025, la cual parece asemejarse a una distribución uniforme.

- 22.
- a) Los gráficos muestran que se registraron un mayor número de delitos entre las 8 pm y las 4 am.  
 b) La ventaja del gráfico circular es que se puede apreciar mejor el lapso de tiempo en los que se registraron más delitos, de manera continuando, permitiendo observar claramente este lapso correspondiente a la noche. Su desventaja es que no es claro para comparar las distintas clases, y que al ser un tipo de gráfico muy común, se requiere un esfuerzo extra para entenderlo e interpretarlo. La ventaja del histograma es la facilidad para hacer comparaciones y la claridad con la que se presenta la información.

23.

- a) Es una variable cualitativa.  
 b) Es indistinto, pues no existe un orden natural para los niveles de la variable; de todos modos, la disposición al modo de un diagrama de Pareto facilita la visualización de los datos y la comparación entre los distintos niveles.  
 c) No se debe realizar esos diagramas porque corresponden a variables cuantitativas.

- 24.
- a)

b) Este diagrama nos habla de una distribución asimétrica. a la izquierda

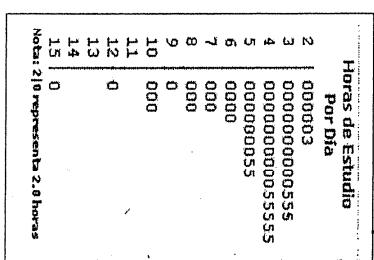
25.

a) A simple vista, el gráfico parece indicar una leve asimetría a la izquierda.

b) Un gráfico en series de tiempo nos podría dar información en la variación de la temperatura a lo largo de los años, si presentó una tendencia creciente o decreciente, si se produjeron ciclos con el correr del tiempo, o si no hubo ningún patrón particular en el comportamiento de esta variable.

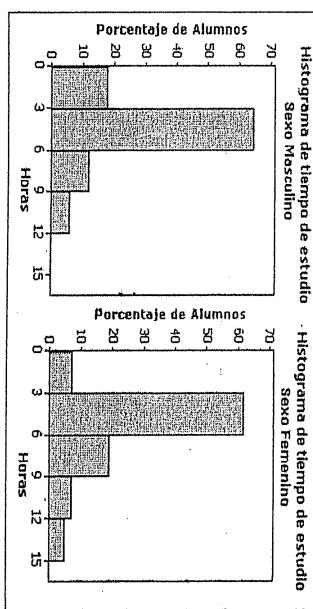
c) El histograma nos dice qué intervalos de temperatura fueron los más frecuentes a lo largo del siglo.

26. Diagrama de tallo y hojas:



Notas: 2|0 representa 2.0 horas

Histogramas para cada sexo:



Se puede observar que la distribución es asimétrica a la derecha, y que al hacer la distinción de género, las distribuciones siguen siendo asimétricas a la derecha y muy similares.

27.

a)

28. El porcentaje para el intervalo 200-500 debe ser del 15%.

Quiz - autoevaluaciones

Quiz 1

1. A ✓

2. C ✓

3. A ✓

4. B ✓

5. B ✓

6. C ✓

7. C ✓

8. A ✓

9. C ✓

10. B ✓

11. A ✓

12. D ✓

13. B ✓

14. C ✓

15. B ✓

Quiz 2

1.	D ✓
2.	B ✓
3.	A ✓
4.	C ✓
5.	A ✓
6.	A ✓
7.	C ✓
8.	A ✓

9. B  
 10. D  
 11. B  
 12. C  
 13. A  
 14. D  
 15. A

### TÉRMINOS CLAVES



- UNIDAD. OBSERVACIÓN. VARIABLE
- CONJUNTO DE DATOS
- VARIABLES CUANTITATIVA. VARIABLE CUANTITATIVA.
- DISCRETA Y CONTINUA.
- ESCALA DE MEDICIÓN
- DISTRIBUCIÓN DE UNA VARIABLE
- GRÁFICO DE SECTORES. GRÁFICO DE BARRAS
- GRÁFICO DE FRECUENCIA. DIAGRAMA DE TALLO Y HOJA SIMPLE Y COMPARATIVO
- DISTRIBUCIÓN DE FRECUENCIA ABSOLUTA Y RELATIVA
- DISTRIBUCIÓN DE FRECUENCIA RELATIVA ACUMULADA
- SIMETRÍA / ASIMETRÍA. UNIMODAL / BIMODAL
- HISTOGRAMA
- DIAGRAMA DE SERIE DE TIEMPO. TENDENCIA. VARIACIÓN: ESTACIONAL Y CÍCLICA

### RECORDAR



- HISTOGRAMAS CON INTERVALOS DE CLASES DE DISTINTA AMPLITUD
- SUGERENCIAS DE TABLAS Y GRÁFICOS. EDA