

## Trabajo Práctico N°1 Estadística descriptiva

### Presentación de datos

Luego de recopilar y organizar los datos provenientes de una investigación, los mismos pueden presentarse recurriendo a tablas y gráficos.

### **Tablas**

Se emplean para presentar datos en forma ordenada. Toda tabla debe contener:

- a- Título
- b- Encabezado, contiene el contenido de las columnas que conforman la tabla.
- c- Concepto o columna matriz, indica los títulos de las filas, para tablas de doble entrada.
- d- Cuerpo, es el contenido, la superficie en la cual se vuelcan los datos
- e- Notas al pie, corresponden a todas las aclaraciones referidas a los datos
- f- Fuente, indica la procedencia de la información

CEOP (Centro de Estudios de Opinión Pública) ha elaborado por segunda vez el ranking de marcas (el primero fue realizado en el año 2001). Para ello la empresa pidió a 1300 personas que dijeran, en forma espontánea, las marcas más prestigiosas que le venían a la mente, una modalidad llamada *top of mind*. Los resultados fueron expuestos en el diario Clarín el 23 de enero de 2005. Algunos de los resultados se exponen a continuación.

### **Hiper-supermercados**

2001	2004	Empresa	Puntaje
1	1	Carrefour	54.21
5	2	Coto	50.18
4	3	Disco	45.29
2	4	Jumbo	42.40
6	5	Norte	41.19
*	6	Libertad	40.85
*	7	Plaza Veá	39.01
3	8	Wal-Mart	35.65
*	9	La Anónima	29.47
*	10	La Gallega	22.12
*	11	Topsy	16.45

Fuente: CEOP

\*No ingresó entre las primeras marcas en el 2001

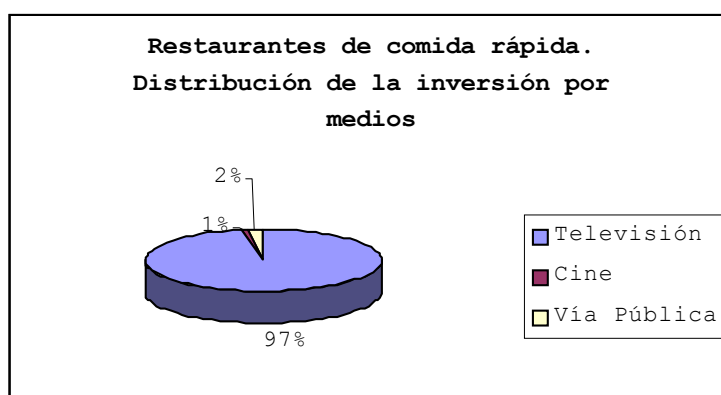
## Gráficos

Los gráficos también son una herramienta empleada en la presentación de los datos. Los mismos permiten una representación visual de la información. Existen muchos tipos de gráficos. La elección de alguno determinado depende del dato y el concepto a transmitir. Vamos a mencionar los más usuales en las ciencias de la comunicación.

### Datos cualitativos

- 1- **Gráficos circulares o diagramas de pastel:** Muestran la cantidad o el porcentaje de datos que pertenecen a una categoría determinada como una parte proporcional de un círculo.

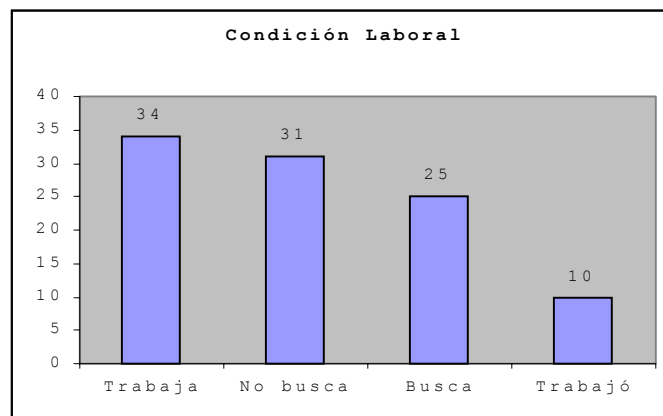
También puede ser empleado para datos cuantitativos discretos



Fuente: EGM (Estudio General de Medios)

- 2- **Gráficos de barras:** Muestran la cantidad de datos que pertenecen a cada categoría como áreas rectangulares de tamaño proporcional. El principio para la construcción de estos gráficos reside en que los rectángulos deben tener el mismo ancho, mientras que su altura marcará la medida establecida

Una investigación realizada por Orígenes y presentada en el Diario Clarín en enero de 2005, analiza las dificultades de los jóvenes para acceder a su primer empleo.

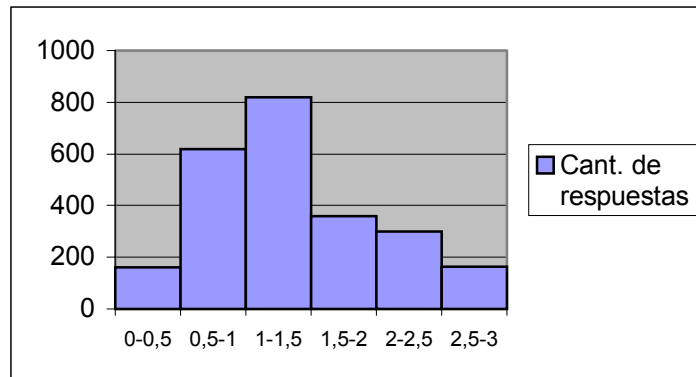


Fuente: Orígenes

### Datos cuantitativos

**Histograma:** Es un gráfico de barras que representa una distribución de frecuencias, y esto es un listado que asocia a cada valor de variable su frecuencia determinada. La diferencia que existe con el gráfico de barras está en que el histograma no presenta espacio entre barras y se emplea frecuentemente para representar distribuciones de frecuencias agrupadas

Se realizó una encuesta para conocer la distribución de los ingresos de las personas con sueldos de hasta \$3000 de la ciudad de Buenos Aires. Los datos se presentan en miles de pesos.



**Ejercitación**

- 1) En función al gráfico de barras presentado en la página 2,
  - a) ¿Qué porcentaje de encuestados está trabajando o ha trabajado?
  - b) ¿Considera que dicho gráfico presentado en el diario Clarín es claro y fue armado de manera correcta? Justifique.
- 2) Proponga un gráfico que mejore la lectura de la información que figura en la tabla de la página 1. Justifique su elección.
- 3) Con base en el gráfico de sectores circulares de la página 2, exprese en el lenguaje coloquial lo que en el gráfico se observa como: “97% Televisión”, ¿Es posible conocer la cantidad de dinero destinado a TV?. Justifique.
- 4) Los siguientes datos corresponden a la estructura de la demanda de los hogares correspondientes a una localidad del norte de Brasil.

<b>RUBRO</b>	<b>Facturación (miles de U\$S)</b>
Productos alimenticios	12.789
Bebidas no alcohólicas	5.670
Bebidas alcohólicas	4.530
Limpieza e higiene	2.679
Limpieza para el hogar	2.240
Infusiones	2.123
Golosinas	1.432
Indumentaria	1.143
Cosmética capilar	1.102
Prod. De librería	1.100

Fuente: EGM

- a) ¿Cuántas variables hay en este conjunto de datos? Clasifique cada una de ellas.
  - b) ¿Qué porcentaje de rubros tiene un facturación superior a los tres millones de dólares?
- 5) Durante el mes de Diciembre la contratación de locales para realización de eventos aumenta el 300%, esto significa que:
    - La cantidad de gente que contrata se multiplica por 300.
    - La cantidad de gente que contrata se multiplica por 3.
    - La cantidad de gente que contrata se multiplica por 4.
    - Ninguna de las anteriores.
  - 6) El salario de una persona que aumentó 50% actualmente es de \$300. Esto significa que anteriormente era de:
    - \$100
    - \$150
    - \$200
    - Ninguna de las anteriores.

- 7) Si una cantidad se duplica significa que aumentó:
- 20%
  - 2%
  - 50%
  - 100%
- 8) Si una cantidad se reduce a la mitad significa que disminuyó:
- 20%
  - 2%
  - 50%
  - 100%
- 9) Si a una cantidad se le resta el 50% y al resultado se lo aumenta en un 50%, ¿se obtiene nuevamente la cantidad original?
- 10) Una empresa discográfica realizó una encuesta de 11 preguntas a sus clientes más importantes. Algunas de ellas fueron:
- i) ¿Cuántos álbumes (CDs, cassettes, DVDs) compró usted en los últimos 12 meses?
  - ii) ¿Es actualmente miembro de un club nacional de pedidos por correo? (Si/No)
  - iii) ¿Qué edad tiene?
  - iv) Incluyéndose usted. ¿cuántas personas viven en su casa?
  - v) ¿Qué tipo de música le gusta comprar? (rock, contemporánea, heavy metal, electrónica, clásica, rap, country, etc)

Indique en cada pregunta si los datos que se piden son cualitativos o cuantitativos.

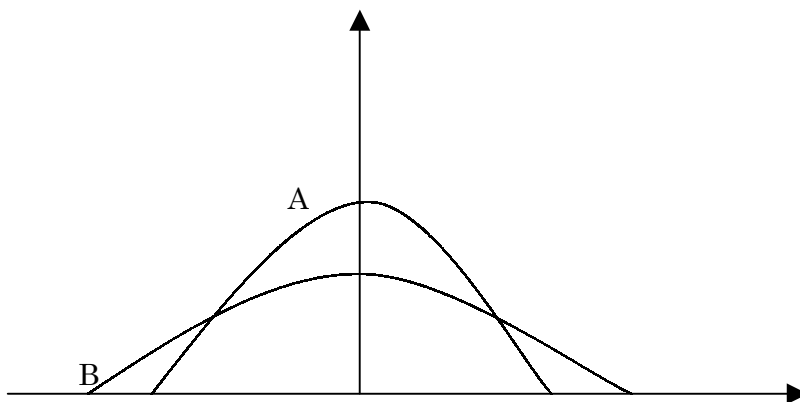
- 11) La Oficina de Turismo de Bariloche recopila datos sobre los turistas que ingresan al lugar. Entre las 16 preguntas de un cuestionario otorgado a los pasajeros de una compañía aérea, figuran:
- i) Este viaje a Bariloche es mi: 1,2,3,4.....
  - ii) El motivo principal de este viaje es: (se presentan 12 categorías entre las cuales figuran, vacaciones, trabajo, estudio, etc)
  - iii) Me pienso hospedar en: (10 categorías entre las cuales figuran hotel, cabaña, campamento, parientes, etc)
  - iv) Días de estadía previstos.
- a) Indique cuál es la población bajo estudio.
  - b) ¿Considera que el empleo de un cuestionario es un buen método para contactar a los pasajeros en vuelos de llegada?. Justifique.
  - c) Indique cuáles son las variables abordadas en el cuestionario y clasifíquelas.

- 12) Clasifique las siguientes variables como cualitativas, cuantitativas discretas o cuantitativas continuas.
- i) Sexo.
  - ii) Costo anual derivado de regalos empresariales.
  - iii) Ingreso mensual.
  - iv) Marca de gaseosas que consume el target.
  - v) Cantidad de dinero mensual destinado a insumos de computación.
  - vi) Edad de personas que concurren a un gimnasio entre 18hs y 21hs.
  - vii) Número de suscripciones a una revista de deporte.
  - viii) Cantidad de quejas diarias.
  - ix) Cantidad de Revistas de tecnología.
  - x) Tipo de regalo empresarial.
- 13) Indique cuáles de las siguientes proposiciones son de naturaleza descriptiva y cuáles de naturaleza inferencial:
- a) El promedio de altura de las personas observadas es de 1,68 metros.
  - b) El 35 % de las personas encuestadas consume diarios y revistas.
  - c) El 35% de los argentinos consume diarios y revistas.
  - d) El edad más frecuente entre los lectores del diario “La Nación” es de 47 años
- 14) Una editorial está analizando el consumo mensual de dos de sus revistas para la ciudad de Córdoba. La revista “El deporte” tiene un consumo promedio de 12.000 revistas con un desvío de 5.000 revistas, mientras que la revista “Salud física” tiene un consumo promedio de 25.000 revistas con un desvío de 9.000 revistas, ¿cuál de las dos revistas tiene un consumo más parejo?
- 15) Se desea comparar la variabilidad de los sueldos de una compañía con sucursales en Argentina, Francia y Brasil, ¿qué medida de resumen propone para esta labor? Justifique.
- 16) La encuesta de usuarios de zapatillas para 2003, entre 100 usuarios arrojó los siguientes datos:
- Nike acaparó el 56% del valor del mercado.
  - La cifra promedio anual invertida en zapatillas por consumidor es de 250\$.
  - En una escala de 1 (muy descontento) a 5 (muy satisfecho), el nivel promedio de satisfacción para con la calidad de zapatillas fue de 3,73.

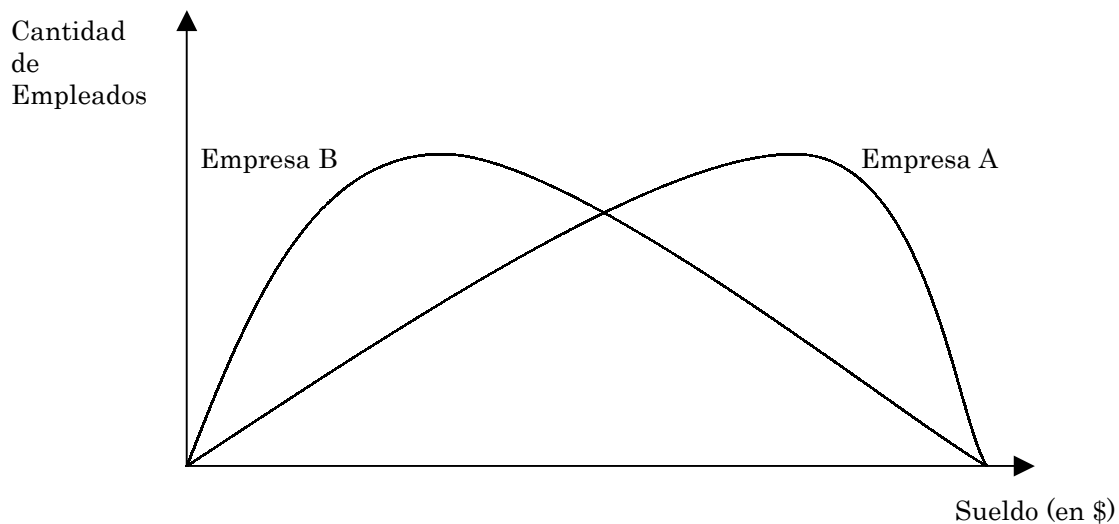
Cite dos medidas descriptivas derivadas de la investigación

17) Contestar V ó. F para las dos figuras que se exponen a continuación . Justifique.

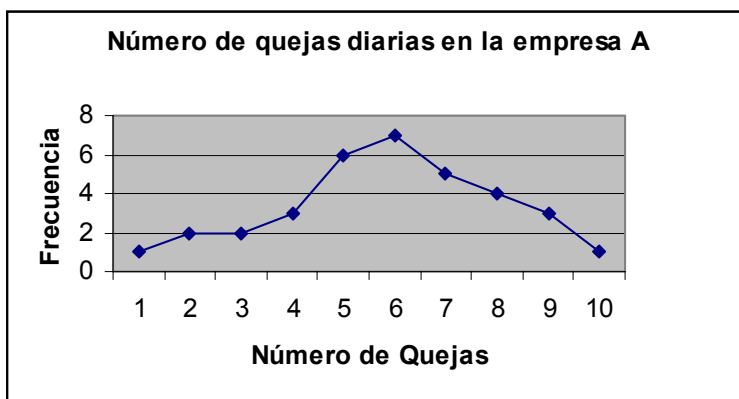
- a) La distribución A tiene mayor promedio que la B.
- b) La distribución A tiene mayor desvío que la B.
- c) La distribución A es más homogénea que la B



18) Los siguientes gráficos representan la distribución sueldos de dos empresas competidoras, si a usted le ofrecieran un puesto en las dos, ¿en qué empresa aceptaría el cargo y porqué?



- 19) Se tomaron datos referidos al número de quejas diarias por mala atención recibidas en dos empresas de telefonía celular. Se diseñaron los siguientes gráficos utilizando la información relevada:



Determine cuál o cuáles de las siguientes afirmaciones es correcta:

- Los datos de ambas empresas tienen la misma media.
- La distribución de la empresa B tiene menor desvío estándar.
- El coeficiente de Variación de ambas distribuciones es igual.

- 20) Sin utilizar calculadora, resuelva:

- a)  $\frac{30 + 50}{10} \cdot 2$
- b)  $\frac{30}{10} + 50 \cdot 2$
- c)  $\left( \frac{30}{10} + 50 \right) \cdot 2$



- 21) Los siguientes datos corresponden a la cantidad de quejas efectuadas por un grupo de asociados a una compañía de seguros durante la primera semana del mes de marzo:

29, 27, 28, 20, 21, 29.

- Defina y clasifique la variable.
- Calcule el número promedio de quejas.
- Calcule el número más frecuente de quejas.
- Estudie la representatividad del promedio.
- Halle la queja mínima de la mitad de los días que más quejas presentaron.
- Grafique la distribución.

- 22) Los siguientes datos corresponden a la cantidad de suscripciones a una revista tecnológica durante 17 días:

43, 54, 36, 37, 38, 40, 38, 54, 42, 33, 39, 45, 43, 28, 27, 45, 49

- Defina y clasifique la variable.
- Calcule el número promedio de suscripciones.
- Calcule el número más frecuente de suscripciones.
- Estudie la representatividad del promedio.
- Halle la suscripción máxima de la mitad de los días que menos suscripciones se realizaron.
- Repita los ítems b, c, d, e, f usando excell. (Optativo)

- 23) Los siguientes datos corresponden a las compras realizadas de una materia prima para los últimos 3 meses. Un administrador pidió información sobre el costo promedio por kilogramo de la materia prima. Hallar el costo promedio mencionado, empleando los siguientes datos:

Compra	Costo,\$/ kilogramo	Cantidad de libras
1	3.00	2200
2	3.40	400
3	2.80	1500
4	2.90	1200
5	3.25	800

- 24) El promedio de calificaciones para los alumnos de una Universidad Americana se basa en el cálculo de media ponderada. En la mayoría de estas instituciones se asignan los siguientes valores a las calificaciones: A (4 puntos), B (3 puntos), C (2 puntos), D (1 puntos), F (0 punto). Luego de acreditar 68 horas en curso un alumno ha obtenido 10 horas de A, 9 de B, 43 de C y 2 de D.

- Hallar la calificación promedio para el alumno
- Los alumnos de un curso de una Universidad con este sistema deben tener un promedio de 3 puntos en sus primeras 68 horas para pasar al año siguiente, ¿pasará el alumno mencionado?

25) En una encuesta de imagen para candidato a Intendente se contemplaron los siguientes ítems con la ponderación correspondiente:

Formación profesional (7 puntos), obras realizadas a la comunidad (9 puntos), antecedentes positivos (5 puntos), gestión y desarrollo de proyectos (5 puntos).

El candidato A recolectó 30 votos para la formación profesional, 15 para las obras realizadas a la comunidad, 40 para antecedentes positivos y 10 para gestión y desarrollo.

- Hallar el puntaje promedio para el candidato A
- Si usted quisiera privilegiar la formación del candidato y el impulso hacia nuevos proyectos, ¿diseñaría la misma puntuación?

26) Los siguientes datos corresponden a las edades de un grupo de 50 personas que asistieron a un espectáculo.

26 25 23 24 25 25 26 22 21 23  
 25 20 20 21 28 20 29 24 26 27  
 20 21 20 25 29 24 22 27 20 28  
 23 28 27 21 21 21 22 25 27 29  
 24 28 22 20 23 20 28 23 24 25

- Defina y clasifique la variable
- Complete la siguiente tabla

x: valor	f. frecuencia absoluta	fr. Frecuencia relativa	F. Frecuencia absoluta acumulada.
----------	------------------------	-------------------------	-----------------------------------

27) Complete la siguiente tabla agrupando los datos del ejercicio 26 en 5 intervalos.

li - ls	f. frecuencia absoluta	F. Frecuencia absoluta acumulada	PM. Punto medio del intervalo
---------	------------------------	----------------------------------	-------------------------------

28) Con los datos del ejercicio 26

- Calcule medidas de resumen con los datos sin agrupar y con los datos agrupados.
- Interprete cada medida en términos del problema.
- ¿Existe diferencia significativa entre las medidas calculadas en los datos sin agrupar y en los datos agrupados? Justifique.
- ¿Cuál es a su criterio la medida más representativa? Justifique.
- La agencia de publicidad encargada de publicitar la película dirigirá la campaña de promoción al segmento etario 24-26, ¿le parece acertada esta decisión? Justifique.

29) Los siguientes datos corresponden a los tiempos de espera (en días) de personas que iniciaron su trámite para adquirir el DNI.

14	11	23	21	16	19	22	25	11	16
25	24	35	19	16	19	18	19	21	12
16	17	18	33	25	29	23	16	30	19
24	26	15	22	34	20	22	24	22	20

- Construya intervalos de clase y realice la tabla de frecuencias absolutas simples.
- Realice un histograma y un polígono de frecuencias.
- Halle el tiempo promedio de espera para la realización del trámite.
- Suponiendo que el empleado de mesa de entradas es nuevo en su puesto y cada persona tardará 1 día más en el trámite, ¿cuál será el nuevo promedio de espera?
- Halle la cantidad de días más frecuente que una persona debe esperar para recibir su DNI.
- Halle la cantidad de días de espera superada por la mitad de las personas.
- Halle la cantidad de días que supera al 15% de las personas.
- Halle la cantidad mínima de días del 30% de las personas que más esperaron.
- Calcule CV e interprete en términos del problema.
- Halle el porcentaje de personas que esperaron más de 20 días.

30) En una encuesta de suscriptores a una revista tecnológica se preguntó lo siguiente: “¿cuántos de los últimos 4 números ha leído usted?”. ¿Cuál de las dos distribuciones corresponde al conteo de 500 respuestas? Justifique.

Tabla A

Cant. leída	0	1	2	3	4
frecuencia	15	10	40	85	350

Tabla B

Cant. leída	0	1	2	3	4	5
frecuencia	10	25	35	90	120	220

Para la distribución que corresponde:

- ¿Cuál es la media de la cantidad de revistas que lee un suscriptor de revistas tecnológicas?
- Calcule CV e interprete.
- Determine la asimetría de la distribución. Interprete en términos del problema.

- 31) A continuación se proporcionan las edades un grupo de lectores de la revista Cosmopolitan, con el objetivo de analizar la personalidad del medio.

x(edad)	10-20	20-30	30-40	40-50	50-60
f(lectores)	10	14	17	7	2

- Defina y clasifique la variable.
  - Determine la forma de la distribución interpretando en términos del problema. Justifique.
  - Interprete en términos del problema media, mediana y modo.
  - Con el objetivo de cambiar la personalidad del medio la gerencia de la revista decidió ofrecer una promoción al 13% de los lectores más adultos, ¿a partir de que edad el lector recibirá la promoción?
  - Calcule medidas de dispersión e interprete.
- 32) Sea la variable: horas semanales dedicadas a ver TV de lunes a domingo de 15 a 22hs; para estudiarla se han tomado dos muestras distintas de dos targets correspondientes a las ciudades de Córdoba y Buenos Aires.

Córdoba

x	4	7	10	15	20	25
f	1	8	7	7	10	6

Buenos Aires

x	5	8	12	15	20	30
f	6	10	9	7	2	2

- Interprete cuál es la información que, en el cuadro correspondiente a la ciudad de Córdoba muestra:  $x:7$ ,  $f:8$ .
  - Si tuviera que poner una publicidad de una marca en la TV de alguna de las dos ciudades, ¿cuál elegiría y por qué?
- 33) Los siguientes datos corresponden a los tiempos de auditoría (en horas) que se emplearon en el control de 20 empresas del rubro alimenticio.

x: tiempo de auditoria	10-15	15-20	20-25	25-30	30-35
f: cant. de empresas	4	8	5	2	1

- Calcule la cantidad promedio de horas que dura una auditoría en el rubro indicado.
- ¿Es representativo el promedio? Justifique.
- Al 10% de las empresas en las que menos duró la auditoría, se les otorgará un certificado de habilitación por 2 años, mientras que el 10% de las empresas en las que más tiempo duró la auditoría recibirán un certificado por sólo 6 meses. ¿Cuánto tiempo debe durar la auditoría para que una empresa reciba un certificado por 2 años, y cuánto para que una empresa reciba el certificado por sólo 6 meses?
- Determine la asimetría de la distribución e interprete en términos del problema.

- 34) Un productor cinematográfico realiza un preestreno de una película e interroga a los espectadores para averiguar sus reacciones. Por grupo de edades obtiene los siguientes resultados, donde  $x$  representa las edades y  $f(x)$  la frecuencia de personas a las que les gustó la película.

x (edad)	10-20	20-30	30-40	40-50
f(x)	15	13	4	2

- Al 18% de las personas más jóvenes se les pedirán sugerencias para la realización de un comercial de TV que presente la película, ¿entre qué edades sucederá esto?
- Al 10% siguiente se les otorgará un 15% de descuento, ¿entre qué edades recibirán tal descuento?
- El Director afirma que las personas más jóvenes son las que más se interesaron por la película, es cierta esta afirmación?. Justifique con el cálculo de medidas de resumen.
- Con el propósito de comparar estos resultados con los de su película anterior, decide estudiar los siguientes datos:

x (edad)	10-20	20-30	30-40	40-50
f(x)	12	10	9	3

¿En cuál de las dos películas la recepción de la misma fue más homogénea?, justifique.

- 35) En una compañía cervecera se procesaron semanalmente las siguientes cantidades de cerveza blanca en los últimos nueve meses. Se obtuvieron los siguientes datos:

x (litros)	100-120	120-140	140-160	160-180	180-200
f (cant. de semanas)	3	8	10	9	6

- Defina y clasifique la variable.
- Halle la cantidad de litros promedio de cerveza blanca procesada.
- Halle la cantidad de litros de cerveza producida que supera a la producción del 20% de las semanas en que menos cerveza se produce.
- La fábrica de cervezas está analizando la cantidad de avisos de cervezas que se registraron en 30 revistas del mes de junio. Se obtuvieron los siguientes datos:

x	0	4	10	15
f	3	15	6	6

- Defina la variable
- Interprete:  $x:0$   $f:3$
- Sugiera y calcule las medidas de posición necesarias para analizar la comunicación de las empresas de cerveza en base a los datos expuestos.

- 36) Completar el siguiente cuadro correspondiente a la variable  $x$ : cantidad de dinero (en pesos) que un grupo de personas destina a las fiestas navideñas.

$x$ : Pesos	$f$ : nro de personas. Frecuencia absoluta simple	$F$ : Frecuencia absoluta acumulada	$fr$ : frecuencia relativa simple
100-200		18	
200-300			0,20
300-400	12		
400-500		70	
500-600			
600-700		100	0,08

Con los datos hallados:

- Calcule el gasto promedio y analizar su representatividad.
  - Halle el monto más usual que las personas gastan en las fiestas.
  - Halle el monto que no es superado por el 20% de las personas encuestadas.
  - Grafique la distribución, indicando en el gráfico modo, mediana y media. Clasifique la asimetría.
- 37) Una empresa dedicada a la comercialización de carteles luminosos sabe que el precio promedio de las gigantografías es de 25.000\$ (el circuito mensual) con un coeficiente de variación de 0,08%. Para el mes previo al día de la madre decide hacer un descuento del 5% a todos los precios de las distintas gigantografías.
- Halle el nuevo precio promedio, el desvío y el coeficiente de variación.
  - Efectúe los mismos cálculos para un descuento fijo de 1500\$.
- 38) Considerando una muestra de 5 elementos, dos de los cuales son 50 y 70. Agregue los tres elementos faltantes de forma tal que:
- La media sea 70.
  - La media sea 60.
  - La mediana sea 85.
  - El modo sea 40.
  - La media sea 50 y la mediana 60.
  - La mediana sea 70 y el modo 75.
- 39) Elabore un pequeño informe (no más de 5 renglones) teniendo en cuenta los siguientes datos.
- Una consultora obtuvo los siguientes datos de la variable: cantidad de quejas diarias recibidas en una empresa de TV por cable (los datos se tomaron durante 45 días)
- Promedio: 23  
 Modo: 15  
 Mediana: 20  
 CV: 0,26
- La consultora propone un plan para disminuir la cantidad de quejas diarias en un 3%, recalcule promedio y CV.

40) Se desea realizar un estudio acerca de las calificaciones obtenidas por alumnos de primer año en distintas universidades.

En la Universidad “Sapiens” se tomó la siguiente muestra y luego del análisis se obtienen las medidas:

$$\bar{x} = 6,48 \text{ puntos} \quad \sigma = 2,43 \text{ puntos}$$

Si uno de los alumnos que dijo haber obtenido un 5 (cinco) en realidad tuviera una calificación de 4(cuatro), determine cuál o cuáles de las siguientes afirmaciones es correcta, justificando su respuesta:

- La media de la muestra aumenta.
- La media de la muestra disminuye.
- La media de la muestra no se ve afectada.
- No puede saberse qué ocurrirá con la media sin calcularla.
- La mediana de la muestra aumenta.
- La mediana de la muestra disminuye.
- La mediana de la muestra no se ve afectada.
- No puede saberse qué ocurrirá con la mediana con los datos del ejercicio.
- El modo de la muestra aumenta.
- El modo de la muestra disminuye.
- El modo de la muestra no se ve afectada.
- No puede saberse qué ocurrirá con el modo con los datos del ejercicio.