ALGUNAS TÉCNICAS DE MEDICIÓN ESTADÍSTICA
(FERRIS RITCHEY).

- La estadística trata de observar y organizar información numérica sistemáticamente adquirida. La *información sistemáticamente adquirida que se organiza siguiendo los procedimientos de la ciencia y la estadística se llama* **dato** *o* **datos**.
- El análisis estadístico implica precisión, debe serguir procedimientos y realizar mediciones confiables sobre cómo ocurren los eventos en el mundo.
- Por eso, los **errores estadísticos** no son equivocaciones. La expresión “error estadístico” se refiere al *grado conocido de imprecisión en los procedimientos utilizados para reunir y procesar información*.

- La **estadística descriptiva** informa cuántas observaciones fueron registradas y qué tan frecuentemente ocurrió en los datos cada puntuación o categoría de observaciones.
- Por ejemplo: datos tomados de 291 encuestados muestran que 40 por ciento son varones y tienen una edad promedio de 21 años, siendo el más joven de 19 años y el más viejo de 51.
- En ciencias, la estadística descriptiva es utilizada como primer paso en el análisis de hipótesis de investigación científica.
- La **estadística inferencial** procura mostrar relaciones de causa-efecto, así como probar hipótesis y teorías científicas.

- La **Población o Universo** es el grupo gran de personas de interés particular que deseamos estudiar y entender.
- La **Muestra** es un subgrupo pequeño de la población. La muestra se observa y se mide y después se utiliza para obtener conclusiones sobre la población.
- Para adquirir información completamente correcta sobre una población entera, mediríamos todos sus miembros y resumiríamos los resultados en términos matemáticos, reportando porcentajes, tasas y promedios. Al cálculo resumido de mediciones realizadas en todos los sujetos de una población se le llama **parámetro**. El porcentaje de mujeres en cargos directivos en la Universidad es un parámetro. El promedio de edad de los internos en una cárcel es un parámetro.
- Las muestras sólo proporcionan **estimaciones** de los parámetros. La única manera de conocer un parámetro verdadero es mediante el sondeo de la población entera.

- Algunas muestras están más cerca del valor del parámetro verdadero que otras. Hay dos fuentes de error en el muestreo:
- **El tamaño de la muestra:** se refiere al número de casos u observaciones que constituyen una muestra: el número de personas u objetos observados.
- Cuanto mayor sea la muestra, mayor será el rango del error. Para lidiar con esta fuente de error se debe calcular el "intervalo de confianza", con una cantidad exacta de error para cualquier tamaño de muestra dada. Por ejemplo: en una muestra de 1000 estudiantes encontraríamos que la edad promedio en el campus es de 22,4 años, más o menos 0,3 años. El "más o menos" se basa en probabilidades matemáticas resultado de la teoría de la probabilidad.
- **La representatividad de la muestra:** una muestra es representativa cuando en ella están incluidos todos los segmentos de la población y en las proporciones correctas. Por ejemplo, si la población en el campus es realmente 54% hombres y 46% mujeres, la muestra tendrá que acercarse a esos porcentajes para ser representativa.
- El diseño de muestreo más común es la **muestra aleatoria simple**, en la cual cada persona (u objeto) de la población tiene la misma oportunidad de ser seleccionado(a) para formar parte de la muestra.

CONTROL DEL ERROR DE MEDICIÓN

- No sólo debemos evitar los errores de muestreo. También hay que definir con precisión cómo se harán las mediciones y cómo se codificarán las respuestas una vez que se recopilen los datos.
- El conjunto de procedimientos y operaciones para medir una variable se llama **definición operacional**.
- Por su parte, el **nivel de medición de una variable** identifica las propiedades de medición de la variable y determina el tipo de operaciones matemáticas que puede usarse apropiadamente con dicho nivel, así como las fórmulas estadísticas que utiliza para probar las hipótesis teóricas.

CARACTERÍSTICAS DE LOS CUATRO NIVELES DE MEDICIÓN

Nivel de medición	Ejemplos	Cualidades	Operaciones matemáticas permitidas
Nominal	Género, raza, preferencia religiosa, estado civil.	Clasificación en dos categorías; denominación de las categorías.	Conteo del número de casos (o sea, frecuencia) de cada categoría de la variable, comparación de tamaños de categorías.
Ordinal	Rango de clase social, preguntas de actitud y opinión.	Clasificación de categorías; ordenamiento de rangos de categorías de bajo a alto.	Todo lo anterior más juicios de mayor que y menor que, y cálculos de diferencias y promedios de rangos.

CARACTERÍSTICAS DE LOS CUATRO NIVELES DE MEDICIÓN

Nivel de medición	Ejemplos	Cualidades	Operaciones matemáticas permitidas
Intervalo	Temperatura, índices resumidos, escalas de actitud y opinión.	Todo lo anterior más distancias entre puntuaciones con una unidad fija de medida.	Todo lo anterior más operaciones matemáticas como suma, resta, multiplicación, división y raíces cuadradas.
Razón	Peso, ingresos, edad, escolaridad, tamaño de población.	Todo lo anterior con un punto cero real.	Todo lo anterior, más el cálculo de razones significativas.

EJEMPLO

- La siguiente tabla presenta diversas variables nominales que se transformaron en una variable de nivel de razón llamada índice de comportamiento de riesgo de salud. Las variables nominales se registran como cero para *no* y 1 para *sí*. Esto se llama codificación prototipo porque las puntuaciones numéricas son artificiales; 0 y 1 n o se usan para distinguir montos o cantidades. En cambio, cero significa que el factor de riesgo no está presente y 1 que sí está presente. La variable de razón ***RIESGO*** es el número total de factores de riesgo de un individuo y esta variable tiene un punto cero real. Veamos:

EJEMPLO DE ÍNDICE PARA TRANSFORMAR VARIABLES NOMINALES EN UNA VARIABLE DE RAZÓN.

Número y nombre de la variable	Definición operacional (cómo se mide la variable)	Nivel de medición	Código (cómo se registra).
1. Fuma	Es fumador habitual	Nominal	0 = no; 1= sí.
2. Alcohólico.	Ha consumido cinco o más bebidas alcohólicas en el último mes.	Nominal	0 = no; 1= sí.
3. Ejercicio	Su estilo de vida es sedentario.	Nominal	0 = no; 1= sí.
4. Drogas	Ha usado drogas ilícitas en el último mes.	Nominal	0 = no; 1= sí.
5. Condebr	Ha conducido en estado de ebriedad	Nominal	0 = no; 1= sí.
6. Riesgo	Número de conductas de riesgo que reportó.	Razón	Suma de respuestas "sí" para las variables 1 a 5.

MEDIDAS DE TENDENCIA CENTRAL

- Una medida de tendencia central proporciona una estimación de la puntuación típica, común o normal encontrada en una distribución de puntuaciones.
- Una medida de tendencia central sirve para:
  - Mostrar en qué lugar se ubica un individuo promedio o típico.
  - Comparar o interpretar cualquier puntaje en relación con el puntaje central o típico.
  - Comparar el puntaje obtenido por una misma persona en dos momentos diferentes.
  - Comparar los resultados medios de dos o más grupos.
- Las medidas de tendencia central más comunes son:
  - La media.
  - La mediana.
  - La moda.

LA MEDIA

- Es la suma de todas las puntuaciones dividida entre el número de puntuaciones observadas (o sea, el tamaño de la muestra). Para calcularla, sumamos todas las puntuaciones y dividimos el resultado entre el tamaño de la muestra, así:

$$\bar{x} = \frac{\text{suma de todos los valores}}{\text{cantidad total de datos}} = \frac{x_1 + x_2 + x_3 + x_4 + \ldots + x_n}{N}$$

En el siguiente ejemplo se calcula la media de las calificaciones obtenidas por un estudiante en la clase de matemáticas:

$$\bar{x} = \frac{4 + 7 + 7 + 2 + 5 + 3}{6} = \frac{28}{6} = 4{,}8$$

CÁLCULO DE LA MEDIA COMBINADA DE DOS MUESTRAS DE TAMAÑO DIFERENTE

- Cuando tenemos las medidas de los datos de dos grupos de tamaño diferente podemos cometer el error de combinarlas sumándolas y dividiendo el resultado entre 2. Por ejemplo: tenemos el número medio de días de vacaciones por año (X) para el grupo 1, las ocho secretarias de un banco local y para el grupo 2, los tres vicepresidentes. Para las ocho secretarias:

$$\bar{X}(grupo\ 1) = \frac{\sum X\ (grupo\ 1)}{n\ (grupo\ 1)} = \frac{7+10+7+12+16+7+14+10}{8}$$

$$= \frac{83\ días}{8} = 10.38\ días\ de\ vacaciones.$$

Para los tres vicepresidentes:

$$\bar{X}(grupo\ 2) = \frac{\sum X\ (grupo\ 2)}{n\ (grupo\ 2)} = \frac{60+30+30}{3}$$

$$= \frac{120\ días}{3} = 40{,}00\ días\ de\ vacaciones.$$

- Si calculamos mal la media de la oficina completa sumando estas dos medias y dividiendo entre 2, obtendríamos la respuesta errónea de 25,19 días de vacaciones. El cálculo correcto para esta media combinada es:

$$\bar{X}(grupos\ y\ 2\ combinados) = \frac{\sum X\ (grupo\ 1) + \sum X\ (grupo\ 2)}{n\ (grupo\ 1) + n\ (grupo\ 2)}$$

$$= \frac{83+120}{8+3} = \frac{203}{11} = 18{,}45 \text{ días de vacaciones}$$

EJERCICIO

- Supongamos X = calificación en examen final. La calificación media para los 13 estudiantes del último año del grupo es 87 y la calificación media para los 16 estudiantes del penúltimo año es 79. ¿Cuál es la calificación media para los dos grupos combinados?

LA MEDIANA

- La Mediana (Mdn) es la puntuación central en una distribución ordenada, es decir, el valor de una variable que divide en mitades a la distribución de las puntuaciones; la puntuación por encima de la cual queda la mitad de los casos y por debajo de la cual queda la otra mitad.
- Por ejemplo, si la media del ingreso familiar en una ciudad es de USD$26 000, la mitad de las familias de esa ciudad tienen ingresos mayores a USD$26 000 y la otra mitad ingresos menores a USD$26 000.
- La mediana es un punto de localización, la puntuación de la mitad.
- La puntuación mediana también es igual al percentil 50, o sea, el punto bajo en el cual caen el 50% de las observaciones.
- Cuando una distribución de datos tiene pocas puntuaciones hacia un lado, la mediana es una medida muy útil. Por ejemplo, la mediana del precio de las ventas recientes de viviendas es preferible a la media del precio, porque unas cuentas ventas de alto precio incrementarían el valor de la media.

¿CÓMO CALCULAR LA MEDIANA?

- Primero debemos ordenar las puntuaciones para una variable X, es decir, las puntuaciones deben colocarse en orden de tamaño, de menor a mayor o de mayor a menor.
- Después debemos dividir entre 2 el tamaño de la muestra *n* para acercarnos a la puntuación de la mitad de la distribución. Si *n* es un número impar, la mediana será un caso real en la muestra.
- Supongamos, por ejemplo, que tenemos una muestra de cinco familias con los siguientes ingresos mensuales (X):

Orden de los casos	1	2	3	4	5
Valores de X	\$3 540	\$4 675	\$7 350	\$9 860	\$19 000

- La mediana de ingreso es de USD \$7350, el valor de X para la tercera puntuación ordenada.

- Si *n* es un número par, la mediana se localiza entre las dos puntuaciones de la mitad y se calcula tomando la media de esas dos puntuaciones.

Orden de los casos	1	2	3	4	5	6
Valores de X	\$3 540	\$4 675	\$7 350	\$9 860	\$19 000	\$20 000

Mdn = \$8 605

- Con una muestra grande, la mediana si sitúa matemáticamente al dividir entre 2 el tamaño muestra y sumar 0.5. Este resultado nos dirá cuál es la ubicación ordenada de la mediana, no la mediana en sí. Para hallar esta última tendríamos que ordenar las puntuaciones y luego contar hasta esa posición. La puntuación X de esa posición es la mediana.

EJERCICIO

- **Hallar la mediana** de la siguiente serie de datos
  4, 5, 2, 7, 5, 9, 5, 2, 8.

- Para **obtener la mediana** primero debemos ordena la serie:
  2, 2, 4, 5, 5, 5, 7,8, 9.
- A continuación determinamos el **número que queda en el centro de la muestra**, en este problema los datos son impares, y la Mdn = 5.

EJERCICIO

- **Hallar la mediana** de la siguiente serie de datos
  2,5,4,8,9,6,4,3.

- Lo primero que hacemos es **ordenar los datos**.

  Observamos que **la muestra es par**, por lo que sumaremos los dos números centrales y los dividiremos entre dos.

  2,5,4,8,9,6,4,3.

  2,3,4,**4,5**,6,8,9.

  4+5=9; 9/2= 4,5

LA MODA

- La Moda (Mo) es la *puntuación que se presenta con más frecuencia en una distribución*. Conceptualmente, la moda es la puntuación "más popular".
- La siguiente tabla muestra la distribución de edades para una muestra de estudiantes universitarios. La moda es 19 años porque la mayoría de las personas (49 de ellas) tiene esa edad. Obsérvese que la moda es una puntuación X (19 años), *no* una frecuencia, *f* (49 casos).

Edad	Frecuencia	Porcentaje %
18	31	24,8
19	49	39,2
20	20	16,0
21	18	14,4
22	7	5,6
Total	125	100,0

- **ADVERTENCIA:** No se debe confundir la moda (la “puntuación que ocurre con mayor frecuencia”) con “la mayoría de las puntuaciones”. Una mayoría simple sería “más de la mitad” o 50 por ciento de los casos de una más uno, por lo menos. Obsérvese que, en el ejemplo anterior, aunque la puntuación que ocurre más frecuentemente es 19 años, la mayoría de la muestra no tiene 19 años; sólo 39,2 por ciento de la muestra tiene esa edad. Ninguna edad de esta distribución tiene mayoría.

- Para obtener la moda en datos agrupados se usa la siguiente fórmula:

$$M = L_i + \left(\frac{D_1}{D_1 + D_2} \right) A_i$$

Donde:

L_i = L-inferior de la clase modal.

D_1 = es el delta de frecuencia absoluta modal y la frecuencia absoluta premodal.

D_2 = es el delta de frecuencia absoluta modal y la frecuencia absoluta postmodal.

A_i = Amplitud del intervalo modal

EJEMPLO

- **Calcular** la **moda** de una distribución estadística que viene dada por la siguiente tabla:

	f_i
[60, 63)	5
[63, 66)	18
[66, 69)	42
[69, 72)	27
[72, 75)	8
	100

$$M = L_i + \left(\frac{D_1}{D_1 + D_2}\right) A_i$$

Donde:

L_i = L-inferior de la clase modal.

D_1 = es el delta de frecuencia absoluta modal y la frecuencia absoluta premodal.

D_2 = es el delta de frecuencia absoluta modal y la frecuencia absoluta postmodal.

A_i = Amplitud del intervalo modal

$$Mo = 66 + \frac{(42-18)}{(42-18)+(42-27)} \cdot 3 = 67.846$$

EJERCICIO

Calcular la moda de una distribución estadística que viene dada por la siguiente tabla:

$$Mo = 20 + \frac{2}{2+3} \cdot 5 = 22$$

	f_i
[10, 15)	3
[15, 20)	5
[20, 25)	7
[25, 30)	4
[30, 35)	2

$$M = L_i + \left(\frac{D_1}{D_1 + D_2} \right) A_i$$

Donde:

L_i = L-inferior de la clase modal.

D_1 = es el delta de frecuencia absoluta modal y la frecuencia absoluta premodal.

D_2 = es el delta de frecuencia absoluta modal y la frecuencia absoluta postmodal.

A_i = Amplitud del intervalo modal

VOLVAMOS SOBRE EL PROCESO DE INVESTIGACIÓN

- Identificar una situación problemática
- Revisar la literatura sobre el tema.
- Especificar una pregunta de investigación.
- Formular una hipótesis.
- Volver sobre la literatura para operacionalizar las variables.
- Elaborar un diseño de investigación:
  - Cómo se hará el muestreo.
  - Cómo se medirán las variables.
  - Cómo se reunirán los datos.
- Recolectar los datos.
- Organizar y analizar los datos.
- Elaborar conclusiones.
- Divulgar los resultados.

PENSAMIENTO PROPORCIONAL: PROPORCIONES, PORCENTAJES Y TASAS.

- El pensamiento proporcional nos permite sopesar la parte contra el todo y calcular la probabilidad de que, a la larga, ocurra el fenómeno.
- Las **proporciones matemáticas** son problemas de división que comparan una parte (el numerador) con el todo (el denominador). Se expresan mediante una fracción, así:

$$\frac{Numerador}{Denominador}$$

- **Ejemplo:** La población total de una cárcel es 149 internos. 112 de ellos fueron acusados de delitos relacionados con las drogas. ¿Es esta una gran parte de la población de la cárcel?

$$P = \frac{112}{149} = 0{,}75417$$

Este cociente se entiende mejor, si lo presentamos como un porcentaje. Para eso, multiplicamos el cociente por 100 (o sea, movemos el punto decimal dos lugares a la derecha. El resultado sería 75,17%.

- Otra forma de estandarizar es calculando una **tasa**, la frecuencia de que ocurra un fenómeno en relación con un número base especificado para los sujetos de una población.
- El número base se coloca en el denominador, para que la tasa represente los casos por mil, por cien, por diez mil, por cien mil, etc. **Ejemplo:**
- **En el año 2000, en EEUU, 51 personas murieron por rayos. La proporción (*p*) fue:**

$$\frac{51}{281\,421\,906} = 0{,}00000018$$

Expresado en porcentaje, obtenemos un 0,000018%. Esta cifra sigue siendo difícil de entender. Entonces calculamos la tasa de muertes por rayos para una población de 10 millones, o sea, tomamos la cifra de la proporción y movemos el punto 7 lugares a la derecha, así: 0,00000018 x 10 000 000 = 1,8.

La casa de muertes por rayos, para cada 10 millones de habitantes, es de 1,8 (suponiendo, claro, que el año 2000 haya sido un año típico.

- Completen los espacios en blanco en la siguiente tabla:

Fracción	Proporción	Porcentaje
$\frac{51}{207}$		
$\frac{24}{503}$		
$\frac{663}{13200}$		

EJERCICIOS

- Ustedes están interesados en realizar un proyecto de investigación que comprende niveles de logros educativos en Ibagué. El DANE indica que, para el año 2018, la población en Ibagué que tiene 25 años o más, es de 379 556. completen la siguiente tabla insertando la proporción (p) y el porcentaje de esa población que haya completado diversos niveles de educación.

Máximo nivel educativo alcanzado	***n***	***p***	**%**
Básica primaria	15663		
Básica secundaria	28619		
Media académica	105812		
Pregrado universitario	108442		
Especialización	27213		
Maestría	61196		
Doctorado	32611		
Total	379556		

- Como parte de una investigación sobre viviendo urbana en la ciudad de Cali, deben examinar las edades de los propietarios de casas en toda la ciudad. Hay 1 656 053 unidades ocupadas en la ciudad. Completen la siguiente tabla con la proporción (*p*) y el porcentaje de propietarios de casas en cada una de las categorías de edades.

Edades	n	p	%
15 a 24 años	102760		
25 a 34 años	282345		
35 a 44 años	367556		
45 a 54 años	335157		
55 a 64 años	228754		
65 años y más	339481		
Totales			

- Completen la siguiente tabla al calcular la proporción, el porcentaje y la tasa de internamiento (en prisiones y hospitales de salud mental) por cada 100 000 habitantes en ciudades de los EE.UU.

Ciudad	Población	Número de personas internas	Proporción	Porcentaje	Tasa por 100000 habitantes
Anderson, Indiana	130669	3981			
Bellingham, Washington	127780	1602			
Duluth, Minnesota	239971	4610			
Modesto, California	370552	4456			

ORGANIZACIÓN Y RESUMEN GRÁFICO DE DATOS

- Antes de hacer una representación gráfica de un conjunto de datos, conviene organizarlos de forma tabular, con el fin de realizar el proceso de manera más eficiente.
- Ejemplo: con el objeto de decidir si es necesario o no prestar un servicio de transporte a cierta comunidad universitaria, se quiere determinar cuál es el medio de transporte más utilizado por los estudiantes de la universidad para llegar a clase diariamente. Se hizo una encuesta con algunos estudiantes, seleccionados al azar, y los resultados obtenidos son:

Buseta	Bus	A pie	A pie	A pie
Bus	Buseta	Moto	A pie	Moto
A pie	Buseta	A pie	Auto	Buseta
Buseta	Moto	Buseta	A pie	Bus
Buseta	Moto	A pie	Buseta	Buseta
A pie	Auto	Auto	Buseta	Bus
Buseta	Buseta	Bus	Bus	Moto

- Esta forma de presentar los datos no es clara ni es eficiente. Organicen la información en la siguiente tabla, escribiendo en la columna 1 los valores que asume la variable y en la columna 2 el número de veces que se presenta dicho valor:

Valores	**Frecuencia**
Total	

- Una **tabla de distribución de frecuencias** contiene dos columnas. Una de ellas, la primera, muestra todos los posibles valores que asume la variable y la segunda muestra el número de veces que se presenta dicho valor, para cada variable. Este último número se denomina **frecuencia**.
- La distribución de frecuencias, entonces, es la lista de todas las puntaciones observadas de una variable y la frecuencia (*f*) de cada puntuación o categoría.
- En algunos casos es más informativo reportar la frecuencia de una categoría, bajo la forma de proporción o porcentaje con respecto al número total de sujetos de la muestra. Por eso, podemos hacer distribuciones de frecuencia proporcional y distribuciones de frecuencia porcentual.
- Completen la tabla de frecuencias sobre los medios de transporte, con las columnas correspondientes a la frecuencia proporcional y la frecuencia porcentual.

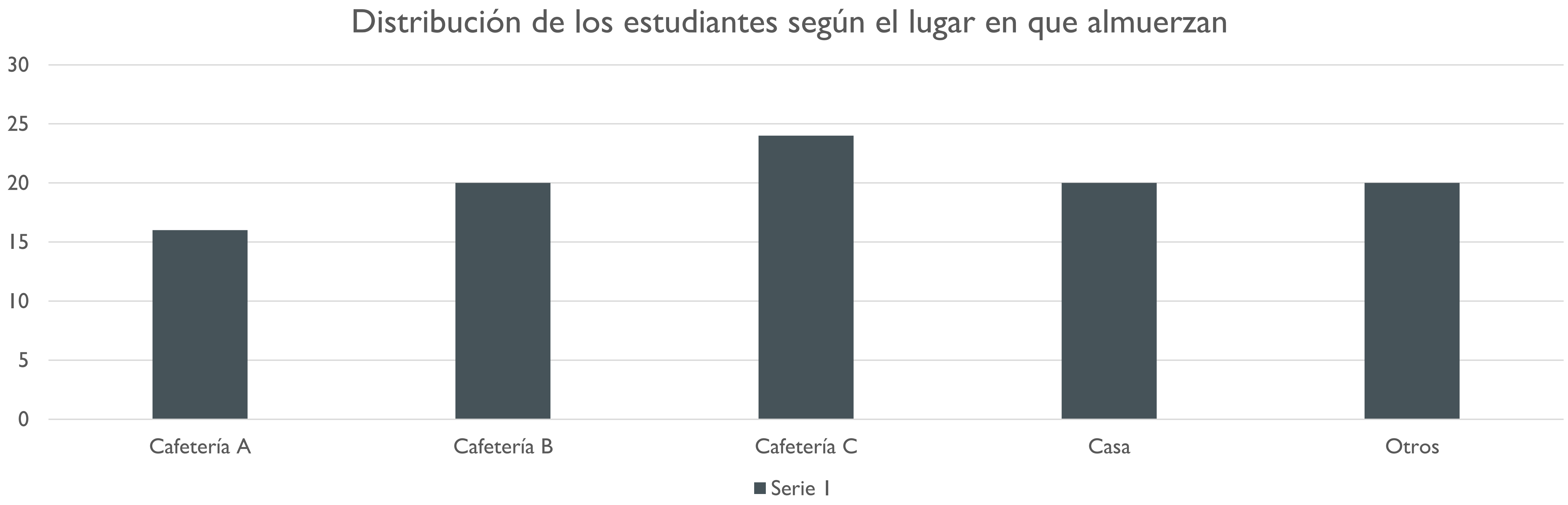
DIAGRAMAS

- Consideremos este problema: durante los días de clase, un mayor número de estudiantes de la Universidad almuerza en la cafetería A que en la cafetería B. el trabajo consiste en tomar una muestra de estudiantes de la Universidad y determinar si la afirmación es verdadera. Para una muestra aleatoria de 100 estudiantes, la tabla de frecuencias es la siguiente:

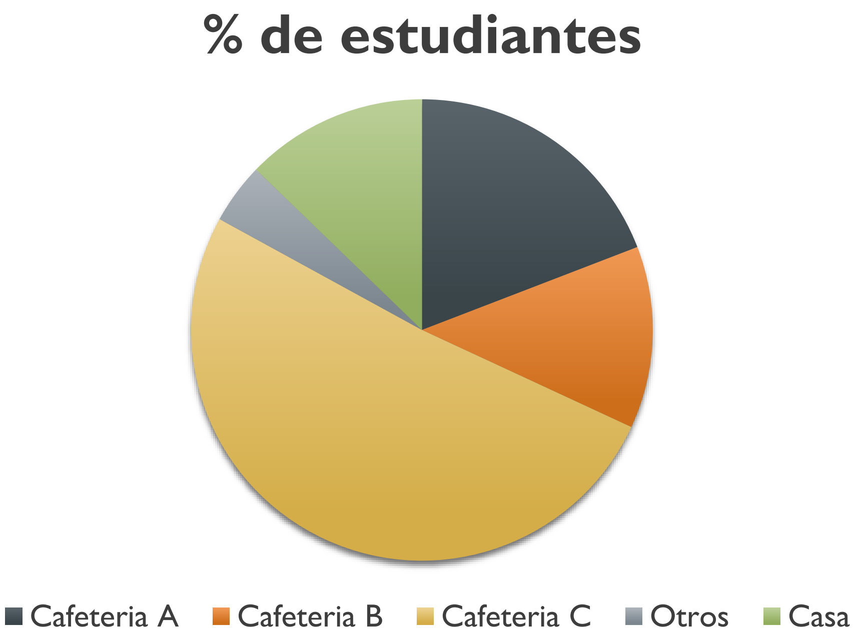
LUGAR	FRECUENCIA
A	16
B	20
C	24
Casa	20
Otros	20

Esta tabla nos permite evidenciar que no es verdad que el número de estudiantes que almuerzan en la cafetería A sea mayor al número correspondiente a la cafetería B.

DIAGRAMA DE BARRAS



- Consideremos otro problema: en los días de clase, la mayoría de los estudiantes de la Universidad almuerza en la cafetería C. la gráfica para una muestra de 100 estudiantes es la siguiente:



- Un **diagrama de bloques** es una gráfica que se emplea para representar la distribución de una variable categórica. Consta de una serie de rectángulos, cada uno de los cuales representa una categoría de la variable. Las bases de los rectángulos están sobre una misma recta y se nombran con los valores que toma la variable cuya distribución se quiere representar. Las bases de todos los rectángulos tienen la misma longitud y la altura de cada uno de ellos es proporcional al número de observaciones de la muestra. Los rectángulos que conforman la gráfica están separados entre sí para indicar que entre uno y otro valor de la variable no hay más valores.
- Un **diagrama circular** es una gráfica que se emplea para representar la distribución de una variable categórica. Para construirlo, se utiliza un círculo: se divide en tantos sectores como categorías tenga la variable. El tamaño de cada sector (es decir, el ángulo central correspondiente) debe ser proporcional al número de observaciones de la muestra que están incluidas en cada clase.

EJERCICIO:

- Laura y Daniela tomaron dos muestras de estudiantes de la Universidad, les preguntaron en dónde almuerzan frecuentemente los días de clase y quieren comparar los resultados obtenidos en las dos muestras. A continuación se presenta la tabla de distribución de frecuencias:

Lugar	Frecuencia (Laura)	Frecuencia (Daniela)
Cafetería A	24	36
Cafetería B	10	15
Cafetería C	14	21
Casa	12	18
Otros	20	30

Representen gráficamente los resultados obtenidos para cada una de las dos muestras. Compare el aspecto de las dos gráficas y exprese esa comparación en palabras. ¿Cuál es el tamaño de cada una de las dos muestras? ¿Qué porcentaje de los encuestados por Daniela prefieren la cafetería C? ¿Qué porcentaje de los encuestados por Laura prefieren la cafetería C? ¿Reflejan las gráficas que hicieron el hecho de que esas proporciones son iguales? ¿Qué cambio tendría que hacer en sus gráficas para que, al compararlas, sea evidente la comparación?

MEDIDAS DE POSICIÓN

- **Cuantiles:** puntuaciones que separan una fracción de los casos de una distribución.
- **Percentiles:** cuantiles comúnmente utilizados. El **rango percentilar** es el porcentaje de casos que caen en o están debajo de un valor específico de *X*.
- **Por ejemplo**, en un examen de admisión a la universidad, un estudiante con una calificación que corresponda al percentil 90 o mayor calificaría para la admisión en una universidad de prestigio, ya que significa que está por encima del 90% del resto de los alumnos.

ILUSTRACIÓN DE UNA DISTRIBUCIÓN DE FRECUENCIAS PORCENTUALES ACUMULADAS: AÑOS DE ESCOLARIDAD ENTRE CUIDADORES DE PACIENTES ANCIANOS CON ALZHEIMER.

Especificaciones		**Cálculos**	
Años de educación formal (*X*)	Frecuencia (*f*)	Frecuencia porcentual	Porcentaje acumulado
5	1		5
6	1		10
7	1		15
9	2		25
10	1		30
11	1		35
12	10		85
14	2	10	95
16	1	5	100
Total	20	100%	100

Un cuidador con 14 años de escolaridad tiene un nivel educativo igual o superior al 95% de la muestra, un rango percentilar de 95.

PASOS PARA CALCULAR PERCENTILES

- Ordenar los datos.
- Calcular la proporción y el porcentaje de casos con valores iguales o menores que el caso de interés.
- Indicar el percentil en porcentajes enteros.

Nota: Recordar que los percentiles se obtienen fácilmente de una distribución de porcentajes acumulada.

EJEMPLO:

Lugar de estudiante	Nombre de estudiante	Calificación de examen (ordenado)	Lugar de estudiante	Nombre de estudiante	Calificación de examen (ordenado)
1	Kevin	54	15	Shannel	79
2	Carl	58	16	William	80
3	Robert	61	17	Angie	82
4	Brian	61	18	Akilah	83
5	Maria	65	19	Daniel	85
6	Sean	69	20	Kaitlin	88
7	Jim	70	21	Marcy	90
8	Jessica	72	22	John	91
9	Carol	73	23	Barry	91
10	Brooke	75	24	Wnda	93
11	Kia	75	25	Sarah	95
12	Terry	77	26	Charles	96
13	Jackie	77	27	Elisa	97
14	Taylor	78			

- Calculemos el rango percentilar de John en este examen. Su calificación es igual o más alta que la de 23 de los 27 estudiantes. Primero, calculamos la proporción de casos iguales o menores a 91 y luego el porcentaje:

$$p\ [\text{de calificaciones} \leq 91] = \frac{23}{27} = 0{,}8518$$

$$\%\ [\text{de calificaciones} \leq 91] = 85{,}18\% = 85\%$$

Entonces el rango percentilar es 85.

- **ACTIVIDAD:** calculen los rangos percentilares de Taylor, Jackie, William y Brian.

- Los **cuartiles** son cuantiles que identifican las puntuaciones que dividen una distribución en cuatro grupos de igual tamaño (es decir, 25 por ciento de los casos en cada grupo). Cuando una distribución tiene un rango grande de puntuaciones, los cuartiles se obtienen fácilmente a partir de distribuciones de frecuencias de porcentajes acumulados. El primer cuartil, Q_1, es el 25° percentil; el segundo cuartil, Q_2, es el 50° percentil; y el tercero, Q_3, es el 75° percentil.
- **Ejemplo:**

TABLA 2-9 | Cuartiles de una distribución de calificaciones de un examen de mitad de curso

	Especificaciones		Cálculo		
	Calificación de examen (X)	f	Porcentaje f	Porcentaje acumulado (%) f	Ubicación de los cuartiles (Q)
	31	1	5.0%	5.0	
	58	1	5.0	10.0	
	63	1	5.0	15.0	
	68	1	5.0	20.0	
Q_1 →	69	1	5.0	25.0	← Q_1 = 25o. percentil
	72	1	5.0	30.0	
	76	1	5.0	35.0	
	77	1	5.0	40.0	
	82	1	5.0	45.0	
Q_2 →	84	1	5.0	50.0	← Q_2 = 50o. percentil
	85	1	5.0	55.0	
	86	2	10.0	65.0	
	88	1	5.0	70.0	
Q_3 →	91	1	5.0	75.0	← Q_3 = 75o. percentil
	93	2	10.0	85.0	
	94	1	5.0	90.0	
	95	1	5.0	95.0	
	97	1	5.0	100.0	
	Total	20	100.0%		

ACTIVIDAD

- Funk et al (2003) examinaron la relación entre jugar en juegos de video violentos y los sentimientos de empatía en una muestra de niños. Supongamos que los datos siguientes representan puntuaciones de escala de empatía para una muestra de 20 niños, después de jugar un videojuego violento: 4, 1, 3, 1, 2, 5, 3, 1, 2, 4, 2, 3, 5, 0, 2, 3, 4, 2, 1, 2.
  - Compilen los datos en una tabla de distribución de frecuencia con columnas para la frecuencia, la frecuencia proporcional, la frecuencia porcentual y la frecuencia de porcentaje acumulada.
  - Si un niño obtiene una calificación de 2 en la escala experimental de empatía, ¿Cuál es el rango percentilar de ese niño?

- Pearson y otros (1990) demostraron que cuando una abuela vive con una familia, es probable que participe en actividades de los padres. Supongamos que los siguientes datos representan el número de órdenes paternales dadas a 25 niños por sus abuelas: 5, 4, 3, 3, 6, 5, 3, 2, 4, 7, 5, 6, 2, 3, 4, 8, 7, 5, 6, 4, 2, 1, 5, 7, 3.
    - Compilen los datos en una tabla de distribución de frecuencia con columnas para la frecuencia, frecuencia proporcional, la frecuencia porcentual y la frecuencia porcentual acumulada.
    - Si una abuela dio dos órdenes, ¿cuál es su rango percentil? Justifiquen la respuesta.

- A continuación aparece una lista de calificaciones del Graduate Record Examination (GRE) para un grupo de 20 estudiantes graduados que solicitan inscribirse en una universidad:

Nombre del estudiante	Calificación GRE
Jones	1380
Jackson	1400
Schmidt	1220
González	1410
Roper	1100
Filer	1190
Miller	1200
Wong	1470
McMillan	1420
Polanski	1510
Nicholson	1450
Zimmerman	1520
Fitzsimmons	1210
van der Bergh	1180
Andropov	1550
Sheets	1110
Chang	1400
McKee	1450
Johnson	1380
Lucie	1220

En dónde se ubicarían los cuartiles 1, 2 y 3 en esta distribución? Calculen los rangos percentilares de Chang y Wong.

RECORDEMOS LAS MEDIDAS DE TENDENCIA CENTRAL

- Dados los datos siguientes donde X = Edad, calculen la edad modal, la edad mediana y la edad media. Empiecen por organizar los datos en una tabla con las puntuaciones ordenadas.

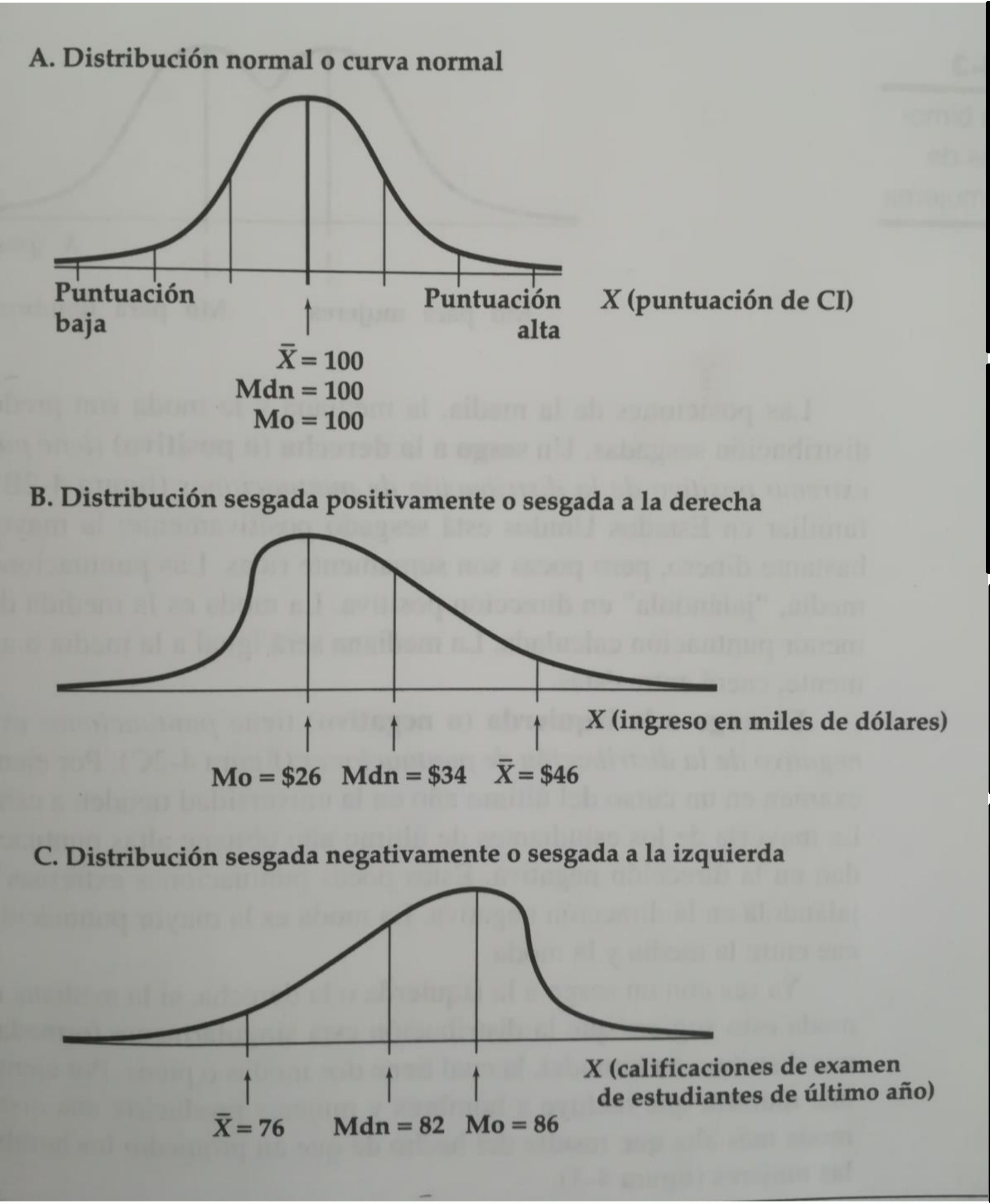
X
14
15
19
19
22
14
17
19
22
28

- Siete trabajadores de oficinas entraron en un concurso de reducción de peso. Tras unas semanas de someterse a dieta, los pesos que bajaron (en libras) fueron como sigue: 5, 7, 3, 0, 2, 4, 3. Calculen la pérdida de peso mediana y media. Los datos de X = libras perdidas.

- Los cinco miembros de una familia trabajan. Sus salarios por hora son: \$30, \$10,50, \$5,15, \$12 y \$6. Los datos de X = salario por hora.
  - Calculen la media y la mediana.
  - En comparación con otras puntuaciones, ¿qué ocurre con el salario de \$30 por hora? ¿cuál es su efecto en el cálculo de la media?
  - Ajusten esa peculiaridad, recalculando la media sin ella.

CURVAS DE DISTRIBUCIÓN DE FRECUENCIAS

- Las curvas de distribución de frecuencias son una forma de representar gráficamente los resultados de las medidas de tendencia central. El área bajo la curva representa el número total de sujetos en la población y es igual a una proporción de 1.00 a un porcentaje de 100 por ciento.
- Las puntuaciones de una variable se ilustran de izquierda (la más baja) a derecha (la más alta), es decir, las puntuaciones se ordenan sobre el eje horizontal.
- El eje vertical (el cual a veces no nos molestamos en dibujar) representa la frecuencia, la frecuencia proporcional o la frecuencia porcentual. Así, la altura de la curva en cualquier valor de X representa la proporción de una muestra o población con esa puntuación.
- Lo que nos interesa es evaluar la forma de una distribución y examinar las posiciones relativas de la media, la mediana y la moda, para estimar la forma de una distribución de frecuencias. Las curvas de distribución de frecuencias se aplican sólo a variables de intervalo o razón.



Una distribución normal es aquella donde la media, la mediana y la moda de una variable son iguales entre sí y la distribución de las puntuaciones tiene forma de campana. También nos referimos a ella como una "curva normal". La figura A. ilustra puntuaciones de coeficiente intelectual, que están distribuidos normalmente con una media de 100. una distribución normal es simétrica, es decir, equilibrada en cada lado. Su media, mediana y moda se localizan en el centro de la distribución.

Una distribución sesgada es aquella en la cual la media, la mediana y la moda de una variable son desiguales y algunos de los sujetos tienen puntuaciones sumamente altas o bajas. Cuando este es el caso, la distribución se alarga hacia un lado.

Un sesgo a la derecha (o positivo) tiene puntuaciones extremas en el extremo positivo de la distribución de puntuaciones. Por ejemplo, el ingreso familiar en Estados Unidos está sesgado positivamente: la mayoría de las familias ganan bastante dinero, pero pocas son sumamente ricas.

Un sesgo a la izquierda tiene puntuaciones extremas en el final bajo o negativo de la distribución de puntuaciones. Por ejemplo, las puntuaciones del examen en un curso de último año en la universidad tienden a estar sesgadas a la izquierda. La mayoría de los estudiantes obtiene altas puntuaciones, pero pocos se quedan en la dirección negativa. Estas pocas puntuaciones extremas bajas desinflan la media, inclinándola en dirección negativa. La moda es la mayor puntuación calculada, y la mediana cae entre la media y la moda.

- Supongamos que nos interesa saber con qué frecuencia es que los estudiantes de artes de la universidad estudian su disciplina asistiendo a cines con películas de estreno. Recolectamos una muestra aleatoria de 19 estudiantes y pedimos a cada uno mencionar las nuevas películas que vieron en el cine el mes pasado, y registramos los siguientes resultados: 2, 6, 4, 5, 2, 3, 4, 3, 6, 4, 3, 3, 5, 4, 5, 2, 3, 4, 3.
  - Elaboren una tabla con los datos ordenados de mayor a menor.
  - Transformen esa tabla en una tabla de frecuencias.
  - Calculen la media.
  - Calculen la mediana.
  - Calculen la moda.
  - Elaboren la curva de distribución de frecuencias.
  - Interpreten los resultados.
- Realizar el mismo ejercicio con el problema en la diapositiva No. 57.