



GERMÁN FROMM

CONSULTOR EN GESTIÓN EDUCACIONAL

Estadísticas Descriptivas

Cuantitativa

La metodología cuantitativa:

- Realiza mediciones en muchos casos.
- Requiere medir en forma de variables.
- Establece relaciones entre variables para someter a análisis estadístico.
- Utiliza técnicas estadísticas de lo más rudimentario (estadística descriptiva) a lo más sofisticado (estadística inferencial).

Variables

Variables: Fenómenos medibles que varían a través del tiempo o que difieren de un lugar a otro o de un individuo a otro. Las variables son características de los sujetos u objetos bajo estudio.

- **Variable dependiente:** Es la variable de interés principal cuándo realizamos un estudio. La variable cuya variación queremos comprender.
- **Variables independientes:** Las variables que sospechamos que tienen alguna relación con la variación de la variable dependiente. Las variables independientes son las variables predictoras.

Casos: Corresponden a los sujetos o unidades de análisis observados que forman parte de nuestro estudio.

Datos: Corresponden a los valores de las variables en cada caso perteneciente al estudio.

Variables, casos y datos ...

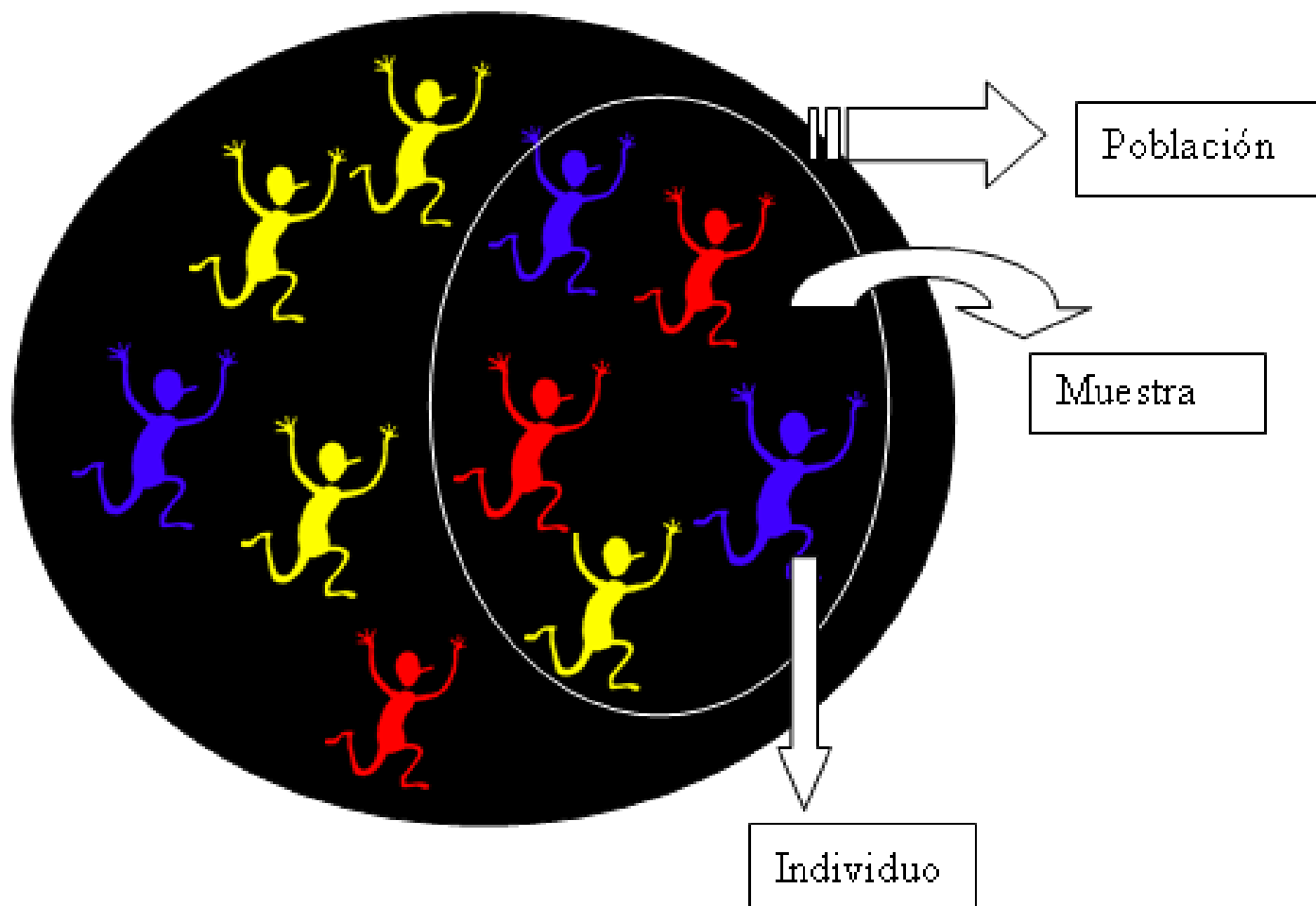
Ejemplo: base de datos encuesta de apoderados SIMCE 4° Básico

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana Ayuda												
Variables												
	RBD	Curso	Codigo	P01	P02	P03	P04	P05	P06_1	P06_2	P06_3	P06_4
1	18237	4°	35011	2	1984	3	3	4	0	0	0	0
2	8438	4°	40001	2	1978	1	1	3	1	0	0	0
			40002	2	1974	2	2	3	1	0	0	0
			40003	2	1960	1	2	3	1	0	0	0
			40004	2	1953	1	1			0	0	0
6	8455	4°	40005	2	1957	1	1			0	0	0
7	8455	4°	40006	2	1964	3	3			0	0	0
8	11678	4°	40007	1	1985	2	2	4	1	0	0	0
9	24324	4°	40008	2	1979	1	1	0	1	0	0	0
10	8421	4°	40009	2	1966	2	1	3	1	0	0	0
11	8421	4°	40010	2	1957	1	1	3	1	0	0	0
12	12502	4°	40011	2	1983	1	1	3	1	0	0	0
13	12539	4°	40012	1	1967	1	1	3	1	0	0	0

Casos

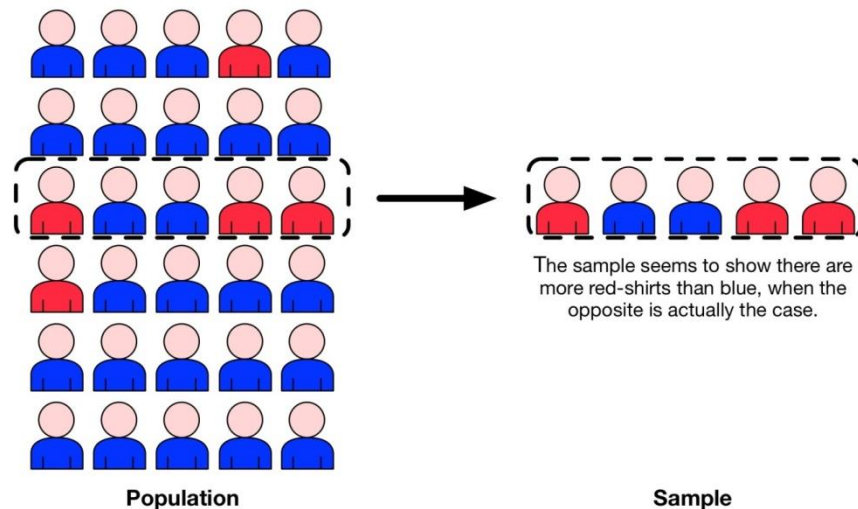
Datos

Población, muestra y unidad de análisis



Representatividad

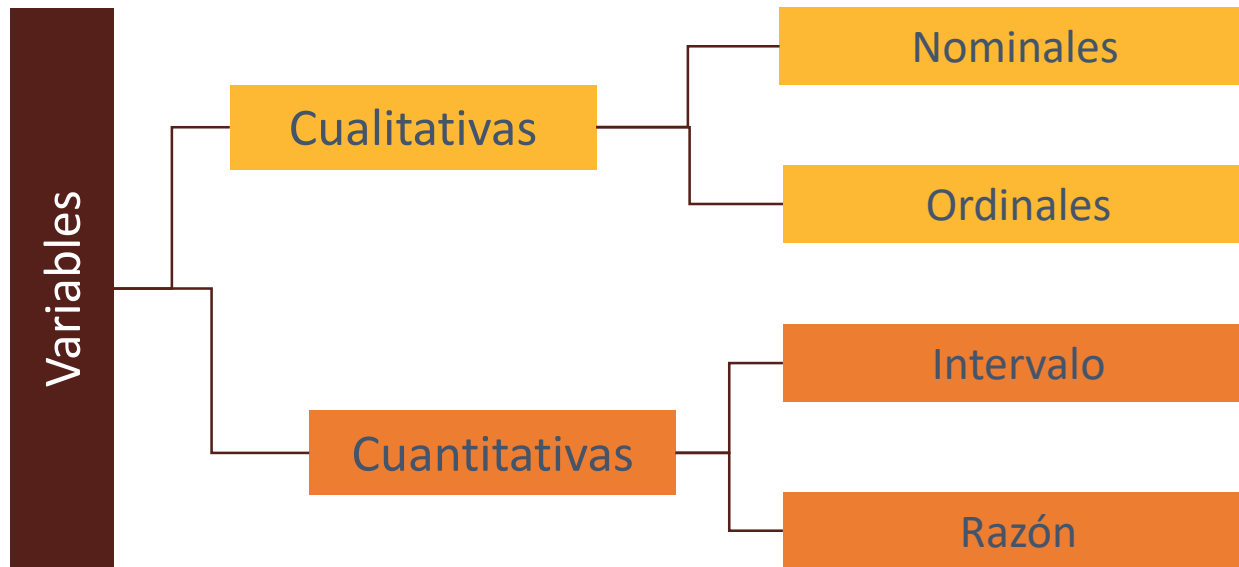
- La representatividad NO es parte del análisis estadístico, sino del andamiaje metodológico que tiene el plan de análisis.



A evitar: Sesgos, muestras pequeñas, poblaciones “iguales”.

Tipos de Variables

“Identifica las propiedades de medición de la variable y determina el tipo de operaciones matemáticas (suma, multiplicación, etc.) que puede usarse apropiadamente con dicho nivel, así como las fórmulas estadísticas que utiliza para probar las hipótesis teóricas” (Ritchey et al, 2002: 43)



Variables Nominales

Indican una diferencia en categoría, clase, calidad o tipo.

Ej: Nombre del establecimiento, RBD, curso, sexo, comuna, dependencia, etc.

Las variables nominales no admiten puntuaciones numéricas ordenadas de manera significativa. Ej: Mujer=1; Hombre =2

Una variable dicotómica con 2 categorías se llama “dummy” y se usa para indicar la presencia (1) y ausencia (0) de una categoría.

13. ¿Pertenece el Padre y/o la Madre del estudiante a alguno de los siguientes pueblos originarios de Chile?

Considere que una persona pertenece a un pueblo originario si cumple con al menos una de las siguientes condiciones:

- *es hijo/a de Padre o Madre indígena*
- *posee al menos un apellido indígena*
- *es cónyuge de una persona indígena*
- *mantiene formas de vida, costumbres o religión de estas etnias de un modo habitual*

Marque con una equis (X) la alternativa que corresponda a Padre y Madre.

	Padre	Madre
Sí pertenece (Aymará, Rapa Nui, Quechua, Mapuche, Atacameño, Coya, Kawaskar, Diaguita o Yagán)	☐	☐
No pertenece	☐	☐

Variables Ordinales

Designan categorías, pero permiten además clasificar las categorías de mayor a menor, de mejor a peor, de primero a último. Las variables ordinales admiten puntuaciones numéricas ordenadas de manera significativa.

Ej: Clase social: C3=1; C2=2; ABC1=3

Ej: Quintiles de ingreso, nivel educativo, periodos de tiempo, etc.

¿Qué pasa con la escala Likert? Es utilizada para medir actitudes y opiniones.
¿Es también de tipo ordinal?

31. ¿Qué tan satisfecho se siente Ud. con el profesor o profesora jefe del estudiante?

Marque con una equis (X) una sola alternativa.

☐

Muy satisfecho

☐

Satisfecho

☐

Ni satisfecho
ni insatisfecho

☐

Insatisfecho

☐

Muy insatisfecho

Variables Escalares

Estas variables además de clasificar y ordenar categorías, tienen una unidad numérica específica. Identifican las diferencias en monto, cantidad, grado o distancia. Las distancias entre las puntuaciones son las mismas entre cualquier par de puntos en la escala de medición.

Ej: tº, CI, índice de desarrollo humano.

Variables ordinales tipo intervalo: Son variables ordinales con varias categorías que se consideran como si fueran de tipo intervalar. La literatura sugiere una dispersión continua (no discreta) o un mínimo 7 categorías discretas. La variable se debe distribuir normalmente.

Revisen las preguntas del cuestionario enviado a modo de ejemplo.

Variables de Razón

Poseen las características de una variable intervalo y un punto cero verdadero (cero significa ninguno) y las proporciones en las magnitudes se mantienen (por ejemplo: el doble efectivamente significa el doble).

Ej: Peso, altura, edad, distancia, tamaño de la población, duración de tiempo, cantidad de respuestas correctas, etc.

2. ¿Cuál es su edad?

Escriba el número de años, por ejemplo 45.

años

Análisis con Variables Cualitativas

- Básicamente Nominales y Ordinales

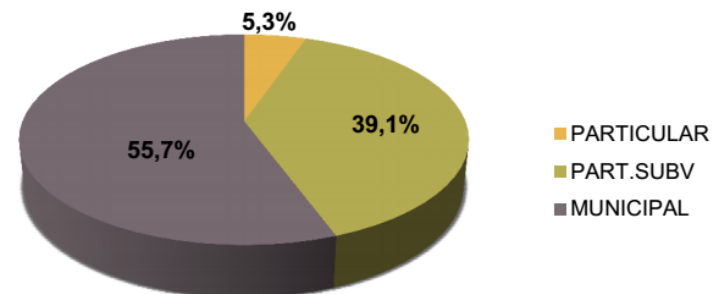
Distribución de Frecuencia

Las variables cualitativas pueden ser analizadas mediante tablas de frecuencias así como también mediante representaciones gráficas (barra, torta o diagrama de Pareto entre otros).

Dependencia establecimientos SIMCE 4° Básico 2010

Dependencia		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	MUNICIPAL	4412	55,7	55,7	55,7
	PART.SUBV	3098	39,1	39,1	94,7
	PARTICULAR	417	5,3	5,3	100,0
	Total	7927	100,0	100,0	

Dependencia establecimientos
SIMCE 4° básico 2010



Análisis con Variables Cualitativas

Tablas de Contingencia

Las tablas de contingencia muestran la relación entre dos o más variables categóricas (nominales u ordinales). Busca explorar relaciones entre variables. Se examina cada subgrupo en la variable independiente y se compara en función de la variable dependiente.

Tabla de contingencia PROMEDIO MATEMATICA * Dependencia

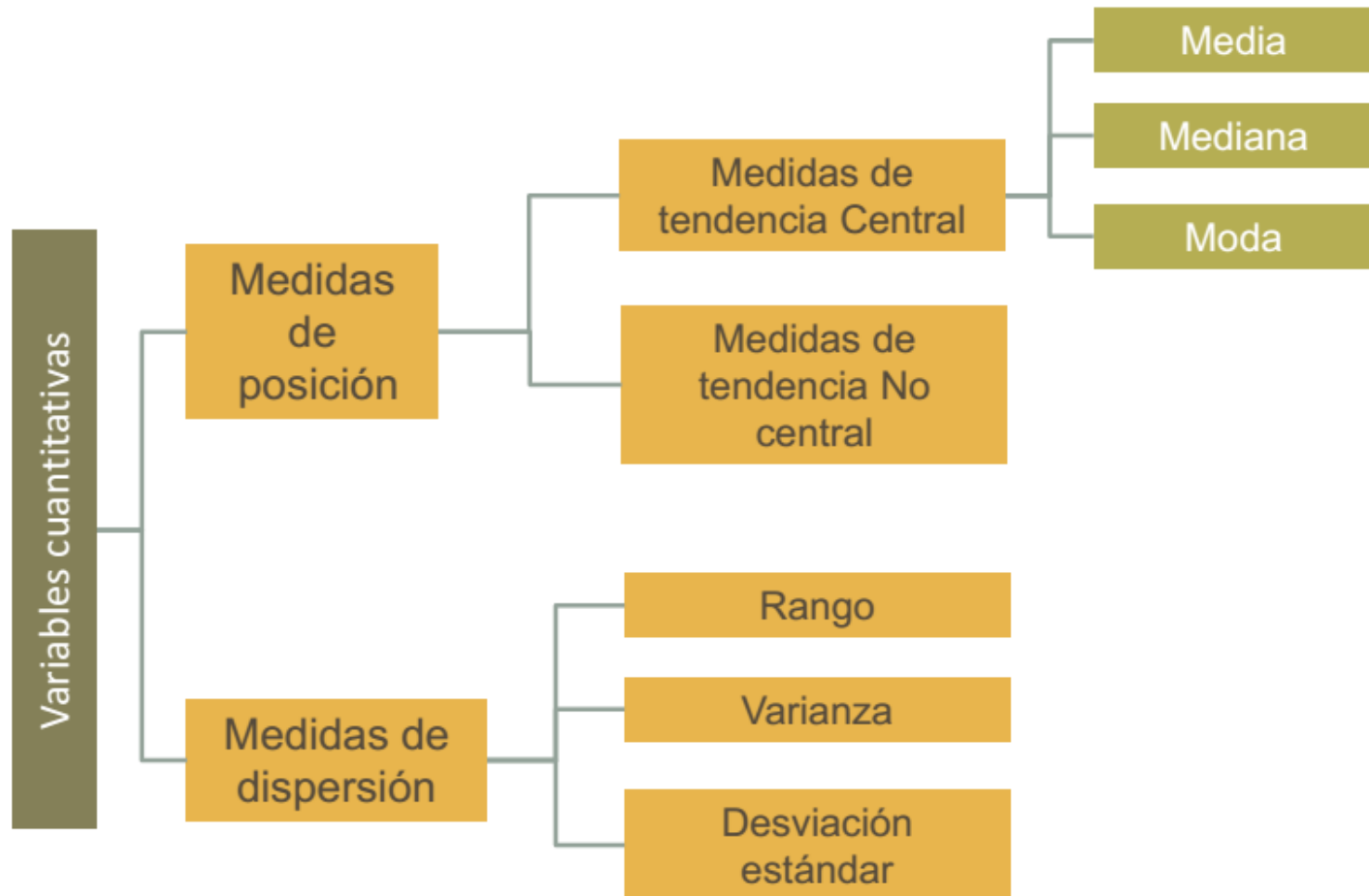
			Dependencia			Total
			MUNICIPAL	PART.SUBV	PARTICULAR	
PROMEDIO MATEMATICA	155-225	Recuento	983	468	7	1458
		% dentro de Dependencia	35,7%	18,2%	1,8%	25,5%
	226-245	Recuento	873	591	15	1479
		% dentro de Dependencia	31,7%	23,0%	3,9%	25,9%
	246-266	Recuento	617	728	21	1366
		% dentro de Dependencia	22,4%	28,3%	5,5%	23,9%
	267-340	Recuento	281	786	339	1406
		% dentro de Dependencia	10,2%	30,5%	88,7%	24,6%
Total		Recuento	2754	2573	382	5709
		% dentro de Dependencia	100,0%	100,0%	100,0%	100,0%

Resumen

Nivel de medición	Cualidades	Operaciones matemáticas
Nominal	Clasificación en categorías	Conteo (frecuencia) Comparación de los tamaños de las categorías
Ordinal	Clasificación en categorías; ordenamiento de los rangos de categorías de menor a mayor	Todas las anteriores; juicios de mayor que y menor que y el cálculo de diferencias y promedios de rango
Intervalo	Todas las anteriores y las distancias entre las puntuaciones tienen una unidad determinada de medición	Todas las anteriores y operaciones como suma, resta, multiplicación, división, raíz cuadrada.
Razón	Todo lo anterior y un punto cero real	Todas las anteriores y el cálculo de razones con significado.

(Ritchey et. al,2002)

Análisis de Estadística Descriptiva



Medidas de Tendencia Central

Una medida de tendencia central es aquel valor que se localiza en el centro o a la mitad de un determinado conjunto de datos.

En general, los estadísticos de tendencia central proporcionan una estimación de la puntuación típica, común o normal encontrada en una distribución de puntuaciones en bruto.

- Media (=promedio)
- Mediana
- Moda

Promedio o Media

Media aritmética...

- **Desventajas:** La media puede verse distorsionada por los valores extremos (máximos o mínimos), por eso es conveniente reportarla junto con otros estadísticos de tendencia central.

220 247 262 275 278 280 296 305

$$\chi = \frac{\sum x}{n} = \frac{247 + 262 + 275 + 278 + 280 + 296 + 305}{7} = 277,5$$

- Mientras menor sea el tamaño de la muestra, mayor será la distorsión que genere un valor extremo.

Mediana

Para lidiar con la sensibilidad de la media aritmética a los valores extremos o una distribución sesgada, podemos utilizar la **mediana (Mdn)** que puede ser entendida como el *valor intermedio* de una distribución ordenada. En específico, es aquel **valor que se encuentra en la mitad de la distribución**, por encima de la cual queda la mitad de los casos y por debajo queda la otra mitad.

Cálculo de la Mediana (Mdn)

1. Ordenar la distribución de puntuaciones (valores) de menor a mayor
2. Ubicar la posición de la mediana ($n/2$)
3. Si n es impar, la mediana será un caso real de la distribución
4. Si n es par, la media será el promedio de las dos puntuaciones que están a la mitad.

Mediana (un ejemplo)

Mediana. Un ejemplo

- Se tienen los resultados SIMCE 2010 en lectura 4º básico de 7 escuelas. Los datos son:

247 275 305 278 280 296 262

- Ordenar las puntuaciones: 247 262 275 278 280 296 305
 - Ubicar la posición $\frac{n}{2} + 0.5 : \frac{7}{2} + 0.5 = 4$
 - Mdn siendo n impar es **278**
-

Cuál es la Mdn con los datos de 8 escuelas, cuyos datos son:

247 275 220 305 278 280 296 262

- Ordenar las puntuaciones: 220 247 262 275 278 280 296 305
- Ubicar la posición $\frac{n}{2} + 0.5 : \frac{8}{2} + 0.5 = 4,5$
- Mdn siendo n par es $\frac{275+278}{2} = \mathbf{276,5}$

Mediana (un ejemplo)

Mediana...

- **Desventajas:** La mediana es insensible a los valores de la distribución, por lo que reportarla sola nos puede llevar a conclusiones erradas o imprecisas.

Ej: Puntajes Simce en lenguaje 4° básico en dos comunas con igual cantidad de escuelas

Comuna 1: 247 262 275 **278** 280 296 305

Comuna 2: 272 274 276 **278** 287 302 322

Frente a distribuciones con valores muy distintos, la mediana podría llevarnos a plantear erróneamente que las escuelas de ambas comunas tuvieron desempeños similares.

- **Desventajas:** Al depender del número de casos de la distribución, cualquier cambio en el tamaño de la muestra puede afectar este estadístico

Moda

Moda

Es aquella puntuación que se repite con mayor número de veces en una distribución determinada, es decir, **que posee la mayor frecuencia entre nuestros datos**. A este valor lo denominamos como **moda (M_o)**. La moda es un valor determinado de la variable x y no la frecuencia de esa categoría.

Cálculo de la Moda (M_o)

1. Compilar las puntuaciones en una distribución de frecuencias
2. Identificar la moda, que es el valor de x con mayor frecuencia

Desventajas: La moda no es muy útil por sí sola, porque es insensible tanto a los valores de la distribución como al tamaño de la muestra. Es útil reportarla por sí sola cuando las puntuaciones de x son en esencia del mismo valor para todos los casos. Ej: edad de los estudiantes de un curso.

Moda. Un ejemplo

15. ¿Con qué tipo de horas está hecho su contrato? ¿Cuántas horas del tipo que señaló tiene su contrato a la semana?

Marque con una equis (X) el tipo de horas que tiene su contrato y luego indique el número total semanal de horas contratadas según ese tipo de horas.

- ☐ Horas pedagógicas (45 minutos) ☐ ☐ Horas a la semana
- ☐ Horas cronológicas (60 minutos) ☐ ☐ Horas a la semana

Horas pedagógicas contratadas	Frecuencia
19	5
20	9
21	6
22	12
23	9
24	21
25	33
26	31
27	22
28	56
29	15
30	484
31	26
32	88
33	52
34	94
35	55
36	167
37	73
38	637
39	169
40	245
41	43
42	114
43	35
44	633

Medidas de Tendencia no Central



Medidas de Tendencia no Central

Percentiles, cuartiles, deciles

El rango percentilar o **percentil** es el porcentaje de casos que caen o están debajo de un valor específico de x (el percentil es el **valor de la puntuación**, no el porcentaje acumulado)

Los **cuartiles**, fraccionan una distribución en 4 grupos del mismo tamaño (Q1=25%; Q2=50%; Q3=75%)

Los **deciles** fraccionan una distribución en 10 grupos del mismo tamaño.

Estadísticos

PROM_MAT

N	Válidos	5709
	Perdidos	2218
Media		246,15
Desv. típ.		29,900
Mínimo		155
Máximo		340
Percentiles	25	225,00
	50	245,00
	75	266,00

Medidas de Tendencia no Central

