



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



DIRECCIÓN GENERAL DE EDUCACIÓN  
TECNOLÓGICA INDUSTRIAL Y DE SERVICIOS



2023  
Año de  
**Francisco  
VILLA**



## PENSAMIENTO MATEMÁTICO



1er semestre  
Agosto 23 – Enero 24

**UNIDAD II**

**M**ate  
maycko



## PENSAMIENTO MATEMÁTICO

### PROGRESIÓN No. 6

Se lección: una problemática o situación de interés con la finalidad de recolectar información y datos de fuentes confiables e identificar las variables relevantes para su estudio

METAS	CATEGORÍAS	SUBCATEGORÍAS
M1 Ejecuta cálculos y algoritmos para resolver problemas matemáticos de las ciencias y de su entorno.	C1 Procedural	S4 Manejo de Datos e Incertidumbre
M1 Observa y obtiene información de una situación o fenómeno para establecer estrategias o formas de visualización que ayuden a entenderlo	C1 Procesos de intuición y razonamiento	S1 Capacidad para observar y categorizar S2 Pensamiento intuitivo
M2 Socializa con sus pares sus conjeturas descubriendo los procesos en la solución de un problema tanto teórico como de su entorno	C4 Interacción y Lenguaje matemático	B Ambiente matemático de comunicación



## PROGRESION No. 6



### VARIABLES

Desde el punto de vista conceptual, existen tres tipos de variables:

- Nominales
- Ordinales
- Cardinales

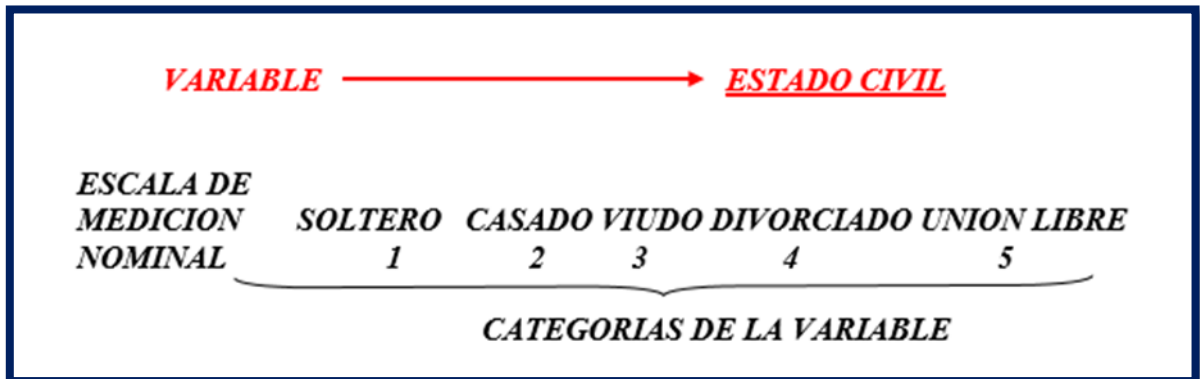
### VARIABLES NOMINALES

Son las más simples y abundantes y su única función es clasificar. Su variable operacional correspondiente es una escala nominal que sirve para clasificar las observaciones en un conjunto de categorías mutuamente excluyentes, cuyo orden de colección es indistinto. A estas se les puede asignar cifras u otros símbolos arbitrarios con el fin de distinguirlas; si son cifras, no tienen ningún valor intrínseco ni propiedades numéricas como en la aritmética, ejemplos:

<i>Distrito de Riego</i>	<i>Hectáreas sembradas</i>
1	680
2	1200
3	300
4	500

En esta tabla observamos que los símbolos 1, 2, 3, 4, si bien, son los mismos que se emplean para representar números, no representan sino distritos de riego; es decir, carecen de propiedades numéricas. Además, el orden que se les dé en la tabla es indistinto, ya que solo sirven para distinguir un distrito de otro.

Veamos ahora otro ejemplo:



A nivel de mediación, estado civil en este ejemplo es un conjunto de cinco categorías mutuamente excluyentes (La expresión mutuamente excluyentes significa que si un elemento del universo es, digamos, soltero, no puede ser casado o cualquier otra modalidad del estado civil al mismo tiempo), cuyo orden de colocación es indistinto, ya que pudimos haber puesto primero viudo o casado y terminar en soltero. Además, si a “soltero” le llamamos 1, a “casado” 2, etc., estas cifras carecen de propiedades numéricas, ya que solo sirven para distinguir un estado civil de otro.

\*\*Es común también confundir la variable con sus categorías. Suele oírse que en un ejemplo como el anterior hay 5 variables; ESTO ES UN ERROR. La variable es solo una: estado civil, que en este caso tiene cinco categorías o posibilidades de respuesta en un universo determinado. Otras variables nominales serían: sexo, nacionalidad, color de automóvil, tipo de lámpara, lugar de nacimiento, etc.

## VARIABLES ORDINALES

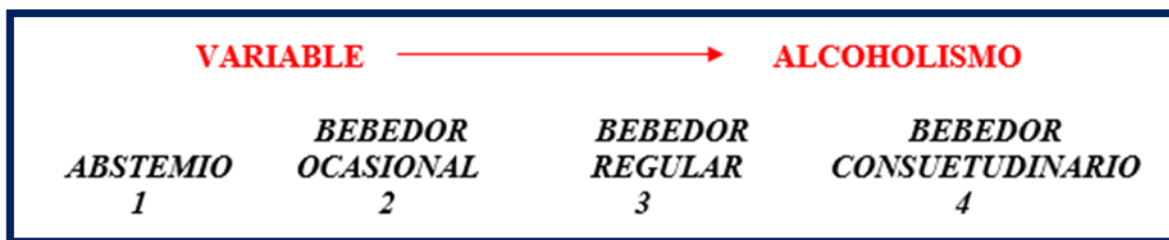
Clasifican las observaciones en categorías mutuamente excluyentes que exigen ordenación, ya que guardan entre sí relaciones de “mayor que”. Su variable operacional es una escala ordinal que va desde la categoría más baja a la más alta o viceversa, de modo que las observaciones queden en el orden apropiado. Estas categorías tampoco tienen propiedades numéricas, aunque se las represente por cifras, ejemplos:

**En este ejemplo nótese que es preciso ordenar las cifras 1, 2, y 3, puesto que representan la gravedad de las quemaduras de menor a mayor.**



<i>Grado de las Quemaduras</i>	<i>No de Casos</i>
1	70
2	40
3	10

La **variable alcoholismo**, definida como el grado de dependencia respecto a las bebidas embriagantes, es un buen ejemplo de variable ordinal.



Nos damos cuenta que, a nivel de la medición, “**alcoholismo**” es un conjunto de categorías mutuamente excluyentes, que van desde la posibilidad de no beber nunca hasta la de beber consuetudinariamente, dos extremos entre los cuales cabrían un sinnúmero de gradaciones. Si definimos al bebedor regular como el que ingiere bebidas alcohólicas con más frecuencia que el consuetudinario, podemos afirmar que aquel tiene un grado de dependencia respecto al alcohol, mayor que el bebedor ocasional, y menor que el consuetudinario. Estas categorías tienen que estar ordenadas, sea que se le llame por su nombre o por medio de cifras que carecerían de propiedades numéricas: la cifra 3 indicaría un grado de dependencia menor que la 4 mayor que la 2, pero nada más.

Otras variables del tipo ordinal serian “grado de escolaridad” “rango militar” o “jerarquía eclesiástica”.

## VARIABLES CARDINALES

Son las más complejas. Su variable operacional es una escala cardinal que se caracteriza porque las diferencias iguales entre dos de sus puntos son iguales




entre sí. Las cifras asociadas a las categorías son **efectivamente cuantitativas** y, en consecuencia, se puede efectuar con ellas operaciones aritméticas.

## Sabias que....


### LAS VARIABLES CARDINALES SE DIVIDEN EN CONTINUAS O DISCRETAS

**VARIABLE CONTINUA**



Es aquella que puede tomar cualquier valor, ya sea entero o fraccionario, por ejemplo: el número de kilómetros que puede recorrer un automóvil con el tanque de gasolina lleno, a diferentes alturas sobre el nivel del mar, podría ser variable 353.5, 348.76, 350.0, 351.20 Km, etcétera, otros ejemplos, Producción anual de azúcar, salarios, estatura, etc.

**VARIABLE DISCRETA**



Es aquella que tan solo puede tomar como valores unidades enteras, por ejemplo; una familia puede tener 0, 1, 2, 3, 4, etcétera, hijos, pero jamás podrá tener 2.5 o 3.2 hijos, otros ejemplos; producción anual de automóviles, número de huelgas anuales.

*Ejemplos:*

**VARIABLE** → **EDAD**

6    7    8    9    10    11    12 años

CATEGORIAS DE LA VARIABLE

Sin embargo, aunque por razones prácticas se acostumbre reportar las edades de las personas a partir de cierto momento en años cumplidos, bien podríamos decir que un niño tiene 7.25 años; es decir, 7 años 3 meses o 7.5 años. Con esto queremos destacar que la variable puede tomar cualquier valor entre los límites 6-12. Por lo tanto, "edad" es una variable continua.



Supongamos ahora que investigamos en una comunidad el número de hijos por familia. Esta variable podría admitir las siguientes respuestas:

Es evidente que entre los límites 0-12 no puede haber cualquier valor; no podríamos registrar 4.25 hijos. Por tanto, "número de hijos" es una variable discontinua o discreta.

## VARIABLE CONSTANTE

Es aquella que solo puede tomar un valor, por ejemplo; una persona solo puede tener una madre; o el número de razones sociales que puede tener una sociedad siempre sería uno.

Estos conceptos pertenecen a una de las ramas de la estadística denominada teoría del muestreo, que es muy importante, pues si se hace un muestreo correcto, se obtendrán resultados representativos de la población; en caso contrario podemos incurrir en errores u omisiones debidas a estimaciones que están fuera de la realidad.

Se diferencian dos usos del método estadístico: estadística descriptiva o deductiva y estadística inferencial o inductiva



En la lista siguiente, anota a que tipo de variable corresponde cada una de las que se te presentan y anota de que tipo son, posteriormente escribe los posibles valores que puede tomar, anótala en tu cuaderno para que la puedas guardar en tu portafolio de evidencias.



Variable de estudio	Tipo de variable	Posibles valores que puede tomar
Número de computadoras por hogar.		
Mes de nacimiento.		
Cantidad de agua que ingiere durante un día		
Nivel de hemoglobina.		
Deporte preferido.		
Peso de cada miembro de mi familia.		
Índice de Masa Corporal.		
Horas de estudio semanales.		
Veces que se acude al cine durante un mes.		
Promedio escolar.		
Tipo de sangre.		
Número de hermanos.		
Estatura de los alumnos.		
Lugar que ocupa en la lista de asistencia.		
Monto de recibo del consumo de agua.		
Tiempo diario dedicado al WhatsApp.		
Color de ojos.		
Materia preferida.		
Talla de calzado.		
Pulsaciones por minuto.		
Carrera universitaria a elegir.		
Lugar de nacimiento.		
Partido político con el que simpatiza.		

 **Actividad de Aprendizaje No 24**

Realiza una búsqueda bibliográfica que incluya conceptos básicos y ejemplos de variables aleatorias discretas y continuas. Al encontrar el significado de estos conceptos, sintetízalos y desarrolla un esquema en que se relacionen cada uno de ellos. Puedes usar otros libros de texto o páginas de internet. El objetivo es desarrollar un esquema conceptual para reforzar el contenido de los conceptos y su vinculación. Al finalizar comenta tus resultados con tu profesor (a) y compañeros de la comunidad. Recuerda integrar tus documentos en el portafolio de evidencias.





## Reflexionemos sobre la actividad

### ¿De qué te das cuenta?

Señala con el número 1 las variables que consideres son aleatorias discretas, y con el número 2 las que sean aleatorias continuas. Comparte con tus compañeros (as) y tu profesor(a) las respuestas.

1. Número de páginas de un libro \_\_\_\_\_
2. Tiempo que tarda en fundirse una bombilla \_\_\_\_\_
3. Número de preguntas en una clase de una hora \_\_\_\_\_
4. Cantidad de agua consumida en un mes \_\_\_\_\_
5. La edad de las mujeres en tu localidad \_\_\_\_\_
6. La estatura de los hombres de tu salón \_\_\_\_\_
7. El tiempo que tarda en madurar una manzana \_\_\_\_\_
8. Kilowatts gastados por familia en tu población \_\_\_\_\_



### Actividad de Aprendizaje No 25.

Como actividad final de esta Progresión, se realizará una actividad en común con los demás grupos y los compañeros de Lengua y Comunicación.



## PENSAMIENTO MATEMÁTICO

### PROGRESIÓN No. 7

Analiza datos categóricos y cuantitativos de alguna problemática o situación de interés para el estudiantado, a través de algunas de sus representaciones gráficas más sencillas como las gráficas de barras (variables cualitativas) o gráficos de puntos e histogramas (variables cuantitativas).

METAS	CATEGORÍAS	SUBCATEGORÍAS
M1 Ejecuta cálculos y algoritmos para resolver problemas matemáticos de las ciencias y de su entorno.	C1 Procedural	S2 Elementos Geométricos S4 Manejo de Datos e Incertidumbre
M2 Analiza los resultados obtenidos al aplicar los procedimientos algorítmicos propios del pensamiento matemático en la resolución de problemáticas teóricas como de su entorno.		
M3 Socializa con sus pares sus conjeturas descubrimientos o procesos en la solución de un problema tanto teórico como de su entorno.	C2 Procesos de Intuición y razonamiento	S1 Capacidad para observar y conjeturar S2 Pensamiento intuitivo



## GRAFICOS ESTADISTICOS



Organización de los datos. Existen muchas formas de organizar los datos. Para poder decidir cuál es la mejor debes tener muy claro para qué quieres la información. Recuerda a Bachelard

*Lo importante no es medir para pensar, sino pensar para medir.*

¿Para qué te sirve tener organizada la información? Imagínate que tu profesor te pide que hagas una presentación de tres minutos frente a tu grupo sobre tu ciudad. ¿Crees que puedes hablar de todo lo que hay en tu ciudad?

**Si No**

¿de qué hablarías?

¿Hablarías de su tamaño, de sus habitantes, de su vegetación y clima, de sus actividades productivas, de lo que haces como miembro de tu ciudad o hablarías de tu escuela, de lo que más te gusta, de lo que te gusta menos?

Toda esta información sobre tu ciudad es valiosa, pero

¿a qué características les darías prioridad al hablar?

¿Por qué?

¿Cómo la organizarías?





Supón que decides trabajar sobre las edades de los habitantes de la comunidad del llanito de tu ciudad y que su población es de 2,500 personas. Como no tienes el tiempo de ir a preguntarle su edad a cada uno de ellos, decides salir a la comunidad y entrevistar a las primeras 50 personas que te encuentras en el camino

Edades de 50 encuestados de la comunidad				
2	17	6	5	14
6	43	3	26	26
31	9	24	12	8
33	13	6	39	32
23	32	30	6	30
18	11	22	36	21
27	11	16	24	13
13	14	35	23	20
37	28	21	22	29
21	20	29	13	23

 **Actividad de Aprendizaje No 26**

¿Qué observas en este conjunto de datos? Describe al menos cinco características que encuentres. Comparte tus reflexiones con tus compañeros.

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_
4. \_\_\_\_\_
5. \_\_\_\_\_

Formas de organización de datos Tomemos los datos de edades de la sección anterior. ¿Cómo ordenarías la información? No existe una forma única de ordenar



los datos. Puede ser de menor a mayor, de mayor a menor, por decenas, entre otras.

¿Cuál es la mejor?

La que te convenga, de acuerdo con el propósito de tu investigación



**Actividad de Aprendizaje No. 27**

En comunidad selecciona una de las formas de ordenar la información de edades en la siguiente tabla, la cual puede ser de: mayor a menor, menor a mayor, por grupos de edades repetidas (por columna). Una vez terminada la tabla de tu equipo, intercambien las tablas y discutan cuál de ellas permite conocer de manera más directa:

- ¿Cuántas personas están en el rango de edad de 20 a 30 años?
- ¿Cuál es el mayor grupo de personas con la misma edad?
- ¿Encuentras de forma más fácil las cinco características que describiste en la actividad anterior?

Comparte tus conclusiones con tus compañeros de otros equipos



Edades Ordenadas de las 50 encuestas en la comunidad del Llanito				
EDAD	FRECUCIA ABSOLUTA (fa.)	FRECUCIA ABSOLUTA ACUMULADA (faa.)		
<b>2</b>			FRECUCIA EN PERSONAS DE 18 AÑOS O MAS	
<b>43</b>				
	<b>fat</b>			

El cuadro de conteo te permite obtener información para organizar y facilitar el conteo de los datos. Al total de observaciones de cada renglón (edad en nuestro ejemplo) de tu cuadro de conteo se le llama frecuencia absoluta y se denota como **fa**. En tu cuadro puedes observar, por ejemplo, que la frecuencia absoluta para la edad de 26 es de dos, mientras que para la de 27 es sólo de uno. Como puedes observar, la columna de total del cuadro de conteo se llama **fat** (frecuencia acumulada total).



Actividad de Aprendizaje No. 28

A partir de tu cuadro, identifica los siguientes grupos de personas:

- a) La población que está cursando **algún grado de primaria**, considerando que, en condiciones óptimas, **inicias la primaria a los 6 años y la concluyes a los 11.**
- b) La población que está cursando **algún grado de secundaria**. En condiciones óptimas **inicias la secundaria a los 12 años y terminas a los 14.**
- c) La población en edad de votar.

1. ¿Cuántas personas hay en cada grupo?


2. ¿Consideras que obtuviste del cuadro de conteo información más precisa de la que ya tenías sobre la comunidad y sus habitantes? Argumenta tu respuesta.

---



---

Número de habitantes que:		
Cursan la primaria	Cursan la secundaria	Están en edad de votar



3. Ahora supón que tu profesor te solicita que le presentes un breve informe al presidente municipal sobre el número de habitantes que pudieran recibir un apoyo económico del programa “Reducción de la brecha económica y social”. Este



programa del gobierno del estado atiende a las personas menores de 17 años.  
¿Cómo obtendrías esta información?

---

---

4. ¿En qué fue diferente la forma en que obtuviste la información sobre el programa “Reducción de la brecha económica y social. ” y sobre la población en edad de votar?

---

---

En el caso de la población en edad de votar, si hubieras tomado la frecuencia acumulada total y le hubieras restado la frecuencia acumulada hasta antes de la edad de 18 años tendrías el mismo resultado que obtuviste llenando la cuarta columna del cuadro. En este ejercicio resolvimos dos preguntas: en el caso del programa “Crece más” contestamos a la pregunta **menos de**, directamente con la tabla de frecuencias acumuladas. En el segundo caso, el de la población en edad de votar, resolvimos la pregunta **mayor de** con la fórmula:

**edad mayor de = edad total – observaciones acumuladas hasta antes del dato que buscamos.**

El informe que presentaste al presidente municipal le gustó. Por ello te ha pedido que realices el mismo ejercicio para los habitantes de la comunidad “Rio Laja”. Te pide que compares los resultados que obtengas con los de tu comunidad.

Con la información que le proporciones piensa establecer el monto de recursos que van a utilizar en el programa “Reducción de la brecha económica y social.” en esa localidad. Como en el caso anterior te asignó un monto para que realices la investigación. Sin embargo, otra vez no te alcanzan ni el tiempo ni los recursos para entrevistar a todos los habitantes de “Rio Laja”. Como la población de esta





comunidad es cinco veces mayor que la del Llanito, decidiste encuestar cinco veces más personas que en el caso de la comunidad del Llanito, es decir 250 personas. Para hacer tu investigación tomaste el camión hasta la comunidad de "Rio Laja", te bajaste en la terminal y entrevistaste a las primeras 250 personas que te encontraste. Los datos que obtuviste en la encuesta son los siguientes:

Edades de 250 Habitantes de " Rio Laja "																											
51	7	89	74	53	69	22	83	7	11	75	83	17	24	60	61	87	84	63	8	7	52	31	48	84			
74	70	3	37	35	12	52	29	30	49	3	75	19	5	77	63	80	37	37	69	90	48	50	45	10			
1	83	87	20	55	5	63	59	38	27	84	8	79	31	36	61	833	38	88	26	79	17	51	86	9			
72	82	30	88	69	19	43	52	44	76	48	73	12	76	85	7	28	5	48	6	68	13	79	81	89			
8	8	17	78	13	43	76	86	4	28	84	55	81	40	11	78	83	45	32	67	50	18	78	2	88			
85	65	11	63	14	81	86	76	3	25	44	21	19	68	70	41	28	63	65	84	14	45	80	44	76			
68	37	65	16	33	20	10	81	32	38	76	76	49	71	53	6	32	17	19	86	8	43	77	27	7			
73	61	62	67	4	44	28	11	88	88	9	64	90	56	10	21	34	53	37	89	37	5	54	73	84			
48	71	15	33	35	50	56	71	21	21	9	90	36	60	72	43	47	16	50	17	42	58	57	81	44			
80	51	27	66	50	89	13	62	44	44	64	19	50	5	34	84	88	23	4	60	26	20	77	53	86			

**Actividad de Aprendizaje No. 29**

1. A partir de los datos obtenidos, haz el cuadro de conteo, desarrolla, cuenta y resume tu información en la siguiente tabla de frecuencia absoluta y absoluta acumulada. (apóyate con los cuadros siguientes, puedes imprimirlos)



Edades Ordenadas de las 50 encuestas en la comunidad de RIO LAJA			Edades Ordenadas de las 50 encuestas en la comunidad de RIO LAJA		
EDAD	FRECUENCIA ABSOLUTA (fa.)	FRECUENCIA ABSOLUTA ACUMULADA (faa.)	EDAD	FRECUENCIA ABSOLUTA (fa.)	FRECUENCIA ABSOLUTA ACUMULADA (faa.)
<b>1</b>			<b>48</b>		
<b>47</b>			<b>90</b>		



2. ¿Qué semejanzas y diferencias encuentras al comparar ambas poblaciones: la de la comunidad del Llanito y la de “Rio Laja?”

---

---

---

3. ¿Consideras que esta muestra de 250 personas tiene el mismo valor que el de las edades de las 50 personas que realizaste en el Llanito? Argumenta tu respuesta.

---

---

4. ¿Cómo puedes comparar los resultados de dos comunidades de tamaños diferentes?

---

---

5. ¿Tiene más valor hacer muestras más grandes?

---

---

---

La comparación de resultados en las dos comunidades de tamaño diferente la puedes hacer obteniendo para ambas los porcentajes de frecuencia de cada una de las edades registradas. Hacer esto se llama obtener las frecuencias relativas.

Ahora responde las siguientes preguntas:

- a) ¿Cuál es el porcentaje de habitantes en edad de votar en ambas comunidades?
- b) ¿Cuál es el porcentaje de habitantes que son beneficiarios del programa “60y + años” en ambas comunidades?



	Porcentaje de habitantes que:	
	Están en edad de votar	Son beneficiarios del programa "60 y + años"
<b>El Llanito</b>		
<b>Rio Laja</b>		

5. ¿Encuentras alguna similitud entre los datos obtenidos en ambas comunidades?  
Explica Ampliamente:

---

---

---

---

---

---

---

---

---

---

---

---

---

---

6. ¿Qué ventajas tiene comparar en términos relativos las poblaciones de el Llanito y de "Rio Laja "?

- a) \_\_\_\_\_  
\_\_\_\_\_
- b) \_\_\_\_\_  
\_\_\_\_\_
- c) \_\_\_\_\_  
\_\_\_\_\_



### Sabías que...

La frecuencia relativa se obtiene aplicando una regla de tres. Y la frecuencia relativa acumulada se obtiene sumando todas las frecuencias relativas.

Como ya dijimos, la frecuencia relativa es uno de los datos que te permite hacer comparables dos o más poblaciones de diferente tamaño que comparten una misma característica por ejemplo en nuestro ejercicio la edad de 84 años su frecuencia acumulada es de 6 ( $250/100=6/x$ ).

Partes de una tabla. A la tabla que contiene la información de las frecuencias: absolutas, absolutas acumuladas, relativas y relativas acumuladas se conoce como tabla de distribución de

frecuencias. Ahora bien, una misma información puede ser representada de diferentes maneras. Como hemos visto, una de ellas es a través de tablas.



### Actividad de Aprendizaje No. 30

Observa las siguientes tablas

**OBESIDAD**

**TABLA 1. PORCENTAJE DE LA POBLACIÓN MUNDIAL MAYOR DE 15 AÑOS CON UN ÍNDICE DE MASA CORPORAL SUPERIOR A 30.**

EEUU  31%	México  24%	Gran Bretaña  23%	Rusia  22%	Grecia  22%	Australia  22%	Nueva Zelanda  21%	Hungría  19%	República Checa  15%
Canadá  14%	España  13%	Alemania  13%	Portugal  13%	Finlandia  13%	Turquía  12%	Bélgica  12%	Polonia  11%	
Holanda  10%	Suecia  10%	Dinamarca  10%	Francia  9%	Austria  9%	Italia  9%	Noruega  9%	Japón  3%	Corea  3%

Fuente: OCDE 1996-2003



**Tabla de distribución de frecuencias:** tabla donde se agrupan los datos en categorías mutuamente excluyentes que indican el número de observaciones en cada categoría.

Disponible en:

[http://www.censia.salud.gob.mx/descargas/infancia/2010/2.9. Estrategia\\_Salud\\_Alim.pdf](http://www.censia.salud.gob.mx/descargas/infancia/2010/2.9. Estrategia_Salud_Alim.pdf)

Consultado el 21 de abril de 2015



### Sobrepeso y o besidad

Grupos	México		Distrito Federal	
	%	No. Personas	%	No. Personas
Mujeres mayores de 20 años	72	20.52 millones	75.4	2.3 millones
Hombres mayores de 20 años	66	16.96 millones	69.8	69.8 millones
Niños en edad escolar	26	5.54 millones	35	481,785 millones

Fuente: Olaiz-Fernández G, Rivera-Domínguez J, Shamah-Levy T, Rojas R, Villalpando-Hernández S, Hernández-Avila M, Sepúlveda-Amor J. ENSANUT 2006. Cuernavaca, México: INSP, 2006.  
\*Según datos de INEGI, 2005

### Responde las siguientes preguntas:

1. ¿Qué información obtienes de cada tabla? \_\_\_\_\_

\_\_\_\_\_

2. Identifica las variables en cada tabla y explica de qué tipo son. \_\_\_\_\_

\_\_\_\_\_

3. ¿Qué elementos tienen en común ambas tablas?

\_\_\_\_\_

\_\_\_\_\_

4. Encierra en un círculo las frecuencias absolutas y las relativas

\_\_\_\_\_

\_\_\_\_\_

5. ¿De dónde se obtuvo la información para hacer las dos tablas?

\_\_\_\_\_

\_\_\_\_\_



Como pudiste observar, toda tabla debe contener los siguientes elementos



 **Actividad de Aprendizaje No. 31**

Una vez presentado tu informe, el presidente municipal te comenta que la información es demasiado detallada y que no le permite analizar lo que le interesa. Te comenta que a él lo que le preocupa es la educación y que quiere saber de la comunidad de Rio laja :

¿cuántas niñas y niños aún no están en edad de estudiar la primaria; cuántos estudian la primaria, cuántos la secundaria; cuántos jóvenes tienen edad de estudiar el bachillerato y cuántos pueden cursar estudios superiores?

Ante esta situación, tú, recordando una vez más a Bachelard, reflexionas:

¿Cómo resolver esta petición?

¿Cómo dividir en grupos de edades tu población?

¿Cuántos grupos proponer?

¿Deben ser del mismo tamaño cada grupo?



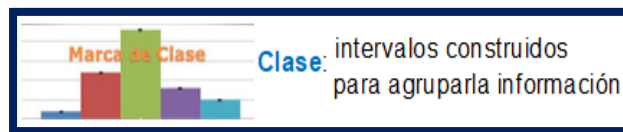
Anota tus reflexiones y discútelas con tus compañeros.

---

---

---

Como pudiste observar, toda tabla debe contener los siguientes elementos: Título Que identifica de



qué se está hablando. Valores numéricos Información que permite al lector tener un panorama general sobre el conjunto de datos. Renglones y columnas Que dividen la información en categorías. Fuente que refiere el origen de la información.

Con lo que has aprendido hasta este momento, ya puedes llenar todas las columnas, **excepto la última**, la de la **marca de clase**, que veremos un poco más adelante. Cuando termines de llenar tu cuadro estarás en condiciones de responder al presidente municipal las siguientes dudas:

1. ¿Qué proporción de la población todavía no está en edad de estudiar primaria?

---

---

---

2. Qué proporción de la población está en edad de estudiar en educación básica (primaria + secundaria)?

---

---

---

---





3. ¿Qué proporción de la población está en edad de realizar estudios de bachillerato y estudios superiores?

4. ¿Qué proporción de la población no está en edad típica para estudiar?

---

---

---

Toda esta información, aunque es mucho más reducida que el total de datos que tenías inicialmente, le permitirá al presidente municipal tomar decisiones más precisas en materia de educación. ¡Felicidades! Tu cuadro anterior contiene varios conceptos que son muy importantes:

- **Clase:** es cada uno de los grupos en los que concentras la información. (En nuestro ejemplo hay seis clases). Se denota como un intervalo.

- **Límite inferior de una clase:**

Es el valor de la variable con que inicia una clase.

- **Límite superior de una clase:** es el valor de la variable con que termina una clase.

- **Intervalo de clase:** es la distancia entre el valor del límite superior y el del límite inferior de una clase.

- **Marca de clase:** es el resultado de sumar los valores de los límites inferior y superior de cada clase y dividirlos entre dos. En pocas palabras, es el promedio de los valores de los límites inferior y superior de una clase. Con esta información, ahora sí ya puedes llenar la columna de marca de clase. Hazlo y comparte tus conclusiones con tus compañeros.

Ahora que has terminado, te darás cuenta de que en tu cuadro puedes observar que:

- Tres de las clases tienen un mismo intervalo, de 0 a menos de 6 años, de 6a menos de 12 años y de 18 a menos de 24 años.

- Dos de las clases tienen un intervalo de tres años: la que va de 12 a menos de 15 años y de 15a menos de 18 años.



•Las cinco clases anteriores se llaman clases “cerradas”, pues tienen de límite inferior y uno superior.

•En cambio, la última clase que definimos es abierta, pues incluye las edades de 24 años o más, que en nuestro ejemplo llega a 43 años.

De acuerdo con Bachelard, definimos el número de clases, su tamaño (intervalo) y las características que interesan para nuestra investigación. Eso es lo correcto. Sin embargo, en términos generales, algunos expertos proponen que es conveniente seguir, en la medida de lo posible, las siguientes reglas:

1. Hay que definir clases del mismo tamaño.
2. Puedes obtener el intervalo de clases restando la distancia entre el dato mayor y el dato menor de toda tu información y dividiendo ese resultado entre el número de clases que deseas formar.
3. La definición de los límites superior e inferior de una clase es distinta para variables discretas y variables continuas. En el primer caso, como la variable sólo toma valores “a saltos”, el cuadro podría quedar como sigue

Población que reporta número de años de estudio terminados		
Clases		
Límite inferior		Límite superior
0	-	6
7	-	12
13	-	15
16	-	21

En el caso de las variables continuas, el límite superior de una clase debe concluir en “menos de \_\_\_\_\_”, o el límite inferior en “más de \_\_\_\_\_”



1. Completa la siguiente tabla de frecuencias para la comunidad “Rio Laja”
2. ¿Cómo puedes comparar la distribución de las edades de la población de la comunidad del Llanito y la de “Rio Laja”?

---



---



---



---

 **Actividad de Aprendizaje No. 32**



En esta actividad tendrás que realizar 30 encuestas ( por comunidad) , puede ser entre ustedes o bien, con quienes ustedes quieran. Es importante tener las encuestas con las respuestas, posteriormente se les darán las indicaciones para realizar sus tablas ya que con estas iniciaremos a realizar los gráficos en base a tus tablas.

Población escolar en edad preescolar y en edad típica de estudiar en la comunidad de Rio Laja ( Distribucion de Frecuencias Organizadas en Clase)								
Clases			Cuadro de conteo	Frecuencia absoluta $f_n$	Frecuencia absoluta acumulada $f_{na}$	Frecuencia relativa $f_r$	Frecuencia relativa acumulada $f_{ra}$	Marca de clase $m_i$
Límite inferior	-	Límite superior						
0	-	menos de 6						
6	-	menos de 12						
12	-	menos de 15						
15	.	menos de 18						
18	-	menos de 24						
24	-	mayores de 24						
Suma / Total								



## Encuesta de entretenimiento.

Agradecería que contesten esta encuesta para realizar un trabajo de investigación de mi materia de Pensamiento Matemático

Instrucciones: Contesten marcando una opción en cada pregunta.

### 1 ¿Cuál es el tipo de música que te gusta escuchar?

- a. grupera
- b. rock
- c. cumbia
- d. clásica
- e. balada

### 2 ¿Cuál es el tipo de programa que te gusta ver en la televisión?

- a. noticias
- b. comedias
- c. caricaturas
- d. musicales
- e. concursos

### 3 ¿Cuál es el deporte que te gusta practicar?

- a. basquetbol
- b. futbol
- c. natación
- d. volibol

### 4 ¿Cuál es el tipo de película que te gusta ver?

- a. suspenso
- b. terror
- c. comediad.
- d. drama
- e. infantil

**Sabias que .....**

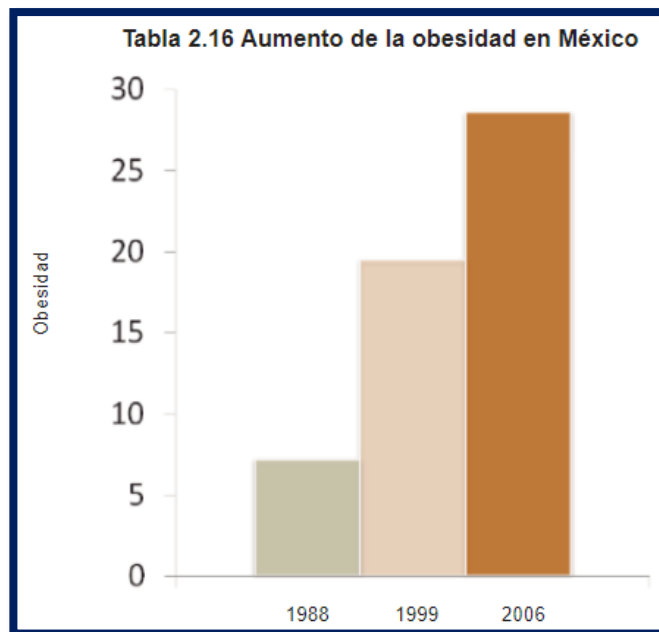


**Has escuchado la frase:** “una imagen vale más que mil palabras” en este caso, una imagen vale más que mil números. Como ya sabemos, no existe una forma única de representar los datos, pero el impacto visual de una gráfica permite sintetizar y explicar fácilmente lo que te interesa transmitir de la masa de datos.

### Tipos de tablas y su uso

Todo valor de tu conjunto de datos debe pertenecer exclusivamente a una clase. No puede haber un valor en más de una clase y no puede quedar fuera del cuadro ningún valor.

Observa las gráficas que te presentamos a continuación y responde:



Disponible en:

<http://www.censia.salud.gob.mx/descargas/infancia/2010/2.9. Estrategia Salud>



Actividad de Aprendizaje No. 33



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE  
TECNOLOGÍA, INVESTIGACIÓN Y SERVICIOS



2023  
**Francisco  
VILLA**



Alim.pdf Consultada el 21 de abril de 2015.

Disponible

en:

<http://www3.inegi.org.mx/sistemas/temas/default.aspx?s=est&c=17484>



Tabla 2.17 Población total por miles por grupo de edad y sexo en 2008 y 2012 de un estado de la República Mexicana

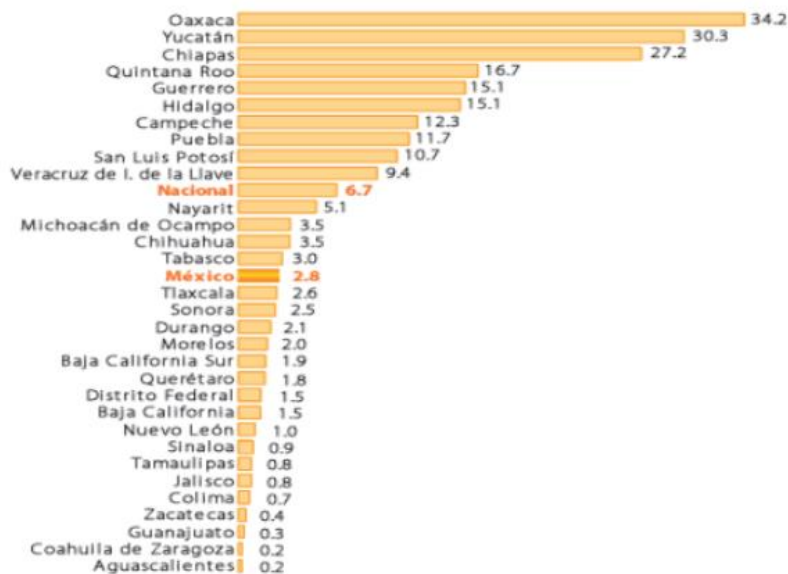


Consultado el 4 de junio de 2015

Disponible en:

<http://cuentame.inegi.org.mx/monografias/informacion/mex/poblacion/diversidad.aspx?tema=me&e=15>

Tabla 2.18 Porcentaje de la población que habla lengua indígena por entidad federativa (año 2010)





Consultado el 4 de junio de 2015.! Tabla 2.18 Porcentaje de la población que habla lengua indígena por entidad federativa (año 2010)

1. ¿Qué variables hay en cada una de las gráficas?

---

---

2. ¿Qué información te brinda cada una de estas gráficas?

---

3. ¿Qué tipo de frecuencia utilizan en cada una de ellas?

---

---

4. ¿Qué similitudes y diferencias hay entre las gráficas?

---

---

Concentrados los datos en una tabla y dependiendo de su naturaleza y finalidad, se puede hacer la representación gráfica correspondiente. Un gráfico estadístico es la representación de datos estadísticos por medio de figuras geométricas (puntos, líneas, rectángulos, etc.) cuyas dimensiones son proporcionales al valor numérico de los datos. Su fin principal es permitir, de un solo vistazo la captación rápida del conjunto de características presentadas y evidenciar sus variaciones en intensidad.

El gráfico es útil para dar una rápida idea de la situación general que se está analizando; permite determinar, por un simple examen, el máximo y el mínimo de las variaciones de un fenómeno, poniendo de relieve lo que debe tomarse en cuenta especialmente.





Aunque existen muchos tipos de gráficos, estudiaremos únicamente los más usuales: algunos gráficos de barra y línea, el gráfico circular y el pictograma.

Salvo el gráfico circular y el pictograma, los de barra y de línea se dibujan en el marco de un sistema rectangular de coordenadas, en el cual, dicho sea de paso, el cuadrante I es el más utilizado, debido a que la inmensa mayoría de las variables que se estudian generan cantidades positivas.

## GRAFICA DE BARRA

Es uno de los mejores para realizar comparaciones de datos estadísticos, porque además de representar los valores absolutos o relativos de los datos en sí, da una imagen de cómo se reparten los elementos del conjunto respecto al total. La construcción de este gráfico se basa en la representación de un valor numérico por un rectángulo, cuya longitud es proporcional a ese valor.

Lo más importante es la determinación de la longitud o altura de los rectángulos que representan los valores de los datos, ya que del cálculo correcto de esta medida depende la comparación de los valores.

El cálculo se realiza mediante Regla de tres simples, una vez establecida la correspondiente igualdad entre una unidad de valor y una unidad de medida. La igualdad se establece entre el dato mayor y una medida determinada.

**Ejemplo 1:** Realiza un grafica de barras con tabla siguiente



2. Con los datos siguientes que representan el tiempo dedicado al estudio fuera de clases en horas semanales, por estudiantes universitarios.

Hrs. Semanales	F
1 – 3	18
4 – 6	18
7 – 9	7
10 – 12	12
13 – 15	5
16 – 18	6
19 – 23	4

### Ejemplo 2:

1. De la tabla siguiente construir un gráfico de barras

Escolaridad de la Población  
de los mayores de seis años  
México 1970

Escolaridad	Total (millones)
	38.37
Primaria	21.39
Sexto año	4.95
Posterior	3.60
Ninguna	13.36

Fuente: Vol. 8 Enciclopedia de México 3ª. Ed. 1978 p. 495



De la tabla de frecuencias para la comunidad “Rio Laja” realizar su grafica

Población escolar en edad preescolar y en edad típica de estudiar en la comunidad de Rio Laja ( Distribucion de Frecuencias Organizadas en Clase)	
INTERVALOS	f
0 - menos de 6	
6 - menos de 12	
12 - menos de 15	
15 - menos de 16	
18 - menos de 24	
24 - mayores de 24	

### GRAFICO CIRCULAR

Se le llama también gráfico de pastel, y es bastante útil para representar proporciones o porcentajes. Es de hecho, una forma alternativa al gráfico de barras para representar una distribución.

En su construcción se utiliza una circunferencia, cuyo círculo se divide en sectores tales que sus medidas angulares sean proporcionales a los valores que representan.

#### Ejemplo 1:

Construir un gráfico circular que represente la información de la tabla siguiente



Presupuestos para 6 municipios de los  
Más importantes del Estado de Veracruz. 1991  
(millones de pesos)

Municipio	Total
	137 437
Coatzacoalcos	35,363
Córdoba	13,309
Orizaba	12,885
Poza Rica	10,644
Veracruz	42,592
Xalapa	22,644

Aprobado por la Legislatura del Estado

Fuente: Legislatura del Estado de Veracruz Dirección de contaduría y Glosa, 1991

Para construir el grafico circular del caso de la comunidad del Llanito, sería muy complejo el realizarlo sobre 50 casos, por lo que estableceremos como obtener los intervalos de clase para una mejor comodidad en la gráfica.

De la Comunidad del Llanito tenemos el siguiente procedimiento:

1) **Recorrido de la variable:** Se calcula restando:

$$\text{Dato Mayor} - \text{Dato Menor}$$

$$43 - 2 = \underline{41}$$

2) **Datos Potenciales:** se Calcula

$$\text{Dato Mayor} - \text{Dato Menor} + 1$$

$$43 - 2 = 41 + 1 = \underline{42}$$

3) **Numero de intervalos que tendrá su tabla de frecuencias** (es un número que divida exactamente a sus datos potenciales), **No debe ser menor de 5 ni mayor de 20**

$$42 / 7 = 6 \text{ a esto se le conoce como anchura y su símbolo es } j = 6$$

4) Para los intervalos de la tabla **el Limite Inferior siempre es el dato menor**, en este caso es = 2 para el Limite mayor se calcula así:

$$\text{Límite Inferior} + (j - 1) = \text{Limite Superior}$$



$2 + (6-1) = 2 + 5 = 7$  Limite Superior

Por lo tanto, la tabla quedaría de esta manera

Edades Comunidad del Llanito	
intervalos	f
2--7	
8--13	
14--19	
20--25	
26--31	
32--37	
38--43	

Construir un gráfico circular que represente la información de la tabla siguiente

Escolaridad de la Población de los mayores de seis años México 1970	
Escolaridad	Total (millones)
	38.37
Primaria	21.39
Sexto año	4.95
Posterior	3.60
Ninguna	13.36

Fuente: Vol. 8 Enciclopedia de México 3ª. Ed. 1978 p. 495



## HISTOGRAMA Y POLIGONO DE FRECUENCIAS

Se da nombre de histogramas a los gráficos de barras cuando representan variables cardinales principalmente continuas. Si se unen con segmentos de recta los puntos medio de los techos de los rectángulos, resulta un polígono de frecuencias.

Aunque el procedimiento de construcción de un histograma es básicamente el mismo que para cualquier gráfico de barras, tiene algunas características que se muestran a continuación

**Construir tanto el histograma como el polígono de frecuencias de la tabla siguiente.**

<i>Años</i>	<i>Total</i>	<i>%</i>
18 – 21	10	14
22 – 25	9	13
26 – 29	19	27
30 – 33	12	17
34 – 37	9	13
38 – 41	4	6
42 – 45	7	10

**Fuente: Datos Supuestos**

Procediendo como en los casos anteriores

- a) Representa categoría cardinal por lo que se *usaran siete rectángulos sin separarse*
- b) Fijamos el ancho 1.5 cm medida entre el origen de coordenadas y el extremo 12 cm
- c) Se determina los  $\frac{3}{4}$  con la medida resultante

$$\frac{(3)}{4} (12) = \frac{36}{4} = 9 \text{ cm altura de la barra}$$

- d) Del dato mayor (frecuencia más grande) con la altura de la barra

$$\frac{9}{19} = .47 \quad \text{Factor} = .47$$



e) División del eje vertical, múltiplo más cercano por el factor (20)

(20) (.47) = 9.4  
9.4 entre 5 partes  $\frac{9.4}{5} = 1.88$  de 4 a 20

5

Factor	Altura
(.47) (10) =	4.7 cm
(.47) (9) =	4.2 cm
(.47) (19) =	9 cm
(.47) (12) =	5.6 cm
(.47) (9) =	4.2 cm
(.47) (4) =	1.8 cm
(.47) (7) =	3.2 cm

A eje horizontal (Marca de Clase)

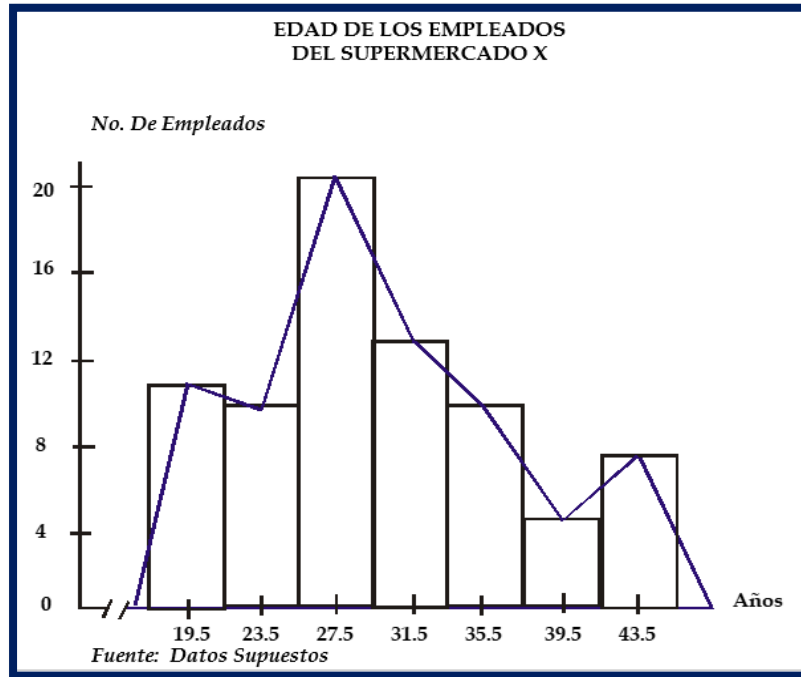
A eje horizontal (Marca de Clase)

Marca de Clase =  $\frac{\text{Frontera superior} + \text{frontera inferior}}{2}$

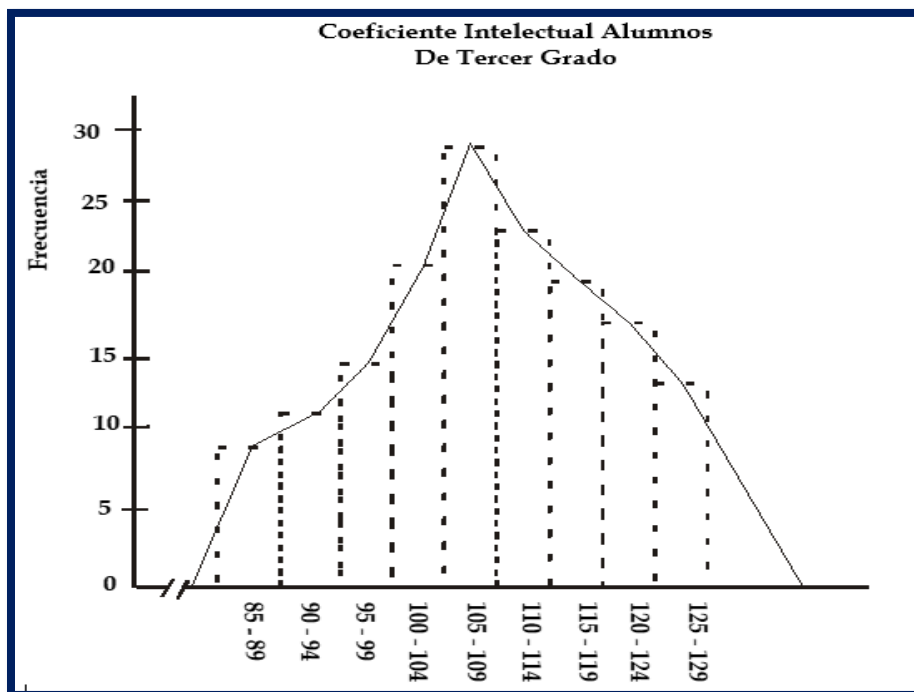
de los intervalos de los años tenemos:

Marcas de clase

18 – 21	mc = $\frac{21 + 18}{2} = 19.5$
22 – 25	mc = $\frac{25 + 22}{2} = 23.5$
26 – 29	mc = $\frac{29 + 26}{2} = 27.5$
30 – 33	mc = $\frac{33 + 30}{2} = 31.5$
34 – 37	mc = $\frac{37 + 34}{2} = 35.5$
38 – 41	mc = $\frac{41 + 38}{2} = 39.5$
42 – 45	mc = $\frac{45 + 42}{2} = 43.5$



Se puede construir un polígono de frecuencia dibujando primero un histograma y luego conectando por medio de líneas rectas los puntos medios de la parte superior de cada una de las celdas, el boceto del histograma se suprime generalmente en los polígonos de frecuencia, del ejemplo siguiente tenemos.







## GRAFICO DE LINEAS

Es bastante útil para comparar los datos de dos o más distribuciones.

Consiste en unir, por medio de segmentos de recta, los puntos de coordenadas determinados por los datos de dos variables que se corresponden o de variables que dependen del tiempo.

En su construcción también se aplica la regla de los  $\frac{3}{4}$  de altura.

El trazo del gráfico puede o no comenzar en el eje de coordenadas; esto no tiene importancia y depende del diseño del gráfico, ejemplo:

Mediante un gráfico de línea, presentar la información de la tabla siguiente:

### Población Rural y Urbana

Año	Total	Urbana	Rural
1900	13.6	2.6	11.0
1910	15.2	3.7	11.5
1921	14.4	4.5	9.9
1930	16.5	5.5	11.0
1940	19.7	6.9	12.8
1950	25.8	11.0	14.8
1960	34.9	17.7	17.2
1970	50.7	29.8	20.9
1979	67.9	44.6	23.3

1970: Población corregida y proyectada al 30 de junio de 1970

1979: Estimaciones del Consejo Nacional de Población

Población rural: Menos de 2500 habitantes

Fuente: "México Demográfico, Breviario, 1979, 1979", p. 44



**Solución:**

A) La población es función del tiempo. Es decir, la magnitud tiempo es la variable independiente; por lo tanto; según las normas convencionales, se debe ubicar las unidades de tiempo en el eje horizontal y los datos de la población en el vertical.

B) Trazados los ejes de coordenadas, se marcan puntos igualmente espaciados para los años dados. Luego se mide la distancia entre el origen de coordenadas y el punto correspondiente al último dato en el eje de las abscisas (supongamos, 16 cm)

$$\frac{\text{(altura)}}{\text{(dato mayor)}} = \frac{12}{67.9} = 0.1767 \text{ (factor) (No. Em. Por cada millón de habitantes)}$$

C) Se determinan los  $\frac{3}{4}$  de la distancia anterior y se establece una relación entre la longitud resultantes y el de mayor de la serie, que se localiza en la columna de totales.  $(\frac{3}{4})(16) = \frac{48}{4} = 12 \text{ cm (altura mayor)}$

E) Se calculan las alturas de los datos de población usando el factor constante.

Totales	Urbana	Rural
13.6 (.1767) = 2.4 cm	2.6 (.1767) = .5	11.0 (.1767) = 2
15.2 (.1767) = 2.7 cm	3.7 (.1767) = .6	11.5 (.1767) = 2.03
14.4 (.1767) = 2.5 cm	4.5 (.1767) = .8	9.9 (.1767) = 1.7
16.5 (.1767) = 3	5.5 (.1767) = 1	11.0 (.1767) = 2
19.7 (.1767) = 3.5	6.9 (.1767) = 1.2	12.8 (.1767) = 2.3
25.8 (.1767) = 4.5	11.0 (.1767) = 1.9	14.8 (.1767) = 2.6
34.9 (.1767) = 6.1	17.7 (.1767) = 3.1	17.2 (.1767) = 3
50.7 (.1767) = 9	29.8 (.1767) = 5.3	20.9 (.1767) = 3.7
67.9 (.1767) = 12	44.6 (.1767) = 7.9	23.3 (.1767) = 4.1

D) Se aproxima 67.9 a 70 y se multiplica este último número por el factor encontrado (0.1767), para saber el número de unidades de longitud que le corresponden

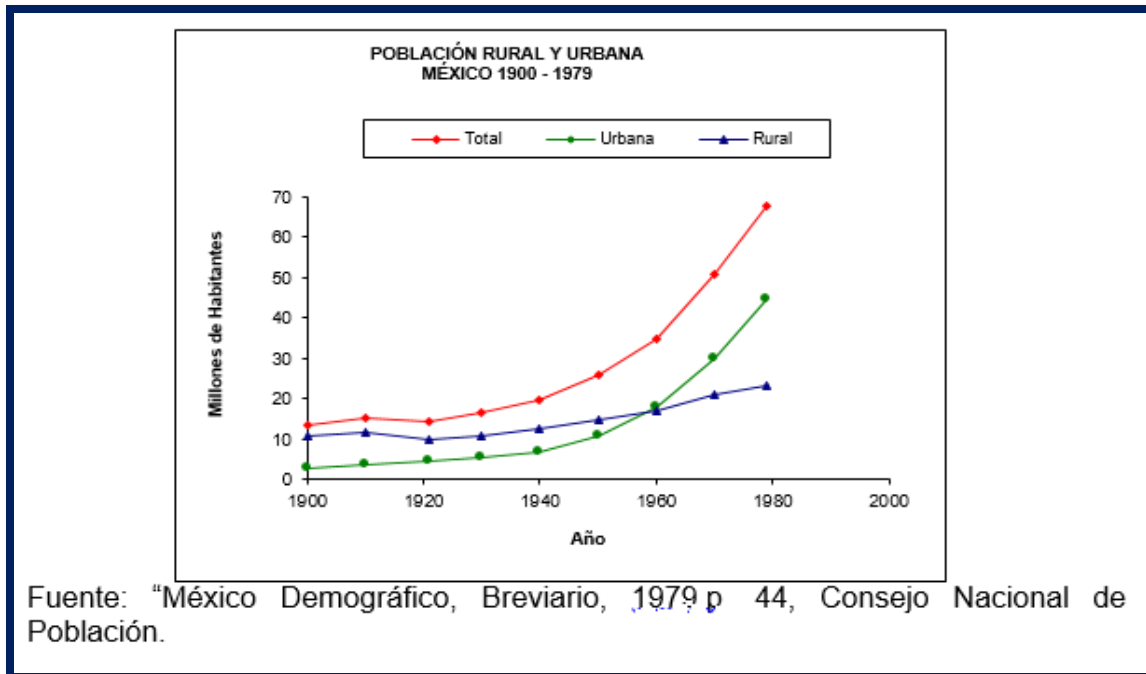
$$(70)(0.1767) = 12.4 \text{ cm}$$



- E) Los 12.4 cm los partimos en 7 partes iguales, para hallar la longitud equivalente 10 millones de habitantes.

$$\frac{12.4}{7} = 1.8 \text{ de } 10 \text{ a } 70$$

- F) Fijamos los puntos correspondientes a las tres distribuciones de datos (total, urbana, y rural) se traza el gráfico.



## Ejercicio

Mediante un gráfico de línea, representar la información de la tabla siguiente:



EGRESADOS POR GENERACIÓN, FECHA DE EGRESO Y SEX  
FAC. DE SOCIOLOGÍA, U.V. (Feb 81 – Ago 83)

Generación	Fecha	Total 137	Hombres 80	Mujeres 57
1 <sup>a</sup> .	Feb. 81	28	17	11
2 <sup>a</sup> .	Ago. 81	38	23	15
3 <sup>a</sup> .	Feb. 82	21	11	10
4 <sup>a</sup> .	Ago. 82	6	5	1
5 <sup>a</sup> .	Ago. 83	44	24	20

Fuente: Anuario Estadístico 1984, Fac. de Sociología, U.11



## PENSAMIENTO MATEMÁTICO

### PROGRESIÓN No. 8

Analiza cómo se relacionan entre sí dos o más variables categóricas a través del estudio de alguna problemática o fenómeno de interés para el estudiante, con la finalidad de identificar si dichas variables son independientes.

METAS	CATEGORÍAS	SUBCATEGORÍAS
M3 Compara hechos, opiniones o afirmaciones para organizarlos en formas lógicas útiles en la solución de problemas y explicación de situaciones y fenómenos	C2 Procesos de intuición y razonamiento	S1 Capacidad para observar y conjeturar
M4 Argumenta a favor o en contra de afirmaciones acerca de situaciones, fenómenos o problemas propios de la matemática, de las ciencias o su contexto.		S2 Pensamiento intuitivo S3 Pensamiento formal

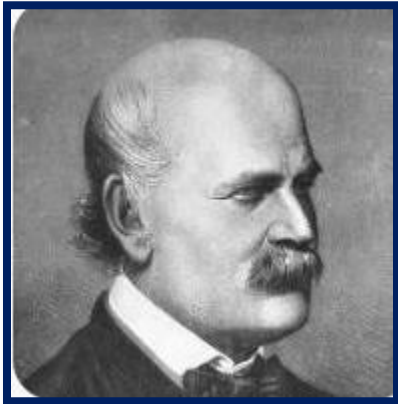


## Progresión No. 8



Para iniciar reflexiona

### Ignaz Semmelweis



Fue un obstetra húngaro que a mediados del siglo XIX, precediendo los hallazgos de Pasteur y Lister, logró descubrir la naturaleza infecciosa de la fiebre puerperal, logrando controlar su aparición con una simple medida de antisepsia. Debió luchar con la reticencia de sus colegas que no aceptaron sus observaciones que, por primera vez en la historia, fueron apoyadas con datos estadísticos. Esta comunicación describe datos biográficos de este

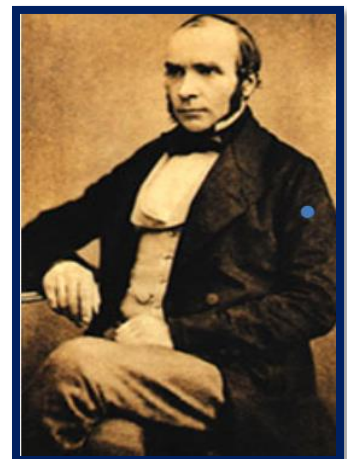
trascendente científico y las circunstancias que rodearon su extraña muerte, apoyada en información infrecuentemente divulgada.

Disponible en : [https://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0716-10182008000100011#:~:text=Ignaz%20Semmelweis%2C%20fue%20un%20obstetra,una%20simple%20medida%20de%20antisepsia.](https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0716-10182008000100011#:~:text=Ignaz%20Semmelweis%2C%20fue%20un%20obstetra,una%20simple%20medida%20de%20antisepsia.)

Consultado el 15/04/2023

Demostró que el cólera era causado por el consumo de aguas contaminadas con materias fecales; William Farr fue el pionero en el desarrollo de las estadísticas y bases de datos sobre salud y mortalidad; e Ignaz Semmelweis hizo un aporte trascendental al desarrollo de la epidemiología al hacer el primer trabajo donde se evaluaron los factores de riesgo asociados al desarrollo de una enfermedad.

### John Snow





La epidemiología es una parte de la medicina que estudia la distribución, frecuencia, determinantes, relaciones, predicciones y control de factores relacionados con la salud y enfermedad en poblaciones humanas determinadas, así como la aplicación de este estudio a los problemas de salud.

Disponible en : <https://www.savalnet.cl/mundo-medico/reportajes/9108.html>

Consultado el 15/04/2023



La epidemiología es una parte de la medicina que estudia la distribución, frecuencia, determinantes, relaciones, predicciones y control de factores relacionados con la salud y enfermedad en poblaciones humanas determinadas, así como la aplicación de este estudio a los problemas de salud.

En sus inicios, esta ciencia centró su interés en el estudio de procesos infecciosos transmisibles (pestes) que afectaban a grandes grupos humanos. Estas enfermedades, llamadas epidemias, causaron un número importante de muertes frente a las cuales la medicina de entonces no tenía soluciones efectivas para ofrecer.

Se le reconoce al **médico inglés John Snow** la paternidad de esta rama de la medicina al utilizar espléndidamente el método científico, aportando importantes avances al conocimiento de la epidemia de cólera que, en aquella época, afectaba a la ciudad de Londres.

**John Snow** fue el primero en postular que las aguas servidas eran el vehículo transmisor del cólera, cuando en el otoño de 1848 llegó a Londres el vapor alemán Elba, procedente de Hamburgo, que diseminó este mal a lo largo de la ribera del Támesis. Su primera publicación apareció en 1849 y tanto su precisión, la lógica de su discurso y la belleza de algunas frases, así como las decididas acciones del autor tuvieron una gran acogida, en una época en la que se pensaba que las enfermedades se transmitían por la inhalación de los vapores de personas enfermas.



Cuenta la historia que Snow “una tarde se sentó con el mapa de Londres, en donde un reciente brote epidémico había matado a más de 500 personas en un espantoso período de 10 días. Marcó las localizaciones de los hogares de aquellos que habían muerto. Por las marcas sobre su mapa pudo ver que todas las muertes habían ocurrido en el área de Goleen Square. La más marcada diferencia entre este distrito y el resto de Londres era la fuente de su agua potable (...) que estaba obtenido su agua de una sección del río Támesis donde era conocido que estaba especialmente contaminado (...) una vez que la bomba fue puesta fuera de uso, se abatió la epidemia”.

Con esta investigación, el médico inglés demostró una estrecha relación entre los casos de cólera y las fuentes de suministro de agua. Su identificación de grupos supuso una selección de intervenciones para mitigar o eliminar los factores de riesgo y su evaluación. El caso de la epidemia de cólera investigada por Snow, implicó la clausura de la bomba de agua de la calle Broad (en la actualidad Broadwick Street) al descubrir que el agua potable estaba contaminada con aguas de drenaje que contenían descargas de personas que sufrían de cólera, por lo que concluyó que las descargas intestinales de los pacientes de cólera llevaban algo que había contaminado el agua y que se había convertido en causa directa de la epidemia.

Lo inusual en este caso fue que Snow fue más allá de examinar pacientes individuales y decidió tratar la aparición de la enfermedad como un problema de salud pública. Salió de su oficina, pasó delante de los pacientes en la sala de espera y empezó a pasearse por la calle Broad. Así nació el mito sobre el que se fundamenta el movimiento de la salud pública, la promoción de la salud y la epidemiología como ciencia.

Asimismo, las acertadas conclusiones de Snow acerca de la etiología, forma de transmisión y control de la enfermedad se anticiparon a los progresos que en este sentido hicieron la microbiología, la infectología y la clínica.





Actualmente todos somos conscientes de la importancia que tiene lavarse las manos para prevenir infecciones. Pero un gesto que ahora nos parece tan lógico (y que ha salvado tantas vidas), no lo era hace apenas 150 años.

**El autor de esta brillante y maravillosa idea fue un médico húngaro de origen alemán llamado Ignaz Philipp Semmelweis.** Nacido el 1 de julio de 1818, este médico es mundialmente conocido como "el salvador de las madres", ya que fue él quien **descubrió que desinfectarse las manos antes de tratar a las parturientas disminuía drásticamente los casos de muerte** de mujeres y recién nacidos a causa de la fiebre puerperal.

## UNA BATALLA PERDIDA

En 1847, Semmelweis propuso a sus colegas **lavarse las manos antes de atender a las pacientes en el hospital obstétrico de Viena** donde ejercía, ya que las cifras de muertes de mujeres que acababan de dar a luz eran aterradoras. Pero **a pesar de demostrar la eficacia de su método**, ya que las muertes disminuían drásticamente tras un minucioso lavado de manos, **sus ideas fueron denostadas por sus colegas**, que lo tildaron de loco y de charlatán. Además, jamás le perdonaron que de algún modo los acusara de ser los responsables de la muerte de sus pacientes

**Al final, fue despedido y años después cayó en una profunda depresión.** Se dio a la bebida, y su comportamiento errático preocupó a sus allegados y a su esposa, que creyeron que estaba perdiendo la razón (se ha dicho que tal vez padecía de demencia precoz o Alzheimer). Al final, **en 1865, Semmelweis fue internado en un hospital psiquiátrico donde murió dos semanas después de su ingreso a causa de una septicemia, a los 47 años.**

Representar cartográficamente un brote histórico de cólera



En 1854, Londres estaba inmersa en una **pandemia de cólera**. En ese momento, existía información limitada de cómo los microbios y otros microorganismos pueden causar enfermedades. La mayoría de la población creía que el cólera se propagaba por miasmas o un vapor tóxico, a través del aire. El Dr. John Snow hizo algo que los investigadores y analistas de SIG hacen hoy en día: recopiló datos y cartografió información para buscar patrones. Explorará los datos capturados por John Snow con ventanas emergentes y simbología de mapa de calor para ilustrar si el cólera se propaga por el aire o de otra forma.

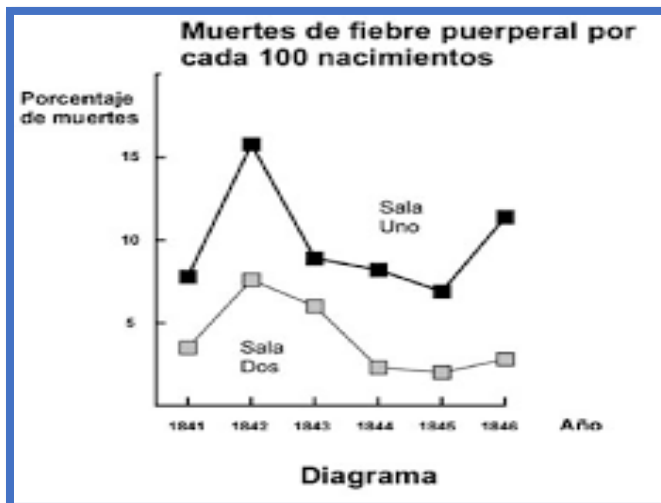
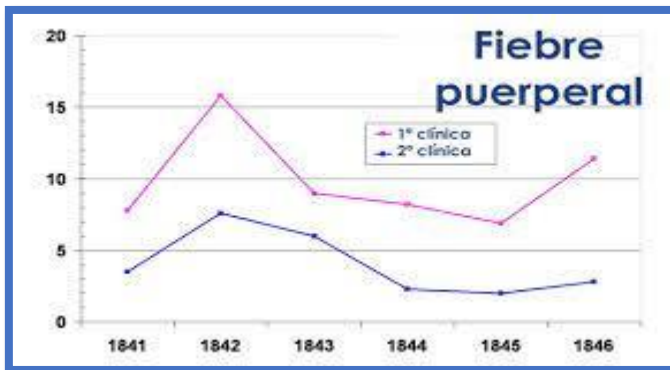
## Explorar datos del cólera

El SIG le permite comparar datos geográficos en forma de capas. Cada capa de un mapa contiene información diferente que puede comparar para ver y entender mejor los patrones y cómo influyen entre sí. Empezará por usar Visor de mapas para explorar las capas de un mapa existente que muestran datos relacionados con el brote del cólera de 1854.





## Graficos de Pandemias Ignaz Philipp Semmelweis





## Graficos de Pandemias

### John Snow



Mapa del Colera

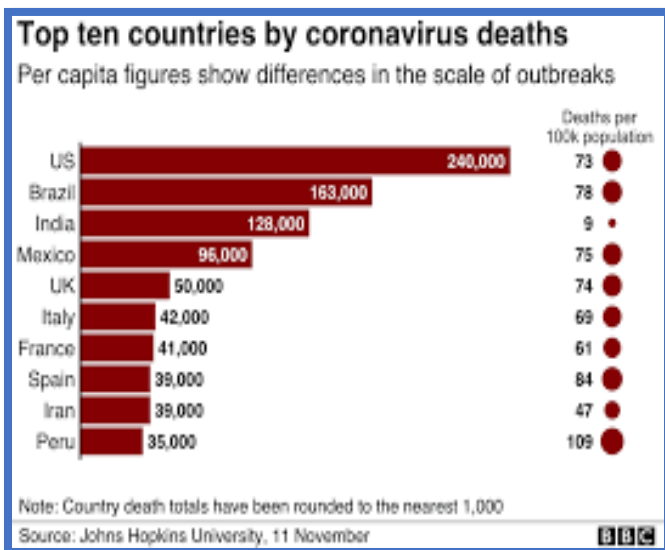
Tabla 1. Análisis de John Snow de la epidemia de cólera de 1853-54<sup>1</sup>

Compañía de agua	Hogares	Muertes por cólera	Muertes por 10.000 hogares
S-V	40.046	1.263	315
L	26.107	98	37
Londres (resto)	256.423	1.422	59

S-V = Southwark and Vauxhall Water Company  
L = Lambeth Water Company.

## Graficos de Pandemias

### Otros





 **Actividad de Aprendizaje No 34**

 **Aprende Mas**

Después de conocer a nuestros personajes, te corresponde contestar las siguientes preguntas de cada uno de ellos para complementar nuestra progresión:

1.- ¿Cuál fue el aporte de John Snow a la epidemiología?

---

---

---

---

---

---

2.- ¿Qué importancia tuvo el estudio de John Snow sobre el cólera?

---

---

---

---

---

---

3.- ¿Quién fue John Snow y cuál fue su teoría?

---

---

---

---

---

---



4.- ¿Quién fue John Snow en epidemiología?

---

---

---

---

---

---

5.- ¿Qué resultados obtuvo Semmelweis?

---

---

---

---

---

---

6.- ¿Qué factores influyeron en la investigación de Semmelweis?

---

---

---

---

---

---

7.- ¿Cuál fue el aporte de Semmelweis a la medicina?

---

---

---

---

---

---

8.- ¿Como hizo Semmelweis para poner a prueba sus hipótesis?

---

---

---

---

---

---

9.- ¿Qué método aplicó Semmelweis?

---

---

---

---

---

---



## PENSAMIENTO MATEMÁTICO

### PROGRESIÓN No. 9

Analiza cómo se relacionan entre sí dos o más variables categóricas a través del estudio de alguna problemática o fenómeno de interés para el estudiante, con la finalidad de identificar si dichas variables son independientes.

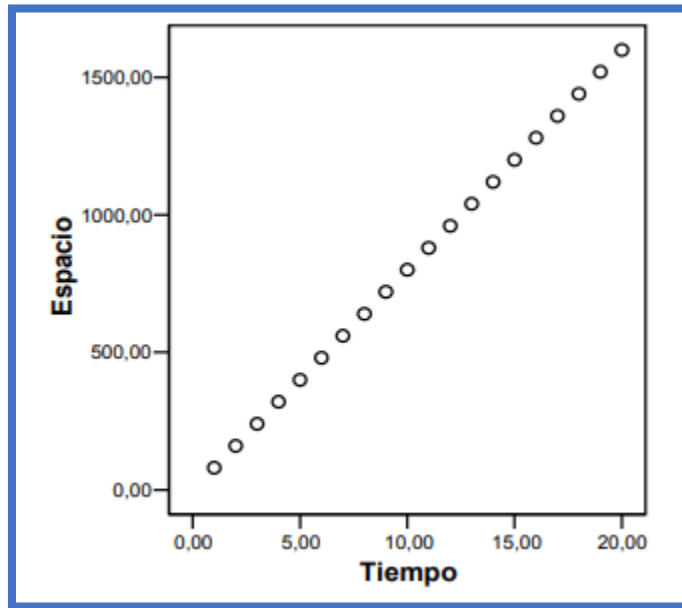
METAS	CATEGORÍAS	SUBCATEGORÍAS
M3 Compara hechos, opiniones o afirmaciones para organizarlos en formas lógicas útiles en la solución de problemas y explicación de situaciones y fenómenos	C2 Procesos de intuición y razonamiento	S1 Capacidad para observar y conjeturar
M4 Argumenta a favor o en contra de afirmaciones acerca de situaciones, fenómenos o problemas propios de la matemática, de las ciencias o su contexto.		S2 Pensamiento intuitivo S3 Pensamiento formal



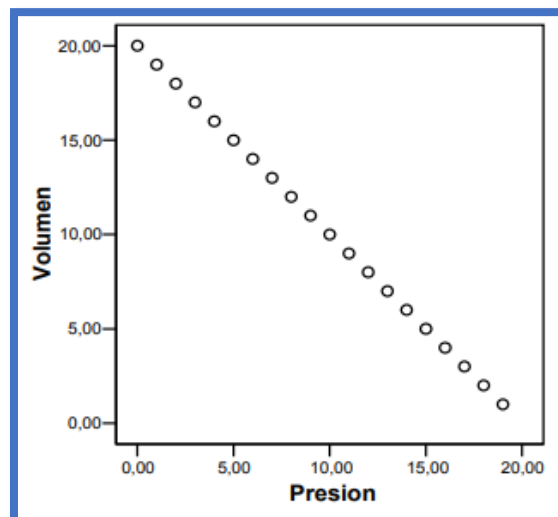
## Coeficiente de correlación lineal de Pearson

El coeficiente de correlación de Pearson, pensado para variables cuantitativas (escala mínima de intervalo), es un índice que mide el grado de covariación entre distintas variables relacionadas linealmente. Adviértase que decimos "variables relacionadas linealmente". Esto significa que puede haber variables fuertemente relacionadas, pero no de forma lineal, en cuyo caso no proceder a aplicarse la correlación de Pearson. Por ejemplo, la relación entre la ansiedad y el rendimiento tiene forma de U invertida; igualmente, si relacionamos población y tiempo la relación será de forma exponencial. En estos casos (y en otros muchos) no es conveniente utilizar la correlación de Pearson. Insistimos en este punto, que parece olvidarse con cierta frecuencia. El coeficiente de correlación de Pearson es un índice de fácil ejecución e, igualmente, de fácil interpretación. Digamos, en primera instancia, que sus valores absolutos oscilan entre 0 y 1. Esto es, si tenemos dos variables X e Y, y definimos el coeficiente de correlación de Pearson entre estas dos variables como  $r$  entonces: Hemos especificado los términos "valores absolutos" ya que en realidad si se contempla el signo el coeficiente de correlación de Pearson oscila entre  $-1$  y  $+1$ . No obstante ha de indicarse que la magnitud de la relación vienen especificada por el valor numérico del coeficiente, reflejando el signo la dirección de tal valor. En este sentido, tan fuerte es una relación de  $+1$  como de  $-1$ . En el primer caso la relación es perfecta positiva y en el segundo perfecta negativa. Pasamos a continuación a desarrollar algo más estos conceptos. Decimos que la correlación entre dos variables X e Y es **perfecta positiva** cuando exactamente en **la medida que aumenta una de ellas aumenta la otra**. Esto sucede cuando la relación entre ambas variables es funcionalmente exacta. Difícilmente ocurrirá en psicología, pero es frecuente en las ciencias físicas donde los fenómenos se ajustan a leyes conocidas, Por ejemplo, la relación entre espacio y tiempo para un móvil que se desplaza a velocidad constante. Gráficamente la relación ser del tipo:





Se dice que la **relación es perfecta negativa** cuando exactamente **en la medida que aumenta una variable disminuye la otra**. Igual que en el caso anterior esto sucede para relaciones funcionales exactas, propio de las ciencias físicas. Por ejemplo, la relación entre presión y volumen se ajusta a este caso. El gráfico que muestra la relación sería del tipo:





## Ejemplos

### ESTUDIOS DE MERCADO

Cuando estamos analizando los datos de un estudio de mercado, muchas veces nos encontramos con la necesidad de saber si entre dos variables de tipo cuantitativo existe algún tipo de relación. Por ejemplo, a la hora de evaluar un producto o un servicio de una compañía, podemos querer saber si existe alguna relación entre la puntuación que se le ha dado a ese producto y el nivel de ingresos. Hay diferentes formas de analizar estos datos, pero una de ellas es comprobar si existe correlación entre esas dos variables.

#### Artículos relacionados:

- [¿Qué métodos científicos diferentes existen?](#)
- [¿Qué es la escala de Likert y cómo utilizarla en tus encuestas?](#)
- [Métodos para calcular precios en productos nuevos](#)

Más específicamente, **las preguntas que podemos hacernos son tres:**

1. ¿Existe alguna relación entre esas dos variables? Dicho de otro modo, si cambian los valores en una de esas dos variables, ¿cambian también en la otra? Una respuesta afirmativa a esta cuestión significaría que las variables están “asociadas”, es decir, que existe una correlación entre ellas.
2. ¿En qué sentido se da esa correlación? Por ejemplo, si el nivel de ingresos es más alto, ¿es también más alta la valoración que se le da al producto o servicio? Si es así, la asociación entre ellas es positiva, es decir, las dos variables no sólo están asociadas, sino que, además, cambian o se mueven en la misma dirección. También puede ocurrir lo contrario: un incremento en los valores de una, se vincula con un descenso en los valores de la otra. Hay correlación, pero es negativa, de sentido inverso.



3. Finalmente, ¿qué fuerza tiene esa correlación? ¿Es alta o es baja? Hay variables que están solo ligeramente correlacionadas, o lo están a un nivel medio, pero en otras ocasiones podemos encontrar una asociación muy alta (el ejemplo más extremo de esto es la correlación de una variable consigo misma, que como es lógico, es la más alta posible).

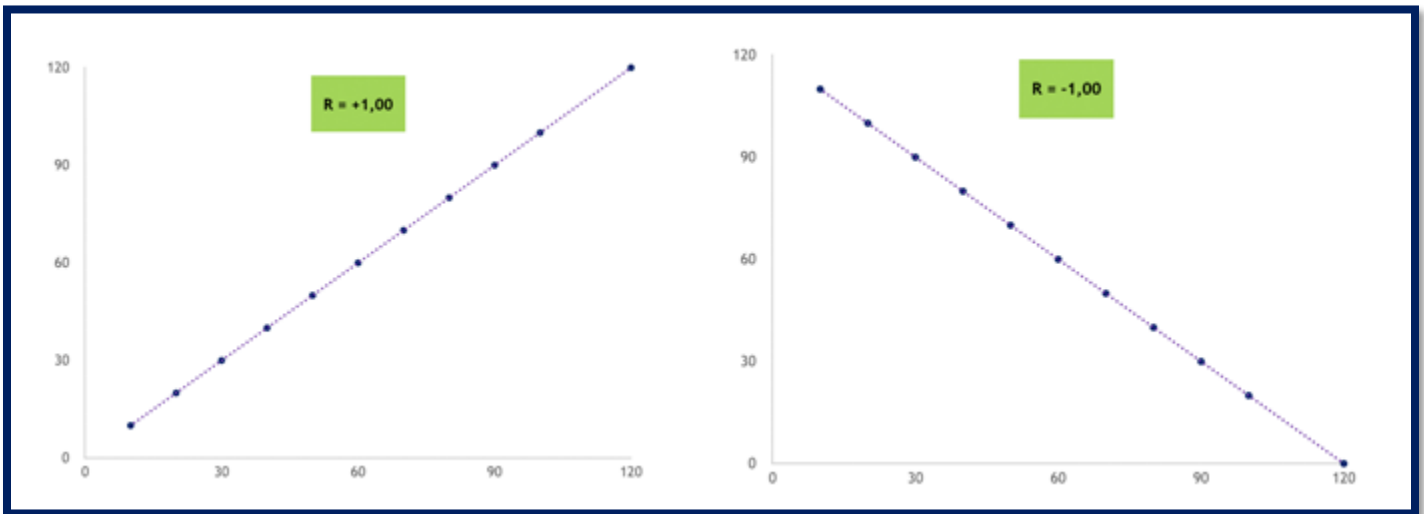
**Una forma de contestar a estas cuestiones es utilizando el Coeficiente de Correlación de Pearson**, un método propio de la estadística paramétrica, que no solo se usa para conocer la relación entre dos variables cuantitativas, sino también para la predicción de una variable (la regresión lineal, tema que no trataremos aquí).

### **Cómo se interpreta el coeficiente de correlación de Pearson**

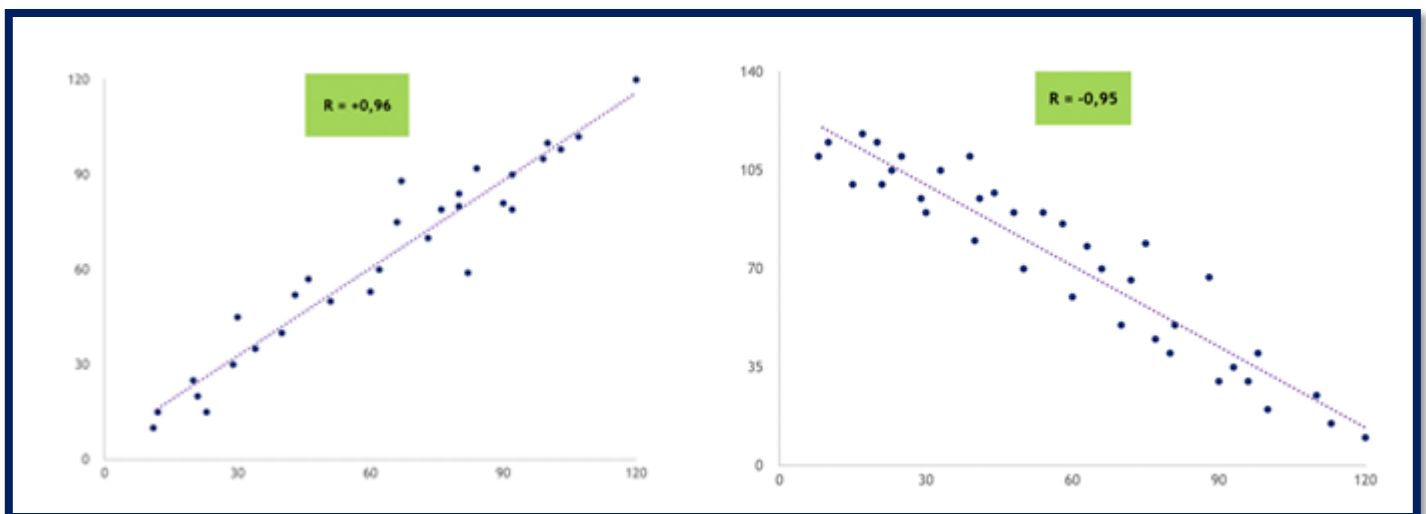
- Un valor mayor que 0 indica que existe una correlación positiva. En este caso las variables estarían asociadas en sentido directo. Cuanto más cerca de +1, más alta es su asociación. Un valor exacto de +1 indicaría una relación lineal positiva perfecta.
- Finalmente, una correlación de 0, o próxima a 0, indica que no hay relación lineal entre las dos variables.

La representación gráfica de los datos es muy útil para visualizar la relación existente entre las variables, ya que hay que tener en cuenta que a veces existen relaciones entre variables que no son lineales.

**Por ejemplo**, los dos siguientes gráficos muestran relaciones lineales entre dos variables. La primera tiene un coeficiente de correlación de +1, indicando una relación lineal perfecta y positiva (el crecimiento en una variable se asocia con un crecimiento en la otra variable). La segunda tiene un coeficiente de correlación de -1, indicando también una relación lineal total, pero negativa, de sentido inverso (el crecimiento en una se asocia con un descenso en la otra).



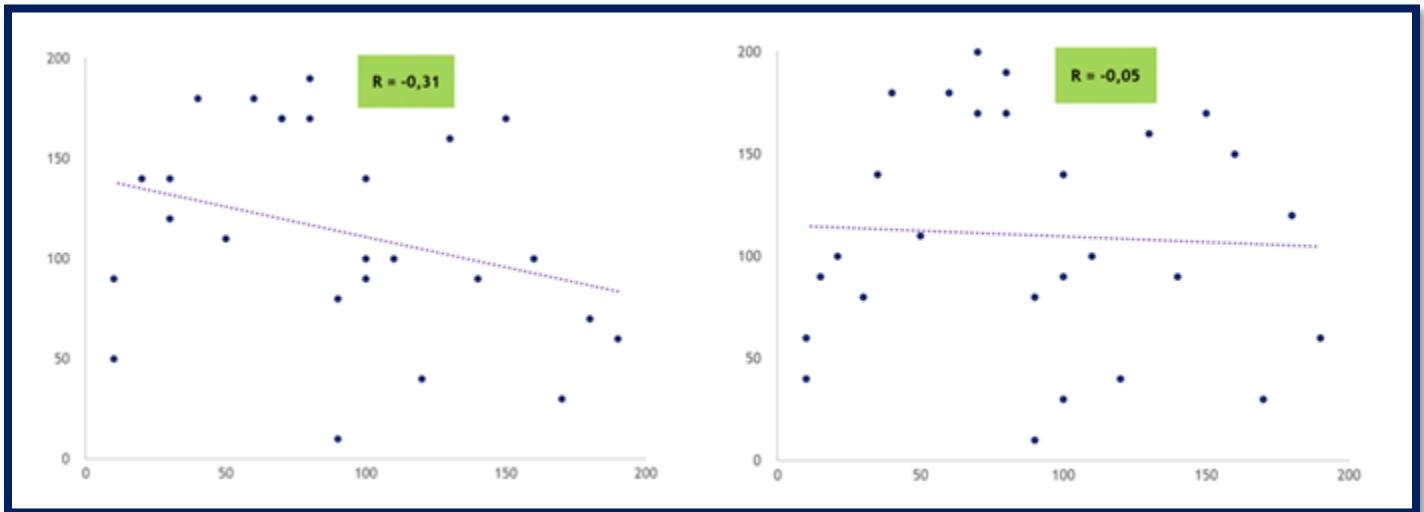
Estos niveles de correlación lineal tan perfectos no suelen darse en la realidad, pero si que encontramos muchas veces alta correlación, bien en un sentido o en otro, con representaciones gráficas similares a las que siguen, donde el valor de “r” es cercano a 1. Puede verse como las nubes de puntos (pares de datos de valores para las dos variables), siguen un sentido y se acumulan en su recorrido cerca de



la línea de tendencia, que es positiva en un caso y negativa en el otro:



Por el contrario, en el primero de los dos siguientes ejemplos podemos comprobar que la nube de puntos está bastante más dispersa, con un coeficiente de correlación de  $-0,31$ . Intuitivamente nos damos cuenta de que la correlación es baja, a pesar de que se apunta una ligera tendencia de correlación negativa. En el caso de la



derecha, la dispersión es total y se ve perfectamente que se trata de dos variables no correlacionadas. Su coeficiente, valor “r” es de  $-0,05$ .

Existe bastante consenso a la hora de interpretar los valores del coeficiente de correlación de Pearson utilizando los siguientes criterios (y considerando los valores absolutos):

- Entre 0 y 0,10: correlación inexistente
- Entre 0,10 y 0,29: correlación débil
- Entre 0,30 y 0,50: correlación moderada
- Entre 0,50 y 1,00: correlación fuerte

## Algunas características del coeficiente de correlación de Pearson

1. El valor del coeficiente de correlación es independiente de cualquier unidad usada para medir las variables, lo que sin duda es una ventaja.



2. Cuando hay valores extremos en alguna de las variables, el valor del coeficiente de correlación puede verse alterado de forma importante. En estas situaciones, si a pesar de ello queremos aplicarlo, conviene realizar una transformación de datos (por ejemplo, la transformación logarítmica), cambiando la escala de medición para moderar el efecto de valores extremos.

3. Cuanto más grande es la muestra, más preciso será el coeficiente de correlación. Con muestras muy pequeñas puede que esté indicado aplicar otro coeficiente de correlación, de tipo no paramétrico (el Coeficiente de Correlación de Spearman, por ejemplo).

4. Del mismo modo, se presupone normalidad en la distribución bivalente de las dos variables en análisis. Si esta condición no se cumple, será mejor aplicar otro (también el Coeficiente de Correlación de Spearman).

5. El coeficiente de correlación mide solo la relación con una línea recta, pero puede que nuestras dos variables tengan una relación diferente. Por ejemplo, al representarlas gráficamente podemos encontrar que tienen una relación curvilínea, y no en línea recta. En este caso el coeficiente de correlación de Pearson, que mide la relación lineal, nos daría un valor pequeño, a pesar de que en realidad las dos variables tienen asociación entre sí (solo que ésta es de tipo no lineal).

6. Debemos tener claro que decir que hay correlación no es lo mismo que decir que existe “causalidad”. Es decir, la “causa” de que una variable varíe en cierto sentido no es que la otra “se mueva” (o viceversa). Puede haber otros factores ajenos u otras variables, al margen de estas dos, que en realidad sean la causa de estos movimientos.

Como ocurre con muchos otros estadísticos, existen multitud de programas y opciones que facilitan el cálculo del coeficiente de correlación de Pearson.



Realmente lo más importante, en el fondo, es saber cuándo y cómo utilizarlo para darle sentido al análisis de los datos, convirtiéndolos así en información relevante para el cliente.

## La fórmula del coeficiente de correlación de Pearson

La fórmula del coeficiente de correlación descubre la relación entre las variables.

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

Dónde:

**N** = Número de valores o elementos

**$\sum xy$**  = la suma de los productos de las puntuaciones emparejadas

**$\sum x$**  = la suma de puntuaciones **x**

**$\sum y$**  = la suma de puntuaciones **y**

**$\sum x^2$**  = Suma de cuadrados Puntuación **x**

**$\sum y^2$**  = Suma de cuadrados Puntuación **y**

## Cómo calcular el coeficiente de correlación de Pearson

Aquí hay una guía paso a paso para calcular el coeficiente de correlación de Pearson:

**Paso uno:** Crear una tabla de coeficientes de correlación de Pearson. Hacer una tabla de datos, incluyendo ambas variables. Etiqueta estas variables como «x» e «y». Añade tres columnas adicionales – (xy), (x<sup>2</sup>), y (y<sup>2</sup>). Checa esta simple tabla de datos.



Persona	Edad(x)	Ingresos(y)	(xy)	(x <sup>2</sup> )	(y <sup>2</sup> )
1					
2					
3					
4					

**Paso dos:** Usa una multiplicación básica para completar la tabla.

Persona	Edad(x)	Ingresos(y)	(xy)	(x <sup>2</sup> )	(y <sup>2</sup> )
1	20	1,500	30,000	400	2,250,000
2	30	3,000	90,000	900	9,000,000
3	40	5,000	200,000	1600	25,000,000
4	50	7,500	375,000	2500	56,250,000

**Paso tres:** Sume todas las columnas de abajo a arriba.

Persona	Edad(x)	Ingresos(y)	(xy)	(x <sup>2</sup> )	(y <sup>2</sup> )
1	20	1,500	30,000	400	2,250,000
2	30	3,000	90,000	900	9,000,000
3	40	5,000	200,000	1600	25,000,000
4	50	7,500	375,000	2500	56,250,000
<b>Total</b>	140	17,000	695,000	5400	92,500,000

**Paso cuatro:** Usar la fórmula de correlación para conectar los valores.

Si el resultado es negativo, hay una relación de correlación negativa entre las dos variables. Si el resultado es positivo, hay una relación de correlación positiva entre las variables.

Los resultados también pueden definir la fuerza de una relación lineal, es decir, una relación positiva fuerte, una relación negativa fuerte, una relación positiva media, etc.





El coeficiente de Pearson (también llamado coeficiente de correlación del producto-momento), determina la fuerza de la relación lineal entre dos variables. Cuanto más fuerte sea la asociación entre las dos variables, más se inclinará la respuesta hacia 1 o -1.

Alcanzar valores de 1 o -1 significa que el cambio de factores de cualquier variable no debilita la correlación con la otra variable. Cuanto más se acerque la respuesta a 0, mayor será la variación de las variables. Si el resultado es negativo, hay una relación de correlación negativa entre las dos variables. Si el resultado es positivo, hay una relación de correlación positiva entre las variables.

Los resultados también pueden definir la fuerza de una relación lineal, es decir, una relación positiva fuerte, una relación negativa fuerte, una relación positiva media, etc.

El coeficiente de Pearson (también llamado coeficiente de correlación del producto-momento), determina la fuerza de la relación lineal entre dos variables. Cuanto más fuerte sea la asociación entre las dos variables, más se inclinará la respuesta hacia 1 o -1.

Alcanzar valores de 1 o -1 significa que el cambio de factores de cualquier variable no debilita la correlación con la otra variable. Cuanto más se acerque la respuesta a 0, mayor será la variación de las variables.

A continuación, se proponen las directrices para la interpretación de la correlación del coeficiente de Pearson:

Fuerza de la asociación	Coeficiente, r	
	Positiva	Negativa
Chica	.1 a .3	-0.1 a -0.3
Mediana	.3 a .5	-0.3 a -0.5
Grande	.5 a 1.0	-0.5 a 1.0



**Actividad de Aprendizaje No 35**

Aplicando la fórmula establece que tipo de asociación tiene el problema anterior, CHICA, MEDIANA O GRANDE, si es positiva o negativa.



## PENSAMIENTO MATEMÁTICO

### PROGRESIÓN No. 10

Cuestiona afirmaciones estadísticas y gráficas, considerando valores atípicos (en el caso de variables cuantitativas) y la posibilidad de que existan factores o variables de confusión.

METAS	CATEGORÍAS	SUBCATEGORÍAS
M1 Observa y obtiene información de una situación o fenómeno para establecer estrategias o formas de visualización que ayuden a entenderlo	C2 Procesos de intuición y razonamiento	S1 Casidad para observar y conjeturar
M2 Socializa con sus pares sus conjeturas descubrimientos o procesos en la solución de un problema tanto teórico como de su entorno	C4 Interacción y lenguaje matemático	S1 Registro escrito, simbólico, algebraico e icográfico S3 Ambiente matemático de comunicación



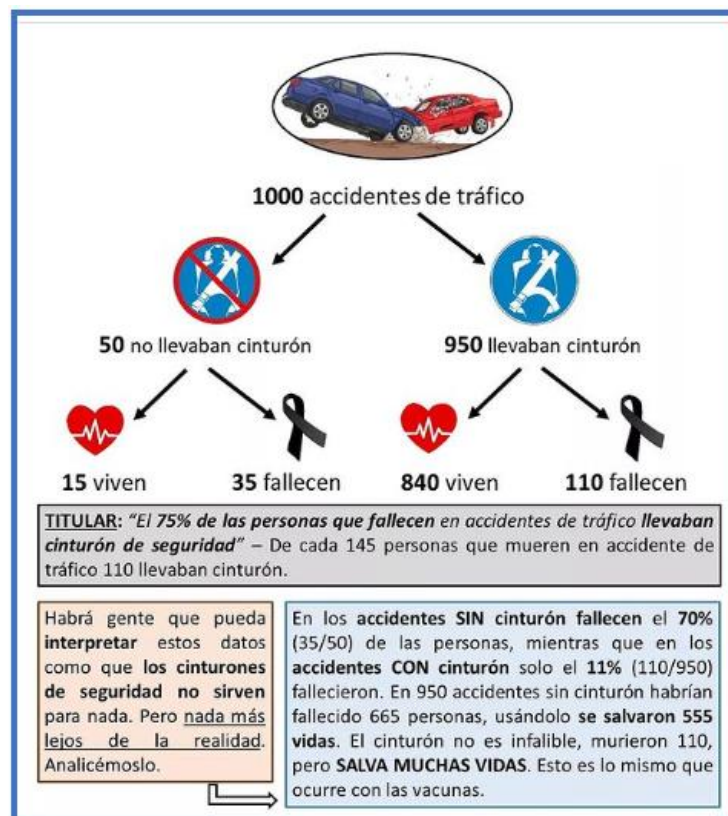
## Progresión No. 10

**Afirmaciones estadísticas y gráficas, considerando valores atípicos (en el caso de variables cuantitativas) y la posibilidad de que existan factores o variables de confusión.**



### ¿Qué es la paradoja de Simpson?

La paradoja de Simpson ocurre cuando al analizar grupos de datos de forma conjunta se producen resultados diferentes o contrarios que al analizarlos por subgrupos. Esto ocurre porque existe una (o más de una) **variable confusora**, es decir, un factor oculto que tiene mucho peso dentro de los datos y desequilibra la balanza cuando estos se analizan de forma agregada, llevando a resultados engañosos. Por ello, es necesario dividir los datos en subgrupos para poder analizarlos según las variables de confusión y **sin que los resultados nos induzcan a error.**





Disponible en: [https://www.instagram.com/p/CSOroPojYoi/?img\\_index=1](https://www.instagram.com/p/CSOroPojYoi/?img_index=1)

Consultado el 23 de Mayo del 2023

\*\* Si analizamos los datos totales, vemos que un 75% de las personas que fallecieron en accidentes de tráfico llevaba el cinturón de seguridad **(110 de 145)**, llegando a la conclusión (errónea) **de que el cinturón de seguridad no es eficaz.**

Sin embargo, si analizamos los datos por subgrupos, vemos que en los accidentes sin cinturón mueren el **70%** de las personas **(35 de 50)** y en aquellos donde las personas llevaban el cinturón mueren solo el **11% (110 de 950)**. Es la conclusión contraria a la anterior.

¿Por qué ocurre esto? Porque en el total de accidentes analizados (1000), el grupo de accidentes donde las personas llevaban cinturón es **mucho más grande (950)** que el **de quienes no lo llevaban (50)**. Por lo tanto, a la hora de analizar los datos de fallecidos de forma conjunta ese subgrupo pesa mucho. De esta forma, en el grupo de fallecidos había muchas más personas con cinturón que sin cinturón, pero no quiere decir que no sea eficaz o que las personas con cinturón en accidentes fallezcan en mayor proporción que las que no lo llevaban\*.

## Ejemplo 2

### Fresa vs Durazno

“Supongamos que estamos en la industria de los refrescos y estamos tratando de elegir entre dos nuevos sabores que hemos producido. Podríamos probar la opinión pública sobre los dos sabores; digamos que elegimos hacerlo instalando dos puestos de muestreo para cada sabor en un área concurrida y preguntando a 1000 personas en cada puesto si les gusta el nuevo sabor.



	Tamaño de Muestra	Sabor Gustado
Fresa	1000	800
Durazno	1000	750

Podemos ver que el **80%** de las personas disfrutó de **fresa** mientras que solo el **75%** de las personas disfrutó de **durazno** . Por lo tanto, es más probable que **fresa** sea el sabor preferido.

\*\*\*\*\*Ahora, supongamos que nuestro equipo de marketing recopiló otra información mientras realizaba la encuesta, como el sexo de la persona que prueba la bebida. ¿Qué sucede si dividimos nuestros datos por sexo?

Sabor	Hombres	Sabor Gustado hombres	Mujeres	Sabor Gustado Mujeres
Fresa	900	760	100	40
Durazno	700	600	300	150



Esto sugiere que al **84,4%** de los hombres y al **40%** de las mujeres les **gustó fresa** , mientras que al **85,7%** de los hombres y al **50%** de las mujeres les **gustó durazno** . Si nos detenemos a pensar, esto puede parecer un poco extraño: según los datos de nuestra muestra, generalmente la gente prefiere fresa , pero tanto hombres como mujeres por separado prefieren durazno . ¡Este es un ejemplo de la paradoja de Simpson!

Nuestra intuición nos dice que el sabor que se prefiere tanto cuando una persona es hombre como mujer también se debe preferir cuando se desconoce su sexo, y es bastante extraño descubrir que esto no es cierto: este es el corazón de la paradoja.”

Disponible en: <https://towardsdatascience.com/simpsons-paradox-and-interpreting-data-6a0443516765> Consultado el 24 de mayo del 2023



Resuelve y concluye con este ejemplo imaginario de la paradoja de Simpson

En este ejemplo imaginario pensamos en la realización de consultas sobre la aceptación de un conjunto de candidatos/as, entre quienes se encuentran Armando y Bárbara. En cada encuesta las consultas se ordenaron de manera aleatoria y se hicieron tantas encuestas como fueran necesarias, hasta obtener 2200 respuestas para cada candidato/a (no todas las personas respondieron por todos los/las candidatos/as).

Los resultados de las consultas arrojaron que de las 2200 personas que contestaron por el candidato A, un 55% dijeron que les gusta; mientras que de las 2200 que contestaron por B, este porcentaje aumenta a 65%. Bárbara es la más aceptada si consideramos las proporciones totales.



Sobre las edades de las personas encuestadas, se observa que entre los adultos consultados por A, el 51% manifiesta aceptación, mientras que entre los adultos consultados por B ese porcentaje es de solo el 40%. Por otra parte, entre los jóvenes consultados por A, el porcentaje de aceptación es del 91%, y entre los jóvenes consultados por B, esa proporción se reduce a 80%. Tanto en el segmento de jóvenes como en el de adultos, Armando es favorito.

Esto se refleja en la siguiente tabla, que respalda igual de bien a las afirmaciones del tío y la sobrina. Pero entonces, ¿quién tiene razón?

<b>Pregunta</b>	<b>¿ Gusta A?</b>	<b>¿ Gusta B ?</b>
<b>Adultos</b>	<b>51% (1010/1980)</b>	<b>40% (329/824)</b>
<b>Jovenes</b>	<b>91% (200/220)</b>	<b>80% (1101/1376)</b>
<b>Totales</b>	<b>55% (1210/2200)</b>	<b>65% (1430/2200)</b>

Esto se llama la **paradoja de Simpson** y puede darse en muchas situaciones, induciéndonos a pensar que estamos “siguiendo los datos”. En realidad, los datos nos pueden llevar a conclusiones distintas dependiendo de la forma en que los agrupemos. Y en algunos casos esas conclusiones pueden ser erróneas cuando intentamos responder a preguntas específicas.