



Variables

Tasas

Frecuencias

Variable	f	Fr (f / 50)	fra
1	12	0.24	0.24
2	11	0.22	0.46
3	9	0.18	0.64
4	5	0.10	0.74
5	5	0.10	0.84
6	8	0.16	1.00
TOTAL	50	1.00	

VARIABLES CUALITATIVAS

Se caracterizan por su naturaleza conceptual. Los datos se clasifican en categorías o clases que deben ser exhaustivas y mutuamente excluyentes.

- **Nominales:** los datos se clasifican en categorías que no tienen un orden implícito. Ejemplos: sexo, religión, profesión.
- **Ordinales:** los datos se clasifican en categorías que tienen un orden implícito. Ejemplos: estado de una vivienda, nivel de educación.

VARIABLES CUANTITATIVAS

Se caracterizan por su naturaleza numérica. Denominamos recorrido de la variable al conjunto de todos los valores que ella puede asumir.

- **Discretas:** Su recorrido es finito o infinito numerable. Ejemplos: cantidad de empleados de una empresa, población mundial a una determinada fecha, cantidad de autos que circulan en el mundo.
- **Continuas:** Aunque sólo sea en teoría, su recorrido es infinito no numerable. Ejemplos: tasa de interés efectiva mensual, ingresos de una empresa, estatura de una persona.

Nota: La distinción entre continuas y discretas puede ser considerada teórica en tanto los instrumentos de medición tienen precisión finita.

PRESENTACIÓN DE DATOS EN TABLAS DE FRECUENCIAS

Para un conjunto de datos (de tamaño n), una tabla de frecuencias muestra las distintas categorías o valores observados de una variable (dependiendo de si la variable es cualitativa o cuantitativa) junto con sus frecuencias observadas.

- **FRECUENCIA ABSOLUTA (f):** Cantidad de veces que aparece una categoría o valor en el conjunto de datos.
- **FRECUENCIA RELATIVA ($h = \frac{f}{n}$):** Proporción de veces que aparece una categoría o valor en el conjunto de datos.

1, 3, 3, 3, 2, 3, 2, 4, 1, 5, 3, 6, 6, 1, 1,
4, 3, 4, 4, 1, 1, 3, 3, 2, 3, 2, 2, 3, 2, 5,
3, 3, 3, 5, 3, 2, 3, 3, 3, 4

Valor	Frecuencia Absoluta (f)	Frecuencia relativa ($h = \frac{f}{n}$)
1	6	$6/40 = 0,15$
2	7	$7/40 = 0,175$
3	17	$17/40 = 0,425$
4	5	$5/40 = 0,125$
5	3	$3/40 = 0,075$
6	2	$2/40 = 0,05$
Total	40	1

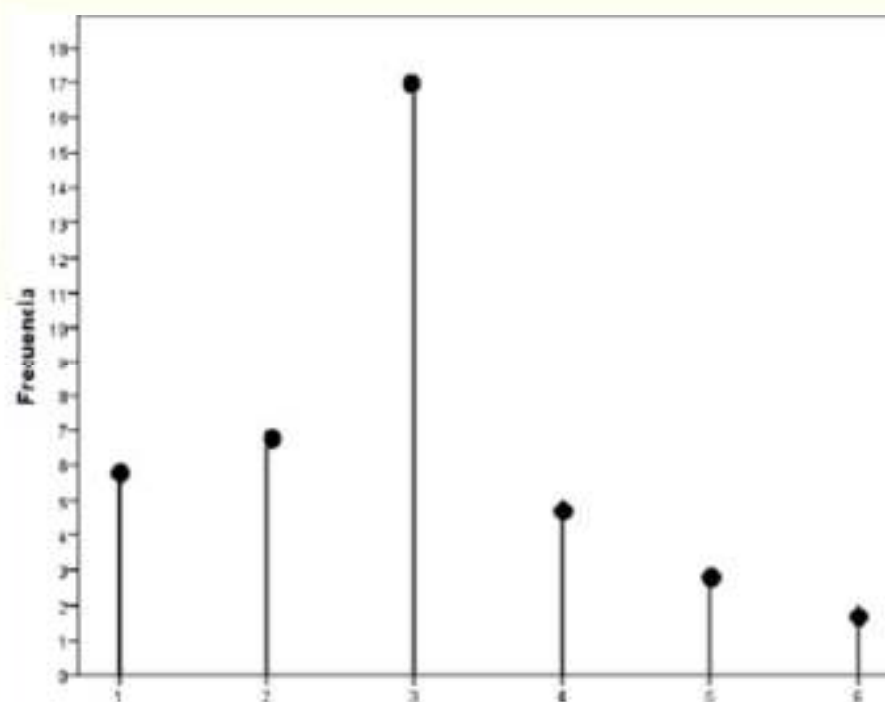
GRAFICO DE LINEAS O BASTONES



CONCEPTO

Es una representación gráfica de una tabla de frecuencias.

En el eje de abscisas se representan las categorías o valores observados de la variable, y la frecuencia (absoluta o relativa) de tal categoría o valor se identifica con la altura de una línea vertical.



DATOS AGRUPADOS EN INTERVALOS

- Las tablas de frecuencias y los gráficos de barras o líneas constituyen una forma efectiva de representar las frecuencias de los valores de una variable cuantitativa cuando la cantidad de valores distintos observados es reducida.
- Sin embargo, cuando el número de valores distintos es demasiado grande éstas representaciones gráficas pierden utilidad.
- En su lugar, es posible agrupar los valores observados en intervalos para luego representar la frecuencia que corresponde a cada intervalo.

Frecuencia

- ***Frecuencia absoluta:*** (n_i) El número de veces que se repite cada valor o dato de la variable.
- ***Frecuencia relativa:*** (f_i) La frecuencia absoluta dividida por el número de datos.

$$f_i = \frac{n_i}{N}$$

donde N es el número de datos.

- Al agrupar datos en intervalos utilizaremos la convención de inclusión sólo por la izquierda excepto para el último intervalo en el que se incluye el extremo derecho.
- Notación: [extremo inferior, extremo superior).
- Por ejemplo el intervalo 20-30 incluye los valores que son mayores o iguales que 20 y menores que 30.

EJEMPLO

Sean los siguientes datos de ingresos (en miles de pesos) de un conjunto de personas: 1, 1, 1, 2, 2, 3, 4, 4, 5, 6, 8, 8, 9, 9, 10, 11, 12, 15, 17, 19, 20, 21, 23, 24, 25, 27, 30, 35, 41, 45

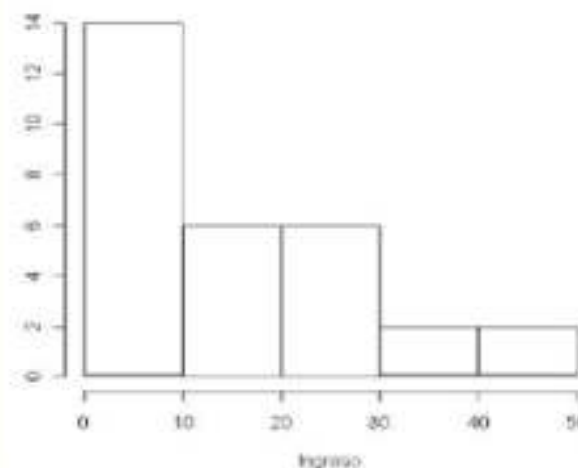
Intervalo	Frecuencia Absoluta (f)	Frecuencia relativa ($h = \frac{f}{n}$)
0-10	14	14/30
10-20	6	6/30
20-30	6	6/30
30-40	2	2/30
40-50	2	2/30
Total	30	1

HISTOGRAMA: REPRESENTACIÓN DE DATOS AGRUPADOS EN INTERVALOS.

CONSTRUCCIÓN

- 1 Ordene los datos en forma creciente y agrúpelos en intervalos de igual amplitud de manera que cada dato sea clasificado en uno y sólo uno de los intervalos.
- 2 Construya una tabla de frecuencias de los intervalos.
- 3 En el eje de abscisas represente los intervalos.
- 4 En el eje de ordenadas represente las frecuencias (absolutas o relativas) de los intervalos.
- 5 Sobre los intervalos dibuje barras adyacentes con alturas iguales a sus frecuencias.

EJEMPLO



ADVERTENCIA

Es usual que las barras del histograma tengan área igual a la frecuencia relativa del intervalo. Sin embargo, siendo los intervalos de igual amplitud, lo anterior genera la misma interpretación que surge de representar barras de altura igual a la frecuencia del intervalo.

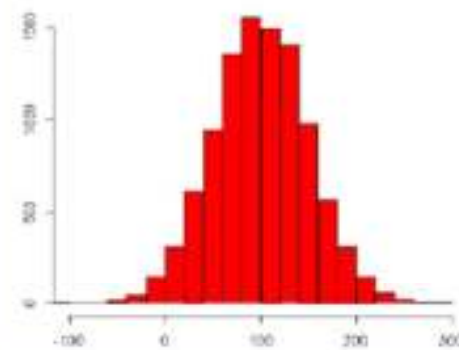
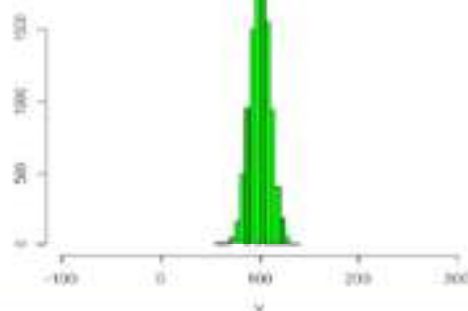
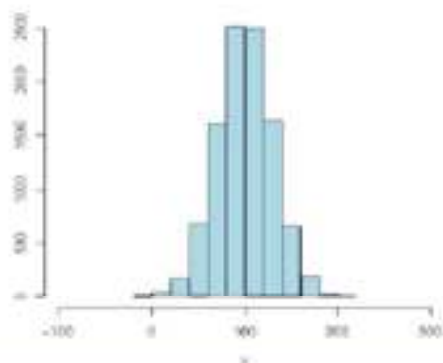
CÁLCULO DE \bar{x} A PARTIR DE UNA TABLA DE FRECUENCIAS

- Supongamos que los datos se dan mediante una tabla de frecuencias, donde se incluyen los k valores distintos x_1, x_2, \dots, x_k junto con sus respectivas frecuencias absolutas f_1, f_2, \dots, f_k .
- El conjunto de datos consiste de n observaciones. Por tanto, $n = \sum_{j=1}^k f_j$ y cada valor x_j aparece f_j veces, para $j = 1, \dots, k$.
- Por consiguiente, la media muestral de este conjunto de datos será:

$$\bar{x} = \frac{\overbrace{x_1 + \dots + x_1}^{f_1} + \overbrace{x_2 + \dots + x_2}^{f_2} + \dots + \overbrace{x_k + \dots + x_k}^{f_k}}{n} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{n} = \frac{\sum_{j=1}^k x_j f_j}{n}$$

- La expresión anterior puede reescribirse como $\bar{x} = \sum_{j=1}^k x_j \frac{f_j}{n}$. Así, podemos afirmar que la media muestral es una media ponderada del conjunto de valores distintos observados en los datos cuyos ponderadores son sus frecuencias relativas.

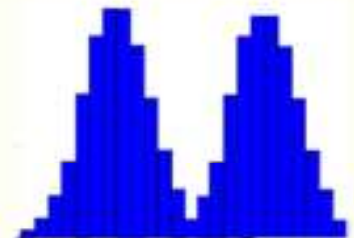
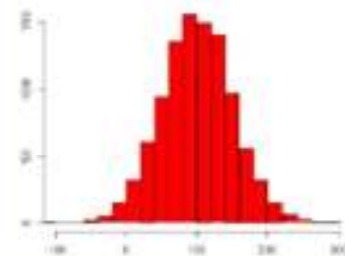
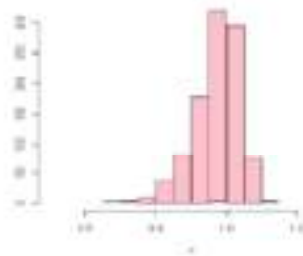
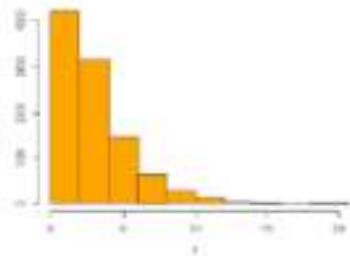
¿Cómo distinguir entre los casos siguientes, que presentan la misma media, mediana y moda?



A las medidas de posición las complementamos con medidas de dispersión.

CONJUNTOS DE DATOS NORMALES Y NO NORMALES

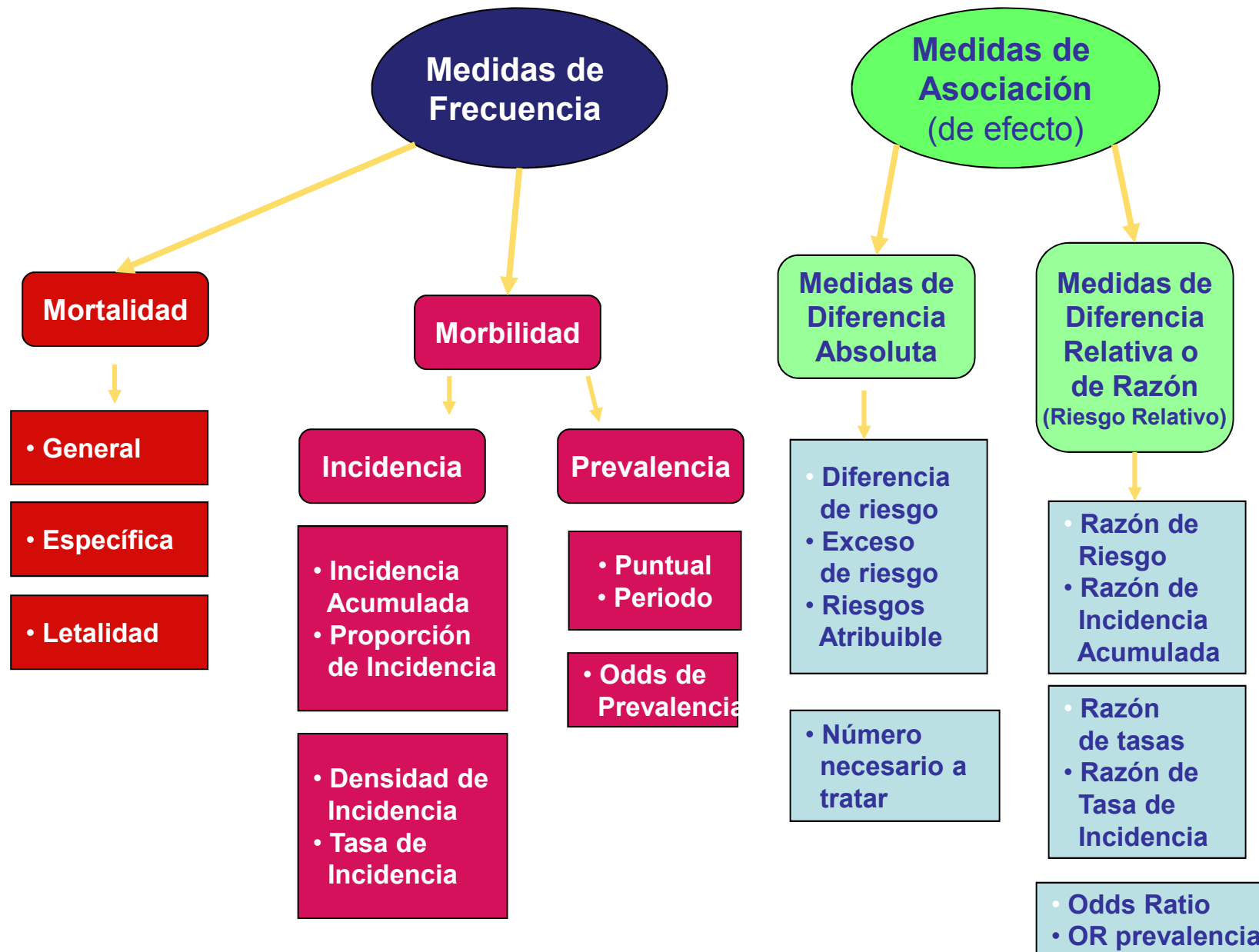
¿Cuáles de los siguientes conjuntos de datos es aproximadamente normal?



SIMETRÍA O ASIMETRÍA

- El primer conjunto de datos tiene distribución asimétrica con rama a la derecha.
- El segundo conjunto de datos tiene distribución asimétrica con rama a la izquierda.
- Los dos últimos conjuntos de datos son aproximadamente simétricos.
- Sólo el tercer conjunto de datos puede tener distribución aproximadamente normal.
- El cuarto conjunto de datos puede ser una mezcla de datos normales.

EPIDEMIOLOGÍA DESCRIPTIVA



Cálculo de proporciones



Proporciones

Las proporciones son medidas que expresan la frecuencia con la que ocurre un evento en relación con la población total en la cual éste puede ocurrir. (el numerador esta incluido en el denominador)



Esta medida se calcula dividiendo el número de eventos ocurridos entre la población en la que ocurrieron

NUMERADOR/DENOMINADOR

Ejemplo: 3 muertes/100 personas= 0.03 = 3%

Cálculo de tasas



Tasa

Es similar a una proporción, con la diferencia que llevan incorporado el concepto de **tiempo y lugar**, o sea numerador y denominador deben estar referidos, al mismo tiempo y lugar.

$$\frac{\text{numerador}}{\text{denominador}} \times K$$

El numerador expresa el número de eventos acaecidos durante un periodo en un número determinado de sujetos observados.

El denominador de una tasa expresa el tiempo durante el cual tales sujetos estuvieron en riesgo de sufrir el evento.

La constante, por la cual se multiplica el cociente para una lectura entendible, fácil y comúnmente aplicada..



Tasa

$$\text{Tasa} = \frac{\text{número de eventos ocurridos en una población en un periodo } t}{\text{sumatoria de los periodos durante los cuales los sujetos de la población libres del evento estuvieron expuestos al riesgo de presentarlo en el mismo periodo}} \times \text{una potencia de } 10$$

Cuando en el denominador se trata de población general, para fines del cálculo de la población expuesta, se usa como convención la existente al 30 de junio en ese lugar durante ese año (mitad de año).

Cálculo de razones



Razones

Son magnitudes que expresan la relación aritmética existente entre dos eventos en una misma población, o un solo evento en dos poblaciones.

$$\text{Razón hombre: mujer} = \frac{4\,000}{5\,000} = 0.8$$

$$\text{RTM} = \frac{\text{tasa de mortalidad en la ciudad B}}{\text{tasa de mortalidad en la ciudad A}} = \frac{50 \times 1\,000}{25 \times 1\,000} = 2.0$$

Prevalencia e Incidencia

¿Qué se requiere conocer?:

- N° pacientes que experimentan la enfermedad (CASOS)
- N° de sujetos o población en quienes hubiese podido ocurrir la enfermedad (Población en riesgo)

Diferencias entre Prevalencia Puntual y de Período



Prevalencia Puntual

Proporción de personas que experimentan el evento clínico (Caso), en un punto dado en el tiempo.
Es un corte en un momento dado, en el que cuento los **CASOS**

Prevalencia de Período

Proporción de casos. Se considera los que enfermaron en períodos anteriores y que aun siguen siendo **CASOS** (Existentes y Nuevos).

Cada persona representada en el Numerador tuvo la enfermedad en algún momento durante el Período especificado


Incidencia: N° de casos **NUEVOS en relación a la Población en Riesgo**

Si dividimos por el intervalo de tiempo,
podemos comparar a través de los intervalos

N° de casos nuevos

Incidencia Acumulada = $\frac{\text{N° de casos nuevos}}{\text{Población en Riesgo}}$

Incidencia calculada usando un período de tiempo durante el cual todos los individuos en la población son considerados en

RIESGO para el evento de interés  **INCIDENCIA ACUMULADA**

Características de la Incidencia Acumulada

- Se estima en una cohorte fija. El t de observación de la población será el mismo para todas las personas del estudio y el n° de personas susceptibles será el mismo que al inicio.
- Las personas observadas deben ser susceptibles y estar libres de la enfermedad al comienzo de la observación.
- Se debe fijar el período de observación de ocurrencia del **EVENTO** a medir por un período específico de **TIEMPO**.
- Mide **RIESGO**.

Característica	Incidencia	Prevalencia
Numerador	Nuevos casos que ocurren en un período de tiempo en un grupo inicialmente libre de la enfermedad	Todos los casos existentes
Denominador	Todos los susceptibles al comienzo del período de observación	Todas las personas examinadas: Casos y No casos
Tiempo	Duración del período de observación	Puntual
Diseño para medir	Estudio de Cohorte	Estudio de Prevalencia (Cross-Sectional)
Tipo de Medición	Riesgo	Proporción

Densidad de Incidencia



Número de casos **NUEVOS** en relación al **TAMAÑO** y el **TIEMPO** de observación de la población en riesgo

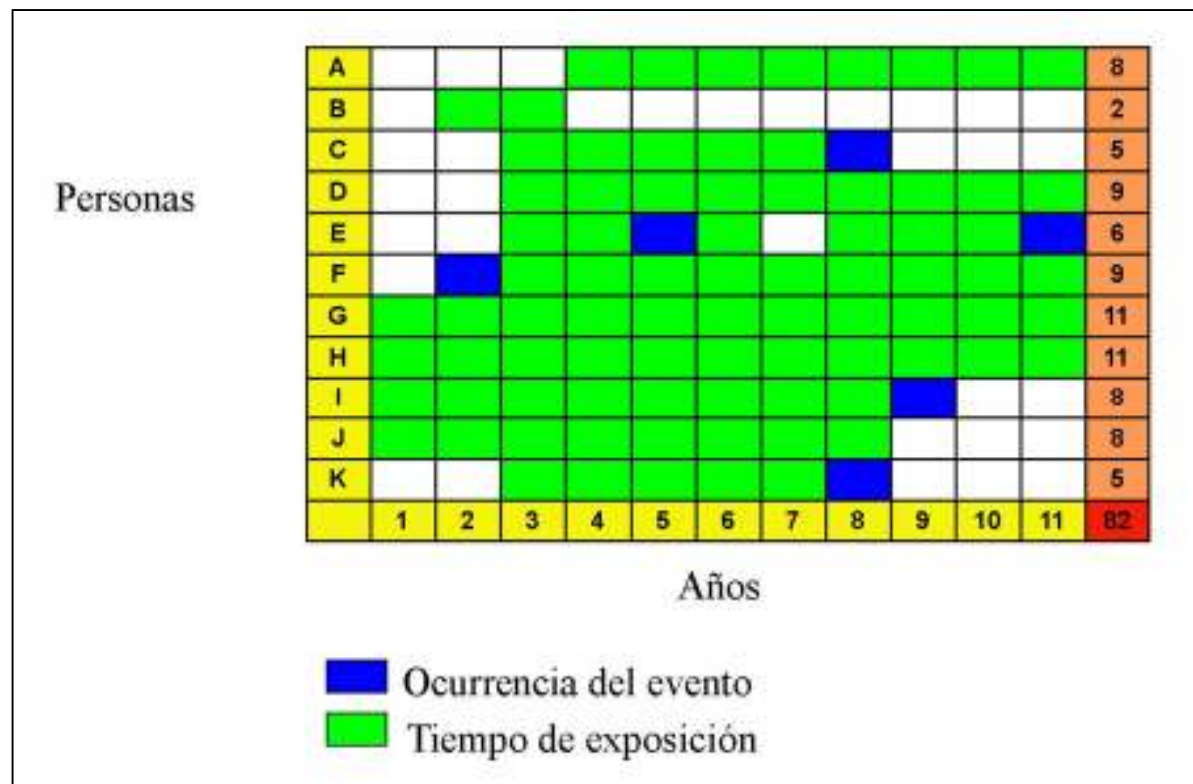
$$\frac{\text{Número de casos } \mathbf{NUEVOS}}{\text{Población promedio en riesgo} \times \text{Intervalo de Tiempo}}$$

$$\text{Número de casos } \mathbf{NUEVOS} = \frac{\quad}{\text{Personas-tiempo}}$$

Cada individuo en el denominador no es seguido por el mismo período de tiempo. Corresponde a **la suma de los diferentes tiempos** que cada individuo estuvo en riesgo.

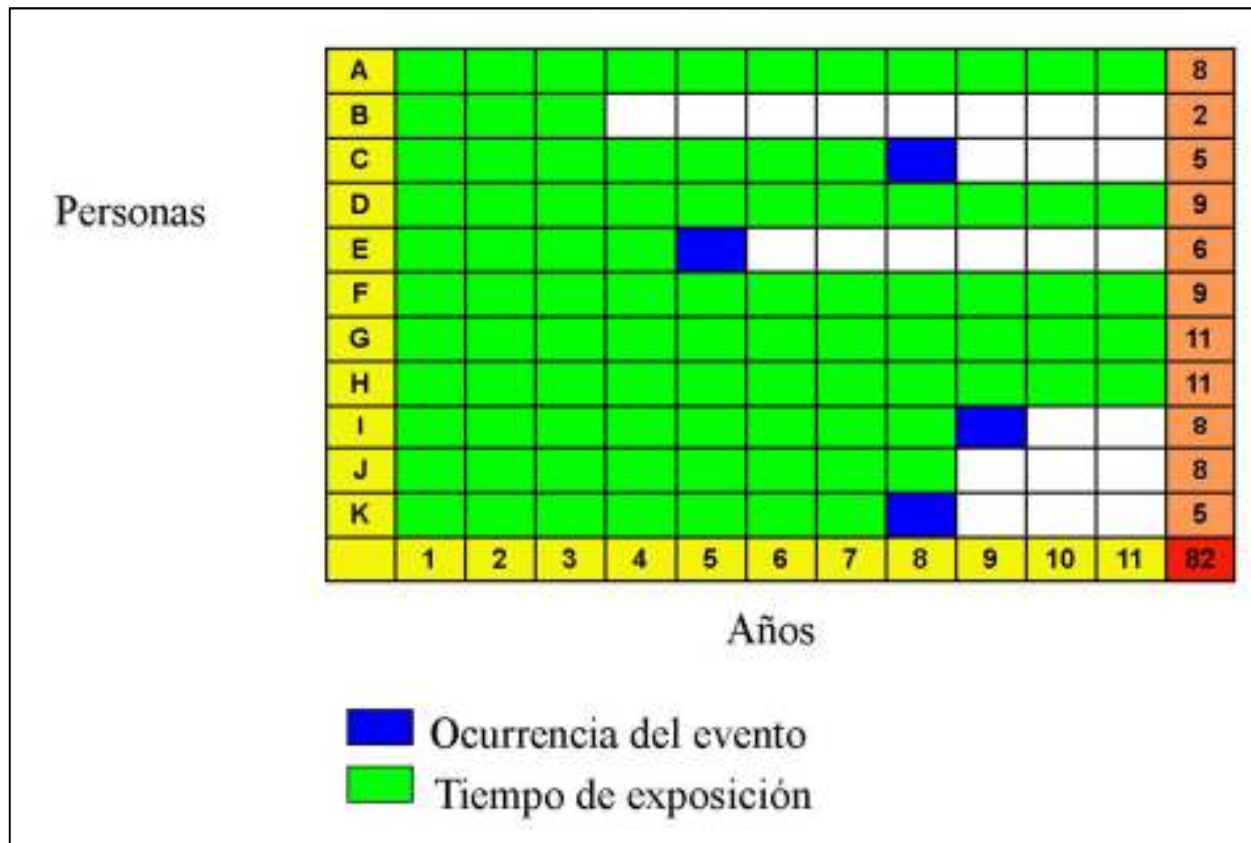


$$\text{Tasa de incidencia} = \frac{\text{número de casos nuevos}}{\text{suma de todos los periodos libres de la enfermedad durante el periodo definido en el estudio (tiempo-persona)}}$$



Tasa de Incidencia o Densidad de Incidencia= 6/82 años-persona= 0,073 = 7,3 eventos 100 años-persona

$$IA = \frac{\text{número de personas que contraen la enfermedad en un periodo determinado}}{\text{numero de personas libres de la enfermedad en la población expuesta al riesgo en el inicio del estudio}}$$



Incidencia Acumulada= $\frac{4}{11}$ persona= 0,36 = 36 casos por 100 personas

■ Tasa de ataque:

- Incidencia en el curso de un brote epidémico, limitada a la duración del brote.
- $\frac{\text{Nº de casos observados} \times 100 \text{ ó } \times 1000}{\text{Población realmente expuesta al riesgo del brote}}$

■ Tasa de letalidad:

- $\frac{\text{Defunciones por una causa} \times 100}{\text{número de enfermos por la misma causa.}}$
- Infecciones agudas = Tasa de incidencia (casos nuevos)
- Infecciones crónicas = Tasa de prevalencia (casos existentes)

$$p = \frac{\text{número total de casos existentes al momento } t}{\text{total de la población en el momento } t} \times 10n$$

Prevalencia = Incidencia x Duración de la Enfermedad

■ Tasa de Mortalidad:

Nº de Muertes por todas las causas en 1 año *1000

Población estimada al 31 de Junio de ese año

$$\text{Tasa mortalidad general} = \frac{\text{número de muertes en el periodo } t}{\text{población total promedio en el mismo periodo}} \times 10n$$

$$\text{TME} = \frac{\text{total de muertes en un grupo de edad y sexo específicos de la población durante un periodo dado}}{\text{población total estimada del mismo grupo de edad y sexo en el mismo periodo}} \times 10n$$

■ Tasa de Mortalidad todas las causas <10 años:

Nº de Muertes por todas las causas en 1 año en <10 años *1000

Nº niños < 10 años al 31 de Junio de ese año



Medidas de razón

Estas medidas cuantifican discrepancias para un solo evento en dos poblaciones.

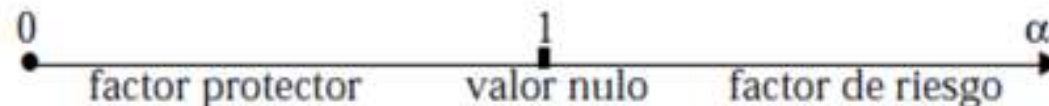
Son las razones que con mayor frecuencia se calculan y se obtienen con la siguiente fórmula:

$$\text{Razón} = \frac{\text{medida de frecuencia en un grupo expuesto (E}_1\text{)}}{\text{medida de frecuencia de un grupo no expuesto (E}_0\text{)}}$$

Valor =1 indica ausencia de asociación, no-asociación o valor nulo.

Valores <1 indica asociación negativa, factor protector.

Valores >1 indica asociación positiva, factor de riesgo.



Razón de incidencia acumulada o riesgo relativo RR



Compara el riesgo de enfermar del grupo de expuestos (IA_i) con el riesgo de enfermar del grupo de no expuestos (IA_o). Es útil si lo que se desea es conocer la probabilidad de padecer la enfermedad en función de la exposición, y es la medida que mejor refleja su asociación.

$$RR = \frac{IA_i}{IA_o} = \frac{a/n_i}{c/n_o}$$

donde,

IA_i es la incidencia acumulada o riesgo de enfermar entre los expuestos, y

IA_o es la incidencia acumulada o riesgo de enfermar entre los no expuestos (para observar gráficamente la ubicación de las celdas a, c, n_i y n_o , véase la tabla de 2 X 2).

	Mal control	Buen control	
Infarto	10	7	17
No Infarto	250	461	711
	260	468	728

- De acuerdo con los datos de este ejemplo, el riesgo de infarto en el grupo de mal control es $10 / 260 = 0.04$, mientras que el riesgo en el grupo de buen control es de $7 / 468 = 0.015$.
- Existen diferentes procedimientos para cuantificar la importancia de la asociación. Uno de ellos utiliza el cociente de esos dos riesgos $0.04 / 0.015 = 2.6$, y nos indica cuánto más probable es que ocurra el suceso en el primer grupo frente al segundo:
Riesgo Relativo (RR).

Razón de productos cruzados (RPC u OR)



Se utiliza en estudios de casos y controles donde los sujetos son elegidos según la presencia o ausencia de enfermedad, desconociéndose el volumen de la población de donde provienen por lo que no es posible calcular la incidencia de la enfermedad.



Es un buen estimador de la RDI, sobre todo cuando los controles son representativos de la población de la que han sido seleccionados los casos. El OR también puede ser un buen estimador del RR. Esta medida se calcula obteniendo el cociente de los productos cruzados de una tabla 2x2.

$$RPC = \frac{a/c}{b/d} = \frac{ad}{bc}$$

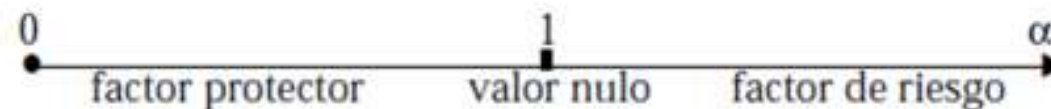
donde,

Exposición	casos		controles	
	presente	a	b	Total de expuestos (n_p)
	ausente	c	d	Total de no expuestos (n_o)
	Total (m_j)		Total (m_o)	Total de sujetos (n)

Valor =1 indica ausencia de asociación, no-asociación o valor nulo.

Valores <1 indica asociación negativa, factor protector.

Valores >1 indica asociación positiva, factor de riesgo.



Calcular el riesgo relativo e interpretar

	Grupo A	Grupo B	Grupo C
Tasa incidencia expuestos	0,3	0,4	0,5
Tasa incidencia no expuestos	0,4	0,4	0,4

- Cuando el Riesgo Relativo es menor de uno significa que aquellos sujetos expuestos al factor en estudio (variable independiente) tienen un menor riesgo de presentar el resultado (variable dependiente), mientras que un valor mayor de uno significa que la exposición confiere un riesgo mayor.
 - Un valor de 1 significa que el riesgo es el mismo en ambos grupos.
-

	Mal control	Buen control	
Infarto	10	7	17
No Infarto	250	461	711
	260	468	728

- El "**odds ratio**" (**OR**), es otra forma de representar un riesgo, mediante el cociente entre el número de veces que ocurre el suceso frente a cuántas veces no ocurre. Así un odds de 3 indica que 3 personas sufrirán el suceso frente a 1 que no lo hará.
- En la tabla anterior el odds de infarto en el grupo de Mal Control es $10 / 250 = 1 / 25 = 0.04$, es decir que ocurre 1 vez por 25 veces que no ocurre.
- En el grupo de buen control el odds es $7 / 461 = 0.015$.
- El cociente de los odds de los dos grupos es lo que se denomina odds ratio y constituye otra forma de cuantificar la asociación entre dos variables dicotómicas. En la tabla anterior el valor del odds ratio es aproximadamente 2.6, que coincide con el riesgo relativo.

Medidas de asociación

		Enfermedad	
		Sí	No
Exposición	Sí	a	b
	No	c	d

Incidencia expuestos (I_{e+}) = $a / a+b$

Incidencia no expuestos (I_{e-}) = $c / c+d$

$$\text{Riesgo relativo (RR)} = I_{e+} / I_{e-} = \frac{(a / a+b)}{(c / c+d)}$$

EJEMPLO		Enfermedad Coronaria			
		Enfermos	No Enfermos	Totales	Tasa de Incidencia
HABITO DE FUMAR	Fumador	84	2.916	3.000	84/3.000 (0,028 ó 28 x 1.000)
	No Fumador	87	4.913	5.000	87/5.000 (0,0174 ó 17.4 x 1000)
	Totales	171	7.829	8.000	

$$RR = 28 / 17.4 = 1.61$$

Odds Ratio (OR)



		Casos	Controles
Exposición	Sí	a	b
	No	c	d

Odds enfermedad expuestos = a / b

Odds enfermedad no expuestos = c / d

Odds ratio o razón de odds (OR) = $(a/b) / (c/d) = a \cdot d / b \cdot c$

EJEMPLO		Cáncer de endometrio		
		Presente	Ausente	Totales
Uso de Estrógenos	Presente	55	19	74
	No Presente	128	164	292
	Totales	183	183	366

$$OR = \frac{55 \times 164}{19 \times 128} = 3.71$$