BIOESTADÍSTICA BÁSICA

MEDIDAS DE FORMA & DISTRIBUCIÓN NORMAL ESTANDAR

PROBABILIDADES

CLASE 3

MEDIDAS DE FORMA

Permiten conocer qué forma tiene la curva que representa la serie de datos de la muestra.

Se clasifican en: asimetría - curtosis

Asimetría: mide si la curva tiene o no, una forma simétrica - Dispersión

Curtosis: mide si los valores de la distribución están más o menos concentrados alrededor de los valores medios de la muestra- Dispersión

ASIMETRÍA

Mide el grado de asimetría de la distribución de sus datos en torno a la media.

Las colas de la variable están constituidas por los valores alejados de la media (valores extremos).

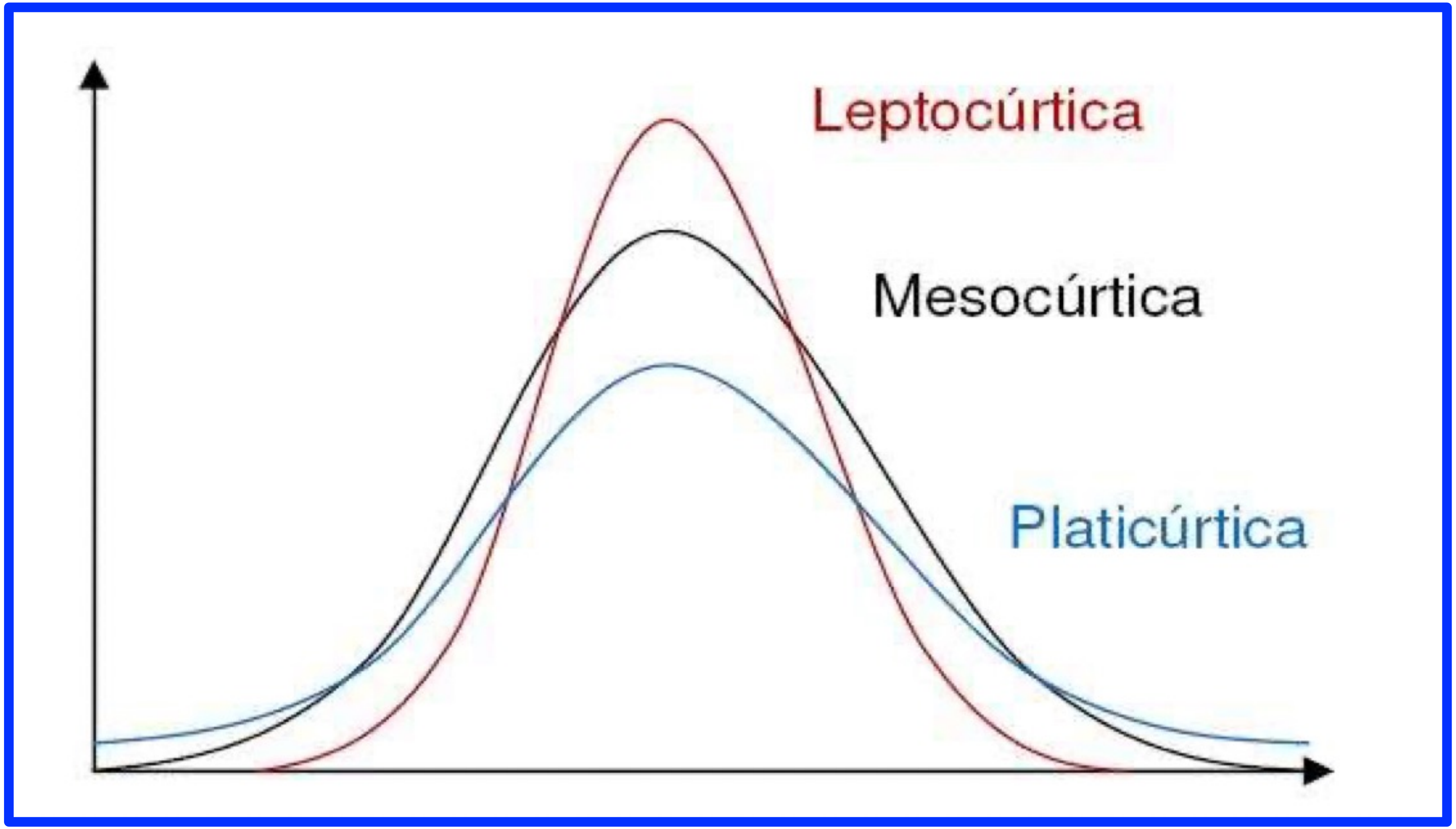
Una variable es asimétrica si su cola tiene un lado más largo que otro, y simétrica si ambas colas son iguales.

Figura 5-1

- Si As > 0 / derecha / la cola derecha es más larga / +
- Si AS = 0 / simétrica / ambas colas son iguales
- Si AS < 0 / izquierda / la cola izquierda es más larga / -

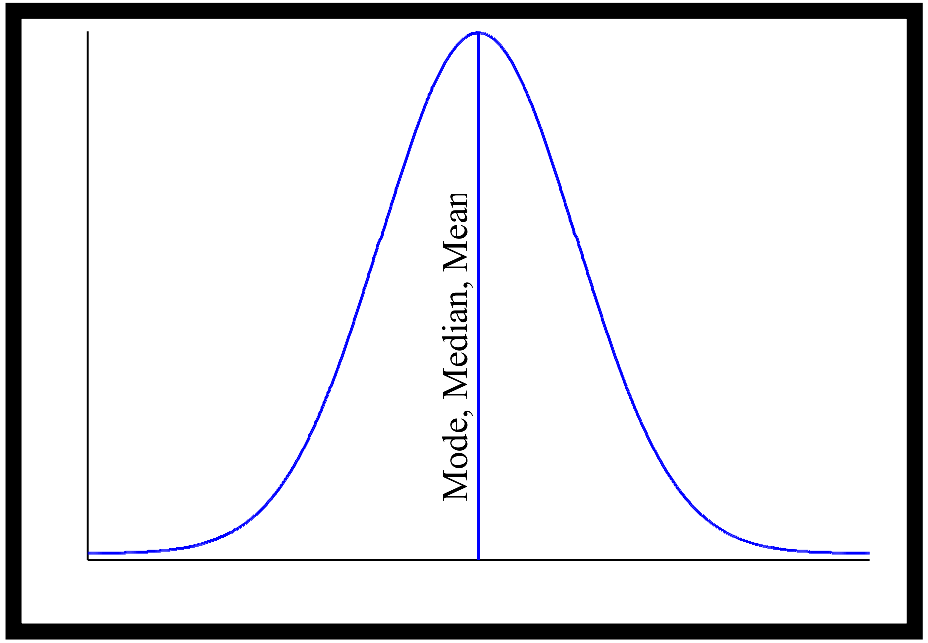
CURTOSIS O APUNTAMIENTO

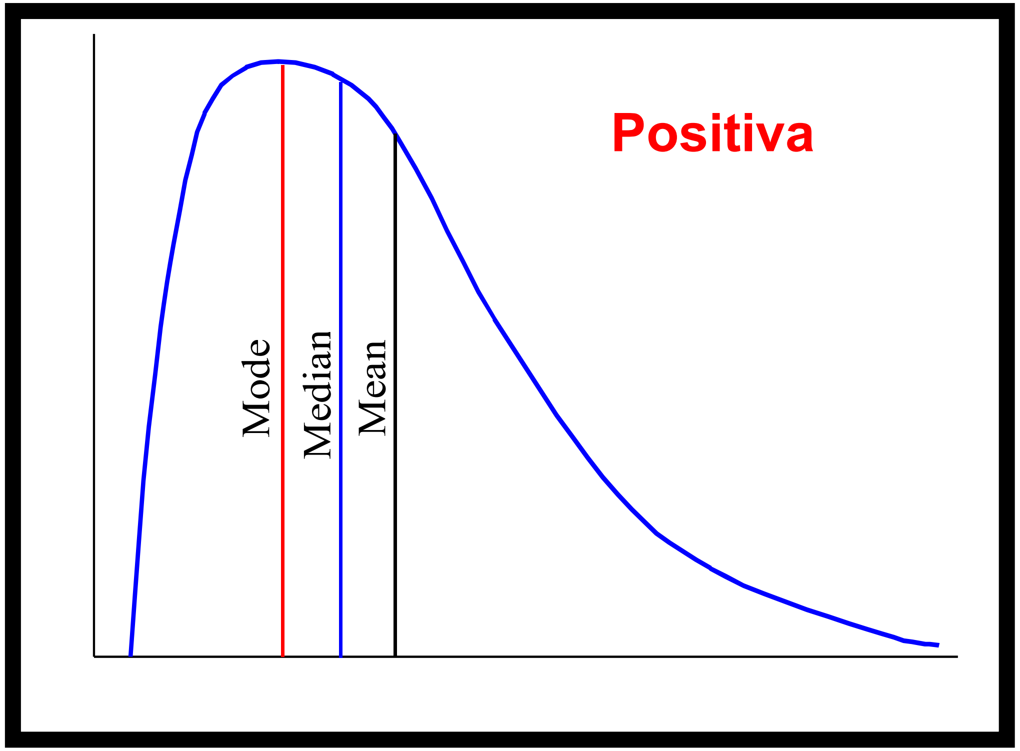
Mide el grado de concentración de los valores que toma en torno a la media.

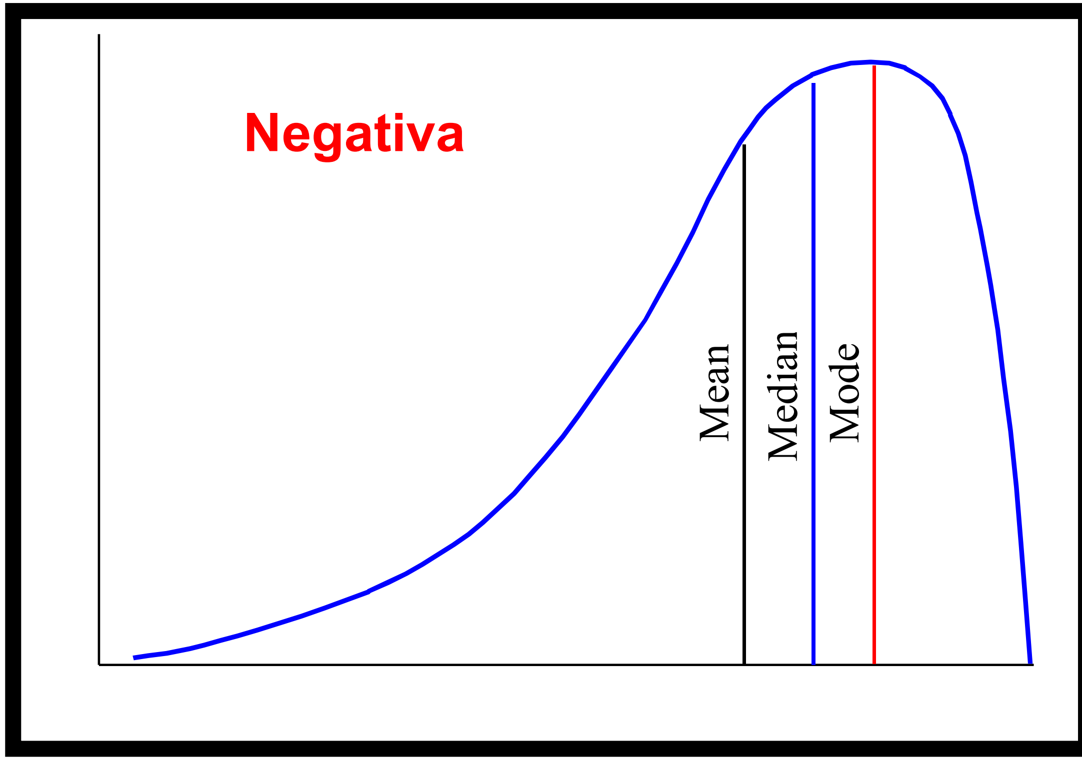


- Leptocúrtica K>0 : los valores están muy concentrados en torno a la media, hay pocos valores extremos (apuntamiento pronunciado)
- Mesocúrtica K=0 : Los valores son simetricos - una distribución normal.
- Platicúrtica K<0 : Hay muchos valores extremos (dispersión).

¿QUÉ SUCEDE SI LA DISTRIBUCIÓN DE DATOS ES PERFECTAMENTE SIMÉTRICA?


Mode, Median, Mean


Positiva
Mode
Median
Mean


Negativa
Mean
Median
Mode

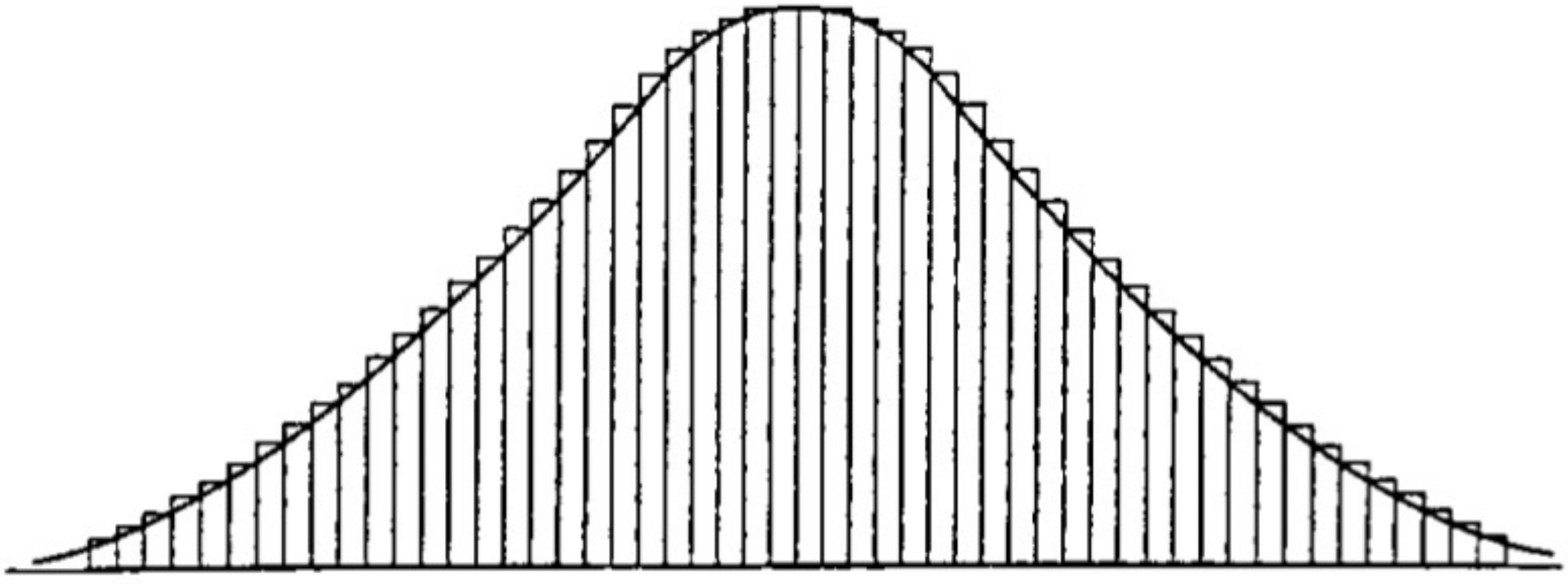


Figure 3.2 Histogram based on a large data set of weights.

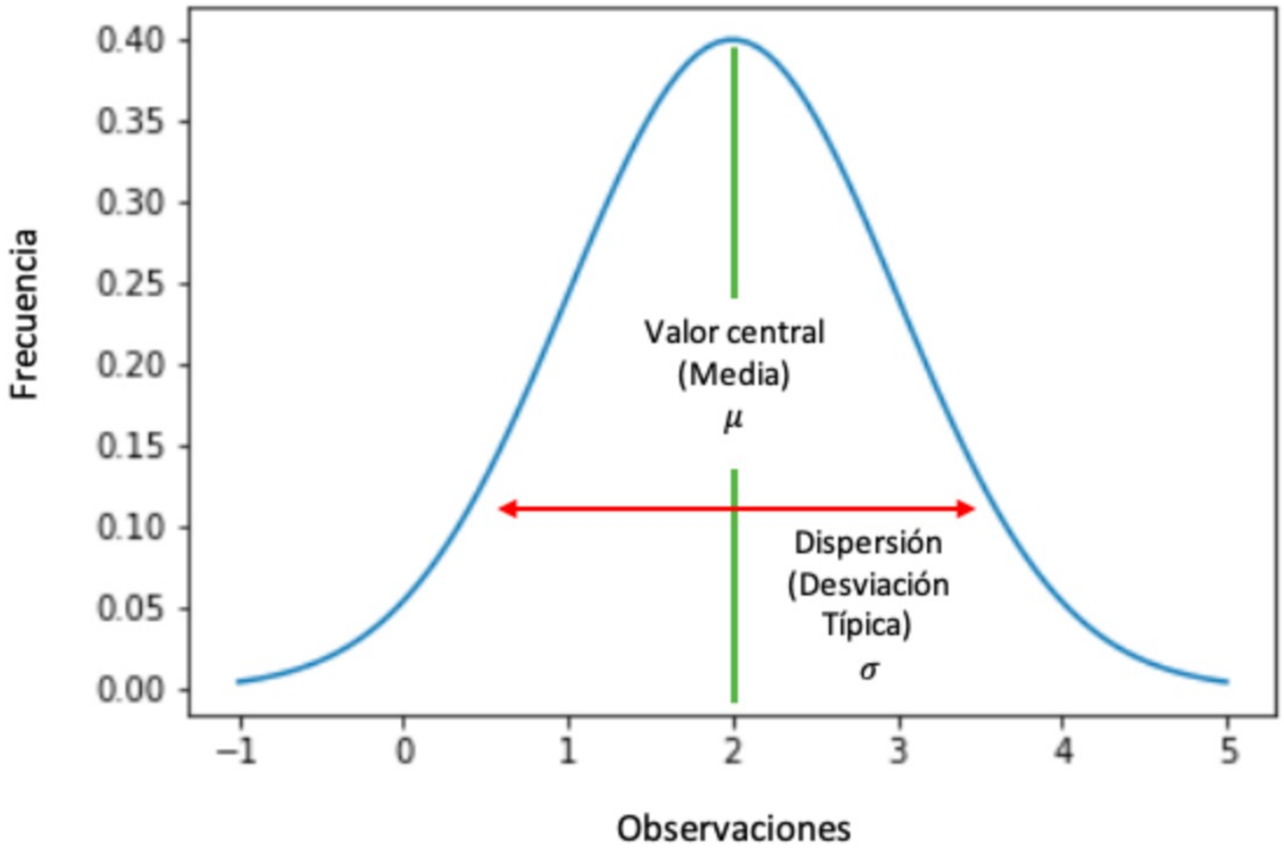
DISTRIBUCIÓN NORMAL

INTRODUCTORY BIOSTATISTICS /CHAP T. LE
CAPITULO 3

¿PARA QUÉ NOS SIRVE?

Se puede calcular la probabilidad de que varios valores ocurran dentro de ciertos rangos o intervalos.

La distribución normal o distribución de Gauss representa la forma en la que se distribuyen los diversos valores númericos de las variables cuantitativas.



Función de densidad de una distribución normal.

PROPIEDADES DE LA CURVA NORMAL

1

- Toma en cuenta la **media(μ) y la desviación estándar(σ).**

2

- El **área bajo la curva** es igual a 1.

3

- Es **simétrica** respecto al centro, o a la media. El 50% de los valores son mayores que la media, y 50% de los valores son menores que la media.

4

- La **media es igual a la mediana y a la moda.**

EXISTEN MILES DE PROBABILIDADES PARA ARMAR UNA DISTRIBUCIÓN NORMAL:

Media 7 – Desviación estándar de 1,5
Media 3 – Desviacón estándar de 2,8

Esto equivale a miles tablas para encontrar la probablilidad. Dado que eso no es práctico se estandariza la distribución con la tipificación de la variable aleatoria X a Z la cual sigue una distribución normal de media 0 y desviación estándar 1. Por lo tanto se utiliza una tabla estandarizada, la tabla Z.

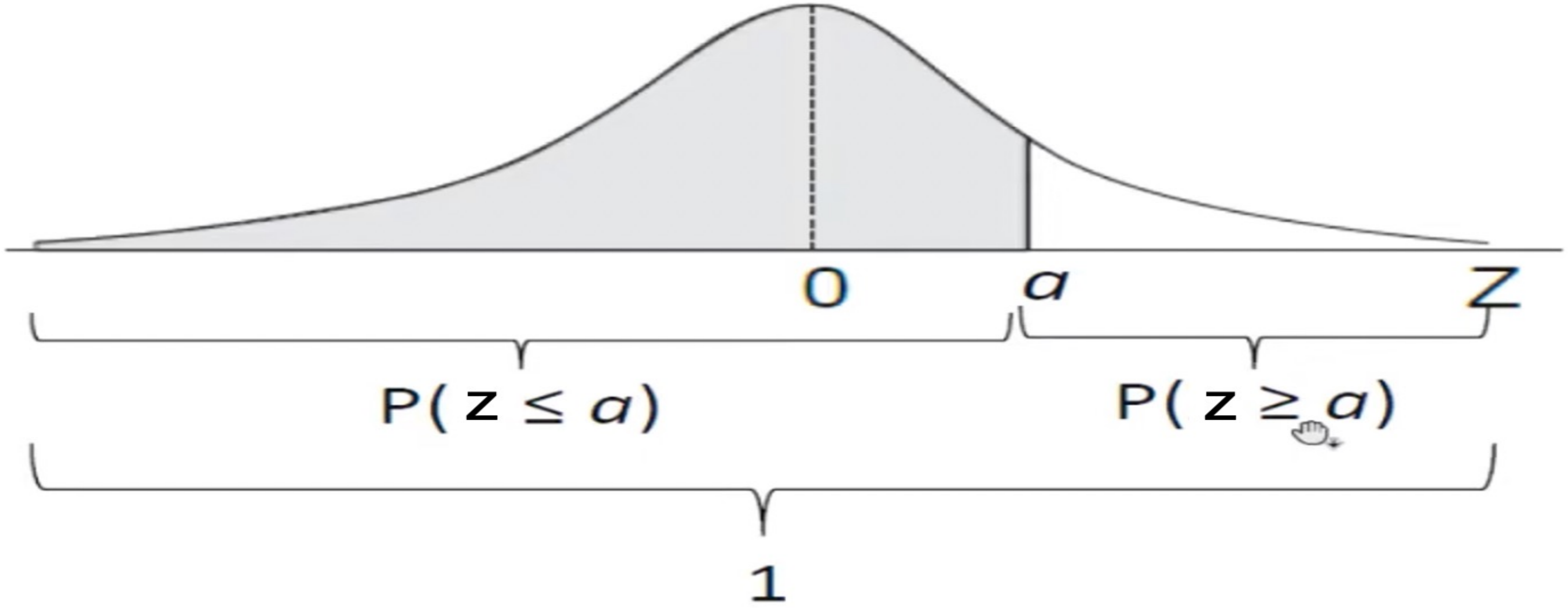
Las puntuaciones z transformadas se pueden utilizar para determinar las áreas bajo la curva para cualquier distribución normal.

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

PROPIEDADES PARA RESOLVER LA CURVA NORMAL

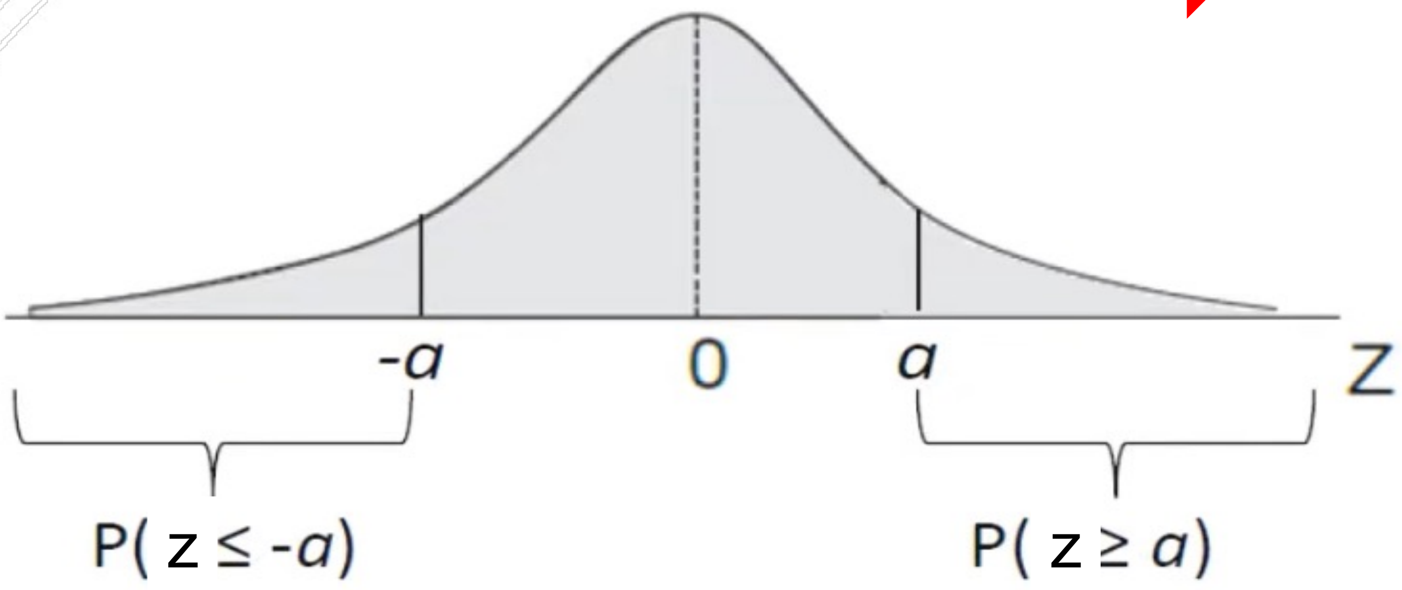
- <u>Propiedades</u>

1) $P(Z \geq a) = 1 - P(Z \leq a)$



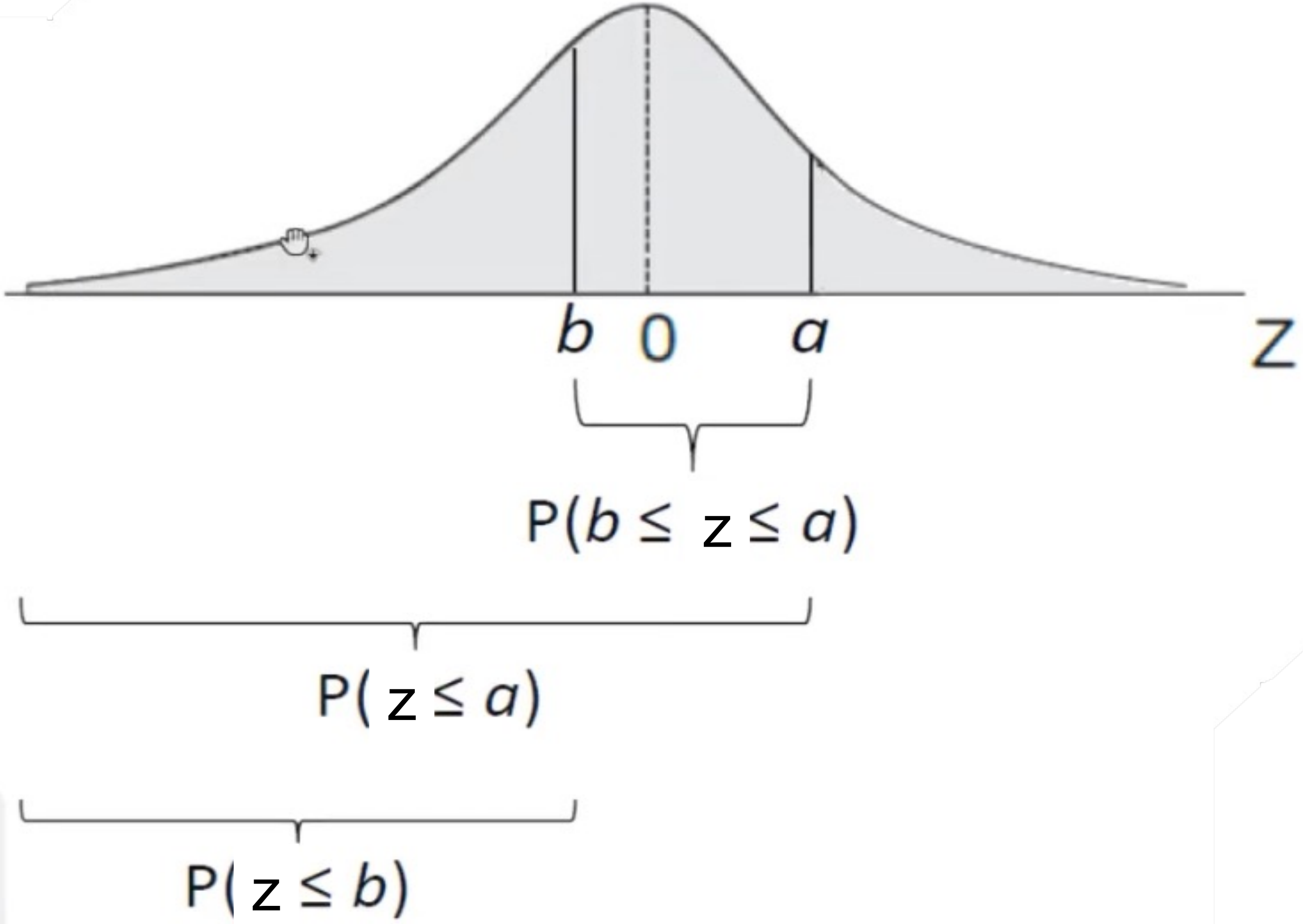
La P (Z<=a) + P (Z>=a) = 1 porque la tabla no contiene todos los valores.

2) $P(Z \leq -a) = P(Z \geq a)$



SON SIMÉTRICOS RESPECTO A 0

3) $P(b \leq Z \leq a) = P(Z \leq a) - P(Z \leq b)$



EJERCICIOS

Supongamos que se sabe que el peso de los sujetos de una determinada población sigue una distribución aproximadamente normal, con una media de 80 Kg y una desviación estándar de 10 Kg.

1. ¿Cuál es la probabilidad de que una persona, elegida al azar, tenga un peso superior a 100 Kg?

2. ¿Cuál la probabilidad de que el peso de un sujeto esté entre 60 y 100 Kg

$$Z = \frac{X - 80}{10} \Rightarrow Z = \frac{100 - 80}{10} = 2$$

$$P(Z) > 2 \Rightarrow P(Z > 2) = 1 - P(Z < 2)$$

$$P\ (Z > 2) = 1 - P(0{,}9772)$$

Por lo tanto, la probabilidad buscada de que una persona elegida aleatoriamente de esa población tenga un peso mayor de 100 Kg , es de:

$$1 - 0.9772 = 0.0228$$

Es decir, aproximadamente de un 2.3%.

2. La probabilidad de que el peso de un sujeto esté entre 60 y 100 Kg:

$$Z = \frac{X - 80}{10}$$

$$Z = \frac{60 - 80}{10} = -2$$

$$Z = \frac{100 - 80}{10} = 2$$

$$\boxed{P(b \leq \mathbf{Z} \leq a) = P(\mathbf{Z} \leq a) - P(\mathbf{Z} \leq b)}$$

P (Z<= 2) – P (Z<=-2)

P (0,9772) - P (0,0228)

La probabilidad buscada de que una persona elegida al azar tenga un peso entre 60 y 100 Kg es de:

0.9772-0.0228=0.9544

Es decir, de un 95%.

PROBABILIDADES VALIDEZ-CONDICIONALES-BAYES

PROBABILIDADES

Es un valor entre 0 y 1 que indica la posibilidad relativa de que ocurra un evento.

Entre 0 y 1 puede ser: 0 - 0,5 - 0,7, 1.

El valor de la probabilidad es 0 cuando es imposible que ocurra dicho evento.

Por ejemplo, tenemos un dado de 6 caras y nos dicen

¿cuál es la probabilidad de tener un 8, al lanzar el dado una vez?

¿Es posible tener un 8? No, no es imposible.

Mientras más cerca al 0, menos probabilidades hay.

FÓRMULA ESTÁNDAR DE PROBABILIDAD

$$\boxed{P\,(x) = \frac{Número\ de\ casos\ favorables\ del\ evento\ x}{Número\ de\ casos\ posibles}}$$

La salud pública y epidemiología, están dirigidas por **probabilidades;** y, al mismo tiempo manejan la **incertidumbre.**

1 historia clínica

2 sintomas

3 exploración física

4 realización de pruebas complementarias

Ante varias hipótesis diagnósticas, se realiza el diagnóstico diferencial y **las pruebas complementarias** para aclarar las dudas.

VALIDEZ

La validez parte de una prueba dicotómica, que clasifica a cada persona como sano o enfermo en función de que el resultado de la prueba sea positivo o negativo.

Generalmente un resultado positivo se asocia con presencia de enfermedad y un resultado negativo con la ausencia.

Cuando se estudia una muestra de personas, los datos obtenidos permiten clasificar a los sujetos en cuatro grupos; según, una tabla de 2 x 2.

Relación entre el resultado de una prueba diagnóstica (tamizaje) y la presencia o ausencia de una enfermedad (confirmación)		
Resultado de la prueba (tamizaje)	**Verdadero diagnóstico**	
	Enfermo	**Sano**
Reactivo (Positivo)	Verdaderos Positivos (VP) (a)	Falsos Positivos (FP) (b)
No reactivo (Negativo)	Falsos Negativos (FN) (c)	Verdaderos Negativos (VN) (d)

SENSIBILIDAD

1. Probabilidad de clasificar correctamente a un individuo enfermo, es decir, la probabilidad de que para un sujeto enfermo se obtenga en la prueba de tamizaje un resultado positivo.

2. La sensibilidad es la capacidad del test para detectar la enfermedad.

$$\frac{VP}{VP + FN} = \frac{a}{a + c}$$

ESPECIFICIDAD

1. Probabilidad de clasificar correctamente a un individuo sano, la probabilidad de que para un sujeto sano se obtenga un resultado negativo.
2. Capacidad para detectar a los sanos:

$$\frac{VN}{VN + FP} = \frac{d}{b + d}$$

SEGURIDAD: VALOR PREDICTIVO

La sensibilidad y la especificidad informan acerca de la probabilidad de obtener un resultado concreto (positivo o negativo) en función de la verdadera condición de la persona respecto de la enfermedad.

VALOR PREDICTIVO POSITIVO (VPP)

- Probabilidad de padecer la enfermedad si se obtiene un resultado positivo (reactivo) en la prueba.
- El VPP se puede estimar a partir de la proporción de pacientes con un resultado positivo en la prueba que finalmente resultaron estar enfermos

$$\frac{VP}{VP + FP} = \frac{a}{a + b}$$

VALOR PREDICTIVO NEGATIVO (VPN)

- Es la probabilidad de que un sujeto con un resultado negativo (no reactivo) en la prueba, esté realmente sano.
- Se estima dividiendo el número de verdaderos negativos entre el total de pacientes con un resultado negativo en la prueba:

$$\frac{VN}{FN + VN} = \frac{d}{c + d}$$

La siguiente tabla muestra el número de mujeres que recibieron y no recibieron un diagnostico de cáncer en el último año, y el resultado de su mamografia.

	Cáncer	No cáncer	Total
Mamografia +	15 (a)	135 (b)	150
Mamografia -	45 (c)	470 (d)	515
Total	60	605	665

Sensibilidad = VP/VP+FN 15/60 = 0,25 x 100 = 25%

Especificidad = VN/VN+FP 470/605 = 0,78 x 100 = 78%

VALOR PREDICTIVO

	Cáncer	No cáncer	Total
Mamografia +	15 (a)	135 (b)	150
Mamografia -	45 (c)	470 (d)	515
Total	60	605	665

VPP = VP/VP+FP $15/150 = 0{,}10 \times 100 = 10\%$

VPN= VN/VN+FN $470/515 = 0{,}91 \times 100 = 91\%$

PROBABILIDAD CONDICIONAL

La probabilidad de que ocurra un evento dado que ha ocurrido otro evento

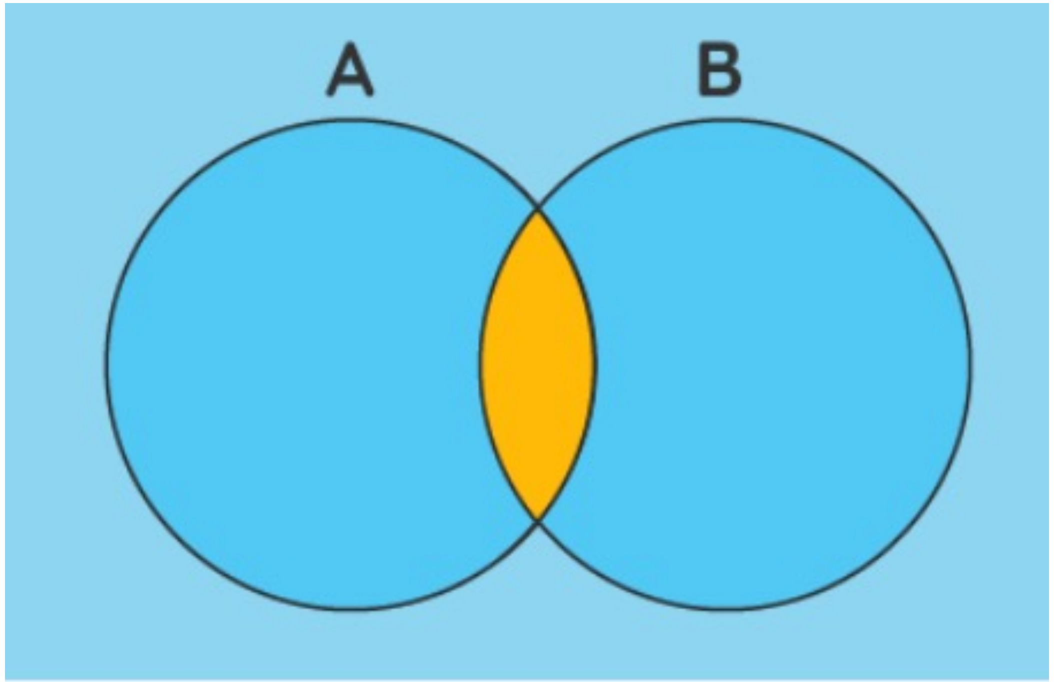
$P(A \mid B)$

se lee como "dado que”

La probabilidad de que ocurra el evento A, dado que ha ocurrido el evento B

LA FÓRMULA PARA LA PROBABILIDAD CONDICIONAL:

$$P(A|B) = P(B\&A) / P(B)$$



$$P(A \mid B) = P(B \cap A) / P(B)$$

Los eventos no son independientes

La siguiente tabla muestra el número de mujeres que recibieron y no recibieron un diagnostico de cáncer en el último año, y el resultado de su mamografia. Encuentra la probabilidad de que una mujer elegida al azar:

a) Tiene cáncer dado que tiene una mamografia con resultado positivo (tumor)
b) Tiene una mamografia con resultado positivo (tumor) dado que tiene cáncer

	Cáncer	**No cáncer**	**Total**
Mamografia +	15	135	150
Mamografia -	45	470	515
Total	60	605	665

P(cáncer | mamografía +) = $15/150 = 0{,}1 \times 100 = 10\%$

P(mamografía + | cáncer) = $15/60 = 0{,}25 \times 100 = 25\%$

$P(A \mid B) = P(B \cap A) / P(B)$

P(A | B) no es igual a P(B | A)

Creemos que el orden los factores no altera el producto eso sirve para la multiplicación aquí no tenemos esto, tenemos **dado que** es decir, un condición no funciona con la propiedad conmutativa.

Las probabilidades condicionales son utilizadas por las compañías de seguros privados de salud para determinar sus tarifas.

Analizan la probabilidad condicional de que tenga un accidente, dada su edad, dada enfermedades pre existente, etc.

Y fijan el precio de su póliza en función de esa probabilidad.

TEOREMA DE BAYES

Se aplica en bioestadística para calcular probabilidades para detección de enfermedades, NO OBSTANTE, no se evalúa el test diagnóstico, sino, un evento con otro

La probabilidad de que ocurra el evento A dado que ha ocurrido el evento B es igual a: probabilidad del evento A por la probabilidad de que ocurra el evento B dado que ha ocurrido el evento A y todo ello dividido para la probabilidad del evento B

PRIMER PASO: Para no confundir el evento colocamos el esqueleto del teorema

$$P(A|B) = \frac{P(A)\times P(B|A)}{P(B)}$$

La probabilidad de B nunca debe ser 0

Se colocan los elementos en ese orden, primero A luego B, luego A y luego B, así sucesivamente.

Calcular la probabilidad de una persona que tiene fiebre (F) dado que tiene gripe (G)

$$P\ (F|G) = \frac{P(F) x\ P\ (G/F\ )}{P\ (G\ )}$$

EJEMPLO

En el consultorio de Ma Alejandra, el 40% de los pacientes fingen tener gastroenteritis para obtener un descanso médico. El 10% de los pacientes del consultorio, son hombres. La probabilidad de que el paciente finja una enfermedad dado que es hombre, es del 50%. Calcular la probabilidad de que un paciente sea hombre, dado que finge una enfermedad.

Aquí me dice que 40 de 100 fingen. Entonces 40 entre 100 = 0,40

$P(F)=0,4$

$P(H)=0,1$

$P(F \mid H)=0,5$

Calcular la probabilidad de que un paciente sea hombre, dado que finge una enfermedad.

$$P\,(H/F) = \frac{P(H)x\,P\,(F/H)}{P\,(F)}$$

$$P\,(H/F) = \frac{P(H)\,x\,P\,(F/H)}{P\,(F)}$$

P(H/F) = 0,1 X 0,5 / 0,4 = 0,125

Si quiero trabajar con porcentaje, lo multiplico por 100% :
0,125 x 100% = 12,5%

La probabilidad de que el paciente sea hombre
dado que finge es 12,5%

Y ¿SI TENGO MÁS ELEMENTOS?

1. MAPA DE ÁRBOL

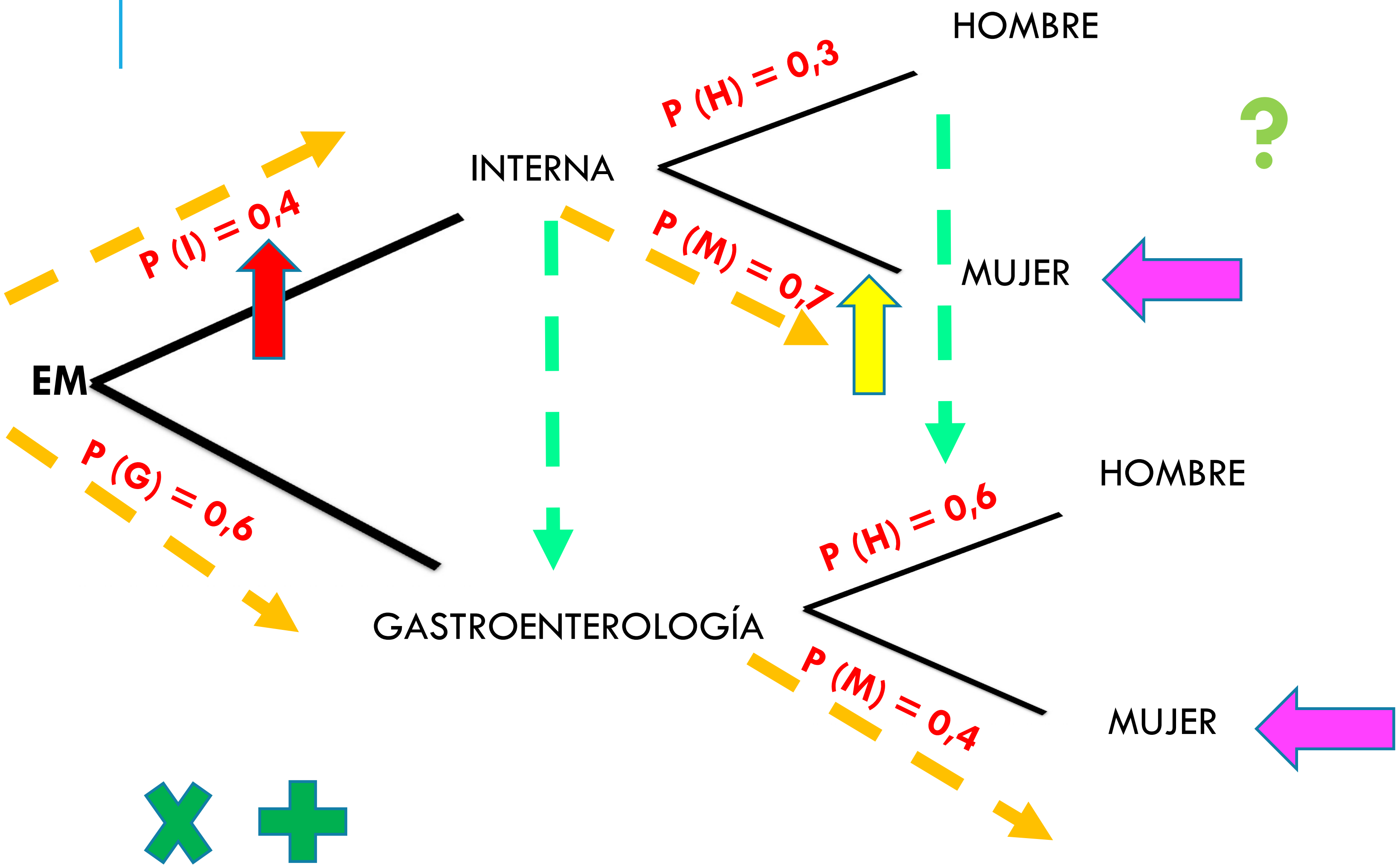
2. LAS PROBABILIDADES DEBEN, SIEMPRE, SUMAR 1

- Una clínica de Quito, cuenta solo con dos especialidades médicas. El 40% de los médicos son internistas, y el 60% son gastroenterólogos.

- De los internistas, el 30% son hombres. Mientras que de los gastroenterólogos el 40% son mujeres.

- Se reporta un accidente cerca de la Clínica. En ese caso, si se selecciona una médica para colaborar con los accidentados, cuál es la probabilidad de que sea internista.

¿Cuál es la probabilidad de que sea internista dado que sea mujer?



$$\text{P (I)} = \frac{P(\quad) x\, P(\ /\ )}{P(\ )}$$

$$\text{P (I/M)} = \frac{P(I) x\, P(M/I)}{P(M)}$$

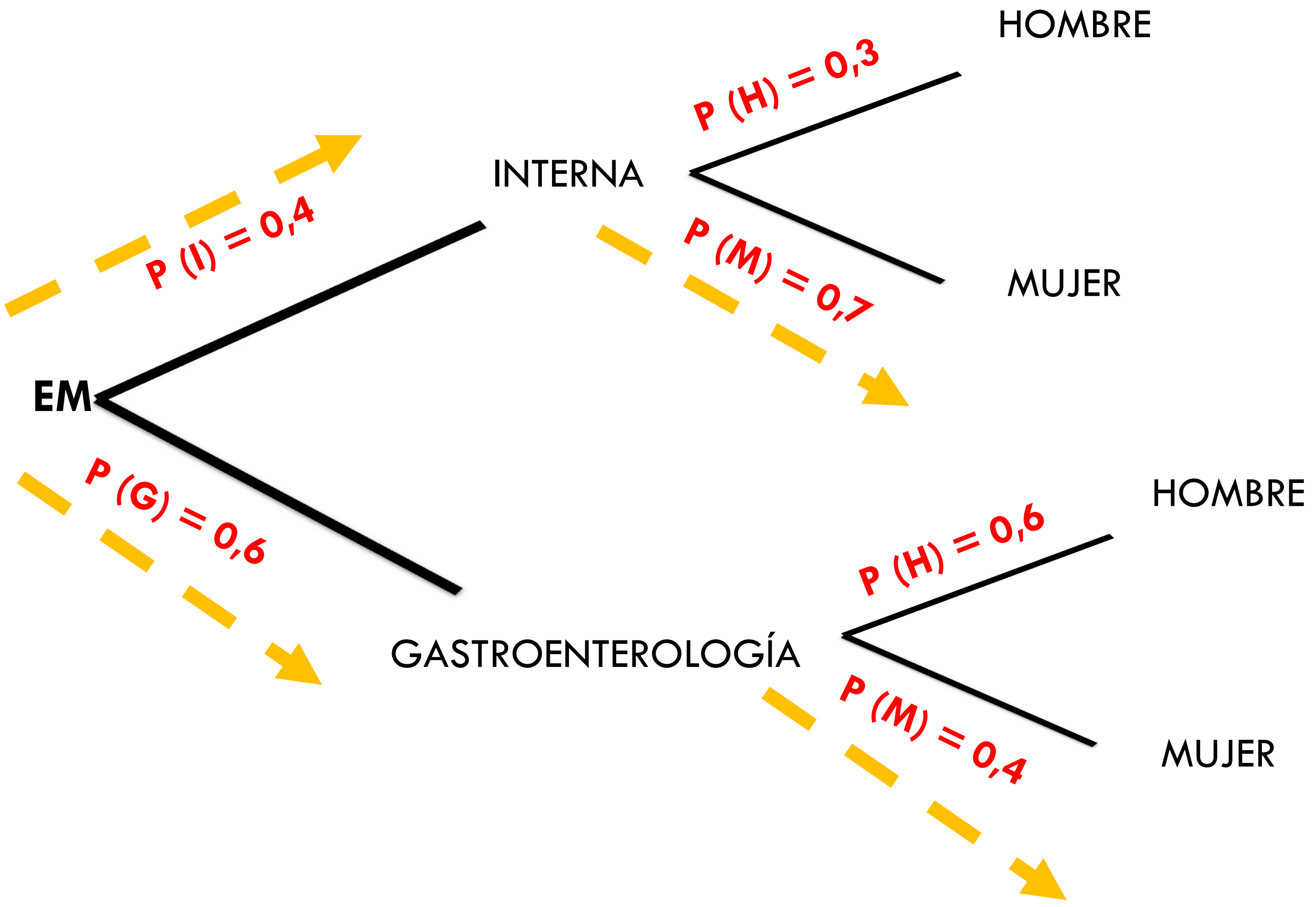
- Interna a mujer = 0,4 x 0,7
- Gastroenterología a mujer = 0,6 x 0,4

0,4 x 0,7. + 0,6 x 0,4

$$\text{P (I/M)} = \frac{P(0,4) x\, P(0,7)}{P(0,4\ X\ 0,7 + 0,6\ X\ 0,4)}$$

P (I/M)= 0,5385
P (I/M)= 53,85%

Cuál es la probabilidad de que sea internista dado que sea mujer?



$$P\ (x) = \frac{\textit{Número de casos favorables del evento } x}{\textit{Número de casos posibles}}$$

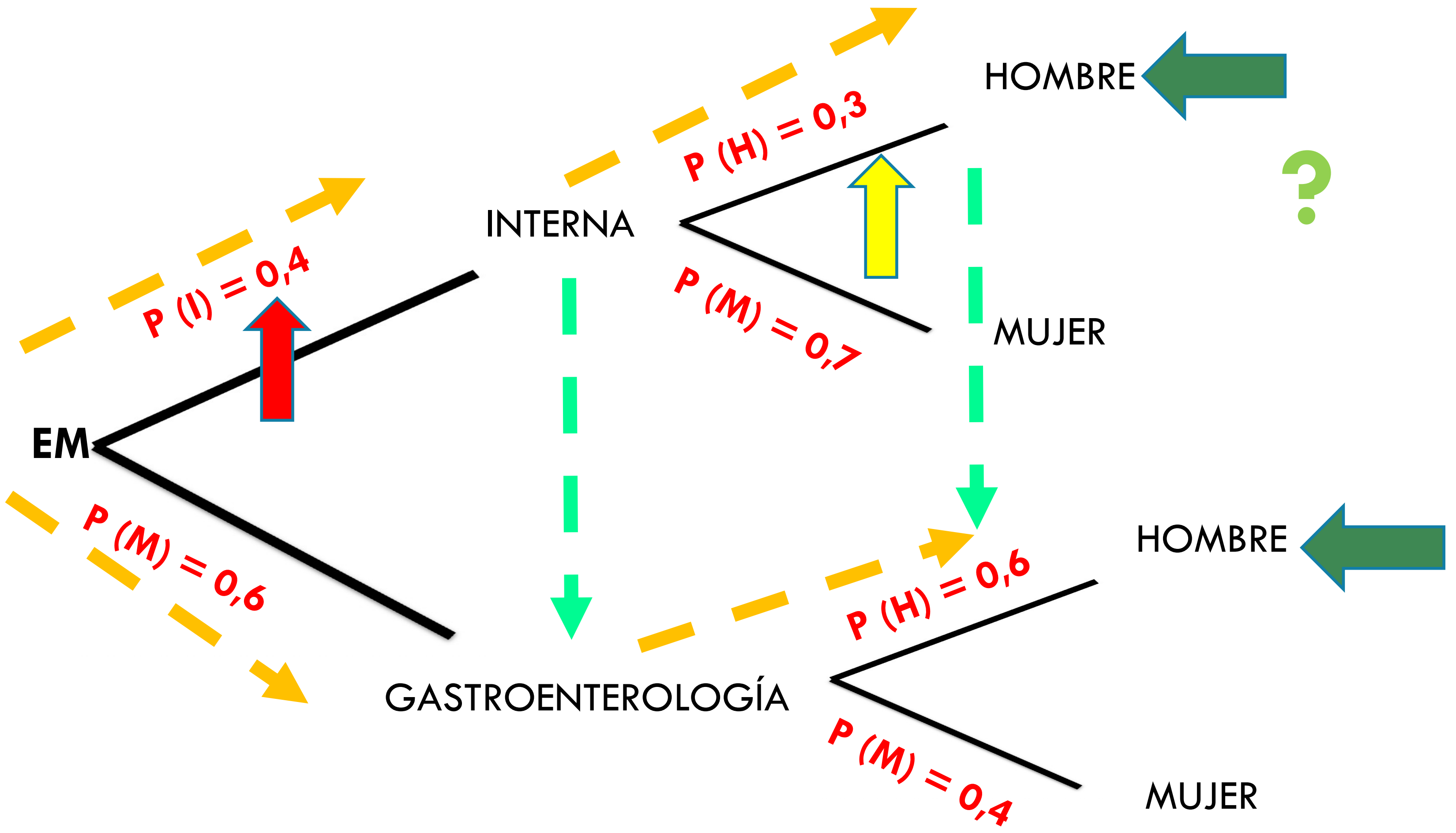
$$P\ (x) = \frac{0{,}4\ x\ 0{,}7}{0{,}4\ x\ 0{,}7 + 0{,}6 * 0{,}4}$$

P (I/M)= 0,5385

P (I/M)= 53,85%

B) SI SE SELECCIONA UN HOMBRE, CUÁL ES LA PROBABILIDAD DE QUE SEA INTERNISTA

Cuál es la probabilidad de que sea internista dado que sea hombre?



$$P\ (I/H) = \frac{P(I)x\ P\ (\ H/I\ )}{P\ (\ H)}$$

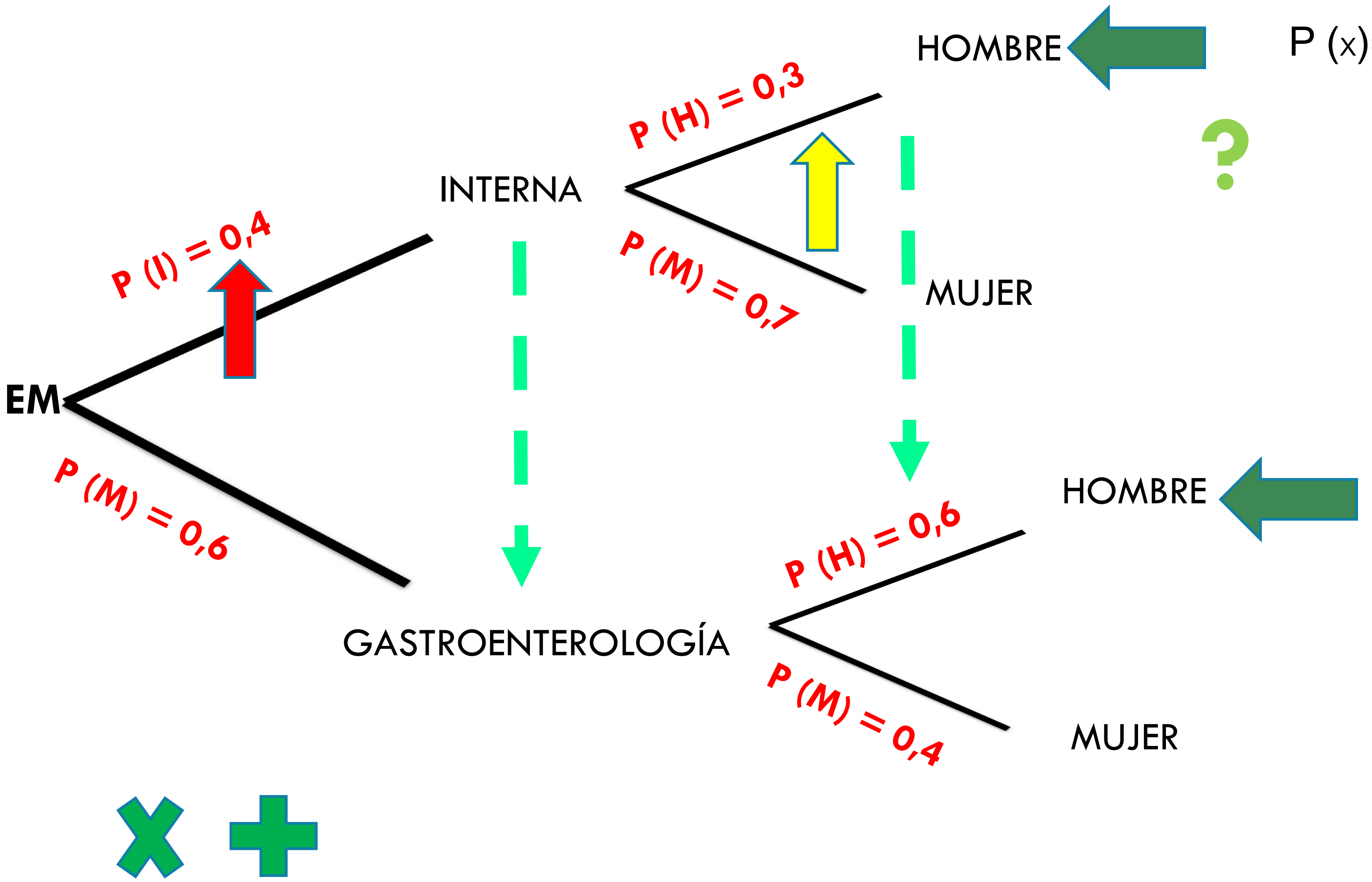
- Interna a hombre =
- Gastroenterología a hombre =

$$P\ (_{I/M}) = \frac{P(0{,}4)x\ P\ (0{,}3)}{P\ (0{,}4X0{,}3+0{,}6\ X\ 0{,}6)}$$

$P\ (_{I/h}) = 0{,}25$

$P\ (I/h) = 25\%$

Cuál es la probabilidad de que sea internista dado que sea hombre?



$$P\ (x) = \frac{Número\ de\ casos\ favorables\ del\ evento\ x}{Número\ de\ casos\ posibles}$$

$$P\ (_{I/M}) = \frac{P(0{,}4X0{,}3)}{P\ (0{,}4X0{,}3+0{,}6\ X\ 0{,}6)}$$

$P\ (_{I/h}) = 0{,}25$

$P\ (I/h) = 25\%$

Probabilidad condicional	Teorema de Bayes
La probabilidad condicional es la probabilidad de un evento A que se basa en la aparición de otro evento B.	El teorema de Bayes se deriva utilizando la definición de probabilidad condicional. La fórmula del teorema de Bayes incluye dos probabilidades condicionales.
Fórmula: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$	Fórmula: $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$