

VARIABLES ALEATORIAS Y **DISTRIBUCIONES DE** **PROBABILIDAD**

VARIABLES ALEATORIAS

En el capítulo de probabilidades se estudiaron fundamentos de la teoría de probabilidades. Es posible combinar estas ideas para obtener distribuciones de probabilidad que se parecen bastante a las distribuciones de frecuencias relativas, la diferencia más importante entre las distribuciones de probabilidad y las de frecuencia relativa, es que las distribuciones de probabilidad son probabilidades teóricas (modelo), mientras que las distribuciones de frecuencias relativas son probabilidades empíricas o proporciones (muestras). En el capítulo anterior vimos que los espacios muestrales no son necesariamente numéricos. Cuando por ejemplo lanzamos una moneda tres veces, podremos registrar un resultado como CCS. En estadística, sin embargo nos interesan los resultados numéricos, tal como el número de caras al lanzar una moneda tres veces.

VARIABLES ALEATORIAS

Definición

Una variable aleatoria es una variable cuyo número que depende del resultado aleatorio de un experimento. Más formalmente, una variable aleatoria es una regla que asigna un valor numérico (sólo uno) a cada punto en el espacio muestral de un experimento aleatorio.

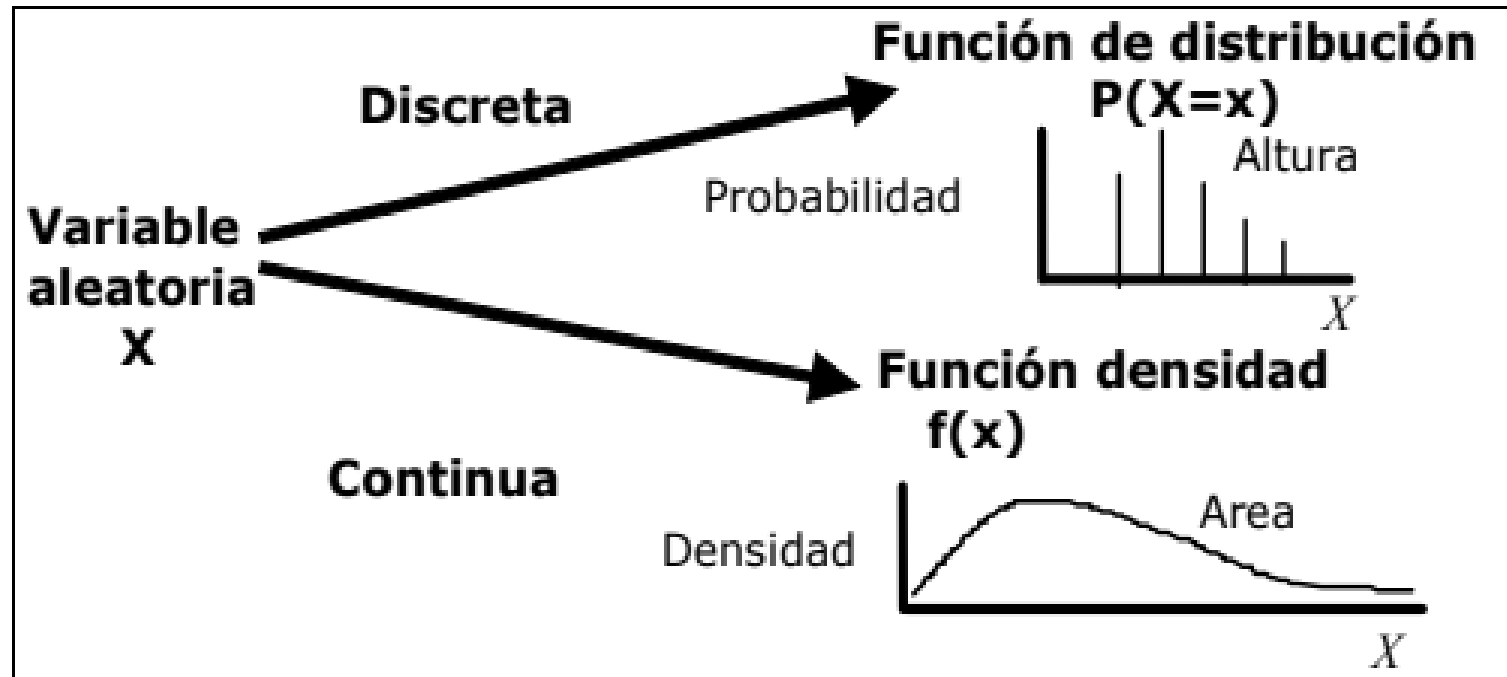
Nota: normalmente se usan letras mayúsculas y del final de abecedario (X, Y, o Z), para denotar variables aleatorias.

Se van a distinguir dos tipos de variables aleatorias:

VARIABLES ALEATORIAS

Una **variable aleatoria discreta** puede tomar valores finitos o contables.

Una **variable aleatoria continua** puede tomar cualquier valor en un intervalo.



VARIABLES ALEATORIAS DISCRETAS

Variables aleatorias discretas

Si la variable aleatoria es discreta la describimos según su **distribución de probabilidades**, que consiste en una lista de valores posibles de la variable y la proporción de veces que esperamos que ocurran:

X	x_1	x_2	\dots	x_k
$p(x)$	p_1	p_2	\dots	p_k

☒ Ejemplo

Volvamos al experimento de lanzar una moneda 3 veces.

Espacio muestral $S = \{CCC, CCS, CSS, CSC, SSS, SSC, SCC, SCS\}$

Sea $X = \text{número de caras} = 0, 1, 2, 3$

x				
$p(x)$				

VARIABLES ALEATORIAS DISCRETAS

☒ Ejemplo

Modelo para el **número de libros** en mochilas de estudiantes.

Sea X una variable aleatoria que representa el número de libros que llevan en la mochila los estudiantes de esta Universidad:

x	0	1	2	3
$p(x)$	0,5	0,2	0,2	0,1

- a) Describa la forma de la distribución.
- b) ¿Qué proporción de estudiantes llevan 3 o menos libros ($X \leq 3$)?
- c) ¿Qué proporción de estudiantes llevan más de 2 libros ($X > 2$)?
- d) ¿Qué proporción de estudiantes llevan entre 2,1 y 2,8 libros ($2,1 < X < 2,8$)?
- e) ¿Qué proporción de estudiantes llevan entre 1 y 2 libros ($1 \leq X \leq 2$)?

La **distribución de probabilidades** de una variable aleatoria **discreta** X es una función (tabla o regla), denotada por $p(x)$ o $P[X=x]$, que asigna una probabilidad a cada valor posible de la variable aleatoria X .

Propiedades de una función de distribución:

1. Los valores de las probabilidades están entre 0 y 1 ($0 \leq p(x) \leq 1$) para todo x .
2. La suma de las probabilidades es 1 ($\sum p(x) = 1$).

VARIABLES ALEATORIAS DISCRETAS



Ejemplo

Tamaño familiar.

Sea X el número de personas de hogares en el censo 2002:



En Internet <http://www.ine.cl/cd2002/index.php>

X	1	2	3	4	5	6	7 y más
$P(x)$	0,11	0,18	0,22	0,23	0,14	0,07	

- a) ¿Cuánto debe ser la probabilidad de que el tamaño familiar sea de 7 personas para que esta sea una distribución de probabilidades discreta legítima?
- b) Muestre gráficamente la distribución de probabilidades.
- c) ¿Cuál es la probabilidad de que un hogar elegido al azar tenga un tamaño familiar de más de 5 personas?
- d) ¿Cuál es la probabilidad de que un hogar elegido al azar tenga un tamaño familiar de no más de 2 personas?
- e) ¿Cuál es $P(2 < X \leq 4)$?

VARIABLES ALEATORIAS DISCRETAS



Pensemos

No todas las tablas representan un modelo discreto.

Un modelo discreto puede servir para describir la distribución de una variable cualitativa, pero no cualquier tabla representa una variable aleatoria. Considere por ejemplo la siguiente tabla que contiene información acerca del tipo de mascota que poseen en cierto barrio:

Mascota	Perro	Gato	Otras
Proporción	0,70	0,40	0,20

¿Es esta una distribución discreta legítima?

Definición:

Si X es una variable aleatoria discreta que toma valores x_1, x_2, \dots, x_k , con probabilidad p_1, p_2, \dots, p_k , entonces la **media o el valor esperado de X** está dado por:

$$E(X) = \mu_X = x_1 p_1 + x_2 p_2 + \dots + x_k p_k = \sum_{i=1}^k x_i p_i ,$$

la **varianza de X** está dada por:

$$\begin{aligned} Var(X) &= \sigma_X^2 = E[(X - \mu)^2] \\ &= E(X^2) - [E(X)]^2 \\ &= \sum x_i^2 p_i - \mu^2 \end{aligned}$$

y la **desviación estándar de X** está dada por: $\sigma_X = \sqrt{\sigma_X^2}$

VARIABLES ALEATORIAS DISCRETAS

✓ Ejemplo

En el caso del número de caras al lanzar 3 monedas, la **distribución de probabilidades** de X es:

x	0	1	2	3
p(x)	1/8	3/8	3/8	1/8

La **media** de X es:

$$E(X) = \mu_X = 0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right)$$

$$E(X) = 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = 1,5 \text{ caras}$$

$$Var(X) = 0^2\left(\frac{1}{8}\right) + 1^2\left(\frac{3}{8}\right) + 2^2\left(\frac{3}{8}\right) + 3^2\left(\frac{1}{8}\right) - (1,5)^2$$

$$Var(X) = 0 + \frac{3}{8} + \frac{12}{8} + \frac{9}{8} - (1,5)^2$$

$$Var(X) = \frac{24}{8} - 2,25 = 0,75$$

Luego, la **desviación estándar** de X es: $\sigma_X = \sqrt{0,75} = 0,866 \text{ caras}$.

✓ **Ejercicio:** Calcule las medias y desviaciones estándar en los demás ejercicios.

MODELOS PROBABILISTICOS DISCRETOS

Modelos discretos

- Ensayos de Bernoulli
- Distribución Binomial
- Distribución de Poisson

MODELO BERNOULLI

Descripción.

Partimos de un experimento aleatorio con sólo dos posibles resultados, que calificamos de éxito/fracaso.

Definimos la variable aleatoria:

$$X = \begin{cases} 1 & \text{si éxito} \\ 0 & \text{si fracaso} \end{cases}$$

Sea p la probabilidad de éxito.

El experimento se llama **ensayo de Bernoulli** y la variable aleatoria se dice que sigue una **distribución Bernoulli** de parámetro p .

Se escribe $X \sim \text{Ber}(p)$.

MODELO BERNOULLI

Ejemplo

Tirar una moneda al aire

$$X = \begin{cases} 1 & \text{Si sale Cara} \\ 0 & \text{Si sale Sello} \end{cases}$$

Es un ensayo Bernoulli, y X sigue una distribución Bernoulli de parámetro $1/2$.

Ejemplo

Una línea aérea estima que los pasajeros que compran un billete para un vuelo tienen una probabilidad igual a $0,05$ de no presentarse al embarque de dicho vuelo.

Definamos

$$Y = \begin{cases} 1 & \text{si el pasajero se presenta} \\ 0 & \text{si no lo hace} \end{cases}$$

Y sigue una distribución Bernoulli con parámetro $0,95$.

MODELO BERNOULLI

Función de Probabilidad:

$$P[X = 0] = 1 - p \quad P[X = 1] = p$$

Función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Propiedades

- $E[X] = p \times 1 + (1 - p) \times 0 = p$
- $E[X^2] = p \times 1^2 + (1 - p) \times 0^2 = p$
- $\text{Var}[X] = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$
- $\sigma[X] = \sqrt{p(1 - p)}$

DISTRIBUCION BINOMIAL

Descripción

Un ensayo Bernoulli de parámetro p se repite n veces de manera independiente. La variable *número de éxitos obtenidos*, sigue una **distribución Binomial** (de parámetros n y p).

Definición

Una variable X sigue una **distribución binomial** con parámetros n y p si

$$P[X = x] = \binom{n}{x} p^x (1 - p)^{n-x}$$

para $x = 0, 1, \dots, n$ donde

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Se escribe $X \sim B(n, p)$.

DISTRIBUCION BINOMIAL

Ejemplo

La línea aérea del ejemplo anterior ha vendido 80 billetes para un vuelo. La probabilidad de que un pasajero no se presente al embarque es de 0,05.

Definimos $X = \text{número de pasajeros que se presentan}$. Entonces (suponiendo independencia)

$$X \sim B(80, 0,95)$$

- La probabilidad de que los 80 pasajeros se presenten

$$P[X = 80] = \binom{80}{80} 0,95^{80} \times (1 - 0,95)^{80-80} = 0,0165$$

DISTRIBUCION BINOMIAL

Propiedades

- $E[X] = np$
- $\text{Var}[X] = np(1 - p)$
- $\sigma[X] = \sqrt{np(1 - p)}$

DISTRIBUCION DE POISSON

Descripción

Modela el número de sucesos *raros* que ocurren en un determinado periodo de tiempo o en una cantidad de espacio determinada.

Ejemplos: llamadas de telefono en una hora, erratas en una página, accidentes de tráfico, ...

Definición

Una variable X sigue una distribución Poisson de parámetro λ si

$$P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{para } x = 0, 1, 2, \dots$$

Se escribe $X \sim \mathcal{P}(\lambda)$.

DISTRIBUCION DE POISSON

Propiedades

- $E[X] = \lambda$
- $\text{Var}[X] = \lambda$
- $\sigma[X] = \sqrt{\lambda}$

Propiedad de la Poisson

Si $X \sim \mathcal{P}(\lambda)$ y representa el número de sucesos raros en una unidad de tiempo, e Y es una variable aleatoria que representa el número de dichos sucesos raros en un tiempo t , se tiene que:

$$Y \sim \mathcal{P}(t\lambda)$$

DISTRIBUCION DE POISSON

Ejemplo

El número medio de erratas por transparencias es de 0,2. Es decir, si X es el número de erratas por transparencia, entonces

$$X \sim \mathcal{P}(0,2)$$

¿Cuál es la probabilidad de que en una transparencia no haya erratas?

$$P[X = 0] = \frac{0,2^0 e^{-0,2}}{0!} = e^{-0,2} = 0,8187.$$

¿Cuál es la probabilidad de que en 4 transparencias haya exactamente una errata?

Sea Y el número de erratas en 4 transparencias. Sabemos que

$$Y \sim \mathcal{P}(0,2 \cdot 4) = \mathcal{P}(0,8)$$
$$P[Y = 1] = \frac{0,8^1 e^{-0,8}}{1!} = 0,8e^{-0,8} = 0,3595.$$

VARIABLES ALEATORIAS CONTINUAS

Variables Aleatorias Continuas

Definición:

Una **función de densidad** es una función o curva que describe la forma de una distribución.

El área total bajo la curva es igual a *uno* y calculamos probabilidades como áreas bajo la curva de densidad.

Propiedades de una función densidad:

La **función densidad** de una variable aleatoria **continua** X es una función, denotada por $f(x)$, que satisface:

1. $f(x) \geq 0$ La densidad es siempre mayor o igual a cero.
2. El área bajo la curva de densidad es uno.
3. $P(a < X < b) = P(a \leq X \leq b)$ = es el área o proporción entre a y b.

MODELOS PROBABILISTICOS CONTINUOS

Modelos continuos

- Distribución normal

DISTRIBUCION NORMAL

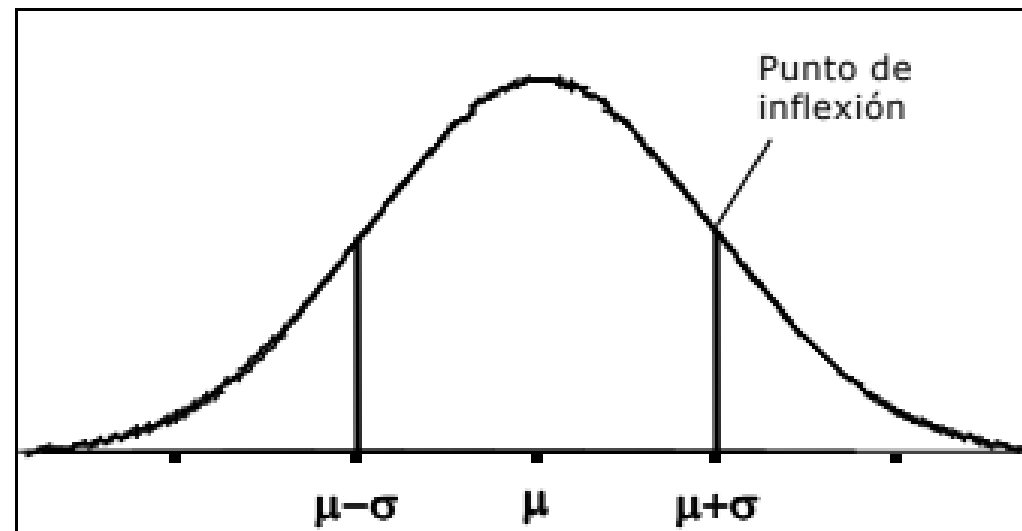
La distribución Normal

Existe una distribución de frecuencias teórica llamada distribución normal, que puede considerarse como modelo adecuado para la distribución de un gran número de variables en el campo biológico.

Notación: $X \sim N(\mu, \sigma)$ se lee: X es una variable aleatoria continua con distribución Normal, con media μ y desviación estándar σ .

La **función densidad** de una variable aleatoria Normal está dada por:

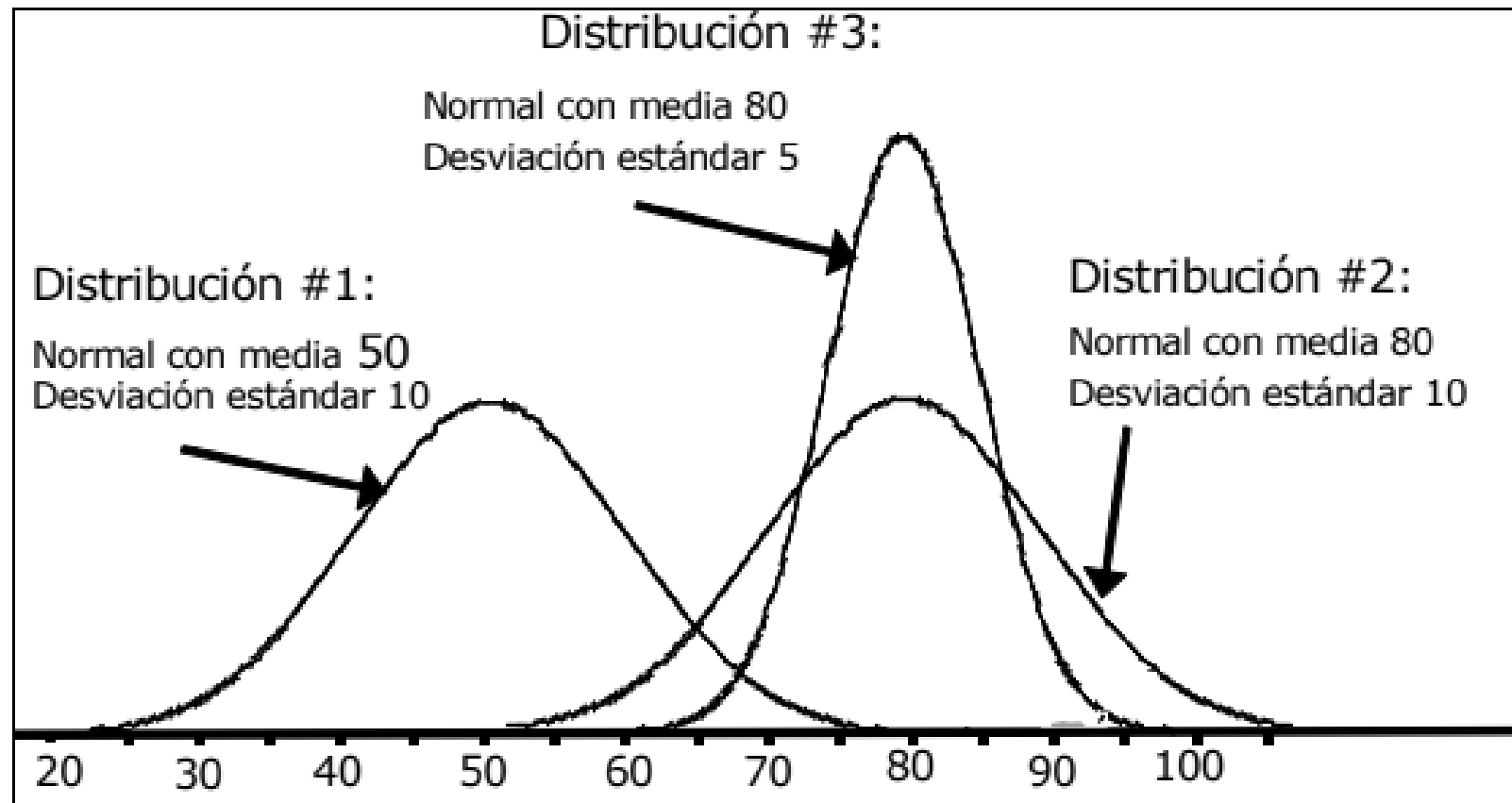
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$



DISTRIBUCION NORMAL

Características:

- Su gráfico semeja una campana simétrica, cuyas colas se extienden hacia el infinito tanto en dirección negativa como en la positiva.
- El promedio, la mediana y la moda de la distribución tienen el mismo valor.
- La distribución queda completamente definida por el **promedio** y la **desviación estándar**. El promedio nos informa sobre la **posición** o ubicación de la distribución en el eje horizontal y la desviación estándar refleja la **dispersión** de los valores con respecto al promedio.

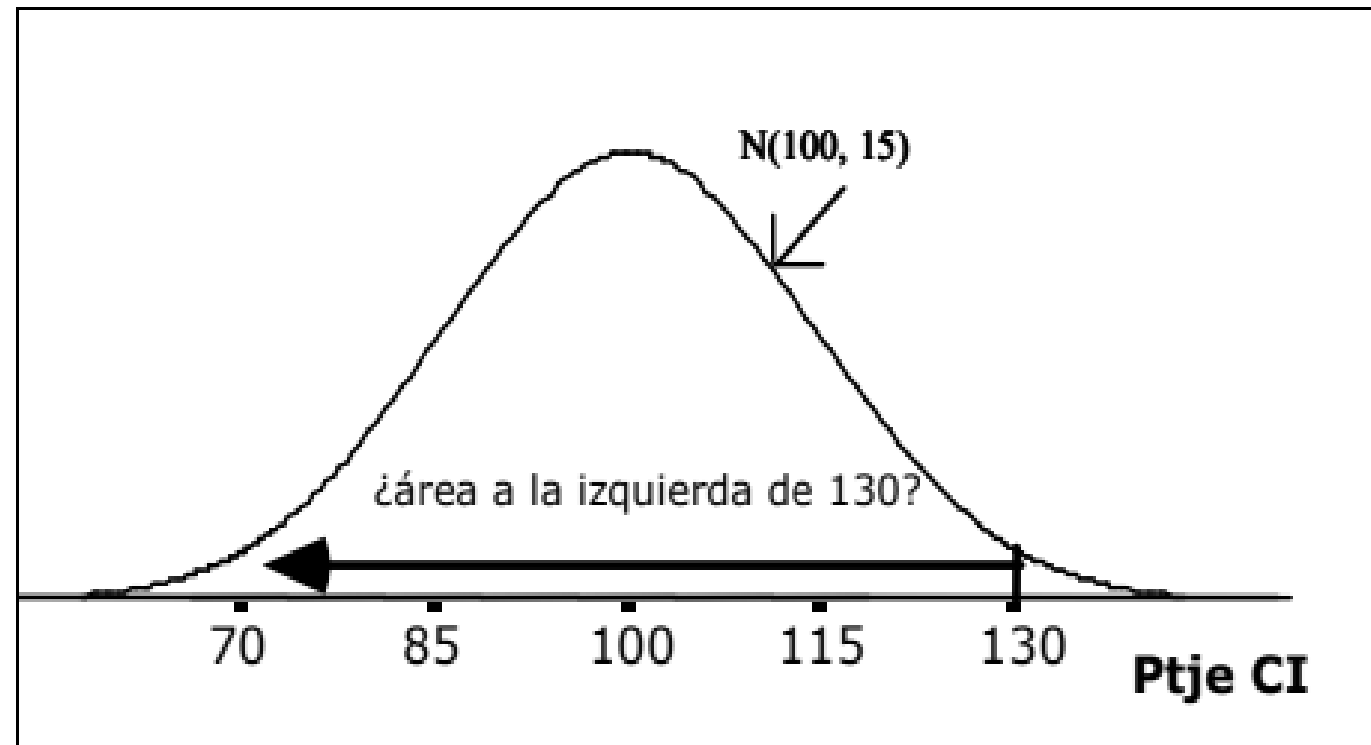


DISTRIBUCION NORMAL

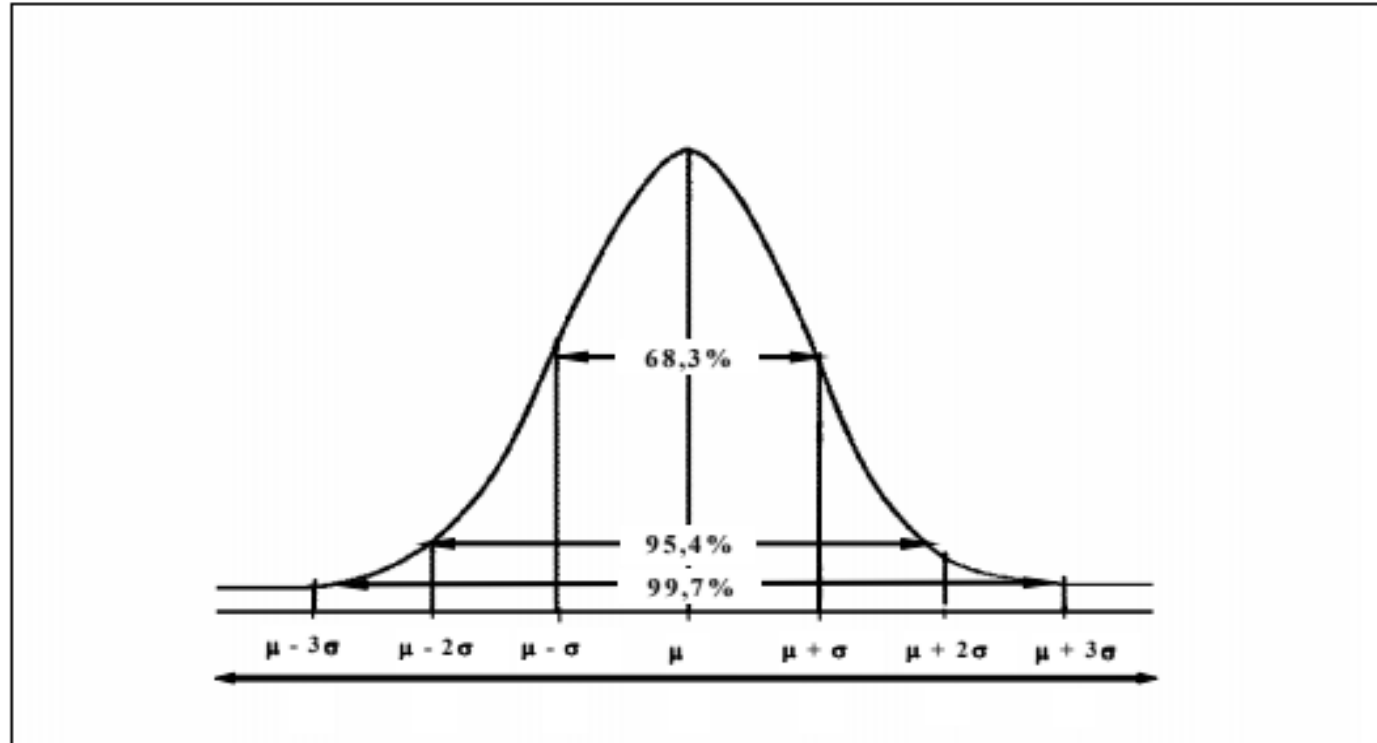
✓ Ejemplo **CI**

Los puntajes del test de inteligencia para adultos WAIS-R (Wechsler Adult Intelligence Scale-Revised) siguen una distribución Normal con media 100 y desviación estándar de 15.

A partir de este modelo podemos hacernos preguntas tales como ¿qué proporción de adultos tendrán un CI menor que 130?



DISTRIBUCION NORMAL



Para cualquier distribución Normal $N(\mu, \sigma)$ se cumple que:

- **68,3%** de las observaciones se encontrarán a **una** desviación estándar de la media, es decir dentro del intervalo:

$$(\mu - \sigma, \mu + \sigma)$$

- **95,4%** de las observaciones se encontrarán a **dos** desviaciones estándar de la media, i.e. dentro del intervalo:

$$(\mu - 2\sigma, \mu + 2\sigma)$$

- **99,7%** de las observaciones se encontrarán a **tres** desviaciones estándar de la media, i.e. dentro del intervalo:

$$(\mu - 3\sigma, \mu + 3\sigma)$$

DISTRIBUCION NORMAL

Estandarización

Si $X \sim \mathcal{N}(\mu, \sigma)$, considero

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Se llama **distribución normal estándar**.

DISTRIBUCION NORMAL

☒ Cálculo de áreas.

- a) Encuentre el área de la distribución Normal estándar que se encuentra a la izquierda de $z = 1,22$.
- b) Encuentre el área de la distribución Normal estándar que se encuentra a la derecha de $z = 1,22$.
- c) Encuentre el área de la distribución Normal estándar que se encuentra entre $z = 0$ y $z = 1,22$.
- d) Encuentre el área de la distribución Normal estándar que se encuentra a la izquierda de $z = -2,55$.
- e) Encuentre el área de la distribución Normal estándar que se encuentra entre $z = -1,22$ y $z = 1,22$.

☒ Puntajes de CI.

Suponga que definimos a X como los puntajes de CI del test de inteligencia WAIS-R, con distribución $N(100, 15)$.

- a) ¿Qué proporción de adultos tendrá un CI menor a 85?
- b) ¿Qué proporción de adultos tendrá un CI mayor a 85?
- c) ¿Qué proporción de adultos tendrá un CI entre 85 y 115?

APROXIMACIONES A LA DISTRIBUCION NORMAL

Binomial

Si $X \sim B(n, p)$ con n suficientemente grande

$$\frac{X - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0, 1)$$

Poisson

Si $X \sim \mathcal{P}(\lambda)$ con λ suficientemente grande

$$\frac{X - \lambda}{\sqrt{\lambda}} \sim \mathcal{N}(0, 1)$$

EJEMPLOS

Ejemplo 1: Sea X una variable aleatoria que representa el número de partes por estar mal estacionado que sacan en una hora en días laborales en Talca. La distribución de probabilidades de X es:

x	1	2	3	4	5
Probabilidad		0,10	0,20	0,30	

- Complete la tabla, suponga que $P(X=1)=P(X=5)$.
- ¿Cuál es el número esperado de partes por estar mal estacionado en Talca?
- ¿Cuál es la desviación estándar de X ?
- ¿Cuál es la probabilidad de que en un día de la semana cualquiera se saquen menos de 3 partes?

Ejemplo 2:

Los estudiantes en general manifiestan que tienen mayor dificultad en los cursos que involucren cálculos matemáticos. Experiencias anteriores han consistido en exponer 5 palabras y 5 números ante los estudiantes durante 10 segundos al comienzo de la clase y luego preguntar por ellos al final de la clase, obteniéndose las siguientes distribuciones de probabilidades:

Cantidad de palabras que recuerdan	0	1	2	3	4	5
P(X=x)	0,05	0,15	0,20	0,25	0,30	0,05

Cantidad de números que recuerdan	0	1	2	3	4	5
P(Y=y)	0,10	0,30	0,20	0,20	0,10	0,10

- a. ¿Qué es más probable que recuerden al menos 3 palabras o al menos 3 números? Justifique.
- b. En término medio (valor esperado), ¿Qué cantidad de palabras y qué cantidad de números recuerdan los estudiantes?
- c. ¿Cuál es la desviación estándar de la cantidad de palabras y de la cantidad de números?

Ejemplo 3:

Suponga que los puntajes en la PSU de matemáticas se distribuyen en forma normal con media 505 y desviación estándar 107. Si 10 mil alumnos rinden la PSU y para postular a la Universidad se exige un mínimo necesario de 400 puntos:

- a. ¿Cuántos alumnos podrán postular a la Universidad?
- b. Si a los alumnos que obtengan puntajes superiores al percentil 90 se les ofrece una beca, ¿Qué puntaje será el mínimo para acceder a la beca?