

Variables Aleatorias

Hugo S. Salinas

Introducción

En Estadística Descriptiva, se estudiaron las distribuciones de frecuencias de conjuntos de datos y posteriormente se vimos los 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)**.

También vimos que los espacios muestrales no necesariamente son numéricos. Por ejemplo cuando lanzamos una moneda tres veces, podremos registrar un resultado como **CCS**. En estadística, sin embargo interesan los resultados numéricos, tal como el número de caras al lanzar una moneda tres veces.

Ejemplo. Se tiene el experimento aleatorio: Lanzar una moneda 3 veces

El espacio muestral que corresponde a este experimento es:

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

Sea $X :=$ número de caras. ¿Qué valores puede tomar X ?

Introducción cont.

Espacio Muestral	Variable aleatoria X
SSS	$X_1 = 0$
CSS, SSC, SCS	$X_2 = 1$
CCS, CSC, SCC	$X_3 = 2$
CCC	$X_4 = 3$

Sea $Y :=$ número de caras menos número de sellos. ¿Qué valores puede tomar Y ?

Sea $Z :=$ número de caras hasta que sale sello. ¿Qué valores puede tomar Z ?

Definición

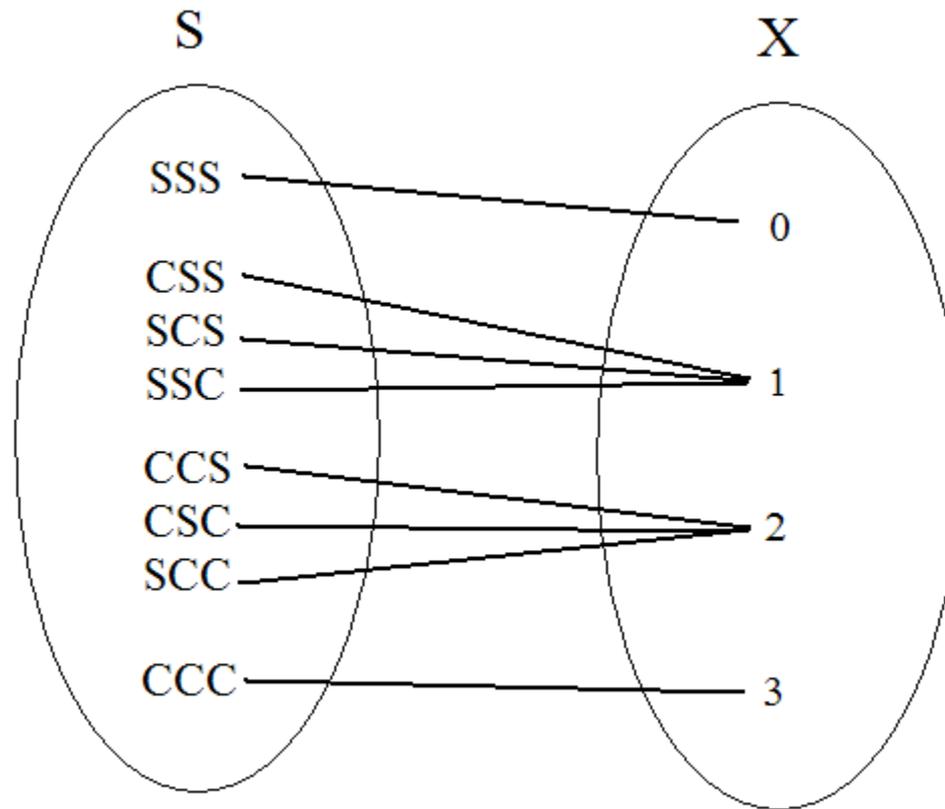
Una **variable aleatoria** es una variable cuyo número que depende del resultado aleatorio de un experimento.

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.

Ejemplo. Supongamos que se aplicará una encuesta a los estudiantes de la UDA donde se preguntará por el número de cursos inscritos este semestre. Identificar la variable aleatoria de interés y enumerar sus valores posibles.

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

Variables aleatorias



Ahora nos interesa aprender cómo asignar probabilidades a eventos, y para eso vamos a distinguir dos tipos de variables aleatorias:

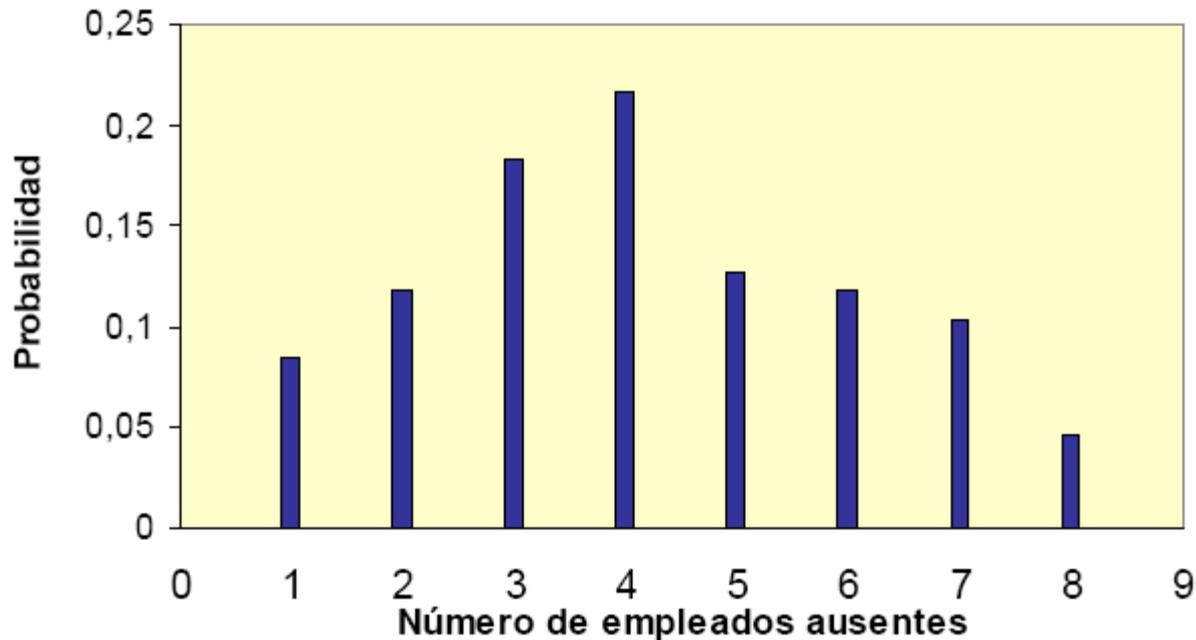
Una variable **aleatoria discreta** puede tomar valores finitos o contables.

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

Función de distribución

Variables aleatorias discretas

$$P(X=x)$$



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:

Variables aleatorias discretas

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

x				
$p(x)$				

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

- Describir la forma de la distribución.
- ¿Qué proporción de estudiantes llevan 3 o menos libros?
- ¿Qué proporción de estudiantes llevan más de 2 libros?
- ¿Qué proporción de estudiantes llevan entre 2,1 y 2,8 libros?
- ¿Qué proporción de estudiantes llevan entre 1 y 2 libros?

Variables aleatorias discretas cont.

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 ($\sum p(x) = 1$).

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

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 y más personas para que esta sea una distribución de probabilidades discreta legítima?
- b) Mostrar 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 $P(2 < X < 4)$?

VARIABLES ALEATORIAS DISCRETAS CONT.

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:

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

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

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

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:

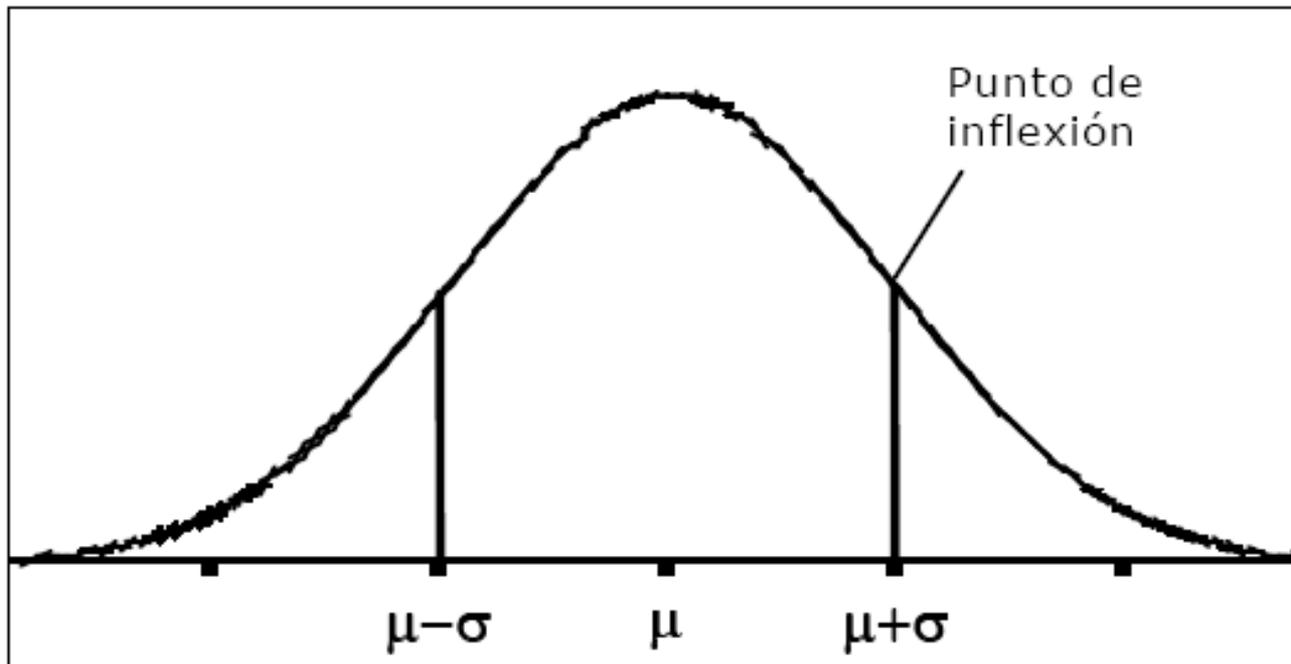
1. $f(x) \geq 0$ La densidad es siempre mayor o igual a cero.
2. El área bajo la curva de densidad es uno.
3. $P(a < X < b) = P(a \leq X \leq b)$ = es el área o proporción entre a y b.

La distribución Normal

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:

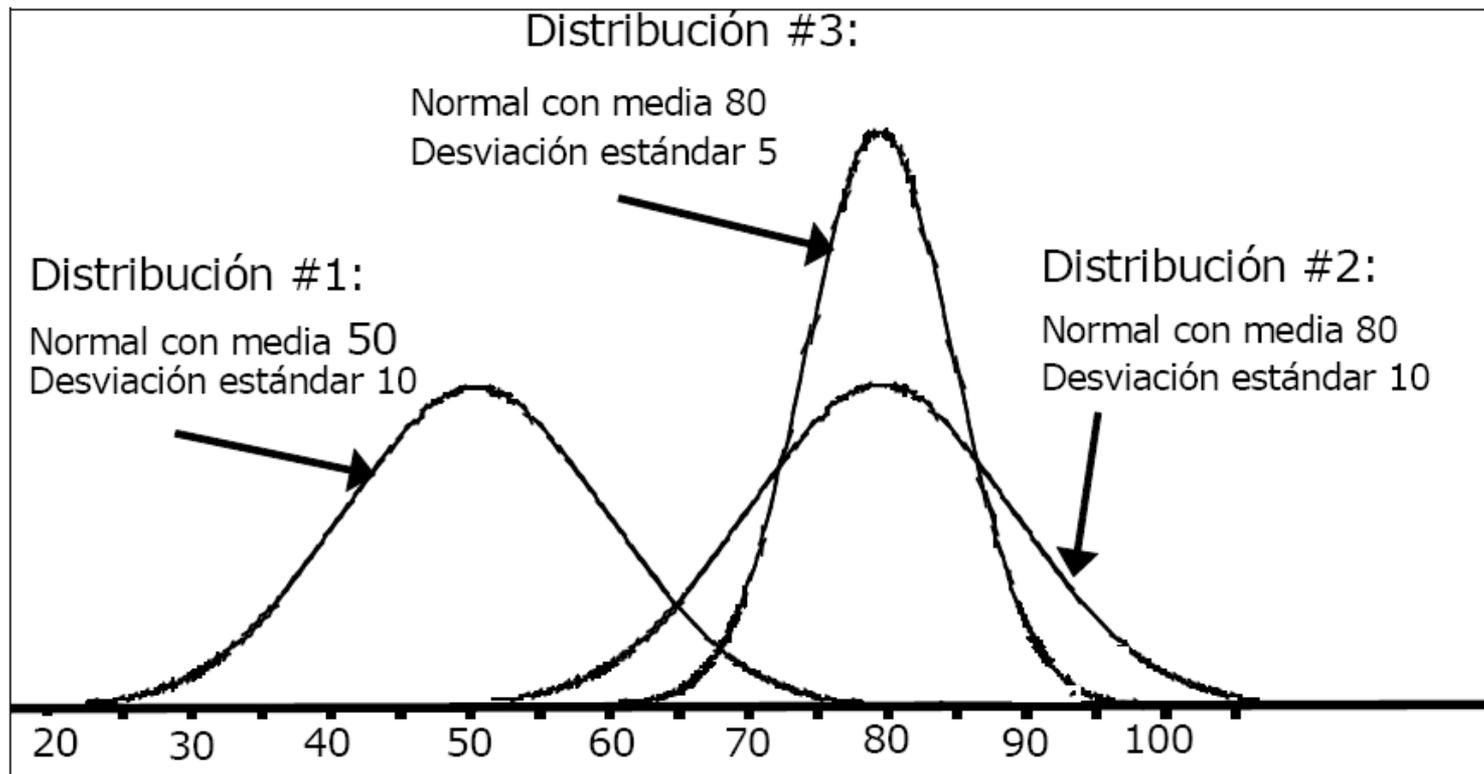
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$



La distribución Normal cont.

Características:

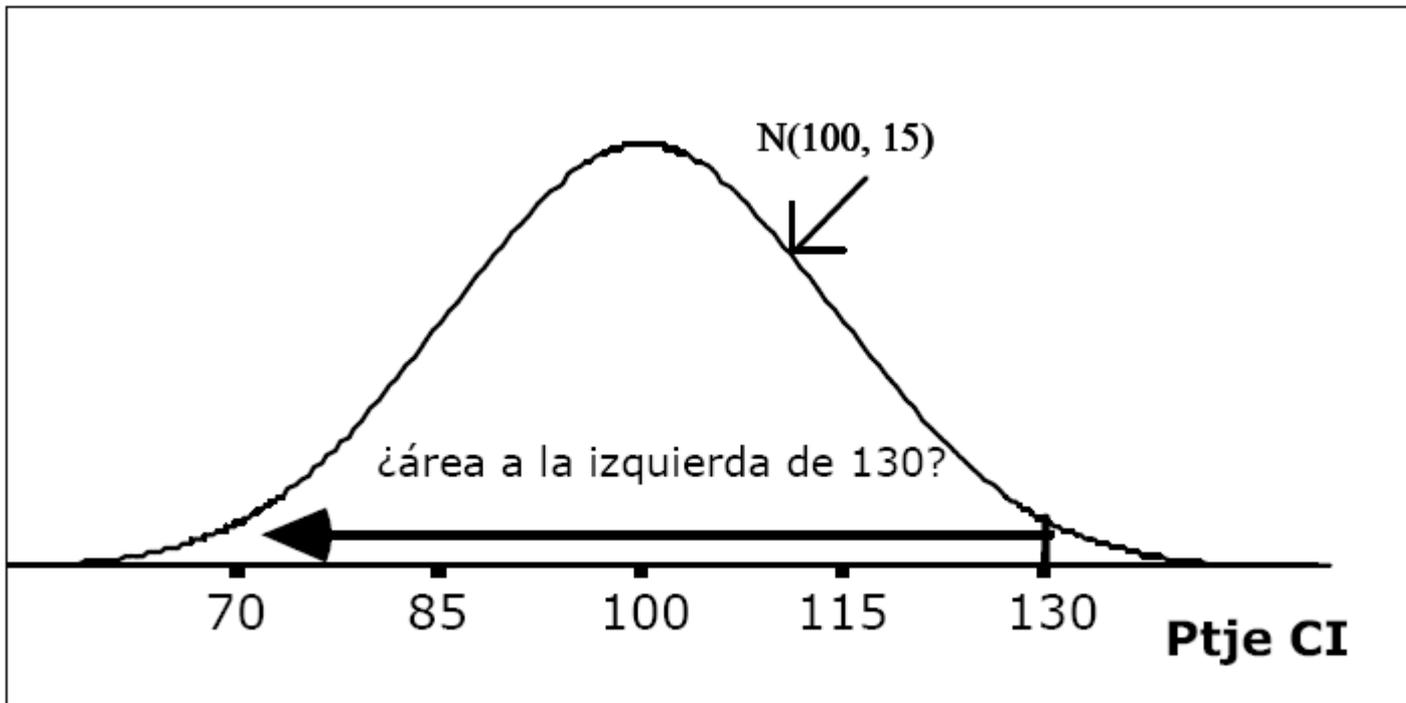
- Su gráfico semeja una campana simétrica, cuyas colas se extienden hacia el infinito tanto en dirección negativa como en la positiva.
- El promedio, la mediana y la moda de la distribución tienen el mismo valor.
- La distribución queda completamente definida por la **media (promedio)** y la **desviación 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.



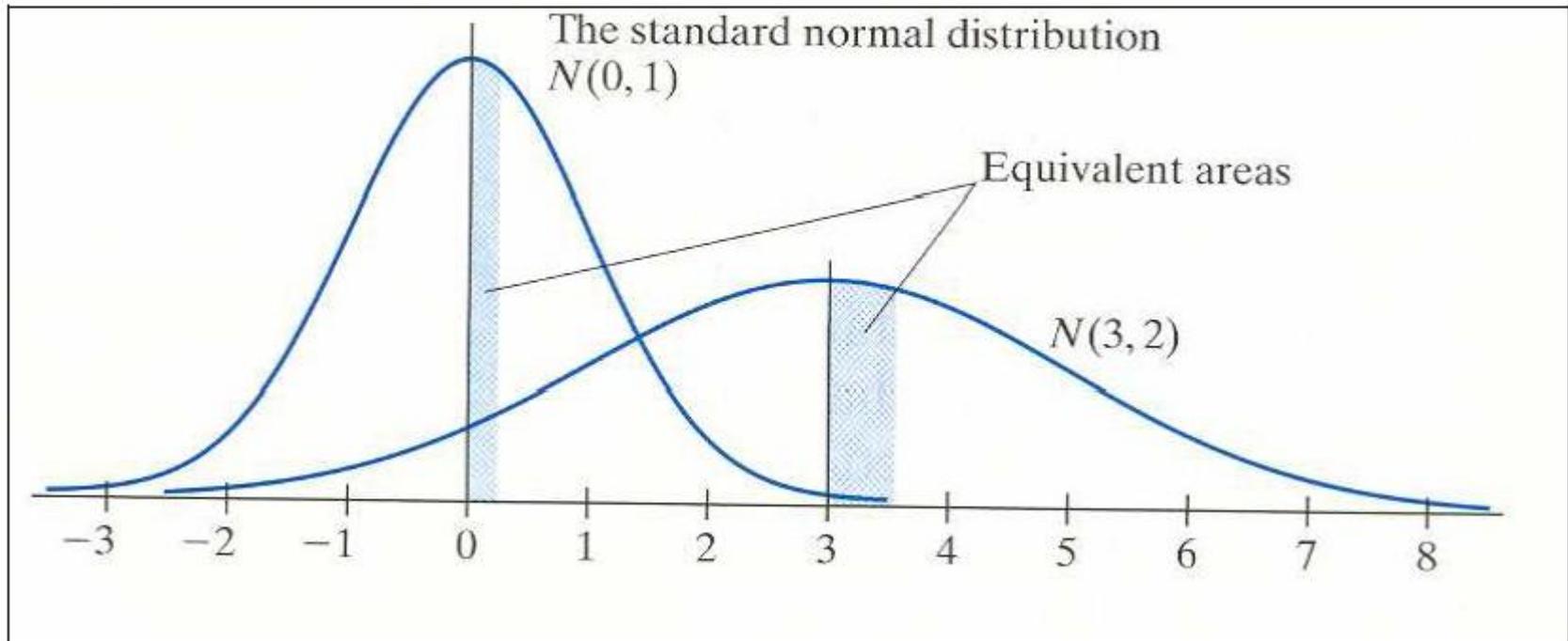
La distribución Normal cont.

Ejemplo . 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?



Cálculo de áreas de una distribución normal



Definición:

Si $X \sim N(\mu, \sigma)$, la variable **normal estandarizada** es: $Z = \frac{X - \mu}{\sigma}$ y tiene distribución Normal con media cero y varianza igual a uno: $Z \sim N(0, 1)$.

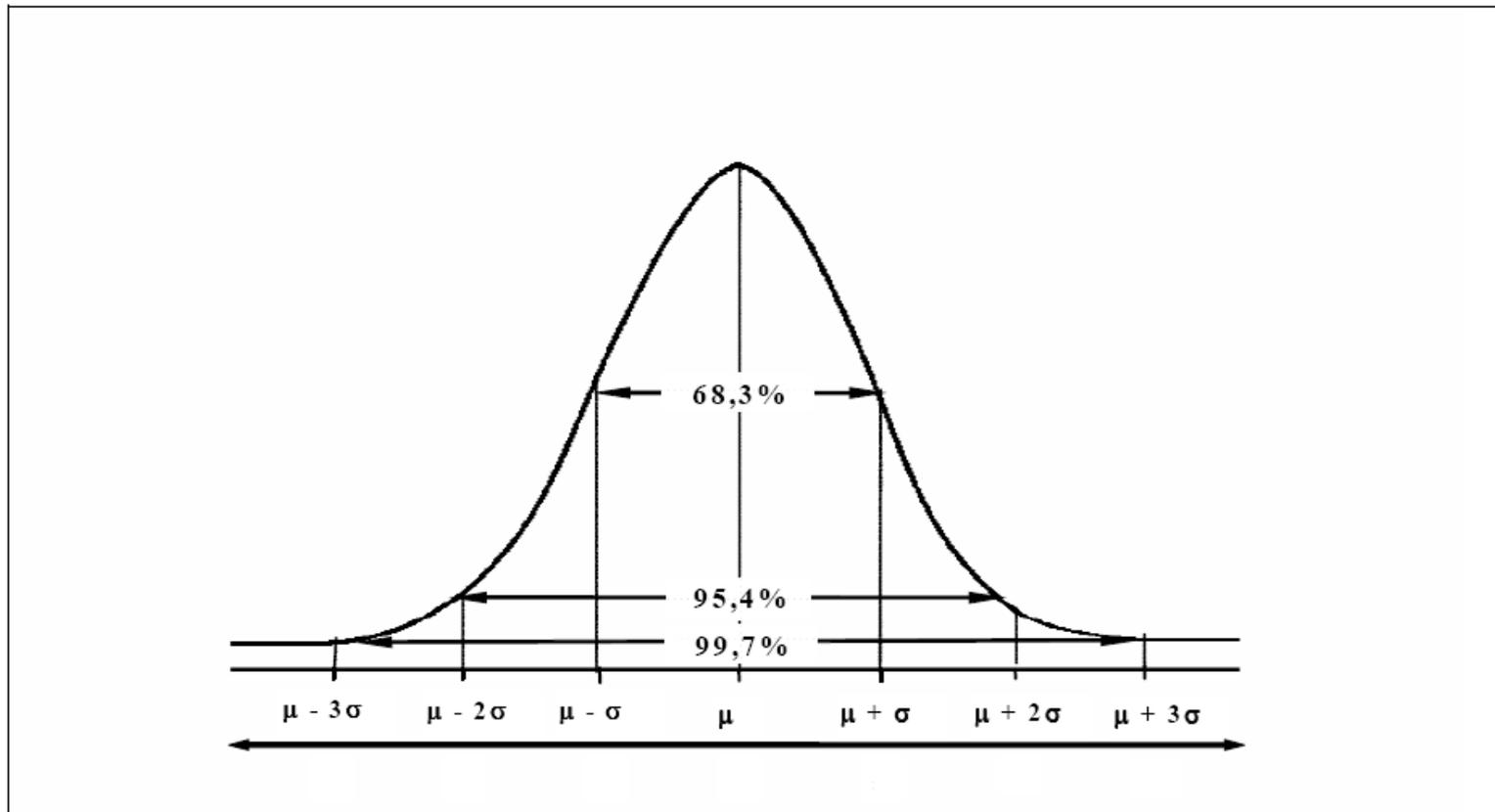
Cálculo de áreas de una distribución normal cont.

Z es el número de desviaciones estándar que x difiere de la media μ :

Si $Z > 0$ entonces x es mayor a la media μ .

Si $Z < 0$ entonces x es menor a la media μ .

Si $Z = 0$ entonces x es igual a la media μ .



Variables aleatorias

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:

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

Aunque teóricamente la distribución llega a $-\infty$ y a $+\infty$, en la práctica es muy raro encontrar valores a más de 3 desviaciones estándar del promedio.

Tabla de la normal

TABLA NORMAL: Valores de la función de distribución acumulativa normal estándar.

$$P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp(-t^2 / 2) dt$$

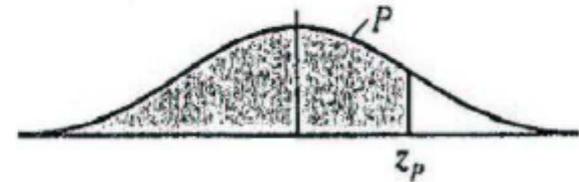


Tabla de la normal

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Tabla de la normal cont.

2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

La normal cont.

Cálculo de áreas.

- Encuentre el área de la distribución Normal estándar que se encuentra a la izquierda de $z = 1,22$.
- Encuentre el área de la distribución Normal estándar que se encuentra a la derecha de $z = 1,22$.
- Encuentre el área de la distribución Normal estándar que se encuentra entre $z = 0$ y $z = 1,22$.
- Encuentre el área de la distribución Normal estándar que se encuentra a la izquierda de $z = -2,55$.
- 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 $N(100, 15)$.

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

Continuando con el modelo $N(100, 15)$ para el puntaje de CI para adultos, considere la siguiente pregunta: ¿Qué puntaje de CI debe tener un adulto para ubicarse entre el 1% con más alto puntaje?

La normal cont.

Natación. El tiempo que demoran los nadadores de 100 metros mariposa sigue una normal con media 55 segundos y desviación estándar de 5 segundos.

- a) Los organizadores de un campeonato deciden dar certificados a todos los nadadores que terminen antes de 49 segundos. Si hay 50 nadadores en los 100 metros mariposa, ¿Cuántos certificados se necesitarán?
- b) ¿Con qué tiempo debe terminar un nadador para estar entre el 2% más rápido de la distribución de tiempos?

Ejemplo. Sea X es $N(3, 2)$:

- a) Mostrar gráficamente esta distribución en particular
- b) Encontrar el rango entre cuartiles de la distribución
- c) Encontrar $P(X > 3)$.
- d) Encontrar $P(X = 3)$.

La normal cont.

Se cree que la altura de los **pinos** en un bosque tiene distribución Normal. Queremos docimar las siguientes hipótesis:

$$H_0 : X \sim N(15, 3)$$

$$H_1 : X \sim N(10, 3)$$

Las alturas de los pinos son medidas en metros. Decidiremos rechazar la hipótesis nula si la altura de un pino seleccionado al azar del bosque mide menos de 8 metros.

- Calcule la probabilidad del Error Tipo I, α .
- Calcule la probabilidad del Error Tipo II, β .
- Calcule el valor-p si la altura del pino seleccionado fue de 8,5 metros.