

Clase 01: Modelos de Probabilidad¹

Hugo S. Salinas

Universidad de Atacama

hugo.salinas@uda.cl

¹José R. Berrendero de la Universidad Autónoma de Madrid

Estructura de este tema

- ▶ Conceptos básicos de probabilidad.
- ▶ Modelos discretos: la distribución binomial y la distribución de Poisson.
- ▶ Modelos continuos: la distribución exponencial y la distribución normal.
- ▶ Estimación de una media en poblaciones normales.

Parámetro y estimador

Un **parámetro** es un número que describe alguna característica de interés de una **población**. En la práctica, siempre tiene un valor desconocido.

Algunos parámetros que nos van a interesar en este curso son:

- ▶ La *media poblacional* (μ) de una variable.
- ▶ La *varianza poblacional* (σ^2) de una variable.
- ▶ La *proporción poblacional* (p) de individuos que presentan cierta característica.

La población no es conocida en su totalidad, pero suponemos que se dispone de una **muestra** x_1, \dots, x_n .

Un **estimador** es una cantidad que se puede calcular con los datos muestrales y que aproxima el valor de un parámetro de interés.

Ejemplos

Ejemplo 1: Se seleccionan aleatoriamente 200 personas de una ciudad y se les pregunta si han seguido alguna dieta en los últimos cinco años. De las personas seleccionadas 20 responden afirmativamente.

Ejemplo 2: Se seleccionan aleatoriamente 200 personas de una ciudad y se mide su índice de masa corporal (IMC). La media de los IMC medidos es de 22.3.

Determina en los ejemplos anteriores un parámetro poblacional de interés y su correspondiente estimador.

Para determinar la calidad de una estimación se utilizan conceptos de probabilidad.

Variables aleatorias

Una **variable aleatoria** (v.a.) representa numéricamente el resultado de un experimento aleatorio.

En el ejemplo 1, el experimento consiste en seleccionar a una persona aleatoriamente y preguntarle si ha seguido o no una dieta. Una variable aleatoria que representa el resultado es:

$$X = \begin{cases} 1, & \text{si la respuesta es afirmativa;} \\ 0, & \text{si la respuesta es negativa.} \end{cases}$$

Espacio muestral y sucesos

El **espacio muestral**, Ω , es el conjunto de todos los posibles resultados finales de un experimento aleatorio.

Ejemplo: Sea X el número de personas que afirman haber seguido una dieta en los últimos 5 años en una encuesta realizada a 200 personas. ¿Cuál es el espacio muestral correspondiente?

Un **suceso** A es un subconjunto del espacio muestral Ω , $A \subset \Omega$.

Ejemplo: Sea X el resultado obtenido al lanzar un dado ($\Omega = \{1, 2, 3, 4, 5, 6\}$).

- ▶ $P = \{2, 4, 6\}$ es un suceso (sacar un número par).
- ▶ $S_4 = \{4\}$ es un suceso (sacar 4).
- ▶ $I = \{1, 3, 5\}$ es un suceso (sacar un número impar).
- ▶ $\Omega = \{1, 2, 3, 4, 5, 6\}$ es un suceso (suceso seguro).
- ▶ \emptyset es un suceso (suceso imposible).

Probabilidad de un suceso

La **probabilidad** es una función P que, a cada suceso A , le hace corresponder un número $P(A)$ entre 0 y 1 y que refleja el grado de seguridad con el que el suceso ocurre.

En muchas ocasiones, los sucesos presentan un comportamiento regular a largo plazo. Por ejemplo, al tirar muchas veces una moneda, el porcentaje de veces que sale cara se aproximará al 50%

En estos casos, la probabilidad de un suceso puede interpretarse como el valor al que converge la frecuencia relativa de veces que ocurre ese suceso al aumentar el número de veces que se repite el experimento.

Principales propiedades de la probabilidad

- ▶ $P(\Omega) = 1$

- ▶ La probabilidad de que un suceso no ocurra es 1 menos la probabilidad de que ocurra:

$$P(A^c) = 1 - P(A)$$

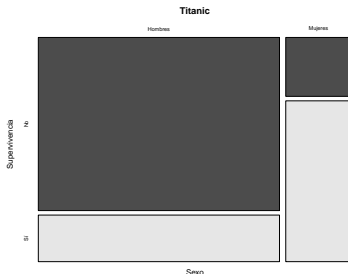
- ▶ Si A y B son dos sucesos cualesquiera (no necesariamente incompatibles), entonces

$$P(A \text{ ó } B) = P(A) + P(B) - P(A \text{ y } B)$$

Probabilidad condicionada

La siguiente tabla muestra los 2201 individuos que viajaban en el Titanic clasificados por sexo y según sobrevivieron o no:

	Hombres	Mujeres
No	1364	126
Si	367	344



Al seleccionar aleatoriamente a un individuo de los que viajaban en el Titanic, sea H el suceso *el individuo es un hombre* y S el suceso *el individuo sobrevive*. Calcula las probabilidades siguientes:

- ▶ $P(S)$
- ▶ $P(S \text{ y } H)$
- ▶ $P(S \mid H)$ y $P(S \mid H^c)$
- ▶ $P(H \mid S)$

Probabilidad condicionada

En general, la probabilidad de un suceso A condicionado a otro B (tal que $P(B) > 0$) se define

$$P(A \mid B) = \frac{P(A \text{ y } B)}{P(B)}$$

Sensibilidad y especificidad

La prevalencia del virus HIV es del 5% en una población. Se ha diseñado una prueba para detectar la presencia del virus. La prueba tiene una **sensibilidad** (porcentaje de positivos entre los enfermos) del 95% y una **especificidad** (porcentaje de negativos entre los sanos) del 99%

Definimos los sucesos:

- ▶ A , la prueba da positivo ($B = A^c$, da negativo).
- ▶ E , el individuo está enfermo ($S = E^c$ está sano).

Probabilidad condicionada

$$P(E) = 0.05$$

$$\text{Sensibilidad} = P(A \mid E) = \frac{P(A \text{ y } E)}{P(E)} = 0.95$$

$$\text{Especificidad} = P(B \mid S) = \frac{P(B \text{ y } S)}{P(S)} = 0.99$$

- ¿Cuál es la probabilidad de que al seleccionar aleatoriamente un individuo de la población y hacerle la prueba, el resultado sea positivo?
- Añade los porcentajes adecuados en cada casilla:

	A	B	
S			95
E			5
			100

Independencia de sucesos

Sabemos que

$$P(A \text{ y } B) = P(A | B)P(B) = P(B | A)P(A)$$

Definición: A y B son **independientes** si

$$P(A \text{ y } B) = P(A) \cdot P(B)$$

Intuitivamente, dos sucesos son independientes si el conocimiento de que ha ocurrido uno de ellos no modifica la probabilidad de que ocurra el otro, $P(A | B) = P(A)$ y $P(B | A) = P(B)$.

Ejemplo: Calcula la probabilidad de que al tirar un dado 3 veces, se obtenga al menos una vez un seis.

Variables aleatorias discretas

Una v.a. es **discreta** si toma un número finito (o una sucesión) de valores.

Su distribución viene dada por una función que asigna a cada valor su probabilidad:

Valores	x_1	\cdots	x_k
Probabilidades	p_1	\cdots	p_k

Ejemplos:

- ▶ ¿Cuál es la distribución de la v.a. que representa el resultado de tirar un dado?
- ▶ Si X es la v.a. del ejemplo 1, entonces su distribución es:

Valores	0	1
Probabilidades		

Media y varianza de una v.a. discreta

Sea X una v.a. discreta con distribución:

Valores	x_1	\cdots	x_k
Probabilidades	p_1	\cdots	p_k

La **media o esperanza** de X es:

$$\mu = E(X) = p_1 x_1 + \cdots + p_k x_k$$

La **varianza** de X es:

$$\sigma^2 = \text{Var}(X) = p_1(x_1 - \mu)^2 + \cdots + p_k(x_k - \mu)^2,$$

donde $\mu = E(X)$.

La **desviación típica** de X es $\sigma = \sqrt{\text{Var}(X)}$

Ejemplos: calcula la media y la varianza de X

- (a) X es la v.a. que representa el resultado de tirar un dado.
(b) X tiene distribución

Valores	0	1
Probabilidades	0.25	0.75

- (c) X tiene distribución

Valores	0	1
Probabilidades	$1 - p$	p

- (d) Sea X el número de aleteos por segundo de una cierta especie de mariposa cuando vuela. Su distribución es:

Valores	6	7	8	9	10
Probabilidades	0.05	0.1	0.6	0.15	

Calcula $P(X \geq 8)$, $E(X)$ y $\text{Var}(X)$.

La distribución de Bernoulli

Una prueba de Bernoulli consiste en un experimento aleatorio con dos posibles resultados: éxito y fracaso.

Una v.a. de Bernoulli (con parámetro p) es aquella que toma el valor 1 (éxito) con probabilidad p y el valor 0 (fracaso) con probabilidad $1 - p$.

Siempre que examinamos a n individuos de una población para ver si presentan o no cierta característica tenemos una muestra x_1, \dots, x_n de variables de Bernoulli.

La media y la varianza son las que se han calculado en el ejemplo anterior.

La distribución binomial

Realizamos n pruebas de Bernoulli independientes tales que la probabilidad de éxito es p .

Consideramos la v.a. que corresponde al número de éxitos obtenidos en las n pruebas. Se dice que X tiene **distribución binomial** de parámetros n y p :

$$X \equiv B(n, p)$$

- ▶ ¿Qué valores puede tomar una v.a. $B(10, 0.2)$? ¿Y una v.a. $B(10, 0.5)$?
- ▶ En el primer caso, ¿cuánto valen $P(X = 10)$, $P(X = 0)$ y $P(X = 1)$?
- ▶ Intuitivamente, ¿cuánto crees que vale $E(X)$ en los dos casos anteriores?
- ▶ Responde a las mismas preguntas en el caso general $X \equiv B(n, p)$.

La distribución binomial

El porcentaje de individuos de una población cuya sangre es del grupo B es del 10%. Se seleccionan aleatoria e independientemente cuatro individuos de esta población.

Determina las probabilidades de los siguientes sucesos:

- ▶ La sangre de los cuatro es del grupo B.
- ▶ Ninguno de los cuatro tiene sangre del grupo B.
- ▶ Exactamente dos de ellos tienen sangre del grupo B.

En la misma población anterior se selecciona una muestra de 234 individuos. ¿Cuál es el número esperado de ellos cuya sangre pertenece al grupo B?

La distribución de Poisson

La v.a. X sigue una **distribución de Poisson** de parámetro λ ($\lambda > 0$), y se denota $X \equiv \mathcal{P}(\lambda)$, si

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{para } k = 0, 1, 2, \dots$$

Se puede comprobar que $E(X) = \lambda = V(X)$.

La distribución de Poisson resulta útil en muchos procesos en los que ocurren determinados sucesos por unidad de tiempo o espacio:

- El número de plaquetas en un ml. de sangre.
- El número de mutaciones en un fragmento de ADN después de una cierta cantidad de radiación.
- Número de goles en un partido de fútbol.

La distribución de Poisson

La distribución de Poisson aparece como límite de la binomial bajo ciertas condiciones:

- ▶ Dividimos el intervalo en un gran número n de pequeños intervalos.
- ▶ En cada pequeño intervalo la probabilidad de que ocurra dos o más veces el suceso se puede despreciar.
- ▶ La ocurrencia o no del suceso en cada pequeño intervalo es independiente de lo que ocurra en el resto de intervalos.
- ▶ La probabilidad de que el suceso ocurra en cada pequeño intervalo es p , donde si $n \rightarrow \infty$ se verifica $p \rightarrow 0$
- ▶ Además, $np \rightarrow \lambda$ para cierto valor $\lambda > 0$. Es decir, el número medio de veces que ocurre el suceso en el intervalo grande se estabiliza en torno a λ .

Relación entre la distribución de Poisson y la binomial

Cuando $n \rightarrow \infty$, $p \rightarrow 0$ y $np \rightarrow \lambda$,

$$\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

Esto significa que cuando n es grande y p es pequeña (muchos intentos con poca probabilidad de éxito en cada uno), podemos calcular las probabilidades relativas al número de éxitos usando la fórmula de Poisson en lugar de la binomial, ya que:

$$B(n, p) \approx \mathcal{P}(\lambda = np)$$

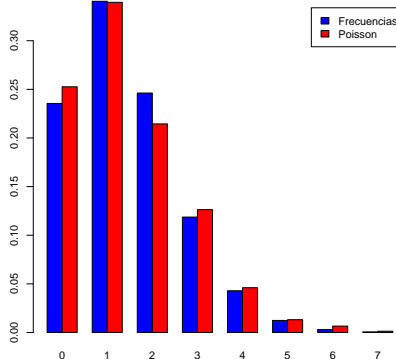
Se igualan los valores esperados de ambas distribuciones ($\lambda = np$)

La distribución de Poisson

Datos de la liga 2008-09

Media de goles por partido y equipo: 1,446

Goles	0	1	2	3	4	5	6	7
Frecuencias	192	258	163	96	35	10	5	1
Poisson	178.98	258.81	187.13	90.2	32.61	9.43	2.27	0.47



Variables aleatorias continuas

Una v.a. es **continua** si puede tomar cualquier valor en un intervalo.

La distribución de una v.a. continua X está determinada por una **función de densidad** f .

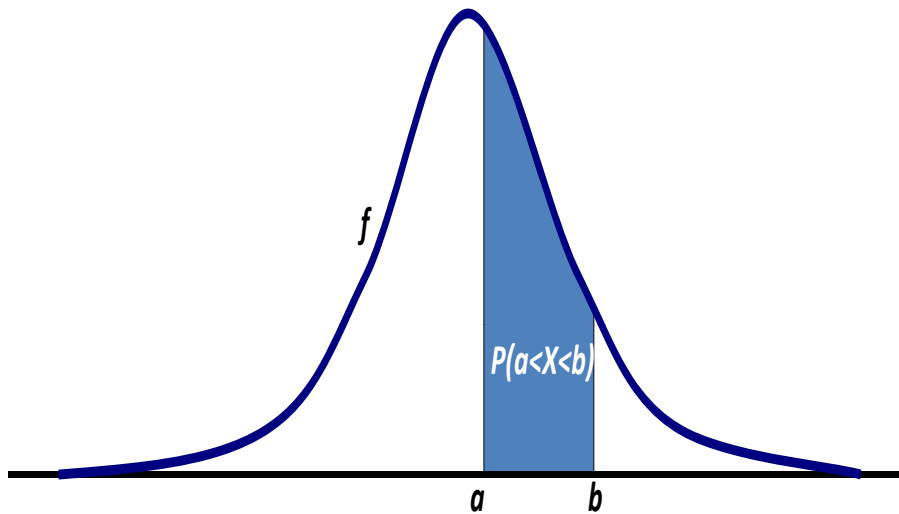
La probabilidad de un intervalo (a, b) es el área entre a y b bajo la función de densidad, es decir,

$$P(a < X < b) = \int_a^b f(x) dx$$

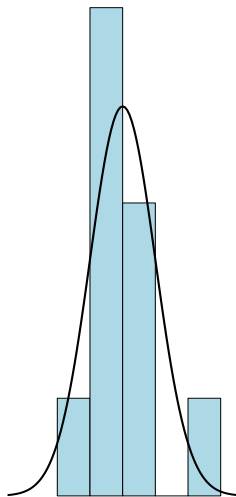
La función de densidad debe cumplir:

- ▶ $f(x) \geq 0$ para todo x .
- ▶ $\int_{-\infty}^{\infty} f(x) dx = 1$.

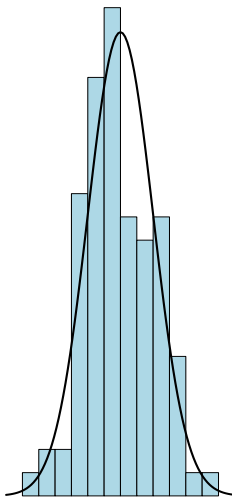
Densidad y probabilidad de un suceso



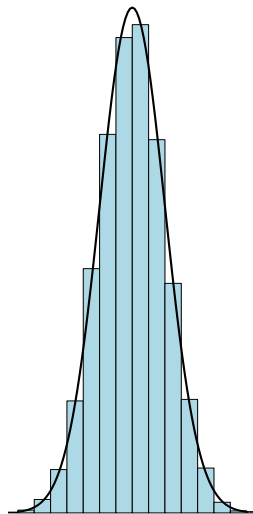
La densidad como límite de histogramas



$n=10$

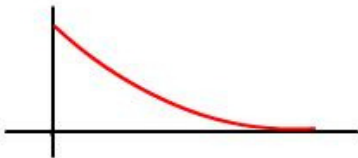
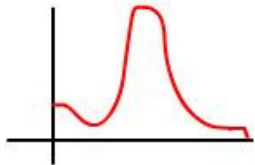
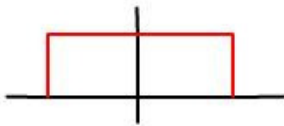
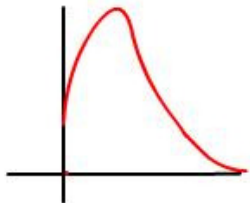
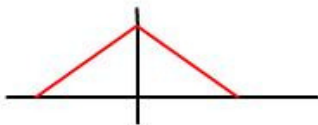
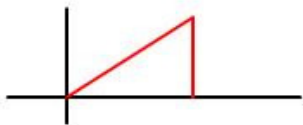


$n=100$



$n=10000$

Algunas funciones de densidad



Cuestiones

De las seis funciones de densidad para una v.a. X , indica cuáles verifican cada una de las condiciones siguientes:

- ▶ X sólo toma valores positivos.
- ▶ X puede tomar cualquier valor positivo.
- ▶ La probabilidad de que X tome valores en (a, b) es la misma que la de que tome valores en $(-b, -a)$.
- ▶ La v.a. X es un modelo adecuado para el tiempo de vida humana.

Ejemplo

La proporción X de un cierto aditivo en la gasolina es una v.a. con función de densidad

$$f(x) = \begin{cases} 2x & \text{si } 0 \leq x \leq 1 \\ 0 & \text{si } x \notin [0, 1]. \end{cases}$$

Dibuja la densidad y calcula las siguientes probabilidades:

► $P(0 < X < 0.5)$

► $P(X > 3)$

► $P(X < 0.75)$

Media y varianza de una v.a. continua

Las v.a. continuas también tienen su media, varianza y desviación típica. Se definen mediante integrales:

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

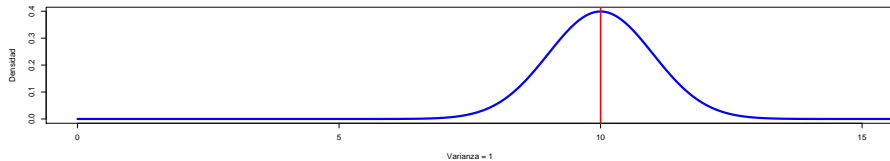
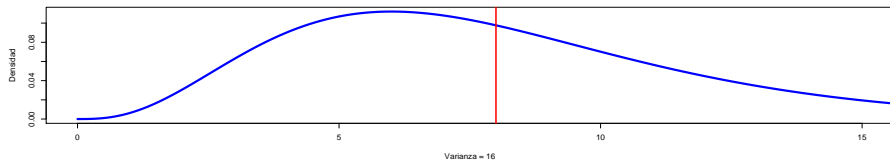
$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx.$$

Puede comprobarse que:

$$\text{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Ejercicio: Calcula la esperanza y la varianza de X , la variable que describe el aditivo en la gasolina.

Densidades, media y varianza



Distribución exponencial

Una v.a. X tiene **distribución exponencial** de parámetro $\lambda > 0$, $X \equiv \exp(\lambda)$, si su función de densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

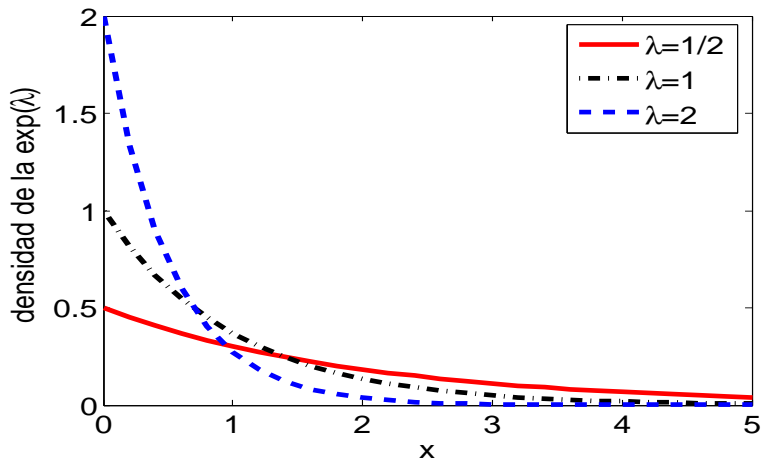
$$P(X \leq t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}.$$

Su media y varianza son

$$\mu = \frac{1}{\lambda} \quad y \quad \sigma^2 = \frac{1}{\lambda^2}.$$

Esta distribución se suele utilizar para representar el tiempo de supervivencia o duración de un sistema biológico o mecánico.

Distribución exponencial



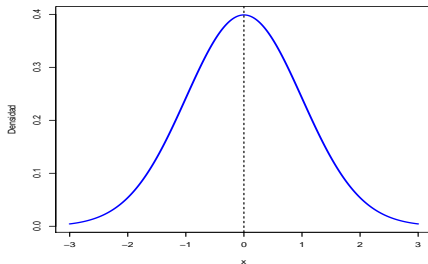
Distribución exponencial

Se ha comprobado que el tiempo de vida de cierto tipo de marcapasos sigue una distribución exponencial con media de 16 años.

- ▶ ¿Cuál es la probabilidad de que a una persona a la que se le ha implantado este marcapasos se le deba reimplantar otro antes de 20 años?
- ▶ ¿Cuál es la probabilidad de que el marcapasos que le han implantado le dure más de 10 años?
- ▶ Determina el tiempo T tal que la probabilidad de que el marcapasos dure más que T es igual a la probabilidad de que dure menos que T .
- ▶ Compara el valor obtenido en el apartado anterior con la duración media.

Distribución normal

Muchos histogramas tienen la siguiente forma aproximada:

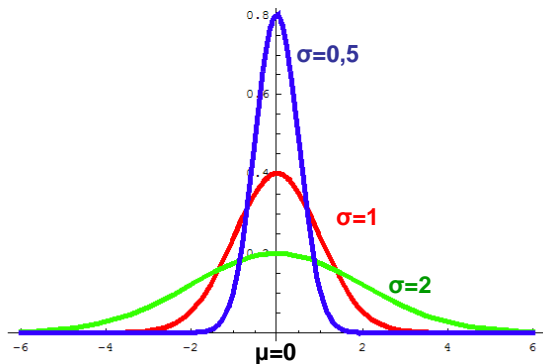


- ▶ Simétrica alrededor de un valor central μ .
- ▶ A medida que los valores se alejan del centro las frecuencias disminuyen rápidamente.
- ▶ La dispersión viene dada por la desviación típica poblacional σ . Los puntos de inflexión se sitúan en los valores $\mu - \sigma$ y $\mu + \sigma$.

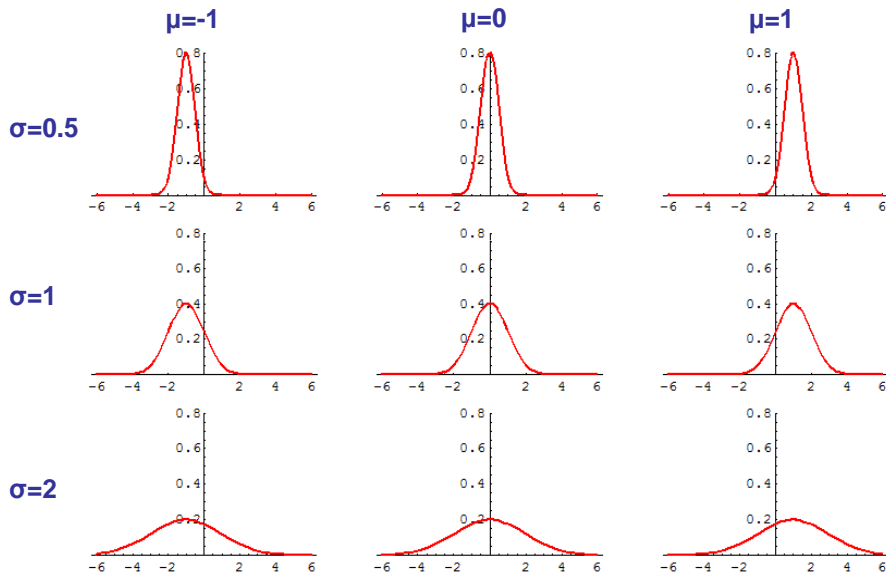
Distribución normal: definición

La v.a. continua X sigue una **distribución normal** $N(\mu, \sigma)$ de parámetros μ y σ ($\sigma > 0$), si su densidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right]$$



La curva de densidad normal según μ y σ



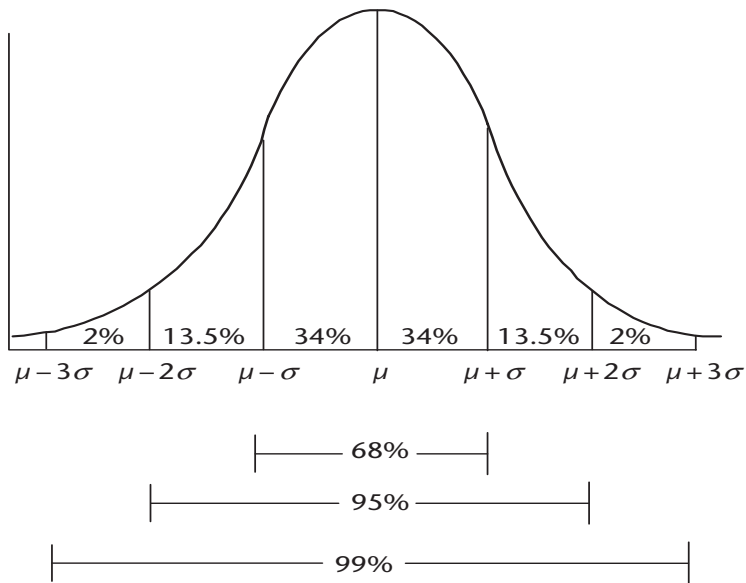
Propiedades importantes de una población normal

(1) Regla 68-95-99

En una población con distribución $N(\mu, \sigma)$:

- ▶ Aproximadamente el 68% de los datos está entre $\mu - \sigma$ y $\mu + \sigma$.
- ▶ Aproximadamente el 95% de los datos está entre $\mu - 2\sigma$ y $\mu + 2\sigma$.
- ▶ Más del 99% de los datos está entre $\mu - 3\sigma$ y $\mu + 3\sigma$.

Propiedades importantes de una población normal



Ejemplo

Sea X la v.a. que representa la cantidad diaria de kcal que toma una persona elegida al azar en una población. Se sabe que la población es normal con media $\mu = 2500$ kcal y desviación típica $\sigma = 100$ kcal. Usando las propiedades anteriores da respuestas aproximadas a las preguntas siguientes:

- ▶ ¿Cuál es la probabilidad de que X esté entre 2300 y 2700 kcal?
- ▶ ¿Cuál es la probabilidad de que X sea mayor que 2700 kcal?
- ▶ ¿Cuál es la probabilidad de que X sea mayor que 2500 kcal?
- ▶ ¿Cuál es la probabilidad de que X sea mayor que 2300 kcal?

Propiedades importantes de una población normal

(2) Estandarización

Si una v.a. X tiene distribución $N(\mu, \sigma)$, entonces la variable estandarizada

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

tiene distribución $N(0, 1)$ (normal estándar).

Como consecuencia, solo necesitamos tablas para la normal estándar.

Tablas de la distribución normal estándar

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

Ejercicios para usar las tablas

Z una v.a. normal estándar. Mira en las tablas para hacer los ejercicios siguientes:

$P(Z > 1)$	$P(Z > c) = 0.025$
$P(Z < -1)$	$z_{0.05}$
$P(-1 < Z < 1)$	$z_{0.95}$
$P(-2 < Z < 1)$	$z_{0.1}$

X es una v.a. normal con $\mu = 1$ y $\sigma = 2$. Mira en las tablas para hacer los ejercicios siguientes:

$P(X > 3)$	$P(X > c) = 0.96$
------------	-------------------

Propiedades importantes de una población normal

(3) Producto de una normal por una constante

Si $a \in \mathbb{R}$ y $X \equiv N(\mu, \sigma)$, entonces:

$$aX \equiv N(a\mu, |a|\sigma)$$

Ejemplo: Si $X \equiv N(5, 1)$, determina la distribución de

- (a) $-X$
- (b) $10X$
- (c) $-10X$
- (d) $X/2$

Propiedades importantes de una población normal

(4) Suma de v.a. normales independientes

Si $X_1 \equiv N(\mu_1, \sigma_1)$ y $X_2 \equiv N(\mu_2, \sigma_2)$ independientes, entonces

$$X_1 + X_2 \equiv N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

Ejemplo: Si $X_1 \equiv N(5, \sigma = 1)$ y $X_2 \equiv N(5, \sigma = 1)$ y ambas variables son independientes, determina la distribución de

- (a) $X_1 + X_2$
- (b) $X_1 - X_2$
- (c) $2X_1 + X_2$
- (d) $(X_1 + X_2)/2$

Estimación de la media poblacional

Para estimar la media de una población, μ , el estimador más natural es la media muestral \bar{x} .

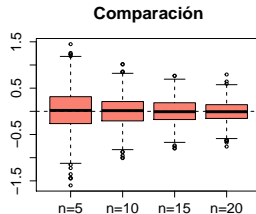
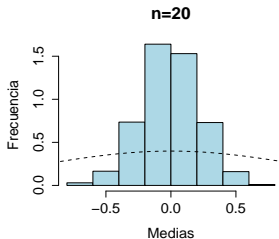
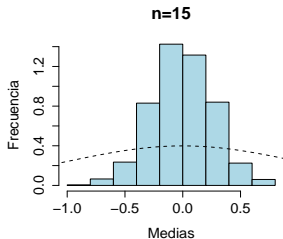
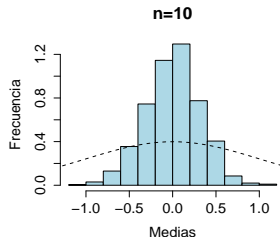
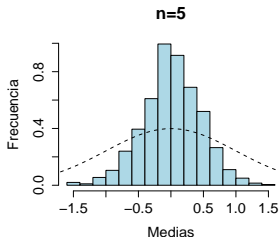
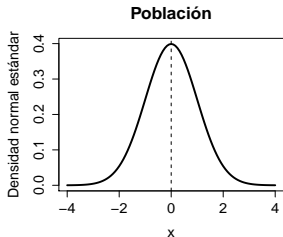
¿Cuál es la calidad de la estimación?

Un estimador es una variable aleatoria ya que su valor depende de la muestra concreta de la que se dispone y la selección de la muestra es aleatoria.

La precisión de la estimación se mide analizando lo que ocurriría si dispusiéramos de muchas muestras y pudiéramos evaluar la media para cada una de ellas.

Tenemos que estudiar la distribución de \bar{x} .

Distribución de la media muestral en una población normal



Distribución de la media muestral en una población normal

Si x_1, \dots, x_n son datos independientes procedentes de una población normal de media μ y desviación típica σ ,

$$\bar{x} \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Observaciones:

- ▶ Para cualquier n , el valor esperado de \bar{x} coincide con la media de la población.
- ▶ Al aumentar n , la desviación típica disminuye y la probabilidad de obtener valores de \bar{x} cercanos a μ aumenta.

Ejemplos

- Un laboratorio pesa el filtro de una mina de carbón para medir la cantidad de polvo ambiental en la mina. Debido a imprecisiones en los aparatos, las medidas tienen distribución normal con media el verdadero peso $\mu = 123$ mg y desviación típica $\sigma = 0.08$ mg.

(a) Se calcula la media de 3 medidas realizadas con el filtro:

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3}.$$

¿Cuál es la distribución de \bar{x} ?

(b) ¿Cuál es la probabilidad de que \bar{x} diste de μ menos de 0.05 mg?

(c) ¿Cuál es la probabilidad de que una única medida del filtro diste de μ menos de 0.05 mg?

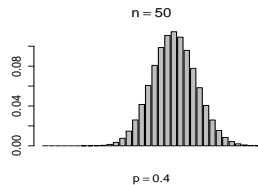
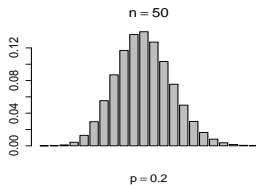
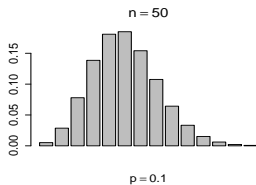
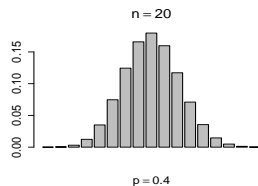
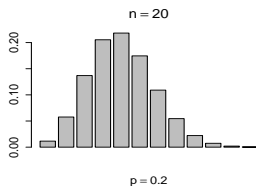
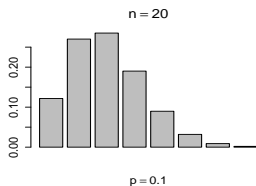
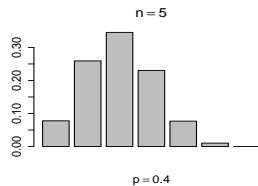
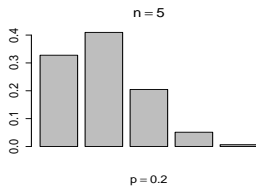
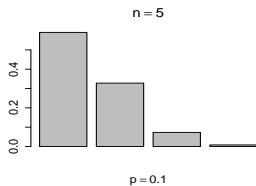
- De acuerdo con la Organización Mundial de la Salud un individuo tiene sobrepeso si su índice de masa corporal (IMC) es superior a 25. Se sabe que el IMC de una población es una variable con distribución normal de media $\mu = 26$ y desviación típica $\sigma = 6$.

(a) Calcula la probabilidad de que un individuo seleccionado al azar en esta población presente sobrepeso.

(b) Calcula el valor x tal que el IMC del 25% de la población es menor que x .

(c) Si se seleccionan aleatoriamente 100 individuos y se calcula la media de sus IMC, ¿cuál es la probabilidad de que esta media sea superior a 25.5?

Relación entre la distribución binomial y la normal



Relación entre la distribución binomial y la normal

Para valores grandes de n ,

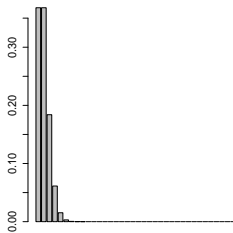
$$B(n, p) \approx N(\mu = np, \sigma = \sqrt{np(1-p)})$$

La aproximación es mejor cuanto mayor es n y el valor de p es más cercano a $1/2$.

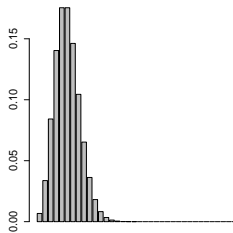
Ejemplos:

- ▶ Al lanzar una moneda 500 veces, calcula la probabilidad aproximada de obtener entre 250 y 265 caras.
- ▶ La prevalencia de cierta infección en una población es del 20%. En un estudio se analizan 200 individuos. ¿Con qué probabilidad se obtendrán entre 30 y 40 individuos infectados en esta muestra?

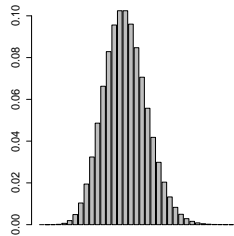
Relación entre la distribución de Poisson y la normal



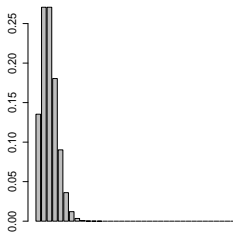
$\lambda = 1$



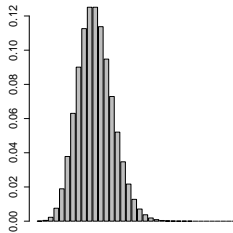
$\lambda = 5$



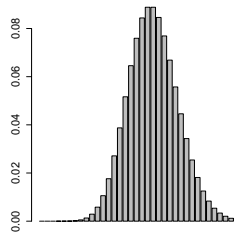
$\lambda = 15$



$\lambda = 2$



$\lambda = 10$



$\lambda = 20$

Relación entre la distribución de Poisson y la normal

Para valores grandes de λ ,

$$\mathcal{P}(\lambda) \approx N(\mu = \lambda, \sigma = \sqrt{\lambda})$$

Completa qué valores hay que escoger para μ y σ en la aproximación anterior.

Cuanto mayor es el valor de λ , mejor es la aproximación.

Ejemplo:

Supongamos que X es una v.a. con distribución de Poisson de parámetro $\lambda = 29$. Determina el valor aproximado c tal que

$$P(29 - c < X < 29 + c) \approx 0.95$$