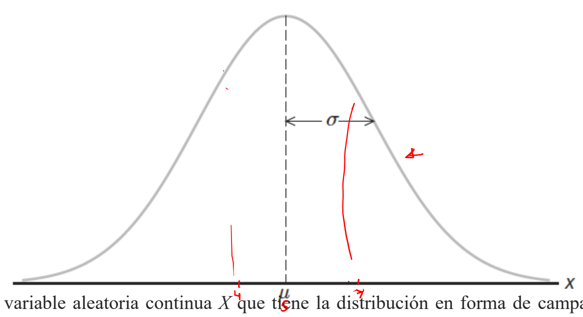
Distribución Normal

Distribuciones de VAC
Distribución normal

- La distribución de probabilidad continua más importante en todo el campo de la estadística es la **distribución normal**, que a menudo se denomina **distribución gaussiana** en honor de Karl Friedrich Gauss (1777-1855). Su gráfica, denominada **curva normal**, es la curva con forma de campana, la cual describe de manera aproximada muchos fenómenos que ocurren en la naturaleza, la industria y la investigación.



Una variable aleatoria continua X que tiene la distribución en forma de campana de la figura anterior, se denomina **variable aleatoria normal**. La ecuación matemática para la distribución de probabilidad de la variable normal depende de los dos parámetros μ y σ, su media y su desviación estándar, respectivamente. Por ello, denotamos los valores de la densidad de X por $n(x; \mu, \sigma)$. $X \sim n(x, \mu, \sigma)$

$$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty,$$

donde $\pi = 3.14159\ldots$ y $e = 2.71828\ldots$

- En la figura 6.3 aparecen dos curvas normales que tienen la misma desviación estándar pero diferentes medias. Las dos curvas son idénticas en forma, pero están centradas en diferentes posiciones a lo largo del eje horizontal.

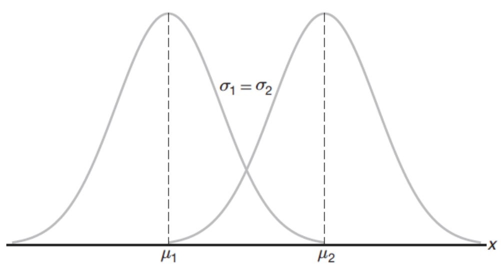


Figura 6.3: Curvas normales con $\mu_1 < \mu_2$ y $\sigma_1 = \sigma_2$.

- En la figura 6.4 se muestran dos curvas normales con la misma media pero con desviaciones estándar diferentes.

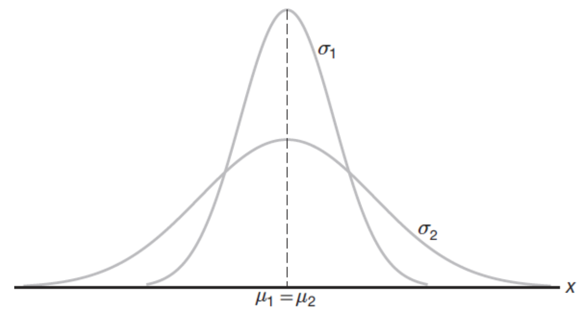


Figura 6.4: Curvas normales con $\mu_1 = \mu_2$ y $\sigma_1 < \sigma_2$.

- La figura 6.5 muestra dos curvas normales que tienen diferentes medias y diferentes desviaciones estándar. Evidentemente, están centradas en posiciones diferentes sobre el eje horizontal y sus formas reflejan los dos valores diferentes de σ.

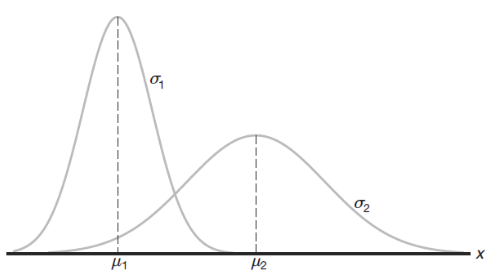


Figura 6.5: Curvas normales con $\mu_1 < \mu_2$ y $\sigma_1 < \sigma_2$.

- Con base en lo que observamos en las figuras 6.3 a 6.5, y en el examen de la primera y la segunda derivadas de $n(x; \mu, \sigma)$, listamos las siguientes propiedades de la curva normal:
- La moda, que es el punto sobre el eje horizontal donde la curva tiene su punto máximo, ocurre en $x = \mu$.
- La curva es simétrica alrededor de un eje vertical a través de la media μ.
- La curva tiene sus puntos de inflexión en $x = \mu \pm \sigma$, es cóncava hacia abajo si $\mu - \sigma < X < \mu + \sigma$, y es cóncava hacia arriba en otro caso.
- La curva normal se aproxima al eje horizontal de manera asintótica, conforme nos alejamos de la media en cualquier dirección.
- El área total bajo la curva y sobre el eje horizontal es igual a uno.

- La media y la varianza de *n* (x; μ, σ) son μ y σ^2 respectivamente, por lo tanto la desviación estándar es σ.

Para determinar el área bajo la curva, es decir la probabilidad, se emplea la siguiente fórmula:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma)\, dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx,$$

- La dificultad que se enfrenta al resolver las integrales de funciones de densidad normal exige tabular las áreas de la curva normal para una referencia rápida. Sin embargo, sería inútil tratar de establecer tablas separadas para cada posible valor de μ y σ. Por fortuna, podemos transformar todas las observaciones de cualquier variable

- aleatoria normal X en un nuevo conjunto de observaciones de una variable aleatoria normal Z con media 0 y varianza 1. Esto se puede realizar mediante la transformación

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

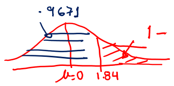
En consecuencia, podemos escribir

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2}\, dz$$

$$= \int_{z_1}^{z_2} n(z; 0, 1)\, dz = P(z_1 < Z < z_2),$$

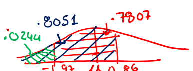
$n(x; \mu, \sigma)$

- Ahora hemos reducido el número requerido de tablas de áreas de curva normal a una, la de la distribución normal estándar. La tabla indica el área bajo la curva normal estándar que corresponde a $P(Z < z)$ para valores de z que van de –3.49 a 3.49.
- Ejemplos:
- Dada una distribución normal estándar, calcule el área bajo la curva que se localiza:
- .8729 → a) A la izquierda de z=1.170   0.8790



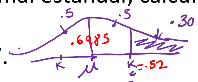
1 − .9671 = .0329

- .0329 → *b*) a la derecha de z = 1.84, y
- .7807 → *c*) entre z = –1.97 y z = 0.86.

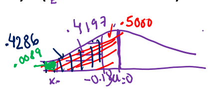


.8051 − 0.244 = 0.7807

- Dada una distribución normal estándar, calcule el valor de k tal que
- k = 0.52 *a)* $P(Z > k) = 0.3015$, y
- k = -2.37 *b)* $(k < Z < -0.18) = 0.4197$.



K = 1 − .3015 = 0.6985
K = 0.52



.5   .4286 − .4197 = .0089

- Dada una variable aleatoria X que tiene una distribución normal con μ = 50 y σ = 10, calcule la probabilidad de que X tome un valor entre 45 y 62.

$\mu = 50$

$\sigma = 10$

$\frac{1}{\sqrt{2\pi\sigma}} \int_{45}^{62} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{2\pi(100)}} \int e^{-\frac{1}{2(100)}(x-50)^2} dx$

$P(45 < X < 62)$ con $x_1 = 45$, $x_2 = 62$

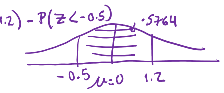
$z_1 = \frac{45 - 50}{10}$, $z_2 = \frac{62 - 50}{10}$

$z_1 = \frac{-5}{10} = \frac{-1}{2} = -0.5$ $z_2 = \frac{12}{10} = 1.2$

$P(z_1 < Z < z_2)$

$P(-.05 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5)$

=



$P(-.05 < Z < 1.2)$ = 0.8849 − 0.385

= .5764

$P(45 < X < 62) = 57.64\%$

- Dado que X tiene una distribución normal con $\mu = 300$ y $\sigma = 50$, calcule la probabilidad de que X tome un valor mayor que 362.

- Dada una distribución normal con $\mu = 40$ y $\sigma = 6$, calcule el valor de x que tiene 45% del área a la izquierda, y 14% del área a la derecha.

- Cierto tipo de batería de almacenamiento dura, en promedio, 3.0 años, con una desviación estándar de 0.5 años. Suponga que la duración de la batería se distribuye normalmente y calcule la probabilidad de que una batería determinada dure menos de 2.3 años.

- Una empresa de material eléctrico fabrica bombillas de luz cuya duración, antes de quemarse, se distribuye normalmente con una media igual a 800 horas y una desviación estándar de 40 horas. Calcule la probabilidad de que una bombilla se queme entre 778 y 834 horas.

- Se utilizan medidores para rechazar todos los componentes en los que cierta dimensión no esté dentro de la especificación $1.50 \pm d$. Se sabe que esta medida se distribuye normalmente con una media de 1.50 y una desviación estándar de 0.2. Determine el valor d tal que las especificaciones “cubran” 95% de las mediciones.

- Cierta máquina fabrica resistencias eléctricas que tienen una resistencia media de 40 ohms y una desviación estándar de 2 ohms. Si se supone que la resistencia sigue una distribución normal y que se puede medir con cualquier grado de precisión, ¿qué porcentaje de resistencias tendrán una resistencia que exceda 43 ohms?

- Calcule el porcentaje de resistencias que excedan 43 ohms para el ejemplo anterior si la resistencia se mide al ohm más cercano.

- La calificación promedio para un examen es 74 y la desviación estándar es 7. Si 12% del grupo obtiene *A* y las calificaciones siguen una curva que tiene una distribución normal, ¿cuál es la *A* más baja posible y la *B* más alta posible?

- Si *X* es una variable aleatoria binomial con media $\mu = np$ y varianza
- σ^2= npq, entonces la forma limitante de la distribución es

$$Z = \frac{X - np}{\sqrt{npq}},$$

conforme $n \to \infty$, es la distribución normal estándar $n(z; 0, 1)$.

Aproximación normal a la distribución binomial

- Sea *X* una variable aleatoria binomial con parámetros *n* y *p*. Para una *n* grande, *X* tiene aproximadamente una distribución normal con μ = *np* y σ^2 =npq= np(1-p) y

$$\begin{aligned} P(X \le x) &= \sum_{k=0}^{x} b(k; n, p) \\ &\approx \text{área bajo la curva normal a la izquierda de } x + 0.5 \\ &= P\left(Z \le \frac{x + 0.5 - np}{\sqrt{npq}}\right), \end{aligned}$$

Tabla 6.1: Aproximación normal y probabilidades binomiales acumulativas reales

	$p = 0.05, n = 10$		$p = 0.10, n = 10$		$p = 0.50, n = 10$	
r	**Binomial**	**Normal**	**Binomial**	**Normal**	**Binomial**	**Normal**
0	0.5987	0.5000	0.3487	0.2981	0.0010	0.0022
1	0.9139	0.9265	0.7361	0.7019	0.0107	0.0136
2	0.9885	0.9981	0.9298	0.9429	0.0547	0.0571
3	0.9990	1.0000	0.9872	0.9959	0.1719	0.1711
4	1.0000	1.0000	0.9984	0.9999	0.3770	0.3745
5			1.0000	1.0000	0.6230	0.6255
6					0.8281	0.8289
7					0.9453	0.9429
8					0.9893	0.9864
9					0.9990	0.9978
10					1.0000	0.9997

	$p = 0.05$					
	$n = 20$		$n = 50$		$n = 100$	
r	**Binomial**	**Normal**	**Binomial**	**Normal**	**Binomial**	**Normal**
0	0.3585	0.3015	0.0769	0.0968	0.0059	0.0197
1	0.7358	0.6985	0.2794	0.2578	0.0371	0.0537
2	0.9245	0.9382	0.5405	0.5000	0.1183	0.1251
3	0.9841	0.9948	0.7604	0.7422	0.2578	0.2451
4	0.9974	0.9998	0.8964	0.9032	0.4360	0.4090
5	0.9997	1.0000	0.9622	0.9744	0.6160	0.5910
6	1.0000	1.0000	0.9882	0.9953	0.7660	0.7549
7			0.9968	0.9994	0.8720	0.8749
8			0.9992	0.9999	0.9369	0.9463
9			0.9998	1.0000	0.9718	0.9803
10			1.0000	1.0000	0.9885	0.9941

Ejemplo

- Un paciente que padece una rara enfermedad de la sangre tiene 0.4 de probabilidad de recuperarse. Si se sabe que 100 personas contrajeron esta enfermedad, ¿cuál es la probabilidad de que sobrevivan menos de 30?

- Un examen de opción múltiple tiene 200 preguntas, cada una con 4 respuestas posibles, de las que sólo una es la correcta. ¿Cuál es la probabilidad de que solamente adivinando se obtengan de 25 a 30 respuestas correctas para 80 de los 200 problemas sobre los que el estudiante no tiene conocimientos?