Universidad Nacional de Colombia
Departamento de Estadística

DIPLOMADO EN PROBABILIDAD E INFERENCIA BÁSICA

Módulo

PROBABILIDAD

TERCERA SESIÓN

Modelos continuos

El rango de una variable aleatoria continua es un intervalo real o la unión de varios de ellos

Generalidades

Concepto de VAC

Una variable aleatoria X es continua si su función de distribución acumulada $F_X(x)$ es continua. Si la función $F_X(x)$ es además diferenciable, entonces

$$F_X(x) = \int_{-\infty}^{x} f_T(t)dt$$

La función $f_X(x)$ se conoce como *función de densidad de probabilidad*.

- Para todo $[x_1;\ x_2) \in \mathbb{R}$

$$P(x_1 \leq X < x_2) = \int_{x_1}^{x_2} f_X(x)dx$$

- Si $x_1 \to -\infty$ y $x_2 \to \infty$, entonces

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

- $P(X = x) = 0$

Atributos numéricos de una VAC

El *valor esperado* de una VAC X es una medida de localización y se define como

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

La *varianza* de una VAC X es una medida de dispersión y se define como

$$\sigma^2 = V(X) = E(X^2) - \mu^2$$

donde

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx$$

Tanto $E(X)$ como $V(X)$ son constantes y tienen unidades de medida. Usualmente, se toma como medida de dispersión la desviación estándar dada por $\sigma = \sqrt{V(X)}$.

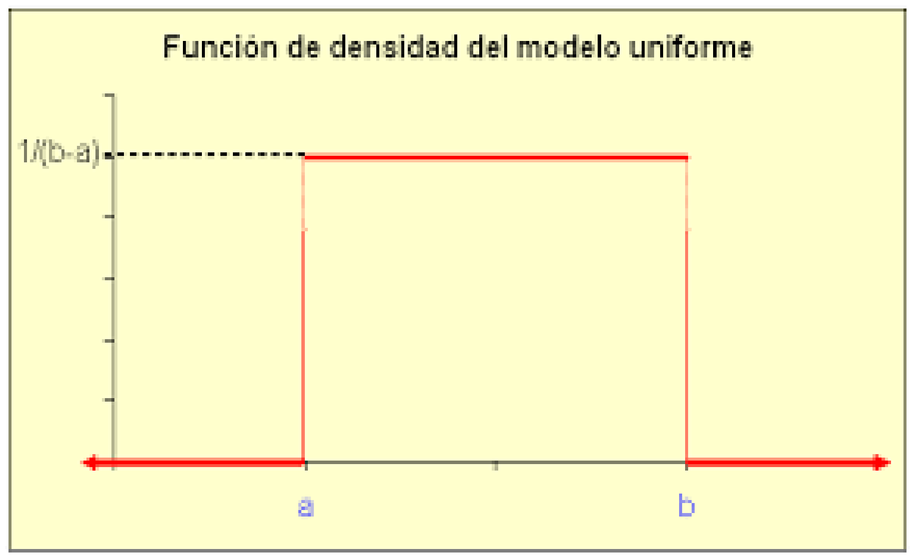
Distribución uniforme continua

Distribución uniforme continua

La VAC X se distribuye uniformemente en el intervalo $[a, b]$, y se simboliza por $X \sim U(a, b)$, si su función de densidad es

$$\frac{1}{b-a}\int_a^b dx = \frac{b-a}{b-a} = 1$$

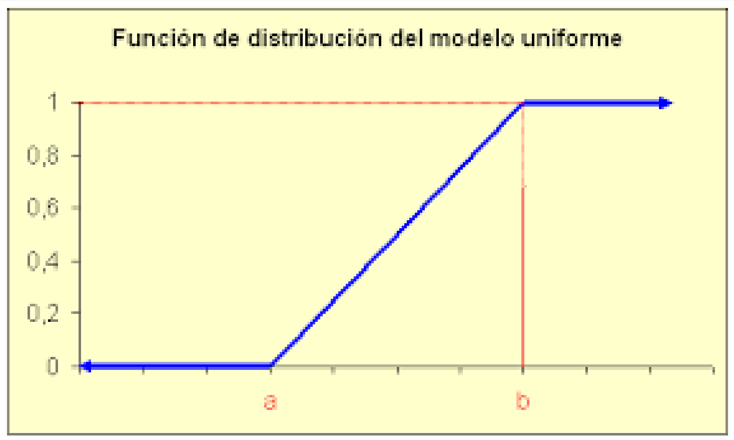
$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$$



Distribución uniforme

La función de distribución acumulada de la variable X es

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$



La probabilidad de que la variable aleatoria X tome valores en el intervalo real $[x_1, x_2] \subseteq [a, b]$ es igual a

$$P(x_1 < X \le x_2) =$$

$$= F_X\,(x_2) - F_X(x_1) =$$

$$= \frac{x_2 - x_1}{b - a}$$

Atributos de una distribución uniforme

El valor esperado de la variable aleatoria uniforme es

$$EX = \frac{1}{b-a}\int_a^b x dx = \frac{x^2}{2(b-a)}\bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{1}{2}(a+b)$$

La varianza de la variable aleatoria uniforme es

$$V(X) = E(X^2) - (EX)^2 = \frac{(b-a)^2}{12}$$

donde

$$E(X^2) = \frac{1}{b-a}\int_a^b x^2 dx = \frac{x^3}{3(b-a)}\bigg|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{1}{3}(a^2+ab+b^2)$$

Caso particular

Caso particular. Si $X \sim U(0,1)$, entonces su función de densidad es

$$f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

La función de distribución acumulada es

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Luego,

$$P\left(\frac{1}{4} \leq X < \frac{1}{2}\right) = F_X\left(\frac{1}{2}\right) - F_X\left(\frac{1}{4}\right) = \frac{1}{4}$$

Además, se tiene que

$$EX = \frac{1}{2}$$

$$V(X) = \frac{1}{12}$$

Ejemplo

Dos amigos, Fernando y Antonio deben encontrarse en una parada de autobús entre las 09:00 y las 10:00 a.m. Cada uno esperara al otro por no más de 10 minutos. ¿Cuál es la probabilidad de que no se encuentren si Fernando llega a las 09:30?

Ambas personas llegaran a la parada del autobús en un tiempo T, medido en minutos, tal que $T \sim \mathcal{U}(0; 60)$. Por consiguiente, la probabilidad de que los amigos no se encuentren es

$$P(0 < T < 20) + P(40 < T < 60) = \frac{1}{60}\left(\int_0^{20} dt + \int_{40}^{60} dt\right) = \frac{2}{3}$$

Distribución exponencial

Distribución exponencial

La variable aleatoria X tiene distribución exponencial de probabilidades con parámetro de escala $\lambda > 0$, si su función de densidad está dada por

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

Simbólicamente, se escribe $X \sim \mathcal{E}xp(\lambda)$.

La función de distribución acumulada de X es

$$F_X(x) = 1 - e^{-\lambda x}, \qquad x > 0$$

Los atributos numéricos de la distribución exponencial son tales que

$$EX = \sigma_X = \lambda^{-1}$$

Propiedad de la distribución exponencial

Sea $X \sim \mathcal{E}xp(\lambda)$. Se tiene que

$$P(X > x_1 + x_2 | X > x_1) = \frac{P(X > x_1 + x_2)}{P(X > x_1)} = \frac{e^{-\lambda(x_1+x_2)}}{e^{-\lambda x_1}} = e^{-\lambda x_2}$$

Luego,

$$P(X > x_1 + x_2 | X > x_1) = P(X > x_2)$$

Es decir, no importa cómo un fenómeno exponencial evoluciona hasta su realización particular x_1.

Ejemplo

El tiempo, en minutos, transcurrido entre dos llegadas sucesivas a un cajero automático se puede representar a través de una variable aleatoria X que se distribuye exponencialmente con parámetro $\lambda = 0{,}5$.

(a) La probabilidad de que el próximo usuario llegue en menos de 3 minutos es

$$P(X \leq 3) = 1 - e^{-1.5} = 0.7769$$

(b) Se espera que el próximo cliente llegue, en promedio, en $2\ min$ con una desviación estándar de $2\ min$.

(c) La probabilidad de que el siguiente usuario llegue al cajero en un tiempo que se encuentra a una desviación estándar del tiempo medio es igual a

$$P(0 \leq X \leq 4) = 1 - e^{-2} = 0.8647$$

(d) La probabilidad de que el próximo usuario llegue después de 4 minutos si ya se sabe que el cajero ha estado vacío por tres minutos desde del último usuario que lo utilizó es

$$P(X \geq 4 | X \geq 3) = e^{-0.5} = 0.6065$$

Distribución normal de probabilidades

Distribución normal

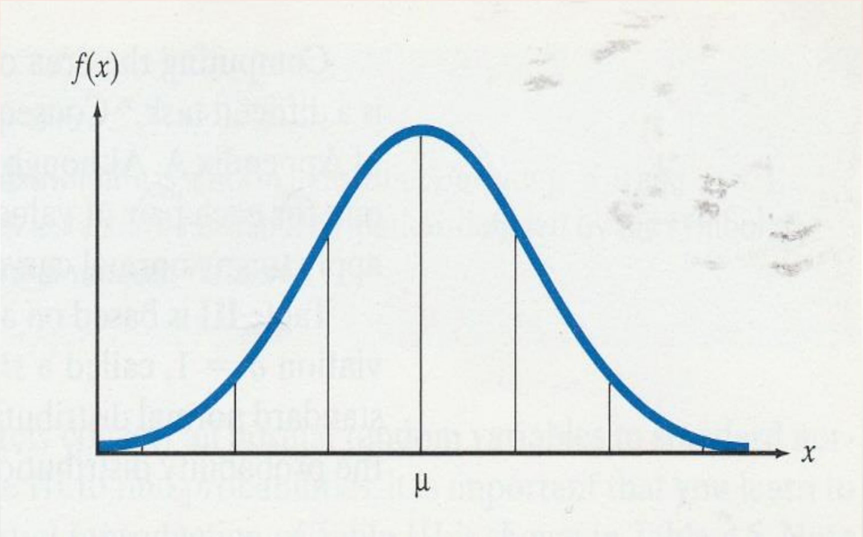
La VAC X tiene distribución normal de probabilidades si su función de densidad es

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

con $-\infty < x < \infty$, parámetro de localización $-\infty < \mu < \infty$ y parámetro de escala $\sigma > 0$.

Simbólicamente, $X \sim N(\mu; \sigma)$.

Para μ y σ dados, se tiene que



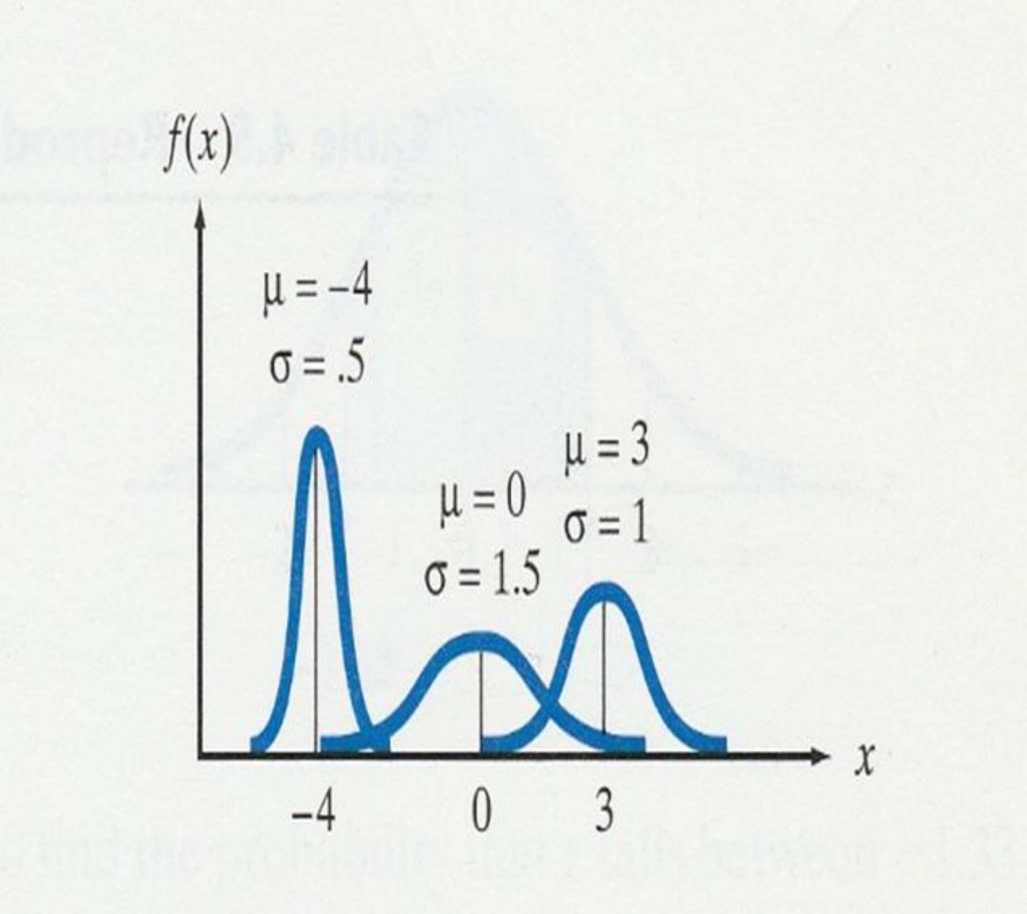
$$E(X) = \mu$$

$$V(X) = \sigma^2$$

Distribución normal

El valor $X = \mu$ (valor esperado o media de la variable normal X) muestra la posición del eje con respecto al cual $f_X(x)$ es simétrica.

La desviación estándar σ de la variable normal X muestra el grado de dispersión de los valores x con respecto al valor medio μ.



Distribución normal estándar

La función de distribución acumulada de la variable aleatoria normal X está dada por

$$\Phi_X(x) = P(X < x) =$$

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

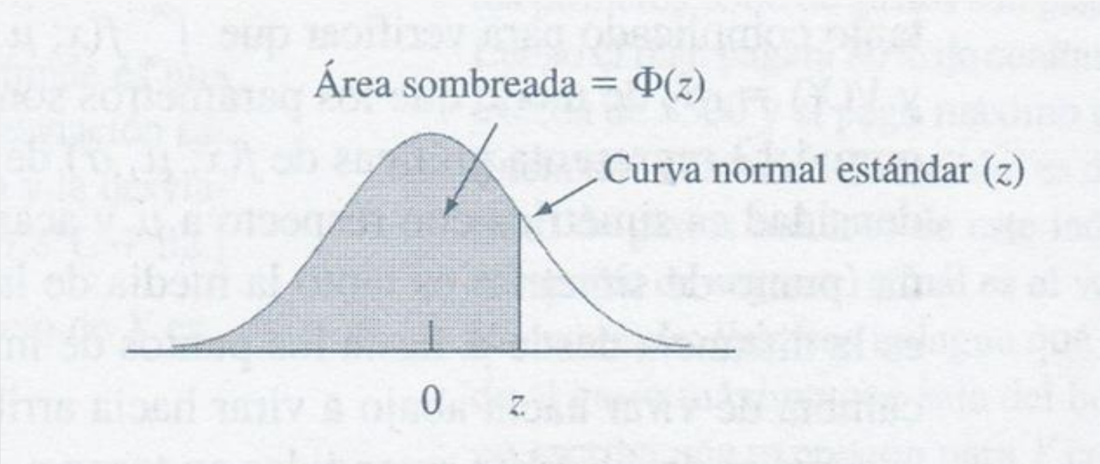
Luego, para todo $(a, b) \subseteq \mathbb{R}$ se tiene

$$P(a < X < b) = \Phi_X(b) - \Phi_X(a)$$

En particular, sea $Z \sim N(0; 1)$. Luego,

$$\Phi_Z(z) = P(Z < z) =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left[-\frac{1}{2}t^2\right] dt$$



Estandarización

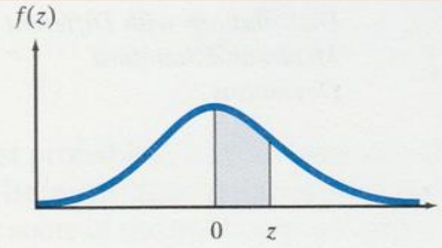
TEOREMA. Si la variable $X \sim N(\mu;\ \sigma)$, entonces

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

Luego, para todo $(a,b) \subseteq \mathbb{R}$ se tiene

$$P(a < X < b) = \Phi_X(b) - \Phi_X(a) =$$

$$= \Phi_Z\left(\frac{b-\mu}{\sigma}\right) - \Phi_Z\left(\frac{a-\mu}{\sigma}\right)$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429

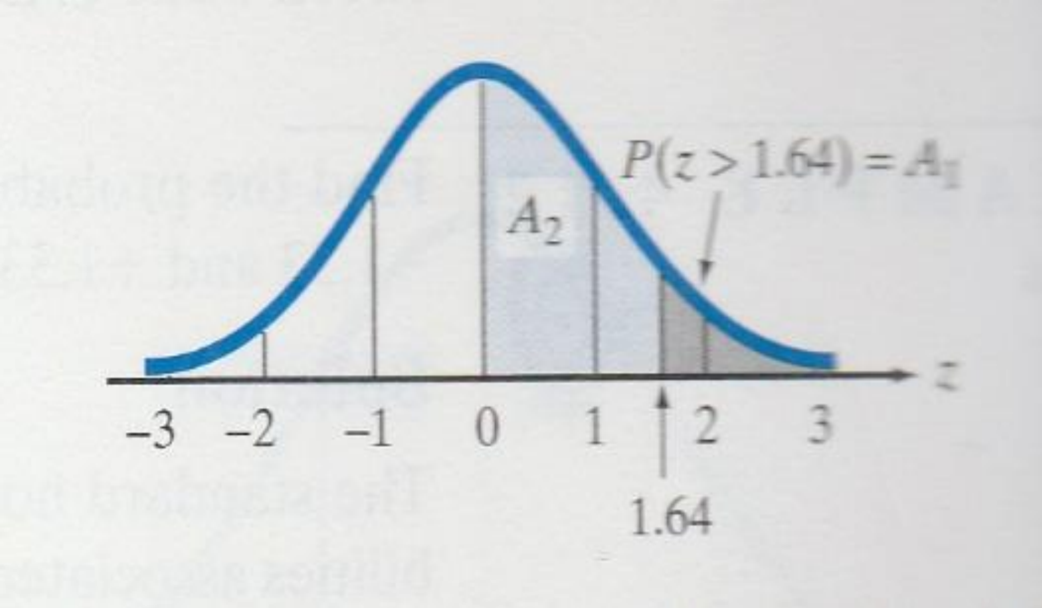
Cálculo de probabilidades

Sea $X \sim N(\mu = 3{,}0;\ \sigma = 1{,}5)$.
Se pide:

$$P(X > 5{,}46) =$$

$$= P\left(\frac{X - 3{,}0}{1{,}5} > \frac{5{,}46 - 3{,}0}{1{,}5}\right) =$$

$$= P(Z > 1{,}64)$$



$$P(Z > 1{,}64) = A_1 = 0{,}5 - A_2 =$$

$$= 0{,}5 - 0{,}4495 = 0{,}0505$$

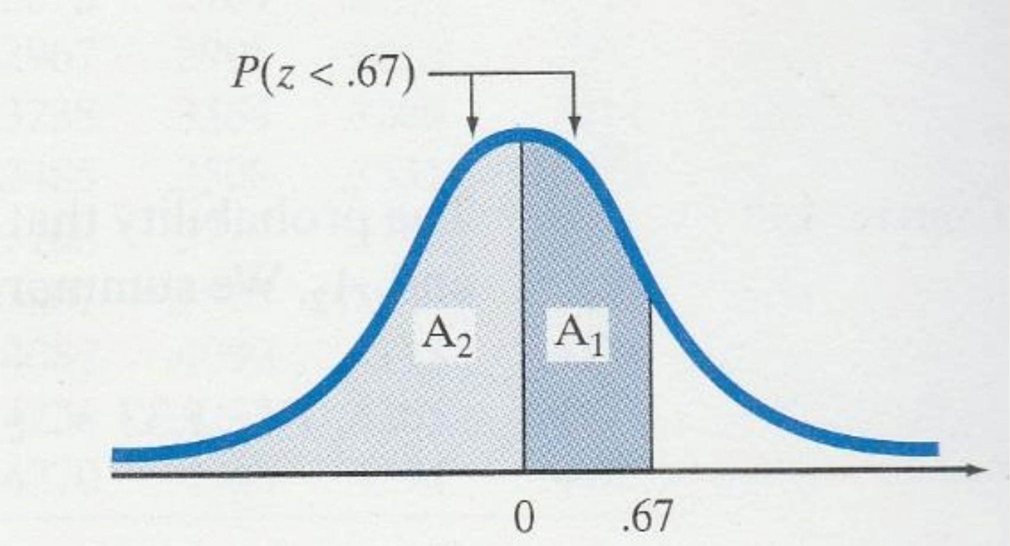
Cálculo de probabilidades

Sea $X \sim N(\mu = 3{,}0;\ \sigma = 1{,}5)$.
Se pide:

$$P(X < 4{,}0) =$$

$$= P\left(\frac{X - 3{,}0}{1{,}5} < \frac{4{,}0 - 3{,}0}{1{,}5}\right) =$$

$$= P(Z < 0{,}67)$$



$$P(Z < 1{,}64) = A_1 + A_2 =$$

$$= 0{,}5 + 0{,}2486 = 0{,}7486$$

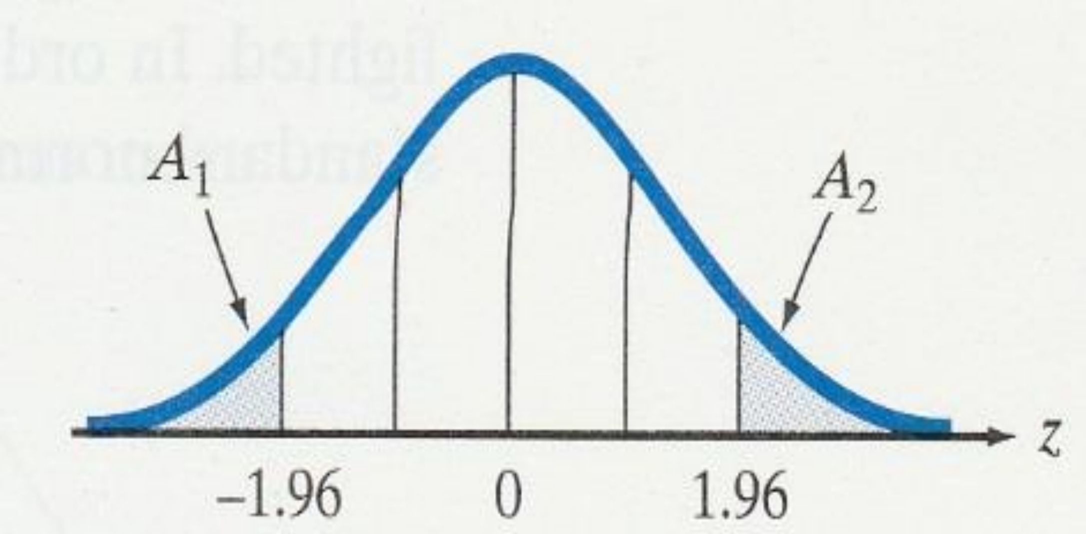
Cálculo de probabilidades

Sea $X \sim N(\mu = 3{,}0;\ \sigma = 1{,}5)$.
Se pide:

$$P(|X| > 5{,}94) =$$

$$= P(X < -5{,}94) + P(X > 5{,}94)$$

$$= P(Z < -1{,}96) + P(Z < 1{,}96)$$



$$P(|X| > 5{,}94) = A_1 + A_2 =$$

$$= 2 \times 0{,}025 = 0{,}05$$

Ejemplo

Supóngase que los puntajes (variable X) obtenidos en la prueba de admisión a una universidad son tales que $X \sim (\mu = 550;\ \sigma = 100)$.

(a)La proporción de aspirantes que obtienen más de 500 puntos en la prueba de admisión es

$$P(X > 500) = P(Z > -0{,}50) =$$

$$= 0{,}5 + 0{,}1915 = 0{,}6915$$

(b) La probabilidad de que un aspirante a ser admitido en la universidad obtenga entre 450 y 700 puntos es

$$P(450 < X < 700) =$$

$$= P(-1{,}00 < Z < 1{,}50) =$$

$$= 0{,}3413 + 0{,}4332 = 0{,}7745$$

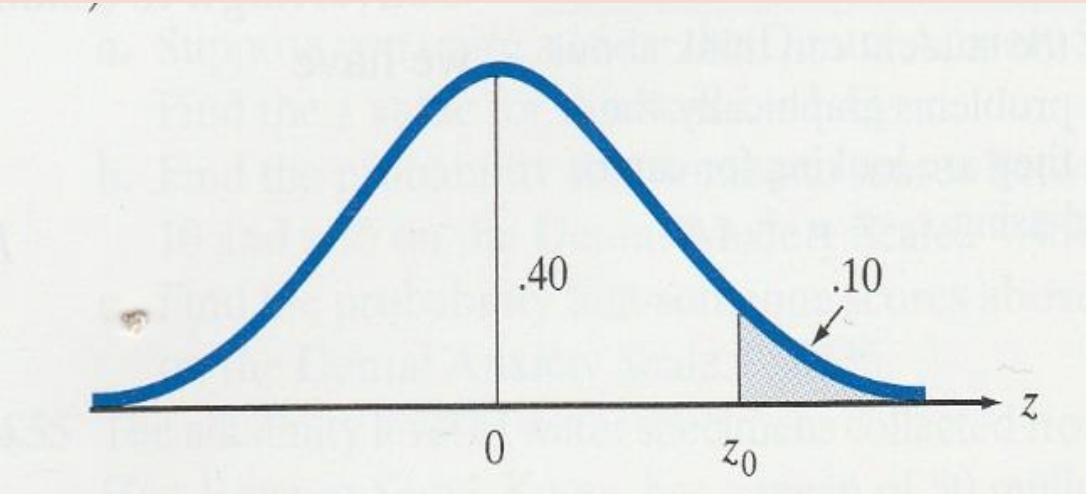
Cálculo de percentiles

Sea $X \sim N(\mu = 3{,}0;\ \sigma = 1{,}5)$. Se pide encontrar el valor x_0 de X tal que $P(X \geq x_0) = 0{,}1$. Es claro que

$$P(X \geq x_0) = P(Z \geq z_0) = 0{,}1$$

donde

$$z_0 = \frac{x_0 - 3{,}0}{1{,}5}$$



Si $P(Z \geq z_0) = 0{,}1$ entonces se tiene que $P(0 \leq Z \leq z_0) = 0{,}4$ y $z_0 \cong 1{,}28$. Luego,

$$x_0 = 1{,}28 \times 1{,}5 + 3{,}0 = 4{,}92$$

Cálculo de percentiles

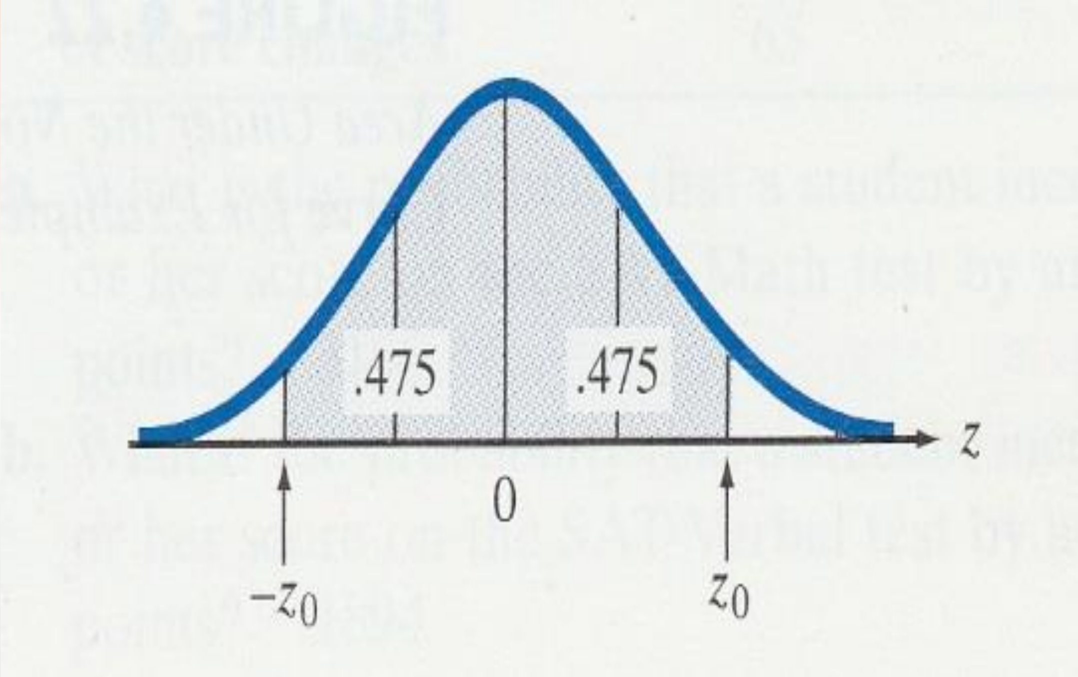
Sea $X \sim N(\mu = 3{,}0;\ \sigma = 1{,}5)$. Se pide encontrar el valor x_0 de X tal que $P(|X| \leq x_0) = 0{,}95$

$$P(|X| \leq x_0) = P(-x_0 \leq X \leq x_0) =$$

$$= P(-z_0 \leq Z \leq z_0) = 0{,}95$$

donde

$$z_0 = \frac{x_0 - 3{,}0}{1{,}5}$$



Es claro que $z_0 \cong 1{,}96$. Luego,

$$x_0 = 1{,}96 \times 1{,}5 + 3{,}0 = 5{,}94$$

Ejemplo

En el ejemplo de la prueba, la universidad considerará sólo la admisión de aquellos aspirantes cuyo puntaje supere el 90% de los que se presentan a la prueba. Luego, el puntaje mínimo de admisión x_0 se obtiene de la condición

$$P(X \leq x_0) = 0{,}9$$

o equivalentemente

$$P(X \geq x_0) = 0{,}1$$

Por tanto,

$$P(X \geq x_0) = P(Z \geq z_0) = 0{,}1$$

Luego, $z_0 \cong 1{,}28$ y el puntaje mínimo de admisión es

$$x_0 = \sigma z_0 + \mu =$$

$$= 1{,}28 \times 100 + 550 \cong 678$$

Propiedad de una distribución normal

Sea $X \sim N(\mu;\sigma)$. A una desviación estándar del valor medio se tiene que

$$P(\mu - \sigma \leq X \leq \mu + \sigma) =$$

$$= P(-1{,}0 \leq Z \leq 1{,}0) = 0{,}6826$$

Luego, en una distribución normal aproximadamente el 68% de los datos se encuentra a una desviación estándar del valor medio.

De igual modo, a dos desviaciones estándar del valor medio en una distribución normal se tiene que

$$P(-2{,}0 \leq Z \leq 2{,}0) = 0{,}9545$$

y a tres desviaciones estándar del valor medio en una distribución normal se tiene que

$$P(-3{,}0 \leq Z \leq 3{,}0) = 0{,}9986$$

GRACIAS

Diagramado en

Power Point 2013