Diplomado en Análisis de Información Geoespacial

Estadística inferencial

Autor:
M. en G. Alberto Porras Velázquez

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.

Introducción

La estadística se refiere a un conjunto de métodos enfocados a la obtención, presentación y análisis de observaciones numéricas. Como se ha visto anteriormente, la rama llamada estadística descriptiva tiene como fin de describir al conjunto de datos obtenidos.

Por otro lado, la estadística inferencial se enfoca en la toma de decisiones o realización de generalizaciones acerca de las características de todas las observaciones bajo consideración con base en información parcial o incompleta.

A continuación se definen algunos conceptos que es necesario conocer para el estudio de la estadística inferencial.

Un experimento es un proceso de medición o de realización de observaciones que tiene resultados bien definidos. En estadística, al igual que en el ámbito científico en general, los experimentos involucran mediciones e instrumentación científica, como es el caso de un termómetro o un pluvímetro.

Sin embargo un experimento estadístico también puede contemplar el registro de datos, no necesariamente provenientes de mediciones, como en el caso de las respuestas ante las preguntas referentes a las preferencias de las personas hacia un producto o su posición ante un tema en particular.

En la estadística se suelen estudiar fenómenos aleatorios, es decir, cuyos resultados están determinados por factores fortuitos. A pesar de esto, cierto tipo de regularidad es inherente al proceso, de manera que a cada uno de los posibles resultados del fenómeno se le puede asignar una fracción de probabilidad.

Un ensayo es un acto que lleva a uno de los posibles distintos resultados de un experimento. En caso de lanzar una moneda, un ensayo puede arrojar sólo uno de los dos resultados posibles, cara o cruz.

El espacio muestral de un experimento es el conjunto de todos los posibles resultados distintos del experimento y suele representarse con la letra S. Se dice que un espacio muestral es un conjunto universal. A cada resultado en un espacio muestral se le llama elemento, miembro del espacio muestral o, simplemente, punto muestral.

Un concepto directamente relacionado con el de espacio muestral es el de evento, que es algo que ocurre y que tiene como producto uno o varios resultados. Un evento aleatorio simple es el resultado de un ensayo único en cualquier experimento particular, mientras que un evento compuesto consiste en un subconjunto del espacio muestral con dos o más resultados asociados.

Por ejemplo, al lanzar un dado, un evento aleatorio simple es el obtener como resultado del lanzamiento un 3; un evento compuesto consiste en obtener un número par, es decir, hay tres eventos simples (obtener un 2, un 4 o un 6 en el lanzamiento de un dado) relacionados con él.

Existen varias maneras de representar espacios muestrales y eventos, una de ellas implica un listado de los elementos que lo componen, por ejemplo, el espacio muestral del experimento que consiste en lanzar un dado es:

$S = \{1, 2, 3, 4, 5, 6\}$

El evento A, que consiste en obtener un número par al lanzar un dado, se representa como:

$A = \{2, 4, 6\}$

Los eventos también se pueden representar mediante un enunciado que comprenda una característica común para todos ellos. Por ejemplo, supongamos el conjunto de ciudades de más de 500,000 habitantes:

S = { x | x es una ciudad de más de 500,000 habitantes}

Otra representación implica una regla o expresión matemática que defina sus características, como son los puntos sobre una circunferencia de radio 2:

$S = \{(x,y) \mid x^2 + y^2 = 2^2\}$

Probabilidad

Existen dos interpretaciones comúnmente aceptadas sobre el significado del término probabilidad.

Por un lado, la interpretación frecuentista ve la probabilidad como la fracción de veces que un evento ocurrirá si repetimos un experimento indefinidamente.

Si lanzáramos una moneda de manera indefinida, esperaríamos que 50 por ciento de los resultados fueran caras y el otro 50 por ciento cruces.

La visión frecuentista falla cuando consideramos eventos como, “va a llover mañana” debido a que esperamos que este evento ocurra solo una vez. En relación con este tipo de eventos, no hay claridad en cómo definir su frecuencia. Se han realizado intentos por definir esta probabilidad mediante la creación de una clase para casos similares, sin embargo no hay soluciones completamente satisfactorias.

Otra alternativa es interpretar la probabilidad en función de los grados subjetivos de creencia sobre la posibilidad que de ocurra uno o varios eventos.

La afirmación de que la probabilidad de lluvias para la tarde es de 60 por ciento, desde este punto de vista, es una afirmación subjetiva del meteorólogo.

En este caso el problema radica en explicar cómo los grados subjetivos de creencia (que son completamente personales) se reflejan en nuestras acciones.

En cuanto a esta interpretación, esperaríamos que una persona tenga grados de creencia que satisfagan las reglas de probabilidad. Por ejemplo, sería erróneo afirmar que la probabilidad de que llueva por la tarde es de 200%, o afirmar que al lanzar una moneda la probabilidad de que caiga cara es del 60% y la probabilidad de que caiga cruz es de 70%.

Ahora podemos ligar los conceptos de espacio muestral y eventos con la probabilidad.
La primera propiedad de la probabilidad para un evento (A), denotada por la función P(A), es que siempre tendrá un valor entre 0 y 1. Cero significa que un evento es imposible, mientras que1 indica que un evento siempre ocurrirá.

$0 \leq P(A) \leq 1$

Una segunda propiedad radica en que la suma de las probabilidades de todos los posibles resultados (o eventos simples) de un experimento es igual a 1.

$\Sigma\ P(A) = 1$

En la expresión que se muestra, la letra griega sigma (Σ) se utiliza para representar la suma de las probabilidades de los eventos simples que constituyen el espacio muestral del experimento.

Debido a que todos los eventos simples de un experimento conforman el espacio muestral, la segunda propiedad nos dice que la probabilidad del espacio muestral es 1, es decir, es completamente seguro que ocurra alguno de los resultados posibles del espacio muestral.

La probabilidad de ocurrencia P del evento A, denotada por P(A), se calcula como la división del número de elementos del espacio muestral que componen el evento A entre el tamaño del espacio muestral.

P(A) = Número de elementos que conforman A / Tamaño del espacio muestral

Por ejemplo, la probabilidad del evento B, que consiste en obtener un número par al lanzar un dado, es:

$P(B) = 3/6 = \frac{1}{2}$

En donde B está conformado por tres eventos y el tamaño del espacio muestral es 6.

En ocasiones requerimos combinar dos o más eventos en uno compuesto. Este evento se puede formar a partir de uniones o intersecciones de eventos simples.

Por definición decimos que la unión entre dos eventos A y B, simbolizada mediante la expresión (AUB), ocurre si en una sola realización del experimento ocurre el evento A, o el evento B o ambos eventos.

La intersección de dos eventos A y B, denotada por ($A \cap B$), se presenta si el evento A y el evento B suceden en una sola realización del experimento. La letra U invertida se utiliza para representar la intersección entre conjuntos. La manera más simple de expresar una intersección mediante una frase es A y B, es decir, que se presentan ambos eventos.

Dos eventos A y B son mutuamente excluyentes o disjuntos si la intersección de ellos es el conjunto vacío, es decir, si A y B no tienen elementos en común. Esto se expresa como:

$A \cap B = \emptyset$

en donde Ø representa el conjunto vacío.

Por ejemplo, en el lanzamiento de un dado definimos el evento A como aquel en el que se obtiene un número par, en tanto que en el evento B se obtiene un número impar.
No hay elementos que sean pares y nones al mismo tiempo, por lo que la intersección es nula, indicando que son eventos mutuamente excluyentes.

La probabilidad de la unión de dos eventos A y B se calcula como:

$$\mathbf{P(AUB) = P(A) + P(B) - P(A \cap B)}$$

El procedimiento para calcular la probabilidad $P(A \cap B)$ se abordará más adelante.
El complemento de un evento A respecto de S, denotado como A'**,** es el subconjunto de todos los elementos de S que no están en A.

Por otro lado, la regla de la complementariedad nos dice que la suma de las probabilidades de un evento y su complemento es 1.

P(A)+P(A’) = 1 …. Regla de complementariedad.

A partir de la regla de la complementariedad se puede deducir que la probabilidad de un evento A es igual a la probabilidad de todo el espacio muestral (1), menos la probabilidad del complemento del evento A, esto es:

P(A) = 1 – P(A’)

En ocasiones deseamos calcular la probabilidad de un evento bajo ciertas circunstancias. Por ejemplo, sabemos que la probabilidad de obtener un número impar en el lanzamiento de un dado es 0.5 (el evento A se muestra en la ilustración 1 como la circunferencia con contorno azul).

Sin embargo, supongamos que en un lanzamiento se obtuvo un número mayor que tres, representado por el evento B mostrado en la circunferencia con contorno rojo. ¿Se mantendrá la probabilidad de observar un número impar en ese caso particular?

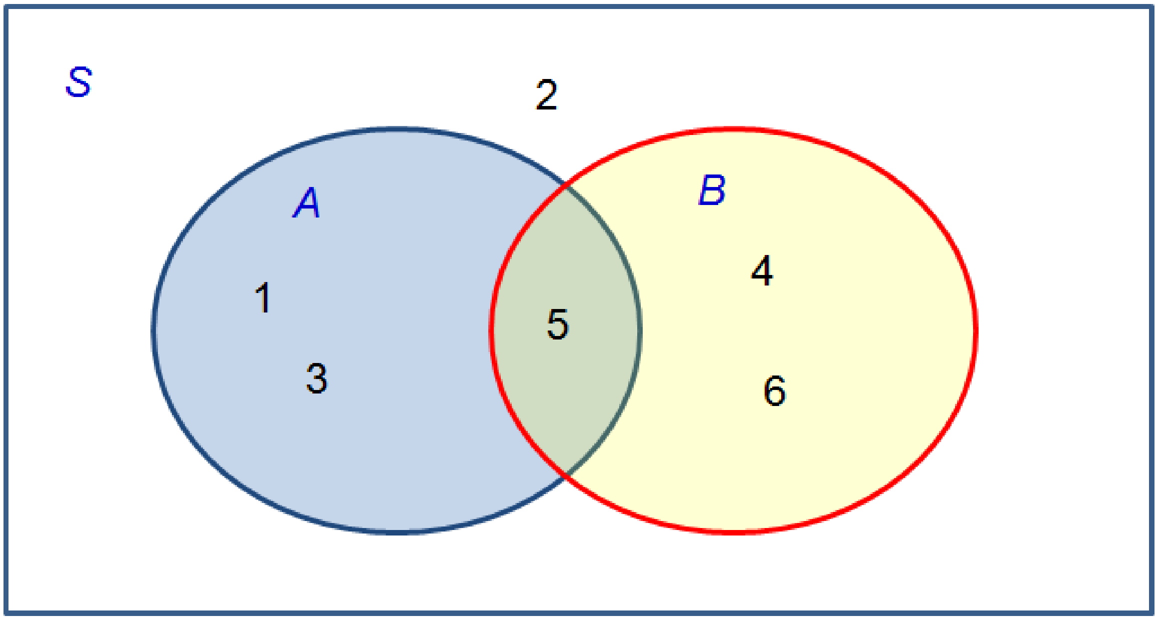

Ilustración 1. Diagrama de Venn

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.

La respuesta es negativa, pues al realizarse el evento B se reduce el tamaño del espacio muestral de seis eventos simples a tres eventos simples, como se observa en la ilustración 1.

El único número impar en el espacio muestral del evento B es el 5. De manera que la probabilidad de que ocurra el evento A cuando se presenta el evento B, expresado esto en la notación matemática **P(A|B)**, es de uno entre tres casos posibles, es decir, un tercio.

$P(A|B) = 1/3$

La expresión matemática para calcular la probabilidad condicional de que el evento A ocurra dado que ocurre el evento B , conocida como teorema de Bayes, establece que el resultado se obtiene al dividir la probabilidad de que ambos eventos sucedan, es decir, la intersección de A y B, entre la probabilidad de que el evento B se realice.

$P(A|B) = P(A \cap B)/ P(B)$

De la fórmula anterior se puede deducir que la probabilidad de que ocurra la intersección de dos eventos es:

$P(A \cap B) = P(A|B)\ P(B)$

Si el evento A es independiente del evento B, es decir, si la ocurrencia de A no se ve influida por la ocurrencia de B, entonces $P(A|B) = P(A)$.

Por ejemplo, si tomamos una baraja de 52 cartas y queremos conocer la probabilidad del evento A, que consiste en obtener uno de los cuatro reyes en la primera extracción, entonces:

$P(A) = 4/52$

Dado que se obtuvo un rey en la primera extracción, ¿cuál es la probabilidad del evento B, que consiste en obtener otro rey en la segunda extracción de una carta de la baraja? Después de la primera extracción quedan 51 cartas, entre las cuales solo hay tres reyes. Entonces:

$P(B|A) = 3/51$

Finalmente, ¿cuál es la probabilidad de obtener dos reyes en las primeras dos extracciones de cartas de una baraja?

$P(A \cap B) = P(A)*P(B|A)=(4/52)*(3/51)$

Si después de extraer una carta se hubiera reemplazado en la baraja, entonces los eventos A y B serían independientes, es decir:

$P(A \cap B) = P(A)*P(B|A) = P(A)*P(B) = (4/52)*(4/52)$

Funciones probabilísticas

En matemáticas se define una función como una asociación entre un elemento (x) de un conjunto y un elemento (y) de otro conjunto, en la cual a cada elemento x se le asocia uno y solo uno de los elementos de y. A los elementos x y y se les suele llamar par ordenado.

Con mucha frecuencia, x y y son valores numéricos. Los dos conjuntos de elementos representan todos los posibles valores que x y y pueden tomar y cualquier regla que defina a la relación entre ellos será una ecuación. Por ejemplo, la ecuación de la recta:

$$y = x + 3$$

En donde a cada valor de x le corresponde un valor de y.

La mayoría de las funciones analizadas en estadística son funciones probabilísticas. Como se ha visto, siempre que se habla de probabilidad, el término se utiliza asociado con un evento aleatorio. En el análisis estadístico, a menudo lo que se asigna es un número a un evento aleatorio y a tal número se le denomina como valor de la variable aleatoria.

Si los valores que toma un símbolo tal como **x** están asociados con los eventos aleatorios simples de un experimento dado y, por lo tanto, dependen de ocurrencias aleatorias, al símbolo se le denomina variable aleatoria.

Considera un experimento que consiste en lanzar tres monedas no balanceadas simultáneamente. En este caso imagina que la probabilidad de que aparezca cara (C) es de 0.6, mientras que la probabilidad de una cruz (Z) es de 0.4. Los ocho eventos aleatorios simples y su probabilidad asociada se muestran en la tabla 1.

Eventos simples	Probabilidad
CCC	**0.6*0.6*0.6 = 0.216**
CCZ	**0.6*0.6*0.4 = 0.144**
CZC	**0.144**
CZZ	**0.6*0.4*0.4 =0.096**
ZCC	**0.144**
ZCZ	**0.096**
ZZC	**0.096**
ZZZ	**0.4*0.4*0.4= 0.064**

Tabla 1. Eventos simples y su probabilidad asociada para el experimento de lanzar tres monedas de manera simultánea.

Nota que la suma de las probabilidades asociadas a los eventos simples debe ser igual a 1.

En este experimento, el punto es relacionar un conjunto de eventos (simples o compuestos) que sean mutuamente excluyentes con los valores de una variable aleatoria.

Cada uno de los distintos valores de la variable se asocia solamente con un número real denominado probabilidad. De este modo se cuenta con una función probabilística que asocia una fracción probabilística a los distintos valores de la variable aleatoria.
En el experimento de las tres monedas, la variable aleatoria puede ser el número de caras que aparece en cada evento (estos son eventos mutuamente excluyentes), de manera que la probabilidad asociada a cada valor de la variable aleatoria (X) es igual a la suma de probabilidades de los eventos aleatorios simples que tengan X número de caras (tabla 2).

En el lanzamiento simultáneo hay sólo un evento aleatorio simple en el que no se obtienen caras y otro en el que se obtienen tres caras; por otro lado, hay tres eventos simples asociados a una cara y dos caras.

Valores de X	Probabilidad
0	**0.064**
1	**0.096*3 = 0.288**
2	**0.144*3 = 0.432**
3	**0.216**

Tabla 2. Función probabilística para la variable aleatoria X, el número de caras que se pueden obtener en tres lanzamientos no simultáneos.

Los distintos valores que X puede tomar y su probabilidad asociada constituyen la distribución probabilística de X. Se les denomina distribuciones porque, en cada caso, la probabilidad total de 1 se reparte entre todos los posibles valores diferentes de la variable aleatoria.

En la ilustración 2 se muestra la gráfica de distribución de probabilidad para el experimento del lanzamiento simultáneo de tres monedas no balanceadas.

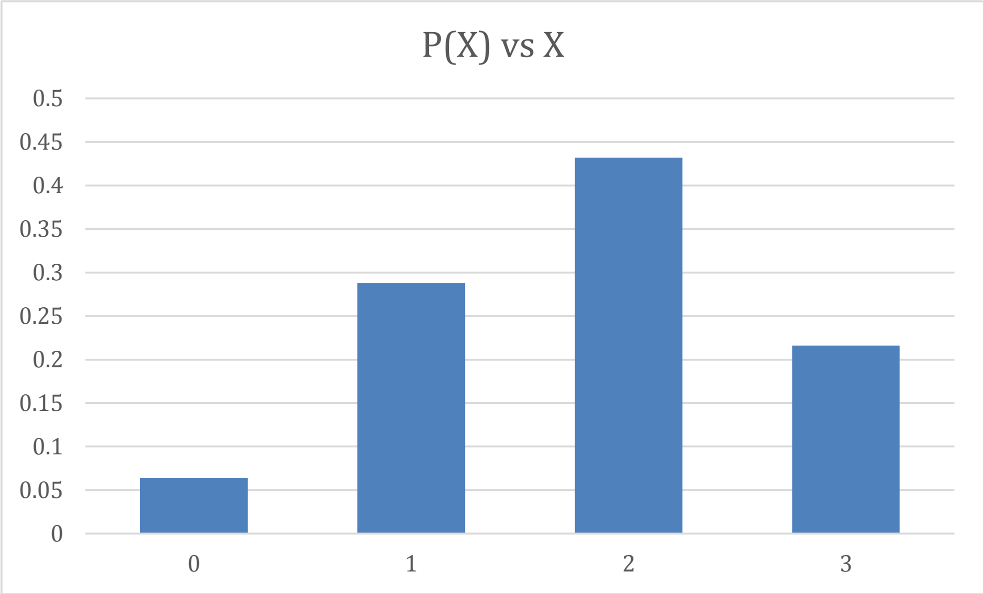

Ilustración 2. Distribución de probabilidad de la variable X, número de caras en el lanzamiento de tres monedas.

Una distribución probabilística puede ser discreta o continua. La primera está relacionada con una variable discreta; en el caso del lanzamiento de las tres monedas la distribución es discreta porque la variable puede tener como valor tan solo números enteros (0,1,2,3). Las variables de este tipo se obtienen como resultado de un proceso de conteo.

Si una variable aleatoria es continua, le corresponderá una distribución de probabilidad continua y será producto del proceso de medición (como es el caso de la estatura). La distribución normal es una distribución continua.

Una distribución probabilística es una distribución teórica. En el ejemplo construimos la distribución de probabilidad de un experimento sin necesidad de realizar ensayos reales. Esto es diferente al caso de la distribución de frecuencias relativas (que se puede visualizar en un histograma de frecuencias relativas), en donde la distribución se construye con datos reales, es decir, es una distribución empírica.

En una distribución de probabilidad, la probabilidad de un valor particular de la variable aleatoria X es una proporción de la población; mientras que en el caso de la distribución de frecuencias relativas, la frecuencia de una clase particular es una proporción de la muestra.

Ambas distribuciones tienen una relación, pues se considera que una distribución probabilística es una distribución de frecuencias relativas a largo plazo.

Dicho de otro modo, a mayor número de ensayos de un experimento, el comportamiento del histograma de frecuencias relativas se asemejará cada vez más a la distribución probabilística de la variable.

Distribución binomial y distribución normal

Durante mucho tiempo, se ha logrado identificar y caracterizar mediante funciones probabilísticas el comportamiento de una gran diversidad de fenómenos naturales.
En este apartado tan sólo se presentarán dos de las distribuciones más utilizadas, llamadas binomial y normal.

Un experimento de Bernoulli consiste en una función probabilística en la cual sólo se pueden asociar dos valores a la variable aleatoria discreta: el éxito (1), al que se le asigna una probabilidad p, y el fracaso (0), con una probabilidad asociada de 1-p= q.
La distribución binomial se construye al calcular la probabilidad de obtener x éxitos en n ensayos de Bernoulli independientes en sucesión (la probabilidad p permanece inalterada en todos los ensayos).

La fórmula para calcular la probabilidad de obtener x éxitos en n ensayos con una probabilidad de éxito p se expresa como:

$$\mathbf{b(x; n, p) = {}_nC_x\, p^x q^{n-x}}$$

en donde ${}_nC_x$ son las combinaciones de n elementos tomados en subconjuntos de x elementos y no importa el orden de los elementos del subconjunto, los cuales se calculan mediante la fórmula:

$${}_nC_x = n!/(x!(n-x)!)$$

n! es el factorial de n y se calcula como:

$$n! = n(n-1)(n-2)\ldots(1)$$

también por definición 0! = 1

En la distribución binomial la variable aleatoria x puede tomar valores entre 0 y n, es decir, se pueden tener desde cero éxitos en n ensayos hasta n éxitos en los n ensayos.
Hay que poner énfasis que en el caso de la distribución binomial el “éxito” no tiene una connotación de bueno o malo, sino que está asociado al resultado que queremos estudiar.

Por ejemplo, planteemos el experimento que consiste en lanzar simultáneamente tres monedas no balanceadas. En este ejercicio la probabilidad de obtener cara es de 0.6 y la probabilidad de obtener cruz es de 0.4.

Si en este caso identificamos como éxito obtener cruz, ¿cuál es la probabilidad de obtener dos cruces en los tres lanzamientos?

$P(X = 2) = b(2; 3, 0.4) = {}_3C_2\, p^2q^1 = 3!/(2!1!)(0.4^2)(0.6^1) = 3(0.4^2)(0.6^1) = 0.288$

Como en toda distribución, la suma de probabilidades asociadas a cada valor de la variable aleatoria es 1. Para este ejemplo:

$b(0; 3, 0.4) + b(1; 3, 0.4) + b(2; 3, 0.4) + b(3; 3, 0.4) =$
$0.216 + 0.432 + 0.288 + 0.064 = 1$

Los valores de la media y la varianza para una distribución binomial son:

$\mu = np$
$\sigma^2 = npq$

La distribución normal o distribución gaussiana es la distribución continua más utilizada en estadística. Muchos fenómenos de la naturaleza como la distribución de estaturas de una población o la de errores de medición tienen un comportamiento que coincide con la normal.

La ecuación de la distribución normal queda determinada de manera única por el valor de los parámetros correspondientes a la media μ y la varianza σ^2:

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

La gráfica de la distribución normal tiene una forma de campana con las siguientes características:

1. Es una distribución simétrica: la moda, media y mediana coinciden.
2. La curva normal se aproxima al eje horizontal de manera asintótica, conforme nos alejamos de la media en cualquier dirección.
3. 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 cualquier otro caso.

El rango de valores para la variable que se encuentra hasta una desviación estándar con respecto a la media abarca 68 por ciento de la población, en el rango entre la media y dos desviaciones estándar está 94.45 por ciento, y entre la media y tres desviaciones estándar se ubica 99.73 por ciento de la población (ilustración 3).

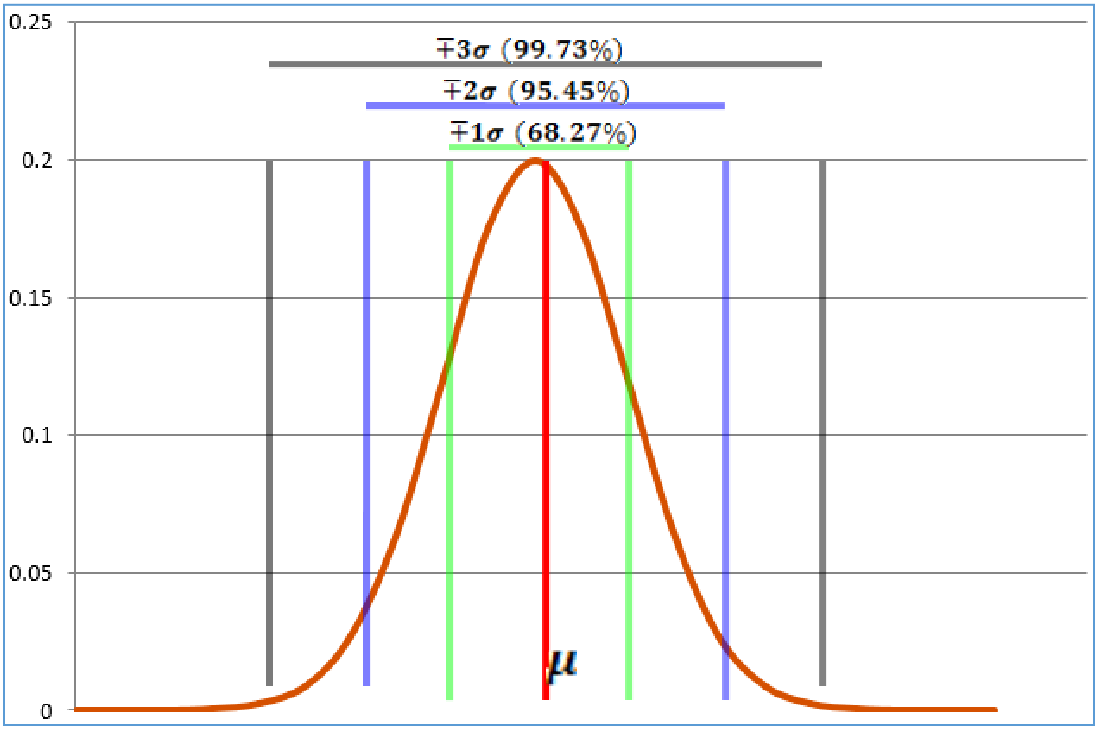

Ilustración 3. Distribución normal. Distribución de probabilidad con respecto a la media.

La curva de distribución de probabilidad o función de densidad se construye de manera que el área bajo la curva, limitada por las dos ordenadas x_1 y x_2, sea igual a la probabilidad de que la variable aleatoria X tome un valor entre x_1 y x_2 (ilustración 4). Esto se calcula a través de la integral:

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

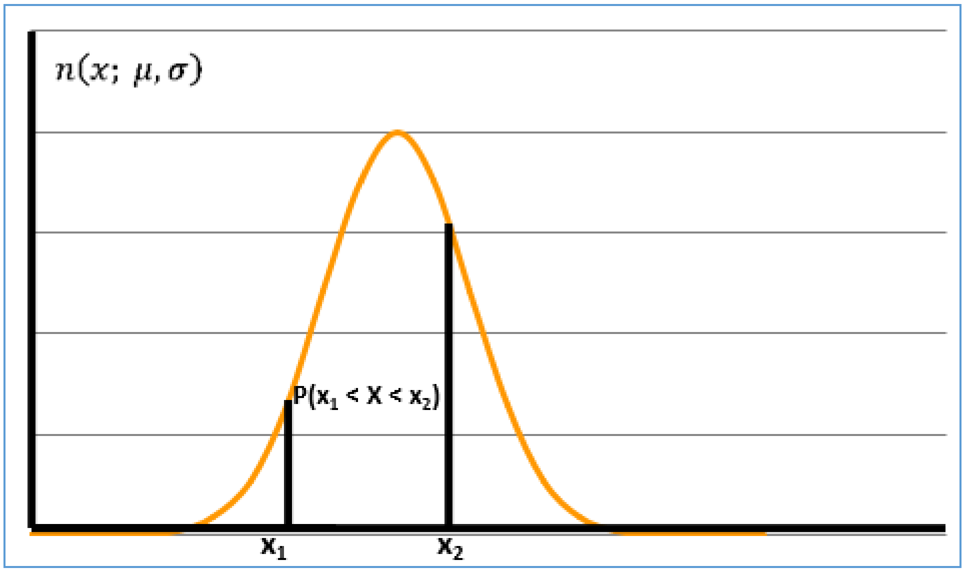

Ilustración 4. Área correspondiente a $P(x_1 < X < x_2)$ en la distribución normal.

No es necesario realizar el cálculo directo de una integral cada vez que se quiere conocer la probabilidad de que la variable tome un rango de valores entre x_1 y x_2 para una distribución normal. Existen funciones específicas que arrojan estos resultados en diversas aplicaciones (Excel) y lenguajes de programación (R).

Otra forma de encontrar estas probabilidades es mediante la transformación de los valores de la variable **X** de la curva normal que estudiamos, a los valores equivalentes en la variable **Z** en una curva normal con media igual a cero y desviación estándar igual a uno (conocida como distribución normal estándar). Los libros de estadística suelen tener una tabla de probabilidades para el estadístico Z en la distribución normal estándar. La trasformación de una x a una z determinada se da mediante la ecuación:

Z = (X - μ)/ σ

Para ilustrar el uso de la transformación al estadístico **z** imagina que deseamos conocer la probabilidad de obtener una muestra con un valor menor a x= 20, es decir, P(X < 20) para una variable que se comporta como una distribución normal, con media igual a 15 y desviación estándar de cinco.

La ecuación Z = (X - μ)/ σ se utiliza para obtener el valor correspondiente de z, que en este caso es 1 (ilustración 5). Entonces sabemos que P(X < 20) = P(Z < 1).

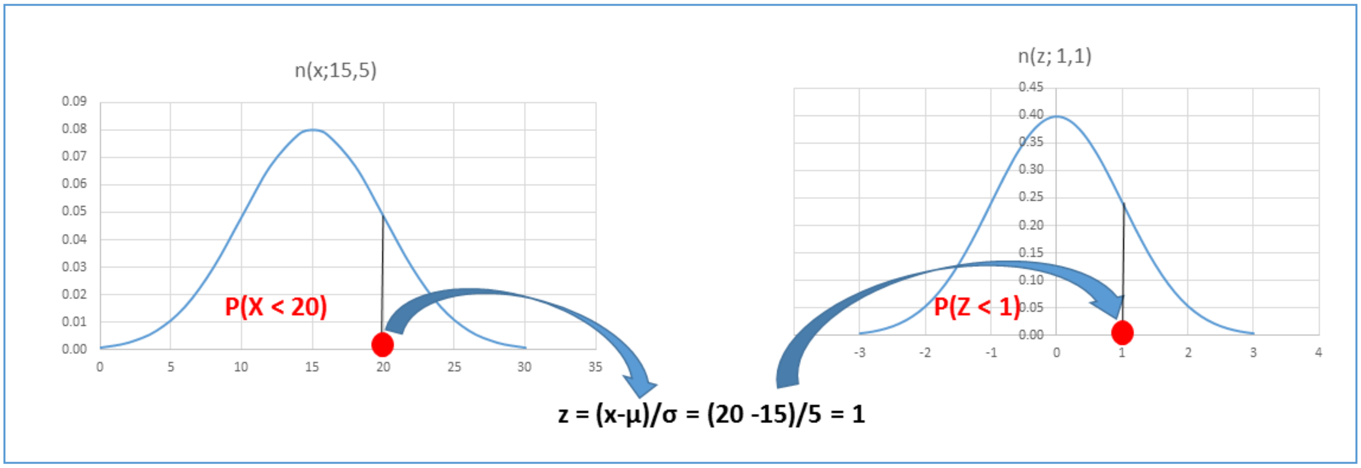

Ilustración 5. Transformación de x al estadístico z.

Posteriormente, en una tabla de áreas bajo la curva normal, se busca el valor correspondiente a P(Z < 1) = 0.88413 (tabla 3). Otro ejemplo, P(Z < 1.43) = 0.9236 se encuentra en las coordenadas 1.4 en la primera columna + 0.03 en el primer renglón.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	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	0.7704	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
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

Tabla 3. Área bajo la curva normal.

Teorema del límite central

Para las distribuciones probabilísticas se tienen tres elementos importantes: la media (medida de tendencia central), la desviación estándar (medida de dispersión) y el patrón de la distribución.

El teorema del límite central proporciona información referente a estas tres características de la distribución muestral en las medias muestrales. Por ello, si de una población se extraen todas las posibles muestras de un mismo tamaño y para cada muestra se obtiene la media, entonces la distribución de estas medias muestrales tiene las siguientes características:

- La media de la distribución muestral de la media muestral es igual a la media de la población.

$$\mu_{\bar{x}}=\mu$$

- La desviación estándar de las medias muestrales, conocido como error estándar, es igual a la desviación estándar de la población dividida entre la raíz cuadrada de n.

$$\sigma_{\overline{X}}=\frac{\sigma_X}{\sqrt{n}}$$

- La distribución muestral de la media muestral es casi normal, independientemente del patrón de la distribución de la población.

El punto dos indica que, a mayor tamaño de la muestra, el error estándar será menor y que aumentará la precisión para calcular el parámetro de una población a partir de la estimación de la media de la distribución muestral de las medias muestrales.

En la ilustración 6 la línea superior representa la distribución de una población con media μ. La línea inferior revela la distribución de medias muestrales para 200 muestras independientes, cada muestra de tamaño n= 2 (a), n= 20 (b) y n= 200 (c). Observa como el error estándar decrece en la medida que el tamaño de la muestra aumenta.

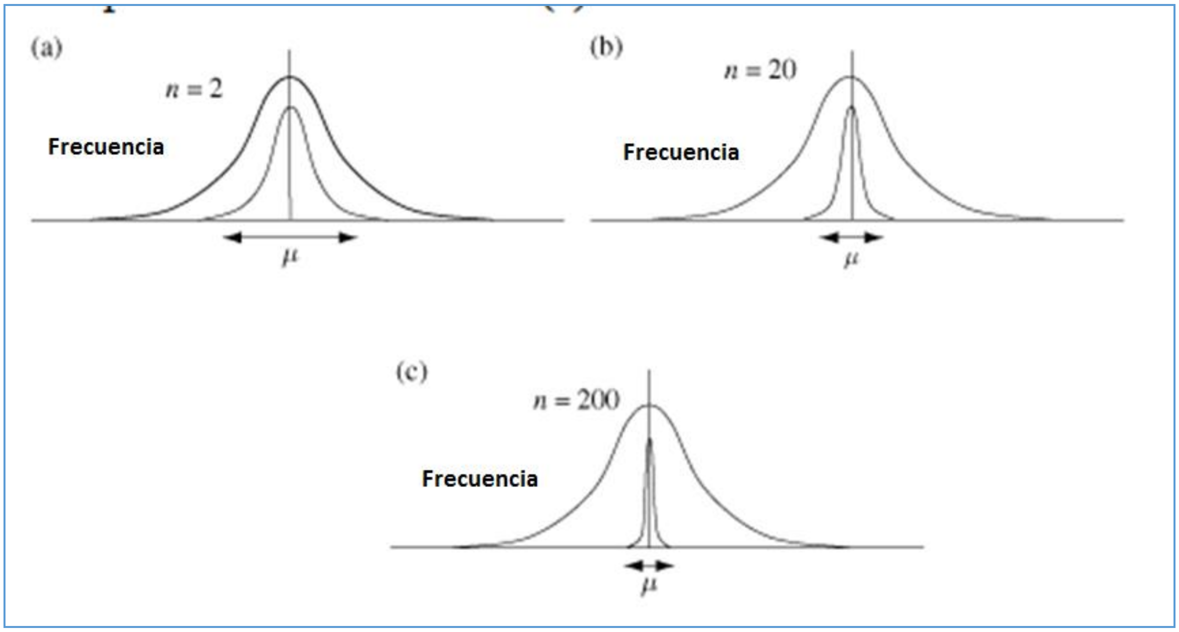

Ilustración 6. Efecto del tamaño de la muestra en la precisión de los valores de X como estimadores de μ. Adaptada "7 Working from samples: data, populations and statistics," por McKillup, S. y Darby, M, 2010, Geoestatistics Explained. An Introductory Guide for Earth Scientists, Copyright 2010 por Steve McKillup y Melinda Darby Dyar.

El teorema del límite central se puede enunciar de la siguiente manera:

Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n tomada de una población con media μ y varianza finita σ^2, entonces la forma límite de la distribución de

$$Z = (\bar{X} - \mu) / (\sigma/\sqrt{n})$$

conforme n tiende a infinito, es la distribución normal estándar n(z;0,1). (Walpone, 2010, p.245).

Inferencia estadística (pruebas de hipótesis).

Una hipótesis estadística es una afirmación o conjetura acerca del valor de un parámetro o parámetros de una población. Tal declaración se considera tentativa pues, a menos que examinemos a toda la población, los verdaderos valores de los parámetros en cuestión se desconocen. Las pruebas de hipótesis pueden mostrar si una declaración tentativa se ve apoyada o rechazada por la evidencia de la muestra.

Para realizar una prueba de hipótesis se pueden seguir los siguientes pasos.

1. **Identificar el patrón de distribución de la población que se ilustra en el problema**

¿Se trata de una distribución discreta o continua? ¿Es una distribución binomial, normal o se sigue algún otro patrón de distribución?

2. **Planteamiento de la hipótesis**

Para realizar una prueba de hipótesis se deben plantear realmente dos hipótesis. La primera de ellas es la hipótesis nula (H_0), una declaración tentativa de que un parámetro de la población es igual a un valor específico.

Por lo regular, la hipótesis nula se plantea de manera que no hay diferencia o cambio en el parámetro de la población, pues el objetivo de la prueba es rechazarla.

Por otro lado tenemos la **hipótesis alternativa (H_1),** una declaración tentativa de que el valor del parámetro de la población tiene un valor diferente al planteado por la hipótesis nula. La hipótesis alternativa se acepta cuando la hipótesis nula se rechaza.

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.

Por ejemplo, puede convenir para determinar si la edad promedio de la población mexicana es de 25 años. Debido a que debe suponerse que no es diferente de 25 años, se tiene la hipótesis nula:

H_0: $\mu = 25$ años

Ante esta hipótesis nula se pueden plantear diversas hipótesis alternativas. Por ejemplo, que el promedio de edad de la población mexicana es de menos de 25 años, es decir:

H_1: $\mu < 25$ años.

Otra alternativa es plantear que el promedio de edad de la población mexicana es mayor que 25 años.

H_1: $\mu > 25$ años.

Finalmente, una tercera opción para la hipótesis alternativa es que el promedio de edad de la población mexicana es distinto de 25 años, o

H_1: $\mu \neq 25$ años.

Los dos primeros planteamientos para la hipótesis alternativa implican una prueba de hipótesis de una cola de la distribución, el tercer planteamiento requiere una prueba de hipótesis de dos colas.

3. Especificar el nivel de significación

Las diferencias entre muestras extraídas de la misma población se deben al azar y rara vez son idénticas.

Supóngase:

H_0: μ = 25 años

Y una hipótesis alternativa

H_1: μ = 22 años

Para probar esta hipótesis, se extrae una muestra y se calcula su media. Sabemos que difícilmente la muestra tendrá una media igual a 25. ¿Qué tanto más pequeña deberá ser una media muestral que la media esperada para justificar o rechazar la hipótesis nula?

La respuesta depende del nivel de error que se desee tolerar, es decir, de la probabilidad de que la muestra haya proporcionado una media lo suficientemente mayor que el valor hipotético debido a factores aleatorios.

El nivel de significación es la probabilidad de rechazar una hipótesis nula verdadera o de cometer lo que se denomina error tipo I. A esta probabilidad se le denomina con la letra α.

Ya que α es la probabilidad de cometer un error tipo I, ¿por qué no seleccionar el menor valor posible? Observa en la ilustración 7 que conforme α disminuye (desplazando la línea roja hacia la línea verde punteada), aumenta la probabilidad de aceptar una hipótesis nula falsa.

El error de no rechazar la hipótesis nula cuando es falsa se denomina error tipo II y por lo general se representa con la letra β.

En la gráfica, el área bajo la curva azul, correspondiente a H_1, aumenta al disminuir el área α bajo la curva roja correspondiente a H_0. El valor de β puede determinarse solamente si la hipótesis alternativa es exacta (de la forma $\mu = X$ y no de la forma $\mu < X$ o $\mu > X$).

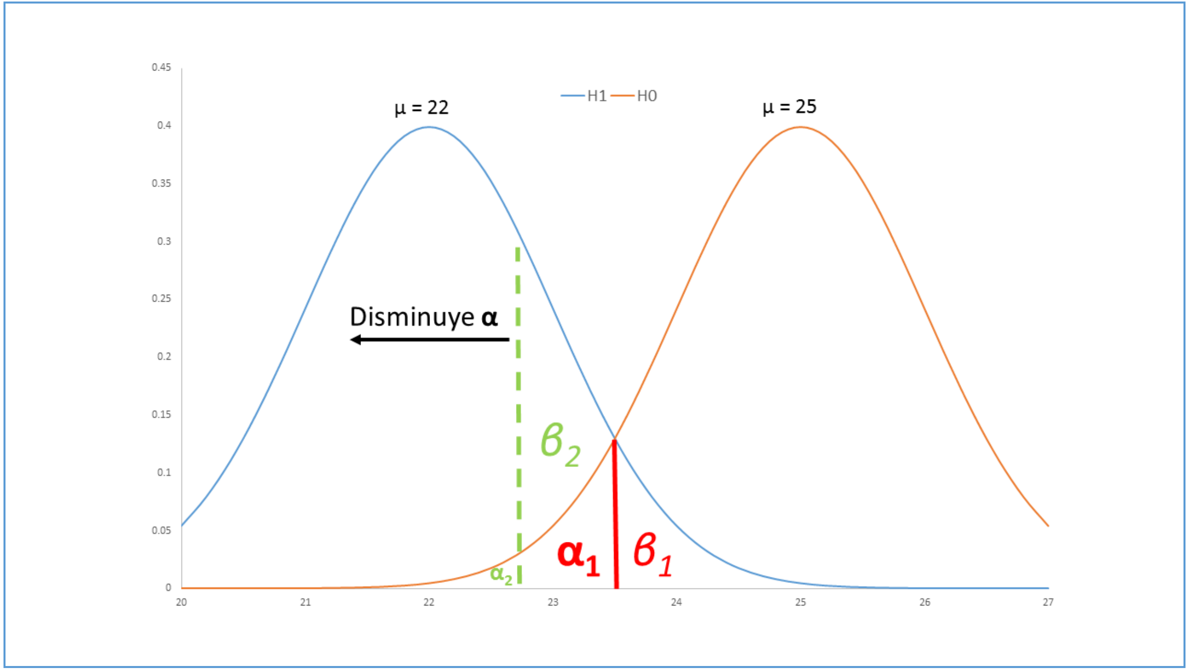

Ilustración 7. Errores tipo I y tipo II.

4. Plantear la regla de decisión

Además del nivel de significación, el criterio de decisión consta de dos factores más: el estadístico de prueba y la región crítica.

Un estadístico de prueba es una variable aleatoria, cuyo valor se utiliza para decidir si se rechaza o no una hipótesis nula. Puede ser un estadístico muestral como la media muestral $\bar{X}$ o alguna otra variable Z.

La región crítica es el conjunto de valores para el estadístico de prueba que llevará a rechazar la hipótesis nula. La región de no rechazo es el conjunto de valores para el estadístico de prueba que provocará la aceptación de la hipótesis nula.

Cuando se utiliza Z como estadístico de prueba, el valor crítico para Z se toma de la tabla normal estándar. Sea Z_α el valor crítico que marca la α inferior bajo la curva normal estándar, un caso típico es α= 0.05 al que le corresponde una $z_{0.05} = -1.645$.

La regla de decisión se establecería de la siguiente forma:

$$Rechazar\ H_0\ si\ \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq\ -1.645$$

En la ilustración 8 se muestran los distintos tipos de regiones de aceptación y rechazo, dependiendo de la prueba de hipótesis que se trate. Nota que en el caso de una prueba de dos colas, como en el inciso c, α se divide en dos colas, cada una con probabilidad α/2.

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.

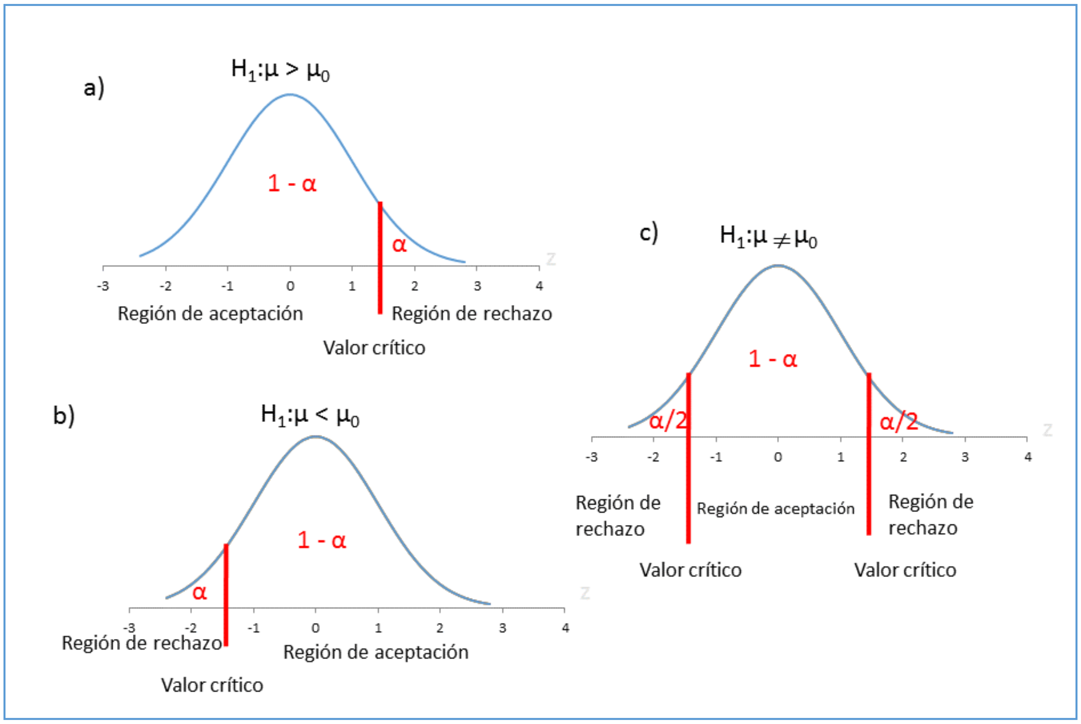

Ilustración 8. Regiones de aceptación y rechazo para distintos tipos de pruebas de hipótesis a) de una cola por el lado derecho, b) de una cola por el lado izquierdo y c) de dos colas.

5. Toma de decisiones

Una vez planteada la regla de decisión para la prueba, se puede calcular el valor del estadístico y compararlo con el valor especificado en la regla de decisión.

Hay cuatro posibles resultados de decisión para cualquier problema de pruebas de hipótesis:

- Cuando la hipótesis nula es verdadera y se rechaza, se trata de un error tipo I.
- Cuando la hipótesis nula es verdadera y deja de rechazarse, se trata de una decisión correcta.

- Cuando la hipótesis nula es falsa y no se rechaza, se trata de un error tipo II.
- Cuando la hipótesis nula es falsa y se rechaza, se trata de una decisión correcta.

La tarea está concluida cuando se decide si se rechaza o no la hipótesis nula.

Referencias

- Chao, L. L. (2006). *Introducción a la estadística.* México: Cecsa.
- McKillup, S. y Darby, M. (2010). *Geoestatistics Explained. An Introductory Guide for Earth Scientists*. United Kingdom: Cambridge University Press.
- Walpole, R. E; Myers, R. H; Myers, S. L. y Ye, K. (2007). *Probabilidad y estadística para ingeniería y ciencias*. México: Pearson Educación.