Decisiones Estadísticas: Bases Teóricas

(Statistical Decision Making: Theoretical Basis)

Badii, M.H. & A. Guillen*

Resumen. Se mencionan las bases de tomar decisiones en estadística. Se presentan una explicación breve de las pruebas de hipótesis con énfasis en intervalo de confianza y poder estadística.

Abstract. The basics of decision making in statistics are noted. A brief explanation of hypothesis testing and error calculation with emphasis on confidence interval and statistical power are described.

Keywords. Hypothesis testing, statistical decisions, statistical power.

Palabras claves. Decisiones estadística, poder estadístico, pruebas de hipótesis.

1. Introducción

En la práctica nos vemos obligados con frecuencia a tomar decisiones relativas a una población sobre la base de información proveniente de muestras. Tales decisiones se llaman *decisiones estadísticas*. Al intentar alcanzar una decisión, es útil hacer hipótesis sobre la población implicada (Badii et al., 2004, Badii et al., 2007a, 2007b). Tales hipótesis, que pueden ser o no cierto se llaman *hipótesis estadísticas*. Son, en general, enunciados a cerca de las distribuciones de probabilidad de las poblaciones (Badii & Castillo, 2007, 2009a). En muchos casos formulamos una hipótesis estadística con el único propósito de rechazarla o invalidarla. Análogamente, si deseamos decidir si un procedimiento es mejor que otro, formulamos la hipótesis de que no hay diferencia entre ellos (o sea, que cualquier diferencia observada se debe simplemente a fluctuaciones en el muestreo de la misma población) (Badii et al, 2007c). Tales hipótesis se suelen llamar *hipótesis nula* y se denotan por H_o (Badii, 1989). Al contrario toda hipótesis que difiera de una dada se llamará una *hipótesis alternativa*.

Los principales objetivos de la *Teoría de Estadística de Decisiones* son:

- **a)** Aprender cómo usar las muestras para decidir si una o unas poblaciones poseen características particulares.
- **b)** Determinar que tan improbable es que una o más muestras observadas haya provenido de una población hipotética.
- **c)** Comprender los tipos de errores posibles que se producen al probar las hipótesis.

d) Aprender cómo la prueba de hipótesis para las diferencias existentes entre medias de población toman diferentes formas.
e) Diferenciar entre muestras independientes y dependientes cuando se comparan dos mediad.
f) Aprender cómo y cuándo usar las distribuciones t y normal para probar hipótesis sobre medias y proporciones de población.

2. Prueba de hipótesis: caso una sola muestra

En la prueba de hipótesis, debemos establecer el valor supuesto o hipotético del parámetro de población antes comenzar a tomar la muestra. La suposición que deseamos probar se conoce como *hipótesis nula*, H_0 (H sub-cero). El término hipótesis nula surge de las primeras aplicaciones agrícolas y medicas de la estadística. Siempre que rechazamos la hipótesis nula, la conclusión que si aceptamos se llama *hipótesis alternativa* y se simboliza como H_a o H_1 (H sub-uno). El propósito de la prueba de hipótesis no es cuestionar el valor calculado de la estadística de muestra, sino hacer un juicio con respecto a la *diferencia* entre esa estadística de muestra y un parámetro de población hipotética. Si suponemos que la hipótesis es correcta, entonces el nivel de significancia indicará el porcentaje d medias de muestra que está fuera de ciertos límites. En la Figura 1 se ilustra cómo interpretar un nivel de significancia de 5%. Observe que 2.5% del área bajo la curva está localizado en cada extremo. De la tabla normal (en cualquier libro de estadística) del apéndice, podemos determinar que 95% de toda el área bajo la curva está incluido en un intervalo que se extiende $1.96\sigma_{\bar{x}}$ de cada lado de la media hipotética. Así pues, en 95% del área, no hay diferencia significativa entre el valor observado de la estadística de muestra y el valor hipotético del parámetro de población. En el restante 5% no existe una diferencia significativa.

En la Figura 2 se examina la misma situación de una manera distinta. Aquí, 0.05 del área bajo la curva cae en la región donde aceptamos la hipótesis nula. Las dos partes sombreadas bajo la curva, que representan un total de 5% del área, son las regiones donde rechazaríamos la hipótesis nula. Se puede ser una advertencia en este punto, incluso si nuestra estadística de muestra cae en la región no sombreada, esto no prueba que nuestra hipótesis nula sea cierta; simplemente no nos proporciona evidencia estadística para rechazarla.

El empleo del término *aceptar*, en vez de *no rechazar*, se ha vuelto de uso común. Significa simplemente que cuando los datos de la muestra no hacen que rechacemos una hipótesis nula, nos comportamos como sí esa hipótesis fuera cierta.

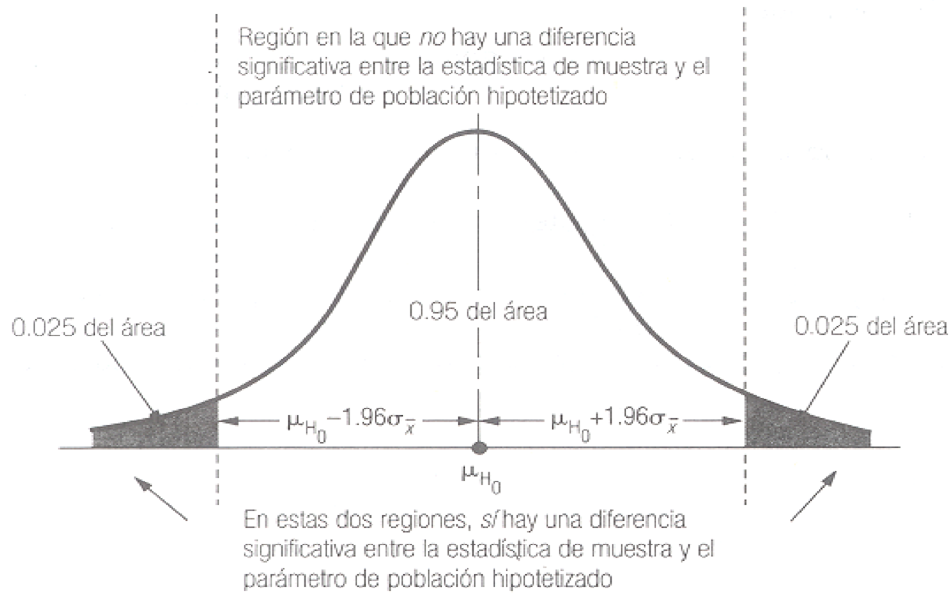


Figura 1. Regiones de diferencia significativa y de diferencia no significativa ($\alpha = 0.05$).

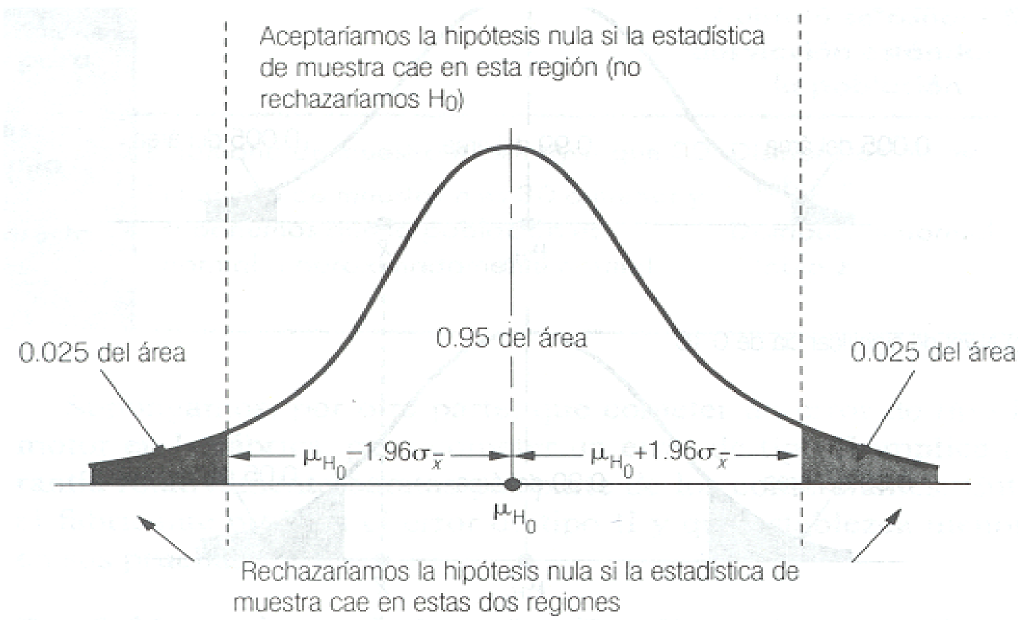


Figura 2. Un nivel de significancia de 5% con regiones designadas de aceptación y de rechazo.

3. Contrastes de hipótesis y nivel de significancia

Si suponemos que una hipótesis particular es cierta pero vemos que los resultados hallados en una muestra aleatoria difieren notablemente de los esperados bajo tal hipótesis, entonces diremos que las diferencias observadas son significativas y nos veríamos inclinados a rechazar la hipótesis nula. Los procedimientos que nos capacitan para determinar si las muestras observadas difieren significativamente de los resultados esperados, y por tanto nos ayudan a decidir si aceptamos o rechazamos una hipótesis, se llaman *contrastes* (*o tests*) de *hipótesis* o de *significación o reglas de decisión*.

No existe un nivel de significancia única estándar o universal para probar hipótesis. En algunos casos, se utilizan un nivel de confianza de 5%. Algunos resultados de investigaciones publicados a menudo prueban hipótesis al nivel de significaría de 1%. Es posible probar una hipótesis a cualquier nivel de significancia. Pero recordamos que nuestra elección del estándar mínimo para una población aceptable, es también el riesgo que asumimos al rechazar una hipótesis nula cuando es cierta. Mientras más alto sea el nivel de significancia, mayor será la probabilidad de rechazar una hipótesis nula cuando es cierta (Badii & Castillo, 2009b).

Presentamos una prueba de hipótesis a tres niveles de significancia diferentes: 0.01, 0.10 y 0.50, mediante la ilustración de la Figura 3. También hemos indicado la ubicación de la misma media de muestra $\bar{x}$ en cada distribución. En las partes *a* y *b*, aceptaríamos la hipótesis nula de que la media de población es igual al valor hipotético. Pero observe en la parte *c*, rechazaríamos esta misma hipótesis nula. ¿Por qué? Nuestro nivel de significancia de 0.50 en esta parte es tan alto que raramente aceptaríamos la hipótesis nula cuando no sea cierta, pero, al mismo tiempo, con frecuencia la rechazaríamos cuando es cierta, en este caso se puede cometer errores de tipo I y Tipo II.

4. Errores de tipo I y de tipo II

El rechazo de una hipótesis nula cuando es cierta se denomina error de tipo I, y su probabilidad (que, como hemos visto, es también el nivel de significancia de la prueba) se simboliza como α (alfa), de manera alternativa, al, hecho de aceptar una hipótesis nula cuando es falsa se le denomina error de tipo II, y su probabilidad se simboliza como β (beta). Existe un equilibrio entre estos dos tipos de errores. La probabilidad de cometer un tipo de error puede reducirse sólo si deseamos incrementar la posibilidad de cometer el otro tipo de error.

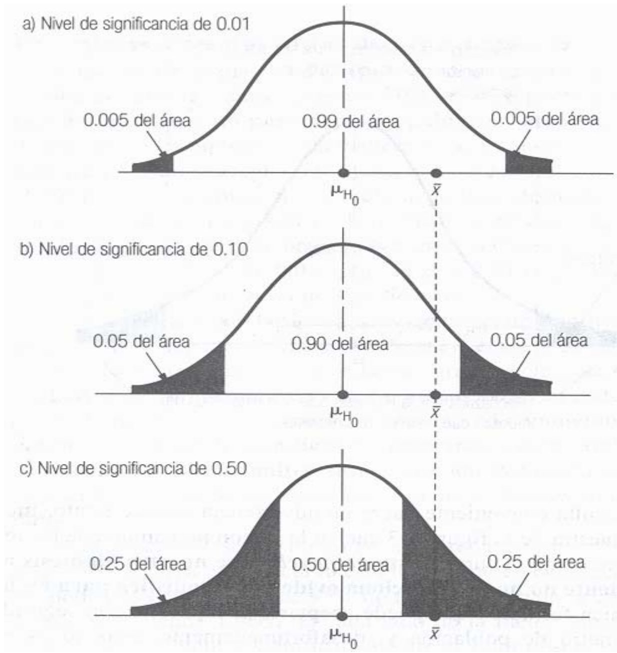


Figura 3. Diferentes niveles de significancia distintos: a) α = 0.01, b) α = 0.10, y c) α = 0.50.

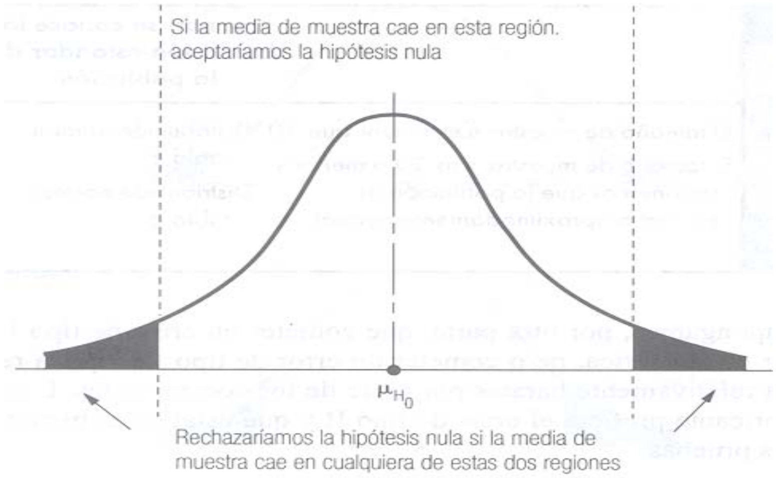
Observamos en la parte c de la Figura 3 que muestra región de aceptación es bastante pequeña (0.50 del área bajo la curva). Con una región de aceptación así de pequeña, rara vez aceptaremos una hipótesis nula cuando no sea cierta, pero como precio de esta seguridad, a menudo realizaremos una hipótesis nula cuando es cierta. Puesto de otra manera, con el fin de obtener una β baja, tendremos que tolerar una α alta. Para lidiar con este equilibrio en situaciones personales y profesionales los responsables de la toma de decisiones deciden el nivel de significancia adecuado, al examinar los costos o desventajas vinculadas con ambos tipos de errores. Para que las reglas de decisión sean buenas, deben diseñarse de modo que minimicen los errores de la decisión. Y no es una cuestión sencilla, porque para cualquier tamaño de la muestra, un intento de disminuir un tipo de error suele ir acompañado de un crecimiento del otro tipo. En la práctica, un tipo de error puede ser más grave que el otro, y debe alcanzarse un compromiso que disminuya el error más grave. La única forma de disminuir ambos a la vez es aumentar el tamaño de la muestra, que no siempre es posible.

5. Prueba de hipótesis de uno y de dos extremos

Una *prueba de dos extremos* de una hipótesis rechazará la hipótesis nula sí la media de muestra es significativamente mayor o menor que la media de población hipotética. Por tanto, en una prueba de dos extremos, existen dos regiones de rechazo (Figura 4).

Una prueba de dos extremos es apropiada cuando la hipótesis nula es $\mu = \mu_{H_0}$ (en donde μ_{H_0} es algún valor especificado) y la hipótesis alternativa es $\mu \neq \mu_{H_0}$.

Figura 4. Prueba de dos extremos de una hipótesis.



Sin embargo, hay situaciones en las que no es apropiada una prueba de dos extremos, por lo que debemos usar una *prueba de un extremo* que se conoce como prueba de extremo izquierdo (prueba de extremo inferior) o prueba de extremo derecho (prueba de extremo superior).

En general, se utiliza una prueba de extremo izquierdo si las hipótesis son H_0: $\mu = \mu_{H_0}$, y H_A: $\mu < \mu_{H_0}$. En una situación semejante, la evidencia muestral con la media de muestra significativamente por debajo de la media de población hipotética es la que nos lleva a rechazar la hipótesis nula a favor de la hipótesis alternativa (Figura 5). Una prueba de extremo superior utiliza cuando las hipótesis son H_0: $\mu = \mu_{H_0}$, y H_A: $\mu > \mu_{H_0}$. Sólo los valores de la media de muestra que están significativamente por encima de la media de población hipotética harán que rechazamos la hipótesis nula a favor de la hipótesis alternativa.
Cabe mencionar que cada prueba de hipótesis, cuando aceptamos una hipótesis nula con base en la información de la muestra, en realidad estamos diciendo que no hay evidencia estadística para rechazarla. No estamos afirmando que la hipótesis nula sea cierta. La única forma de probar una hipótesis nula es conociendo el parámetro de población, y eso no es posible al tomar una muestra. Por consiguiente, aceptamos la

hipótesis nula y nos comportamos como si fuera cierta, simplemente porque no podemos encontrar evidencia para rechazarla.

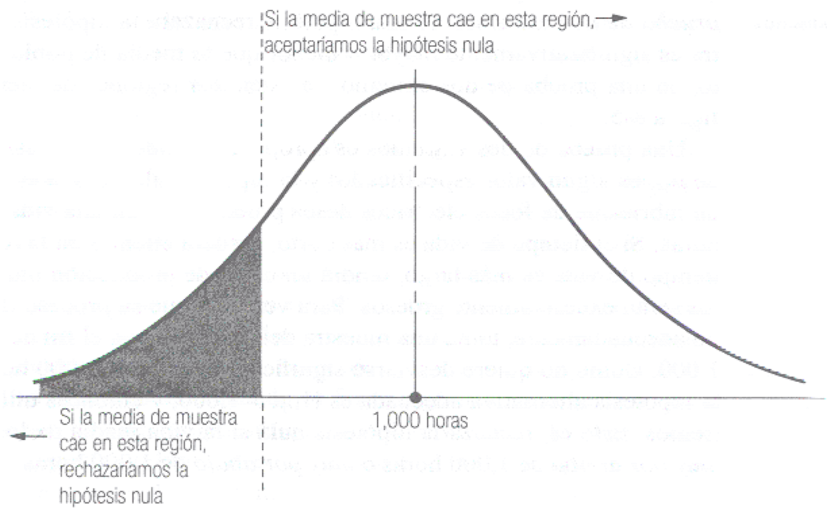


Figura 5. Prueba de extremo izquierdo (prueba de extremo inferior) con la región de rechazo.

6. Prueba de hipótesis de medias: desviación estándar de la población conocida

Un productor de camarones cuenta con producción de camarones cuya la longitud promedio es de 80 milímetros. La larga experiencia del productor indica que la desviación estándar de la longitud es 4.00 milímetros. El productor selecciona una muestra de 100 camarones, las determine la longitud y encuentra una longitud promedio de la muestra igual a 79.60 milímetros. Escritos simbólicamente, los datos en este caso son:

$\mu_{H_0} = 80$ = valores hipotéticos de la media de población
$\sigma = 4.00$ = desviación estándar de la población
$n = 100$ = tamaño de muestra
$\bar{x} = 79.60$ = media de muestra

Si el productor utiliza un nivel de significancia (α) de 0.05 en la prueba, ¿satisfará la producción sus requerimientos de longitud? Para verificar esto se plantea las siguientes hipótesis:

H_0: $\mu = 80$ = hipótesis nula: la media real es 80 mm

H_A: $\mu \neq 80$ = hipótesis alternativa: la media real no es 80 mm

Cuando se conoce la desviación estándar de la población, y como el tamaño de la poblaciones lo suficientemente grande como para considerarlo infinito, podemos utilizar la distribución normal en nuestra prueba. Primero calculamos el error estándar de la media usando la siguiente ecuación:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{4.00}{\sqrt{100}} = 0.4 \text{ mm}$$

De la tabla normal podemos ver que el valor z apropiado para 0.475 del área bajo la curva es 1.96. Ahora podemos determinar los límites de la región de aceptación:

$$\text{LSC} = \mu_{H_0} + 1.96\sigma_{\bar{x}} = 80 + 1.96\ (4.00) = 80.784 \text{ mm}$$
$$\text{LIC} = \mu_{H_0} - 1.96\sigma_{\bar{x}} = 80 - 1.96\ (4.00) = 79.216 \text{ mm}$$

Observe que hemos definido los límites de la región de aceptación (80.784 y 79.216) y la media de la muestra (79.60), y que las hemos ilustrado en la Figura 6 en la escala de la variable original.

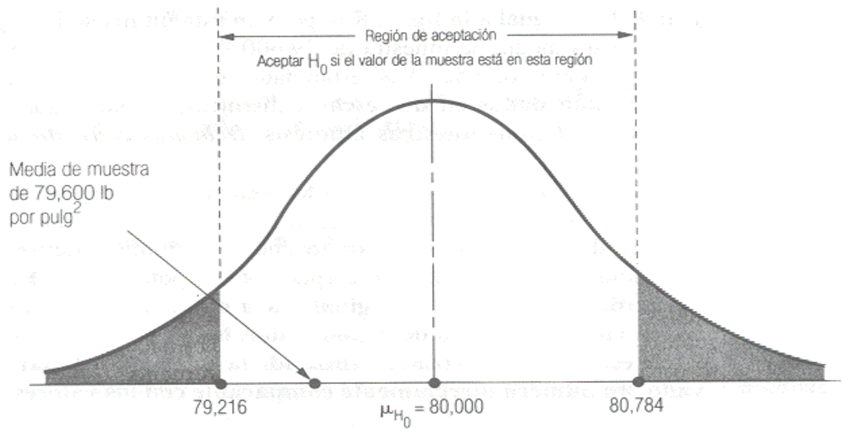


Figura 6. Prueba de hipótesis de dos extremos al nivel de significancia de 0.05.

7. Prueba de hipótesis usando la escala estandarizada

Recordemos nuestro análisis de las variables normales de estandarización, donde en vez de medir la variable en sus unidades originales, la variables estandarizada z nos dice a cuantas desviaciones estándar por arriba (z>0) o por debajo (z<0) de la media se encuentra nuestra observación. La Figura 7 muestra la conversión de nuestro valor observado de $\bar{x}$ a la escala estandarizada, utilizando la ecuación 1 para obtener un valor z observado, un número directamente comparable con los valores críticos z:

$$z = \frac{\bar{x} - \mu_{H_0}}{\sigma_{\bar{x}}} \qquad (1)$$

Donde:

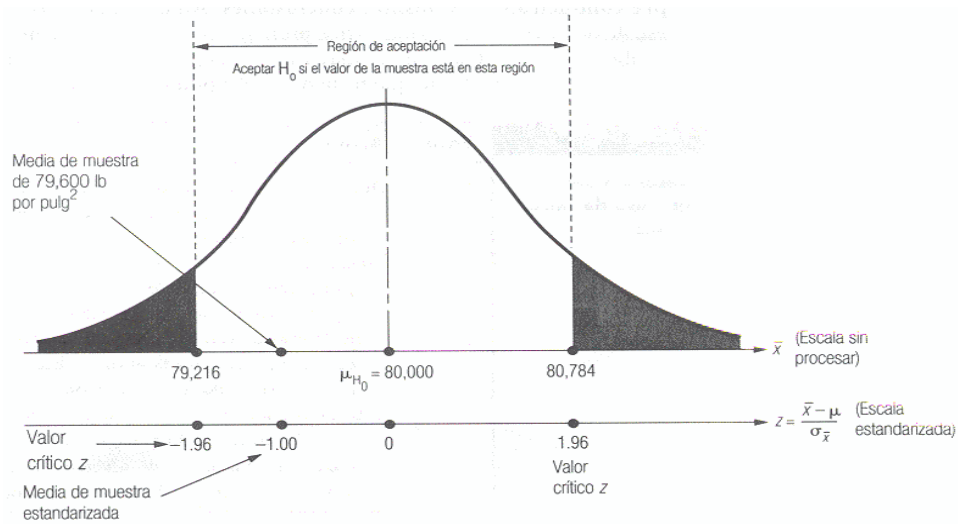
z = media de la muestra (valor estandarizado)

$\sigma_{\bar{x}}$ = error estándar de la media

$$z = \frac{79.60 - 80.00}{0.4} = -1.00$$

En la Figura 7 también hemos ilustrado este valor observado en la escala estandarizada. Observe que el valor cae entre ±1.96 de los límites inferior y superior de la región de aceptación de esta escala. El conjunto *z* fuera del rango ±1.96 se llama la *región crítica de la hipótesis* ó *región de rechazo de la hipótesis*, o *región de significación*. El conjunto de *z* en el rango ±1.96 se conoce como *región de aceptación de la hipótesis o región de no significación.*

Figura 7. Prueba de hipótesis de dos colas al nivel de significancia de 0.05 (la región de aceptación y la $\bar{x}$ de la muestra en las escalas sin procesar y estandarizada).



Como podemos observar el 95% de confianza de que la hipótesis sea verdadera, entonces el valor de *z* para un estadístico muestral estará entre ±1.96. Sin embargo, al escoger una sola muestra al azar hallamos que el valor de *z* de su estadístico está *fuera* de ese rango, debemos concluir que tal suceso podría ocurrir con una probabilidad de sólo 0.05 si la hipótesis dada fuera cierta. Diremos entonces que esta *z* difiere de *forma significativa* de lo que sería de esperar bajo la hipótesis, y se rechaza la hipótesis.

8. Prueba de un extremo de medias

Para una prueba de un extremo de una media, supongamos que un hospital usa grandes cantidades de dosis envasadas de un medicamento particular. La dosis individual de esta medicina es 100 cm^3. La acción del medicamento es tal que el cuerpo tolerará inocuamente dosis excesivas. Por otra parte, las dosis insuficientes no producen el efecto médico deseado. El hospital ha adquirido la cantidad de dicho medicamento durante varios años con una desviación estándar de la población igual a 2 cc. El hospital inspecciona aleatoriamente 50 dosis tomadas de un suministro muy grande y encuentra que la media de estas dosis es de 99.75 cc.

Si el hospital establece un nivel de significancia de 0.10 y nos pregunta si las dosis de esta remesa son demasiado pequeñas, ¿cómo podemos hallar la respuesta?

H_0: $\mu = 100$ = la media de las dosis de la remesa es 100 cc

$H_A = \mu < 100$, la media es menor que 100 cc

Como conocemos la desviación estándar de población, y *n* es mayor que 30, podemos utilizar la distribución normal. Por la tabla z del apéndice, podemos determinar que el valor crítico de z para 40% del área bajo la curva es 1.28. El hospital desea saber si las dosis reales son de 100 cc o si, por el contrario, las dosis son demasiado pequeñas. El hospital debe determinar que las dosis contienen más de una cierta cantidad, o debe rechazar la remesa. Ésta es una prueba de extremo izquierdo, que se muestra gráficamente en la Figura 8. En este caso la región de aceptación consta del 40% en el lado izquierdo de la distribución y todo el lado derecho (50%), para un área total de 90%.

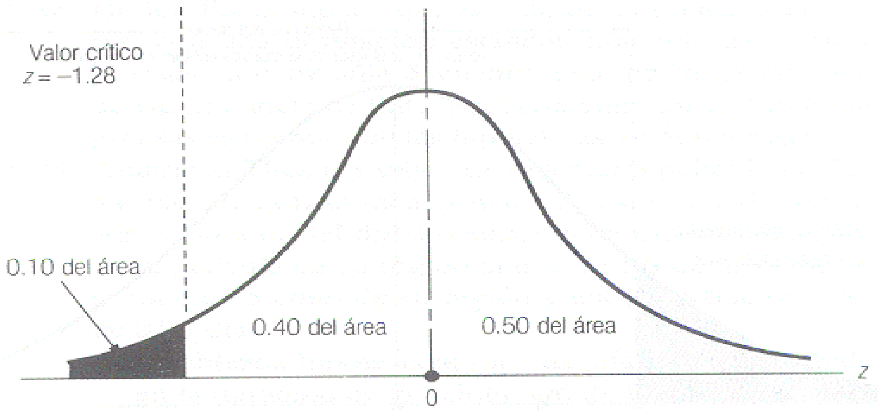


Figura 8. Prueba de hipótesis de extremo izquierdo al nivel de significancia de 0.10.

Para la verificación de hipótesis, calculamos el error estándar de la media:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{50}} = 0.2829 \text{ cc}$$

A continuación determinamos el valor de z para estandarizar la media de la muestra:

$$z = \frac{\bar{x} - \mu_{H_0}}{\sigma_{\bar{x}}} = \frac{99.75 - 100}{0.2829} = -0.88$$

Como se muestra en la Figura 9, al situar el valor estandarizado en la escala z se observa que esta media de muestra cae de lleno en la región de aceptación.

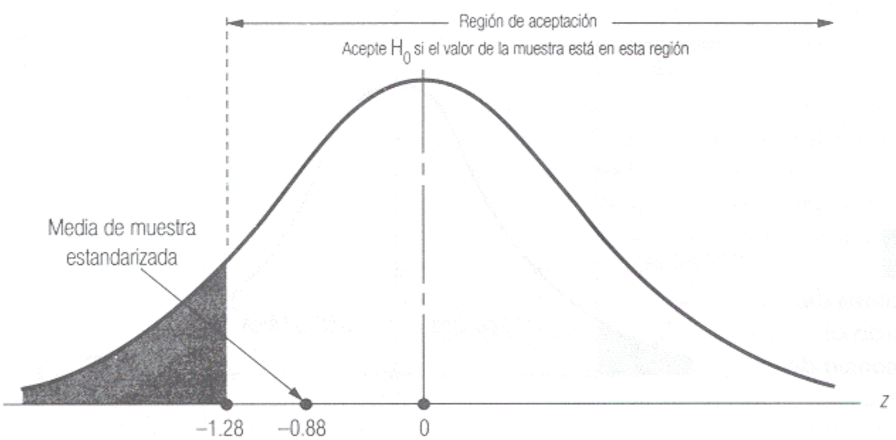


Figura 9. Prueba de hipótesis de extremo izquierdo al nivel α = 0.10.

9. Medición de la potencia de una prueba de hipótesis

Supongamos que la hipótesis nula es falsa. Entonces los investigadores desearían que la prueba de hipótesis la rechazara siempre. Desafortunadamente, las pruebas de hipótesis no pueden ser infalibles; algunas veces, cuando la hipótesis nula es falsa, una prueba no la rechaza, y por tanto se comete un error de tipo II. Cuando la hipótesis nula es falsa, μ no es igual a μ_{H_0}, en vez de esto, μ es igual a algún otro valor. Por cada valor posible de μ para el que la hipótesis alternativa es cierta, hay una probabilidad diferente (β) de aceptar incorrectamente la hipótesis nula. Claro que esta β (la probabilidad de aceptar una hipótesis nula cuando es falsa) fuera más pequeña posible, nos gustaría que 1 - β (la probabilidad rechazar una hipótesis nula cuando es falsa) fuera lo más grande posible.

Puesto que rechazar una hipótesis nula cuando es falsa es exactamente lo que debe hacer una buena prueba, un valor alto de 1 - β (algo cerca de 1.0) significa que la prueba esta trabajando bastante bien (está rechazando la hipótesis nula cuando es falsa); un valor bajo de 1 - β significa que la prueba está trabajando muy mal (no esta rechazando la hipótesis nula cuando es falsa). Puesto que el valor de 1-β es la media de qué tan bien trabaja la prueba, se lo conoce como *potencia de la prueba*. Si representamos gráficamente los valores de 1-β por cada valor de μ para el que la hipótesis alternativa es cierta, la curva resultante se conoce como *curva de potencia*.
Tomando en cuenta la curva correspondiente de la Figura 10, reprodujimos la curva de potencia que está asociada con esta prueba. Hemos calculado los valores de 1-β para representar gráficamente la curva de potencia como indica la Figura 10: b-e.

10. Prueba de hipótesis de porción para muestras grandes: Prueba de dos extremos

Recordamos que la binomial es la distribución teóricamente correcta para usarse al trabajo con porciones, por que los datos son discretos, no continuos. Al aumentar el tamaño de la muestra, la distribución binomial se aproxima a la normal en sus características, y podemos utilizar la distribución normal para aproximar la distribución de muestreo. Específicamente, np y nq cada una debe ser al menos 5 antes de que podamos utilizarla distribución normal como aproximación de la distribución binomial.

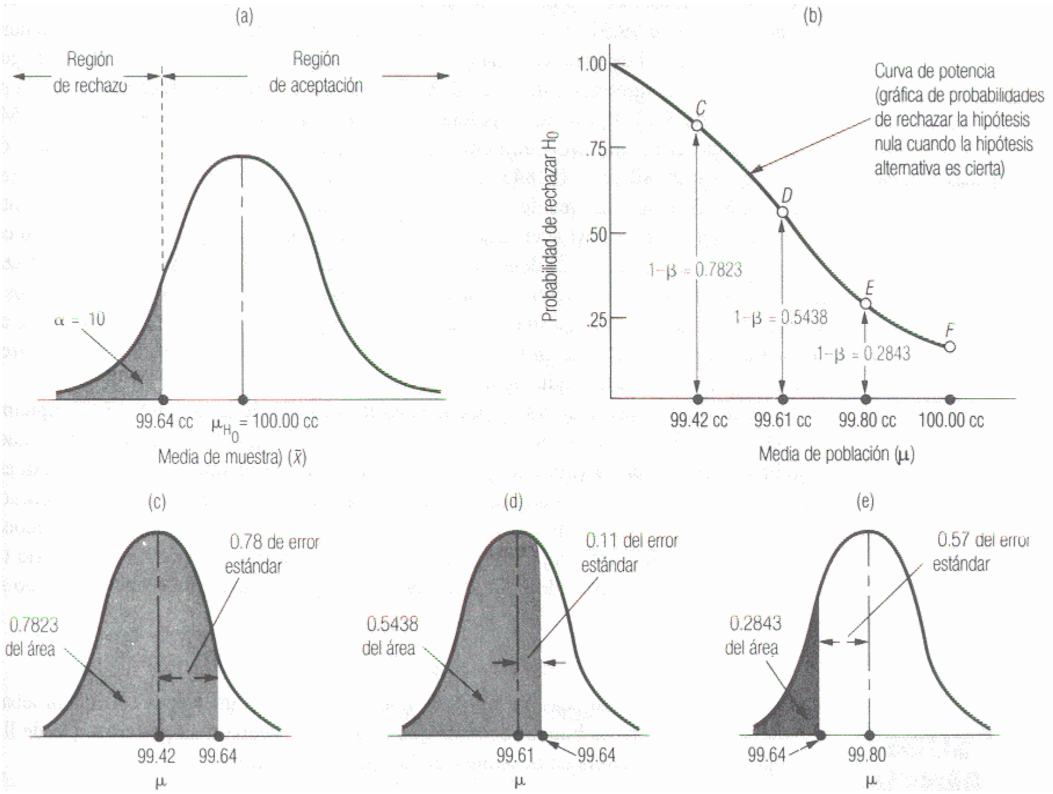


Figura 10. Prueba de hipótesis de extremo izquierdo, curva de potencia asociada y tres valores de μ.

Consideramos, como ejemplo, un investigador de un laboratorio biomédica reporta al director de un hospital que 80% de sus ensayos sobre los pacientes concluyen con éxitos. El director del hospital reúne un comité de especialista para evaluar los éxitos de ensayos efectuados por el investigador. Este comité efectúa entrevistas a 150 pacientes y encuentra que, a su juicio sólo 70% de la muestra está calificada.

El presidente desea probar al nivel de significancia de 0.05 la hipótesis de que 80% de los pacientes adquirieron mejoría en su salud, por lo tanto tenemos:

$p = 0.8$ valor hipotético de la porción de población de éxitos

$q = 0.2$ valor hipotético de la porción de la población de fracaso

$n = 150$ tamaño de muestra
$\bar{p} = 0.7$ porción de muestra de mejorados
$\bar{q} = 0.3$porción de muestra de no mejorados
H_0: p = 0.8 hipótesis nula: 80% de los pacientes mejoraron su estado de salud

H_A: P ≠ 0.8 hipótesis alternativa: la porción de pacientes mejorados no es 80%

Podemos calcular el error estándar de la porción

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.8)(0.2)}{150}} = 0.0327$$

A continuación estandarizamos la porción de muestra

$$z = \frac{\bar{p} - p}{\sigma_p} = \frac{0.7 - 0.8}{0.0327} = -3.06$$

Al señalar la porción de la muestra estandarizada (-3.06), es claro que esta muestra cae fuera de región de aceptación, como se muestra en la Figura 11, por lo tanto se concluye que se debe rechazar la hipótesis nula y concluir que existe una diferencia significativa entre la porción hipotética de éxitos (0.8) del investigador y la porción observada por el director del hospital.

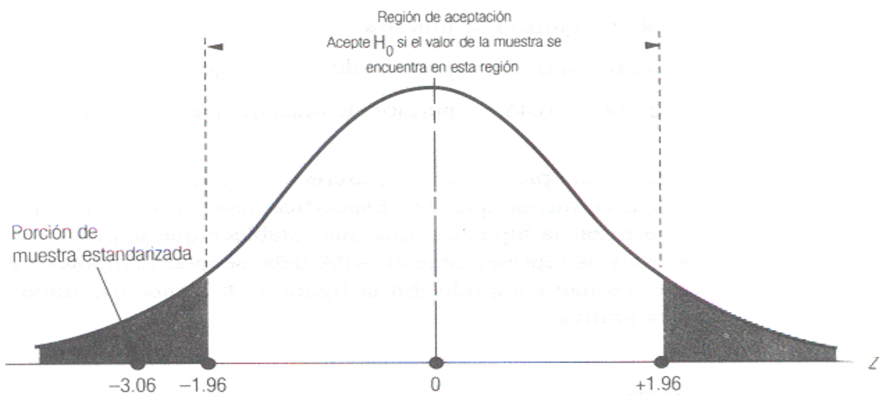


Figura 11. Prueba de hipótesis de dos extremos de una porción al nivel de α = 0.05.

11. Prueba de un extremo de porción

Una prueba de un extremo de una porción es conceptualmente equivalente a una prueba de un extremo de una media. Por ejemplo, la Secretaria del medio ambiente pública que “menos de 60% de las plantas industriales cumple con los estándares de contaminación ambiental”. Sin embargo, un funcionario de Agencia de Protección Ambiental (APA) cree que 60% de las plantas sí cumplen con los estándares. El funcionario decide probar esa hipótesis al nivel de significancia de 0.02. Él tomó una muestra de 60 plantas de una población de más de 10,000 plantas, donde encontró que 33 cumplen con los estándares de contaminación ambiental. ¿Es válida la afirmación del funcionario?

Resumen simbólico

H_0: p = 0.6 hipótesis nula: la porción de plantas que cumplen con los estándares
H_A: p< 0.6 hipótesis alternativa: la porción que cumple con los estándares es<0.6
p = 0.6 valor hipotético de la porción que cumple con los estándares
q = 0.4 valor hipotético que no está cumpliendo con los estándares
N = 60 tamaño de muestra
$\bar{p}$ = 33/60 = 0.55 porción de muestra que no contamina
$\bar{q}$ = 27/60 = 0.45 porción de muestra que contamina

Como *np* y *nq* están cada uno por arriba de 5, podemos usar la aproximación normal de la distribución binomial. El valor crítico de z para (0.5-0.02=0.48) o 0.48 del área bajo la curva es 2.05. A continuación, podemos calcular el error estándar de la porción, usando la porción de población hipotética de la siguiente manera:

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.6)(0.4)}{60}} = 0.0632$$

Y estandarizamos la porción de muestra mediante la aplicación de la siguiente fórmula:

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{0.55 - 0.6}{0.0632} = -0.79$$

Podemos concluir que el funcionario de APA debe aceptar la hipótesis nula de que la porción real de plantas que cumplen es 0.6. Aunque la porción de muestra observada está por debajo de 0.6, no está significativamente por debajo de 0.6, es decir, no está lo bastante debajo de 0.6 para hacer que aceptamos la afirmación del miembro del grupo de interés público.

12. Prueba de hipótesis para diferencias entre medias y porciones

En muchas situaciones de toma de decisiones, le gente necesita determinar si los parámetros de dos poblaciones son parecidos o diferentes. Bajo dicha circunstancia, deseamos estudiar dos poblaciones, la distribución de muestreo que nos interesa, es decir, *la distribución de muestreo de la diferencia entre medias de muestra*. La Figura 12 nos ayuda para conceptuar esta distribución de muestreo particular. En la parte superior de la figura hemos presentado dos poblaciones. Éstas tienen medias μ_1 y μ_2 y desviaciones estándar σ_1 y σ_2, respectivamente. En la parte inferior de la figura se encuentra la distribución de muestreo de la diferencia entre las medias de las muestras.

Las dos distribuciones de muestreo teóricas de la media están construidas a partir de todas las muestras posibles de un tamaño dado que se pueden tomar de la distribución de población correspondiente. Supongamos que se toma una muestra aleatoria de la población 1 y otra de la población 2. La media de la distribución de muestreo de la diferencia entre las medias de las muestras se presenta con el símbolo:

$$\mu_{x_1 - x_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

Si $\mu_1 = \mu_2$, entonces $\mu_{\bar{x}_1} - \mu_{\bar{x}_2} = 0$

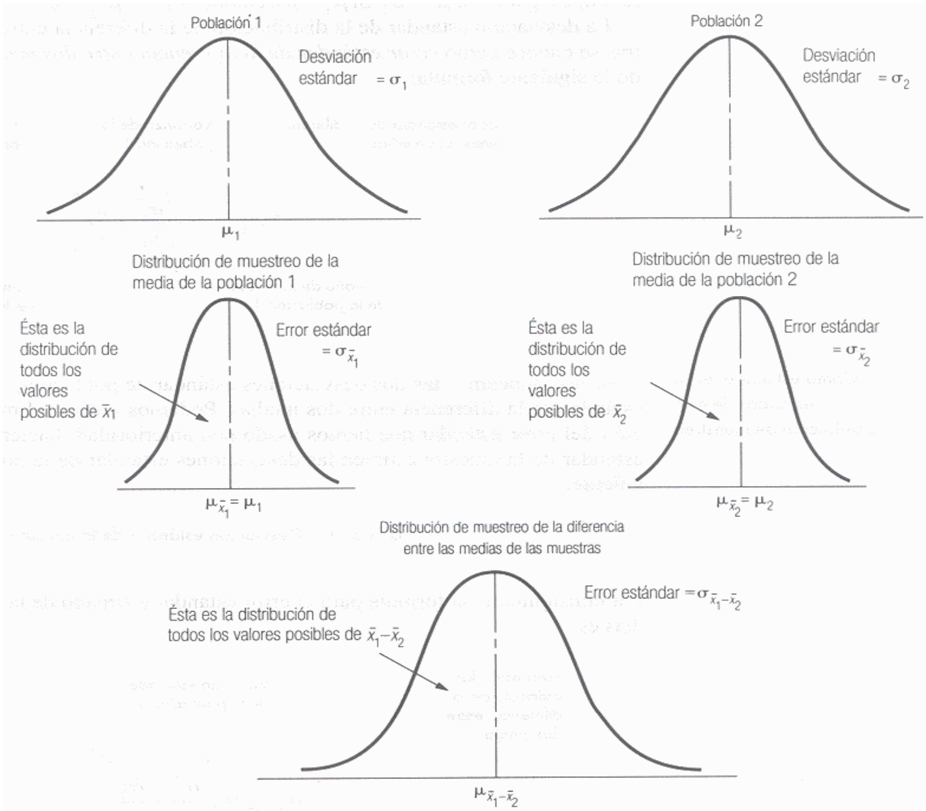


Figura 12. Distribuciones de muestreo de la media y de la diferencia entre las medias de las muestras.

La desviación estándar de la distribución de la diferencia entre las medias de las muestras se conoce como *error estándar de la diferencia entre dos medias* y se calcula mediante la siguiente fórmula:

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (2)$$

Donde:

$\sigma_{\bar{x}_1-\bar{x}_2}$ = error estándar de la diferencia entre dos medias.

σ_1^2 y σ_2^2 = variancia de la población 1 y 2 respectivamente.
n_1 y n_2 = tamaño de la muestra de las poblaciones 1 y 2 respectivamente.

Si no conocemos las dos desviaciones estándar de población, podemos estimar el error estándar de la diferencia entre dos medias. Podemos utilizar el mismo método de estimación del error estándar que hemos usado con anterioridad, haciendo que las desviaciones estándar de la muestra estimen las desviaciones estándar de la población de la manera siguiente: $\hat{\sigma} = s$ Desviación estándar de la muestra. Por consiguiente, la fórmula para el error estándar estimado de la diferencia entre dos medias es:

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \qquad (3)$$

$\hat{\sigma}_{\bar{x}_1-\bar{x}_2}$ = error estándar estimado de la diferencia entre dos medias.

$\hat{\sigma}_1^2$ y $\hat{\sigma}_2^2$ = variancia estimada de las poblaciones 1 y 2 respectivamente.

Ejemplo 1. Los datos de una encuesta sobre salarios por hora son como sigue en la Tabla 1.

Tabla 1. Datos de la encuesta.

Ciudad	Salario/hora	Desviación estándar	Tamaño de la muestra
A	8.95	0.40	200
B	9.10	0.60	175

Calcular el error estándar estimado y el valor estandarizado para la diferencia de las medias de las muestras.

Solución

Cálculo del error estándar estimado: $\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(0.4)^2}{200} + \frac{(0.60)^2}{175}} = \$0.053$$

A continuación, estandarizamos la diferencia de las medias de las muestras, $\bar{x}_1 - \bar{x}_2$.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{z}_1-\bar{x}_2}}$$

$$z = \frac{(8.95 - 9.10) - 0}{0.053} = -2.83$$

Señalamos la diferencia estandarizada en una gráfica de la distribución de muestreo y comparamos con el valor crítico, como se muestra en la Figura 13. En ésta se muestra que la diferencia estándar entre las dos medias de las muestras se encuentra fuera de la región de aceptación. Así pues, rechazamos la hipótesis nula de no diferencia y llegamos a la conclusión de que las medias de las poblaciones son diferentes.

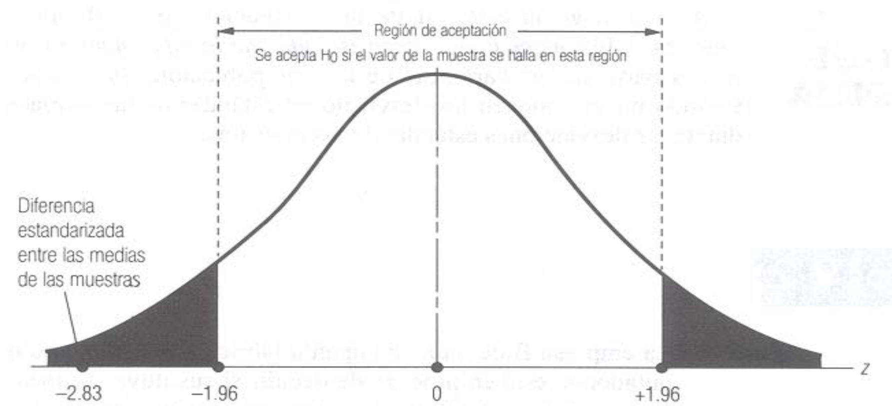


Figura 13. Prueba de hipótesis de dos extremos de la diferencia entre dos medias.

13. Pruebas para diferencias entre porciones: tamaños de muestra grandes

Pruebas de dos extremos

Considere que un laboratorio de farmacología fabrica productos medicinales que está probando nuevos compuestos destinados a reducir los niveles de presión sanguínea. Los compuestos son suministrados a dos diferentes grupos de animales. En el grupo uno, 71 de 100 animales probados respondieron a la droga 1 con niveles menores de presión arterial. En el grupo dos, 58 de 90 animales probados respondieron a la droga 2. El laboratorio desea probar a un nivel de significancia de 0.05 si existe una diferencia entre la eficiencia de las dos medicinas.

$\bar{p}_1 = 0.71$	porción de la muestra de éxitos con la medicina 1
$\bar{q}_1 = 0.29$	porción de la muestra de fracasos con la medicina 1
$n_1 = 100$	tamaño de la muestra para probar la medicina 1
$\bar{p}_2 = 0.644$	porciones de la muestra de éxitos con la medicina 2
$n_2 = 90$	tamaño de la muestra para probar la medicina 2
$\bar{q}_2 = 0.356$	porciones de la muestra de fracasos con la medicina 2
H_0: $p_1 = p_2$	hipótesis nula: No existe diferencia entre ellas
H_A: $p_1 \neq p_2$	hipótesis alternativa: Si existe diferencia entre ellas

Deseamos encontrar el error estándar estimado de la diferencia entre dos porciones, pero antes de realizar esta operación es necesario calcular la *porción completa de éxitos de las poblaciones combinadas de ambas muestras* ($\hat{p}_{común}$):

$$\hat{p}_{común} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} = \frac{(100)(0.71) + (90)(0.644)}{100 + 90} = 0.6789$$

$\hat{q} = 1 - 0.6789 = 0.3211$ Presenta la *porción completa de fracasos de las poblaciones combinadas de ambas muestras*:

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$$

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{(0.6789)(0.3211)}{100} + \frac{(0.6789)(0.3211)}{90}} = 0.0678$$

Estandarizamos la diferencia entre las dos porciones de muestra observadas, $\bar{p}_1 - \bar{p}_2$, dividiéndola entre el error estándar estimado de la diferencia entre dos porciones:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)_{H_0}}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}} \qquad (4)$$

$$z = \frac{(0.71 - 0.644) - 0}{0.0678} = 0.973$$

En la Figura 14 podemos ver que la diferencia estandarizada entre las dos porciones de muestra se encuentra de la región de aceptación. Así pues, aceptamos la hipótesis nula y concluimos que estos dos nuevos medicamentos producen efectos en la presión sanguínea que no son significativamente diferentes.

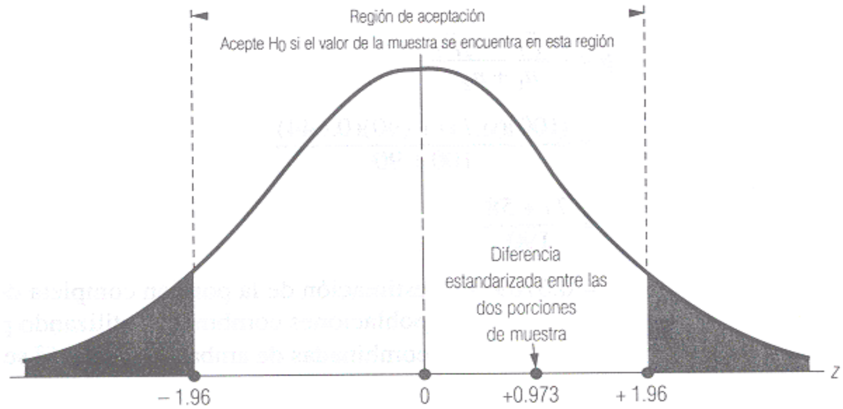


Figura 14. Prueba de dos extremos de la diferencia entre dos porciones (α = 0.05).

14. Prueba de un extremo para diferencias entre porciones

Conceptualmente hablando, la prueba de un extremo para la diferencia entre dos porciones de población es parecida a la prueba de un extremo para la diferencia entre dos medias. Por ejemplo, en un laboratorio de análisis industrial, se determinó la toxicidad de 50 productos tóxicos mediante el método uno y 75 determinaciones a través del método dos. 10% de las determinaciones del método uno y 13.3% del método dos resultaron con errores. Probar la hipótesis de que el método uno de análisis produce una porción menor de error con un nivel de confianza igual al 0.15. Los resultados del muestreo se resumen de la siguiente manera:

$\bar{p}_1 = 0.10$ porción de análisis realizada por el método uno que contienen errores

$\bar{q}_1 = 0.90$ porción de análisis que no contienen errores

$n_1 = 50$ tamaño de muestra correspondiente a los análisis realizados (método 1)

$\bar{p}_2 = .133$ porción de análisis por el método dos que contienen errores

$\bar{q}_2 = .867$ porción de análisis por el método 2 que no contienen errores

$n_2 = 75$ tamaño de muestra (método 2)

H_0: $p_1 = p_2$ hipótesis nula: no existe diferencia entre los dos métodos

H_A: $p_1 < p_2$hipótesis alternativa: el método 1 con una porción menor de error que el método 2.

Para estimar el *error estándar estimado entre dos porciones*, primero utilizamos las porciones combinadas de ambas muestras para estimar la porción completa de éxitos.

$$\hat{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} \qquad (5)$$

$$\hat{p} = \frac{(50)(0.10) + (75)(0.133)}{50 + 75} = 0.12$$

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$$

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{(0.12)(0.88)}{50} + \frac{(0.12)(0.88)}{75}} = 0.0593$$

Ahora se puede calcular la diferencia observada entre las dos porciones de muestra $\bar{p}_1 - \bar{p}_2$, en un valor estandarizado:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}}$$

$$z = \frac{(0.10 - 0.133) - 0}{0.0593} = -0.556$$

Por lo tanto se acepta la hipótesis nula de que no existe diferencia entre los dos métodos de análisis de residuos tóxicos (Figura 15).

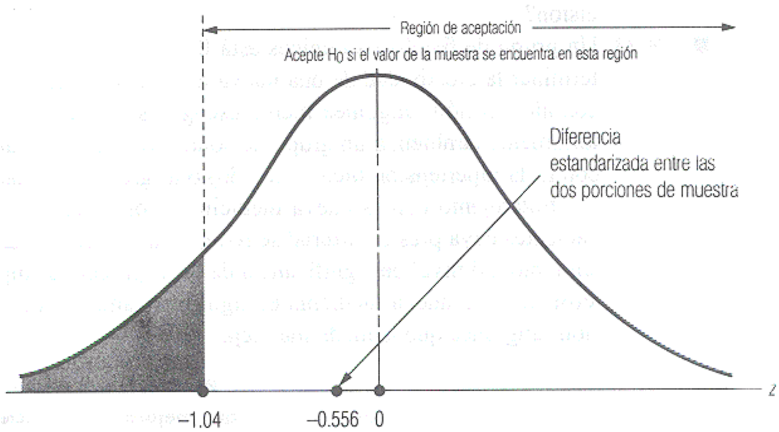


Figura 15. Prueba de hipótesis unilateral de la diferencia entre dos porciones (α = 0.15).

Conclusiones

La vida es una colección de experiencias y decisiones. El punto esencial es tratar de tomar las decisiones apropiadas y correctas. En cualquier tipo de decisión donde se trata de la inversión de algún tipo de recursos (financiero, emocional, estructural, etc.) una persona es sujeta a uno de dos clases de errores. El primero es probable que la persona niega y por tanto rechaza una idea correcta y consecuentemente, cometa el error tipo I, por ejemplo, uno puede rechazar la idea de invertir en algún negocio redituable en donde sí va a tener ganancias. El segundo caso se trata de estar de acuerdo y por tanto aceptar una idea que verdaderamente es falsa, y en este caso la persona comete el error tipo II, aquí contrario al error tipo I, la persona puede involucrarse e invertir su patrimonio en un negocio que lo pueda llevar a una ruina financiera. La ciencia de la estadística se trata de cuantificar las probabilidades de cometer estos dos tipos de errores y por tanto alertar a la persona a intentar de hacer decisiones cruciales con inteligencia y lógica estadística. No se debe olvidar si aceptamos que la lenguaje universal entre diferentes tipos de seres es la matemática, también uno debe tomar con mucha seriedad que el lenguaje de la ciencia es la estadística lo cual se encarga de medir la validez probabilística de todos los fenómenos, procesos, sucesos, objeto, eventos, etc, en la escala del tiempo y el espacio.

Referencias

Anderson, D.R; D.J. Sweeney & T.A. Williams, 1999. Estadística para Administración y Economía. Ed. Thompson Internacional 909 pp. + apéndice.

Badii, M.H. 1989. Ciencia y generación de hipótesis. Boletín de División General de Estudios de Postgrado, UANL. 3(31): 1-2.

Badii, M.H. & J. Castillo. (eds.). 2007. Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H. & J. Castillo. 2009a. Muestreo Estadístico: Conceptos y Aplicaciones. UANL, Monterrey.

Badii, M.H. & J. Castillo. 2009b. Distribuciones probabilísticas de uso común. Daena, 4(1): 149-178.

Badii, M.H., A.R. Pazhakh, J.L. Abreu & R. Foroughbakhch. 2004. Fundamentos del método científico. InnOvaciOnes de NegOciOs, 1(1): 89–107.

Badii, M.H., J. Castillo, R. Rositas & G. Ponce. 2007a. Experimental designs. Pp. 335-348. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H., J. Castillo, J. Rositas & G. Alarcón. 2007b. Uso de un método de pronóstico en investigación. Pp. 137-155. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H., J. Castillos, R. Foroughbakhch & K. Cortez. 2007c. Probability and scientific research. Daena, 2(2): 358-369.

Bamett, V. & T. Lewis, 1978. Outliers in statistical data. Wiley, Nueva Cork.

Box, G.E.P., 1957. Evolutionary Operation: A Method for icreasing industrial productivity. Applied Statistics, vol. 6:81-101.

Daniel, C. 1976. Applications of statistics to industrial experimentation. Wiley, New -York.

Newman, D. 1959. The Distribution of the Range in Simples from a normal population, expressedin termes of an independent estimate of standard deviation. Biometrika, Vol. 31: 20-30.

***Acerca de los autores**

El Dr. Mohammad Badii es Profesor e Investigador de la Facultad de Administración y Contaduría Pública de la U. A. N. L. San Nicolás, N. L., México, 66450. mhbadii@yahoo.com.mx

La Dra. Amalia Guillen es egresada de la Facultad de Administración y Contaduría Pública de la U. A. N. L. San Nicolás, N. L., México, 66450. a_guillen_g@hotmail.com