CAPÍTULO 2-3
LAS PRUEBAS DE HIPÓTESIS[*]

2-3.1 INTRODUCCIÓN A LAS PRUEBAS DE HIPÓTESIS

Existen varios elementos que forman parte de la formulación de un test de hipótesis:

- La población (con más exactitud, los parámetros de la población), la cual corresponde a la realidad desconocida a propósito de la cual queremos probar una hipótesis...
- La hipótesis, la cual es un enunciado con relación a esta población (con más exactitud, con relación a uno o varios parámetros de esta población) y de la cual no se sabe si es verdadera o falsa.
- La muestra, es decir el conjunto de las observaciones obtenidas de la población a partir de las cuales buscaremos decidir[108] si consideramos la hipótesis como verdadera o falsa.

[*] Referencias: Wonnacott y Wonnacott (1992, cap. 9) presentan los tests de hipótesis una vez que presentaron la estimación por intervalos (intervalos de confianza). En estas condiciones, no es posible establecer un paralelismo perfecto entre el manual de estos autores y el presente documento.

[108] Observe que decimos claramente "decidir" y no "determinar". De hecho, para "determinar", necesitaríamos llegar a una certeza. Por lo con-

- La variable-test, la cual es una estadística de la muestra que se usará para decidir si consideramos la hipótesis como verdadera o falsa.
- La probabilidad, la cual, en este contexto, es una medición inversa[109] del grado de incompatibilidad del valor observado de la variable-test con la hipótesis.
- El nivel de significancia, el cual es el umbral de probabilidad crítica abajo del cual se decide que se juzgarán las observaciones (en sus formas resumidas en la variable-tests) lo suficientemente improbables como para ser incompatibles con la hipótesis.

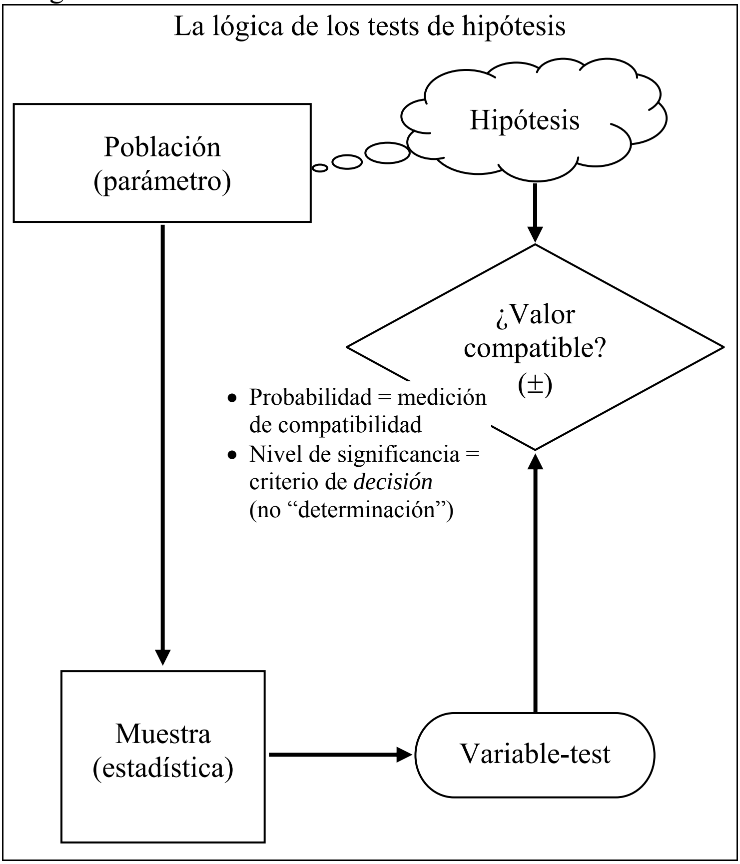
Estos elementos así como las relaciones que los unen se representan en el diagrama 2a.

---

trario, podríamos, en dado caso, "decidir" rechazar una hipótesis sin estar seguro de no cometer un error.

[109] ¡Cuidado con la doble negación! Cuando más grande está la probabilidad, más el valor observado es compatible con la hipótesis, por consiguiente, cuando más pequeña está la probabilidad, más el valor observado tiende a ser incompatible con la hipótesis. La probabilidad es efectivamente una medición inversa de la incompatibilidad.

Diagrama 2a



En caso de pretender formalizar el ejemplo de los dromedarios australianos, podemos decir que:

- la población estudiada es la fauna australiana salvaje;
- la hipótesis para probar es que el número de dromedarios en esta fauna es nulo;
- la muestra se constituye del conjunto de animales observados hasta el momento del test;

- la estadística usada es el número de dromedarios observados.

Pero dejemos de un lado este ejemplo, pues desarrollarlo más no sería muy congruente puesto que lo que hace falta en este ejemplo es la posibilidad de medir el grado de incompatibilidad entre el valor observado de la variable-test y la hipótesis. De manera concreta, en el caso de los dromedarios, es imposible construir un enunciado del tipo "Ubiqué *X* dromedarios hasta el momento. Supongo que el dromedario no es parte de la fauna salvaje australiana (o sea que supongo que los que vi eran animales escapados de un zoológico o de un circo). Si mi suposición es cierta, la probabilidad de observar *X* dromedarios escapados son de *Y* en un millón"... En un test de hipótesis, es necesario poder cuantificar este *Y*.

Durante la formulación de un test de hipótesis, el meollo del problema es la selección de una variable-test. Es tan cierto que varias de las variables-test que se usan con frecuencia llevan el nombre de su inventor (Student,[110] Fisher, Durbin-Watson,...). Una variable-test debe poseer varias propiedades indispensables:

1. El valor de una variable-test depende, al mismo tiempo, de los datos de la muestra y de la hipótesis que se quiere probar.
   En efecto, la variable-test constituye el enlace entre el modelo (la hipótesis que se quiere probar) y las observaciones. Mide, en cierto modo, la distancia o la disimilitud entre las observaciones y las predicciones del modelo o de la hipótesis.[111] La variable-test no tendría ninguna utilidad si no fuera que incorpora la información contenida en la muestra, es decir si su valor fuera

---

[110] Student (estudiante) es seudónimo matemático W.S. Gossett (1876-1937).
http://www-gap.dcs.st-and.ac.uk/~history/Mathematicians/Gosset.html

[111] Es evidente con el Chi-cuadrado de Pearson que se utiliza para el test de independencia en los cuadros de contingencia. Vea 4-1.

independiente de las observaciones. No sería tampoco de gran utilidad si su valor fuera lo mismo no importando la hipótesis específica que se quisiera probar. Más bien, la variable-test debe permitir diferenciar entre las hipótesis que decidimos rechazar y las que decidimos no rechazar.

2. Se debe poder calcular el valor de la variable-test, es decir que no debe depender de valores desconocidos sino, más bien, únicamente de las observaciones y de los datos de la hipótesis.
3. La variable-test debe tener una distribución de muestreo cuya forma generalizada es conocida y cuya forma específica depende del propio contenido de la hipótesis que se quiere probar.

Veremos más tarde con un ejemplo el significado preciso y concreto de esta tercera propiedad. Mientras, examinemos un poco el aspecto aleatorio de la variable-test. De hecho, para tener una distribución de probabilidad (su distribución de muestreo), es necesario que la variable-test sea un variable aleatoria. En cambio, acabado el sorteo de la muestra, los valores observados ya no son aleatorios sino, más bien, fijos (así como el número de dromedarios detectados, una vez que hayan sido contados). La contradicción existe solamente en las apariencias en cuanto recordemos la distinción entre una variable aleatoria y los valores que puede tomar. En efecto, la muestra sorteada es sólo una de las muestras posibles. A cada una de ellas corresponde un valor de la variable-test (es poco probable que otros viajeros o el mismo viajero en otro momento hubieran visto el mismo número de dromedarios). Antes de sortear la muestra, existía por lo tanto una multitud (y en algunos casos, una infinidad) de valores posibles de la variable-test. En otras palabras, imaginando que nos encontramos justo antes del sorteo, entonces la variable-test es, por lo tanto, claramente una variable aleatoria a la cual se asocia una distribución de probabilidad (la distribución de muestreo).

Este concepto no es tan exótico como parece y para demostrarlo, los lingüistas gustan citar este ejemplo de dos títulos de periódico:

Hombre mordido por un perro

y

Perro mordido por un hombre

En los dos casos se emplean las mismas palabras; sólo se modificó un tanto su orden. Entonces, ¿por qué el segundo título es digno de la portada de la sección de policía de *La Prensa*[112] y no el primero? Claro está que el segundo relata un evento sorprendente, sorprendente porque su probabilidad ex ante era muy pequeña.

De la misma manera, consideramos a una persona que acaba de ganar la lotería como una persona con mucha suerte sólo porque, ex ante, la probabilidad de que fuera ella era muy pequeña.

Resumiendo, la distinción entre el valor observado y su distribución ex ante es análoga a la distinción entre lo que efectivamente aconteció y lo que esperábamos. Siendo poetas, podríamos decir que el evento que se realiza no borra el recuerdo de lo que se esperaba de él sino todo lo contrario, la sorpresa nace del choque entre los dos.

Para efectuar un test de hipótesis es necesario, por lo tanto, poder medir la sorpresa, es decir, determinar, suponiendo que la hipótesis que queremos probar sea verdadera, cuál era la probabilidad de observar lo que observamos antes de observarlo (lo cual se resume con la variable-test). Para poder determinar esta probabilidad, se tiene que definir un modelo de muestreo, es decir un modelo de la relación entre la pobla-

[112] Diario de circulación nacional en México.

ción y la muestra. Esto implica que la selección de una variable-test y la especificación del modelo de muestreo van de la mano. El modelo de muestreo contiene, usualmente dos elementos:

- Una hipótesis en cuanto a la forma general de las leyes de probabilidad que rigen el fenómeno estudiado en la población.
- La especificación del proceso de muestreo.

Una vez que se determinó la probabilidad de lo que se observó, siempre y cuando se supone que la hipótesis es verdadera, sólo falta decidir si el resultado conduce o no al rechazo de la hipótesis. Para determinar esto se compara esta probabilidad con el umbral de probabilidad crítica, escogido previamente, abajo del cual pensamos que las observaciones son lo suficientemente improbables para ser incompatibles con la hipótesis. Este umbral crítico se llama nivel de significancia porque es el nivel de probabilidad debajo del cual se decide considerar el desacuerdo entre las observaciones y la hipótesis como estadísticamente significativo.

A fin de quedar con las ideas bien claras, se formaliza el argumento lógico del test de hipótesis clásico en el cuadro que sigue. En este enunciado, los términos en cursivas y entre corchetes son las "variables" del argumento. Para aplicar el argumento a un caso particular, se reemplaza estas variables con los datos pertinentes del caso particular. Por consiguiente, se presenta un poco el argumento como una fórmula matemática donde calculamos el resultado con reemplazar las variables por su valor. Una tabla subsiguiente presentará el "valor" que debe tomar cada "variable" para aplicar el argumento al test de una hipótesis simple sobre un promedio.

Argumento del test de hipótesis clásico

1. Modelo de muestreo, hipótesis y implicaciones (silogismo)

   Si es cierto {*modelo de muestreo*},
   entonces {*variable*}[113] tiene la distribución {*distribución de muestreo*}.

   Ahora bien

   si es cierta {*hipótesis*},
   entonces {*variable*} es igual a la estadística {*variable-test*}.

   Por lo tanto

   si son ciertos a la vez {*modelo de muestreo*} y {*hipótesis*},
   entonces {*variable-test*} tiene la distribución {*distribución de muestreo*}.

2. Regla de decisión: definición de la zona de rechazo
   Se rechazará {*hipótesis*} si el valor observado de {*variable-test*} pertenece a un conjunto de valores extremas cuya probabilidad es inferior o igual a {*nivel de significancia*}.[114]
   Teniendo

   - la distribución {*distribución de muestreo*},
   - la orientación del test (bilateral o unilateral, a la derecha o a la izquierda, dependiendo de la

[113] Esta variable, ni es una estadística, ni es un parámetro. No es un estadística porque su valor depende de parámetros pero, tampoco es un parámetro ya que su valor depende también de una estadística.
[114] Sería más sencillo hablar en términos de la probabilidad del valor observado. Sin embargo, se trata de una variable aleatoria continua lo que implica que la probabilidad de un valor específico es infinitamente pequeña. Ésta es la razón por la cual se razona en términos de un conjunto de valores extremos que se define con uno o dos valores críticos (dependiendo si se hace un test unilateral o bilateral).

hipótesis complementaria H_A),

el conjunto de valores extremos que tiene una probabilidad igual al {*nivel de significancia*} se define con {*zona de rechazo*}.

3. Decisión
   Ahora bien, el valor observado de {*variable-test*} {*forma / no forma*} parte del conjunto de valores extremos que se define con {*zona de rechazo*}.
   Por lo tanto, la regla de decisión seleccionada lleva a {*rechazar / no rechazar*} {*hipótesis*}.

2-3.2 CASO MODELO: UN TEST DE HIPÓTESIS SIMPLE SOBRE UN PROMEDIO

Ahora que expusimos la lógica fundamental de los tests de hipótesis, examinemos cuáles son las etapas a seguir para aplicar esta lógica. Lo haremos basándonos en un caso modelo: un test de hipótesis simple sobre un promedio.

Ejemplo:

queremos estudiar el tiempo que pasan los habitantes de la Isla de Montreal en escuchar la radio. El indicador posible (la variable x) podría ser el número de minutos durante los cuales un individuo escuchó la radio el miércoles 23 de septiembre 1998. El promedio desconocido μ_x podría ser el número promedio de minutos de audiencia radiofónica de los habitantes de la Isla de Montreal ese día.[115] En

[115] Observe que el promedio buscado podría ser, por ejemplo, el número promedio de audiencia radio durante un miércoles cualquiera del periodo del primero de septiembre al 30 de octubre de 1998. Este promedio diferente se refiere, también, a una población diferente. Sin embargo, si todas las observaciones se efectuaron el miércoles 23 de septiembre, la muestra

cuanto a la muestra, supondremos que encierra 25 observaciones y que el tiempo promedio de audiencia de la muestra es igual a 110 minutos, con una desviación estándar $s_x = 20$. La hipótesis que se quiere probar podría ser que, en promedio, los habitantes de Montreal escucharon la radio este día durante cien minutos:[116]

$$H_0 : \mu_x = 100$$

De manera más general, queremos estudiar en una población dada, una característica que se representa con la variable x. Nos interesa μ_x, el promedio de x en la población. Este promedio μ_x es desconocido. Sin embargo, disponemos de una muestra obtenida en la población y podimos calcular m_x, el promedio de x en la muestra. Se trata, ahora, de probar la hipótesis H_0 que el "verdadero" valor del promedio es igual a un valor especifico dado el cual se reconocerá con la letra griega gamma: γ.

$$H_0 : \mu_x = \gamma$$

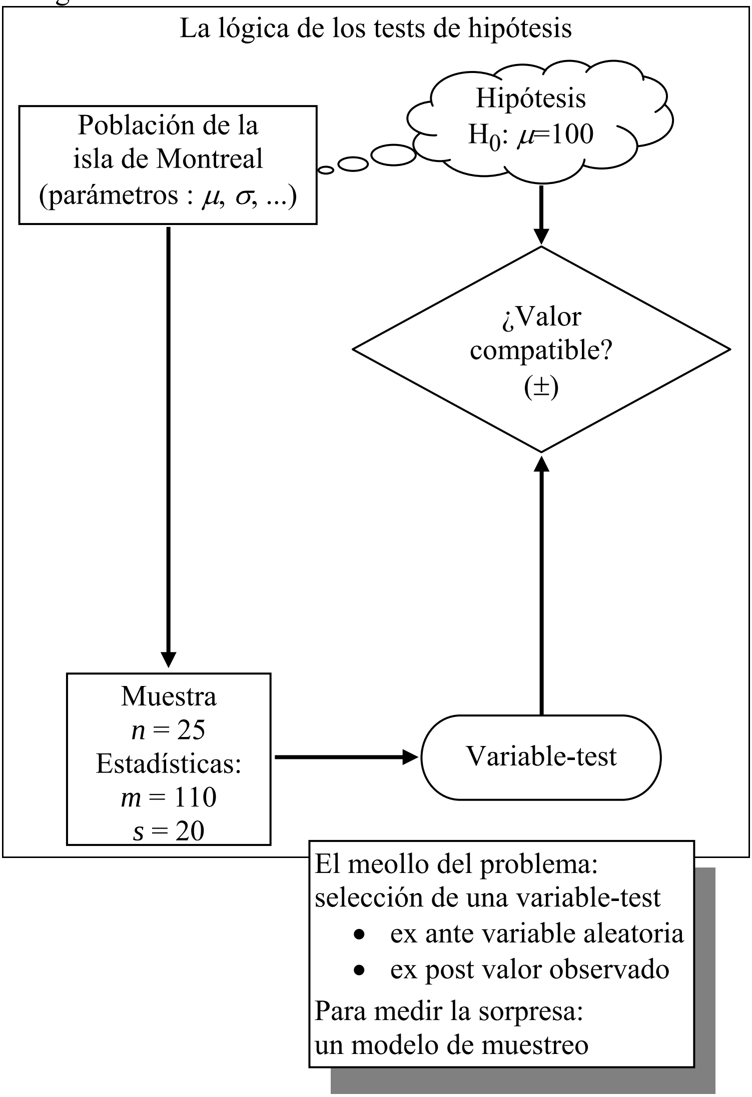
Para complementar la presentación, el diagrama 2b, como anexo, es una copia del diagrama 2a pero con los datos de nuestro ejemplo de test de hipótesis simple sobre un promedio.

---

corre el riesgo de no ser representativa de esta población más amplia, al menos que se crea que los comportamientos son similares durante todos los miércoles del periodo seleccionado (aunque hubo mal tiempo el miércoles, pero no así el 16...).

[116] Está claro que nadie nos impide testar la hipótesis que $\mu_x = m_x = 110$. Sin embargo, esta hipótesis específica no es más que una entre una infinidad de posibilidad.

Diagrama 2b


La lógica de los tests de hipótesis
Población de la isla de Montreal (parámetros : μ, σ, ...)
Hipótesis H_0: μ=100
¿Valor compatible? (±)
Muestra $n = 25$ Estadísticas: $m = 110$ $s = 20$
Variable-test
El meollo del problema: selección de una variable-test
• ex ante variable aleatoria
• ex post valor observado
Para medir la sorpresa: un modelo de muestreo

Las diferentes etapas a seguir para probar una hipótesis son las siguientes:

1. Seleccionar una variable-test;
2. Verificar que el modelo de muestreo asociado a esta variable es aceptable;
3. calcular el valor de la variable-test;
4. seleccionar un nivel de significancia;
5. detectar los valores críticos de la variable-test (zona de rechazo);
6. comparar el valor de la variable-test con los valores críticos y tomar la decisión de rechazar o no la hipótesis

Veamos ahora con más detalle en qué consiste cada una de estas etapas en nuestro ejemplo.

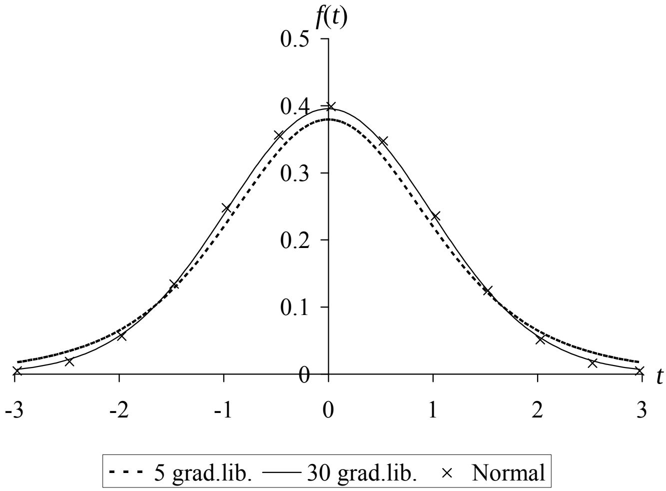
2-3.2.1 Primera etapa: selección de la variable-test

Por el tipo de usuarios que somos, no podemos pensar en inventar por completo una variable-test. Más bien, se trata de seleccionar una entre las que pone a nuestra disposición la estadística. En este caso particular, se aplicará el test de Student, el cual usa la variable-test que sigue:

$$t_{n-1} = \frac{m_x - \gamma}{\left(s_x / \sqrt{n} \right)}$$

La selección de esta variable se justifica puesto que, bajo ciertas condiciones (las examinaremos más tarde), la variable $\frac{m_x - \mu_x}{\left(s_x / \sqrt{n} \right)}$ posee una distribución conocida la cual designamos con el nombre de distribución de Student. La distribución de Student se parece a la distribución normal pero su forma cambia un poco con el valor de *n* tal como lo ilustra la figura 3: Se dice que esta variable posee una distribución de Student con $n - 1$ grados de libertad.[117].

Figura 3
Comparación entre la distribución de Student y la normal según el número de grados de libertad
Función de densidad



Se calcula, por consiguiente, el valor de la variable-test con simplemente sustituir el valor γ con μ_x en la fórmula anterior. De esta manera se puede afirmar que *si la hipótesis H_0 es verdadera*, entonces $\gamma = \mu_x$ y la variable-test posee, una distribución de Student con $n - 1$ grados de libertad:

[117] No obstante, para valores de n superiores a 30, la distribución de Student se asemeja lo suficiente a la normal para considerar, con frecuencia, que la variable posee una distribución aproximadamente normal.

Si H_0 es verdadera, entonces $t_{n-1} = \dfrac{m_x - \gamma}{\left(s_x \big/ \sqrt{n} \right)} = \dfrac{m_x - \mu_x}{\left(s_x \big/ \sqrt{n} \right)}$

Es importante observar que, en fórmula de cálculo de la variable-test t_{n-1}, la desviación estándar que se emplea es efectivamente la desviación estándar de la muestra

$$s_x^2 = \frac{1}{n-1} \sum_i \left(x_i - m_x \right)^2$$

y no la desviación estándar de la población

$$\sigma_x^2 = \frac{1}{n} \sum_i \left(x_i - m_x \right)^2$$

Es posible verificar que esta variable-test posee las cualidades de requisito. Para empezar, su valor depende al mismo tiempo de los datos de la muestra (m_x, s_x, y n) y de la hipótesis que se quiere probar (γ). Luego, este valor no es una incógnita puesto que se puede calcular. Finalmente, esta variable-test posee una distribución de muestreo cuya forma general es conocida (distribución de Student) y cuya formar particular depende de que trata la hipótesis (el promedio μ_x).

2-3.2.2 Segunda etapa: ¿Es aceptable el modelo de muestreo?

Escogimos el test de Student porque, bajo ciertas condiciones, la variable

$$\frac{m_x - \mu_x}{\left(s_x \big/ \sqrt{n} \right)}$$

posee una distribución de Student. ¿Cuáles son, pues, estas condiciones?

Las condiciones que siguen son suficientes[118]

- En la población, la variable x posee una distribución (aproximadamente[119]) normal con un promedio μ_x y una diferencia tipo σ_x desconocido.
- La población es de gran tamaño y en ella se sorteó una muestra aleatoria simple de tamaño n.

Estas condiciones constituyen un modelo de muestreo que especifica la forma general de la distribución de la probabilidad de x en la población y el tipo de muestreo. En cuanto a la distribución de probabilidad de x, puede ser un hecho declarado o una hipótesis dependiendo del contexto. En cuanto al tipo de muestreo, se tomó claramente la decisión al momento de la constitución de la muestra: en una muestra aleatoria simple, cada individuo tenía la misma probabilidad de formar parte de la muestra.

Es responsabilidad del investigador decidir si las condiciones que constituye el modelo de muestreo son aceptables. No se aplica el test sobre el modelo de muestreo, por lo tanto no se cuestionará más en el marco de este test.[120] El test se aplica únicamente sobre la hipótesis H_0.

---

[118] Estas condiciones son suficientes pero no necesarias. En caso que estas condiciones se realicen y que H_0 sea verdadera, entonces la variable-test t_{n-1} tendrá una distribución de Student. Sin embargo, existen otros grupos de condiciones bajo las cuales la variable-test t_{n-1} tendrá también una distribución de Student.

[119] Es imposible que la variable tenga una distribución exactamente normal puesto que no puede tomar valores negativos cuando una variable normal sí puede.

[120] Es cierto que existen tests de "nivel superior", para nombrarlos de alguna manera, que se aplican sobre algunos aspectos del modelo de muestreo. Sin embargo, estos mismos tests se basan en modelos aleatorios más generales, los cuales a este nivel, no se cuestionan. Es posible imaginar un test del modelo de muestreo del test del modelo de muestreo... No obstante, poco importa la "altura" del nivel al cual nos elevamos, siempre existirá en el nivel superior un modelo de muestreo que no se cuestiona.

2-3.2.3 Tercera etapa: cálculo del valor de la variable-test

Una vez seleccionada la variable-test, solo basta calcular su valor al reemplazar los símbolos por su valor numérico.

En nuestro ejemplo del tiempo de audiencia radiofónica, el tamaño de la muestra es de $n = 25$, el promedio de x en la muestra $m_x = 110$ y la desviación estándar $s_x = 20$; la hipótesis que se quiere probar es

$$H_0 : \mu_x = 100$$

Entonces, $$t_{24} = \frac{m_x - \gamma}{\left(s_x / \sqrt{n}\right)} = \frac{110 - 100}{\left(20/5\right)} = 2.5$$

2-3.2.4 Cuarta etapa: selección del nivel de significancia

Debemos, ahora, seleccionar un umbral de probabilidad crítico abajo del cual juzgaremos que las observaciones son lo suficiente improbables como para ser incompatibles con la hipótesis. Los valores que más se emplean en ciencias sociales son 1%, 5% y 10%. Para nuestro ejemplo, se tomará 5%.

2-3.2.5 Quinta etapa: detectar los valores críticos de la variable-test (zona de rechazo)

Siguiendo, consultemos una tabla estadística (vea al final del capítulo la “Tabla de valores críticos del test de Student”). Con esta tabla, nos enteramos que, con $n – 1 = 24$ grados de libertad, existe una probabilidad de 0.05 (o sea, de 5%) que

$$t_{24} < -2.064 \text{ o que } t_{24} > 2.064$$

De manera general, la tabla estadística del t de Student nos entrega los valores críticos $\theta_{n-1}(\alpha)$ para los cuales, con $n – 1$ grados de libertad, existe una probabilidad de α que

$$t_{n-1} < -\theta_{n-1}(\alpha) \text{ o } t_{n-1} > +\theta_{n-1}(\alpha)$$

Nota: Algunos autores emplean la anotación $t_{\alpha,n-1}$ para designar los valores críticos de la distribución de Student;. En esta anotación, α es el nivel de significancia (aquí, 0.05) y $n - 1$ es el número de grados de libertad (aquí, 24). Para evitar cualquier confusión entre los valores críticos y la variable-test misma, evitaremos esta anotación y, más bien, designaremos los valores críticos con $\theta_{n-1}(\alpha)$: aquí en nuestro ejemplo.

$\theta_{24}(0.05) = 2.064$

2-3.2.6 Sexta etapa: comparación del valor de la variable-test con los valores críticos y toma de decisión

En este momento tenemos en nuestras manos todos los elementos necesarios para concluir. Calculamos la variable-test $t_{n-1} = 2.5$. Con la tabla estadística, nos enteramos que *si H_0 es verdadera*, este valor es bastante improbable, es decir que la probabilidad de observar un valor tan alejado de cero es de menos de 5%. Puesto que seleccionamos 5% como nivel de significancia, decidimos rechazar H_0. Esto significa como conclusión que el promedio de x en la población no es igual a 100 porque pensamos que nuestras observaciones son probablemente incompatibles con esta hipótesis.

De manera general, rechazamos la hipótesis, teniendo un nivel de significancia de α, si $t_{n-1} < -\theta_{n-1}(\alpha)$ o $t_{n-1} > +\theta_{n-1}(\alpha)$.

Está claro que si el valor de la variable-test no hubiera rebasado los valores críticos (lo que hubiera podido suceder con otra muestra), no hubiéramos rechazado la hipótesis.

Para resumir, seguimos las etapas siguientes:

1. Seleccionamos una variable-test que tuviera las propiedades de requisitos, es decir el *t* de Student.
2. Examinamos las condiciones bajo las cuales el test de Student se aplica (el modelo de muestreo) y decidimos que eran aceptables.
3. Calculamos el valor de esta variable-test ($t_{24} = 2.5$).
4. Seleccionamos un nivel de significancia ($\alpha = 5$ %).
5. Detectamos los valores críticos en la tabla estadística: si la hipótesis es verdadera, existe una probabilidad α de que la variable caiga al exterior del intervalo definido por los valores críticos $-\theta_{n-1}(\alpha)$ y $+\theta_{n-1}(\alpha)$ (en nuestro ejemplo, la probabilidad de que t_{24} sea inferior a –2.064 o superior a +2.064 es de 5%).
6. Comparamos el valor de la variable-test ($t_{24} = 2.5$) con los valores críticos vistos en la tabla. En nuestro ejemplo constatamos que si la hipótesis fuera verdadera, las observaciones tal como se resumieron en la variable-test hubieran sido improbables (probabilidad inferior a 5%). Y, puesto que esta probabilidad era inferior al nivel de significancia seleccionado, rechazamos la hipótesis.

La tabla que sigue da el “valor” que se necesita atribuir a cada “variable” en el argumento del test de hipótesis clásico con el objetivo de aplicar el argumento al test de una hipótesis simple sobre un promedio.

Aplicación del argumento al test
de una hipótesis simple sobre un promedio

Formulación general	Ejemplo: $n = 25$; $m_x = 110$; $s_x = 20$; $\alpha = 0{,}05$
{*hipótesis*}	
H_0: $\mu_x = \gamma$	H_0: $\mu_x = 100$
{*modelo de muestreo*}	
• En la población la variable x tiene una distribución (aproximadamente) normal, con un promedio μ_x y una desviación estándar σ_x desconocidos. • La población es de gran tamaño y en ella se sorteó un muestra aleatoria simple de tamaño…	
n	25
{*variable*}	
$\dfrac{m_x - \mu_x}{\left(s_x \big/ \sqrt{n} \right)}$	$\dfrac{110 - \mu_x}{\left(20 \big/ \sqrt{25} \right)}$
{*distribución de muestreo*}: distribución de Student con...	
n–1 grados de libertad	24 grados de libertad
{*variable-test*}	
$t_{n-1} = \dfrac{m_x - \gamma}{\left(s_x \big/ \sqrt{n} \right)}$	$t_{24} = \dfrac{110 - 100}{\left(20 \big/ \sqrt{25} \right)} = 2.5$
α ←{*Nivel de significancia*}→	0.05
Orientación del test ⇒ {*zona de rechazo*}: test bilateral	
H_A: $\mu_x \neq \gamma$ $\Rightarrow t_{n-1} < -\theta_{n-1}(\alpha)$ o $t_{n-1} > +\theta_{n-1}(\alpha)$	H_A: $\mu_x \neq 100$ $\Rightarrow t_{24} < -2.064$ o $t_{24} > 2.064$
O test unilateral a la derecha	
H_A: $\mu_x > \gamma \Rightarrow t_{n-1} > +\theta_{n-1}(2\alpha)$	H_A : $\mu_x > 100 \Rightarrow t_{24} > 1.711$
O test unilateral a la izquierda	
H_A : $\mu_x < \gamma \Rightarrow t_{n-1} < -\theta_{n-1}(2\alpha)$	H_A : $\mu_x < 100 \Rightarrow t_{24} < -1.711$

Examinemos nuevamente el criterio de selección del nivel de significancia. ¿Qué hubiera pasado si hubiésemos seleccionado un criterio diferente, 1% por ejemplo? La tabla, en anexo, nos informa que $\theta_{24}(0.01) = 2.797$ es decir que, con n $-1 = 24$ grados de libertad, la probabilidad que $t_{24} < -2.797$ o que $t_{24} > 2.797$ es de 0.01 (o sea 1%). Por consiguiente, si hubiésemos seleccionado un nivel de significancia de 1%, el valor de la variable-test (2.5) se hubiera encontrado en el interior del intervalo delimitado por los valores críticos –2.797 y +2.797. Así que, con este criterio más exigente, no podríamos rechazar la hipótesis.¡Sin embargo, esto no significa tampoco que aceptaríamos la hipótesis!

En términos generales, más pequeño es el nivel de significancia seleccionado, más grande es el valor crítico. En caso de comparar las decisiones que se tomarían con dos umbrales de significación diferentes, es evidente que existen hipótesis, las cuales corresponden a valores de la variable-test, que se encontrarán arriba del valor crítico para el nivel de significancia más elevado pero abajo del valor crítico para el nivel de significancia más exigente (el más pequeño). Tales hipótesis se rechazarían con el nivel de significancia más elevado (menos exigente) pero no con el nivel de significancia más pequeño (más exigente).

La relación entre la selección del nivel de significancia, los valores críticos y la zona de rechazo se ilustra en la figura 4.

Como una síntesis, el diagrama 2c es una copia de la estructura del diagrama 2a pero integra el conjunto de conceptos que acabamos de explicitar, los cuales intervienen en un test de hipótesis.

Figura 4
Test de Student (bilateral)

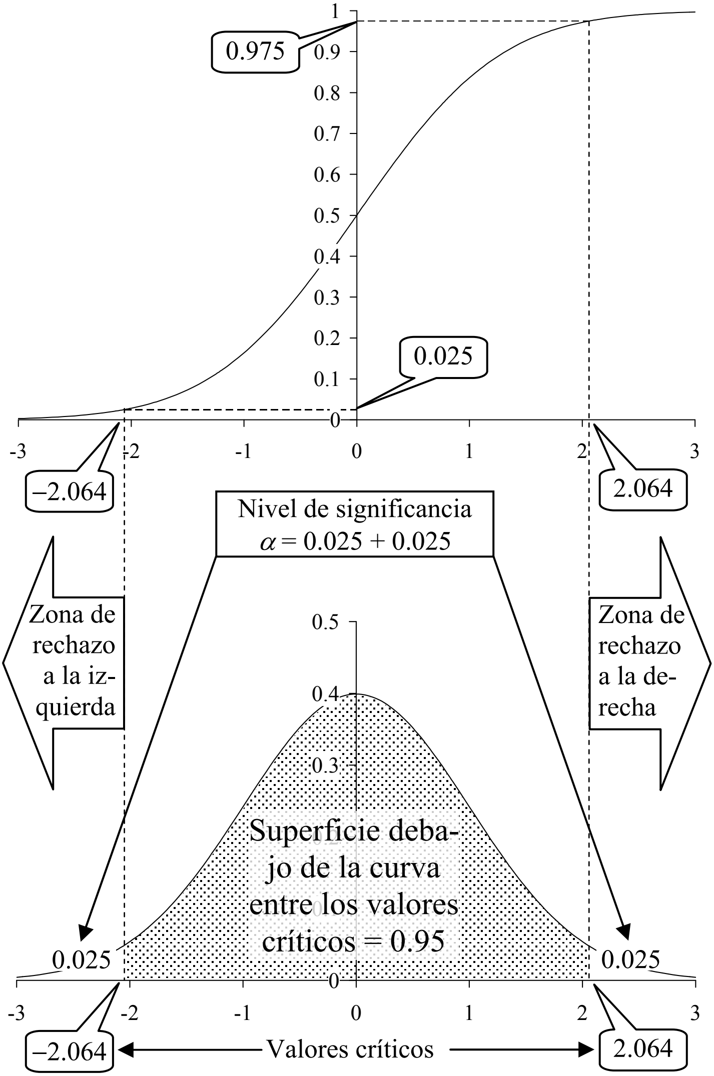
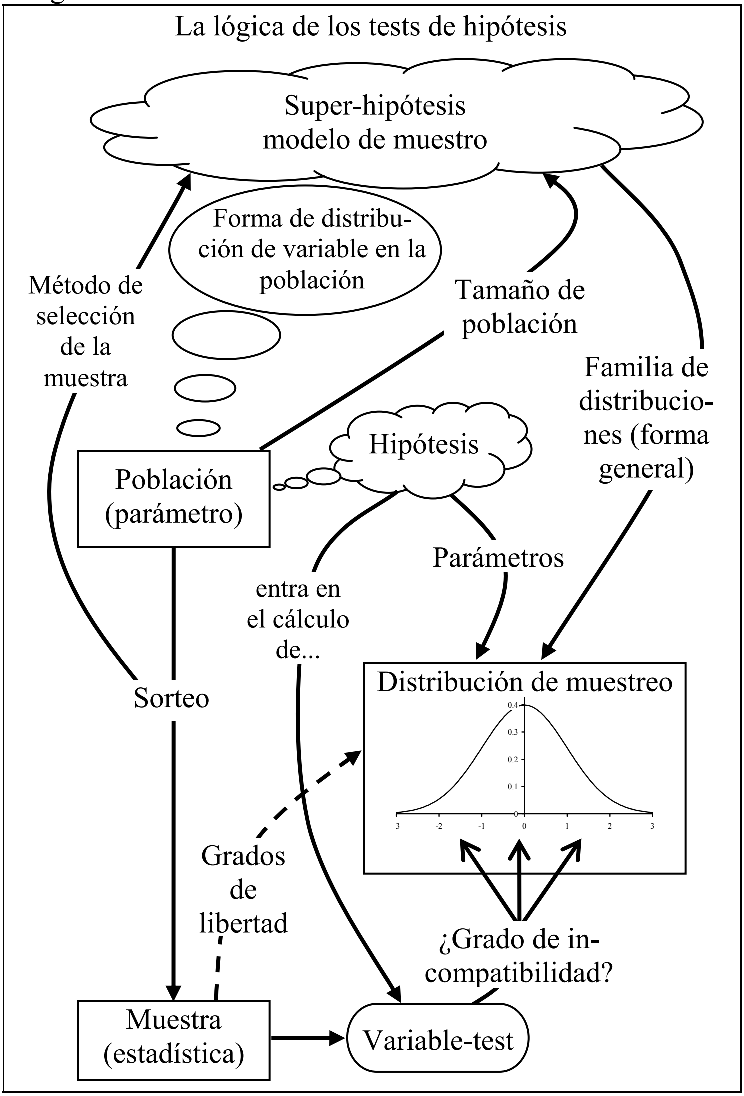


Diagrama 2c


La lógica de los tests de hipótesis
Super-hipótesis modelo de muestreo
Forma de distribución de variable en la población
Método de selección de la muestra
Tamaño de población
Familia de distribuciones (forma general)
Hipótesis
Población (parámetro)
Parámetros
entra en el cálculo de...
Sorteo
Distribución de muestreo
Grados de libertad
¿Grado de incompatibilidad?
Muestra (estadística)
Variable-test

2-3.3 UN POCO DE TERMINOLOGÍA EN RELACIÓN CON LOS TESTS DE HIPÓTESIS*

2-3.3.1 Hipótesis simple, hipótesis compuesta; hipótesis nula, hipótesis complementaria

Acabamos de exponer el proceso de todos los tests de hipótesis simples. Una hipótesis simple es una hipótesis que especifica en su totalidad la distribución de la variable-test: en la práctica, una hipótesis simple atribuye un valor único a un parámetro. Una hipótesis que albarca una serie de valores posibles es una hipótesis compuesta.. Por ejemplo:

- hipótesis simple: $\mu_x = 0$
- hipótesis compuesta: $\mu_x > 0$

En un test de hipótesis simple, la hipótesis que se quiere probar es con frecuencia llamada la hipótesis nula[121] y se designa con H_0. Cuando un test conlleva al rechazo de la hipótesis, implica, lógicamente, que aceptamos la hipótesis complementaria (*alternate hypothesis*) la cual se designa, a menudo, con H_A. La hipótesis complementaria de una hipótesis simple es, por lo general, una hipótesis compuesta. Por ejemplo:

- $H_0 : \mu_x = 0$
- $H_A : \mu_x \neq 0$

* Referencias: Freund (1962), cap. 11.

[121] Según Knapp (1996), esta expresión tiene varias explicaciones: (1) la hipótesis que se quiere testar es, a menudo que el valor del parámetro es nulo; (2) es la hipótesis neutral según la cual nada sale de lo ordinario; (3) con frecuencia, el investigador desea que los datos “anulan” esta hipótesis (por lo personal, considero esta ultima explicación muy ligera).

*2-3.3.2 Nivel de significancia, zona de rechazo y errores del tipo I y II**

Una vez seleccionado el nivel de significancia, el conjunto de los valores de la variable-test cuya probabilidad de realización se encuentra abajo del nivel de significancia se llama la zona de rechazo (o región crítica o zona crítica), del test (vea figura 3).

Un test estadístico se basa en un razonamiento probabilista. Su conclusión no es, por consiguiente, cierta es más bien solamente probable. Siempre habrá ciertos riesgos al momento de tomar decisiones a la luz de un test estadístico.

La tabla que sigue resume los diferentes tipos de errores que se puede cometer:

		Situación (inobservable)	
		H_0 es verdadera	H_0 es falsa
Decisión	Rechazar H_0	Error de tipo I	Buena decisión
	No rechazar H_0	Buena decisión	Error de type II

En cada una de las situaciones posibles, las probabilidades asociadas a estas posibilidades son:

* Referencias: Wonnacott y Wonnacott (1992, p. 344-345, 349-350 y 354-357).

		Situación (inobservable)	
		H_0 es verdadera	H_0 es falsa
Decisión	Rechazar H_0	Nivel de significancia α	Potencia $(1 - \beta)$
	No rechazar H_0	$(1 - \alpha)$	β

En el habla de la estadística, el error del tipo I corresponde al error que se comete cuando se rechaza la hipótesis mientras que, de hecho, era verdadera. La probabilidad de cometer un error del tipo I es la probabilidad de que la variable caiga en la zona de rechazo aunque H_0 sea verdadera. ¿Y cuál es esta probabilidad? ¡Es, por definición, el nivel de significancia seleccionado! Si H_0 es verdadera, la probabilidad α de cometer un error del tipo I es el nivel de significancia seleccionado para el test. La selección del nivel de significancia es, por lo tanto, una selección del nivel que se acepta arriesgar de cometer un error del tipo I en caso que H_0 sea efectivamente verdadera.

Un error del tipo II consiste en aceptar (más bien no rechazar) una hipótesis cuando ésta es falsa. En caso que H_0 sea falsa, la probabilidad β de cometer un error del tipo II es difícil de evaluar dado que existen, en estas condiciones, varias distribuciones posibles para la variable-test. En caso de poder evaluar esta probabilidad, entonces la probabilidad de evitar un error del tipo II $(1-\beta)$ se llama la potencia del test.

En condiciones ideales, desearíamos un test cuyas probabilidades de los dos tipos de errores fueran muy pequeñas (α y β pequeños). Sin embargo, podemos lograr entender de manera intuitiva que, para un test dado, más α es pequeño y más β es grande. En efecto, cuando α es pequeño, la zona de rechazo es pequeña también, lo que aumenta la probabilidad de no rechazar H_0 y, por consiguiente, esto aumenta β. En resumen, la decisión que se toma basándose en un test de hipótesis es una apuesta donde hacemos un compromiso entre dos riesgos de errores: el riesgo de error de tipo I o el riesgo de error de tipo II. Un buen test de estadística es, por lo tanto, un test que, para cualquier nivel dado de probabilidad de error de tipo I, posee la más pequeña probabilidad posible de error de tipo II; en otras palabras, el mejor test es el test más potente para cada nivel de significancia.

2-3.3.3 Distribuciones asintóticas[*]

El modelo de muestreo no siempre permite especificar por completo la distribución de muestreo de una variable-test. A menudo se puede lograr resolver este problema con la distribución asintótica de la variable-test. En efecto, se puede demostrar que varias distribuciones de muestreo tienden a aproximarse de una distribución conocida a medida que el tamaño de la muestra aumenta. Esta distribución conocida se llama una distribución asintótica. Cuando la muestra es lo "suficiente grande" se puede tomar la distribución asintótica como aproximación de la distribución de muestreo exacta.

Por ejemplo, la distribución asintótica de una distribución de Student es la distribución normal (vea figura 3). En este caso particular. No existen verdaderos problemas y se podrá especificar por completo la distribución de muestre de la variable-test para cada valor de número de grados de libertad. No obstante, al momento de rebasar un cierto tamaño de muestra (y de número de grados de libertad que esto implica), se considera que la distribución de Student es tan próxima de la normal que ya no vale la pena referirse a la distribución exacta. En la practica, cuando la distribución de Student tiene

[*] Referencias: Wonnacott y Wonnacott (1992, pp. 224-228).

más de 30 grados de libertad, se estima, con frecuencia, que la muestra es “lo suficiente grande”.

2-3.4 TESTS UNILATERALES (ONE-SIDED TESTS)

El ejemplo que se presentó en el apartado 2-3.2 era un test bilateral (two-side test). En este test, se concede la misma importancia a las desviaciones tanto hacia arriba como hacia abajo con relación a la hipótesis nula.

$H_0 : \mu_x = \gamma$

Sin embargo, en algunas circunstancias, importa sólo una de las dos posibilidades. Por ejemplo, supongamos que un comprador quiera asegurarse que un producto respete una norma de calidad promedio. Digamos que se mide la calidad con un indicador x y que la norma que se debe respetar es que el valor promedio μ_x del indicador de calidad x sea, por lo menos, igual a γ. Para decidir aceptar el lote (la población), el comprador examina una muestra obtenida del lote y calcula la calidad promedio m_x de esta muestra. Está claro que el comprador no se decepcionará si la calidad promedio del producto rebasa la norma. En este caso, la hipótesis complementaria no es $H_A : \mu_x \neq \gamma$ sino, más bien $H_A : \mu_x > \gamma$

Dicho de otra manera, rechazar $H_{0,}$ significa, para el comprador, aceptar el lote, es decir aceptar la hipótesis de que la calidad del lote rebasa la norma. En estas condiciones, la zona de rechazo se sitúa de un solo lado del cero, a la derecha. La lógica es simple: si m_x es lo suficiente grande para que se rechace la hipótesis $H_0 : \mu_x = \gamma$, entonces, con mucha más razón, se rechazará cualquier hipótesis H_0’: $\mu_x = \gamma$ para cualquier valor γ’ < γ. Debemos notar que este razonamiento induce al comprador para aceptar únicamente el lote cuando la calidad promedia de la muestra rebasa la norma con un margen suficiente.[122] En otras palabras, cuando un comprador no acepta el lote, no es porque rechaza la hipótesis que el lote respeta la norma, sino más bien es porque se siente estadísticamente hablando, incapaz de rechazar la hipótesis de que el lote no respeta la norma.

Por otra parte, debemos recordar que la aplicación de un test unilateral cambia la relación entre el nivel de significancia y los valores críticos que definen la zona de rechazo. Por ejemplo, si la zona de rechazo que se usa se define con $t_{24} > 2.064$ al lugar de $t_{24} < -2.064$ o $t_{24} > 2.064$ entonces, la probabilidad de rechazar H_0, mismo si esta hipótesis fuera verdadera, no es de 5% pero de 2.5%. El nivel de significancia de este test unilateral sería, por lo tanto, de 2.5%. Se ilustra esa situación en la figura 5a.

Ejemplo:[123]

Un comprador debe decidir si acepta un lote de 100 000 tubos catódicos. La norma de calidad exigida es que el tiempo de vida promedio de los tubos del lote sea por lo menos de 1200 horas. Se efectúan algunos tests en una muestra de 100 tubos que revelan un tiempo de vida promedio de los tubos de 1265 horas. Teniendo una diferencia tipo de 300 horas, la hipótesis que se quiere probar es

$$H_0 : \mu_x = 1200$$

Se define la estadística t con

$$t_{n-1} = \frac{m_x - \gamma}{\left(s_x / \sqrt{n} \right)} = \frac{1265 - 1200}{\left(300/10 \right)} = 2.17$$

---

[122] Para un test unilateral con un nivel de significancia igual a α, este margen es el margen de error bilateral asociado a un nivel de confianza de $(1-2\alpha)$. Vea 2-3.4.

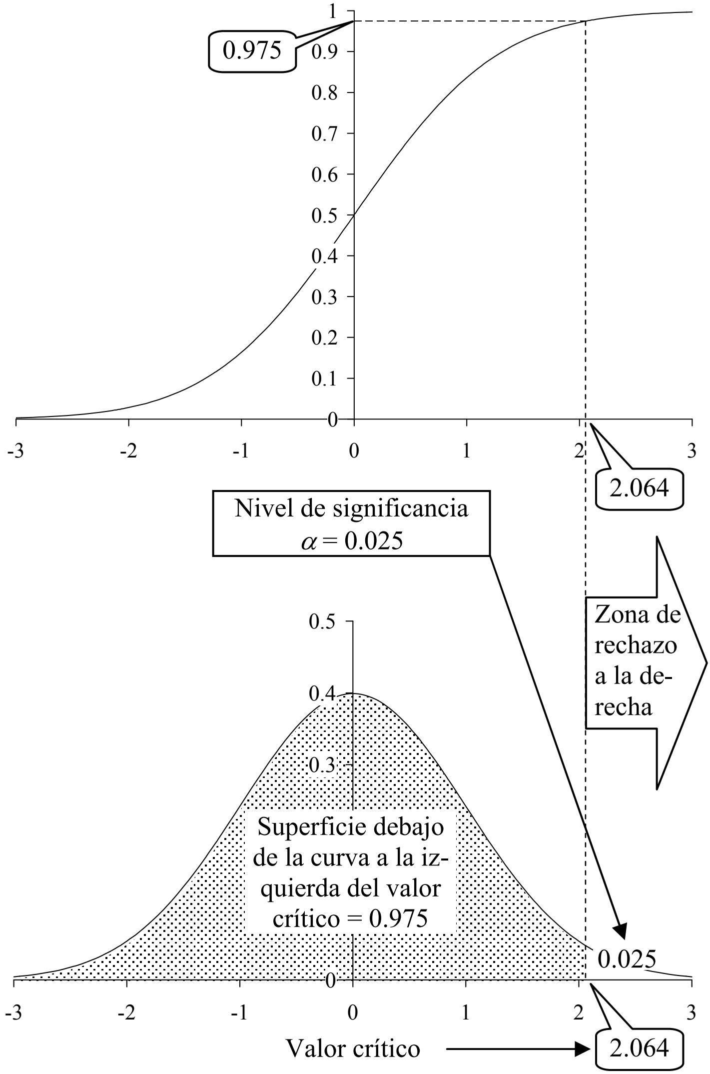
[123] Se toma este ejemplo de Wonnacott y Wonnacott (1991, pp. 333-334).

Si el umbral significación que se escogió es de 0.005, con 99 grados de libertad, el valor crítico de sitúa entre 2.626 y 2.632 (vea la tabla de los valores críticos del test de Student)[124]. No podemos rechazar H_0 y el comprador no aceptará el lote.

En este ejemplo, la zona de rechazo se encuentra a la derecha del cero. Existe, obviamente, circunstancias cuando la zona de rechazo estaría a la izquierda.

[124] Claro está que hubiéramos podido calcular el valor crítico exacto por medio de la función TINV del software Excel de la misma manera que se calculó los valores de la tabla.

Figura 5a
Test de Student (unilateral)



El argumento lógico del test de hipótesis que se formalizó en el cuadro al final del apartado 2-3.1 se aplica también a un test unilateral. Se puede aplicar a un test de hipótesis unilateral a la derecha efectuando en la tabla del apartado 2-3.2 que da el "valor" que es necesario atribuir a cada "variable", la sustitución que sigue:

Para un test unilateral a la derecha ($H_A : \mu_x > \gamma$) con 24 grados de libertad y $\alpha = 0.05$,

$$\{zona\ de\ rechazo\} = t_{n-1} > +\theta_{n-1}(2\ \alpha) = t_{24} > 1.711$$

al lugar de

$$\{zona\ de\ rechazo\} = t_{n-1} < -\theta_{n-1}(\alpha) \text{ o}$$

$$t_{n-1} > +\theta_{n-1}(\alpha) = t_{24} < -2.064 \text{ o } t_{24} > 2.064$$

2-3.5 TEST DE PROBABILIDAD CRÍTICO SIN UMBRAL DE SIGNIFICADO PRE-DETERMINADO (P-VALUE TEST)*

Los tests estadísticos clásicos se efectúan comparando el valor calculado de una variable-test con los valores de referencias que se encuentran en las tablas. Sin embargo, para varias variables-tests que se emplean con frecuencia los paquetes de aplicación de estadística procuran hoy en día el nivel de significancia con el cual el valor de la estadística estaría exactamente al límite de la zona de rechazo.[125] Este nivel de significancia se llama la probabilidad crítica (p-value). En la presentación de resultados, se procura de más en más el valor de esta probabilidad en lugar de indicar si se rechaza o no la hipótesis con el nivel de significancia de 1%, 5% o 10%. Es ésta, una manera de entregar los resultados con un máximo de transparencia lo cual deja el lector libre de escoger el nivel

* Referencias: Wonacott y Wonnacott (1992, pp. 333-337).

[125] En caso de que se trate del t de Student, es posible encontrar su valor con la función del logicial Excel.

de significancia y decidir estar de acuerdo o no con el rechazo de la hipótesis.

En seguida se ilustra el test de probabilidad crítica en las figuras 5b y 5c; luego, se representa su argumento lógico en el cuadro siguiendo el modelo exhibido en el apartado 2-3.1.

Figura 5b
Test de de probabilidad crítica (Student bilateral)

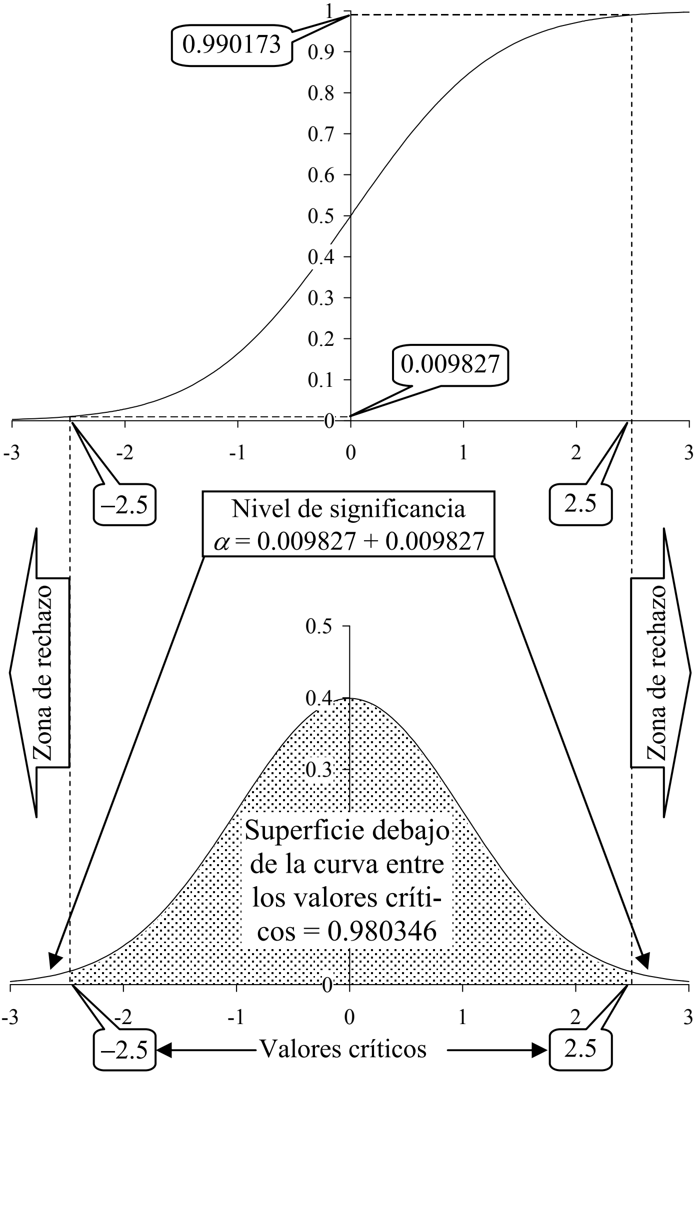
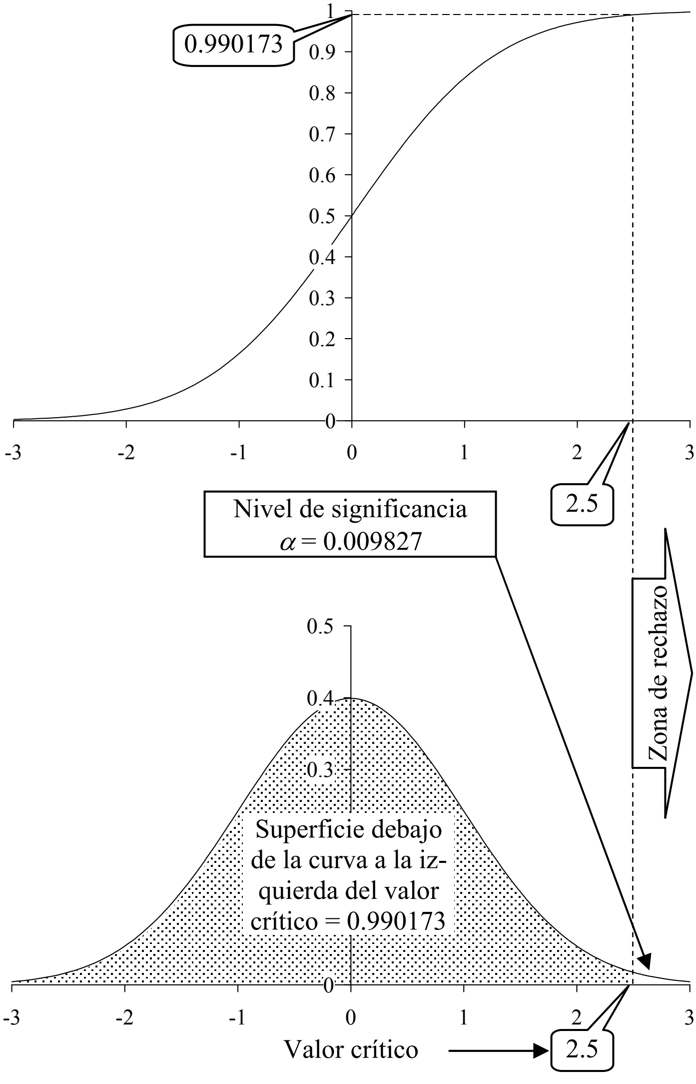


Figura 5c
Test de de probabilidad crítica (Student unilateral)



Argumento del test de probabilidad crítica
(test de hipótesis sin nivel de significancia predeterminado)

1. Modelo de muestreo, hipótesis y implicaciones (silogismo)

   Si es cierto {*modelo de muestreo*},
   entonces {*variable*}[126] tiene la distribución {*distribución de muestreo*}.

   Ahora bien

   si es cierta {*hipótesis*},
   entonces {*variable*} es igual a la estadística {*variable-test*}.

   Por lo tanto

   si son ciertos a la vez {*modelo de muestreo*} y {*hipótesis*},
   entonces {*variable-test*} tiene la distribución {*distribución de muestreo*}.

2. Evaluación de la credibilidad de la hipótesis
   Teniendo

   - la distribución {*distribución de muestreo*},
   - la orientación del test (bilateral o unilateral, a la derecha o a la izquierda, dependiendo de la hipótesis complementaria H_A),

   el conjunto de valores extremos cuyo límite es definido por el valor observado de {*variable-test*} tiene una probabilidad igual a {*probabilidad crítica*}.
   Por lo tanto, si es cierta {*hipótesis*}
   entonces, el valor observado de {*variable-test*} forma parte del conjunto de valores extremos con una probabilidad igual a {*probabilidad crítica*}.

[126] Esta variable, ni es una estadística, ni es un parámetro. No es un estadística porque su valor depende de parámetros pero, tampoco es un parámetro ya que su valor depende también de una estadística.

3. Conclusión:
   Se determina si la {*probabilidad crítica*} {*es* / *no es*} suficientemente pequeña para concluir que las observaciones son con toda probabilidad incompatibles con {*hipótesis*} para, luego, {*rechazar* / *no rechazar*} {*hipótesis*}.

2-3.6 INTERVALOS DE CONFIANZA Y MÁRGENES DE ERRORES (ESTIMACIÓN DEL PROMEDIO)*

En el ejemplo que se exhibió en el apartado 2-3.2, rechazamos la hipótesis que $\mu_X = 100$ con un nivel de significancia de 5%. Podríamos repetir el test con otras hipótesis, con $\mu_X = 101$, $\mu_X = 102$, ..., $\mu_X = 110$, etc. Al efectuar el test para todos los valores posibles, podríamos hacer un inventario de las hipótesis que no se rechazarían (es decir, que serían "aceptables") con un nivel de significancia de 5%. El conjunto de hipótesis que no se rechace con un nivel de significancia dado constituye un intervalo de confianza.[127]

Existe, sin embargo, una solución más directa para llegar al mismo resultado. Se sabe que para cualquier hipótesis posible del tipo $\mu_X = \gamma$, se tendrá la variable-test

$$t_{n-1} = \frac{m_x - \gamma}{\left(s_x / \sqrt{n} \right)}$$

es decir, en nuestro ejemplo,

$$t_{24} = \frac{110 - \gamma}{\left(20/5 \right)}$$

---

* Referencias: Wonnacott y Wonnacott (1992, pp. 286-296).

[127] La terminología estadística tradicional distinguía entre la estimación "puntual" y la estimación "por intervalo", esta ultima refiriéndose a los intervalos de confianza.

Se rechazará todas las hipótesis por las cuales

$$t_{n-1} < -\theta_{n-1}(\alpha) \text{ o } t_{n-1} > +\theta_{n-1}(\alpha)$$

o sea, en el caso de nuestro ejemplo,

$$t_{24} < -2.064 \text{ o } t_{24} > +2.064$$

El total de las hipótesis que NO se rechazaría con un nivel de significancia α se define, por lo tanto, de la manera siguiente:

$$-\theta_{n-1}(\alpha) < \tau_{n-1} < +\theta_{n-1}(\alpha)$$

o sea, en el caso de nuestro ejemplo,

$$-2.064 < t_{24} < +2.064$$

Al reemplazar t_{n-1}, obtenemos

$$-\theta_{n-1}(\alpha) < \frac{m_x - \gamma}{\left(s_x / \sqrt{n} \right)} < +\theta_{n-1}(\alpha)$$

o sea,

$$-\theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right) < \left(m_x - \gamma \right) < +\theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right)$$

$$-m_x - \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right) < -\gamma < -m_x + \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right)$$

$$m_x + \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right) > \gamma > m_x - \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right)$$

$$m_x - \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right) < \gamma < m_x + \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right)$$

o, en nuestro ejemplo,

$$110 - 2.064\left(20/5 \right) < \gamma < 110 + 2.064\left(20/5 \right)$$

$$101.744 < \gamma < 118.256$$

Así que, siempre y cuando γ no forme parte del intervalo

$$\left[m_x - \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right) ; m_x + \theta_{n-1}(\alpha)\left(s_x / \sqrt{n} \right) \right]$$

([101.744; 118.256] en nuestro ejemplo), la hipótesis no se rechazará con un nivel de significancia de α (5%). Esto implica naturalmente que se formule la hipótesis compuesta:

$$\mathrm{C}: \boxed{m_x - \theta_{n-1}(\alpha)\left({s_x}\big/{\sqrt{n}} \right) < \mu_x < m_x + \theta_{n-1}(\alpha)\left({s_x}\big/{\sqrt{n}} \right)}$$

es decir, en nuestro ejemplo, donde $\theta_{n-1}(\alpha) = 2.064$,

$$110 - 2.064\ (20/5) < \mu_x < 110 + 2.064\ (20/5)$$

¿Cuál es la probabilidad de que la condición C sea verdadera? De cierta manera, esta pregunta no tiene sentido puesto que la muestra misma nos da los valores de m_x, s_x y n mientras que μ_x es desconocida pero fija; por lo tanto nada es aleatorio en el enunciado de la condición C. Sin embargo, si imaginamos que estamos justo antes del momento del sorteo de la muestra,[128] sabemos que por cualquier valor fijo pero desconocido de μ_x, existe una probabilidad de α (5% o 0.05 en nuestro ejemplo), que los valores de m_x y de s_x extraídos de la muestra no respeten la condición C. Dicho de otro modo, antes de sortear la muestra, existe una probabilidad de $(1-\alpha)$ de que la condición C se respete (95% o 0.95).

El intervalo

$$\left[m_x - \theta_{n-1}(\alpha)\left({s_x}\big/{\sqrt{n}} \right); m_x + \theta_{n-1}(\alpha)\left({s_x}\big/{\sqrt{n}} \right) \right]$$

se llama un intervalo de confianza cuyo nivel de confianza se define con

$1 - \alpha = 1 -$ nivel de significancia

(en nuestro ejemplo, $0.95 = 1 - 0.05$).

El intervalo de confianza y el nivel de confianza son indisociables. Hablar de un intervalo de confianza sin mencionar su nivel de confianza, es como reportar el resultado “parcial” de un juego deportivo con sólo anunciar el número de goles

[128] Es decir, justo antes de conocer los valores de m_x y s_x.

que contó uno de los dos equipos sin mencionar el número de goles que contó el otro...

De manera paralela, se calcula el margen de error: si se considera m_X como valor estimado de μ_X, diremos que el margen de error es de $\pm\,\theta_{n-1}(\alpha)\left(s_x \big/ \sqrt{n}\right)$ con un nivel de confianza de $1-\alpha$ (en nuestro ejemplo ± 8.256 con un nivel de confianza de 95% o como se acostumbra mencionar en los reportes periodísticos de sondeo, "19 veces de 20"). Tal como el intervalo de confianza, el margen de error pierde todo significado al no ser acompañado de su nivel de confianza.

Puesto que μ_X es fijo, no es una variable aleatoria y su valor no depende de una distribución de probabilidad, no es del todo riguroso afirmar que el valor del parámetro se encuentre en el intervalo de confianza con una probabilidad de 95%. Es, de por si, la razón por la cual la estadística emplea una formulación diferente cuando se trata de "confianza" (probabilidad subjetiva). En cambio, es exacto concluir que, al momento de sortear muestras repetidas de la misma población, la diferencia entre el valor estimado del parámetro y su verdadero valor sería inferior al margen de error[129] en 95% de los casos; es lo que significa el famoso "19 veces de 20" a saber que, en promedio, de 20 muestras diferentes, habría 19 para las cuales no se rebasaría el margen de error.

Es importante notar aquí que el proceso de inducción estadística nos permite formular un enunciado afirmativo en lugar de un no rechazo. Sin embargo, esta afirmación, muy matizada de por sí, se infiere de una lógica de no rechazo: afirmamos que, en un conjunto dado de hipótesis, existe probablemente una que es verdadera y calificamos este "proba-

[129] Este margen es, sin embargo, diferente de una muestra a otra puesto su valor depende de la diferencia tipo de la muestra s_X.

blemente” con una evaluación de la confianza que se dicta de manera afirmativa.

El cuadro de la página siguiente resume el desarrollo que permite definir un intervalo de confianza o un margen de error.

Dos conclusiones se pueden sacar de lo anterior:

1. El ejemplo del promedio muestra con claridad que, mientras más el nivel de confianza seleccionado es alto, más el intervalo de confianza ha de ser amplio y más el margen de error es grande; es decir, cuanto más ganamos en confianza, menos precisión se tiene.
2. Este ejemplo ilustra, también, cómo la precisión de las estimaciones depende del tamaño de la muestra. Cuando se trata de estimar el promedio al momento de aumentar el tamaño de la muestra, el margen de error disminuye con la raíz cuadrada del tamaño de la muestra.[130] La ganancia de precisión es menos que proporcional al aumento del tamaño de la muestra: algo muy parecido a la ley de los rendimientos decrecientes de la economía, trasladada al campo de la estadística.

Presentamos las nociones de intervalos de confianza y de margen de error en el contexto de la estimación del promedio de una variable aproximadamente normal, con la ayuda de una muestra aleatoria simple obtenida de una población de muy gran tamaño. Está claro que estas nociones se pueden aplicar en otras situaciones donde las conclusiones que acabamos de sacar siguen válidas.[131]

---

[130] Hay, también, una ganancia de precisión cuando el número de grados de libertad, asociados al t de Student, aumenta; en la tabla, podemos ver cómo los valores críticos $\theta_{n-1}(\alpha)$ disminuyen cuando el número de grados de libertad aumenta. Sin embargo, a medida que nos aproximamos de 30 grados de libertad, las ganancias son cada vez menores.

[131] Aunque las conclusiones siguen siendo válidas en otras situaciones, es importante recordar que la forma particular de las fórmulas depende del modelo de muestreo que se definió en el apartado 2-3.2.

Intervalos de confianza y márgenes de error

Formulación general	Ejemplo: $n = 25$; $m_X = 110$; $s_X = 20$; $\alpha = 0{,}05$
{*variable-test*}	
$t_{n-1} = \dfrac{m_x - \gamma}{\left(s_x / \sqrt{n}\right)}$	$t_{24} = \dfrac{110 - 100}{\left(20 / \sqrt{25}\right)} = 2.5$
Hipótesis rechazadas con un nivel de significancia de α (5 %)	
$t_{n-1} < -\theta_{n-1}(\alpha)$ o $t_{n-1} > +\theta_{n-1}(\alpha)$	$t_{24} < -2.064$ o $t_{24} > 2.064$
Hipótesis no rechazadas con un nivel de significancia de α (5 %)	
$-\theta_{n-1}(\alpha) < t_{n-1} < +\theta_{n-1}(\alpha)$	$-2.064 < t_{24} < +2.064$
$m_x - \theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right) < \gamma < m_x + \theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right)$	$110 - 2.064\left(20/5\right) < \gamma < 110 + 2.064\left(20/5\right)$ $110 - 8.256 < \gamma < 110 + 8.256$ $101.744 < \gamma < 118.256$
Antes de sortear la muestra e independientemente del valor de m_X, existe una probabilidad de $(1-\alpha)$ (95 %) que se respete la condición C : $m_x - \theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right) < \mu_x < m_x + \theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right)$	
El intervalo de confianza...	
$\left[m_x - \theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right) ; m_x + \theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right)\right]$	$\left[110 - 2.064\left(20/5\right) ; 110 + 2.064\left(20/5\right)\right]$ $[101.744 ; 118.256]$
... y su nivel de confianza	
$1-\alpha = 1-$nivel de significancia	$0.95 = 1 - 0.05$
Margen de error con un nivel de confianza de $(1-\alpha)$ (95 %)	
$\pm\,\theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right)$	$\pm\, 8.256$

2-3.7 Determinación del tamaño requisito de una muestra (estimación del promedio)

¿Cómo determinar el tamaño de la muestra en función del nivel de precisión deseado? En el contexto de la estimación del promedio, observamos que el valor del margen de error depende de la desviación estándar de la muestra s_x, de tal manera que ninguna fórmula permite conocer el grado de precisión mientras no se efectue el sorteo de la muestra. A lo más, se puede determinar el tamaño necesario para que, en los peores de los casos, el margen de error esté aceptable. ¿Pero qué queremos decir con "en los peores de los casos"? Es evidente que el peor de los casos es el caso cuando, en la muestra sorteada, la desviación estándar es la más grande. Examinemos esto en detalle.

Vimos que el margen de error se define con

$$\varepsilon = \pm\theta_{n-1}(\alpha)\left({}^{s_x}\!\big/\!_{\sqrt{n}} \right)$$

Buscamos determinar n, el tamaño de la muestra. Las tablas estadísticas nos procuran lo valores de $\theta_{n-1}(\alpha)$ y el valor de s_x es desconocido en cuanto no se haya sorteado la muestra. Así que éstos son los pasos que se debe seguir:

1. Decidir el margen de error aceptable ε.
2. Escoger el nivel de confianza deseado $(1 - \alpha)$.
3. Detectar en la tabla los valores de $\theta_{n-1}(\alpha)$ para los diferentes tamaños de la muestra n.
4. Formular con relación a s_x la hipótesis del peor, o sea del más grande valor de s_x, que se puede obtener en la muestra.
5. Resolver para n, la ecuación siguiente.

$$\varepsilon = \pm\theta_{n-1}(\alpha)\left({}^{s_x}\!\big/\!_{\sqrt{n}} \right)$$

$$\sqrt{n} = \theta_{n-1}(\alpha)\left({}^{s_x}\!/\!_{\varepsilon} \right)$$

$$n = \left[\theta_{n-1}(\alpha)\left({}^{s_x}\!/\!_{\varepsilon} \right) \right]^2$$

Por ejemplo, supongamos que el margen de error aceptable sea de 10 ($\varepsilon = 10$), el nivel de confianza deseado de 90% ($\alpha = 0.10$) y la hipótesis del peor $s_x = 20$. Tendríamos, entonces:

$$n = \left[\theta_{n-1}(\alpha)\left({}^{s_x}\!/\!_{\varepsilon} \right) \right]^2 = \left[\theta_{n-1}(0.10)\left({}^{20}\!/\!_{10} \right) \right]^2$$

$$n = 4\left[\theta_{n-1}(0.10) \right]^2$$

Puesto que $\theta_{n-1}(\alpha)$ depende de n, se trata de una ecuación en forma implícita. Se puede resolver con aproximaciones sucesivas.

Solución con aproximaciones sucesivas
Podemos iniciar el proceso de aproximaciones suponiendo una muestra de gran tamaño ($n \to \infty$). por el cual la tabla nos da

$$\theta_{\infty}(0,10) = 1.645$$

Entonces

$$n_0 = \left[\theta_{n-1}(\alpha)\left({}^{s_x}\!/\!_{\varepsilon} \right) \right]^2 = \left[\theta_{n-1}(0.10) \right]^2$$

$$n_0 = 4\left[1.645 \right]^2 = 10.8$$

lo que significa que la muestra podría ser más pequeña que el infinito al mismo tiempo que más grande que 10.8. En cambio, con $n = 11$ (11 es el primer número entero superior a 10.8), tenemos

$$\theta_{11-1}(0.10) = 1.812 \text{ y}$$

$$n_1 = 4\left[\theta_{10}(0.10) \right]^2 = 4\left[1.812 \right]^2 = 13.1 > 11$$

lo que significa que la muestra debe ser más grande que 11. Con $n = 13$ (13 es el primer número entero inferior a 13.1). tenemos

$$\theta_{13-1}(0.10) = 1.782 \text{ y}$$

$$n_2 = 4[\theta_{12}(0.10)]^2 = 4[1.782]^2 = 12.7 < 13$$

Esto significa que la muestra podría ser más pequeña. Sin embargo con $n = 12$

$$\theta_{12-1}(0.10) = 1.796 \text{ y}$$

$$4[\theta_{11}(0.10)]^2 = 4 \times (1.796)^2 = 12.9 > 12$$

Por lo tanto, la muestra debe ser más grande.

Conclusión: puesto que 12 no es suficiente y 13 es más que suficiente para obtener el margen de error deseado, se necesita una muestra de tamaño 13.

Se puede verificar el resultado calculando el margen de error:

$$\varepsilon = \pm\theta_{n-1}(\alpha)\left(s_x / \sqrt{n}\right) = 1.782\left(20 / \sqrt{13}\right) = 9.885 < 10$$

2-3.7.1 Caso en que el margen de error aceptable se fija en términos relativos

Por lo general, nos interesa mayormente el margen de error en términos relativos, o sea en fracción del promedio estimado (fracción que expresamos, con más frecuencia, como un porcentaje):

$$\frac{\varepsilon}{m_x} = \pm\theta_{n-1}(\alpha)\frac{\left(s_x / m_x\right)}{\sqrt{n}}$$

Vemos que el método de determinación del tamaño de la muestra requerido es esencialmente lo mismo cuando se desea fijar el margen de error en porcentaje del promedio estimado. La única diferencia es el hecho de formular la hipótesis del peor con relación al coeficiente de variación s_x / m_x, en lugar de con relación a la desviación estándar. La ventaja de este enfoque es de permitir la construcción de una tabla de uso general que da el tamaño de la muestra requerido en función del margen de error relativo aceptable y del coeficiente de variación.

2-3.7.2 Caso en que el promedio buscado es una proporción[*]

Puede ocurrir que no sea fácil formular la hipótesis del peor con relación a la desviación estándar o del coeficiente de varianza. No obstante, existe una clase de situaciones que no requieren de hipótesis: esto sucede al momento de querer estimar una proporción. Por ejemplo, no interesa saber cual proporción de una población es favorable a algún proyecto de planificación urbana. Se realiza un sondeo y se define una variable dicotómica que representa las respuestas a las preguntas sobre el proyecto de planificación.

$x_i = 1$ si el sondeado i es favorable

$x_i = 0$ si el sondeado i no es favorable

En estas condiciones, tenemos

$$m_x = \frac{\sum_i x_i}{n} = \frac{\text{Número de respuestas favorables}}{\text{Número total de respondientes}}$$

El promedio m_x es, por lo tanto, la proporción de personas favorables en la muestra; tal proporción se acostumbra representar con la letra p (por la p e proporción) mejor que con m_x. Se pretende estimar μ_x, la proporción de personas favorables en la población con un cierto margen de error.

Para determinar el tamaño de la muestra requerido, es necesario especificar lo que llamamos la hipótesis del peor.

[*] Referencias: Wonnacott y Wonnacott (1992, pp. 232-240 y 309-311).

Ahora bien, es posible demostrar que, para una variable dicotómica

$$s_x^2 = p\,(1-p)$$

donde el valor de p se encuentra, forzosamente, entre cero y uno. Es posible demostrar también que, para todos los valores de p contenidos entre cero y uno, s_x alcanza su máximo cuando $p = 0.5$ lo cual implica que $s_x^2 = 0.25$ y $s_x = 0.5$. Se resuelve de esta manera el problema para especificar la más grande desviación estándar posible.

> Nota: Puesto que la variable estudiada es una variable dicotómica, no es posible pretender que tenga una distribución normal en la población. Siendo riguroso, esto implica que el test de Student y sus procedimientos respectivos no se aplican en el caso de una proporción. Sin embargo, si se trata con una población de muy grande tamaño y que, en ella, se sortea una muestra aleatoria simple, la estadística matemática nos indica que el test de Student es aproximadamente válido con la condición de que μ_x no sea muy alejado de 0.5.

Obviamente, existen tablas de estadísticas que procuran el tamaño de la muestra requerido en función del margen de error aceptable para diferentes hipótesis emitidas con el más grande valor posible de m_x (o sea de p).

Hay que tener un especial cuidado con no confundir el error relativo sobre un promedio que no es una proporción y el error absoluto sobre una proporción. Por ejemplo, se estima que en promedio el 23 de septiembre de 1998 los habitantes de la Isla de Montreal escucharon la radio durante 120 minutos con un margen de error de doce minutos (con un nivel de confianza de 95%), se calcula un margen de error relativo de 10% (12/120). Por otra parte, si se dice haber estimado al 80%, la población de los habitantes de Montreal que escucharon la radio durante por lo menos diez minutos el 23 de septiembre de 1998, con un margen de error de más o menos 10% (con un nivel de confianza de 95%), hay ambigüedad: ¿Son 10% de 80% o 10% simplemente? Dicho de otra manera, ¿el intervalo de confianza con 95% de nivel de confianza se extiende de 72% a 88% o de 70% a 90%? Es, por lo general, la segunda interpretación correcta, porque hablar de un porcentaje de un porcentaje es un tanto trastornado (además si de hecho el intervalo de confianza se extendiera de 72% a 88%, la empresa de sondeo ganaría mucho con proclamar que su margen de error es de 8% en lugar de decir que es de 10% de 80%).

2-3.8 Otros tests empleados con frecuencia

Hasta el momento, hablamos del test de Student y de un solo uso que corresponde al test de una hipótesis simple sobre una media. Existen otras aplicaciones del test de Student. Por ejemplo, cuando se compara dos muestras, el test de Student sirve a testar la hipótesis que los dos promedio son iguales[132]. En caso de rechazar la hipótesis que los dos promedios son iguales, se rechaza automáticamente que los las dos muestras provienen de la misma población. Tenemos

$$\mathrm{H}_0 : \mu_1 - \mu_2 = \delta$$

En el caso muy particular cuando las dos muestras tienen el mismo tamaño *n*, tenemos:

$$t_{2(n-1)} = \frac{m_1 - m_2 - \delta}{\left(s_1 + s_2 \big/ \sqrt{n} \right)}$$

Por lo general, si las dos muestras son de tamaño n_1 y n_2 respectivamente:

[132] Wonnacott y Wonnacott (1992, pp. 299-307).

$$t_{2(n-1)} = \frac{m_1 - m_2 - \delta}{\left(\frac{\sqrt{\frac{(n_1-1)s_1 + (n_2-1)s_2}{n_1+n_2-2}}}{\sqrt{\frac{n_1+n_2}{n_1 n_2}}}\right)}$$

Otro test empleado con frecuencia es el test de χ^2 (chi al cuadrado). Puede emplearse para, por ejemplo, probar una hipótesis simple sobre una varianza. En efecto, en una muestra aleatoria simple sorteada de una población normal de gran tamaño, la variable

$$\frac{s^2}{\left(\sigma^2 / n-1\right)}$$

posee la distribución del χ^2 con $n - 1$ grados de libertad.[133] El cuadro que sigue procura el "valor" que se debe atribuir a cada "variable" en el argumento del test de hipótesis clásica con el fin de aplicar el argumento al test de una hipótesis sobre la desviación estándar (vea el cuadro ubicado al final del apartado 2-3.1).

Mencionemos también el test F de Fisher, del cual trataremos al momento de estudiar el análisis de regresión. La distribución F de Fisher depende de dos parámetros: el número de grados de libertad del numerador y el número de grados de libertad del denominador (el significado de estas dos expresiones se aclarará un tanto en el contexto de los tests F sobre las regresiones). Puede emplearse para, por ejemplo, probar un coeficiente de correlación simple. En la hipótesis cuando el "verdadero" coeficiente de correlación $\rho = 0$, la variable-test

---

[133] Wonnacott y Wonnacott (1002, cap. 17).

$$\frac{r^2}{\left[\left(1-r^2\right)\big/\left(n-2\right)\right]} = (n-2)\frac{r^2}{1-r^2}$$

posee la distribución F de Fisher con 1 grado de libertad en el numerador y $(n - 2)$ grados de libertad en el denominador. En esta expresión, r es el coeficiente de correlación de la muestra:

$$r = \frac{s_{xy}}{s_x \, s_y}$$

Aplicación del argumento al test de una hipótesis simple sobre la diferencia tipo

Formulación general		Ejemplo: $n = 25$; $m_x = 110$; $s_x = 20$; $\alpha = 0,05$
H_0: $\sigma_x = \gamma$	←{*hipótesis*}→	H_0: $\sigma_x = 70$
	{*modelo de muestreo*} • En la población la variable x tiene una distribución (aproximadamente) normal, con un promedio μ_x y una desviación estándar σ_x desconocidos. • La población es de gran tamaño y en ella se sorteó un muestra aleatoria simple de tamaño…	
n		20
$\dfrac{s_x^2}{\left(\sigma_x^2 / n-1\right)}$	←{*variable*}→	$\dfrac{65^2}{\left(\sigma_x^2 / 19\right)}$
	{*distribución de muestreo*}: Distribución del χ^2 con...	
$n-1$ grados de libertad		19 grados de libertad
	{*variable-test*}	
$\chi^2_{n-1} = \dfrac{s_x^2}{\left(\gamma^2 / n-1\right)}$		$\chi^2_{19} = \dfrac{65^2}{\left(70^2 / (20-1)\right)} = 16.38$
α	←{*Nivel de significancia*}→	0.05
	Orientación del test ⇒ {*zona de rechazo*}: test unilateral a la derecha	
H_A: $\sigma_x > \gamma \Rightarrow \chi^2_{n-1} > \chi^2_{n-1}(\alpha)$		H_A : $\sigma_x > 70 \Rightarrow \chi^2_{19} > 30.144$
	O test unilateral a la izquierda	
H_A: $\sigma_x < \gamma \Rightarrow \chi^2_{n-1} < \chi^2_{n-1}(\alpha)$		H_A : $\sigma_x < 70 \Rightarrow \chi^2_{19} < 30.144$
2α	←{*Nivel de significancia*}→	0.10
	Test bilateral asimétrico	
H_A: $\sigma_x \neq \gamma$ $\Rightarrow \chi^2_{n-1} < \chi^2_{n-1}(1-\alpha)$ o $\chi^2_{n-1} > \chi^2_{n-1}(\alpha)$		H_A : $\sigma_x \neq 70$ $\Rightarrow \chi^2_{19} < 10.117$ o $\chi^2_{19} > 30.144$

Tabla de los valores críticos del test de student (test bilateral)

Grados de libertad	Probabilidad 0.10	0.05	0.01
1	6.314	12.706	63.656
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750
40	1.684	2.021	2.704
50	1.676	2.009	2.678
60	1.671	2.000	2.660
70	1.667	1.994	2.648
80	1.664	1.990	2.639
90	1.662	1.987	2.632
100	1.660	1.984	2.626
∞	1.645	1.960	2.576

Fuente: Valores calculados con la ayuda de la función TINV del logicial Excel.