Menú principal

Contraste de hipótesis

Esquema

---

1. Contraste de hipótesis

Una hipótesis estadística es una afirmación sobre alguna característica desconocida de la población. Probar una hipótesis estadística consiste comprobar si hay suficiente evidencia empírica para sustentar dicha afirmación. Para ello se obtiene una muestra y se calcula la probabilidad de sustentar dicha afirmación en base a los datos muestrales.

A la afirmación que se quiere probar se le llama ***hipótesis nula*** y se designa con H_0.

Definición 1.

La probabilidad de rechazar H_0, en el supuesto de que H_0 es cierta, se conoce como **error de tipo I** y se denota con α.

$$P\{rechazar\ H_0 \mid H_0 es\ cierta\} = \alpha$$

Definición 2.

La probabilidad de no rechazar H_0, en el supuesto de que H_0 es falsa, se conoce como **error de tipo II** y se denota con β.

$$P\{no\ rechazar\ H_0 \mid H_0 es\ falsa\} = \beta$$

Definición 3.

Una **prueba** de una hipótesis estadística con respecto a alguna característica desconocida de la población es cualquier regla para decidir si se rechaza la hipótesis nula con base en una muestra aleatoria de la población.

Definición 4.

Se denomina **región crítica** a los valores de prueba que rechazan la hipótesis nula. El área de la región crítica es igual al tamaño del error de tipo I.

Para construir una regla de decisión apropiada en la prueba de una hipótesis estadística es necesario establecer una hipótesis alternativa que refleje el valor o intervalo de valores posibles del parámetro, si la hipótesis nula es falsa. Esta hipótesis alternativa se designa con H_1. Probar una hipótesis consiste, pues, en proporcionar una decisión entre H_0 y H_1.

El enfoque general es el de aceptar que el error de tipo I es mucho más serio que el error de tipo II, y formular la hipótesis nula y la alternativa de acuerdo con lo anterior.

Una manera simple de obtener reglas de decisión consiste en seleccionar aquel procedimiento que tenga el error más pequeño de tipo II de entre todos los procedimientos que den el mismo tamaño de error de tipo I.

La probabilidad α de error de tipo I también se conoce como el **nivel de significación estadística.**

Ejemplo.

Supóngase que se tiene interés en que el tiempo promedio de armado de una unidad en una planta de un gran almacén, sea de 10 minutos. El gerente decide continuar con el proceso a menos que se encuentre evidencia de que el tiempo promedio no es de 10 minutos. La evidencia se obtendrá de una muestra de tamaño *n* de la población de interés. ¿Cómo debe decidirse si el proceso continúa en operación?

Afinemos un poco más. Supóngase que puede tolerarse un error de tipo I hasta de 0,06 cuando se prueba la hipótesis nula

$$H_0: \mu = 10$$

Contra la hipótesis alternativa

$$H_1: \mu > 10$$

Y supóngase, también, que la distribución del tiempo de armado de una unidad sigue una normal con desviación estándar σ=1,4 minutos.

Para obtener la evidencia empírica se extrae una muestra de aleatoria de tamaño 25 y se escoge la media muestral como estadístico de prueba.

Deseamos comparar las siguientes regiones críticas:

$$\begin{array}{llll} \text{Prueba } A & \text{Rechazar } H_0 & si & \overline{X} > 10{,}65 \\ \text{Prueba } B & \text{Rechazar } H_0 & si & \overline{X} > 10{,}45 \\ \text{Prueba } C & \text{Rechazar } H_0 & si & \overline{X} > 10{,}25 \end{array}$$

Para determinar cuál de estas satisface el tamaño de error de tipo I que puede tolerarse y cuál es el valor más pequeño de β de entre las tres.

Para determinar el error de tipo I, se asumirá que H_0 es cierta y se calculará:

$$P\{\overline{X} > c \mid \mu = 10\} = \alpha$$

Como

$$\overline{X} \to N(10, \frac{1{,}4}{\sqrt{25}})$$

Para la prueba A

$$P\{\overline{X} > 10{,}65\} = P\left\{Z > \frac{10{,}65 - 10}{1{,}4 / \sqrt{25}}\right\} = P\{Z > 2{,}32\} = 0{,}012$$

De manera similar para la prueba B

$$P\{\overline{X} > 10{,}45\} = P\left\{Z > \frac{10{,}45 - 10}{1{,}4 / \sqrt{25}}\right\} = P\{Z > 1{,}607\} = 0{,}0537$$

y para C

$$P\left\{\overline{X} > 10{,}25\right\} = P\left\{Z > \frac{10{,}25 - 10}{1{,}4 / \sqrt{25}}\right\} = P\{Z > 0{,}893\} = 0{,}1867$$

Como C no reúne los requisitos, no será ya considerado.

De entre A y B determinemos cuál de los dos es la que tiene un error de tipo II más pequeño.

El error de tipo II, es el de no rechazar H_0, en el supuesto que H_0 es falsa.

$$\beta(\mu) = P\left\{\overline{X} < c \mid \mu > 10\right\}$$

Para la prueba A

$$\beta(\mu) = P\left\{\overline{X} < 10{,}65 \mid \mu > 10\right\}$$

y para la prueba B

$$\beta(\mu) = P\left\{\overline{X} < 10{,}45 \mid \mu > 10\right\}$$

Estos valores se pueden tabular para los diferentes valores de μ.

Por ejemplo para $\mu = 10{,}4$ se obtiene:

En la prueba A

$$\beta(10{,}4)=0{,}8133$$

Y en la prueba B

$$\beta(10{,}4)=0{,}5714$$

Para este valor particular de μ en la prueba alternativa, la hipótesis B es mejor que la hipótesis A.

Tabulando los resultados para distintos valores de μ en la hipótesis alternativa se obtiene:

μ	10,1	10,2	10,3	10,4	10,5	10,6	10,7	10,8	10,9	11	11,1	11,2
Prueba A	0,9753	0,9460	0,8944	0,8140	0,7039	0,5709	0,4291	0,2961	0,1860	0,1056	0,0540	0,0247
Prueba B	0,8944	0,8140	0,7039	0,5709	0,4291	0,2961	0,1860	0,1056	0,0540	0,0247	0,0101	0,0037

Como puede verse en la tabla conforme el error de tipo I disminuye, el de tipo II aumenta. Si la opción propuesta por la hipótesis nula es falsa, pero difiere muy poco del verdadero valor, la opción de no rechazar H_0 es alta.

2. Regiones críticas y la función potencia.

Considérese la hipótesis simple

$$H_0: \theta=\theta_0$$

cuando se muestrea una población con función de densidad $f(x; \theta)$ en donde θ_0 es el valor propuesto de θ. Si la hipótesis alternativa es de la forma:

$$H_1: \theta>\theta_0.$$

ó

$$H_1: \theta<\theta_0.$$

se dice que H_1 es una hipótesis alternativa unilateral.

Vale la pena observar que la hipótesis alternativa debe formularse solo si el valor de uno de los parámetros que se encuentra en el lado opuesto no tiene sentido para el investigador.

Debe notarse que $\beta(\theta)$ se conoce como función característica de operación.

Dado que $\beta(\theta)$ es la probabilidad de que un valor de la estadística de prueba no se encuentre en la región crítica cuando H_0 es falsa, entonces $1-\beta(\theta)$ representa la probabilidad de que se encuentre en la región crítica cuando es falsa. Esta probabilidad se conoce como **potencia de la prueba**.

Definición 5.

La función $P(\theta)=1-\beta(\theta)$ recibe el nombre de función potencia de la prueba y representa la probabilidad de rechazar la hipótesis nula cuando ésta es falsa; es decir, cuando el valor del parámetro de H_1 es cierto.

3. Las mejores pruebas.

Sea X_1. X_2, …, X_n una muestra aleatoria de tamaño *n* de una población cuya función de densidad es $f(x;\theta)$ y sean

$$H_0: \theta=\theta_0$$

contra

$$H_1: \theta=\theta_1$$

θ_0 y θ_1 dados. Si α es el tamaño máximo del error de tipo I admitido, entonces la mejor prueba para H_0 contra H_1 es aquella que tiene el tamaño más pequeño del error de tipo II, de entre todas las pruebas que tengan un tamaño de error de tipo I no mayor que α.

Se pueden determinar las regiones críticas para estas pruebas mediante el uso del siguiente teorema.

Teorema 1.

Si existe una región crítica C de tamaño α y una constante positiva k tal que:

$$\frac{L_0(x_1, x_2, \ldots, x_n; \theta_0)}{L_1(x_1, x_2, \ldots, x_n; \theta_1)} \leq k \qquad \text{en el interior de} \quad C$$
$$\frac{L_0(x_1, x_2, \ldots, x_n; \theta_0)}{L_1(x_1, x_2, \ldots, x_n; \theta_1)} \geq k \qquad \text{en el exterior de} \quad C$$

entonces C es la mejor región crítica de tamaño α para probar H_0: $\theta=\theta_0$ contra H_1: $\theta=\theta_1$, en donde L_0 y L_1 son las funciones de verosimilitud relativas a H_0 y H_1, respectivamente.

Definición 6.

Se dice que una prueba de la hipótesis H_0: $\theta=\theta_0$ es la prueba uniformemente más potente de tamaño α si ésta es por lo menos tan poderosa, para cualquier valor de θ de la hipótesis alternativa, como cualquier otra prueba de tamaño menor o igual que α.

4. Principios generales para probar una H0 simple contra una H1 unilateral o bilateral.

Se desarrollarán criterios generales de prueba para los siguientes 3 casos.

Caso 1.

H_0: $\theta = \theta_0$

contra

H_1: $\theta \neq \theta_0$.

Caso 2.

H_0: $\theta = \theta_0$

contra

$$H_1: \theta > \theta_0.$$

Caso 3.

$$H_0: \theta = \theta_0$$

contra

$$H_1: \theta < \theta_0.$$

Principios generales para el caso 1.

Dada una muestra aleatoria de tamaño *n* de la distribución de interés, el procedimiento general para probar H_0 es escoger el mejor estimador de θ, T y rechazar H_0 cuando el estimado t obtenido de la muestra, es en forma suficiente, diferente del valor propuesto de θ_0.

Para un tamaño preseleccionado α, del error de tipo I, se puede obtener una región crítica bilateral, de los extremos de la distribución de muestreo de T, de manera que el área en cualquier lado más allá del valor crítico, sea igual a $\alpha/2$.

Entonces se rechaza H_0, a favor de H_1, cuando el estimado t se encuentra en la región crítica.

Cuando el estimado t no se encuentra dentro de la región crítica, no puede rechazarse la hipótesis nula. De esta forma cualquier diferencia con respecto al valor de θ_0, se considera causada por la fluctuación en el muestreo del estimador T.

Ejemplo.

Supóngase que para una ciudad solo hay dos canales de TV, el canal 6 y el 10. Se cree que para las noticias de la tarde el auditorio se encuentra dividido en partes iguales para ambos canales. Una compañía desea probar la hipótesis nula

$$H_0: p = 0{,}5$$

contra

$$H_1: p \neq 0{,}5.$$

Para ello selecciona una muestra de 18 residentes al azar y se pregunta qué canal prefiere. La variable X indica que el canal seleccionado es el 6. Se proponen las siguientes dos pruebas

Prueba A: Rechazar H_0 si $X \leq 4$ ó $X \geq 14$.

Prueba B: Rechazar H_0 si $X \leq 5$ ó $X \geq 13$.

Si la compañía piensa tolerar un tamaño máximo de 0,1 para el error de tipo I, determinar la mejor prueba a emplear para decidir entre H_0 y H_1.

La estadística de prueba X es una variable aleatoria binomial con n=18 y, bajo la hipótesis nula, p=0,5.

Para la prueba A, la probabilidad de error tipo I es

$$\alpha_A = P(X \le 4 \mid p = 0{,}5) + P(X \ge 14 \mid p = 0{,}5) = 0{,}0154 + 0{,}0154 = 0{,}0308$$

Para la prueba B, la probabilidad de error tipo I es

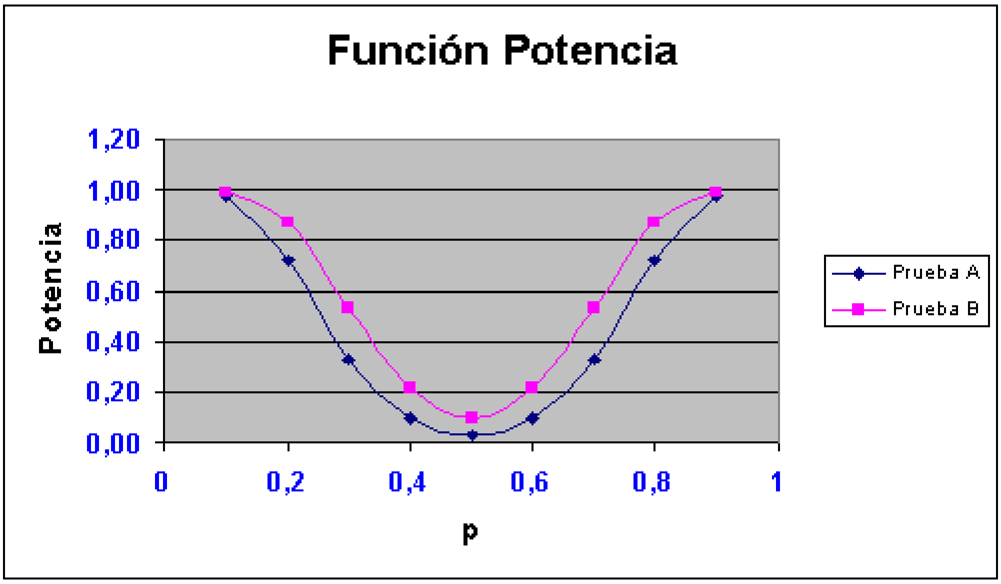
$$\alpha_B = P(X \le 5 \mid p = 0{,}5) + P(X \ge 13 \mid p = 0{,}5) = 0{,}0962$$

Como ambas pruebas tiene valores de α menores al tamaño máximo que puede tolerarse, se compararán sus funciones potencia para decidir cuál es la mejor de las dos.

Utilizamos la hoja de cálculo del Excel.

p	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Prueba A	0,971806	0,716354	0,332694	0,095448	0,030884	0,095448	0,332694	0,716354	0,971806
Prueba B	0,993585	0,867086	0,534649	0,214509	0,096252	0,214509	0,534649	0,867086	0,993585

En la tabla se observa que, para cualquier valor de p, la potencia de la prueba B es mayor que la de la prueba A. De acuerdo con lo anterior, la prueba B es uniformemente más poderosa que la prueba A y es la mejor prueba de las dos indicadas.



Principios generales para el caso 2.

El procedimiento general para probar H_0, es de nuevo escoger el mejor estadístico T de θ y rechazar H_0 cuando el estimado t es en forma suficiente mayor que el valor propuesto de θ_0.

De esta forma para un tamaño α del error de tipo I, la región crítica se encuentra localizada en el extremo superior de la distribución de muestreo de T y H_0 se rechazará si el estimado t no es menor que el valor crítico.

Análogamente se hace para el caso 3.

Se han formulado críticas con respecto a las pruebas de hipótesis estadísticas debido a que la decisión final de rechazar o no una H_0 dada, es demasiado cortante y seca y no proporciona una mediada real de que la decisión sea correcta en términos de la probabilidad.

Para ello se ha sugerido el cálculo del **valor de p**. El valor de p es la probabilidad, dado que H_0 es cierta, de que el estadístico de prueba tome un valor mayor o igual que el calculado con base en la muestra aleatoria.

Se sugiere la siguiente regla: Si el valor de p es menor o igual que α, se rechaza H_0; de otra forma no puede rechazarse la hipótesis nula.

Es de destacar que muchos paquetes estadísticos para computadora, tales como SAS, APSS, BMD y otros, imprimen el valor de p para casi todas las situaciones en las que interviene las pruebas de hipótesis estadísticas.

Ejemplo.

Los siguientes datos representan tiempos de armado para 20 unidades seleccionadas aleatoriamente: 9,8; 10,4; 10,6; 9,6; 9,7; 9,9; 11,1; 9,6; 10,2; 10,3; 9,6; 9,9; 11,2; 10,6; 9,8; 10,5; 10,1; 10,5; 9,7. Supóngase que el tiempo necesario para armar una unidad es una variable aleatoria normal con media μ y desviación típica σ=0,6. Con base en esta muestra, ¿existe alguna razón para creer, a un nivel de 0,05, que el tiempo de armado promedio es mayor de 10 minutos?

Solución

Considérese la hipótesis nula

$$H_0: \mu=10$$

contra la alternativa

$$H_1: \mu>10$$

Si puede rechazarse a H_0 con α=0,05, entonces existe una razón para creer que el tiempo necesario para armar una unidad es mayor que 10 minutos.

Como la media muestral, en el supuesto de que H_0 es cierta, sigue una distribución normal de parámetros

$$\overline{X} \rightarrow N\left(\mu; \frac{0,6}{\sqrt{20}}\right)$$

El valor crítico es aquel para el cual

$$P(\overline{X} > c) = 0,05$$

Buscando en las tablas el valor de c=10,2206803.

De los datos se obtiene que la media muestral vale 10,2. Por tanto no puede rechazarse la hipótesis nula. El valor de p en este caso es la probabilidad de

$$P(\overline{X} > 10,2 \mid \mu = 10) = p = 0,06801859$$

Puesto que p es mayor que α, se concluye que con base en la muestra no existe evidencia suficiente para rechazar la hipótesis de que el tiempo promedio necesario para armar una unidad es de 10 minutos.

Ejemplo

Los pesos en gramos del contenido de 16 cajas de cereal que se seleccionaron para verificar el peso promedio, fueron:

506 508 499 503 504 510 497 512 514 505 493 496 506 502 509 496

Si el peso de cada caja es una variable aleatoria normal con desviación típica σ=5 g.

Demostrar que para cualquier valor propuesto de μ_0 de μ que se encuentre en el interior del intervalo de confianza del 95%, una prueba de hipótesis

H_0: $\mu=\mu_0$

contra

H_1: $\mu\neq\mu_0$.

no llevará al rechazo de H_0 para α=0,05.

Solución

$$\overline{X} \rightarrow N\left(\mu; \frac{5}{\sqrt{16}}\right)$$

Un intervalo de confianza al 95% para μ viene dado por:

503,75 $\pm$ 1.96 (5/4) lo que da unos límites de 500,45 y 507,05.

Es necesario demostrar que los límites 500,45 y 507,05 coinciden con los límites de los valores propuestos μ_0 bajo H_0 que llevan al rechazo de la hipótesis nula.

Para cualquier valor propuesto de μ_0 comprendido entre 500,45 y 507,05, el mejor estimador de μ viene dado por

$$T = \frac{\overline{X} - \mu_0}{S / \sqrt{n}} \quad \textit{que sigue una} \quad t - \textit{Student} \quad \textit{con} \quad n-1 \quad \textit{grados de libertad}$$

Dado que la media muestral es 503,75 y la cuasidesviación típica muestral s =6,2. Para el límite 500,45 se tiene:

$$t = \frac{503{,}75 - 500{,}45}{6{,}2 \big/ \sqrt{16}} = 2{,}131$$

y para el límite 507,05

$$t = \frac{503{,}75 - 507{,}05}{6{,}2 \big/ \sqrt{16}} = -2{,}131$$

Si calculamos la región crítica bilateral para T en una t-Student con 15 grados de libertad, obtenemos:

$$P (\mid T \mid > c) = 0{,}05$$

El valor de c que se obtiene es de 2,131. De esta forma, cualquier valor propuesto de μ_0 interior al intervalo (500,45 ; 507,05) no llevará al rechazo de H_0 con α=0,05.

Para ilustrar el cálculo del valor de p en este contexto, considérese la hipótesis nula

$$H_0: \mu = 508$$

contra la alternativa

$$H_1: \mu \neq 508$$

Dado que 508 está fuera del intervalo (500,45 ; 507,05), H_0 será rechazada a un nivel de 0,05. El valor de p será la probabilidad

$P (\mid T \mid > t_0) = p$, donde

$$t_0 = \frac{503{,}75 - 508}{6{,}2 / \sqrt{16}} = -2{,}742$$

Puesto que es una hipótesis bilateral, se calcula P (| T | > 2,742)=0,016. Por tanto si la hipótesis nula es cierta, existe una probabilidad menor del 2% para observar un valor de la distribución t de Student cuya magnitud sea igual o mayor al valor observado 2,742.

Ejemplo

Un investigador diseña un experimento en el que se pedirá a un determinado número de sujetos que lleven a cabo una tarea específica en un medio controlado y bajo dos niveles diferentes de ruido de fondo. El investigador selecciona 32 personas que son capaces de realizar la misma tarea y de manera práctica en el mismo tiempo. Del total de personas, 16 seleccionadas al azar realizarán esta tarea bajo un nivel modesto de ruido de fondo. Las restantes 16 llevarán a cabo la tarea bajo un nivel de ruido 2, el cuál es más severo que el 1. Los siguientes datos representan los tiempos observados necesarios para completar la tarea para cada una de las 16 personas de cada nivel.

Nivel 1	14	12	15	15	11	16	17	12	14	13	18	13	1
Nivel 2	20	22	18	18	19	15	18	15	22	18	19	15	2

Asumiendo que estos datos constituyen muestras aleatorias de dos distribuciones normales e independientes con varianzas iguales pero desconocidas, ¿existe alguna razón para creer que el tiempo promedio para el nivel 2 es mayor por más de 2 minutos que para el nivel 1 con α=0,01?

<u>Solución</u>

Sean μ_1 y μ_2 las medias desconocidas de los niveles 1 y 2. Se afirma que μ_2 es mayor que μ_1 por más de 2 minutos. En base a lo anterior, consideremos la hipótesis nula

$$H_0: \mu_2-\mu_1=2$$

contra la hipótesis alternativa

$$H_1: \mu_2-\mu_1>2$$

Como

$$\overline{X} \to N(\mu_1, \sigma)$$
$$\overline{Y} \to N(\mu_2, \sigma)$$

Como la varianza es desconocida, el estadístico

$$T = \frac{\overline{Y} - \overline{X} - (\mu_Y - \mu_X)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \quad donde \quad S_p = \sqrt{\frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}}$$

sigue una T-Student con n_X+n_Y-2 grados de libertad.

La media muestral de Y es 18,5 y la de X, 14,375; la varianza muestral de Y es 6 y la de X, 5,1833. Por tanto S_p=2,36467. Luego

$$t = \frac{18{,}5 - 14{,}375 - 2}{2{,}36467\sqrt{\dfrac{2}{16}}}$$

eso da t = 2,5417. Dado que el valor de t = 2,5417 se encuentra por encima de la región crítica de tamaño 0,01, se rechaza la hipótesis nula. El cálculo de la región crítica de tamaño 0,01 se puede hacer con la función t inversa de la hoja de cálculo para una sola cola. Eso da DISTR.T.INV(0,02;30)=2,45726369 (*el valor 0,02 es porque solo tiene una cola*).

Bajo la hipótesis de H_0, el valor de p es: P ($T \geq 2{,}5417$)= 0,00821748. Por tanto, con base en este experimento, puede concluirse que la diferencia entre las medias de los niveles 1 y 2 es mayor de 2 minutos es estadísticamente discernible con valor p de 0,00821748.

Ejemplo.

Un investigador médico se interesa por determinar si un fármaco experimental tiene el efecto colateral no deseable de elevar la presión sistólica sanguínea. En un ambiente controlado de laboratorio se le toman la presión sanguínea de los n sujetos y se les administra el fármaco durante un lapso adecuado de tiempo después del cual se las vuelve a tomar la presión sanguínea.

Sean (X_1, Y_1), (X_2, Y_2), …, (X_n, Y_n) los n pares, donde (X_i, Y_i) denota la presión sanguínea del i-ésimo individuo antes y después de tomar el fármaco. Cada individuo forma un par, así pues al considerar las diferencias de cada par, se eliminan las influencias externas que pudieran haber por otros motivos distintos a la ingesta del fármaco.

Supongamos que las diferencias D_1, D_2, …, D_n constituyen v.a IID normalmente distribuidas con $E(D_i)=\mu_D$ y $Var(D_i)=\sigma_D^2$.

Vamos a probar la hipótesis nula

$$H_0: \mu_D=0$$

contra la hipótesis alternativa

$$H_1: \mu_D>0$$

con α=0,01.

Suponiendo la media y la varianza desconocidas, se tiene que:

$$\frac{\overline{D}-0}{S_D / \sqrt{n}} \rightarrow t_{n-1}$$

La media muestral de las diferencias vale 3,75 y S_D=3,7929 se tiene que t=3,425. El valor crítico c es igual a P(T>=c)=0,01 con T siguiendo una t-Student de 11 grados de libertad, lo que equivale a c=DISTR.T.INV(0,02;11)=2,71807949. Por tanto se rechaza la hipótesis nula de no efecto del medicamento. Por tanto, con base en los resultados de este estudio, un incremento en el valor promedio de la presión sanguínea es estadísticamente discernible con un valor de *p* igual a P(T>=3,425)= 0,00283646.

Ejemplo.

Un investigador cree que ha desarrollado una variedad de maíz nueva que es superior a la mejor variedad disponible. Para verificar lo anterior, el investigador diseña el siguiente experimento: se seleccionan 10 parcelas de igual tamaño cada una en distinta localidad geográfica. Cada parcela se divide en 2 secciones iguales, de manera tal que puedan cultivarse las 2 variedades en cada localidad. En el momento de recoger la cosecha se anotan las toneladas por unidad de área.

Tipo	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}
Variedad X (estándar)	23	35	29	42	33	19	37	24	35	26
Variedad Y (nueva)	26	39	35	40	38	24	36	27	41	27

Con base en estos datos, obténgase un intervalo de confianza del 95% para la diferencia media de la producción entre X e Y.

Diferencias	-3	-4	-6	2	-5	-5	1	-3	-6	-1

Media	Varianza Muestral	Desviación T muestral
-3	8,0000	2,8284

$$\bar{d} \pm t_{0{,}975{,}9} \frac{s_D}{\sqrt{10}}$$

$$-3 \pm 2{,}262 \frac{2{,}8284}{\sqrt{10}} = (-5{,}0232, -0{,}9768)$$

Dado que el valor cero no se incluye en dicho intervalo, se rechaza la correspondiente hipótesis nula de que la diferencia es cero a un nivel de α=0,05.

5. Pruebas de hipótesis con respecto a la varianza cuando se muestrean distribuciones normales.

Pruebas para una muestra

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2 desconocidas. Sea la hipótesis nula:

$$H_0 : \sigma^2 = \sigma_0^2$$

contra una de las hipótesis

$$H_1 : \sigma^2 \neq \sigma_0^2, \qquad H_1 : \sigma^2 > \sigma_0^2, \qquad H_1 : \sigma^2 < \sigma_0^2,$$

El estadístico de interés es la varianza muestral S^2. La hipótesis nula será rechazada si la realización de S^2 calculada a través de la muestra, es en forma suficiente, diferente, mayor o menor que σ_0^2, dependiendo de la hipótesis alternativa.

Pero bajo la suposición de que H_0 es cierta, el estadístico $(n-1)s^2/\sigma_0^2$ es un valor de la variable aleatoria chi-cuadrada con *n-1* grado de libertad. Por tanto, se rechazará H_0 si dicho valor se encuentra dentro de la región crítica de tamaño α. Véase la tabla:

Hipótesis nula	Valor del estadístico de prueba
H_0: $\sigma^2=\sigma_0^2$	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$
Hipótesis alternativa	**Criterios de rechazo**
H_1: $\sigma^2 \neq \sigma_0^2$	$\chi^2 \geq \chi_{1-\alpha/2,n-1}^2$ *o cuando* $\chi^2 \leq \chi_{\alpha/2,n-1}^2$
H_1: $\sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi_{1-\alpha,n-1}^2$
H_1: $\sigma^2 < \sigma_0^2$	$\chi^2 \leq \chi_{\alpha,n-1}^2$

Pruebas para dos muestras.

Sea $X_1, X_2, \ldots, X_{nX}$ y $Y_1, Y_2, \ldots, Y_{nY}$ dos muestras aleatorias de una distribución normal con media μ_X y μ_Y y varianzas σ_X^2 y σ_Y^2 desconocidas. Sea la hipótesis nula:

$$H_0 : \sigma_X{}^2 = \sigma_Y{}^2$$

contra una de las hipótesis

$$H_1 : \sigma_X{}^2 \neq \sigma_Y{}^2, \qquad H_1 : \sigma_X{}^2 > \sigma_Y{}^2, \qquad H_1 : \sigma_X{}^2 < \sigma_Y{}^2,$$

Los estadísticos de interés son las varianzas maestrales S_X^2 y S_Y^2. Se sabe que

$$(n_X\text{-}1)\, S_X^2/\sigma_X^2 \text{ y } (n_Y\text{-}1) S_Y^2/\sigma_Y^2$$

son dos variables aleatorias independientes chi-cuadrado con n_X-1 y n_Y-1 grados de libertad. Entonces el estadístico

$$F = \frac{S_X{}^2/\sigma_X{}^2}{S_Y{}^2/\sigma_Y{}^2}$$

sigue una F con con n_X-1 y n_Y-1 grados de libertad. Por tanto, bajo la hipótesis nula cierta, se reduce a:

$$F = S_X{}^2 \big/ S_Y{}^2$$

Para una hipótesis alternativa bilateral y un tamaño α del error de tipo I, se rechazará la hipótesis nula cuando $f = s_X^2/ s_Y^2 \geq f_{1-\alpha/2,\, nX-1,nY-1}$ o cuando $f \leq 1/ f_{1-\alpha/2,\, nY-1,nX-1}$

Ejemplo

El investigador del experimento del ruido de fondo, con los datos

Nivel 1	14	12	15	15	11	16	17	12	14	13	18	13	1
Nivel 2	20	22	18	18	19	15	18	15	22	18	19	15	2

Asumiendo que estos datos constituyen muestras aleatorias de dos distribuciones normales e independientes con varianzas iguales pero desconocidas, ¿existe alguna razón para creer que las varianzas son efectivamente iguales para el nivel α=0,1?

Sea la hipótesis

$$H_0: \sigma_1^2=\sigma_2^2$$

contra la alternativa

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Se observa que los valores críticos, izquierdo y derecho, son $f_{0,95;\,15,\,15}$ =DISTR.F.INV(0,05;15;15)=2,40 y 1/ $f_{0,95;\,15,\,15}$=DISTR.F.INV(0,95;15;15)=1/2,40=0,42. Con base en los datos de la muestra s_1^2=5,1833 y s_2^2=6. De esta forma el estadístico de prueba es f=5,1833/6=0,8639.

Dado que f no es mayor que 2,40 ni menor que 0,42 no es posible rechazar la hipótesis nula.

6. Inferencias para proporciones de dos distribuciones binomiales independientes.

La población general se subdividió en dos grupos, zurdos y derechos, y cada grupo fue dividido en fumadores y no fumadores. Sea p_1 la proporción de personas zurdas que fuman y p_2 la proporción de personas diestras que fuman. El interés recae en hacer una comparación entre p_1 y p_2.

Supóngase que los zurdos y los derechos constituyen dos distribuciones binomiales independientes tales que la proporción de fumadores en los dos grupos es p_1 y p_2, respectivamente. Con base en muestras aleatorias de tamaño n_1 y n_2, sean X e Y el número observado de personas zurdas y derechas que fuman, respectivamente. las proporciones muestrales

$$\hat{P}_1 = X/n_1 \quad y \quad \hat{P}_2 = Y/n_2$$

son los estimadores de máxima verosimilitud de p_1 y p_2. Dado que X e Y son variables aleatorias binomiales, las varianzas de los estimadores están dadas por

$$Var(\hat{P}_1) = Var(X/n_1) = p_1(1-p_1)/n_1$$
$$Var(\hat{P}_2) = Var(Y/n_2) = p_2(1-p_2)/n_2$$

Supóngase que se desea construir un intervalo de confianza muestral grande para la diferencia entre p_1 y p_2. El estadístico de interés es la diferencia entre las proporciones muestrales. Luego

$$E(\hat{P}_1 - \hat{P}_2) = E(\hat{P}_1) - E(\hat{P}_2) = p_1 - p_2$$
$$y$$
$$Var(\hat{P}_1 - \hat{P}_2) = Var(\hat{P}_1) + Var(\hat{P}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_1(1-p_2)}{n_2}$$

Se demostró que para n_1 y n_2 grandes, la distribución de

$$\hat{P}_1 - \hat{P}_2 \to N\left(p_1 - p_2; \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_1(1-p_2)}{n_2}}\right)$$

es aproximadamente normal. Es decir:

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_1(1-\hat{P}_2)}{n_2}}}$$

es aproximadamente normal N(0,1). Nótese que el denominador es un estimador de la desviación típica, puesto que se han sustituido las proporciones muestrales p_1 y p_2. Por tanto, la probabilidad del intervalo aleatorio

$$\left[(\hat{P}_1 - \hat{P}_2) - z_{1-\alpha/2} d.e(\hat{P}_1 - \hat{P}_2), \quad (\hat{P}_1 - \hat{P}_2) + z_{1-\alpha/2} d.e(\hat{P}_1 - \hat{P}_2)\right]$$

es aproximadamente 1-α, y un intervalo de confianza aproximado del 100(1-α)% para p_1-p_2 es:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_1(1-\hat{p}_2)}{n_2}} \quad donde \quad \hat{p}_1 = x/_{n_1} \; y \, \hat{p}_2 = y/_{n_2}$$

son los estimadores de máxima verosimilitud de p_1 y p_2, respectivamente.

Ejemplo

Siguiendo con el estudio del hábito de fumar, en una muestra de 400 zurdos 190 fumaban, y en una muestra de 800 diestros, 300 fumaban. Con base en esto construir un intervalo de confianza al 98% para la diferencia real entre las proporciones p_1 y p_2.

Los estimadores de las proporciones son

$$\hat{p}_1 = 190/400 = 0{,}475, \quad \hat{p}_2 = 300/800 = 0{,}375$$

Como los tamaños de las muestras son grandes, la aproximación normal es adecuada. Para un intervalo del 98%, $z_{0,99}$=2,33.

$$(0{,}475 - 0{,}375) \pm 2{,}33\sqrt{\frac{0{,}475(1-0{,}475)}{400} + \frac{0{,}375(1-0{,}375)}{800}}$$

lo que da un intervalo (0,0295, 0,1705). Dado que dicho intervalo no incluye al cero, puede concluirse con un 98% de confiabilidad, que el porcentaje de zurdos que fuman es mayor que el correspondiente para las personas derechas.

Supóngase que el interés recae en probar la hipótesis nula

H_0: p_1-p_2 = 0

contra una de las hipótesis alternativas

H_1: p_1-$p_2 \neq 0$, H_1: p_1-$p_2 > 0$, H_1: p_1-$p_2 < 0$

Dadas las muestras aleatorias de tamaños n_1 y n_2, considérese el estadístico

$$\hat{P}_1 - \hat{P}_2$$

La intuición sugiere que debe rechazarse la hipótesis nula si un valor de la estadística es en forma suficiente, diferente, mayor que, o menor que cero, dependiendo de la hipótesis alternativa.

Dado que bajo H_0 se supone que las proporciones son iguales, sea $p = p_1 = p_2$ la proporción común. Entonces, si H_0 es cierta, el estadístico

$$\hat{P}_1 - \hat{P}_2 \to N\left(0, \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}\right)$$

Ya que el valor de p no se conoce, se combina la información de las dos muestras para obtener el estimador combinado

$$\hat{P} = \frac{X+Y}{n_1+n_2}$$

Entonces un estimador de la desviación estándar de

$$d.e.(\hat{P}_1 - \hat{P}_2) = (\sqrt{\hat{p}(1-\hat{p})})\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

en donde

$$\hat{p} = \frac{x+y}{n_1+n_2}$$

es el estimador combinado de p.

Bajo H_0, el estadístico

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\left(\sqrt{\hat{P}(1-\hat{P})}\right)\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)}$$

es aproximadamente normal N(0,1) para valores grandes de n_1 y n_2. Dependiendo de la hipótesis alternativa, no es difícil decidir cuando rechazar H_0, con base en un tamaño de error tipo I.