

# TEMA 8 CONTRASTE DE HIPÓTESIS

Pablo Buenestado

Curso 2019-2020 Primavera

Departamento de Matemáticas (UPC)

# Índice

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

Un parámetro puede ser estimado a partir de datos muestrales o con un solo número (una estimación puntual) o un intervalo completo de valores plausibles (un intervalo de confianza).

Con frecuencia, sin embargo, el objetivo de una investigación no es estimar un parámetro sino decidir cuál de dos pretensiones contradictorias sobre el parámetro es la correcta.

Los métodos para lograr esto comprenden la parte de la inferencia estadística llamada prueba de hipótesis.

En este tema, primero se discuten algunos de los conceptos y terminología básicos en la prueba de hipótesis y luego se desarrollan procedimientos para la toma de decisiones para los problemas de realización de pruebas más frecuentemente encontrados con base en una muestra tomada de una sola población.

# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

Una **hipótesis estadística** o simplemente **hipótesis** es una pretensión o aseveración sobre el valor de un solo parámetro (característica de una población o característica de una distribución de probabilidad), sobre los valores de varios parámetros o sobre la forma de una distribución de probabilidad completa.

Un ejemplo de una hipótesis es la pretensión de que  $\mu = 0.75$ , donde  $\mu$  es el diámetro interno promedio verdadero de un cierto tipo de tubo de PVC.

Otro ejemplo es la proposición  $p < 0.10$ , donde  $p$  es la proporción de tarjetas de circuito defectuosas entre todas las tarjetas de circuito producidas por un cierto fabricante.

No obstante otro ejemplo de una hipótesis es la aseveración de que la distancia de detención en condiciones particulares tiene una distribución normal.

Hipótesis de esta última clase no se considerarán en este tema.

Aquí la atención se concentra en hipótesis en relación con parámetros.

En cualquier problema de prueba de hipótesis, existen dos hipótesis contradictorias consideradas.

Una podría ser la pretensión de que  $\mu = 0.75$  y la otra  $\mu \neq 0.75$  o las dos proposiciones contradictorias podrían ser  $p \geq 0.10$  y  $p < 0.10$ .

El objetivo es decidir, con base en información muestral, cuál de las dos hipótesis es la correcta.

Existe una analogía conocida de esto en un juicio criminal.

Una pretensión es la aseveración de que el individuo acusado es inocente.

En el sistema judicial, esta es la pretensión que inicialmente se cree que es cierta.

Sólo de cara a una fuerte evidencia que diga lo contrario el jurado deberá rechazar esta pretensión a favor de la aseveración alternativa de que el acusado es culpable.

En este sentido, la pretensión de inocencia es la hipótesis favorecida o protegida y el agobio de comprobación recae en aquellos que creen en la pretensión alternativa.

Asimismo, al probar hipótesis estadísticas, el problema se formulará de modo que una de las pretensiones sea inicialmente favorecida.

Esta pretensión inicialmente favorecida no será rechazada a favor de la pretensión alternativa a menos que la evidencia muestral la contradiga y apoye fuertemente la aseveración alternativa.



## Definición

La **hipótesis nula** denotada por  $H_0$ , es la pretensión de que inicialmente se supone cierta (la pretensión de "creencia previa").

La **hipótesis alternativa** denotada por  $H_a$ , es la aseveración contradictoria a  $H_0$ .

La hipótesis nula será rechazada en favor de la hipótesis alternativa sólo si la evidencia muestral sugiere que  $H_0$  es falsa.

Si la muestra no contradice fuertemente a  $H_0$ , se continuará creyendo en la verdad de la hipótesis nula.

Las dos posibles conclusiones derivadas de un análisis de prueba de hipótesis son entonces rechazar  $H_0$  o no rechazar  $H_0$ .

Una **prueba de hipótesis** es un método de utilizar datos muestrales para decidir si la hipótesis nula debe ser rechazada.

Por consiguiente se podría probar  $H_0 : \mu = 0.75$  contra la alternativa  $H_a : \mu \neq 0.75$ .

Sólo si los datos muestrales sugieren fuertemente que  $\mu$  es otra diferente de 0.75 deberá ser rechazada la hipótesis nula.

Sin semejante evidencia,  $H_0$  no deberá ser rechazada, puesto que sigue siendo bastante plausible.

En ocasiones un investigador no desea aceptar una aseveración particular a menos y hasta que los datos apoyan fuertemente la aseveración.

Como ejemplo, supongamos que una compañía está considerando aplicar un nuevo tipo de recubrimiento en los cojinetes que fabrica.

Se sabe que la vida de desgaste promedio verdadera con el recubrimiento actual es de 1000 horas.

Si  $\mu$  denota la vida promedio verdadera del nuevo recubrimiento, la compañía no desea cambiar a menos que la evidencia sugiera fuertemente que  $\mu$  excede de 1000.

Una formulación apropiada del problema implicaría probar  $H_0 : \mu = 1000$  contra  $H_a : \mu > 1000$ .

La conclusión de que se justifica un cambio está identificada con  $H_a$  y se requeriría evidencia conclusiva para justificar el rechazo de  $H_0$  y cambiar al nuevo recubrimiento.

La investigación científica a menudo implica tratar de decidir si una teoría actual debe ser reemplazada por una explicación más plausible y satisfactoria del fenómeno investigado.

Un método conservador es identificar la teoría actual con  $H_0$  y la explicación alternativa del investigador con  $H_a$ .

El rechazo de la teoría actual ocurrirá entonces sólo cuando la evidencia es mucho más compatible con la nueva teoría.

En muchas situaciones,  $H_a$  se conoce como "hipótesis del investigador", puesto que es la pretensión que al investigador en realidad le gustaría validar.

La palabra nulo significa "sin ningún valor, efecto o consecuencia", la que sugiere que  $H_0$  no deberá ser identificada con la hipótesis de ningún cambio (de la opinión actual), ninguna diferencia, ninguna mejora, y así sucesivamente.

Supongamos, por ejemplo, que 10% de todas las tarjetas de circuito producidas por un cierto fabricante durante un periodo reciente estaban defectuosas.

Un ingeniero ha sugerido un cambio del proceso de producción en la creencia de que dará por resultado una proporción reducida de tarjetas defectuosas.

Sea  $p$  la proporción verdadera de tarjetas defectuosas que resultan del proceso cambiado.

Entonces la hipótesis de investigación en la cual recae el agobio de comprobación, es la aseveración de que  $p < 0.10$ .

Por consiguiente la hipótesis alternativa es  $H_a : p < 0.10$ .

En el tratamiento de la prueba de hipótesis,  $H_0$  siempre será formulada como una afirmación de igualdad.

Si  $\theta$  denota el parámetro de interés, la hipótesis nula tendrá la forma  $H_0 : \theta = \theta_0$  donde  $\theta_0$  es un número específico llamado **valor nulo** del parámetro (valor pretendido para  $\theta$  por la hipótesis nula).

Como ejemplo, consideremos la situación de la tarjeta de circuito que se acaba de discutir.

La hipótesis alternativa sugerida fue  $H_a : p < 0.10$ , la pretensión de que la modificación del proceso redujo la proporción de tarjetas defectuosas.

Una opción natural de  $H_0$  en esta situación es la pretensión de que  $p \geq 0.10$  de acuerdo a la cual el nuevo proceso no es mejor o peor que el actualmente utilizado.

En su lugar se considerará  $H_0 : p = 0.10$  contra  $H_a : p < 0.10$ .

El razonamiento para utilizar esta hipótesis nula simplificada es que cualquier procedimiento de decisión razonable para decidir entre  $H_0 : p = 0.10$  y  $H_a : p < 0.10$  también será razonable para decidir entre la pretensión de que  $p \geq 0.10$  y  $H_a$ .

Se prefiere utilizar una  $H_0$  simplificada porque tiene ciertos beneficios técnicos, los que en breve serán aparentes.

La alternativa de la hipótesis nula  $H_0 : \theta = \theta_0$  se verá como una de las siguientes tres aseveraciones:

- ❶  $H_a : \theta > \theta_0$  (en cuyo caso la hipótesis nula implícita es  $\theta \leq \theta_0$ ),
- ❷  $H_a : \theta < \theta_0$  (por consiguiente la hipótesis implícita nula establece que  $\theta \geq \theta_0$ ) o
- ❸  $H_a : \theta \neq \theta_0$ .

Por ejemplo, sea  $\sigma$  la desviación estándar de la distribución de diámetros internos (pulgadas) de un cierto tipo de manguito de metal.

Si se decidió utilizar el manguito a menos que la evidencia muestral demuestre conclusivamente que  $\sigma > 0.001$ , la hipótesis apropiada sería  $H_0 : \sigma = 0.001$  contra  $H_a : \sigma > 0.001$ .

El número  $\theta_0$  que aparece tanto en  $H_0$  como en  $H_a$  (separa la alternativa de la nula) se llama **valor nulo**.

## Procedimientos de prueba

Un procedimiento de prueba es una regla, basada en datos muestrales, para decidir si rechazar  $H_0$ .

Una prueba de  $H_0 : p = 0.10$  contra  $H_a : p < 0.10$  en el problema de tarjetas de circuito podría estar basado en examinar una muestra aleatoria de  $n = 200$  tarjetas.

Sea  $X$  el número de tarjetas defectuosas en la muestra, una variable aleatoria binomial;  $x$  representa el valor observado de  $X$ .

Si  $H_0$  es verdadera,  $E(X) = n \cdot p = 200 \cdot 0.10 = 20$ , en tanto que se pueden esperar menos de 20 tarjetas defectuosas si  $H_a$  es verdadera.

Un valor de  $x$  un poco por debajo de 20 no contradice fuertemente a  $H_0$ , así que es razonable rechazar  $H_0$  sólo si  $x$  es sustancialmente menor que 20.

Un procedimiento de prueba como ése es rechazar  $H_0$  si  $x \leq 15$  y no rechazarla de lo contrario.



Este procedimiento consta de dos constituyentes:

- 1 un estadístico de prueba o función de los datos muestrales utilizados para tomar la decisión y
- 2 una región de rechazo compuesta de aquellos valores  $x$  con los cuales  $H_0$  será rechazada a favor de  $H_a$ .

De acuerdo a la regla que se acaba de sugerir, la región de rechazo se compone de  $x = 0, 1, 2, \dots, 15$ ,  $H_0$  no será rechazada si  $x = 16, 17, \dots, 199, 200$ .

### Procedimientos de prueba

Un procedimiento de prueba se especifica como sigue:

- 1 Un **estadístico de prueba**, una función de los datos muestrales en los cuales ha de basarse la decisión (rechazar  $H_0$  o no rechazar  $H_0$ )
- 2 Una **región de rechazo**, el conjunto de todos los valores estadísticos de prueba por los cuales  $H_0$  será rechazada.

La hipótesis nula será rechazada entonces si y sólo si el valor estadístico de prueba observado o calculado queda en la región de rechazo.

Como otro ejemplo, supongamos que una compañía tabacalera afirma que el contenido de nicotina prometido  $\mu$  de los cigarrillos marca B es (cuando mucho) de 1.5 mg.

Sería imprudente rechazar la afirmación del fabricante sin una fuerte evidencia contradictoria, así que una formulación del problema apropiada es probar  $H_0 : \mu = 1.5$  con  $H_a : \mu > 1.5$ .

Consideremos una regla de decisión basada en analizar una muestra aleatoria de 32 cigarrillos.

Sea  $\bar{X}$  el contenido de nicotina promedio muestral.

Si  $H_0$  es verdadera  $E(\bar{X}) = \mu = 1.5$ , en tanto que si  $H_0$  es falsa, se espera que  $\bar{X}$  exceda de 1.5.

Una fuerte evidencia contra  $H_0$  es proporcionada por un valor de  $\bar{x}$  que exceda considerablemente de 1.5.

Por consiguiente se podría utilizar  $\bar{X}$  como un estadístico de prueba junto con la región de rechazo  $\bar{x} \geq 1.6$ .

Tanto en el ejemplo de tarjetas de circuito como en el de contenido de nicotina, la selección del estadístico de prueba y la forma de la región de rechazo tienen sentido intuitivamente.

Sin embargo, la selección del valor de corte utilizado para especificar la región de rechazo es un tanto arbitraria.

En lugar de rechazar  $H_0 : p = 0.10$  a favor de  $H_a : p < 0.10$  cuando  $x \leq 15$ , se podría utilizar la región de rechazo  $x \leq 14$ .

En esta región,  $H_0$  no sería rechazada si se observaran 15 tarjetas defectuosas, mientras que esta ocurrencia conduciría al rechazo de  $H_0$  si se emplea la región inicialmente sugerida.

Asimismo, se podría utilizar la región de rechazo  $\bar{x} \geq 1.55$  en el problema de contenido de nicotina en lugar de la región  $\bar{x} \geq 1.60$ .

## Errores en la prueba de hipótesis

La base para elegir una región de rechazo particular radica en la consideración de los errores que se podrían presentar al sacar una conclusión.

Consideremos la región de rechazo  $x \leq 15$  en el problema de tarjetas de circuito.

Aun cuando  $H_0 : p = 0.10$  sea verdadera, podría suceder que una muestra inusual dé por resultado  $x = 13$ , de modo que  $H_0$  sea erróneamente rechazada.

Por otra parte, aun cuando  $H_a : p < 0.10$  sea verdadera, una muestra inusual podría dar  $x = 20$ , en cuyo caso  $H_0$  no sería rechazada, de nueva cuenta una conclusión incorrecta.

Por lo tanto es posible que  $H_0$  pueda ser rechazada cuando sea verdadera o que  $H_0$  no pueda ser rechazada cuando sea falsa.

Estos posibles errores no son consecuencia de una región de rechazo tontamente seleccionada.

Cualquiera de estos dos errores podría presentarse cuando se emplea la región  $x \leq 14$ , o cuando se utiliza cualquier otra región.

### Definición

Un **error de tipo I** consiste en rechazar la hipótesis nula cuando es verdadera.

Un **error de tipo II** implica no rechazar  $H_0$  cuando  $H_0$  es falsa.

En el problema de contenido de nicotina, un error de tipo I consiste en rechazar la afirmación del fabricante de que  $\mu = 1.5$  cuando en realidad es cierta.

Si se emplea la región de rechazo  $\bar{x} \geq 1.6$  podría suceder que  $\bar{x} = 1.63$  aun cuando  $\mu = 1.5$ , con el resultado de un error de tipo I.

Alternativamente, puede ser que  $H_0$  sea falsa y no obstante  $\bar{x} = 1.52$ , lo que conduciría a que  $H_0$  no sea rechazada (un error de tipo II).

En el mejor de todos los mundos posibles, se podrían desarrollar procedimientos de prueba en los cuales ningún tipo de error es posible.

Sin embargo, este ideal puede ser alcanzado sólo si la decisión se basa en el examen de toda la población.

La dificultad con la utilización de un procedimiento basado en datos muestrales es que debido a la variabilidad del muestreo, el resultado podría ser una muestra no representativa.

Aun cuando  $E(\bar{X}) = \mu$ , el valor observado  $\bar{x}$  puede diferir sustancialmente de  $\mu$  (por lo menos si  $n$  es pequeño).

Por consiguiente cuando  $\mu = 1.5$  en la situación de la nicotina,  $\bar{x}$  puede ser mucho más grande que 1.5 y el resultado sería el rechazo erróneo de  $H_0$ .

Alternativamente, puede ser que  $\mu = 1.6$  y no obstante que se observe una  $\bar{x}$  mucho más pequeña que este valor, lo que conduce a un error de tipo II.

En lugar de demandar procedimientos sin errores, habrá que buscar procedimientos con los cuales sea improbable que ocurra cualquier tipo de error.

Es decir, un buen procedimiento es uno con el cual la probabilidad de cometer cualquier tipo de error es pequeña.

La selección de un valor de corte en una región de rechazo particular fija las probabilidades de errores de tipo I y tipo II.

Estas probabilidades de error son tradicionalmente denotadas por  $\alpha$  y  $\beta$ , respectivamente.

Como  $H_0$ , especifica un valor único del parámetro, existe un solo valor de  $\alpha$ .

Sin embargo, existe un valor diferente de  $\beta$  por cada valor del parámetro compatible con  $H_a$ .

## Ejemplo

Se sabe que un cierto tipo de automóvil no sufre daños visibles el 25% del tiempo en pruebas de choque a 10 mph.

Se ha propuesto un diseño de defensa modificado en un esfuerzo por incrementar este porcentaje.

Sea  $p$  la proporción de todos los choques a 10 mph con esta nueva defensa en los que no producen daños visibles.

Las hipótesis a ser tratadas son  $H_0 : p = 0.25$  (ninguna mejora) contra  $H_a : p > 0.25$ .

La prueba se basará en un experimento que implica  $n = 20$  choques independientes con prototipos del nuevo diseño.

Intuitivamente,  $H_0$  deberá ser rechazada si un número sustancial de los choques no muestra daños.



Consideremos el siguiente procedimiento de prueba:

- Estadístico de prueba:  $X$  = número de choques sin daños visibles
- Región de rechazo:  $R_8 = 8, 9, 10, \dots, 19, 20$ ; es decir, rechazar  $H_0$  si  $x \geq 8$ , donde  $x$  es el valor observado del estadístico de prueba.

Esta región de rechazo se llama **de cola superior** y se compone de sólo grandes valores del estadístico de prueba.

Cuando  $H_0$  es verdadera, la distribución de probabilidad de  $X$  es binomial con  $n = 20$  y  $p = 0.25$ .

Entonces

$$\alpha = P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando es verdadera})$$

$$\alpha = P(X \geq 8 \text{ cuando } X \sim B(20, 0.25)) = 1 - P(X < 8) = 1 - 0.898 = 0.102$$

Es decir, cuando  $H_0$  en realidad es verdadera, aproximadamente el 10% de todos los experimentos compuestos de 20 choques darían por resultado que  $H_0$  fuera rechazada incorrectamente (un error de tipo I).

En contraste con  $\alpha$ , no hay una sola  $\beta$ .

En su lugar, hay una  $\beta$  diferente por cada  $p$  diferente que exceda de 0.25.

Por consiguiente hay un valor de  $\beta$  con  $p = 0.3$  (en cuyo caso  $X \sim B(20, 0.3)$ ), otro valor de  $\beta$  con  $p = 0.5$  y así sucesivamente.

Por ejemplo,

$$\beta(0.3) = P(\text{error de tipo II cuando } p = 0.3)$$

$$\beta(0.3) = P(H_0 \text{ no es rechazada cuando es falsa porque } p = 0.3)$$

$$\beta(0.3) = P(X \leq 7 \text{ cuando } X \sim B(20, 0.3)) = P(X \leq 7) = 0.772$$

Cuando  $p$  es en realidad 0.3 y no 0.25 (un "pequeño" alejamiento de  $H_0$ ), ¡aproximadamente el 77% de todos los experimentos de este tipo darían por resultado que  $H_0$  fuera incorrectamente rechazada!

La tabla adjunta muestra  $\beta$  para valores seleccionados de  $p$  (cada uno calculado para la región de rechazo  $R_8$ ).

Claramente,  $\beta$  disminuye conforme el valor de  $p$  se aleja hacia la derecha del valor nulo 0.25.

Intuitivamente, mientras más grande es el alejamiento de  $H_0$ , menos probable es que dicho alejamiento no sea detectado.

| $p$        | 0.3   | 0.4   | 0.5   | 0.6   | 0.7   | 0.8   |
|------------|-------|-------|-------|-------|-------|-------|
| $\beta(p)$ | 0.772 | 0.416 | 0.132 | 0.021 | 0.001 | 0.000 |

El procedimiento de prueba propuesto sigue siendo razonable para poner a prueba la hipótesis nula más realista de que  $p \leq 0.25$ .

En este caso, ya no existe una sola  $\alpha$ , sino que hay una por cada  $p$  que sea cuando mucho de 0.25:  $\alpha(0.25)$ ,  $\alpha(0.23)$ ,  $\alpha(0.20)$ ,  $\alpha(0.15)$  y así sucesivamente.

Es fácil verificar, no obstante, que  $\alpha(p) < \alpha(0.25) = 0.102$  si  $p < 0.25$ .

Es decir, el valor más grande de  $\alpha$  ocurre con el valor límite 0.25 entre  $H_0$  y  $H_a$ .

Por consiguiente si  $\alpha$  es pequeña para la hipótesis nula simplificada, también igual o más pequeña para la  $H_0$  realista.

## Ejemplo (Continuación)

Utilizamos el mismo experimento y el estadístico de prueba  $X$  como previamente se describió en el problema de la defensa de automóvil pero ahora consideremos la región de rechazo  $R_9 = \{9, 10, \dots, 20\}$ .

Como  $X$  sigue teniendo una distribución binomial con parámetros  $n = 20$  y  $p$ ,

$$\alpha = P(H_0 \text{ es rechazada cuando } p = 0.25)$$

$$\alpha = P(X \geq 9 \text{ cuando } X \sim B(20, 0.25)) = 1 - P(X \leq 8) = 0.041$$

La probabilidad de error de tipo I se redujo con el uso de la nueva región de rechazo.

Sin embargo, se pagó un precio por esta reducción:

$$\beta(0.3) = P(H_0 \text{ es rechazada cuando } p = 0.3)$$

$$\beta(0.3) = P(X \leq 8 \text{ cuando } X \sim B(20, 0.3)) = P(X \leq 8) = 0.887$$

$$\beta(0.5) = P(X \leq 8 \text{ cuando } X \sim B(20, 0.5)) = P(X \leq 8) = 0.252$$

Ambas  $\beta$  son más grandes que las probabilidades de error correspondientes 0.772 y 0.132 para la región  $R_8$ .

En retrospectiva, no es sorprendente;  $\alpha$  se calcula sumando las probabilidades de los valores estadísticos de prueba **en la región de rechazo**, en tanto que  $\beta$  es la probabilidad de que  $X$  quede **en el complemento** de la región de rechazo.

Al hacerse más pequeña la región de rechazo debe reducirse  $\alpha$  al mismo tiempo que se incrementa  $\beta$  con cualquier valor alternativo fijo del parámetro.

## Ejemplo

Se sabe que el tiempo de secado de un cierto tipo de pintura en condiciones de prueba especificadas está normalmente distribuido con valor medio de 75 min y desviación estándar de 9 min.

Algunos químicos propusieron un nuevo aditivo para reducir el tiempo de secado.

Se cree que los tiempos de secado con este aditivo permanecerán normalmente distribuidos con  $\sigma = 9$ .

Debido al gasto asociado con el aditivo, la evidencia deberá sugerir fuertemente una mejora en el tiempo de secado promedio antes de que se adopte semejante conclusión.

Sea  $\mu$  el tiempo de secado promedio verdadero cuando se utiliza el aditivo.

Las hipótesis apropiadas son  $H_0 : \mu = 75$  contra  $H_a : \mu < 75$ .

Sólo si  $H_0$  puede ser rechazada será declarado exitoso y utilizado.

Los datos experimentales tienen que estar compuestos de tiempos de secado de  $n = 25$  especímenes de prueba.

Sean  $X_1, \dots, X_{25}$  los 25 tiempos de secado, una muestra aleatoria de tamaño 25 de una distribución normal con valor medio  $\mu$  y desviación estándar  $\sigma = 9$ .

El tiempo de secado medio muestral  $\bar{X}$  tiene entonces una distribución normal con valor esperado  $\mu_{\bar{X}} = \mu$  y desviación estándar  $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 9/\sqrt{25} = 1.80$ .

Cuando  $H_0$  es verdadera,  $\mu_{\bar{X}} = 75$ , así que un valor  $\bar{x}$  un poco menor que 75 no contradeciría fuertemente a  $H_0$ .

Una región razonable de rechazo tiene la forma  $\bar{X} \leq c$ , donde el valor de corte  $c$  es adecuadamente seleccionado.



Consideremos la opción  $c = 70.8$ , de modo que el procedimiento de prueba se componga del estadístico de prueba  $\bar{X}$  y una región de rechazo  $\bar{x} \leq 70.8$ .

Debido a que la región de rechazo se compone de sólo valores pequeños del estadístico de prueba, se dice que ésta es **de cola pequeña**.

El cálculo de  $\alpha$  y  $\beta$  ahora implica una estandarización de rutina de  $\bar{X}$  seguida por una referencia a las probabilidades normales estándar.

$$\alpha = P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando es verdadera})$$

$$\alpha = P(\bar{X} \leq 70.8 \text{ cuando } \bar{X} \sim N(\mu_{\bar{X}} = 75, \sigma_{\bar{X}} = 1.8))$$

$$\alpha = P\left(Z \leq \frac{70.8 - 75}{1.8}\right) = P(Z \leq -2.33) = 0.01$$

$$\beta(72) = P(\text{error de tipo II cuando } \mu = 72)$$

$$\beta(72) = P(H_0 \text{ no es rechazada cuando es falsa porque } \mu = 72)$$

$$\beta(72) = P(\bar{X} > 70.8 \text{ cuando } \bar{X} \sim N(\mu_{\bar{X}} = 72, \sigma_{\bar{X}} = 1.8))$$

$$\beta(72) = 1 - P(Z \leq \frac{70.8 - 72}{1.8}) = 1 - P(Z \leq -0.67) = 1 - 0.2514 = 0.7486$$

$$\beta(70) = 1 - P(Z \leq \frac{70.8 - 70}{1.8}) = 0.3300$$

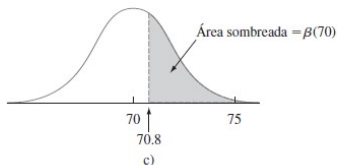
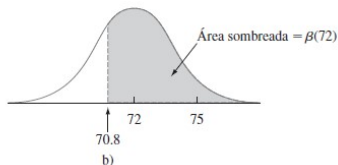
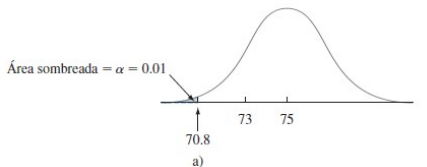
$$\beta(67) = 1 - P(Z \leq \frac{70.8 - 67}{1.8}) = 0.0174$$

Con el procedimiento de prueba especificado, sólo el 1% de todos los experimentos realizados como se describió darán por resultado que  $H_0$  sea rechazada cuando en realidad es verdadera.

No obstante, la probabilidad de un error de tipo II es muy grande cuando  $\mu = 72$  (sólo un pequeño alejamiento de  $H_0$ ), un poco menor cuando  $\mu = 70$  y bastante pequeño cuando  $\mu = 67$  (un alejamiento muy sustancial de  $H_0$ ).

Estas probabilidades de error se ilustran en la figura siguiente.

Observemos que  $\alpha$  se calcula con la distribución de probabilidad del estadístico de prueba cuando  $H_0$  es verdadera, en tanto que la determinación de  $\beta$  requiere conocer la distribución del estadístico cuando  $H_0$  es falsa.



## Ejemplo

$\alpha$  y  $\beta$  ilustradas para el ejemplo:

- a la distribución de  $\bar{X}$  cuando  $\mu = 75$  ( $H_0$  verdadera);
- b la distribución de  $\bar{X}$  cuando  $\mu = 72$  ( $H_0$  falsa);
- c la distribución de  $\bar{X}$  cuando  $\mu = 70$  ( $H_0$  falsa).

Como en el ejemplo anterior, si se considera la hipótesis nula más realista  $\mu \geq 75$ , existe una  $\alpha$  por cada valor de parámetro con el cual  $H_0$  es verdadera:  $\alpha(75)$ ,  $\alpha(75.8)$ ,  $\alpha(76.5)$ , y así sucesivamente.

Es fácil verificar que  $\alpha(75)$  es la más grande de todas estas probabilidades de error de tipo I.

Enfocándose en el valor límite equivale a trabajar explícitamente con el "peor caso".

El uso del valor de corte  $c = 70.8$  en el ejemplo de secado de la pintura dio por resultado un valor de  $\alpha$  muy pequeño (0.01) pero  $\beta$  grande.

Consideremos el mismo experimento y pruebe el estadístico de prueba  $\bar{X}$  con la nueva región de rechazo  $\bar{x} \leq 72$ .

Como  $\bar{X}$  sigue siendo normalmente distribuida con valor medio  $\mu_{\bar{X}} = \mu$  y  $\sigma_{\bar{X}} = 1.8$ ,

$$\alpha = P(H_0 \text{ es rechazada cuando es verdadera})$$

$$\alpha = P(\bar{X} \leq 72 \text{ cuando } \bar{X} \sim N(\mu_{\bar{X}} = 75, \sigma_{\bar{X}} = 1.8))$$

$$\alpha = P\left(Z \leq \frac{72 - 75}{1.8}\right) = P(Z \leq -1.67) = 0.0475 \approx 0.05$$

$$\beta(72) = P(H_0 \text{ no es rechazada cuando } \mu = 72)$$

$$\beta(72) = P(\bar{X} > 72 \text{ cuando } \bar{X} \sim N(\mu_{\bar{X}} = 72, \sigma_{\bar{X}} = 1.8))$$

$$\beta(72) = 1 - P\left(Z \leq \frac{72 - 72}{1.8}\right) = 1 - P(Z \leq 0) = 0.5$$

$$\beta(70) = 1 - P\left(Z \leq \frac{72 - 70}{1.8}\right) = 0.1335$$

$$\beta(67) = 1 - P\left(Z \leq \frac{72 - 67}{1.8}\right) = 0.0027$$

El cambio del valor de corte agrandó la región de rechazo (incluye más valores de  $\bar{x}$ ) y el resultado es una reducción de  $\beta$  por cada  $\mu$  fija menor que 75.

Sin embargo,  $\alpha$  en esta nueva región se ha incrementado desde el valor previo 0.01 hasta aproximadamente 0.05.

Si una probabilidad de error de tipo I así de grande puede ser tolerada, se prefiere la segunda región ( $c = 72$ ) a la primera ( $c = 70.8$ ) debido a las  $\beta$  más pequeñas.



La especificación de un valor de corte para la región de rechazo en los ejemplos que se acaban de considerar fue algo arbitraria.

El uso de  $R_8 = \{8, 9, \dots, 20\}$  en el ejemplo inicial dio por resultado  $\alpha = 0.102$ ,  $\beta(0.3) = 0.772$  y  $\beta(0.5) = 0.132$ .

Muchos pensarán que estas probabilidades de error son intolerablemente grandes.

Quizá puedan reducirse si se cambia el valor de corte.

Los resultados de estos ejemplos pueden ser generalizados de la siguiente manera.

### Proposición

Supongamos que un experimento y un tamaño de muestra están fijos y que se selecciona un estadístico de prueba.

Entonces si se reduce el tamaño de la región de rechazo para obtener un valor más pequeño de  $\alpha$  se obtiene un valor más grande de  $\beta$  con cualquier valor de parámetro particular compatible con  $H_a$ .

Esta proposición expresa que una vez que el estadístico de prueba y  $n$  están fijos, no existe una región de rechazo que haga que al mismo tiempo  $\alpha$  y  $\beta$  sean pequeños.

Se debe seleccionar una región para establecer un compromiso entre  $\alpha$  y  $\beta$ .

Debido a las indicaciones sugeridas para especificar  $H_0$  y  $H_a$ , casi siempre un error de tipo I es más serio que uno de tipo II (esto en general se puede lograr mediante la selección apropiada de las hipótesis).

El método seguido por la mayoría de los practicantes de la estadística es especificar el valor más grande de  $\alpha$  que pueda ser tolerado y encontrar una región de rechazo que tenga valor de  $\alpha$  en lugar de cualquier otro más pequeño.

Esto hace a  $\beta$  tan pequeño como sea posible sujeto al límite en  $\alpha$ .

El valor resultante de  $\alpha$  a menudo se conoce como **nivel de significación** de la prueba.

Niveles tradicionales de significación son 0.10, 0.05 y 0.01, aunque el nivel en cualquier problema particular dependerá de la seriedad de un error de tipo I: mientras más serio es este error, más pequeño deberá ser el nivel de significación.

El procedimiento de prueba correspondiente se llama **prueba de nivel  $\alpha$**  (p. ej., prueba de nivel 0.05 o prueba de nivel 0.01).

Una prueba con nivel de significación  $\alpha$  es una donde la probabilidad de error de tipo I se controla al nivel especificado.

# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

La discusión general en el tema 7 de intervalos de confianza para una media de población se enfocó en tres casos diferentes.

A continuación se desarrollan procedimientos para estos mismos tres casos.

## Caso I: Una población normal con $\sigma$ conocida

Aun cuando la suposición de que el valor de  $\sigma$  es conocido rara vez se cumple en la práctica, este caso proporciona un buen punto de partida debido a la facilidad con que los procedimientos generales y sus propiedades pueden ser desarrollados.

La hipótesis nula en los tres casos propondrá que  $\mu$  tiene un valor numérico particular, el valor nulo, el cual será denotado por  $\mu_0$ . Sean  $X_1, \dots, X_n$  una muestra aleatoria de tamaño  $n$  de la población normal.

Entonces la media muestral  $\bar{X}$  tiene una distribución normal con valor esperado  $\mu_{\bar{X}} = \mu$  y desviación estándar  $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ .

Cuando  $H_0$  es verdadera  $\mu_{\bar{X}} = \mu_0$ .

Consideremos ahora el estadístico  $Z$  obtenido estandarizando  $\bar{X}$  de conformidad con la suposición de que  $H_0$  es verdadera:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Al sustituir la media muestral calculada  $\bar{x}$  se obtiene  $z$ , la distancia entre  $\bar{x}$  y  $\mu_0$  expresada en "unidades de desviación estándar".

Por ejemplo, si la hipótesis nula es  $H_0 : \mu = 100$ ,  $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 10/\sqrt{25} = 2.0$  y  $\bar{x} = 103$ , entonces el valor estadístico de prueba es  $z = (103 - 100)/2.0 = 1.5$ .

Es decir, el valor observado de  $\bar{x}$  es 1.5 desviaciones estándar (de  $\bar{X}$ ) más grande de lo que se espera sea cuando  $H_0$  es verdadera.

El estadístico  $Z$  es una medida natural de la distancia  $\bar{X}$ , el estimador de  $\mu$  y su valor esperado cuando  $H_0$  es verdadera.

Si esta distancia es demasiado grande en una dirección consistente con  $H_a$ , la hipótesis nula deberá ser rechazada.



Supongamos primero que la hipótesis alternativa tiene la forma

$$H_a : \mu > 0.$$

Entonces un valor de  $\bar{x}$  menor que  $\mu_0$  indudablemente no apoya a  $H_a$ .

Tal  $\bar{x}$  corresponde a un valor negativo de  $z$  (puesto que  $\bar{x} - \mu_0$  es negativa y el divisor de  $\sigma/\sqrt{n}$  es positivo).

Del mismo modo, un valor de  $\bar{x}$  que exceda de  $\mu_0$  por sólo una pequeña cantidad (correspondiente a  $z$  la cual es positiva aunque pequeña) no sugiere que  $H_0$  deberá ser rechazada a favor de  $H_a$ .

El rechazo de  $H_0$  es apropiado sólo cuando  $\bar{x}$  excede considerablemente de  $\mu_0$ , es decir, cuando el valor de  $z$  es positivo y grande.

En suma, la región de rechazo apropiada basada en el estadístico de prueba  $Z$  en lugar de  $\bar{X}$  tiene la forma  $z \geq c$ .

Como se discutió anteriormente, el valor de corte  $c$  deberá ser elegido para controlar la probabilidad de un error de tipo I al nivel  $\alpha$  deseado.

Esto es fácil de lograr porque la distribución del estadístico de prueba  $Z$  cuando  $H_0$  es verdadera es la distribución normal estándar (es por eso que  $\mu_0$  se restó al estandarizar).

El valor  $c$  de corte requerido es el valor crítico  $z$  que captura el área de la cola superior  $\alpha$  bajo la curva  $z$ .

Como ejemplo, sea  $c = 1.645$ , el valor que captura el área de cola 0.05 ( $z_\alpha = z_{0.05} = 1.645$ ).

Entonces,

$$\alpha = P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando } H_0 \text{ es verdadera})$$

$$\alpha = P(Z \geq 1.645 \text{ cuando } Z \sim N(0, 1)) = 1 - P(Z \leq 1.645) = 0.05$$

Más generalmente, la región de rechazo  $z \geq z_\alpha$  tiene un probabilidad de error de tipo I  $\alpha$ .

El procedimiento de prueba es de **cola superior** porque la región de rechazo se compone de sólo valores grandes del estadístico de prueba.

Un razonamiento análogo para la hipótesis alternativa  $H_a : \mu < \mu_0$  sugiere una región de rechazo de la forma  $z \leq c$ , donde  $c$  es un número negativo adecuadamente seleccionado ( $\bar{x}$  aparece muy debajo de  $\mu_0$  si y sólo si  $z$  es bastante negativo).

Como  $Z$  tiene una distribución normal estándar cuando  $H_0$  es verdadera, con  $c = -z_\alpha$  da  $P(\text{error de tipo I}) = \alpha$ .

Esta es una prueba de **cola inferior**.

Por ejemplo,  $z_{0.10} = 1.28$  implica que la región de rechazo  $z \leq -1.28$  especifica una prueba con nivel de significación de 0.10.

Por último, cuando la hipótesis alternativa es  $H_a : \mu \neq \mu_0$ ,  $H_0$  deberá ser rechazada si  $\bar{x}$  está muy lejos a ambos lados de  $\mu_0$ .

Esto equivale a rechazar  $H_0$  si  $z \geq c$  o si  $z \leq -c$ .

Supongamos que se desea  $\alpha = 0.05$ .

Entonces,

$0.05 = P(Z \geq c \text{ o } Z \leq -c \text{ cuando } Z \text{ tiene una distribución normal estándar})$

$0.05 = P(Z \geq c) + P(Z \leq -c) = 1 - P(Z \leq c) + P(Z \leq -c) = 2 \cdot [1 - P(Z \leq c)]$

Por consiguiente  $c$  es tal que  $1 - P(Z \leq c)$ , el área bajo la curva  $z$  a la derecha de  $c$ , es 0.025 (¡y no 0.05!).

De acuerdo con la tabla normal estándar,  $c = 1.96$  y la región de rechazo  $z \geq 1.96$  o  $z \leq -1.96$ .

Con cualquier  $\alpha$ , la región de rechazo de **dos colas**  $z \geq z_{\alpha/2}$  o  $z \leq -z_{\alpha/2}$  tiene una probabilidad  $\alpha$  de error de tipo I (puesto que el área  $\alpha/2$  está capturada debajo de cada una de las dos colas de la curva  $z$ ).

De nueva cuenta, la razón clave para utilizar el estadístico de prueba estandarizado  $Z$  es que como  $Z$  tiene una distribución conocida cuando  $H_0$  es verdadera (normal estándar), es fácil de obtener una región de rechazo con probabilidad de error de tipo I mediante un valor crítico apropiado.

El procedimiento de prueba en el caso I se resume en el cuadro adjunto y las regiones de rechazo correspondientes se ilustran en la siguiente figura.

## Caso I

Hipótesis nula:  $H_0 : \mu = \mu_0$

Valor del estadístico de prueba:  $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Hipótesis alternativa      Región de rechazo para la prueba de nivel  $\alpha$

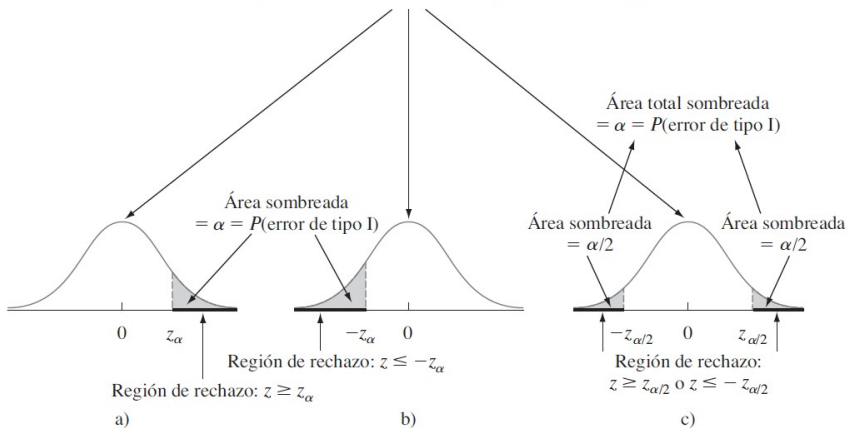
$H_a : \mu > \mu_0$        $z \geq z_\alpha$  (prueba de cola superior)

$H_a : \mu < \mu_0$        $z \leq -z_\alpha$  (prueba de cola inferior)

$H_a : \mu \neq \mu_0$        $z \geq z_{\alpha/2}$  o  $z \leq -z_{\alpha/2}$  (prueba de dos colas)

Regiones de rechazo para pruebas  $z$  : a) prueba de cola superior;  
 b) prueba de cola inferior; c) prueba de dos colas.

Curva  $z$  (distribución de probabilidad del estadístico de prueba  $Z$  cuando  $H_0$  es verdadera)



Se recomienda utilizar la siguiente secuencia de pasos cuando se prueben hipótesis con respecto a un parámetro.

- 1 Identificar el parámetro de interés y describirlo en el contexto de la situación del problema.
- 2 Determinar el valor nulo y formular la hipótesis nula.
- 3 Formular la hipótesis alternativa apropiada.
- 4 Dar la fórmula para el valor calculado del estadístico de prueba (sustituyendo el valor nulo y los valores conocidos de cualquier otro parámetro, pero no aquellos de cualquier cantidad basada en una muestra).
- 5 Establecer la región de rechazo para el nivel de significación seleccionado  $\alpha$ .
- 6 Calcular cualquier cantidad muestral necesaria, sustituir en la fórmula para el estadístico de prueba y calcular dicho valor.
- 7 Decidir si  $H_0$  debe ser rechazada y expresar esta conclusión en el contexto del problema.

La formulación de hipótesis (pasos 2 y 3) deberá ser realizada antes de examinar los datos.



## Ejemplo

Un fabricante de sistemas rociadores utilizados como protección contra incendios en edificios de oficinas afirma que la temperatura de activación del sistema promedio verdadera es de  $130^{\circ}\text{F}$ .

Una muestra de  $n = 9$  sistemas, cuando se somete a prueba, da una temperatura de activación promedio muestral de  $131.08^{\circ}\text{F}$ .

Si la distribución de los tiempos de activación es normal con desviación estándar de  $1.5^{\circ}\text{F}$ , ¿contradicen los datos la afirmación del fabricante a un nivel de significación  $\alpha = 0.01$ ?

- 1.- Parámetro de interés:  $\mu$  = temperatura de activación promedio verdadera.
- 2.- Hipótesis nula:  $H_0 : \mu = 130$  (valor nulo =  $\mu_0 = 130$ ).
- 3.- Hipótesis alternativa:  $H_a : \mu \neq 130$  (un alejamiento del valor declarado en una u otra dirección es de interés).

## Ejemplo (Continuación)

- 4.- Valor de estadístico de prueba:  $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$
- 5.- Región de rechazo: La forma de  $H_a$  implica el uso de una prueba de dos colas con región de rechazo  $\bullet$  de  $z \geq z_{0.005}$   $\bullet$   $z \leq -z_{0.005}$ . De acuerdo con la tabla normal estándar  $z_{0.005} = 2.58$ , así que se rechazaría  $H_0$  si  $z \geq 2.58$  o  $z \leq -2.58$ .
- 6.- Sustituyendo  $n = 9$  y  $\bar{x} = 131.08$ ,

$$z = \frac{131.08 - 130}{1.5/\sqrt{9}} = \frac{1.08}{0.5} = 2.16$$

Es decir, la media muestral observada es de un poco más de 2 desviaciones estándar sobre el valor que era de esperarse si  $H_0$  fuera verdadera.

- 7.- El valor calculado  $z = 2.16$  no queda en la región de rechazo ( $-2.58 < 2.16 < 2.58$ ), así que  $H_0$  no puede ser rechazada al nivel de significación de 0.01. Los datos no apoyan fuertemente la afirmación de que el promedio verdadero difiere del valor de diseño de 130.

## Caso II: Pruebas con muestras grandes

Cuando el tamaño de muestra es grande, las pruebas  $z$  en el caso I son fáciles de modificar para dar procedimientos de prueba válidos sin requerir una distribución de población normal o  $\sigma$  conocida.

El resultado clave se utilizó en el Tema 7 para justificar intervalos de confianza para muestra grande: Un tamaño de muestra grande  $n$  implica que la variable estandarizada

$$Z = \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}}$$

tiene **aproximadamente** una distribución normal estándar.

La sustitución del valor nulo  $\mu_0$  en lugar de  $\mu$  da el estadístico de prueba

$$Z = \frac{\bar{X} - \mu_0}{\hat{S}/\sqrt{n}}$$

la que tiene aproximadamente una distribución normal estándar cuando  $H_0$  es verdadera.

El uso de las regiones de rechazo dadas previamente para el caso I (p. ej.,  $z \geq z_\alpha$  cuando la hipótesis alternativa es  $H_a : \mu > \mu_0$ ) produce entonces procedimientos de prueba con los cuales el nivel de significación es aproximadamente (y no exactamente)  $\alpha$ .

Se utilizará de nuevo la regla empírica  $n > 40$  para caracterizar un tamaño de muestra grande.

## Ejemplo

Se utiliza un penetrómetro cónico dinámico (DCP, por sus siglas en inglés) para medir la resistencia de un material a la penetración (mm/golpe), a medida que el cono es insertado en pavimento o subrasante.

Supongamos que para una aplicación particular, se requiere que el valor penetración cónica dinámica promedio verdadera para un cierto tipo de pavimento sea menor que 30.

El pavimento no será utilizado a menos que exista evidencia concluyente de que la especificación fue satisfecha.

## Ejemplo (Continuación)

Formulemos y probemos las hipótesis apropiadas utilizando los datos siguientes ("Probabilistic Model for the Analysis of Dynamic Cone Penetrometer Test Values in Pavement Structure Evaluation", J. of Testing and Evaluation, 1999: 7-14:

14.1, 14.5, 15.5, 16.0, 16.0, 16.7, 16.9, 17.1, 17.5, 17.8, 17.8, 18.1, 18.2, 18.3, 18.3, 19.0, 19.2, 19.4, 20.0, 20.0, 20.8, 20.8, 21.0, 21.5, 23.5, 27.5, 27.5, 28.0, 28.3, 30.0, 30.0, 31.6, 31.7, 31.7, 32.5, 33.5, 33.9, 35.0, 35.0, 35.0, 36.7, 40.0, 40.0, 41.3, 41.7, 47.5, 50.0, 51.0, 51.8, 54.4, 55.0, 57.0

La penetración cónica dinámica media muestral es menor que 30.

Sin embargo, existe una cantidad sustancial de variación en los datos (coeficiente de variación muestral =  $\hat{s}/\bar{x} = 0.4265$ ), de modo que la media sea menor que el valor de corte de la especificación de diseño puede ser una consecuencia simplemente de la variabilidad muestral.

## Ejemplo (Continuación)

Si se representa el histograma se observa que no se asemeja en absoluto a una curva normal (y una curva de probabilidad normal no exhibe un patrón lineal), aunque las pruebas  $z$  con muestras grandes no requieren una distribución de población normal.

- 1.-  $\mu$  = valor de penetración cónica dinámica promedio verdadero.
- 2.- Hipótesis nula:  $H_0 : \mu = 30$  (valor nulo  $= \mu_0 = 30$ ).
- 3.- Hipótesis alternativa:  $H_a : \mu < 30$  (por consiguiente el parámetro no será utilizado a menos que la hipótesis nula sea rechazada).
- 4.-  $z = \frac{\bar{x} - 30}{\hat{s}/\sqrt{n}}$
- 5.- Una prueba con nivel de significación de 0.05 rechaza a  $H_0$  cuando  $z \leq -1.645$  (una prueba de cola inferior).

### Ejemplo (Continuación)

6.- Con  $n = 52$ ,  $\bar{x} = 28.76$  y  $\hat{s} = 12.2647$

$$z = \frac{28.76 - 30}{12.2647/\sqrt{52}} = \frac{-1.24}{1.701} = -0.73$$

7.- Como  $-0.73 > -1.645$ ,  $H_0$  no puede ser rechazada. No se cuenta con evidencia precisa para concluir que  $\mu < 30$ ; el uso del pavimento no se justifica.



### Caso III: Una distribución de población normal

Cuando  $n$  es pequeño, el teorema del límite central (TLC) ya no puede ser invocado para justificar el uso de una prueba con muestra grande.

Esta misma dificultad se presenta al obtener un intervalo de confianza con muestra pequeña (IC) para  $\mu$  en el tema 7.

El método utilizado aquí es el mismo que el usado allí.

Se supondrá que la distribución de población es por lo menos aproximadamente normal y se describirán los procedimientos de prueba cuya validez se fundamenta en esta suposición.

Se utilizó el resultado clave en el cual están basadas las pruebas con una media de población normal en el tema 7 para derivar el intervalo de confianza  $t$  para una muestra: Si  $X_1, X_2, \dots, X_n$  es una muestra aleatoria de una distribución normal, la variable estandarizada

$$T = \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}}$$

tiene una distribución  $t$  con  $n - 1$  grados de libertad (gl).

Consideremos que ponemos a prueba  $H_0 : \mu = \mu_0$  contra  $H_a : \mu > \mu_0$  utilizando el estadístico de prueba  $T = (\bar{X} - \mu_0)/(\hat{S}/\sqrt{n})$ .

Es decir, el estadístico de prueba resulta de estandarizar  $\bar{X}$  conforme a la suposición de que  $H_0$  es verdadera (utilizando  $\hat{S}/\sqrt{n}$ , la desviación estándar estimada de  $\bar{X}$  en lugar de  $\sigma/\sqrt{n}$ ).

Cuando  $H_0$  es verdadera, el estadístico de prueba tiene una distribución  $t$  con  $n - 1$  grados de libertad.

El conocimiento de la distribución del estadístico de prueba cuando  $H_0$  es verdadera (la "distribución nula") permite construir una región de rechazo para la cual la probabilidad de error de tipo I se controla al nivel deseado.

En particular, el uso del valor crítico  $t$  de cola superior  $t_{\alpha, n-1}$  para especificar la región de rechazo  $t \geq t_{\alpha, n-1}$  implica que

$$P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando es verdadera})$$

$$P(\text{error de tipo I}) = P(T \geq t_{\alpha, n-1} \text{ cuando } T \text{ tiene una distribución } t_{n-1})$$

$$P(\text{error de tipo I}) = \alpha$$

El estadístico de prueba es en realidad el mismo del caso de muestra grande pero se designa  $T$  para recalcar que la distribución nula es una distribución  $t$  con  $n - 1$  grados de libertad en lugar de la distribución normal estándar ( $z$ ).

La región de rechazo para la prueba  $t$  difiere de aquella para la prueba  $z$  sólo en que un valor crítico  $t_{\alpha, n-1}$  reemplaza al valor crítico  $z_{\alpha}$ .

Comentarios similares se aplican a alternativas para las cuales una prueba de cola inferior o de dos colas es apropiada.

### Caso III

Prueba  $t$  con una muestra

Hipótesis nula:  $H_0 : \mu = \mu_0$

Valor del estadístico de prueba:  $t = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}$

Hipótesis alternativa      Región de rechazo para la prueba de nivel  $\alpha$

$H_a : \mu > \mu_0$        $t \geq t_{\alpha, n-1}$  (prueba de cola superior)

$H_a : \mu < \mu_0$        $t \leq -t_{\alpha, n-1}$  (prueba de cola inferior)

$H_a : \mu \neq \mu_0$        $t \geq t_{\alpha/2, n-1}$  o  $t \leq -t_{\alpha/2, n-1}$  (prueba de dos colas)

## Ejemplo

Un entorno de trabajo bien diseñado y seguro puede contribuir en gran medida a incrementar la productividad.

Es especialmente importante que a los trabajadores no se les pida realizar tareas, tales como izar cosas, que excedan sus capacidades.

Los datos adjuntos sobre peso máximo de izamiento (MAWL, por sus siglas en inglés, en kg) para una frecuencia de cuatro izamientos/min fueron reportados en el artículo "The Effects of Speed, Frequency, and Load on Measured Hand Forces for a Floor-to-Knuckle Lifting Task" (Ergonomics, 1992: 833-843); se seleccionaron al azar sujetos de una población de varones saludables de 18 a 30 años de edad.

Suponiendo que el peso máximo de izamiento está normalmente distribuido ¿sugieren los datos siguientes que el peso máximo de izamiento medio de la población excede de 25?

25.8, 36.6, 26.3, 21.8, 27.2

## Ejemplo (Continuación)

Efectuemos una prueba con un nivel de significación de 0.05.

1.-  $\mu$  = peso de izamiento máximo medio de la población.

2.- Hipótesis nula:  $H_0 : \mu = 25$  (valor nulo =  $\mu_0 = 25$ ).

3.- Hipótesis alternativa:  $H_a : \mu > 25$ .

$$4.- t = \frac{\bar{x} - 25}{\hat{s}/\sqrt{n}}$$

5.- Rechazamos  $H_0$  si  $t \geq t_{\alpha, n-1} = t_{0.05, 4} = 2.132$ .

6.-  $\sum x_i = 137.7$  y  $\sum x_i^2 = 3911.97$ , con las cuales  $\bar{x} = 27.54$ ,  $\hat{s} = 5.47$ , y

$$t = \frac{27.54 - 25}{5.47/\sqrt{5}} = \frac{2.54}{2.45} = 1.04$$

7.- Como 1.04 no queda en la región de rechazo ( $1.04 < 2.132$ ),  $H_0$  no puede ser rechazada a un nivel de significación de 0.05. Aún es plausible que  $\mu$  sea (cuando mucho) de 25.

# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

Sea  $p$  la proporción de individuos u objetos en una población que poseen una propiedad especial (p. ej., carros con transmisión manual o fumadores que fuman cigarrillos con filtro).

Si un individuo u objeto con la propiedad es etiquetado como éxito (E), entonces  $p$  es la proporción de población de éxitos.

Las pruebas relacionadas con  $p$  se basarán en una muestra aleatoria de tamaño  $n$  de la población.

Siempre que  $n$  sea pequeño con respecto al tamaño de la población,  $X$  (el número de éxitos en la muestra) tiene (aproximadamente) una distribución binomial.

Además, si  $n$  es grande, tanto  $X$  como el estimador  $\hat{p} = X/n$  están normalmente distribuidos.

Primero se consideran pruebas con muestras grandes basadas en este último hecho y luego se acude al caso de muestra pequeña que usa directamente la distribución binomial.



## Pruebas con muestra grande

Las pruebas con muestra grande relacionadas con  $p$  son un caso especial de los procedimientos con muestra grande para un parámetro  $\theta$ .

Sea  $\hat{\theta}$  un estimador de  $\theta$  que es (por lo menos aproximadamente) insesgado y que tiene aproximadamente una distribución normal.

La hipótesis nula tiene la forma  $H_0 : \theta = \theta_0$ , donde  $\theta_0$  denota un número (el valor nulo) apropiado al contexto del problema.

Supongamos que cuando  $H_0$  es verdadero, la desviación estándar de  $\hat{\theta}$ ,  $\sigma_{\hat{\theta}}$  no implica parámetros desconocidos.

Por ejemplo, si  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$ ,  $\sigma_{\hat{\theta}} = \sigma_{\bar{X}} = \sigma/\sqrt{n}$ , la cual no implica parámetros sólo si se conoce el valor de  $\sigma$ .

Al estandarizar  $\hat{\theta}$  conforme a la suposición de que  $H_0$  es verdadera (de modo que  $E(\hat{\theta}) = \theta_0$  se obtiene un estadístico de prueba con muestra grande):

$$\text{Estadístico de prueba: } Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

Si la hipótesis alternativa es  $H_a : \theta > \theta_0$ , la región de rechazo  $z \geq z_\alpha$  especifica una prueba de cola superior cuyo nivel de significancia es aproximadamente  $\alpha$ .

Las otras dos alternativas,  $H_a : \theta < \theta_0$  y  $H_a : \theta \neq \theta_0$ , se someten a prueba  $z$  de cola inferior y a una prueba  $z$  de dos colas, respectivamente.

En el caso  $\theta = p$ ,  $\sigma_{\hat{\theta}}$  no implicará parámetros desconocidos cuando  $H_0$  es verdadera, aunque esto es atípico.

Cuando  $\sigma_{\hat{\theta}}$  no implica parámetros desconocidos, a menudo es posible utilizar una desviación estándar estimada  $\hat{S}_{\hat{\theta}}$  en lugar de  $\sigma_{\hat{\theta}}$  y seguir teniendo  $Z$  aproximadamente normalmente distribuida cuando  $H_0$  es verdadera (porque  $n$  es grande cuando  $\hat{s}_{\hat{\theta}}$  con la mayoría de las muestras).

La prueba con muestra grande de la sección previa da un ejemplo de esto porque  $\sigma$  casi siempre es desconocida, se utiliza  $\hat{s}_{\hat{\theta}} = \hat{s}_{\bar{X}} = \hat{s}/\sqrt{n}$  en lugar de  $\sigma/\sqrt{n}$  en el denominador de  $z$ .

El estimador  $\hat{p} = X/n$  es ( $E(\hat{p}) = p$ ), insesgado y su distribución es aproximadamente normal y su desviación estándar es

$$\sigma_{\hat{p}} = \sqrt{p \cdot (1 - p)/n}.$$

Estos hechos se utilizaron en el tema 7 para obtener un intervalo de confianza para  $p$ . Cuando  $H_0$  es verdadera  $E(\hat{p}) = p_0$  y

$$\sigma_{\hat{p}} = \sqrt{p_0 \cdot (1 - p_0)/n}, \text{ así que } \sigma_{\hat{p}} \text{ no implica parámetros desconocidos.}$$

Se desprende entonces que cuando  $n$  es grande y  $H_0$  es verdadera, el estadístico de prueba

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 \cdot (1 - p_0)/n}}$$

tiene aproximadamente una distribución normal estándar.

Si la hipótesis alternativa es  $H_a : p > p_0$  y se utiliza la región de rechazo de cola superior  $z \geq z_\alpha$ , entonces

$$\begin{aligned} P(\text{error de tipo I}) &= P(H_0 \text{ es rechazada cuando es verdadera}) = \\ &= P(Z \geq z_\alpha \text{ cuando } Z \text{ tiene aproximadamente una dist. nor. est.}) \approx \alpha \end{aligned}$$

Por consiguiente el nivel de significación deseado  $\alpha$  se obtiene utilizando el valor crítico que capture área crítica  $\alpha$  en la cola superior de la curva  $z$ .

Las regiones de rechazo para las otras dos hipótesis alternativas, cola inferior para  $H_a : p < p_0$  y dos colas para  $H_a : p \neq p_0$  se justifica de manera análoga.

### Pruebas relacionadas con una proporción de población

Hipótesis nula:  $H_0 : p = p_0$

Valor del estadístico de prueba:  $z = \frac{\hat{p} - p_0}{\sqrt{p_0 \cdot (1 - p_0)/n}}$

Hipótesis alternativa      Región de rechazo para la prueba de nivel  $\alpha$

$H_a : p > p_0$        $z \geq z_\alpha$  (prueba de cola superior)

$H_a : p < p_0$        $z \leq -z_\alpha$  (prueba de cola inferior)

$H_a : p \neq p_0$        $z \geq z_{\alpha/2}$  o  $z \leq -z_{\alpha/2}$  (prueba de dos colas)

Estos procedimientos de prueba son válidos siempre que  $n \cdot p_0 \geq 5$  y  $n \cdot (1 - p_0) \geq 5$ .

## Ejemplo

Información reciente sugiere que la obesidad es un problema creciente en Estados Unidos entre grupos de todas las edades.

La Prensa Asociada (9 de octubre de 2002) reportó que 1276 individuos en una muestra de 4115 adultos fueron encontrados obesos (un índice de masa corporal de más de 30; este índice mide el peso con respecto a la estatura).

Una encuesta realizada en 1998 basada en la autoevaluación de las personas reveló que el 20% de los estadounidenses adultos se autoconsideraron obesos.

¿Sugieren los datos más recientes que la proporción verdadera de adultos obesos es más de 1.5 veces el porcentaje de la encuesta de autoevaluación?

Realicemos una prueba de hipótesis utilizando un nivel de significación de 0.10.

## Ejemplo (Continuación)

- 1.-  $p$  = la proporción de todos los adultos estadounidenses obesos.
- 2.- Decir que el porcentaje actual es 1.5 veces el porcentaje de autoevaluación equivale a aseverar que el porcentaje actual es de 30%, de donde se deduce la hipótesis nula como  $H_0 : p = 0.30$ .
- 3.- La frase "más que" en la descripción del problema implica que la hipótesis alternativa es  $H_a : p > 0.30$ .
- 4.- Como  $n \cdot p_0 = 4115(0.3) > 5$  y  $n \cdot (1 - p_0) = 4115(0.7) > 5$ , ciertamente la prueba con muestra grande puede ser utilizada. El valor estadístico de prueba es 
$$z = \frac{\hat{p} - 0.3}{\sqrt{0.3 \cdot (1 - 0.3)/n}}$$
- 5.- La forma de  $H_a$  implica que una prueba de cola superior es apropiada: rechazar  $H_0$  si  $z > z_{0.10} = 1.28$ .
- 6.-  $\hat{p} = 1276/4115 = 0.310$ , de donde se obtiene 
$$z = (0.310 - 0.3)/\sqrt{0.3 \cdot 0.7/4115} = 0.010/0.0071 = 1.40.$$

### Ejemplo (Continuación)

- 7.- Como 1.40 excede el valor crítico 1.28,  $z$  queda en la región de rechazo. Esto justifica el rechazo de la hipótesis nula. Con un nivel de significación de 0.10 parece que más del 30% de los estadounidenses son obesos.



# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

Una forma de reportar el resultado de un análisis de hipótesis es decir simplemente si la hipótesis nula fue rechazada a nivel de significación específico.

Por consiguiente un investigador podría afirmar que si  $H_0$  fue rechazada a nivel de significación de 0.05 o que el uso de una prueba de nivel 0.01 dio por resultado no rechazar  $H_0$ .

Este tipo de afirmación es un tanto inadecuada porque no dice nada sobre si la conclusión es un tanto dudosa o bastante precisa.

Una dificultad relacionada es que semejante reporte traslada el nivel de significación a otros tomadores de decisiones.

En muchas situaciones en las que hay que tomar una decisión, los individuos pueden tener diferentes puntos de vista en cuanto a las consecuencias de un error de tipo I o de tipo II.

Cada individuo desearía entonces seleccionar su propio nivel de significación –algunos seleccionarían  $\alpha = 0.05$ , otros 0.01 y así sucesivamente– y llegar a una conclusión por lo tanto.

Esto conduciría a que algunos individuos rechacen  $H_0$  en tanto que otros concluirían que los datos no muestran una contradicción suficientemente fuerte en contra de  $H_0$  para justificar su rechazo.

## Ejemplo

Se sabe que el tiempo promedio verdadero para el alivio inicial del dolor de un analgésico éxito en ventas es de 10 min.

Sea  $\mu$  el tiempo promedio verdadero para el alivio del dolor de un analgésico recién desarrollado por una compañía.

Ésta desea producir y comercializar este analgésico sólo si proporciona un alivio más rápido que el que más se vende, así que desea probar

$H_0 : \mu = 10$  contra  $H_a : < 10$ .

Sólo si la evidencia experimental conduce al rechazo de  $H_0$  se introducirá el nuevo analgésico.

Tras de ponderar la seriedad relativa de cada tipo de error, se debe acordar un solo nivel de significación y tomar la decisión de –rechazar  $H_0$  e introducir o no el analgésico– a ese nivel.

## Ejemplo (Continuación)

Supongamos que el nuevo analgésico ha sido introducido.

La compañía apoya su afirmación de alivio más rápido declarando que, con base en un análisis de datos experimentales,  $H_0 : \mu = 10$  fue rechazada a favor de  $H_a : \mu < 10$  utilizando un nivel de significación  $\alpha = 0.10$ .

Cualquier individuo que considere cambiar a este nuevo analgésico naturalmente desearía sacar sus propias conclusiones en cuanto a la validez de la afirmación.

Los individuos que están satisfechos con el analgésico que más se vende considerarían un error de tipo I (al concluir que el nuevo producto proporciona un alivio más rápido cuando en realidad no lo hace) tan serio que desearían utilizar  $\alpha = 0.05, 0.01$ , o incluso niveles más pequeños.

Desafortunadamente, la naturaleza de la afirmación de la compañía evita que un individuo llegue a una conclusión a semejante nivel.

### Ejemplo (Continuación)

La compañía ha impuesto su propia selección de nivel de significación en otros.

El reporte podría haber sido elaborado en una manera que permitiera a cada individuo flexibilidad al sacar una conclusión a un nivel  $\alpha$  personalmente seleccionado.

Un **valor  $P$**  transmite mucha información sobre la fuerza de la evidencia en contra de  $H_0$  y permite que un individuo saque una conclusión a cualquier nivel específico  $\alpha$ .

Antes de dar una definición general, consideremos cómo la conclusión en un problema de prueba de hipótesis depende del nivel seleccionado  $\alpha$ .

## Ejemplo

El problema de contenido de nicotina discutido en el ejemplo anterior implicó probar  $H_0 : \mu = 1.5$  contra  $H_a : \mu > 1.5$ .

Debido a la desigualdad presente en  $H_a$ , la región de rechazo es de cola superior, con  $H_0$  rechazada si  $z \geq z_\alpha$ .

Supongamos que  $z = 2.10$ .

La tabla adjunta muestra la región de rechazo con cada uno de cuatro niveles  $\alpha$  diferentes junto con la conclusión resultante.

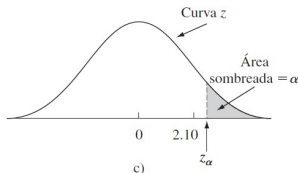
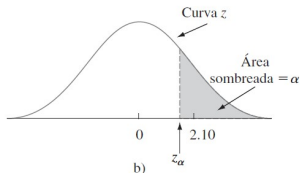
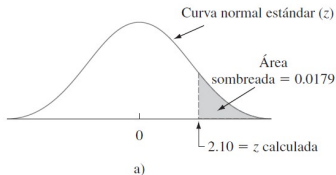
| Nivel de Significación $\alpha$ | Región de rechazo | Conclusión        |
|---------------------------------|-------------------|-------------------|
| 0.05                            | $z \geq 1.645$    | Rechazar $H_0$    |
| 0.025                           | $z \geq 1.96$     | Rechazar $H_0$    |
| 0.01                            | $z \geq 2.33$     | No rechazar $H_0$ |
| 0.005                           | $z \geq 2.58$     | No rechazar $H_0$ |



## Ejemplo (Continuación)

Relación entre  $\alpha$  y área de cola capturada  $z$ :

- área de cola capturada por  $z$  calculada;
- cuando  $\alpha > 0.0179$ ,  $z_\alpha < 2.10$  y  $H_0$  es rechazada;
- cuando  $\alpha < 0.0179$ ,  $z_\alpha > 2.10$  y  $H_0$  no es rechazada.



## Ejemplo (Continuación)

Con un nivel  $\alpha$  relativamente grande, el valor crítico  $z_\alpha$  no está muy alejado en la cola superior; 2.10 excede el valor crítico y por lo tanto  $H_0$  es rechazada.

Sin embargo, a medida que  $\alpha$  disminuye, el valor crítico se incrementa.

Con  $\alpha$  pequeño, el valor crítico  $z$  es grande, 2.10 es menor que  $z_\alpha$  y  $H_0$  no es rechazada.

Recordemos que con una prueba  $z$  de cola superior,  $\alpha$  es exactamente el área debajo de la curva  $z$  a la derecha del valor crítico  $z_\alpha$ .

Es decir, una vez que se especifica  $\alpha$ , se selecciona el valor crítico para capturar área de cola superior  $\alpha$ .

La tabla normal estándar muestra que el área a la derecha de 2.10 es 0.0179.

## Ejemplo (Continuación)

Con un nivel  $\alpha$  más grande que 0.0179 corresponde a  $z_\alpha < 2.10$ . Una  $\alpha$  menor que 0.0179 requiere del uso de un valor crítico de  $z$  que exceda 2.10.

La decisión sobre un nivel particular de  $\alpha$  depende de cómo el nivel  $\alpha$  seleccionado se compara con el área de cola capturada por el valor  $z$  calculado.

Esto se ilustra en la figura anterior.

Observemos en particular que 0.0179, el área de cola capturada, es el nivel más pequeño  $\alpha$  al cual  $H_0$  sería rechazada, porque con cualquier nivel  $\alpha$  más pequeño se obtiene un valor crítico  $z$  que excede de 2.10, de modo que 2.10 no se encuentra en la región de rechazo.

En general, supongamos que se ha determinado la distribución de probabilidad de un estadístico de prueba cuando  $H_0$  es verdadera.

Entonces, con un nivel  $\alpha$  especificado, la región de rechazo se determina encontrando un valor o valores críticos que capturen área de cola  $\alpha$  (de cola superior, inferior o de dos colas, cualquiera que sea apropiada) bajo la curva de distribución de probabilidad.

El valor  $\alpha$  más pequeño con el cual  $H_0$  sería rechazada es el área de cola capturada por el valor calculado del estadístico de prueba.

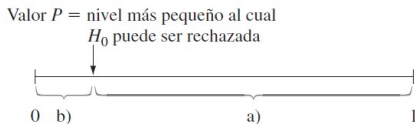
Este valor  $\alpha$  más pequeño es el valor  $P$ .

## Definición

El **valor  $P$**  (o **nivel de significación observado**) es el nivel de significación más pequeño al cual  $H_0$  sería rechazada cuando se utiliza un procedimiento de prueba especificado con un conjunto de datos dado. Una vez que se ha determinado el valor  $P$ , la conclusión a un nivel particular  $\alpha$  resulta de comparar el valor  $P$  con  $\alpha$ :

- 1 Valor  $P \leq \alpha \implies$  rechazar  $H_0$  al nivel  $\alpha$ .
- 2 Valor  $P > \alpha \implies$  no rechazar  $H_0$  al nivel  $\alpha$ .

Se acostumbra a llamar **significativos** a los datos cuando  $H_0$  es rechazada y **no significativos** de lo contrario. El valor  $P$  es entonces el nivel más pequeño al cual los datos son significativos. Una manera fácil de visualizar la comparación del valor  $P$  con el nivel  $\alpha$  seleccionado es trazar una imagen como muestra la figura siguiente. El cálculo del valor  $P$  depende de si la prueba es de cola superior, inferior o de dos colas. No obstante, una vez calculada, la comparación con  $\alpha$  no depende de qué tipo de prueba se utilizó.



Comparación de  $\alpha$  y el valor  $P$ : a) rechazar  $H_0$  cuando  $\alpha$  queda aquí; b) no rechazar  $H_0$  cuando  $\alpha$  queda aquí.

## Ejemplo

Supongamos que cuando se analizaron los datos de un experimento que implican el nuevo analgésico, el valor  $P$  para probar  $H_0 : \mu = 10$  contra  $H_a : \mu < 10$  se calculó como 0.0384.

Como  $\alpha = 0.05$  es más grande que el valor  $P$  (0.05 queda en el intervalo (a) de la figura anterior)

$H_0$  sería rechazada por cualquiera que realice la prueba al nivel 0.05.

Sin embargo, al nivel 0.01,  $H_0$  no sería rechazada porque 0.01 es más pequeño que el nivel más pequeño (0.0384) al cual  $H_0$  puede ser rechazada.

Los programas de computadora más utilizados incluyen automáticamente un valor  $P$  cuando se analiza una hipótesis.

Se puede sacar una conclusión directamente de los datos de salida, sin referencia a una tabla de valores críticos.

Una definición alternativa útil equivalente a la que se acaba de dar es como sigue:

### Definición

El **valor  $P$**  es la probabilidad calculada suponiendo que  $H_0$  es verdadera, de obtener un valor estadístico de prueba por lo menos tan contradictorio a  $H_0$  como el valor que en realidad se obtuvo.

Mientras más pequeño es el valor de  $P$ , más contradictorios son los datos a  $H_0$ .



Por consiguiente si  $z = 2.10$  para una prueba  $z$  de cola superior, el valor  $P = P(Z \geq 2.10 \text{ cuando } H_0 \text{ es verdadera}) = 1 - P(Z \leq 2.10)$ , o sea que el valor  $P = 0.0179$  queda como antes.

Mucho ojo: ¡el valor  $P$  no es la probabilidad de que  $H_0$  sea verdadera, ni es una probabilidad de error!

## Valores $P$ para pruebas $z$

El valor  $P$  para una prueba  $z$  (una basada en un estadístico de prueba cuya distribución, cuando  $H_0$  es verdadera, es por lo menos aproximadamente estándar normal) es fácil de determinar a partir de la información de la tabla normal estándar.

Consideremos una prueba de cola superior y sea  $z$  el valor calculado del estadístico de prueba  $Z$ .

La hipótesis nula es rechazada si  $z \geq z_\alpha$  y el valor  $P$  es el  $\alpha$  más pequeño para el cual este es el caso.

Como  $z_\alpha$  se incrementa a medida que  $\alpha$  disminuye, el valor  $P$  es el valor de  $\alpha$  con el cual  $z = z_\alpha$ .

Es decir, el valor  $P$  es exactamente el área capturada por el valor calculado  $z$  en la cola superior de la curva normal estándar.

El área acumulativa correspondiente es  $P(Z \leq z)$ , así que en este caso valor  $P = 1 - P(Z \leq z)$ .

Un argumento análogo para una prueba de cola inferior demuestra que el valor  $P$  es el área capturada por el valor calculado  $z$  en la cola inferior de la curva normal estándar.

Se debe tener más cuidado en el caso de una prueba de dos colas.

Supongamos primero que  $z$  es positivo.

Entonces el valor  $P$  es el valor de  $\alpha$  que satisface  $z = z_{\alpha/2}$  (es decir,  $z$  calculado = valor crítico de cola superior).

Esto dice que el área capturada en la cola superior es la mitad del valor  $P$ , de modo que valor  $P = 2[1 - P(Z \leq z)]$ .

Si  $z$  es negativo, el valor  $P$  es el  $\alpha$  con el cual  $z = -z_{\alpha/2}$ , o, de forma equivalente,  $-z = z_{\alpha/2}$ , así que valor  $P = 2[1 - P(Z \leq -z)]$ .

Como  $-z = |z|$  cuando  $z$  es negativo, valor  $P = 2[1 - P(Z \leq |z|)]$  con  $z$  positivo o negativo.

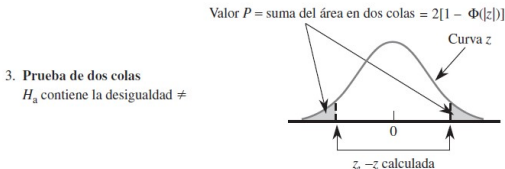
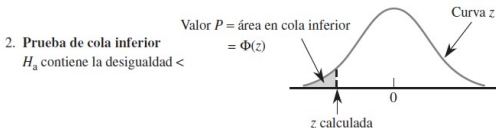
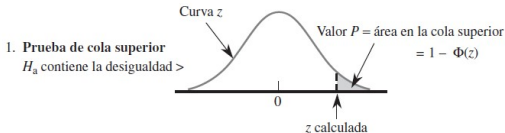
Valor  $P$ 

$$\text{valor } P = \begin{cases} 1 - P(Z \leq z) & \text{para una prueba de cola superior} \\ P(Z \leq z) & \text{para una prueba de cola inferior} \\ 2[1 - P(Z \leq z)] & \text{para una prueba de dos colas} \end{cases}$$

Cada una de éstas es la probabilidad de obtener un valor por lo menos tan extremo como el que se obtuvo (suponiendo que  $H_0$  es verdadera).

Los tres casos se ilustran en la siguiente figura.

El siguiente ejemplo ilustra el uso de la aproximación del valor  $P$  a la prueba de hipótesis por medio de una secuencia de pasos modificados con respecto a la secuencia previamente recomendada.



Determinación del valor  $P$  para una prueba  $z$ .

## Ejemplo

El espesor objetivo de obleas de silicio utilizadas en cierto tipo de circuito integrado es de  $245 \mu\text{m}$ .

De una muestra de 50 obleas, cada una con un espesor determinado, se obtiene una media de espesor de  $246.18 \mu\text{m}$  y una desviación estándar de  $3.60 \mu\text{m}$ .

¿Sugieren estos datos que el espesor de oblea promedio verdadero es algún otro diferente del valor objetivo?

1.- Parámetro de interés  $\mu$  = espesor de oblea promedio verdadero.

2.- Hipótesis nula:  $H_0 : \mu = 245$ .

3.- Hipótesis alternativa:  $H_a : \mu \neq 245$ .

4.- Fórmula para el valor del estadístico de prueba:  $z = \frac{\bar{x} - 245}{\hat{s}\sqrt{n}}$

5.- Cálculo del valor del estadístico de prueba:  $z = \frac{246.18 - 245}{3.60\sqrt{50}} = 2.32$

### Ejemplo (Continuación)

- 6.- Determinación del valor  $P$ : Como la prueba es de dos colas,  
Valor  $P = 2(1 - P(Z \leq 2.32)) = 0.0204$
- 7.- Conclusión: Con un nivel significativo de 0.01,  $H_0$  no sería rechazada puesto que  $0.0204 > 0.01$ . A este nivel de significación, existe suficiente evidencia para concluir que el espesor promedio verdadero difiere del valor objetivo.

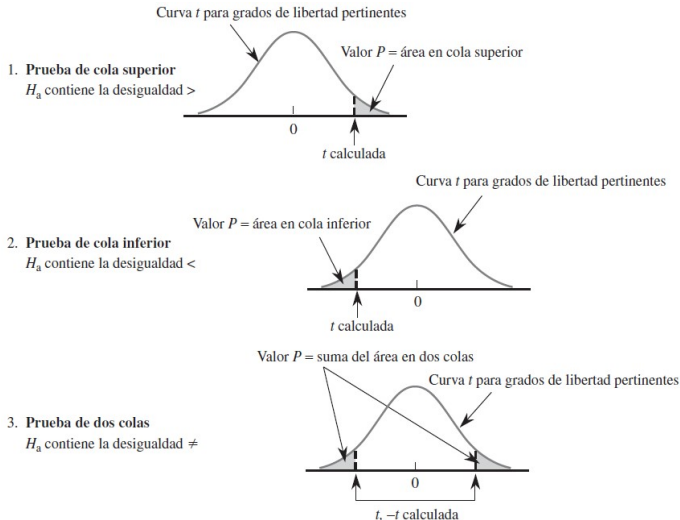
## Valores $P$ para pruebas $t$

Así como el valor  $P$  para una prueba  $z$  es un área de curva  $z$ , el valor  $P$  para una prueba  $t$  será un área de curva  $t$ .

La figura siguiente ilustra los tres casos diferentes.

El número de grados de libertad para la prueba  $t$  con una muestra es  $n - 1$ .





Determinación del valor  $P$  para una prueba  $t$ .

La tabla de valores críticos  $t$  previamente utilizada para intervalos de confianza y predicción no contienen suficiente información sobre cualquier distribución  $t$  particular que permita la determinación precisa de áreas deseadas.

Los valores  $P$  obtenidos con programas de computadora serán más precisos que los obtenidos mediante la tabla  $t$ .

# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

Una vez que el experimentador ha decidido sobre la cuestión de interés y el método de obtención de datos (el diseño del experimento), la construcción de una prueba apropiada se compone de tres pasos distintos:

- 1 Especificar un estadístico de prueba (la función de los valores observados que servirá para tomar una decisión).
- 2 Decidir sobre la forma general de la región de rechazo (típicamente rechazar  $H_0$  con valores apropiadamente grandes del estadístico de prueba, rechazar con valores apropiadamente pequeños o rechazar con valores pequeños o grandes).
- 3 Seleccionar el valor o valores críticos numéricos específicos que separarán la región de rechazo de la región de aceptación (obteniendo la distribución del estadístico de prueba cuando  $H_0$  es verdadera y luego seleccionar un nivel de significación).

En los ejemplos presentados hasta ahora, se realizaron los pasos 1 y 2 en una manera ad hoc mediante intuición.

Por ejemplo, cuando la población subyacente se supuso normal con media  $\mu$  y  $\sigma$  conocida, se procedió desde  $\bar{X}$  hasta el estadístico de prueba estandarizada

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Para probar  $H_0 : \mu = \mu_0$  contra  $H_a : \mu > \mu_0$ , la intuición sugirió entonces rechazar  $H_0$  cuando  $z$  era grande.

Por último, se determinó el valor crítico especificando el nivel de significación  $\alpha$  y utilizando el hecho de que  $Z$  tiene una distribución normal estándar cuando  $H_0$  es verdadera.

La confiabilidad de la prueba para tomar la decisión correcta puede ser evaluada estudiando probabilidades de error de tipo II.

Los temas que tienen que ser considerados al realizar los pasos 1-3 comprenden las preguntas:

- 1 ¿Cuáles son las implicaciones y consecuencias prácticas de seleccionar un nivel de significación particular una vez que se han determinado los demás aspectos de una prueba?
- 2 ¿Existe un principio general, que no dependa de la intuición, que pueda ser utilizado para obtener buenos o mejores procedimientos de prueba?
- 3 Cuando dos o más pruebas son apropiadas en una situación dada, ¿cómo se comparan las pruebas para decidir cuál deberá ser utilizada?
- 4 Si una prueba se deriva con arreglo a suposiciones específicas sobre la distribución o población muestreada, ¿cómo funcionará la prueba cuando se violan las suposiciones?

## Significación estadística contra práctica

Aunque el proceso de llegar a una decisión utilizando la metodología de probar hipótesis clásicas implica seleccionar un nivel de significación y luego rechazar o no rechazar  $H_0$ , a ese nivel  $\alpha$ , reportando simplemente el nivel  $\alpha$  utilizado y la decisión alcanzada da un poco de la información contenida en los datos muestrales.

En especial, cuando los resultados de un experimento han de ser comunicados a una gran audiencia, el rechazo de  $H_0$  a nivel de 0.05 será mucho más convincente si el valor observado del estadístico de prueba excede en gran medida el valor crítico de 5% que si apenas excede ese valor.

Esto es precisamente lo que condujo a la noción de valor  $P$  como una forma de reportar significación sin imponer una  $\alpha$  particular a otros que pudieran desear sacar sus propias conclusiones.

Incluso si se incluye un valor  $P$  en un resumen de resultados, sin embargo, puede haber dificultad al interpretar este valor y al tomar una decisión.

Esto es porque un valor  $P$  pequeño, el que ordinariamente indicaría significación estadística es que sugeriría fuertemente el rechazo de  $H_0$  a favor de  $H_a$ , puede ser el resultado de un tamaño de muestra grande en combinación con un alejamiento de  $H_0$  que tiene poca significación práctica.

En muchas situaciones experimentales, sólo valdría la pena detectar los alejamientos de  $H_0$  de gran magnitud, en tanto que un alejamiento pequeño de  $H_0$  tendría poca significación práctica.



Consideremos como ejemplo probar  $H_0 : \mu = 100$  contra  $H_a : \mu > 100$  donde  $\mu$  es la media de una población normal con  $\sigma = 10$ .

Supongamos que un valor verdadero de  $\mu = 101$  no representaría un alejamiento serio de  $H_0$  en el sentido de no rechazar  $H_0$  cuando  $\mu = 101$  sería un error relativamente costoso.

Con tamaño de muestra razonablemente grande  $n$ , esta conduciría a un valor  $\bar{x}$  próximo a 101, así que no se desearía esta evidencia muestral para argumentar fuertemente a favor del rechazo de  $H_0$  cuando  $\bar{x} = 101$  es observado.

Para varios tamaños muestrales, la tabla siguiente registra tanto el valor  $P$  cuando  $\bar{x} = 101$  y también la probabilidad de no rechazar  $H_0$  al nivel 0.01 cuando  $\mu = 101$ .

Efecto del tamaño de muestra en los valores  $P$  y  $\beta$

| $n$   | Valor $P$ cuando $\bar{x} = 101$ | $\beta(101)$ prueba de nivel 0.01 |
|-------|----------------------------------|-----------------------------------|
| 25    | 0.3085                           | 0.9664                            |
| 100   | 0.1587                           | 0.9082                            |
| 400   | 0.0228                           | 0.6293                            |
| 900   | 0.0013                           | 0.2514                            |
| 1600  | 0.0000335                        | 0.0475                            |
| 2500  | 0.000000297                      | 0.0038                            |
| 10000 | $7.69 \cdot 10^{-24}$            | 0.0000                            |

La segunda columna en la tabla muestra que incluso con tamaños de muestra moderadamente grandes, el valor  $P$  de  $\bar{x} = 101$  argumenta fuertemente a favor del rechazo de  $H_0$ , en tanto que el valor  $\bar{x}$  observado sugiere que en términos prácticos el valor verdadero de  $\mu$  difiere poco del valor nulo  $\mu_0 = 100$ .

La tercera columna señala que incluso cuando existe poca diferencia práctica entre la  $\mu$  verdadera y el valor nulo, con un nivel de significación fijo un tamaño de muestra grande casi siempre conduce al rechazo de la hipótesis nula a ese nivel.

Resumiendo, se debe tener un especial cuidado al interpretar evidencia cuando el tamaño de muestra es grande, puesto que cualquier alejamiento pequeño de  $H_0$  con toda seguridad será detectado por una prueba, aunque semejante alejamiento puede tener poca significación práctica.

# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema

## EJERCICIO

Una compañía tabacalera afirma que el contenido de nicotina prometido de los cigarrillos marca B es (como mucho) de 1.5 mg.

Consideramos  $\mu$  como el contenido de nicotina promedio verdadero de los cigarrillos marca B.

El objetivo es probar  $H_0 : \mu = 1.5$  contra  $H_a : \mu > 1.5$  con base en una muestra aleatoria  $X_1, X_2, \dots, X_{32}$  de contenido de nicotina.

Supongamos que se sabe que la distribución del contenido de nicotina es normal con  $\sigma = 0.20$ .

## Solución

Entonces  $\bar{X}$  está normalmente distribuida con valor medio  $\mu_{\bar{X}} = \mu$  y desviación estándar  $\sigma_{\bar{X}} = \sigma/\sqrt{32} = 0.20/\sqrt{32} = 0.0354$

En lugar de utilizar  $\bar{X}$  como estadístico de prueba, se estandariza  $\bar{X}$  suponiendo que  $H_0$  es verdadera.

## EJERCICIO

### Solución (Continuación)

Estadístico de prueba: 
$$Z = \frac{\bar{X} - 1.5}{\sigma/\sqrt{n}} = \frac{\bar{X} - 1.5}{0.0354}$$

$Z$  expresa la distancia entre  $\bar{X}$  y su valor esperado cuando  $H_0$  es verdadera como algún número de desviaciones estándar.

Por ejemplo,  $z = 3$  resulta de una  $\bar{x}$  que es 3 desviaciones estándar más grande de lo que se habría esperado si  $H_0$  fuera verdadera.

Rechazar  $H_0$  cuando  $\bar{x}$  excede "considerablemente" de 1.5 equivale a rechazar  $H_0$  cuando  $z$  excede "considerablemente" de cero.

Es decir, la forma de la región de rechazo es  $z \geq c$ .

Determinemos ahora  $c$  de modo que  $\alpha = 0.05$ .

Cuando  $H_0$  es verdadera,  $Z$  tiene una distribución estándar normal.

## EJERCICIO

### Solución (Continuación)

Por consiguiente

$$\alpha = P(\text{error de tipo I}) = P(\text{rechazar } H_0 \text{ cuando } H_0 \text{ es verdadera})$$

$$\alpha = P(Z \geq c \text{ cuando } Z \sim N(0, 1))$$

El valor de  $c$  debe capturar el área de la cola superior 0.05 bajo la curva  $z$ .

Encontramos en la tabla normal estándar que  $c = z_{0.05} = 1.645$ .

Observemos que  $z \geq 1.645$  equivale a  $\bar{x} - 1.5 \geq (0.0354)(1.645)$ , es decir,  $\bar{x} \geq 1.56$ .

Entonces  $\beta$  es la probabilidad de que  $\bar{X} < 1.56$  y puede ser calculada con cualquier  $\mu$  mayor que 1.5.

# Esquema

## 1 CONTRASTE DE HIPÓTESIS

- Introducción
- Hipótesis y procedimientos de prueba
  - Procedimientos de prueba
  - Errores en la prueba de hipótesis
- Pruebas sobre una media de población
  - Caso I: Una población normal con  $\sigma$  conocida
  - Caso II: Pruebas con muestras grandes
  - Caso III: Una distribución de población normal
- Pruebas relacionadas con una proporción de población
  - Pruebas con muestra grande
- Valores  $P$ 
  - Valores  $P$  para pruebas  $z$
  - Valores  $P$  para pruebas  $t$
- Algunos comentarios sobre la selección de una prueba
  - Significación estadística contra práctica
- Ejercicio
- Problema



## Problema

Se dan pares de valores  $P$  y niveles de significación,  $\alpha$ . Para cada par, di si el valor  $P$  observado conduciría al rechazo de  $H_0$  en el nivel de significación dado.

- a Valor  $P = 0.084$ ,  $\alpha = 0.05$
- b Valor  $P = 0.003$ ,  $\alpha = 0.001$
- c Valor  $P = 0.498$ ,  $\alpha = 0.05$
- d Valor  $P = 0.084$ ,  $\alpha = 0.10$
- e Valor  $P = 0.039$ ,  $\alpha = 0.01$
- f Valor  $P = 0.218$ ,  $\alpha = 0.10$