

Curso de Estadística Aplicada a las Ciencias Sociales

Tema 12. Contraste de hipótesis

(Cap. 22 del libro)

Tema 12. Contraste de hipótesis

Introducción

1. Tipos de hipótesis
2. La hipótesis nula y la alternativa
3. Metodología del contraste de hipótesis
4. Contraste para una proporción
5. Contraste para una media

Resumen

Ejercicios

Introducción

- En muchas investigaciones no queremos, sin más, saber “cómo son” los datos
- Tenemos sospechas sobre cómo son las cosas y queremos ver si los datos son compatibles con esas sospechas
- Esas sospechas pueden provenir de una “seria teoría” o de información parcial, anecdótica, o dispersa, que queremos verificar
- Esas sospechas se materializan en hipótesis: afirmaciones sobre características de una población

Introducción

- De las hipótesis se derivan predicciones fácticas que podemos contrastar con la realidad de los datos
- Ejemplos de hipótesis:
 - ◆ El tabaco produce cáncer de pulmón
 - ◆ Disminuir los impuestos reduce el fraude fiscal
 - ◆ Las mujeres son más apasionadas que los hombres

Introducción

- Lógicamente, las hipótesis no se refieren a cada individuo, sino al conjunto de las poblaciones
 - ◆ En conjunto, los fumadores tienen una proporción mayor de casos de cáncer de pulmón que los no fumadores. No todos los fumadores tendrán cáncer de pulmón, y algunos no fumadores lo tendrán también
 - ◆ Fraude fiscal: la misma población antes y después de la bajada de impuestos. No todos defraudaban antes, y ninguno después. Sino que la cantidad defraudada disminuye, en conjunto

Introducción

- ◆ Apasionamiento: las poblaciones son los hombres y las mujeres; y se afirma de nuevo que como media, en una variable (cuya definición y medida habría que aclarar), como media, las mujeres tendrán un valor más alto que los hombres
- Contraste de hipótesis requiere, por tanto:
 - ◆ Especificar población o poblaciones
 - ◆ Definir la variable y cómo la medimos
 - ◆ Explicitar qué predice la hipótesis sobre los valores de la variable

Introducción

- Estamos suponiendo que no tenemos todos los datos de toda la población, porque entonces podríamos directamente comprobar la hipótesis
- Estamos haciendo estadística inferencial: tenemos sólo muestras, y a partir de ellas vamos a hacer afirmaciones sobre la población

Tema 12. Contraste de hipótesis

Introducción

- 1. Tipos de hipótesis**
2. La hipótesis nula y la alternativa
3. Metodología del contraste de hipótesis
4. Contraste para una proporción
5. Contraste para una media

Resumen

Ejercicios

1. Tipos de hipótesis

- Las hipótesis más habituales pueden clasificarse en dos tipos:
- Primer tipo de hipótesis. Las que predicen un valor concreto o un intervalo para el parámetro
 - ◆ Una mayoría de los españoles está a favor de ... o está en contra de ...
 - ◆ La nota en selectividad de los colegios públicos es peor que la media, que es de 5,5
 - ◆ Si los españoles votaran hoy, el porcentaje de votos para el PSOE sería mayor que en las últimas elecciones
 - ◆ Los jóvenes tienen mayor riesgo de morir en accidente de tráfico que el conjunto de la población

1. Tipos de hipótesis

- ◆ Ese valor o intervalo puede tener varios orígenes:
 - ★ Un punto de corte “importante” como el 0,5 que distingue la mayoría y la minoría
 - ★ Una información previa sobre el valor en la población, que queremos comparar con el valor en un subgrupo
 - ★ Una información sobre el valor en el pasado, que queremos comparar con el valor actual
 - ★ Una información sobre la proporción que representa el grupo en la población

1. Tipos de hipótesis

- Segundo tipo de hipótesis. Las que predicen que el valor de un parámetro en dos poblaciones es distinto (o igual):
 - ◆ Los chicos y las chicas tienen su primera relación sexual a una edad diferente
 - ◆ La proporción de parejas que conviven antes del matrimonio es mayor en las ciudades que en los pueblos
 - ◆ Los hijos de padres separados o divorciados tienen resultados escolares similares a los de parejas que viven juntas

1. Tipos de hipótesis

- En este caso no tenemos un valor de referencia
- Lo que queremos saber es si el valor es diferente o igual entre dos grupos de población
- Los dos tipos de hipótesis dan lugar a estudios diferentes:
 - ◆ Hipótesis del primer tipo: hacemos un **contraste de hipótesis** (este tema)
 - ◆ Hipótesis del segundo tipo: hacemos una **comparación de poblaciones** (tema 13)

Tema 12. Contraste de hipótesis

Introducción

1. Tipos de hipótesis
- 2. La hipótesis nula y la alternativa**
3. Metodología del contraste de hipótesis
4. Contraste para una proporción
5. Contraste para una media

Resumen

Ejercicios

2. La hipótesis nula y la alternativa

- Al hacer un contraste de hipótesis estamos poniendo a prueba una predicción sobre los valores de una variable, y en consecuencia también, la predicción inversa
- O dicho de otra forma, un contraste de hipótesis tiene en cuenta dos posibles hipótesis, opuestas, sobre el valor de un parámetro
- Por ejemplo, respecto a los centros escolares y la selectividad:
 - ◆ La hipótesis de que las notas de los centros públicos no difieren de la media
 - ◆ La hipótesis de que son menores que la media

2. La hipótesis nula y la alternativa

- Llamamos **hipótesis nula**, y la representamos por H_0 a la hipótesis que se pone a prueba directamente con el contraste
- Llamamos **hipótesis alternativa**, representada por H_1 , a la hipótesis contraria
- La hipótesis alternativa dice que el valor del parámetro es distinto del que dice la hipótesis nula

2. La hipótesis nula y la alternativa

- Si la información que resulta del contraste es contraria a la hipótesis nula, diremos que la hipótesis ha sido rechazada, y aceptamos la hipótesis alternativa
- Si la información que resulta del contraste es compatible con la hipótesis nula, no diremos que ha sido confirmada, o que la aceptamos, sino que no ha podido ser rechazada

2. La hipótesis nula y la alternativa

- ¿Cómo decidimos cuál de las dos es la hipótesis nula?
- Aquella, precisamente, que sí se puede rechazar con un estudio basado en una muestra
- Suele ser la hipótesis más simple, la que presupone un modelo de la realidad más sencillo
- Curiosamente, a pesar de su nombre, en realidad la hipótesis alternativa suele ser la que realmente el investigador está buscando poner a prueba
- Pero se procede “al revés”: la que se pone a prueba es la hipótesis nula

2. La hipótesis nula y la alternativa

- Ejemplos de hipótesis nulas:
 - ◆ Los españoles favorables a la ley tal son la mitad
 - ◆ La nota en selectividad de los colegios públicos es igual que la media, que es de 5,5
 - ◆ El porcentaje de votos para el PSOE sería igual que en 2004
 - ◆ Los jóvenes tienen igual riesgo de morir en accidente de tráfico que el conjunto de la población

2. La hipótesis nula y la alternativa

- La hipótesis nula suele ser del tipo: el parámetro es igual a un valor de referencia

$$H_0: \theta = \theta_0$$

- Ese tipo de hipótesis
 - ◆ Representan un modelo de la realidad más “simple” que el opuesto
 - ◆ Pueden encontrarse informaciones que permiten rechazarlas
- No sucedería lo mismo con las hipótesis opuestas

2. La hipótesis nula y la alternativa

- La hipótesis alternativa puede ser simplemente la negación de la hipótesis nula

$$H_1: \theta \neq \theta_0$$

- Usamos esta hipótesis alternativa (que es en realidad la hipótesis de investigación) cuando esperamos una diferencia en el valor de la variable, pero no sabemos en qué dirección
- En ese caso hacemos un contraste de hipótesis bilateral (la hipótesis alternativa resulta confirmada siempre que el valor de la variable sea diferente del valor previsto)

2. La hipótesis nula y la alternativa

- Otras veces conocemos en qué dirección la hipótesis nula puede ser falsa. Es decir tenemos la hipótesis de que

$$H_1: \theta > \theta_0$$

O bien de que $H_1: \theta < \theta_0$

- Usamos esta hipótesis alternativa cuando “sabemos” que si el valor es distinto, sólo puede serlo en una dirección (sólo puede ser más alto, o más bajo) .

2. La hipótesis nula y la alternativa

- Ejemplo: un experimento para mejorar el rendimiento de unos estudiantes.
- En ese caso hacemos un contraste de hipótesis unilateral
- Salvo que estemos muy seguros, es preferible hacerlo bilateral: las medidas sociales pueden tener a veces efectos opuestos a los esperados. Pueden mejorar o empeorar las cosas.

Tema 12. Contraste de hipótesis

Introducción

1. Tipos de hipótesis
2. La hipótesis nula y la alternativa
- 3. Metodología del contraste de hipótesis**
4. Contraste para una proporción
5. Contraste para una media

Resumen

Ejercicios

3. Metodología del contraste de hipótesis

- Podemos comparar un contraste de hipótesis con un juicio penal, en el que hay presunción de inocencia
- En el contraste de hipótesis hay una “presunción” de que la hipótesis nula H_0 es cierta
- Podemos decir que “salvo prueba en contrario”, aceptamos H_0
- Hay “prueba en contrario” cuando el estimador difiere tanto de θ_0 como para pensar que es imposible o muy poco probable que $\theta = \theta_0$

3. Metodología del contraste de hipótesis

Las etapas de un contraste de hipótesis entonces, son:

1. Definir la hipótesis nula a contrastar, H_0 y la hipótesis alternativa H_1
 - Lo más normal es que H_0 sea del tipo $\theta = \theta_0$
 - Las hipótesis alternativas pueden ser de dos tipos, como hemos visto

$H_1: \theta \neq \theta_0$, que da lugar a un contraste bilateral

$H_1: \theta > \theta_0$ o bien $H_1: \theta < \theta_0$, que da lugar a un contraste unilateral

3. Metodología del contraste de hipótesis

2. Establecer una medida de la distancia entre el dato muestral (el estimador) y el dato predicho por la hipótesis nula, θ_0 .
 - Vamos a calcular la diferencia o discrepancia entre el valor del estimador en la muestra y el valor supuesto del parámetro en la población
 - Vamos a calcular el valor estandarizado z del estimador, si la hipótesis nula fuera cierta
 - Recordamos de los temas anteriores, que las distribuciones muestrales (de una media o de una proporción) tienen como media la media o la proporción de la población

3. Metodología del contraste de hipótesis

- Si la hipótesis nula fuera cierta, la media de la distribución muestral sería θ_0
- Y la desviación típica de una distribución muestral es el ET, que podemos calcular a partir de la muestra (como en temas anteriores)
- Entonces, podemos medir la distancia entre el valor del estimador y el valor predicho por la hipótesis nula, utilizando el valor z del estimador calculado así:

$$z = \frac{\text{estimador} - \theta_0}{ET}$$

3. Metodología del contraste de hipótesis

3. Fijar el nivel de significación

- Tenemos que decidir qué valores de z son compatibles con la hipótesis nula, y cuáles son incompatibles
- Cuando tengamos el valor z, podremos calcular, la probabilidad de que, si la hipótesis nula H_0 fuera cierta, obtuviéramos una muestra con un estimador z como el que tenemos en la muestra, o aún más alejado
- A esa probabilidad la llamamos *Valor-P*

3. Metodología del contraste de hipótesis

- Cuanto más pequeño el *valor-P*, menos probable es que H_0 sea cierta
 - ◆ Si el *valor-P* es grande (por ejemplo 0,25; 0,7) quiere decir que si H_0 fuera cierta no sería raro que obtuviéramos un estimador como el que hemos encontrado. Por tanto, no podemos rechazar la hipótesis nula
 - ◆ En cambio si el *valor-P* es muy pequeño (0,001) quiere decir que si H_0 fuera cierta sería muy poco probable que obtuviéramos un estimador como el que tenemos. En ese caso rechazamos la hipótesis nula
- ¿Cuánto de pequeño ha de ser el *valor-P* para que decidamos rechazar H_0 ?

3. Metodología del contraste de hipótesis

- No hay un criterio universal: depende de nosotros fijar el punto de corte en el que consideramos que el *valor-P* es lo suficientemente pequeño como para descartar la hipótesis nula
- A este punto de corte se le llama **nivel de significación** (a veces representado como α)
- La mayor parte de los estudios en ciencias sociales usan como nivel de significación el *valor-P* 0,05
- Para algunos estudios se usan otros niveles de significación. Si se quiere estar especialmente seguro de no descartar H_0 por error, se usa, por ejemplo 0,01

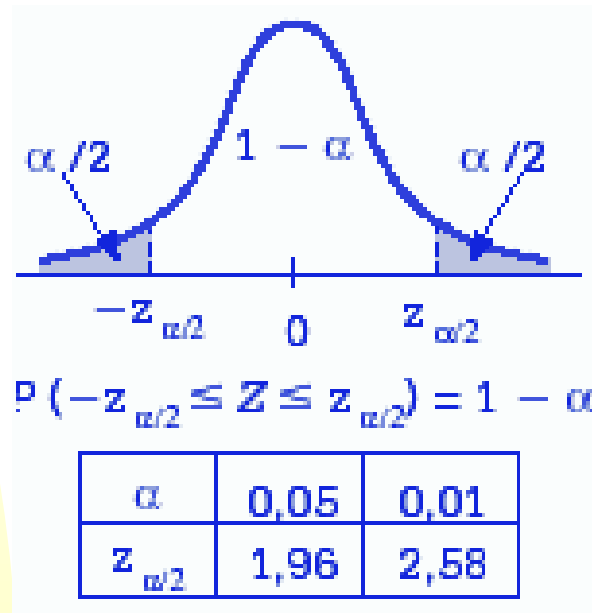
3. Metodología del contraste de hipótesis

- Dado un nivel de significación (escogido por el investigador), por ejemplo 0,05
 - ◆ Si el *valor-P* es mayor que 0,05 se considera que no puede descartarse H_0
 - ◆ Si el *valor-P* es menor o igual que 0,05 se considera que sí puede rechazarse H_0 y aceptar la hipótesis alternativa
- El nivel de significación fija la “región de rechazo” o “región crítica”: el conjunto de valores de la discrepancia para los que se rechaza H_0

3. Metodología del contraste de hipótesis

- El cálculo de la región de rechazo es distinto según vayamos a hacer un contraste de hipótesis unilateral o bilateral,
Si el contraste es bilateral (cuando no sabemos si el valor puede diferir de θ_0 hacia arriba o hacia abajo) la región de rechazo la constituyen en realidad los dos extremos simétricos de la distribución muestral cuyas probabilidades suman el nivel de confianza

3. Metodología del contraste de hipótesis



33

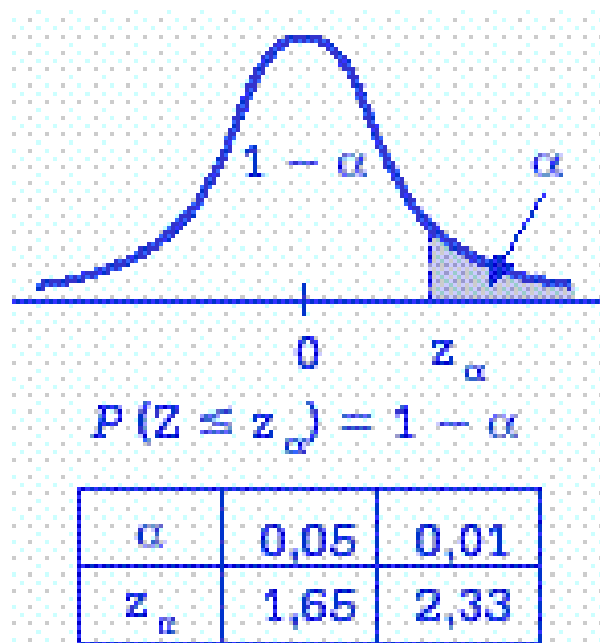
3. Metodología del contraste de hipótesis

- El cálculo de la región de rechazo es distinto según vayamos a hacer un contraste de hipótesis unilateral o bilateral,

Si el contraste es unilateral (esperamos una diferencia solo hacia arriba o hacia abajo) la región de rechazo la constituye sólo una cola de la distribución muestral

34

3. Metodología del contraste de hipótesis



35

3. Metodología del contraste de hipótesis

- Podemos resumir en una tabla los valores z que marcan la región de rechazo según el contraste sea unilateral o bilateral, y según el nivel de significación

Nivel significación (α)	Contraste unilateral	Contraste bilateral
0,05	$z < -1,65$ Ó $z > 1,65$ (según signo hipótesis alternativa)	$z < -1,96$ Y $z > 1,96$
0,01	$z < -2,33$ Ó $z > 2,33$ (según signo hipótesis alternativa)	$z < -2,58$ Y $z > 2,58$

36

3. Metodología del contraste de hipótesis

4. Tomar la muestra, calcular el estimador, medir la diferencia con θ_0 , calcular su probabilidad y compararla con el valor-P

Hasta aquí todavía no habíamos hecho nada más que “preparar” el camino

El último paso es realizar el contraste de hipótesis propiamente dicho:

- ◆ Tomar la muestra
- ◆ Medir en los sujetos la variable relevante
- ◆ Calcular el estimador (una media o una proporción)
- ◆ Calcular el valor z del estimador, suponiendo que H_0 fuera cierta

37

3. Metodología del contraste de hipótesis

- ◆ A partir de ahí tenemos dos opciones:
 - ◆ La más sencilla es simplemente comparar los valores z con los puntos de corte de la zona de rechazo que hemos visto en la transparencia 36
 - ◆ Si el valor z obtenido está dentro de la región de rechazo: rechazamos la hipótesis nula
 - ◆ Si el valor z obtenido no está dentro de la región de rechazo, no podemos rechazar la hipótesis nula

3. Metodología del contraste de hipótesis

- ◆ La segunda opción es calcular exactamente el valor-P correspondiente al valor z que hemos obtenido y comparar el valor-P con el nivel de significación previamente escogido (0,05 ó 0,01)
- ◆ Elaborar una conclusión:
 - ◆ Si el valor-P es mayor que el nivel de significación, no podemos rechazar la hipótesis nula
 - ◆ Si el valor-P es menor que el nivel de significación, rechazamos la hipótesis nula

Tema 12. Contraste de hipótesis

Introducción

1. Tipos de hipótesis
2. La hipótesis nula y la alternativa
3. Metodología del contraste de hipótesis
4. **Contraste para una proporción**
5. Contraste para una media

Resumen

Ejercicios

4. Contraste para una proporción

- Imaginemos la siguiente cuestión de investigación: queremos saber si la mayor parte de la población está a favor de una ley
- Vamos a ir dando los pasos que hemos visto en el punto anterior para realizar el contraste de hipótesis
 1. La hipótesis H_0 nula sería $p=0,5$. Y la hipótesis alternativa H_1 sería $p>0,5$. Estaríamos ante un contraste unilateral

4. Contraste para una proporción

2. La medida de discrepancia será el valor z de la proporción observada

$$z = \frac{\text{estimador} - \theta_0}{ET} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{\hat{p} - 0,5}{\sqrt{\frac{0,25}{n}}}$$

4. Contraste para una proporción

3. Decidimos el nivel de significación, es decir el valor-p máximo para rechazar H_0 . Siguiendo la convención, lo situamos en 0,05

Como el contraste es unilateral, sólo miramos en el lado positivo de la “cola” de la distribución normal

Según la tabla de la transparencia 36, el valor positivo de z que define la región de rechazo es 1,65 (la probabilidad de que z sea mayor que 1,65 es 0,05)

La región de rechazo queda constituida por todos los valores de z mayores que 1,65

4. Contraste para una proporción

Si obtenemos un valor z menor que 1,65 (cualquier valor negativo, ó 0,3 ó 1,25) no podemos rechazar la hipótesis nula (la probabilidad de que p sea menor o igual a 0,5 es mayor que 0,05)

Si obtenemos un valor z mayor que 1,65 (1,75 ó 2,1 ó 3,1) podemos rechazar la hipótesis nula, porque si $p=0,5$, existe menos de un 0,05 de probabilidad de haber obtenido un valor z como el que hemos obtenido

4. Contraste para una proporción

4. Extraemos una muestra aleatoria, pongamos de 500 personas, y obtenemos que una proporción 0,53 está a favor de la ley.

Calculamos el valor z, según la fórmula que hemos visto en el punto 2

$$z = \frac{\hat{p} - 0,5}{\sqrt{\frac{0,25}{500}}} = \frac{0,53 - 0,5}{\sqrt{0,0005}} = \frac{0,03}{0,0224} = 1,34$$

4. Contraste para una proporción

- El valor z que hemos obtenido no está dentro de la región de rechazo, ya que 1,34 es menor que 1,65
 - ◆ Por tanto, no podemos descartar la hipótesis nula
- Manera alternativa de llegar a la misma conclusión:
 - ◆ Calculamos el valor-P de 1,34 (la probabilidad de un valor z mayor o igual que 1,34), y esta es, según la tabla, $1 - 0,9099 = 0,0901$, que es mayor que el nivel de significación (0,05) que habíamos adoptado
- No es posible descartar la hipótesis nula y decir con seguridad que la proporción de personas en la población sea mayor de 0,5

Tema 12. Contraste de hipótesis

Introducción

1. Tipos de hipótesis
2. La hipótesis nula y la alternativa
3. Metodología del contraste de hipótesis
4. Contraste para una proporción
5. **Contraste para una media**

Resumen

Ejercicios

5. Contraste para una media

- Imaginemos la siguiente cuestión de investigación: queremos saber si los alumnos de los institutos públicos tienen como media una nota de selectividad distinta que la del conjunto de los estudiantes, que es de 5,5
- Repetimos los pasos que hemos visto en el punto 3, y en el punto 4 para realizar el contraste de hipótesis
 1. La hipótesis H_0 nula sería $m=5,5$. Y la hipótesis alternativa H_1 sería $m \neq 5,5$. Estaríamos ante un contraste bilateral

5. Contraste para una media

2. La medida de discrepancia será el valor z de la proporción observada, es decir:

$$z = \frac{\text{estimador} - \theta_0}{ET} = \frac{\bar{x} - m}{\hat{s}_x / \sqrt{n}} = \frac{\bar{x} - 5,5}{\hat{s}_x / \sqrt{n}}$$

5. Contraste para una media

3. Decidimos el nivel de significación, es decir el valor- p máximo para rechazar H_0 . Siguiendo la convención, lo situamos en 0,05

- ★ Como el contraste es bilateral, la región de rechazo la constituirán ambas “colas” de la distribución normal, cada una con una proporción de 0,025
- ★ Según la tabla de la transparencia 36, la región de rechazo queda constituida por todos los valores de z menores que -1,96 y los valores mayores de 1,96

5. Contraste para una media

- ◆ Si obtenemos un valor z entre $-1,96$ y $1,96$ (por ejemplo, $-1,3$, ó $-0,8$ ó $1,2$ ó $1,65$) no podemos rechazar la hipótesis nula (existe una probabilidad mayor de $0,05$ de que m sea igual que $5,5$)
- ◆ Si obtenemos un valor z menor que $-1,96$ (por ejemplo, -2 , ó $-2,25$) o mayor que $1,96$ ($2,1$ ó $3,2$) podemos rechazar la hipótesis nula, porque si $m=5,5$ existe menos de $0,05$ de probabilidad de haber obtenido un valor z como el que hemos obtenido

5. Contraste para una media

4. Extraemos una muestra aleatoria de estudiantes de institutos públicos, pongamos de 800 estudiantes, y obtenemos el dato de la variable relevante (nota de selectividad).

Supongamos que la media que obtenemos es $5,3$.

La desviación típica de la muestra es $1,8$

El valor z es entonces:

$$z = \frac{\bar{x} - 5,5}{\hat{s}_x / \sqrt{n}} = \frac{5,3 - 5,5}{1,8 / \sqrt{800}} = \frac{-0,2}{1,8 / 28,28} = \frac{-0,2}{0,0636} = -3,14$$

5. Contraste para una media

- El valor z que hemos obtenido está dentro de la región de rechazo, ya que $-3,14$ es menor que $-1,96$
 - Por tanto, podemos rechazar la hipótesis nula
- Procedimiento alternativo: calculamos el valor-P de los valores z $-3,14$ y $+3,14$ (la probabilidad de un valor z menor o igual que $-3,14$ y mayor que $3,14$).
 - La probabilidad de un valor menor o igual a $-3,14$ es, según la tabla, $0,0008$. Esa misma sería la probabilidad de un valor mayor que $3,14$
 - Si la hipótesis nula fuera cierta, la probabilidad de obtener una muestra con un estimador como el obtenido sería de $0,0016$, que es mucho menor que el nivel de confianza escogido ($0,05$)
- Llegamos a la misma conclusión: podemos descartar la hipótesis nula y decir con casi total seguridad que la nota media de los estudiantes de los institutos públicos es distinta de $5,5$

Tema 02. Contraste de hipótesis

53

Resumen

- Hipótesis
 - ◆ En relación a un valor predeterminado → contraste de hipótesis
 - ◆ En relación al valor de dos grupos → comparación de dos poblaciones (tema siguiente)
- Hipótesis nula (H_0)
- Hipótesis alternativa (H_1)

54

Resumen

- Método del contraste de hipótesis
 - ◆ Definir H_0 y H_1
 - ◆ Definir medida de discrepancia (normalmente z), basada en:
 - ★ Valor de θ_0 , derivado de H_0
 - ★ Valor de ET, derivado de la muestra
 - ◆ Definir nivel de significación, o valor-P, que da lugar a región de rechazo (normalmente 0,05 ó 0,01)
 - ◆ Calcular valor z que delimita región de rechazo
 - ◆ Tomar muestra, medir θ , calcular z , y su valor-P, y comparar con nivel de significación

55

Resumen (2)

■ Contraste de una proporción

– $H_0: \hat{p} = p$

$$z = \frac{\text{estimador} - \theta_0}{ET} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

- Tomar muestra, medir \hat{p} , calcular z
- Dos procedimientos alternativos para rechazar o no hipótesis nula
 - Comparar valor z con límites región de rechazo ó
 - Calcular valor-P de valor z y comparar con nivel de significación

56

Resumen (2)

- **Contraste de una media**

- $H_0: \bar{x} = m$

$$z = \frac{\text{estimador} - \theta_0}{ET} = \frac{\bar{x} - m}{\hat{s}_x / \sqrt{n}}$$

- Tomar muestra, medir \bar{x} y \hat{s}_x , calcular z
- Dos procedimientos alternativos para rechazar o no hipótesis nula
 - Comparar valor z con límites región de rechazo ó
 - Calcular valor-P de valor z y comparar con nivel de significación

57

Ejercicios recomendados

Del manual

- 22.1 (ver mi solución en pag.web)
- 22.2
- 22.3 (cambiar tamaño de la muestra, en lugar de 30, usar 60; ver mi solución en página web)
- 22.5

De exámenes

- Junio 06: 10
- Enero 07: 9