Matemáticas aplicadas a las Ciencias Sociales II

Introducción a la Inferencia

Manuel Molina Fernández y Jacinto Martín Jiménez

Basado en apuntes del
Máster Universitario en Formación del Profesorado en Educación Secundaria

CPR Mérida
24 febrero 2011

Introducción a la Inferencia Estadística

Principales Objetivos de la Estadística Inferencial

- La Estadística Inferencial es la parte de la Estadística que estudia métodos para obtener un conocimiento lo más fiel posible de la distribución de probabilidad de la variable(s) aleatoria(s) investigada(s) en la(s) población(es) bajo estudio.
- Proporciona los procedimientos apropiados para obtener, a partir de la información aportada por la muestra(s) seleccionada(s), conclusiones generales (inferencias científicas) sobre la población(es) objetivo de investigación.

Principales Objetivos de la Estadística Inferencial

- Además de su gran interés teórico, es una de las partes de la Estadística que mayor importancia tiene en las investigaciones de carácter aplicado. La finalidad principal de toda investigación experimental es obtener, con ciertos márgenes de error, unas conclusiones generales válidas para la población(es) bajo estudio.
- Atendiendo al conocimiento que tengamos sobre la distribución de probabilidad de la variable aleatoria bajo investigación, se distinguie entre **Inferencia Paramétrica** e Inferencia No Paramétrica.

Principales procedimientos inferenciales

Los principales procedimientos que estudia la Inferencia Estadística son:

Los Procedimientos de Estimación

Los Procedimientos de Contraste de Hipótesis

Principales procedimientos inferenciales

Procedimientos de Estimación

Su finalidad es proporcionar métodos apropiados para determinar buenas aproximaciones (estimaciones) a los parámetros de interés de la distribución de probabilidad correspondiente a la variable aleatoria bajo estudio. Proporcionan también métodos para determinar los errores máximos cometidos con dichas estimaciones.

La Teoría sobre Estimación Estadística es la parte de la Inferencia Estadística encargada del estudio de tales procedimientos.

Principales procedimientos inferenciales

Procedimientos de Contraste de Hipótesis

Su finalidad es proporcionar las herramientas necesarias para poder decidir, con cierta probabilidad de error, sobre la veracidad o no de determinada afirmación de interés en la investigación a realizar.

La Teoría sobre Contraste de Hipótesis es la parte de la Inferencia Estadística encargada del estudio de tales procedimientos.

Importante distinguir entre Estimación y Contrastes

Estimación de Parámetros

- Sea X una variable aleatoria con función de distribución F_θ dependiente de cierto parámetro desconocido $\theta \in \Theta \subseteq \mathbb{R}$.
- La teoría sobre estimación estadística estudia la metodología para determinar buenas aproximaciones al parámetro θ.
- El proceso se apoya en la observación de una muestra aleatoria simple (m.a.s.) $X_1, \ldots, X_n$ (conjunto de variables aleatorias independientes y todas ellas con la misma función de distribución F_θ asociada a X). Cada observación concreta de dicha muestra proporciona unos datos $x_1, \ldots, x_n$ que combinados de una forma adecuada determinan una aproximación $T(x_1, \ldots, x_n)$ al parámetro θ.
- La teoría sobre estimación estadística estudia también métodos para valorar el error cometido cuando se aproxima θ a través del valor $T(x_1, \ldots, x_n)$.

Estimación de los Principales Parámetros

Estimador para la Media

Supongamos que hemos de estimar la media μ de cierta distribución de probabilidad.
Sea $X_1, \ldots, X_n$ una m.a.s. de dicha distribución.

El mejor estimador $T(X_1, \ldots, X_n)$ para μ es la *Media Muestral*:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$E[\bar{X}_n] = \mu \quad \text{(es centrado).}$$
$$V[\bar{X}_n] = \sigma^2/n \quad (\sigma^2 \text{ es la varianza poblacional).}$$

Estimación de los Principales Parámetros

Estimador para la Proporción

Supongamos que tenemos que estimar la proporción p de cierta característica bajo estudio en la población. La distribución de probabilidad subyacente es la Bernoulli: $B(1, p)$). Los valores que toma dicha variable aleatoria son 1 ó 0.
Sea $X_1, \ldots, X_n$ una m.a.s. de dicha distribución.
El mejor estimador $T(X_1, \ldots, X_n)$ para p es la *proporción muestral*:

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$E[\hat{p}_n] = p$ (es centrado).
$V[\hat{p}_n] = p(1 - p)/n.$

Intervalos de Confianza

En ocasiones, en lugar de proporcionar una estimación puntual de un parámetro θ, suele ser más informativo proporcionar un intervalo numérico basado en la m.a.s.
$I = [T_1(X_1, \ldots, X_n);\; T_2(X_1, \ldots, X_n)]$
que contenga al verdadero valor del parámetro con una probabilidad prefijada suficientemente alta como para proporcionar una confianza razonable de que $\theta \in I$.

Intervalos de Confianza

Intervalo de Confianza

- Sea $X_1, \ldots, X_n$ una m.a.s. $\quad F_\theta, \quad \theta \in \Theta \subseteq \mathbb{R}$
- Se trata de determinar dos funciones de la m.a.s. que denotaremos como $T_1(X_1, \ldots, X_n)$ y $T_2(X_1, \ldots, X_n)$ tales que:

  $$P\left(T_1(X_1, \ldots, X_n) \leq \theta \leq T_2(X_1, \ldots, X_n)\right) = 1 - \alpha$$

  A la probabilidad $1-\alpha$ (puesta por el experimentador) se le denomina nivel de confianza. Se suele expresar en porcentaje y, en la investigación experimental, oscila entre el 90% y el 99%.

  $[T_1(X_1, \ldots, X_n), T_2(X_1, \ldots, X_n)]$ recibe el nombre de *Intervalo de Confianza* al nivel $(1-\alpha)\%$ para el parámetro θ.
- Antes de observar los valores concretos de la muestra, el intervalo $(T_1(X_1, \ldots, X_n), T_2(X_1, \ldots, X_n))$ es aleatorio. Una vez observados los valores $x_1, \ldots, x_n$, pasa a ser un intervalo fijo.

Intervalos de Confianza para la Media en Poblaciones Normales

Varianza conocida

- $X_1, \ldots, X_n \quad$ m.a.s $\quad N(\mu, \sigma), \quad \sigma$ conocida.
- Teniendo en cuenta que $\bar{X}_n \to N(\mu, \frac{\sigma}{\sqrt{n}})$, se deduce:

$$P\left(-z_\alpha \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq +z_\alpha\right) = 1 - \alpha$$

  z_α es el valor en la $N(0,1)$ que deja a su derecha $\alpha/2$.
- En consecuencia, se obtiene como intervalo de confianza al $(1-\alpha)\%$ para μ:

$$\left[\bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}} \;;\; \bar{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right]$$

Intervalos de Confianza para la Proporción

- $X_1, \ldots, X_n \quad$ m.a.s $\quad B(1, p), \quad p$ desconocida.
- p se estima por $\hat{p}_n$. Teniendo en cuenta la aproximación de la Binomial a través de la Normal, se deduce que $\hat{p}_n \to N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.
- Se deduce como intervalo de confianza al $(1-\alpha)\%$ para μ:

$$\left[\hat{p}_n - z_\alpha \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}; \ \hat{p}_n + z_\alpha \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right]$$

$$n\hat{p}_n > 5, \quad n(1-\hat{p}_n) > 5, \quad \hat{p}_n > 0.05, \quad 1-\hat{p}_n > 0.05$$

Objetivo del Contraste de Hipótesis

- Como parte de los objetivos de una investigación, frecuentemente se plantean determinadas preguntas que se traducen en la formulación de unas hipótesis estadísticas sobre los parámetros u otras caracteristicas de la(s) población(es) estudiada(s).
- La Teoría sobre Contraste de Hipótesis, iniciada por Neyman y E. Pearson en 1940, tiene como objetivo desarrollar la metodología necesaria para decidir, con ciertos niveles de error, la hipótesis que debe aceptarse.
- Un contraste (test) de hipótesis es un procedimiento estadístico que, a partir de la información proporcionada por la(s) muestra(s) aleatoria(s) seleccionada(s), permite aceptar o rechazar una hipótesis previamente formulada en la población(es) bajo investigación.

Conceptos básicos

- Hipótesis Nula e Hipótesis Alternativa
- Contraste Bilateral y Contraste Unilateral
- Tipos de Errores
- Probabilidades asociadas a los tipos de errores
- Pruebas de Significación

Conceptos básicos

Hipótesis Nula e hipótesis alternativa Con objeto de distinguir entre las dos hipótesis a contrastar, nos referiremos a ellas con los nombres de *Hipótesis Nula* (la denotaremos como H_0) e *Hipótesis Alternativa* (la denotaremos como H_1).

Usualmente, en la hipótesis alternativa situaremos la sospecha que se trata de confirmar.

Contraste bilateral y contraste unilateral Dependiendo del objetivo de la investigación a realizar, habrá situaciones en las que en la hipótesis alternativa tengamos el signo $\neq$ (no se concreta un sentido) y otras situaciones en las que tengamos los signos $>$ ó $<$ (se concreta un sentido). En el primer caso hablaremos de *Contraste de Hipótesis Bilateral* y en el segundo de *Contraste de Hipótesis Unilateral*.

Conceptos básicos

Tipos de errores

Puesto que hemos de decidir entre las dos hipótesis formuladas en base a la información proporcionada por la(s) muestra(s) seleccionada(s) en la(s) población(es) bajo estudio nuestra decisión estará sujeta a ciertos errores. En concreto, podemos cometer dos tipos de errores a los que nos referiremos como *Error Tipo I* y *Error Tipo II*.

Error Tipo I

Aceptar H_1 cuando en realidad H_0 es cierta.

Error Tipo II

Aceptar H_0 cuando en realidad H_1 es cierta.

Conceptos básicos

Probabilidades Asociadas a los Tipos de Errores

Los errores tipo I y tipo II no es posible cometerlos con probabilidad cero dado que la decisión se toma en base a la información proporcionada por muestras y no en base a la información total. Nuestro objetivo será que las probabilidades de cometer tales errores sean lo más próximas a cero que nos sea posible. Denominaremos:

$$\alpha = P(Cometer\ Error\ Tipo\ I) = P(Aceptar\ H_1 \mid H_0\ cierta)$$

$$\beta = P(Cometer\ Error\ Tipo\ II) = P(Aceptar\ H_0 \mid H_1\ cierta)$$

Conceptos básicos

Consideremos α, β y el tamaño de la muestra(s) que representaremos por n. De esos tres valores siempre podremos fijar dos y encontrar la regla de decisión apropiada que nos permita llegar a una conclusión.

Cabe pensar en los tres siguientes planteamientos:

- Fijar previamente el α y el β que deseemos asumir.
- Fijar previamente el α que deseemos asumir y el n que vamos a tomar.
- Fijar previamente el β que deseemos asumir y el n que vamos a tomar.

Conceptos básicos

Prueba de Significación

De acuerdo con el segundo planteamiento (en el que fijamos el n y el α) el β no lo controlamos. Se trataría de determinar la mejor regla de decisión posible que, garantizando el α prefijado, nos proporcione el menor β posible.

Es el planteamiento mas utilizado en la práctica experimental fundamentalmente por dos razones:

- Debido a que en la mayoría de las ocasiones el n lo fija el experimentador forzado por los recursos y tiempo disponibles.
- Debido a que al experimentador le interesa muy especialmente controlar el α (también denominado *nivel de significación del test*).

Muchos autores se refieren a este planteamiento, considerado originalmente por R.A. Fisher, como *Prueba o Contraste de Significación*.

Resolución práctica de un contraste de hipótesis

Recogidos los datos muestrales, el contraste o test de hipótesis suele resolverse de una forma bastante mecánica. Hay distintas formas de llevar a cabo tal resolución. La más frecuente consiste en:

- Calcular, a partir de los datos muestrales, el denominado *Valor experimental del test*. Se calcula a partir de una formula estadística que depende de cada situación concreta.
- Determinar, a partir de una distribución de probabilidad teórica, el denominado *Valor teórico del test*. Dicho valor depende del α prefijado. La distribución teórica se determina teniendo en cuenta que se está asumiendo que H_0 es cierta.
- Concretar la *regla de decisión* correspondiente a través de la comparación de ambos valores. Dicha regla puede adoptar diferentes formas dependiendo del concraste de hipótesis considerado.

Contrastes con una muestra

Contrastes sobre la Media

Sea X una variable aleatoria con distribución de probabilidad $N(\mu, \sigma)$.

Estudiaremos los siguientes contrastes de hipótesis sobre μ (para los casos σ conocida y σ desconocida):

- $H_0 : \ \mu = \mu_0 \quad H_1 : \ \mu \neq \mu_O$ (Contraste bilateral)
- $H_0 : \ \mu \leq \mu_0 \quad H_1 : \ \mu > \mu_O$ (Contraste unilateral)
- $H_0 : \ \mu \geq \mu_0 \quad H_1 : \ \mu < \mu_O$ (Contraste unilateral)

siendo μ_0 un valor conocido.

Contrastes con una muestra

$H_0 : \ \mu = \mu_0 \quad H_1 : \ \mu \neq \mu_0 \quad (\sigma$ conocida)

$X_1, \dots, X_n$ m.a.s $N(\mu, \sigma)$,
$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to N(\mu, \sigma/\sqrt{n})$

Supuesto que H_0 es cierta: $\bar{X}_n \to N(\mu_0, \sigma/\sqrt{n})$.
Fijado el nivel de significación α:

$$P\left(-z_\alpha \leq \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \leq +z_\alpha\right) = 1 - \alpha$$

$$V_{exp} = \frac{|\bar{X}_n - \mu_0|}{\sigma/\sqrt{n}}, \quad V_\alpha = z_\alpha$$

Regla de decisión:

Cuando $V_{exp} > z_\alpha$ se acepta H_1 al nivel α.
Cuando $V_{exp} \leq z_\alpha$ se acepta H_0 al nivel α.

Contrastes con una muestra

$H_0: \ \mu \leq \mu_0 \quad H_1: \ \mu > \mu_0 \quad$ (σ conocida)

$X_1, \ldots, X_n$ m.a.s $N(\mu, \sigma)$

Fijado el nivel de significación α:

$$P\left(\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > +z_{2\alpha}\right) = \alpha$$

$$V_{exp} = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}, \quad V_\alpha = z_{2\alpha}$$

Regla de decisión:

Cuando $V_{exp} > z_{2\alpha}$ se acepta H_1 al nivel α.
Cuando $V_{exp} \leq z_{2\alpha}$ se acepta H_0 al nivel α.

Contrastes con una muestra

$H_0: \ \mu \geq \mu_0 \quad H_1: \ \mu < \mu_0 \quad$ (σ conocida)

$X_1, \ldots, X_n$ m.a.s $N(\mu, \sigma)$,
$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow N(\mu, \sigma/\sqrt{n})$

Fijado el nivel de significación α:

$$P\left(\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} < -z_{2\alpha}\right) = \alpha$$

$$V_{exp} = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}, \quad V_\alpha = -z_{2\alpha}$$

Regla de decisión:

Cuando $V_{exp} < -z_{2\alpha}$ se acepta H_1 al nivel α.
Cuando $V_{exp} \geq -z_{2\alpha}$ se acepta H_0 al nivel α.

Contrastes con una muestra

Contrastes sobre una Proporción

Estudiaremos los siguientes contrastes:

- $H_0 : \ p = p_0 \quad H_1 : \ p \neq p_0$ (Contraste bilateral)
- $H_0 : \ p \leq p_0 \quad H_1 : \ p > p_O$ (Contraste unilateral)
- $H_0 : \ p \geq p_0 \quad H_1 : \ p < p_O$ (Contraste unilateral)

siendo p_0 un valor conocido.

Contrastes con una muestra

$H_0: \ p = p_0 \quad H_1: \ p \neq p_0$

$X_1, \ldots, X_n \quad \text{m.a.s} \quad B(1, p)$

$np_0 > 5, \quad n(1 - p_0) > 5, \quad p_0 > 0.05, \quad 1 - p_0 > 0.05$

Supuesto que H_0 es cierta,

$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \rightsquigarrow N(p_0, \sqrt{p_0(1 - p_0)/n})$

Fijado el nivel de significación α:

$$P\left(-z_\alpha \leq (\hat{p}_n - p_0)/\sqrt{p_0(1 - p_0)/n} \leq +z_\alpha\right) = 1 - \alpha$$

$$V_{exp} = \frac{|\hat{p}_n - p_0|}{\sqrt{p_0(1 - p_0)/n}}, \quad V_\alpha = z_\alpha$$

Cuando $V_{exp} > z_\alpha$ se acepta H_1 al nivel α.

Cuando $V_{exp} \leq z_\alpha$ se acepta H_0 al nivel α.

Contrastes con una muestra

$H_0 : p \leq p_0 \quad H_1 : p > p_0$

$X_1, \ldots, X_n$ m.a.s $B(1, p)$

$np_0 > 5, \quad n(1-p_0) > 5, \quad p_0 > 0.05, \quad 1 - p_0 > 0.05$

Fijado el nivel de significación α:

$$P\left((\hat{p}_n - p_0)/\sqrt{p_0(1-p_0)/n} > +z_{2\alpha} \right) = \alpha$$

$$V_{exp} = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}, \quad V_\alpha = z_{2\alpha}$$

Cuando $V_{exp} > z_{2\alpha}$ se acepta H_1 al nivel α.

Cuando $V_{exp} \leq z_{2\alpha}$ se acepta H_0 al nivel α.

Contrastes con una muestra

$H_0 : \ p \geq p_0 \quad H_1 : \ p < p_0$

$X_1, \ldots, X_n$ m.a.s $B(1, p)$

$np_0 > 5, \quad n(1 - p_0) > 5, \quad p_0 > 0.05, \quad 1 - p_0 > 0.05$

Fijado el nivel de significación α:

$$P\left((\hat{p}_n - p_0)/\sqrt{p_0(1 - p_0)/n} < -z_{2\alpha} \right) = \alpha$$

$$V_{exp} = \frac{\hat{p}_n - p_0}{\sqrt{(p_0(1 - p_0)/n}}, \quad V_\alpha = -z_{2\alpha}$$

Cuando $V_{exp} < -z_{2\alpha}$ se acepta H_1 al nivel α.

Cuando $V_{exp} \geq -z_{2\alpha}$ se acepta H_0 al nivel α.

Muestras Independientes y Muestras Relacionadas

Muestras Independientes y Muestras Relacionadas (Apareadas)

En general, cuando hemos de comparar dos poblaciones, la recogida de la información puede realizarse atendiendo a los diseños:

- Diseño a través de Muestras Independientes.
- Diseño a través de Muestras Relacionadas (Apareadas)

Muestras Independientes y Muestras Relacionadas

Diseño a través de Muestras Independientes

- Se selecciona una muestra de n_1 unidades de la Población 1. Independientemente, se selecciona otra muestra de n_2 unidades de la Población 2.
- Los tamaños muestrales (n_1 y n_2) no tienen necesariamente que ser iguales.
- Se determina el valor de la variable estudio X en las unidades de ambas muestras.

  En consecuencia, tendremos recogidas dos muestras de datos:

$$\{x_1^1, \dots, x_{n_1}^1\} \quad \text{y} \quad \{x_1^2, \dots, x_{n_2}^2\}$$

- Para que los contrastes comparativos realizados a través de este diseño sean objetivos, la variable X no deberá estar influenciada por otras variables (podrían ser diferentes en las dos muestras seleccionadas).

Comparación de dos Poblaciones

$H_0: \ \mu_1 = \mu_2 \quad H_1: \ \mu_1 \neq \mu_2$

$\bar{X}_1 - \bar{X}_2 \to N(\mu_1 - \mu_2, \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\ )$

Supongamos que $\sigma_1^2 = \sigma_2^2$ y además conocida. Denotemos por σ^2 a la varianza común. Tenemos entonces:

$$\bar{X}_1 - \bar{X}_2 \to N(\mu_1 - \mu_2, \sigma\sqrt{1/n_1 + 1/n_2}\ )$$

Supuesto H_0 cierta, fijado el α, se deduce que:

$$P\left(-z_\alpha \leq \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma^2(1/n_1 + 1/n_2)}} \leq +z_\alpha\right) = 1 - \alpha$$

Comparación de dos Poblaciones

$H_0 : \; \mu_1 = \mu_2 \quad H_1 : \; \mu_1 \neq \mu_2$ (continuación)

$$P\left(-z_\alpha \leq \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma^2(1/n_1 + 1/n_2)}} \leq +z_\alpha\right) = 1 - \alpha$$

Por lo tanto,

$$V_{exp} = \frac{|\bar{X}_1 - \bar{X}_2|}{\sigma\sqrt{(1/n_1 + 1/n_2)}}, \qquad V_\alpha = z_\alpha$$

Cuando $V_{exp} > z_\alpha$ se acepta H_1 al nivel α.
Cuando $V_{exp} \leq z_\alpha$ se acepta H_0 al nivel α.

Comparación de dos Poblaciones

$H_0 : \ \mu_1 \leq \mu_2 \quad H_1 : \ \mu_1 > \mu_2$

Supongamos que $\sigma_1^2 = \sigma_2^2 = \sigma^2$ conocida

Fijado el α, se deduce que:

$$P\left(\frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{(1/n_1 + 1/n_2)}} > +z_{2\alpha}\right) = \alpha$$

$$V_{exp} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{(1/n_1 + 1/n_2)}}, \qquad V_\alpha = z_{2\alpha}$$

Cuando $V_{exp} > z_{2\alpha}$ se acepta H_1 al nivel α.

Cuando $V_{exp} \leq z_{2\alpha}$ se acepta H_0 al nivel α.

Comparación de dos Poblaciones

$H_0: \ \mu_1 \geq \mu_2 \quad H_1: \ \mu_1 < \mu_2$

Supongamos que $\sigma_1^2 = \sigma_2^2 = \sigma^2$ conocida

Fijado el α, se deduce que:

$$P\left(\frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{(1/n_1 + 1/n_2)}} < -z_{2\alpha}\right) = \alpha$$

$$V_{exp} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{1/n_1 + 1/n_2)}}, \qquad V_\alpha = -z_{2\alpha}$$

Cuando $V_{exp} < -z_{2\alpha}$ se acepta H_1 al nivel α.

Cuando $V_{exp} \geq -z_{2\alpha}$ se acepta H_0 al nivel α.

Ejemplos

Media de una Población Normal

Se ha comprobado en repetidos estudios que el número de pulsaciones en reposo de ciertos deportistas sigue una distribución normal con desviación típica de 7 pulsaciones por minuto. En una muestra de 50 de esos deportistas, se obtiene una media de 47 pulsaciones. ¿Se puede rechazar a un nivel de significación de 0.01 que el número medio de pulsaciones por minuto es 45? Justificar la respuesta.

Ejemplos

Media de una Población Normal. Método 1

$H_0 : \ \mu = 45$
$H_1 : \ \mu \neq 45$
Datos:

$$\sigma = 7, \ \ n = 50, \ \ \alpha = 0.01 (\Rightarrow z_\alpha = 2.576)$$

Supuesto que H_0 es cierta: $\bar{X}_{50} \to N(45, 7/\sqrt{50})$.

$$V_{exp} = \frac{|\bar{X}_n - \mu_0|}{\sigma/\sqrt{n}} = \frac{|47 - 45|}{7/\sqrt{50}} = 2.02, \quad V_\alpha = z_\alpha = 2.576$$

Como $V_{exp} \leq z_\alpha$ se acepta H_0 al nivel α.

Ejemplos

Media de una Población Normal. Método 2

$H_0 : \mu = 45; \qquad H_1 : \mu \neq 45$

$$\text{Datos: } \sigma = 7, \ \ n = 50, \ \ \alpha = 0.01 (\Rightarrow z_\alpha = 2.576)$$

Supuesto que H_0 es cierta: $\bar{X}_{50} \to N(45, 7/\sqrt{50})$. Tenemos

$$P\left(-z_\alpha \leq \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \leq +z_\alpha\right) = 1 - \alpha;$$

$$P\left(-z_{0.01} \leq \frac{\bar{X}_n - 45}{7/\sqrt{50}} \leq +z_{0.01}\right) = 1 - 0.01;$$

$$P\left(45 - z_{0.01} * 7/\sqrt{50} \leq \bar{X}_n \leq 45 + z_{0.01} * 7/\sqrt{50}\right) = 1 - 0.01;$$

$$P(42.45 \leq \bar{X}_n \leq 47.55) = 1 - 0.01$$

Como $47 \in [42.45, 47.55]$ se acepta H_0 al nivel α.

Ejemplos

Media de una Población Normal. Método 3

$H_0: \ \mu = 45; \qquad H_1: \ \mu \neq 45$

Datos:

$$\sigma = 7, \quad n = 50, \quad \alpha = 0.01 (\Rightarrow z_\alpha = 2.576)$$

Se crea el I.C. al $(1-\alpha)\%$ para μ:

$$\left[\bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}} \ ; \ \bar{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}} \right]$$

$$\left[47 - 2.576 * 7/\sqrt{50}; \ 47 - 2.576 * 7/\sqrt{50} \right] = [44.45, 49.55]$$

Como 45 está en el I.C. se acepta H_0 al nivel α.

Ejemplos

Proporción

En una amplia población constituida por pequeñas y medianas empresas españolas se selecciona una muestra aleatoria de 180 empresas. Sabiendo que en la muestra seleccionada hay 9 empresas extremeñas, ¿Se puede rechazar a un nivel de significación de 0.05 que la proporción de empresas extremeñas en esa población es 0.09? Justificar la respuesta.

Ejemplos

Proporción. Método 1

$H_0 : \; p = 0.09$
$H_1 : \; p \neq 0.09$
Datos:

$$n = 180, \hat{p}_n = 0.05, \; \alpha = 0.05 (\Rightarrow z_\alpha = 1.96)$$

Supuesto que H_0 es cierta: $\hat{p}_n \rightarrow N(0.09, \sqrt{0.09 * (1 - 0.09)/180})$.

$$V_{exp} = \frac{|\hat{p}_n - p_0|}{\sqrt{p_0 * (1 - p_0)/n}} = \frac{|0.05 - 0.09|}{0.021} = 1.875, \quad V_\alpha = z_\alpha = 1.96$$

Como $V_{exp} \leq z_\alpha$ se acepta H_0 al nivel α.

Ejemplos

Proporción. Método 2

$H_0 : \ p = 0.09 \quad H_1 : \ p \neq 0.09$

$$\text{Datos: } n = 180, \hat{p}_n = 0.05, \ \alpha = 0.05 (\Rightarrow z_\alpha = 1.96)$$

Supuesto que H_0 es cierta: $\hat{p}_n \to N(0.09, \sqrt{0.09 * (1 - 0.09)/180})$. Tenemos

$$\begin{aligned}
P\left(-z_\alpha \leq \frac{\hat{p}_n - p_0}{\sqrt{p_0 * (1 - p_0)/n}} \leq +z_\alpha\right) &= 1 - \alpha; \\
P\left(-z_{0.05} \leq \frac{\hat{p}_n - 0.09}{0.021} \leq +z_{0.05}\right) &= 1 - 0.05; \\
P\left(0.09 - z_{0.05} * 0.021 \leq \hat{p}_n \leq 0.09 + z_{0.05} * 0.021\right) &= 1 - 0.05; \\
P(0.048 \leq \hat{p}_n \leq 0.132) &= 1 - 0.05
\end{aligned}$$

Como 0.05 está en el intervalo $[0.048, 0.132]$ se acepta H_0 al nivel α.

Ejemplos

Proporción. Método 3

$H_0: \ p = 0.09 \quad H_1: \ p \neq 0.09$

$$\text{Datos: } n = 180, \hat{p}_n = 0.05, \ \alpha = 0.05 (\Rightarrow z_\alpha = 1.96)$$

Supuesto que H_0 es cierta: $\hat{p}_n \to N(0.09, \sqrt{0.09 * (1 - 0.09)/180})$.
Se crea el I.C. al $(1-\alpha)\%$ para μ:

$$\left[\hat{p}_n - z_\alpha \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}; \ \hat{p}_n + z_\alpha \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right] =$$

$$\left[0.05 - 1.96 * \sqrt{\frac{0.05 * (1 - 0.05)}{180}}; 0.05 - 1.96 * \sqrt{\frac{0.05 * (1 - 0.05)}{180}}\right]$$

$$= [0.018, 0.081]$$

Como 0.09 no está en el I.C. no se acepta H_0 al nivel α.

Ejemplos

Diferencia de Medias de una Población Normal

La cantidad de impurezas presentes en un lote de una sustancia química se utiliza como medida para evaluar su calidad.

Una fábrica tiene dos líneas de producción. La línea 2 ha debido ser reparada debido a ciertos problemas. El fabricante desea comprobar que la línea mantiene la misma calidad que la línea 1. Para ello selecciona una muestra aleatoria de tamaño 25 en cada línea. La media de impureza para la muestra de la línea 1 es 3 unidades y la de la línea 2, 3.4 unidades. Se sabe de estudios anteriores que la desviación típica de la cantidad de impureza es 1 unidad.

¿Se puede aceptar a un nivel de significación de 0.1 que la cantidad de impurezas en las dos líneas es la misma?

Ejemplos

Diferencia de Medias de una Población Normal. Método 1

$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$

Datos:

$$\bar{X}_1 = 3,\ \bar{X}_2 = 3.4,\ \sigma = 1,\ \ n_1 = 25,\ n_2 = 25\ \alpha = 0.1 (\Rightarrow z_\alpha = 1.645)$$

Supuesto que H_0 es cierta: $\bar{X}_1 - \bar{X}_2 \to N(0, 1 * \sqrt{1/25 + 1/25}\ )$

$$\begin{aligned} V_{exp} = \frac{|\bar{X}_1 - \bar{X}_2|}{\sigma\sqrt{(1/n_1 + 1/n_2)}} = \frac{|3.4 - 3|}{1 * \sqrt{(1/25 + 1/25)}} &= 1.41 \\ V_\alpha &= z_\alpha = 1.645 \end{aligned}$$

Como $V_{exp} \leq z_\alpha$ se acepta H_0 al nivel α.

Ejemplos

Diferencia de Medias. Método 2

$H_0: \ \mu_1 = \mu_2 \quad H_1: \ \mu_1 \neq \mu_2$

$\bar{X}_1 = 3, \ \bar{X}_2 = 3.4, \ \sigma = 1, \ \ n_1 = 25, \ n_2 = 25 \ \alpha = 0.1 (\Rightarrow z_\alpha = 1.645)$

Supuesto que H_0 es cierta: $\bar{X}_1 - \bar{X}_2 \to N(0, 1 * \sqrt{1/25 + 1/25}\ )$
Tenemos

$$
\begin{aligned}
P\left(-z_\alpha \leq \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma^2(1/n_1 + 1/n_2)}} \leq +z_\alpha\right) &= 1 - \alpha \\
P\left(-z_{0.1} \leq \frac{\bar{X}_1 - \bar{X}_2}{1 * \sqrt{(1/25 + 1/25)}} \leq +z_{0.1}\right) &= 1 - 0.1 \\
P\left(-z_{0.1} * \sqrt{(1/25 + 1/25)} \leq \bar{X}_1 - \bar{X}_2 \leq +z_{0.1} * \sqrt{(1/25 + 1/25)}\right) &= 1 - 0.1; \\
P(-0.465 \leq \bar{X}_1 - \bar{X}_2 \leq 0.465) &= 1 - 0.1
\end{aligned}
$$

Como $-0.4 \in [-0.465, 0.465]$ se acepta H_0 al nivel α.