



Qué significa estadística inferencial en Matemáticas

Diccionario > Matemáticas > Estadística > Estadística inferencial

La **Estadística inferencial** o **Inferencia estadística** estudia cómo sacar conclusiones generales para toda la población a partir del estudio de una muestra, y el grado de fiabilidad o significación de los resultados obtenidos.

Muestreo probabilístico

Consiste en elegir una muestra de una población al azar. Podemos distinguir varios tipos de muestreo:

Muestreo aleatorio simple

Para obtener una muestra, se numeran los elementos de la población y se seleccionan al azar los n elementos que contiene la muestra.

Muestreo aleatorio sistemático

Se elige un individuo al azar y a partir de él, a intervalos constantes, se eligen los demás hasta completar la muestra.

Por ejemplo si tenemos una población formada por 100 elementos y queremos extraer una muestra de 25 elementos, en primer lugar debemos establecer el intervalo de selección que será igual a $100/25 = 4$. A continuación elegimos el elemento de arranque, tomando aleatoriamente un número entre el 1 y el 4, y a partir de él obtenemos los restantes elementos de la muestra.

2, 6, 10, 14,..., 98

Muestreo aleatorio estratificado

Se divide la población en clases o estratos y se escoge, aleatoriamente, un número de individuos de cada estrato proporcional al número de componentes de cada estrato.

En una fábrica que consta de 600 trabajadores queremos tomar una muestra de 20. Sabemos que hay 200 trabajadores en la sección A, 150 en la B, 150 en la C y 100 en la D.

$$\frac{20}{600} = \frac{x_1}{200} \quad x_1 = 6.6 \approx 7 \text{ trabajadores de A}$$

$$\frac{20}{600} = \frac{x_2}{150} \quad x_2 = 5 \quad 5 \text{ trabajadores de B}$$

$$\frac{20}{600} = \frac{x_3}{150} \quad x_3 = 5 \quad 5 \text{ trabajadores de C}$$

$$\frac{20}{600} = \frac{x_4}{100} \quad x_4 = 3.3 \approx 3 \text{ trabajadores de D}$$

Un muestreo puede hacerse con o sin reposición, y la población de partida puede ser infinita o finita.

En todo nuestro estudio vamos a limitarnos a una población de partida infinita o a muestreo con reposición.

Si consideremos todas las posibles muestras de tamaño n en una población, para cada muestra podemos calcular un **estadístico (media, desviación típica, proporción, ...)** que variará de una a otra.

Así obtenemos una distribución del estadístico que se llama **distribución muestral**.

Si una población tiene media μ y desviación típica σ , y tomamos muestras de tamaño n ($n > 30$, ó cualquier tamaño si la población es "normal"), las medias de estas muestras siguen aproximadamente la distribución:

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Estimación de parámetros

Es el procedimiento utilizado para conocer las características de un parámetro poblacional, a partir del conocimiento de la muestra.

Con una muestra aleatoria, de tamaño n , podemos efectuar una estimación de un valor de un parámetro de la población; pero también necesitamos precisar un:

Intervalo de confianza

Se llama así a un intervalo en el que sabemos que está un parámetro, con un nivel de confianza específico.

Nivel de confianza

Probabilidad de que el parámetro a estimar se encuentre en el intervalo de confianza.

El nivel de confianza (p) se designa mediante $1 - \alpha$.

Error de estimación admisible

Que estará relacionado con el radio del intervalo de confianza.

Estimación de la media de una población

El intervalo de confianza, para la media de una población, con un nivel de confianza de $1 - \alpha$, siendo \bar{x} la media de una muestra de tamaño n y σ la desviación típica de la población, es:

$$\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

El error máximo de estimación es:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Cuanto **mayor** sea el tamaño de la muestra, n , **menor** es el error.

Cuanto **mayor** sea el nivel de confianza, $1-\alpha$, **mayor** es el error.

Tamaño de la muestra

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

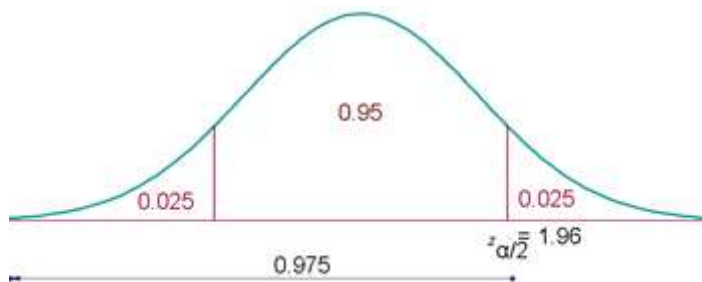
Si **aumentamos** el nivel de confianza, **aumenta** el tamaño de la muestra.

Si **disminuimos** el error, tenemos que **aumentar** el tamaño de la muestra.

El tiempo que tardan las cajas de un supermercado en cobrar a los clientes sigue una ley normal con media desconocida y desviación típica 0,5 minutos. Para una muestra aleatoria de 25 clientes se obtuvo un tiempo medio de 5,2 minutos.

1.

Calcula el intervalo de confianza al nivel del 95% para el tiempo medio que se tarda en cobrar a los clientes.



$$\left(5.2 - 1.96 \cdot \frac{0.5}{\sqrt{25}}, 5.2 + 1.96 \cdot \frac{0.5}{\sqrt{25}} \right) = (5.004, 5.396)$$

2.

Indica el tamaño muestral necesario para estimar dicho tiempo medio con un el error de $\pm 0,5$ minutos y un nivel de confianza del 95%.

$$0.5 = 1.96 \cdot \frac{0.5}{\sqrt{n}} \quad \sqrt{n} = 1.96 \quad n = 3.84$$

$$n \geq 4$$

Estimación de una proporción

Si en una **población**, una determinada característica se presenta en una proporción **p**, la proporción **p'**, de individuos con dicha característica en las **muestras** de tamaño **n**, se distribuirán según:

$$N\left(p, \sqrt{\frac{pq}{n}}\right)$$

Intervalo de confianza para una proporción

$$\left(p' - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}, p' + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \right)$$

El error máximo de estimación es:

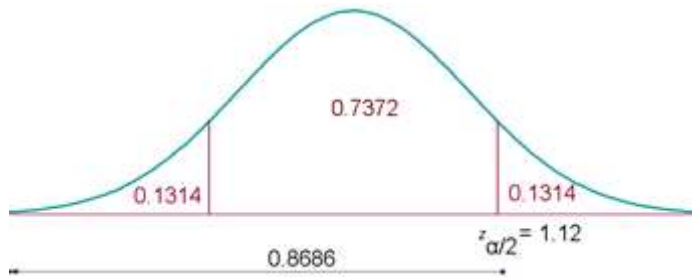
$$E = z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

En una fábrica de componentes electrónicos, la proporción de componentes finales defectuosos era del 20%. Tras una serie de operaciones e inversiones destinadas a mejorar el rendimiento se analizó una muestra aleatoria de 500 componentes, encontrándose que 90 de ellos eran defectuosos. ¿Qué nivel de confianza debe adoptarse para aceptar que el rendimiento no ha sufrido variaciones?

$$p = 0.2 \quad q = 1 - p = 0.8 \quad p' = 90 / 500 = 0.18$$

$$E = 0.2 - 0.18 = 0.02$$

$$0.02 < \sqrt{\frac{0.2 \cdot 0.8}{500}} z_{\alpha/2} \quad z_{\alpha/2} > 1.12$$



$$P(1 - z_{\alpha/2} < 1.12) = 0.86861 - 0.8686 = 0.1314$$

$$0.8686 - 0.1314 = 0.737$$

Nivel de confianza: 73.72%

Contraste de hipótesis

Hipótesis estadísticas

Un **test estadístico** es un procedimiento para, a partir de una muestra aleatoria y significativa, **extraer conclusiones** que permitan **aceptar o rechazar una hipótesis** previamente emitida sobre el valor de un parámetro desconocido de una población. La hipótesis emitida se designa por H_0 y se llama hipótesis nula. La hipótesis contraria se designa por H_1 y se llama hipótesis alternativa.

Contrastes de hipótesis

1.

Enunciar la hipótesis nula H_0 y la alternativa H_1 .

Bilateral	$H_0 = k$	$H_1 \neq k$
Unilateral	$H_0 \geq k$	$H_1 < k$
	$H_0 \leq k$	$H_1 > k$

2.

A partir de un nivel de confianza $1 - \alpha$ o el de significación α . Determinar:

El valor $z_{\alpha/2}$ (bilaterales), o bien z_{α} (unilaterales)

La zona de aceptación del parámetro muestral (\bar{x} o p').

3.

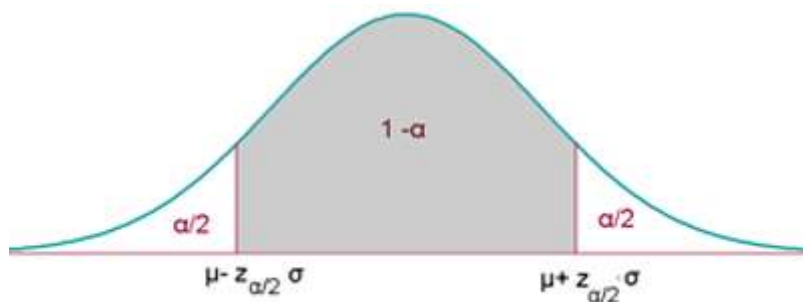
Calcular: \bar{x} o p' , a partir de la muestra.

4.

Si el valor del parámetro muestral está dentro de la zona de la aceptación, se acepta la hipótesis con un nivel de significación α . Si no, se rechaza.

Contraste bilateral

Se presenta cuando la hipótesis nula es del tipo $H_0: \mu = k$ (o bien $H_0: p = k$) y la hipótesis alternativa, por tanto, es del tipo $H_1: \mu \neq k$ (o bien $H_1: p \neq k$).



El nivel de significación α se concentra en dos partes (o colas) simétricas respecto de la media.

La **región de aceptación** en este caso no es más que el correspondiente intervalo de probabilidad para \bar{x} o p' , es decir:

$$\left(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

o bien:

$$\left(p - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}, p + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \right)$$

Se sabe que la desviación típica de las notas de cierto examen de Matemáticas es 2,4. Para una muestra de 36 estudiantes se obtuvo una nota media de 5,6. ¿Sirven estos datos para confirmar la hipótesis de que la nota media del examen fue de 6, con un nivel de confianza del 95%?

1.

Enunciamos las hipótesis nula y alternativa:

$H_0 : \mu = 6$ La nota media no ha variado.

$H_1 : \mu \neq 6$ La nota media ha variado.

2.

Zona de aceptación

Para $\alpha = 0.05$, le corresponde un valor crítico: $z_{\alpha/2} = 1.96$.

Determinamos el intervalo de confianza para la media:

$$(6 - 1,96 \cdot 0,4 ; 6 + 1,96 \cdot 0,4) = (5,22 ; 6,78)$$

3.

Verificación.

Valor obtenido de la media de la muestra: **5,6**.

4.

Decisión

Aceptamos la hipótesis nula H_0 , con un nivel de significación del 5%.

Contraste unilateral

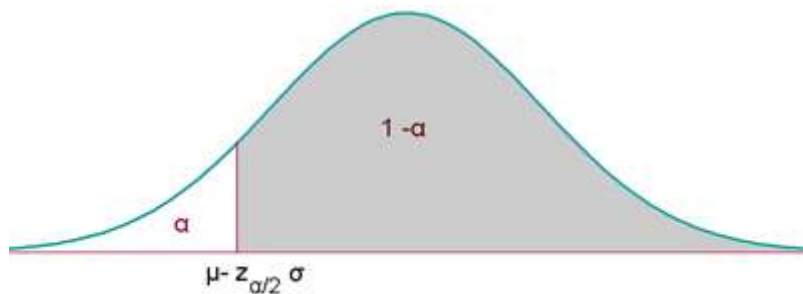
Caso 1

La hipótesis nula es del tipo $H_0: \mu \geq k$ (o bien $H_0: p \geq k$).

La hipótesis alternativa, por tanto, es del tipo $H_1: \mu < k$ (o bien $H_1: p < k$).

Valores críticos

$1 - \alpha$	α	z_α
0.90	0.10	1.28
0.95	0.05	1.645
0.99	0.01	2.33



El nivel de significación α se concentra en una parte o cola.

La región de aceptación en este caso será:

$$\left(\mu - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty \right)$$

o bien:

$$\left(p - z_\alpha \cdot \sqrt{\frac{pq}{n}}, \infty \right)$$

Un sociólogo ha pronosticado, que en una determinada ciudad, el nivel de abstención en las próximas elecciones será del 40% como mínimo. Se elige al azar una muestra aleatoria de 200 individuos, con derecho a voto, 75 de los cuales estarían dispuestos a votar. Determinar con un nivel de significación del 1%, si se puede admitir el pronóstico.

1.

Enunciamos las hipótesis nula y alternativa:

$H_0 : \mu \geq 0.40$ La abstención será como mínimo del 40%.

$H_1 : \mu < 0.40$ La abstención será como máximo del 40%;

2.

Zona de aceptación

Para $\alpha = 0.01$, le corresponde un valor crítico: $z_\alpha = 2.33$.

Determinamos el intervalo de confianza para la media:

$$\left(0.4 - 2.33 \cdot \sqrt{\frac{0.4 \cdot 0.6}{200}}, \infty \right) = (0.3192; \infty)$$

3.

Verificación.

$$p' = \frac{125}{200} = 0.625$$

4.

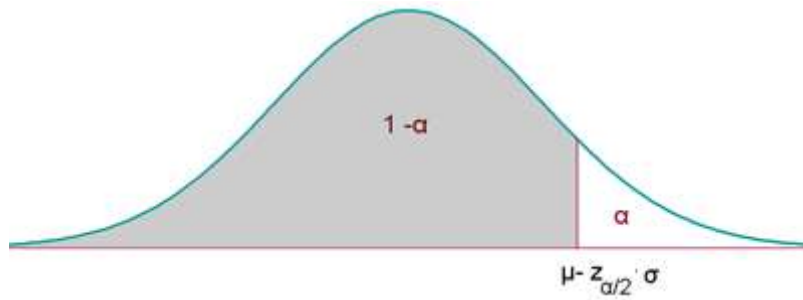
Decisión

Aceptamos la hipótesis nula H_0 . Podemos afirmar, con un nivel de significación del 1%, que la La abstención será como mínimo del 40%.

Caso 2

La hipótesis nula es del tipo $H_0: \mu \leq k$ (o bien $H_0: p \leq k$).

La hipótesis alternativa, por tanto, es del tipo $H_1: \mu > k$ (o bien $H_1: p > k$).



El nivel de significación α se concentra en la otra parte o cola.

La región de aceptación en este caso será:

$$\left(-\infty, \mu + z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

o bien:

$$\left(-\infty, p_0 + z_{\alpha} \cdot \sqrt{\frac{pq}{n}} \right)$$

Un informe indica que el precio medio del billete de avión entre Canarias y Madrid es, como máximo, de 120 € con una desviación típica de 40 €. Se toma una muestra de 100 viajeros y se obtiene que la media de los precios de sus billetes es de 128 €.

¿Se puede aceptar, con un nivel de significación igual a 0,1, la afirmación de partida?

1.

Enunciamos las hipótesis nula y alternativa:

$$H_0 : \mu \leq 120$$

$$H_1 : \mu > 120$$

2.

Zona de aceptación

Para $\alpha = 0.1$, le corresponde un valor crítico: $z_{\alpha} = 1.28$.

Determinamos el intervalo de confianza:

$$\left(-\infty; 120 + 1.28 \frac{40}{\sqrt{100}}\right) = (-\infty; 125.12)$$

3.

Verificación.

Valor obtenido de la media de la muestra: **128 €**.

4.

Decisión

No aceptamos la hipótesis nula H_0 . Con un nivel de significación del 10%.

Errores de tipo I y tipo II

Error de tipo I. Se comete cuando la hipótesis nula es verdadera y, como consecuencia del contraste, **se rechaza**.

Error de tipo II. Se comete cuando la hipótesis nula es falsa y, como consecuencia del contraste **se acepta**.

H_0	Verdadera	Falsa
Aceptar	Decisión correcta Probabilidad = $1 - \alpha$	Decisión incorrecta: ERROR DE TIPO II
Rechazar	ERROR DE TIPO I Probabilidad = α	Decisión correcta

La **probabilidad** de cometer **Error de tipo I** es el **nivel de significación α** .

La probabilidad de cometer **Error de tipo II** depende del verdadero valor del parámetro. Se hace **tanto menor cuanto mayor sea n** .

¿Qué opinas de esta definición?

3,75/5, 4 votes

Otras definiciones en Estadística

Coeficiente de variación

Contraste bilateral

Contraste unilateral

Cuartiles

Dato estadístico

Varianza

Valores críticos

Valor estadístico

Teorema central del límite

Sumatoria

¿Quieres añadir o corregir una definición?

[Añadir](#)

[Corregir](#)

Ir más allá con la teoría en Apuntes

Teoría, ejercicios y resúmenes

Resumen de estadística descriptiva

Ejercicios de estadística II

Estimación de la media de una población

Ejercicios de inferencia estadística

Problemas de inferencia estadística

Resumen de inferencia estadística

Tipos de muestreo

Problemas de contraste unilateral y bilateral de hipótesis

¿Qué es el contraste unilateral?

¿Qué es el contraste bilateral?

Sobre nosotros

Encontrar un profesor particular

¿Quiénes somos?

Material Didáctico

Sigue la aventura

Superprof en el mundo

Dar clases particulares

