Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

Marzo 2021



Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM)

> Jependencias FM

Tipos de inferencia

Estimación d parámetros

Cualidades de l

stimaciones untuales de lo

puntuales de le parámetros

Distribución de la media en el muestr

en el muestreo

Estimaciones p intervalos de lo parámetros

> itervalos de onfianza de la iedia y la variar itervalo de

9.- Inferencia estadística

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM)

Dependencias de la TM

lipos de nferencias Estimación de

Estimación de parámetros

estimadores Estimaciones

puntuales de le parámetros

Distribución de la media en el muestre

en el muestreo Estimaciones por

Estimaciones po intervalos de los parámetros

confianza de la media y la varianza Intervalo de

Introducción

Partimos de la base de que el Método Estadístico se basa en el razonamiento inductivo, o sea que hace proposiciones que van de lo "particular a lo general". Aplicando el método estadístico, la inferencia estadística nos permite sacar conclusiones sobre la población a partir de una muestra representativa.

El proceso nos permite en general, bajo un determinado marco probabilístico, realizar hipótesis sobre parámetros poblaciones desconocidos y desarrollar formas para comprobar la verosimilitud de tales afirmaciones a través de los estimadores.

Fundamentos en Estadística

Esencia de la Teoría del Muestreo(TM)

En esencia y sencillamente, la TM lo que nos tiene qué decir es:

"Cual es el comportamiento del estadístico **u** en el Muestreo".

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

stadística

ntroducción

Dependencias de la TM

nferencias

Estimación de arámetros

estimadores

stimaciones untuales de lo: arámetros

parámetros Distribución de

Distribución de la n el muestreo

stimaciones por itervalos de los arámetros

stervalos de onfianza de la sedia y la varianza stervalo de

Decir cual es el comportamiento, significa saber:

- 1. Cual es el valor esperado de u $\Rightarrow E[u]$
- 2. Cual es la varianza de $\mathbf{u} \Rightarrow V[u] \ o \ Var[u]$
- 3. Cual es **el error de u** $\Rightarrow Error[u]$
- 4. Cual es la distribución de $\mathbf{u} \Rightarrow u \cap ?$

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Dependencias de la ΓΜ

Γipos de

Estimación de arámetros

parámetros Cualidades d

stimaciones untuales de los

ntuales de los rámetros

Distribución de nedia en el mu

nedia en el m

el muestreo timaciones por

Estimaciones por intervalos de los parámetros

media y la var Intervalo de confianza para 1. El valor esperado o esperanza E[u] es la media del estadístico ${\bf u}$ en el muestreo. Media de todos los ${\bf u}$ posibles, de todo el conjunto de medias posibles.

2.
$$V[u] = s_u^2$$

- 3. La desviación típica del estadístico **u** en el muestreo, es el <u>error</u> estadístico². $\sqrt{V[u]} = Error[u] = s_u$
- 4. $u \cap {}^3$? ¿Qué tipo de distribución tiene **u**?.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Dependencias de la TM

Tipos de

nferencias Estimación de

arámetros Cualidades de

estimadores Estimaciones

untuales de lo arámetros

distribución de l nedia en el mues

n el muestreo Estimaciones por

intervalos de los parámetros

media y la varian
Intervalo de
confianza para la

Fundamentos en Estadística

Estas son las cuatro características⁴ principales a las que tiene que responder la Teoría del Muestreo (TM).

Pero a veces la teoría no puede responder(demostrar) de forma analítica a estas cuatro preguntas y aquí es cuando entran de lleno las diferentes técnicas de remuestreo por simulación numérica⁵. Lo que hacen simplemente es simular (generar) numéricamente por ordenador un conjunto de muestras posibles en el mundo del muestreo.

Dependencias de la TM

La TM (E, V, Error, \cap) depende de:

- 1. La población, i.e. de los parámetros de la población.
- 2. Criterio de selección de la muestra.
- 3. Estadístico elegido.
- 4. Tamaño de la muestra.

Fundamentos en Estadística

V. Trujillo

(CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM)

Tipos de

nferencias Estimación de

arámetros Cualidades de lo

Stimaciones untuales de los

ountuales de lo oarámetros

Distribución de la nedia en el muestre

n el muestreo Estimaciones por

Estimaciones por intervalos de los parámetros Intervalos de

media y la varia Intervalo de confianza para la

|--|

La Población

Criterio de selección de la muestra

Estadístico

Tamaño de la muestra

Fundamentos en Estadística

intervalos de los

media v la varianza

Tipos de inferencias

- 1. Estimación de parámetros de la población
- 2. **Prueba de hipótesis**. Hipótesis que hacemos sobre los parámetros de la población

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Tipos de

inferencias

timación de rámetros

rametros Jualidades de

Estimaciones

ountuales de le parámetros

distribución de la nedia en el muestrec

Distribución de en el muestreo

Estimaciones po ntervalos de los parámetros

Intervalos de confianza de la nedia y la varian: Intervalo de

Estimación de parámetros

Fundamentos en Estadística

7. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM)

TM

ipos de ferencia

Cualidades de los estimadores Estimaciones puntuales de los

stribución de la edia en el muestreo

edia en el muest istribución de la

Estimaciones por intervalos de los

Intervalos de confianza de la media y la varianza

confianza pa media Para estimar un parámetro, tengo que seleccionar un estadístico. Pero, ¿qué estadístico tengo que seleccionar?. El estadístico seleccionado se llama estimador, por ejemplo: si quiero estimar μ (media de la población), selecciono \bar{x} (media aritmética de la muestra), a \bar{x} se le llama estimador de μ , pero existen otros posibles estimadores de μ , como por ejemplo la mediana.

Pero todo esto no nos dice nada. Lo que se quiere saber es si la estimación es buena o mala. Para solucionar este problema, recurrimos, como siempre, a la TM que nos da resultados concretos sobre los estimadores.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la TM

l'ipos de nferencia

stimación de arámetros

Estimaciones untuales de los arámetros

Distribución de la nedia en el muest

nedia en el mu

n el muestreo stimaciones po

stimaciones po tervalos de los trámetros

ervalos de nfianza de la dia y la varianzs ervalo de Por ejemplo:

Parámetro	Estimador	TM	Cualidad
μ	\bar{x}	$E[\bar{x}] = \mu$	no sesgado

La TM nos dice que el valor esperado de \bar{x} (media del conjunto de todos las \bar{x} posibles) es μ (la media de la población), y nos dice también que es un estimador no sesgado o centrado, y que además no depende del criterio de selección de la muestra⁶.

Que sea no sesgado o no viciado, es una buena característica para un estimador, quiere decir que con un sólo valor de la muestra, podemos utilizarlo como estimador de la población⁷. Si hiciéramos esto miles de veces, por ejemplo: por simulación; tendríamos la misma probabilidad de que el valor se encontrase tanto a la derecha como a la izquierda de la media poblacional.

Fundamentos en Estadística

Con respecto a los estadísticos de dispersión la TM nos dice que:

Parámetro	Estimador	TM	Cualidad
σ^2	s^2	$E[s^2] = \sigma^2$	no sesgado

La TM dice que sólo es no sesgado si el criterio es un $\mathrm{m.a.s}^8$ sin reposición

Fundamentos en Estadística

7. Trujillo

RC-MERVEX (CO de Vigo. IEO)

stadística

ntroducción

sencia de la Teoría el Muestreo(TM) dependencias de la

ipos de ferencias stimación de

stimaciones untuales de los arámetros

stribución de la dia en el muest

istribución de l n el muestreo stimaciones por

timaciones por tervalos de los rámetros tervalos de

media y la vari Intervalo de confianza para

Parámetro	Estimador	TM	Cualidad
σ	s	$E[s] \neq \sigma$	es sesgado

La TM dice que el valor esperado de s en el muestreo <u>no es</u> la desviación típica poblacional. El estimador <u>es sesgado</u> o no centrado; y el sesgo es grave para un tamaño de muestra menor que 30, si n > 30 el sesgo es muy pequeño. Esta cualidad es llamada <u>consistencia</u> de un estimador.

La TM dice que: $E[s] \neq \sigma$, pero si: $n \to \infty \Rightarrow E[s] = \sigma$; a medida que el tamaño muestral se incrementa, s es mejor estimador de σ . Se dice entonces que s es un estimador consistente, esta cualidad es más importante que la de sesgado.

Fundamentos en Estadística

V. Trujillo

CO-MERVEX CO de Vigo. IEO)

stadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la TM

Γipos de nferencia

Estimación parámetros

stimaciones untuales de los arámetros

Distribución de la nedia en el muestr

Distribución d

estimaciones po ntervalos de los arámetros

ntervalos de onfianza de la ledia y la varianz ntervalo de Decíamos que si queremos estimar μ podíamos utilizar tanto la media como la mediana (\tilde{x}) . Pero, ¿con cuál de las dos nos quedaríamos?.

Parámetro	Estimador	TM	Cualidad
μ	$ar{x}$	$E[\bar{x}] = \mu$	no sesgado
μ	$ ilde{x}$	$E[\tilde{x}] = \mu$	no sesgado

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

'ipos de iferencias Estimación de

timadores timaciones

ristribución de la nedia en el muestreo

Distribución de la nuestreo

stimaciones por itervalos de los arámetros

Intervalos de confianza de la media y la varianza Intervalo de Pero si miramos cómo son las varianzas en el muestreo de estos dos estadísticos, vemos que: $V[\bar{x}] < V[\tilde{x}]$.

Esto quiere decir que la media aritmética como estimador de μ es más <u>eficiente</u> o **preciso** que la mediana. Que sea más <u>eficiente</u> (formalmente en estadística) significa que la varianza en el muestreo es menor.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

del Muestreo(TM)

Dependencias de la

Tipos de

Estimación de parámetros

stimaciones untuales de los arámetros

irametros istribución de la

nedia en el mues: Distribución de la

Estimaciones po ntervalos de los

ntervalos de los parámetros Intervalos de

Intervalo de confianza para la Con estos ejemplos se han visto tres⁹ de las principales cualidades/propiedades de los estimadores puntuales de parámetros de la población, que son:

- 1. $Sesgo^{10}$
- 2. Consistencia
- 3. Eficiencia¹¹

Fundamentos en Estadística

V. Trujillo

(CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

l'ipos de nferencias Estimación de

estimación d parámetros

Estimaciones ountuales de los oarámetros

distribución de la dedia en el muesto

nedia en el muest Distribución de la

n el muestreo Estimaciones po ntervalos de los

Estimaciones por intervalos de los parámetros

> onfianza de la edia y la varianza etervalo de

Estimaciones puntuales de los parámetros

Fundamentos en Estadística

7. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM)

m pos de

stimación de arámetros

estimadores

istribución de l

media en el muestreo Distribución de la s²

Estimaciones po intervalos de los

Intervalos de confianza de la media y la varianza Intervalo de

Aquí veremos sólo las estimaciones puntuales y por intervalos, pero quedan dos métodos de estimación muy importantes a considerar:

- la estimación de máxima verosimilitud.
- la estimación bayesiana.

Fundamentos en Estadística

Distribución de la media en el muestreo

Fundamentos en Estadística

> RC-MERVEX CO de Vigo.

estadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

ipos de iferencias

stimación d arámetros

arametros Cualidades d

stimaciones intuales de l

puntuales de parámetros

istribución de la s²

stimaciones po ntervalos de los

intervalos de los parámetros

ntervalos de onfianza de la nedia y la varianz ntervalo de

]

$$E[\bar{x}] = \mu \tag{1}$$

 $\mu = \text{media de la población}$

 $\bar{x} = \text{media de la muestra}$

E = media del muestreo

Se ve cómo están relacionados los tres mundos de la estadística:

"La media en el muestreo de la media de la muestra es la media de la población"

$$V[\bar{x}] = \frac{\sigma^2}{n} \tag{2}$$

Esto es cierto, cuando población es infinita o cuándo el muestreo es con reposición.

Si N es finita y el criterio del muestreo es un aleatorio simple(a.s.) sin reposición, entonces 12 :

$$V[\bar{x}] = \left(1 - \frac{n}{N}\right) \frac{S^2}{n} \tag{3}$$

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

del Muestreo(TM)

Dependencias de la
TM

Γipos de nferencias

Estimación de parámetros

Cualidades de los estimadores

Estimaciones puntuales de parámetros

> Distribución de la s² n el muestreo

Estimaciones po ntervalos de los

ntervalos de onfianza de la nedia y la varia II'. (N-1) $S^2=N\sigma^2$; S^2 es la varianza modificada. Sabiendo una varianza se puede deducir la otra. Normalmente como la corrección para poblaciones finitas $(1-\frac{n}{N})$ es pequeña, salvo en casos excepcionales, se suele usar el conocido:

$$V[\bar{x}] = \frac{s^2}{n} \tag{4}$$

donde s^2 es la varianza de la muestra.

Como la población suele ser muy grande: $\left(1-\frac{n}{N}\right)*\frac{S^2}{n}\cong\frac{\sigma^2}{n}$ La varianza en el muestreo de \bar{x} es igual a la varianza de la población dividido por el tamaño de la muestra. $V[\bar{x}]$ es más pequeño a medida que el tamaño de la muestra aumenta, es decir: la distribución estará más concentrada entorno a la verdadera media de la población, luego la estimación es más precisa¹³.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

del Muestreo(TM)

Dependencias de la
TM

apos de nferencia

Estimación de parámetros

estimadores
Estimaciones
puntuales de los

Distribución de la s² en el muestreo

Estimaciones po intervalos de los parámetros

ntervalos de onfianza de la nedia y la varianza ntervalo de III.

$$Error[\bar{x}] = \sqrt{V[\bar{x}]} = \frac{\sigma}{\sqrt{n}} \tag{5}$$

La desviación típica en el muestreo $(Error[\bar{x}])$, cuánto mayor sea el tamaño de la muestra, menor será su error¹⁴.

Si desconocemos σ^2 , podremos utilizar la varianza muestral s^2 para calcular el $(\underline{error}[\bar{x}])$. Cuando hacemos esto se denomina $var[\bar{x}]$ en vez de $V[\bar{x}]$.

Si

$$var[\bar{x}] = \sqrt{\frac{s^2}{n}} \quad \Rightarrow \quad error[\bar{x}] = \frac{s}{\sqrt{n}}$$
 (6)

Cuando usamos esta aproximación $(\frac{s^2}{n})$, la \bar{x} se distribuye como una distribución t de Student. En la práctica es lo que se utiliza, pero . . . si n>30 la t-Student se comporta como una distribución Normal¹⁵.

Fundamentos en Estadística

V. Trujillo

CO de Vigo. IEO)

estadística

Introducción

del Muestreo(TM)
Dependencias de la
TM

ipos de iferencias

Estimación de parámetros

Cualidades de lo estimadores

puntuales de l parámetros

Distribución de la s^2 en el muestreo

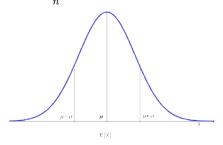
Estimaciones po intervalos de los parámetros

ntervalos de onfianza de la nedia y la varianz ntervalo de

$$\bar{x} \cap N(E, V)$$
 (7)

la media de la muestra tiene una distribución normal con media, el valor esperado de $(E[\bar{x}])$ y varianza $V[\bar{x}]$. Entonces

$$\bar{x} \cap N(E,V) \Rightarrow N(\mu, \frac{\sigma^2}{n})$$
 ; donde $\dot{\cap}$ o \sim \rightarrow "aproximadamente"



Es decir que la TM nos dice también, que la media muestral en el mundo del muestreo se distribuye como una normal de media la esperanza y desviación típica, el error típico. Es fundamental conocer la distribución de los estadísticos en el muestreo.

Para cualquier población, la aproximación a la Normal es mejor cuando n es grande (n>30) y cuando la población es simétrica. Deja de ser aproximada para ser exacta, cuando la <u>Población es Normal</u>.

Fundamentos en Estadística

V. Truiillo

GRC-MERVEX (CO de Vigo.

estadística

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Γipos de nferencia

> stimación o arámetros

Cualidades de lo estimadores

puntuales de parámetros

Distribución de la s² en el muestreo

Estimaciones printervalos de l

Intervalos de confianza de la media y la varianza

Distribución de la s^2 en el muestreo

T

$$E[s^2] = \sigma^2 \tag{8}$$

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoria del Muestreo(TM) Dependencias de la

Tipos de inferencias

estimación de arámetros

Cualidades de los estimadores stimaciones

timaciones intuales de los rámetros

Distribución de la media en el muestreo

Estimaciones por intervalos de los parámetros

Intervalos de confianza de la media y la varianza Intervalo de confianza para la II. Por ahora vamos a dejar de ver como son: $V[s^2]$ y el $Error[s^2]$, siendo de mayor más interés ver cómo es la distribución de s^2 :

$$s^2 = \frac{S_{xx}}{(n-1)}$$
 ; s^2 varianza de la muestra

$$\left(\frac{S_{xx}}{\sigma^2}\right) \cap \chi^2_{(n-1)} \Rightarrow E\left[\frac{S_{xx}}{\sigma^2}\right] = n - 1 \to E\left[\frac{S_{xx}}{(n-1)}\right] = \sigma^2 \qquad (9)$$

Por eso decimos que¹⁶:~ $E[s^2] = \sigma^2$

Nota: igual veis un desarrollo análogo, ya que realmente $E[s^2] = \sigma^2$ es una "aproximación" 17 :

$$E\left[\frac{1}{n}\sum_{i}(x_{i}-\bar{x})^{2}\right] = \frac{(n-1)}{n}\sigma^{2}$$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i} (x_i - \bar{x})^2 \sim \chi^2_{(n-1)}$$
 (10)

Si la variable X tiene una distribución normal, la distribución de $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

Entre otras propiedades, la varianza muestral no es simétrica, tiene asimetría positiva.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

del Muestreo(TM)

Dependencias de la

Tipos de inferencias Estimación de

> Cualidades de los estimadores estimaciones

rámetros stribución de la

media en el muestreo

Estimaciones por intervalos de los parámetros

ntervalos de onfianza de la nedia y la varianza ntervalo de

Estimaciones por intervalos de los parámetros

Fundamentos en Estadística

7. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Esencia de la Teoría del Muestreo(TM)

Tipos de

nferencias Estimación de

Cualidades de los estimadores

Estimaciones untuales de los

Distribución de la nedia en el muestre

Distribución de la en el muestreo

confianza de la media y la varianza Intervalo de confianza para la media

Intervalos de confianza de la media y la varianza

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Esencia de la Teoría del Muestreo(TM)

Dependencias de la TM

lipos de nferencias

arámetros Cualidades de l

estimadores Estimaciones

Estimaciones ountuales de lo oarámetros

Distribución de la media en el muestrec

Distribución de la s en el muestreo

Estimaciones po intervalos de los parámetros

confianza para la media Al estimar los parámetros obteníamos un valor puntual y como nos dice la teoría de muestreo: la distribución de los mismos. Esta distribución nos permite establecer la precisión de nuestra estimación gracias al cálculo de los intervalos de confianza¹⁸(C) de nuestros estimadores.

La teoría del muestreo nos da los datos necesarios para calcular los intervalos de confianza de los estimadores. diciéndonos que la media se distribuye como una normal de media μ y desviación típica, el error de la media.

Si tipificamos, se distribuirá como una normal de media cero y desviación típica 1 ($Z \sim N(0,1)$).

Fundamentos en Estadística

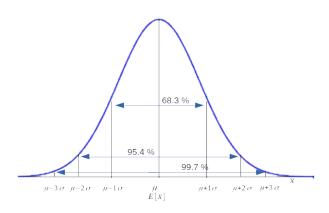
Como vimos en la ecuación sobre la función de densidad (en el capítulo 7), definíamos la función de densidad de probabilidad (pdf) para un intervalo dado como:

$$P(a \le X \le b) = \int_{a}^{b} f(x)dx \tag{11}$$

Que nos permite calcular una densidad entre dos valores dados (a,b) y viceversa, a partir de una determinada área establecer cuales son los valores que acotan tal densidad.

Fundamentos en Estadística

Para una distribución normal tipificada, las densidades correspondientes a 1, 2 y 3 desviaciones(σ), comprenden aproximadamente el 68.3 %, 95.4 % y 99.7 % respectivamente de la densidad. A estas áreas o densidades les podremos denominar **Confianzas** y al resto de la densidad¹⁹ le llamaremos α o **nivel de significación**: $\alpha = 1 - C$, que tiene una enorme importancia en la estadística y que se verá en mayor profundidad en la sección siguiente sobre el Contraste o Prueba de hipótesis.



Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Γipos de

stimación o

Cualidades

Estimaciones ountuales de l

istribución de la iedia en el muest

Distribución d

Estimaciones po intervalos de los

ntervalo de confianza para la nedia

Intervalo de confianza para la media

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM)

Dependencias d FM

Tipos de inferencia

timación de rámetros

Cualidades de los estimadores

Estimaciones puntuales de le parámetros

Distribución de la nedia en el muestr

Distribución d en el muestreo

Estimaciones po intervalos de los parámetros

> ntervalos de confianza de la nedia y la varian

En el caso de la desviación normal tipificada (Z), los intervalos de confianza se construirán habitualmente²⁰ de la siguiente manera:

$$P(-z \le Z \le +z) = 1 - \alpha = C$$

$$P\left(-z \le \frac{\bar{x} - \mu}{s/\sqrt{n}} \le +z\right) = 1 - \alpha = C$$

$$P\left(-z \le \frac{\bar{x} - \mu}{s/\sqrt{n}} \le +z\right) = 1 - \alpha = C$$

$$P\left(-z \le \frac{s}{\sqrt{n}} \le \bar{x} - \mu \le +z \le \frac{s}{\sqrt{n}}\right) = 1 - \alpha = C$$

$$P\left(-z \le \frac{s}{\sqrt{n}} \le \mu - \bar{x} \le +z \le \frac{s}{\sqrt{n}}\right) = 1 - \alpha = C$$

$$P\left(\bar{x} - z \le \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z \le \frac{s}{\sqrt{n}}\right) = 1 - \alpha = C$$

$$P(L_1 \le \mu \le L_2) = 1 - \alpha = C$$

$$L_1 = \bar{x} - z \frac{s}{\sqrt{n}} \qquad y \qquad L_2 = \bar{x} + z \frac{s}{\sqrt{n}}$$
(13)

 L_1 y L_2 , son los límites del intervalo de confianza.

Fundamentos en Estadística

Si el tamaño de las muestras es pequeño y no conocemos la desviación típica paramétrica, entonces debemos utilizar la distribución ${\bf t}$, para usar la desviación típica del muestreo ${s_{\tilde x}}^{21}$:

$$Var[\bar{x}] = \frac{s^2}{n}$$
 ; $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

Donde s^2 y sson la varianza y desviación típica de la muestra. Siendo los límites del intervalo de confianza:

$$L_1 = \bar{x} - ts_{\bar{x}}$$
 ; $L_2 = \bar{x} + ts_{\bar{x}}$ (14)

Debido a que:

$$C = P(-t \le t_{n-1} \le +t)$$

$$C = P(\bar{x} - ts_{\bar{x}} \le \mu \le \bar{x} + ts_{\bar{x}})$$
(15)

Si se supone que los límites son simétricos este intervalo da la mínima amplitud, si se hacen no simétricos la amplitud del intervalo de confianza es mayor para esa confianza. Así que la teoría de muestreo aconseja que sean simétricos.

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

del Muestreo(TM)

vi Vi

ipos de iferencias

Estimación de parámetros

Cualidades de los estimadores

stimaciones untuales de los arámetros

Distribución de la nedia en el muestreo

Distribución de la

Estimaciones po intervalos de los parámetros

ntervalos de confianza de la

Intervalo de confianza para la varianza

Fundamentos en Estadística

7. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM)

m ipos de

ferencias stimación de arámetros

ualidades de los stimadores

stimaciones untuales de los

Distribución de la nedia en el muestre

iedia en el muestreo Distribución de la *s*² n el muestreo

Estimaciones po intervalos de los parámetros

Intervalos de confianza de la media y la varianza Intervalo de La teoría de muestreo en el caso de la varianza nos informa en el muestreo sobre la esperanza, la varianza y el error, así como su distribución. Por lo tanto, se puede calcular perfectamente el intervalo de confianza de la varianza.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducciór

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Tipos de inferenc

stimación de arámetros

Cualidades de lo estimadores

timaciones ntuales de los rámetros

oarametros Distribución de

iedia en el mu Distribución de

stimaciones por stervalos de los arámetros

arámetros ntervalos de onfianza de la nedia y la varianza ntervalo de

Fundamentos en Estadística

 $\frac{S_{xx}}{-2} \cap \chi_{n-1}^2;$

Según la teoría del muestreo:

 $C = P\left(L_1 \le \frac{S_{xx}}{\sigma^2} \le L_2\right);$

 $Error\left(\frac{S_{xx}}{\sigma^2}\right) = \sqrt{2(n-1)}$

 $Como \frac{S_{xx}}{\sigma_x^2} \cap \chi_{n-1}^2 \Rightarrow E\left(\frac{S_{xx}}{\sigma_x^2}\right) = n-1$

 $L_1 = (n-1) - \chi_{n-1(\alpha/2)}^2 Error\left(\frac{S_{xx}}{\sigma^2}\right)$

 $L_2 = (n-1) + \chi_{n-1(\alpha/2)}^2 Error\left(\frac{S_{xx}}{\sigma^2}\right)$

Entonces:

$$\frac{S_{xx}}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \cap \chi_{n-1}^2$$

$$P(-\chi_{n-1}^2 \le \chi^2 \le +\chi_{n-1}^2)$$

$$P\left(\chi_{n-1(1-\alpha/2)}^2 \le \frac{(n-1)s^2}{\sigma^2} \le \chi_{n-1(\alpha/2)}^2\right) = C$$

$$P\left(\frac{(n-1)s^2}{\chi_{n-1(\alpha/2)}^2} \le \sigma^2 \le \frac{(n-1)s^2}{\chi_{n-1(1-\alpha/2)}^2}\right) = C$$

$$Si\ (n-1)s^2 = \sum y^2 \quad \Rightarrow \quad P\left(\frac{\sum y^2}{\chi_{n-1(1-\alpha/2)}^2} \le \sigma^2 \le \frac{\sum y^2}{\chi_{n-1(1-\alpha/2)}^2}\right) = C$$

Fundamentos en Estadística

Contraste o Prueba de hipótesis

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM)

ependencias o M

l'ipos de nferencias

arámetros Cualidades de le

estimadores Estimaciones

puntuales de la parámetros

Distribución de la media en el muestreo

en el muestreo

Estimaciones por intervalos de los parámetros

confianza de la media y la varianza Intervalo de Prácticamente cualquier suposición puede transformarse en una hipótesis que se puede probar. La hipótesis es lo que queremos probar y en cambio una suposición, al igual que una asunción, no necesita ser probada.

Las hipótesis se hacen siempre sobre el mundo de la Población.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la TM

ipos de ferencias

estimación d parámetros

estimadores Estimaciones

ountuales de lo earámetros

Distribución de la media en el muestre

Distribución de en el muestreo Estimaciones po

Estimaciones po ntervalos de los arámetros

ntervalos de onfianza de la nedia y la variar otervalo de Esto no hay que olvidarlo nunca. La hipótesis se denomina tradicionalmente como H_0 , por ej.: podría ser que la media de mi población es igual a $4 \Rightarrow H_0: \mu = 4$; o, que la varianza de mi población es igual a $5 \Rightarrow H_0: \sigma^2 = 5$; o que tengo una población (1) que tiene como media μ_1 y otra población(2) que tiene como media μ_2 y por lo tanto podríamos plantear una hipótesis del tipo $\Rightarrow H_0: \mu_1 = \mu_2$

Siempre sobre la Población, esto es lo que tenemos que probar. Para aceptar (no rechazar)²² o no(rechazar) una hipótesis tenemos que seguir siempre el mismo proceso, que este consta de 5 partes:

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la TM

Tipos de inferencia

estimación de arámetros

Cualidades de l

stimaciones untuales de lo arámetros

stribución de la edia en el muestr

edia en el muest istribución de la

stimaciones po stervalos de los

tervalos de los rámetros

tervalos de nfianza de la edia y la varianz: tervalo de

I.- Proceso de realización:

- 1. Formular la hipótesis $\Rightarrow H_0 / H_1$.
- Definir el nivel de confianza²³.
 Definir un criterio de selección y el tamaño de la muestra.
 Definir el estadístico que se utilizará.
- 3. Definir el criterio de decisión sobre rechazar la H_0 o no hacerlo. Esto se hace en base a la TM.
- 4. Seleccionar la muestra y calcular el valor del estadístico de contraste.
- 5. Toma de decisión: rechazar o no rechazar H_0 .

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Tipos de inferencias

Estimación de parámetros

stimadores

intuales de lo: arámetros

istribución de la edia en el muestr

istribución de l n el muestreo

timaciones por ervalos de los rámetros

tervalos de nfianza de la edia y la varia

Intervalo de confianza para

Cuadro de decisión:

		HIPÓTESIS	
		VERDADERA	FALSA
DECISIÓN	RECHAZAMOS	Incorrecta	Correcta
	NO RECHAZAMOS	Correcta	Incorrecta

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM)

Tipos de

inferencias Estimación de parámetros

Cualidades de l estimadores

Estimaciones ountuales de los oarámetros

istribución de la edia en el muestr

istribución de

en el muestreo
Estimaciones por intervalos de los

parámetros
Intervalos de confianza de la media y la varianza

media y la va Intervalo de confianza par Otra decisión podría ser: ni rechazar, ni no-rechazar 24 . En estadística, nuestras decisiones son tomadas con probabilidad, <u>nunca tenemos la certeza</u> (casi nunca). Lo que tenemos es: la probabilidad de estar en lo correcto o en lo incorrecto ($\underline{\mathbf{no}}$, verdadero o falso). Por eso en estadística, hacemos otro cuadro hablando de probabilidades.

Fundamentos en Estadística

V. Trujillo

(CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Tipos de inferencia

stimación de arámetros

Cualidades de lo estimadores

stimaciones intuales de los arámetros

edia en el mues

nedia en el mues Distribución de l

stimaciones por stervalos de los

oarámetros ntervalos de confianza de la nedia y la varian

II.- Probabilidades asociadas:

		HIPÓTESIS	
		VERDADERA	FALSA
DECISIÓN	RECHAZAMOS	α Error tipo I	$1-\beta$
	NO RECHAZAMOS	1-α	β Error tipo II

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

estadística

ntroducción

Esencia de la Teoría del Muestreo(TM)

тм Tipos de

inferencias Estimación de parámetros

Cualidades de l estimadores

Estimaciones puntuales de los parámetros

istribución de la edia en el muestreo

nedia en el muestre Distribución de la s

en el muestreo Estimaciones po

Estimaciones por intervalos de los parámetros

confianza de la media y la varianza Intervalo de confianza para la Donde α es: la probabilidad de rechazar una hipótesis verdadera y β es: la probabilidad de no rechazar una hipótesis falsa. También son conocidos como error de tipo I (α) y tipo II²⁵ (β) respectivamente.

 $(1-\alpha)$ es la confianza $(C)^{26}$ que se puede interpretar como: tener una confianza del 95 % de no rechazar la hipótesis cuando es verdadera. En prueba de hipótesis, en vez de la confianza, lo habitual es hablar más del error; y a α se le llama <u>nivel de significación</u>. Por lo tanto, tener un α de 5 % implica tener una confianza del 95 %, es decir, significa tener una probabilidad del 5 % de rechazar la hipótesis cuando esta sea verdadera.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la TM

l'ipos de nferencia

stimación de

parámetros

estimadores Estimaciones

untuales de lo arámetros

Distribución de la nedia en el muestreo

Distribución de la en el muestreo

stimaciones p ntervalos de lo arámetros

tervalos de onfianza de la edia y la varia:

$$C = 1 - \alpha = P(no\ rechazar\ /\ H_0\ verdadera)$$

 $\alpha = P(rechazar\ /\ H_0\ verdadera)$

$$Potencia = 1 - \beta = P(rechazar / H_0 \ falsa)$$
$$\beta = P(no \ rechazar / H_0 \ falsa)$$

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM)

Dependencias de la

ipos de ferencias stimación de

Estimación de parámetros Cualidados do

estimadores Estimaciones ountuales de los

ountuales de lo parámetros

Distribución o

nedia en el muestre Distribución de la s

Estimaciones por intervalos de los parámetros

confianza de la media y la varianza Intervalo de confianza para la

Ejemplo 27 :

- 1. $H_0: \mu = 4$, con σ^2 conocida $\Rightarrow \sigma^2 = 12,1$
- 2. .- nivel de confianza y significación, 95 % y 5 % respectivamente. .- criterio de selección de la muestra: m.a.s sin reposición, de tamaño n=10.
 - .- estadístico: \bar{x}

3.

$$C = P(\bar{x} - z * s_{\bar{x}} \le \mu \le \bar{x} + z * s_{\bar{x}})$$

$$C = P(\mu - z * s_{\bar{x}} \le \bar{x} \le \mu + z * s_{\bar{x}})$$

Esta última expresión es la que se usa ya que es la probabilidad de no rechazar H_0 cuando es verdadera y cuando H_0 es verdadera $\mu=4$.

Ahora se calculan los valores que se necesitan, por ejemplo el error de la media muestral.

$$s_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{12,1}{10} = 1,21 \quad \Rightarrow \quad s_{\bar{x}} = 1,1$$

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

estadística

Introducción

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Tipos de inferencias

estimación de arámetros

Jualidades de los stimadores

timaciones ntuales de lo rámetros

stribución de la edia en el muestre

istribución de la

timaciones po tervalos de los rámetros

ntervalos de confianza de la nedia y la varian ntervalo de 3. con't

Como sabemos que:

$$C = P(-z \le Z \le +z)$$

$$0.95 = P(-1.96 \le Z \le +1.96) \quad ; z = 1.96$$

$$0.95 = P(4 - 1.96 * 1.1 \le \bar{x} \le 4 + 1.96 * 1.1)$$

Quedando finalmente:

$$0.95 = P(1.84 \le \bar{x} \le 6.16)$$
 / $H_0: \mu = 4$ $\Rightarrow H_0: es\ verdadera$

Si a partir de la la muestra se obtiene una media comprendida entre 1,84 y 6,16, se tiene un 95 % de confianza de no rechazar H_0 , y esta es la regla de decisión. Esta decisión tiene una probabilidad del 95 % de no rechazarse y un 5 % de rechazarse.

A la región de rechazo se le denomina también **región crítica** y a la de no rechazo, **región de aceptación**.

- 4. Selección de la muestra y cálculo de la media.
- 5. Si la muestra cae en la región de no rechazo entonces H_0 no es rechazada, con un nivel de significación α (H_0 no es rechazada con un nivel de significación del 5%). Si cae en la región de rechazo, ahí sí que se puede decir que es significativamente diferente.

Fundamentos en Estadística

V Truiillo

RC-MERVEX CO de Vigo.

9.- Inferencia estadística

ntroducció

Esencia de la Teoría del Muestreo(TM) Dependencias de la

Γipos de nferencias

Estimación d parámetros

estimadores Estimaciones puntuales de lo

ountuales de lo parámetros

istribución de l edia en el mues

n el muestreo stimaciones por

Estimaciones por intervalos de los parámetros

Intervalos de confianza de la media y la varianza Intervalo de



Fundamentos en Estadística

intervalos de los

media v la varianza