ESTIMACIÓN PUNTUAL Y POR INTERVALOS

Ingeniería

Universidad Mendoza

OBJETIVOS DEL PROCESO INFERENCIAL

- Nuestro objetivo es construir un modelo que nos permita analizar la o las variables de interés.
- Para predecir su comportamiento suponemos que el fenómeno aleatorio poblacional se puede describir mediante un determinado modelo probabilístico.
- La completa especificación de el modelo depende de los valores concretos de sus parámetros.
- El hecho que motiva la necesidad de los métodos de inferencia estadístico es el desconocimiento de los verdaderos valores de estos parámetros.
- ***Buscar el valor de estos parámetros es el objetivo de la estadística inferencial.***
- ***Existen varios caminos para lograrlo, como:***

Áreas que estudiaremos

- Estimación puntual
- Estimación por intervalos
- Pruebas de hipótesis

Proceso inferencial

- El proceso inferencial se basará siempre en la información que nos pueda suministrar una muestra aleatoria extraída de esa población.

ESTIMACIÓN PUNTUAL

- Supongamos que tenemos una muestra aleatoria:

$$X_1, X_2, ..., X_n \sim f_X(., \theta)$$

con $\theta \in \Theta$

- Conocemos la distribución f_X y el espacio de parámetros Θ pero desconocemos exactamente el valor de θ

- Nos proponemos ahora "**estimar**" el valor del parámetro θ o una función del mismo h(θ)a partir de lo que podamos observar en la muestra.

Estimación Puntual

- Deseamos conocer el valor de un parámetro desconocido.
- Para ello se obtiene de la población la información precisa mediante una muestra aleatoria simple, se establece una función de los valores muestrales sin que contenga ningún parámetro de la distribución, ***estimador***, y se asigna al parámetro el valor que tome esta función en la muestra concreta, valor denominado ***estimación puntual***

ESTIMACIÓN DE LOS PARÁMETROS DE LA POBLACIÓN

PARÁMETRO	ESTADÍSTICA	ESTIMACION PUNTUAL
θ	$\hat{\Theta}$	$\hat{\theta}$
μ	$\overline{X}$	$\overline{x}$
σ^2	S^2	s^2
p	$\hat{P}$	$\hat{p}$
	ESTIMADOR	ESTIMACIÓN

PROPIEDADES DESEABLES DE UN BUEN ESTIMADOR

- INSESGADO

$$E(\hat{\Theta}) = \theta \quad ejemplo \quad E\left(\overline{X}\right) = \mu$$

- EFICIENTE

Dados dos estimadores insesgados $\hat{\Theta}_1$ $\hat{\Theta}_2$ de algún parámetro θ el de menor varianza se llama estimador mas eficiente

Ejemplo (1)

- Sea Una población $X \sim Bernoulli(p)$
- Sea la muestra aleatoria $X_1, X_2, ..., X_n \sim Bernoulli(p)$

Como $E(X) = p$ siendo $p = \sum_i \frac{x_i}{n}$

es natural proponer como estimador del parámetro a $\overline{X}$ veamos si es insesgado.

$$E\left(\overline{X}\right) = p$$

- $$E\left(\overline{X}\right) = E(X) = \mu$$

- Por lo que $\overline{X}$ resulta un estimador insesgado para el parámetro.

ESTIMACIÓN POR INTERVALO

- Encontrar un estimador puntual para un parámetro, es muy importante, pero no podemos conocer qué tan alejado del verdadero valor de $\theta\ esta\ \widehat{\theta}$ ya que T es una variable aleatoria continua, la probabilidad de que T tome el valor $\hat{\theta}$ es cero
- $P\left(T = \hat{\theta}\right) = P\left(l(X_1, X_2, \dots, X_n) = \hat{\theta}\right) = 0$

ESTIMACIÓN POR INTERVALO

- Por ejemplo, hemos visto que la media muestral $\bar{X}$, resulto ser un buen estimador puntual de la media poblacional μ, por sus propiedades.
- Sabemos que el valor promedio de todas las medias muestrales observables posibles es la media poblacional μ.
- El estadístico $\bar{X}$ varía de una muestra a otra, es decir depende de los elemento seleccionados en la muestra
- Para lograr una estimación más significativa y característica de la población debemos encontrar una estimación por intervalo del parámetro en cuestión, tomando en consideración la distribución de muestreo del estimador. El intervalo obtenido tendrá una "confianza" especificada de estimar correctamente el verdadero parámetro de la población.

ESTIMACIÓN POR INTERVALO

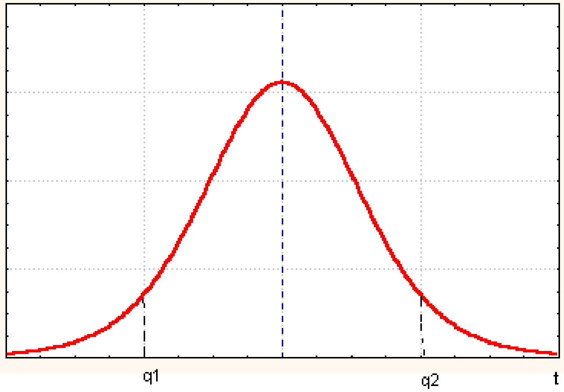
Definición: Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de densidad $f(.,\theta)$

.

Entonces:

$$P\left(T_1 < \theta < T_2\right) = \gamma \qquad (1) \qquad \gamma = 1 - \alpha$$

1. (T_1, T_2) es un "**Intervalo de Confianza"** de $100\gamma\%$ para θ .
2. γ es el coeficiente de confianza o nivel de confianza.
3. T_1 es el extremo inferior del intervalo de confianza y T_2 el extremo superior.
4. El valor observado (t_1, t_2) del intervalo aleatorio (T_1, T_2) es una estimación por intervalo del parámetro



ALGUNAS OBSERVACIONES

1. El intervalo (T_1, T_2) antes de particularizar su valor para una muestra concreta, es un **intervalo aleatorio**, dependiente del vector muestral $X_1, X_2, \ldots, X_n$ fijado el nivel de confianza γ, por lo tanto la expresión **(1) no debe interpretarse** como “la probabilidad de que el parámetro θ tome algún valor entre T_1 y T_2

a) El parámetro θ siempre será desconocido, lo que impide verificar la afirmación anterior.

b) En la expresión $P\left(T_1 < \theta < T_2\right) = \gamma$ las variables aleatorias son T_1 y T_2 y no el parámetro, el cual es una valor desconocido pero fijo.

ALGUNAS OBSERVACIONES

- La interpretación correcta de la expresión **(1)** es que γ es la probabilidad de que el intervalo aleatorio (T_1, T_2) ***incluya*** el verdadero valor del parámetro antes de extraer la muestra, es decir,
  $$P\{(T_1, T_2) incluya\, \theta\} = \gamma$$
- Una vez seleccionada la muestra ,$X_1, X_2, \ldots, X_n$ la probabilidad de que el parámetro θ esté incluido en el intervalo (t_1, t_2)es 1 ó 0, dependiendo de que el parámetro esté o no esté entre los dos números $t_1\ y\, t_2$ en que se convierten los extremos del intervalo al ser calculados para una valor observado de la muestra. Es por eso que hablamos de un **Nivel de Confianza .**

ESTIMACION POR INTERVALO

Si tenemos una muestra y una estadística

$$X_1, X_2, \ldots X_n \overset{IID.}{\sim} f_X(x;\theta)$$

$$P\left(\hat{\Theta}_L < \theta < \hat{\Theta}_U\right) = 1-\alpha \quad para\, 0 < \alpha < 1$$

Para una muestra particular y una distribución de muestreo particular obtenemos

$$\hat{\theta}_L < \theta < \hat{\theta}_U$$

INTERVALOS DE CONFIANZA PARA POBLACIONES NORMALES

$$X \sim N(\mu, \sigma^2)$$

- Entonces suponemos una muestra
-

$$X_1, X_2, \ldots.., X_n \sim N(\mu, \sigma^2)$$

$$\overline{X} \sim N(\mu_{\overline{X}} = \mu; \quad \sigma^2_{\overline{X}} = \sigma^2 / n)$$

1-POBLACIÓN NORMAL- INTERVALO DE CONFIANZA PARA LA MEDIA con varianza (σ^2) conocida

Consideremos la relación

$$Q = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

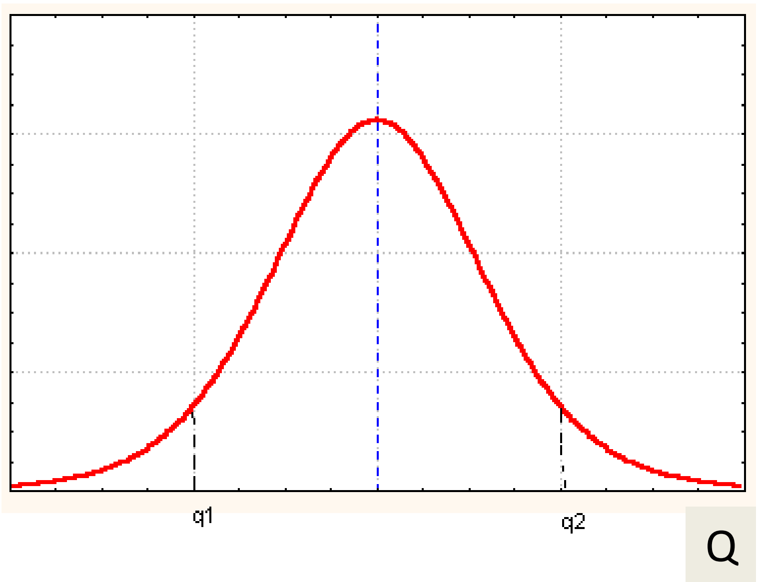
Se plantea $P(q_1 < Q < q_2) = \gamma$,

$$P\left(q_1 < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < q_2 \right) = \gamma$$

$$P\left(q_1 < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < q_2 \right) = \gamma \quad \Rightarrow \quad P\left(q_1 \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < q_2 \frac{\sigma}{\sqrt{n}} \right) = \gamma \quad \Rightarrow$$

$$\Rightarrow \quad P\left(-\overline{X} + q_1 \frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + q_2 \frac{\sigma}{\sqrt{n}} \right) = \gamma$$

$$\Rightarrow \quad P\left(\overline{X} - q_1 \frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - q_2 \frac{\sigma}{\sqrt{n}} \right) = \gamma$$



$$P\left(\overline{X} - q_2 \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} - q_1 \frac{\sigma}{\sqrt{n}} \right) = \gamma$$

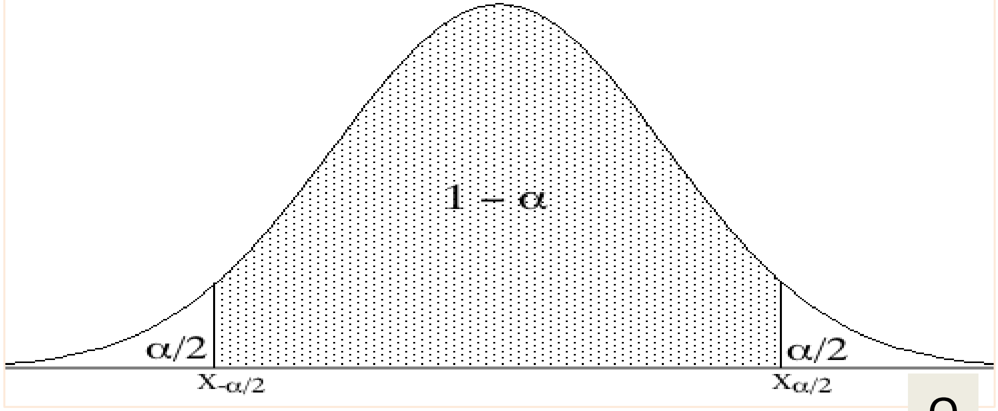
1-POBLACIÓN NORMAL-INTERVALO DE CONFIANZA PARA LA MEDIA, con varianza (σ^2) conocida

- $$P\left(\overline{X} - q_2\,\frac{\sigma}{\sqrt{n}} \langle\, \mu \,\langle\, \overline{X} - q_1\,\frac{\sigma}{\sqrt{n}}\right) = \gamma$$
- *q1* y *q2* son cuantiles de la distribución N(0,1)
- y el orden de cada uno es $q_1 = z_{\frac{1-\gamma}{2}} \quad y \quad q_2 = z_{\frac{1+\gamma}{2}}$
- Se puede probar que para la distribución Normal Estándar el intervalo que tiene longitud mínima es aquel en el cual $q_1 = -q_2$

$$\left(\overline{X} - q_2\,\frac{\sigma}{\sqrt{n}}\,;\,\overline{X} - q_1\,\frac{\sigma}{\sqrt{n}}\right)$$

- Intervalo de confianza aleatorio para μ es con un nivel de confianza γ%. Lo que significa que para todas las muestras de tamaño "*n*" fijo, el γ% de los intervalos obtenidos a partir de ellas, contiene el valor real del parámetro y el (1-γ)% restante, no lo contiene.

1-POBLACIÓN NORMAL- INTERVALO DE CONFIANZA PARA LA MEDIA con varianza (σ^2) conocida



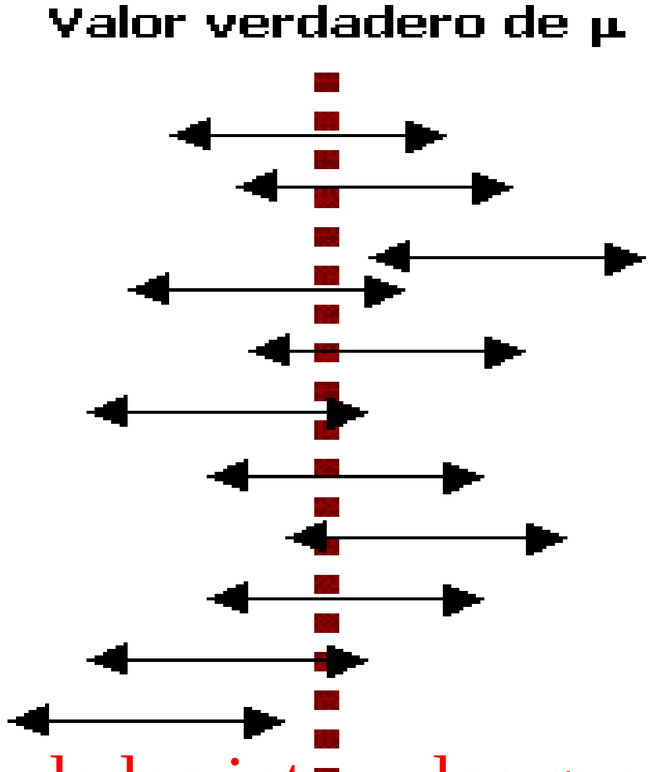
Q

$$P(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

Sí $\overline{x}$ es la media observada de una m.a. de tamañt n de una población con, σ^2conocida, un intervalo de confianza observado para μ está dado por :

$$\overline{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; (1-\alpha)$$

Interpretación de Intervalos de confianza.
Intervalos para distintas muestras del mismo tamaño



¿El centro de los intervalos que valor tiene?

¿Todos los intervalos contienen a μ?

¿La precisión de un intervalo de quien depende?

1-POBLACIÓN NORMAL- INTERVALO DE CONFIANZA PARA LA MEDIA con varianza (σ^2) conocida
Longitud del Intervalo

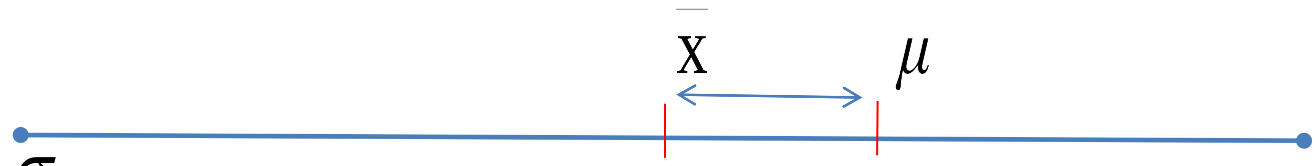
$$\left(\overline{X} - q_2 \frac{\sigma}{\sqrt{n}} ; \overline{X} - q_1 \frac{\sigma}{\sqrt{n}} \right)$$

$$L = \overline{X} - q_1 \frac{\sigma}{\sqrt{n}} - \left(\overline{X} - q_2 \frac{\sigma}{\sqrt{n}} \right)$$

$$L = \overline{X} - q_1 \frac{\sigma}{\sqrt{n}} - \overline{X} + q_2 \frac{\sigma}{\sqrt{n}}$$

$$\boxed{L = \frac{\sigma}{\sqrt{n}} \left(q_2 - q_1 \right)}$$

Error de estimación y tamaño de muestra



$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad\qquad \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\left|\bar{x} - \mu\right| \leq e$$

$$\bar{x} - e \leq \mu \leq \bar{x} + e$$

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ;(1-\alpha)$$

$$e = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

ERROR MAX. DE ESTIMACIÓN

$$n = \left(\frac{z_{1-\alpha/2}\sigma}{e}\right)^2$$

TAMAÑO DE MUESTRA

Importante!!

Para estudiar, comprender y resolver problemas de estimación debemos tener en cuenta:

- ✓ Distribución de la población de la cual proviene la muestra en estudio (normales, no normales o desconocidas).
- ✓ Conocimiento de otros parámetros poblacionales. Por ejemplo, si se quiere estimar un intervalo de confianza para la media de la población, se debe tener en cuenta si se conoce o no la desviación estándar de la población en estudio.
- ✓ Tamaño de la muestra seleccionada (muestras pequeñas o grandes).

2-POBLACIÓN NORMAL- INTERVALO DE CONFIANZA PARA LA MEDIA - varianza σ^2 desconocida

- La relación a considerar es

$$Q = \frac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{(n-1)}$$

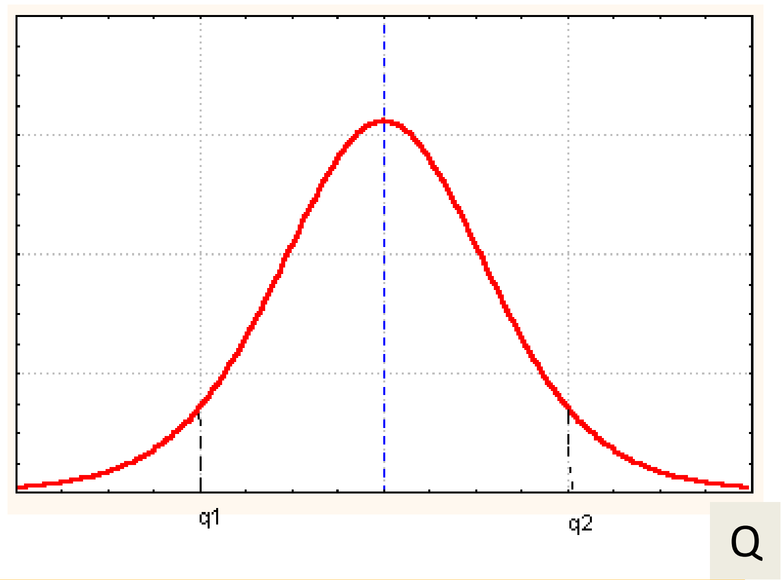
- Se plantea $P\left(q_1 \langle Q \langle q_2\right) = \gamma$

$$P\left(q_1 \langle \frac{\overline{X} - \mu}{S / \sqrt{n}} \langle q_2\right) = \gamma$$

$$P\left(q_1 \langle \frac{\overline{X} - \mu}{S / \sqrt{n}} \langle q_2\right) = \gamma \quad \Rightarrow \quad P\left(q_1 \frac{S}{\sqrt{n}} \langle \overline{X} - \mu \langle q_2 \frac{S}{\sqrt{n}}\right) = \gamma \quad \Rightarrow$$

$$\Rightarrow \quad P\left(-\overline{X} + q_1 \frac{S}{\sqrt{n}} \langle -\mu \langle -\overline{X} + q_2 \frac{S}{\sqrt{n}}\right) = \gamma$$

$$\Rightarrow \quad P\left(\overline{X} - q_1 \frac{S}{\sqrt{n}} \rangle \mu \rangle \overline{X} - q_2 \frac{S}{\sqrt{n}}\right) = \gamma$$



$$P\left(\overline{X} - q_2 \frac{S}{\sqrt{n}} \langle \mu \langle \overline{X} - q_1 \frac{S}{\sqrt{n}}\right) = \gamma$$

2-POBLACIÓN NORMAL- INTERVALO DE CONFIANZA PARA LA MEDIA - varianza σ^2 desconocida

$$P\left(\overline{X} - q_2\frac{S}{\sqrt{n}} \langle \mu \langle \overline{X} - q_1\frac{S}{\sqrt{n}}\right) = \gamma$$

q1 y *q2* son cuantiles de la distribución *t*-Student con (*n-1*) g l., y el orden de cada uno es

$$q_1 = t_{n-1,\frac{1-\gamma}{2}} \quad y \quad q_2 = t_{n-1,\frac{1+\gamma}{2}}$$

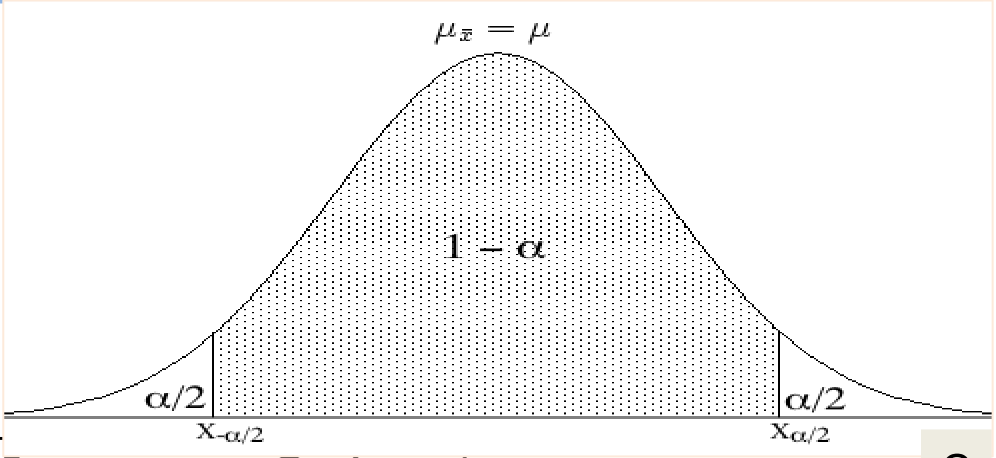
Se puede probar que para la distribución t- Student con (*n-1*) grados de libertad, el intervalo que tiene longitud mínima es aquel en el cual $q_1 = -q_2$

$$\left(\overline{X} - q_2\frac{S}{\sqrt{n}}\,;\,\overline{X} - q_1\frac{S}{\sqrt{n}}\right)$$

El intervalo de confianza aleatorio obtenido para μ con un nivel de confianza γ%. Lo que significa que para todas las muestras de tamaño "*n*" fijo, el γ% de los intervalos obtenidos a partir de ellas, contiene el valor real del parámetro y el (1-γ)% restante, no lo contiene. .

2-POBLACIÓN NORMAL- INTERVALO DE CONFIANZA PARA LA MEDIA- varianza σ^2 desconocida

- Intervalo aleatorio:



$$P(\overline{X} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}}) = 1-\alpha$$

Sí $\overline{x}$ es la media de una m.a. de tamaño n de una población con, σ^2desconocida, un intervalo de confianza observado para μ está dado por;

$$\overline{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} ; (1-\alpha)$$

2-POBLACIÓN NORMAL- INTERVALO DE CONFIANZA PARA LA MEDIA - varianza σ^2 desconocida
Longitud del Intervalo

$$\left(\overline{X} - q_2 \frac{S}{\sqrt{n}} \, ; \, \overline{X} - q_1 \frac{S}{\sqrt{n}} \right)$$

$$L = \overline{X} - q_1 \frac{S}{\sqrt{n}} - \left(\overline{X} - q_2 \frac{S}{\sqrt{n}} \right)$$

$$L = \overline{X} - q_1 \frac{S}{\sqrt{n}} - \overline{X} + q_2 \frac{S}{\sqrt{n}}$$

$$L = \frac{S}{\sqrt{n}} \left(q_2 - q_1 \right)$$

Intervalos con R

- Intervalo de confianza para la media de población normal y varianza desconocida
- >x=c(……….)
- >t.test(x,conf.level=0.99)
- > t.test(x, alternative=“greater”, conf.level=0.99)
- > t.test(x, alternative=“less”, conf.level=0.99

5.6.

Una máquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de nueve piezas y los diámetros medidos son 1,01; 0,97; 1,03; 1,04; 0,99; 0,98; 0,99; 1,01 y 1,03 centímetros. Se sabe que el diámetro de las piezas producidas por la máquina se distribuye normalmente.

a) Estime un intervalo que contenga al diámetro medio de la población de piezas producidas por la máquina, con un nivel de confianza del 99%. Interprete el resultado.

b) Realice una nueva estimación con un nivel de confianza menor, compare los resultados y concluya contrastando los conceptos de precisión y confianza.

5.6.

Una máquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de nueve piezas y los diámetros medidos son 1,01; 0,97; 1,03; 1,04; 0,99; 0,98; 0,99; 1,01 y 1,03 centímetros. Se sabe que el diámetro de las piezas producidas por la máquina se distribuye normalmente.

$$X : \text{Diámetro de las piezas} \quad X \sim N(\mu, \sigma)$$

$$\overline{X} : \text{Díametro medio de las piezas}$$

$$\sigma^2 \text{ DESCONOCIDA} \quad n = 9$$

$$\overline{x} = 1{,}006 \quad \text{y} \quad s = 0{,}025 \rightarrow \text{obtenidos con calculadora}$$

a) Estime un intervalo que contenga al diámetro medio de la población de piezas producidas por la máquina, con un nivel de confianza del 99%. Interprete el resultado.

$$Q \sim t - Student(gl = 8)$$

$$\overline{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}};(1-\alpha)$$

$$1{,}006 - 3{,}355\frac{0{,}025}{\sqrt{9}} < \mu < 1{,}006 + 3{,}355\frac{0{,}025}{\sqrt{9}};(0{,}99)$$

$$(0{,}9780;1{,}034)$$

Tenemos un 99 % de confianza a partir de los datos de muestra que el intervalo (0,9789;1,033) contiene al verdadero promedio del diámetro de las piezas

b) Realice una nueva estimación con un nivel de confianza menor, compare los resultados y concluya contrastando los conceptos de precisión y confianza.

$$Q \rightarrow t - Student(gl = 8)$$

$$\bar{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}};(1-\alpha)$$

$$1{,}006 - 1{,}860\frac{0{,}025}{\sqrt{9}} < \mu < 1{,}006 + 1{,}860\frac{0{,}025}{\sqrt{9}};(0{,}90)$$

$$(0{,}9905;1{,}0215)$$

Tenemos un 90 % de confianza a partir de los datos de muestra que el intervalo (0,9905,10215) contiene al verdadero promedio del diámetro de las piezas.

¿Qué sucede cuando aumentamos la confianza?

$$\left.\begin{array}{l} a)(0{,}9780;1{,}0340) \rightarrow 99\% \\ b)\ (0{,}9905;1{,}0215) \rightarrow 90\% \end{array}\right\}$$

A mayor confianza

Menor precisión

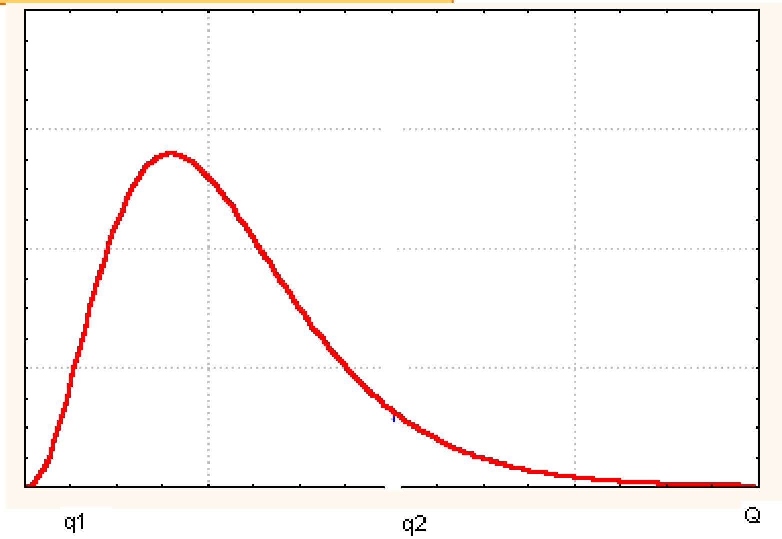
3-POBLACIÓN NORMAL
Intervalo de confianza para la varianza

- La relación a considerar es

$$Q = \frac{S^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$$

Se plantea $P(q_1 \langle Q \langle q_2) = \gamma$

$$P\left(q_1 \langle \frac{S^2(n-1)}{\sigma^2} \langle q_2\right) = \gamma$$



$$P\left(q_1 \langle \frac{S^2(n-1)}{\sigma^2} \langle q_2\right) = \gamma \quad \Rightarrow \quad P\left(\frac{1}{q_1} \rangle \frac{\sigma^2}{S^2(n-1)} \rangle \frac{1}{q_2}\right) = \gamma \quad \Rightarrow$$

$$\Rightarrow \quad P\left(\frac{S^2(n-1)}{q_2} \langle \sigma^2 \langle \frac{S^2(n-1)}{q_1}\right) = \gamma$$

$$P\left(\frac{S^2(n-1)}{q_2} \langle \sigma^2 \langle \frac{S^2(n-1)}{q_1}\right) = \gamma$$

3-POBLACIÓN NORMAL

Intervalo de confianza para la varianza

$$P\left(\frac{S^2(n-1)}{q_2} \langle \sigma^2 \langle \frac{S^2(n-1)}{q_1}\right) = \gamma$$

- *q1* y *q2* son cuantiles de la distribución Chi-cuadrada con (*n-1*) grados de libertad, y el orden de cada uno es

$$q_1 = \chi^2_{n-1,\frac{1-\gamma}{2}} \quad y \quad q_2 = \chi^2_{n-1,\frac{1+\gamma}{2}}$$

- En este caso hemos optado por considerar a γ como el área central quedando de este modo el área de las colas simétrica.
- El intervalo de confianza aleatorio que hemos obtenido para σ2 es
-
- $$\left(\frac{S^2(n-1)}{q_2};\frac{S^2(n-1)}{q_1}\right)$$ con un nivel de confianza γ%.

3-POBLACIÓN NORMAL
Intervalo de confianza para la varianza

$$P\left(\frac{S^2(n-1)}{q_2} \langle \sigma^2 \langle \frac{S^2(n-1)}{q_1}\right) = \gamma$$

- donde los valores *q1* y *q2* son cuantiles de la distribución Chi-cuadrada con (*n-1*) grados de libertad, y el orden de cada uno es

$$q_1 = \chi^2_{n-1,\frac{1-\gamma}{2}} \quad y \quad q_2 = \chi^2_{n-1,\frac{1+\gamma}{2}}$$

- Aquí hemos optado por considerar a γ como el área central quedando de este modo el área de las colas simétrica.
- El intervalo de confianza aleatorio que hemos obtenido para $\sigma 2$ es
-

$$\left(\frac{S^2(n-1)}{q_2}; \frac{S^2(n-1)}{q_1}\right)$$

- con un nivel de confianza γ%.

3-POBLACIÓN NORMAL
Intervalo de confianza para la varianza
LONGITUD DEL INTERVALO

$$\left(\frac{S^2(n-1)}{q_2} ; \frac{S^2(n-1)}{q_1} \right)$$

$$L = \frac{S^2(n-1)}{q_1} - \frac{S^2(n-1)}{q_2}$$

$$L = S^2(n-1)\left(\frac{1}{q_1} - \frac{1}{q_2} \right)$$

3-POBLACIÓN NORMAL
Intervalo de confianza para la varianza

- *Intervalo de confianza para una población normal*
- *(a,b)*
- *x=c(……)*
- *S2=var(x)*
- *n=length(x)*
- *q1=qchisq((1-confianza)/2,n-1)*
- *q2=qchisq((1+confianza)/2,n-1)*
- *a=(n-1)*s2/q2*
- *b=(n-1)*s2/q1*

4-POBLACIÓNES NO NORMALES
INTERVALO DE CONFIANZA PARA LA MEDIA

- Cuando la población de donde se extrae la información muestral no tiene un comportamiento normal, para la construcción de intervalos de confianza, si el tamaño de muestra es grande, se puede hacer uso de las propiedades asintóticas de los estimadores
- Supongamos que tenemos $X \sim f(.,\theta)$ en la población siendo $E(X)=\theta$ y tomamos una m.a $X_1, X_2, \ldots, X_n$ de esta población.
- La variable $Z_n = \dfrac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\operatorname{var}(\bar{X}_n)}}$ sabemos por el T.L.C $f_{Z_n} \xrightarrow[n\to\infty]{} N(0,1)$
-
- Podemos tener en cuenta este hecho para construir Intervalos de Confianza a partir de la siguiente relación

.

$$Q = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\widehat{\operatorname{var}}(\bar{X}_n)}} \xrightarrow[aprox.]{} N(0,1)$$

4-POBLACIÓNES NO NORMALES
INTERVALO DE CONFIANZA PARA LA MEDIA
EJEMPLO: POBLACIÓN BERNOULLI

- $X \sim Bernoulli(p) \qquad E(X) = p$
-

$$E(\overline{X}) = p \qquad \operatorname{var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$$

- como no conocemos el parámetro " p " debemos estimar la varianza resultando

$$\hat{\operatorname{var}}(\overline{X}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{\overline{X}(1-\overline{X})}{n}$$

$$Q = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{\hat{\operatorname{var}}(\overline{X}_n)}} \xrightarrow[aprox.]{} N(0,1)$$

$$\boxed{Q = \frac{\overline{X}_n - p}{\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}}} \to N(0,1)}$$

4-POBLACIÓNES NO NORMALES
INTERVALO DE CONFIANZA PARA LA MEDIA
EJEMPLO: POBLACIÓN BERNOULLI

- Si $X_1, X_2, \ldots\ldots, X_n \sim Bernoulli(p)$ $\qquad \hat{P} = \dfrac{\sum_1 X_i}{n} = \overline{X}$

$$Q = \frac{\overline{X}_n - p}{\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}}} \to N(0,1)$$

- $P\left(q_1 \langle\, Q \,\langle\, q_2\right) = \gamma$

$$P(q_1 < \frac{\overline{X}_n - p}{\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}}} < q_2) = \gamma$$

4-POBLACIÓNES NO NORMALES
INTERVALO DE CONFIANZA PARA LA MEDIA
EJEMPLO: POBLACIÓN BERNOULLI

- Luego trabajando de la misma forma que en los ejemplos anteriores el intervalo aleatorio para la media será:

$$\left(\overline{X} - q_2\sqrt{\frac{\overline{X}(1-\overline{X})}{n}} ; \overline{X} - q_1\sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \right)$$

- Recuerde que ambos cuantiles se obtendrán de la distribución Normal estándar , en particular

$$q_1 = z_{\frac{1-\gamma}{2}} \quad y \quad q_2 = z_{\frac{1+\gamma}{2}}$$

- Como quedaría un intervalo observado ?

4-POBLACIÓNES NO NORMALES
INTERVALO DE CONFIANZA PARA LA MEDIA
EJEMPLO: POBLACIÓN POISSON

$$E(X) = p$$

$$E(\overline{X}) = p$$

$$\operatorname{var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$$

$$p$$

$$\hat{\operatorname{var}}(\overline{X}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{\overline{X}(1-\overline{X})}{n}$$

$$Q = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{\hat{\operatorname{var}}(\overline{X}_n)}} \xrightarrow[aprox.]{} N(0,1)$$

$$Q = \frac{\overline{X}_n - p}{\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}}} \to N(0,1)$$

:

- como no conocemos el parámetro " " debemos estimar la varianza resultando

4-POBLACIÓNES NO NORMALES
INTERVALO DE CONFIANZA PARA LA MEDIA
EJEMPLO: POBLACIÓN POISSON

- $X \sim Poisson(\lambda) \quad E(X) = \lambda$
- $E(\bar{X}) = \lambda \quad Var(\bar{X}) = \dfrac{\sigma^2}{n} = \dfrac{\lambda}{n}$
- $V\hat{a}r(\bar{X}) = \dfrac{\hat{\lambda}}{n} = \dfrac{\bar{X}}{n}$

$$Q = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\hat{var}(\bar{X}_n)}} \xrightarrow[aprox.]{} N(0,1)$$

$$Q = \frac{\bar{X} - \lambda}{\sqrt{\frac{\bar{X}}{n}}} \xrightarrow[n\to\infty]{} \text{N}(0,1)$$

$$\left(\bar{X} - q_2 \frac{\bar{X}}{n};\ \bar{X} - q_1 \frac{\bar{X}}{n}\right),$$

q_1 y q_2 *cuantiles de la distribución normal*

5.34.

En una muestra aleatoria de 500 clientes de una empresa local se encontró que 340 se habían suscripto al servicio de correo electrónico y 160 al de correo electrónico e Internet.

a) Construir un intervalo de confianza del 95% para la proporción actual de clientes suscriptos al correo electrónico de esa empresa de servicios. Interpretar resultado.

b) ¿Qué puede decir acerca del error cometido en la estimación, con un 95% de confianza?

c) A partir de los resultados obtenidos de la muestra preliminar ¿qué tan grande se requiere que sea la muestra, si se desea tener una confianza del 95% de que la estimación de p estará dentro de 0,02? Interpretar resultado.

d) Asumiendo que no se dispone de una muestra preliminar para obtener información acerca de p, ¿qué tan grande se requiere que sea la muestra si se quiere tener una confianza al menos del 95% de que la estimación de p estará dentro de 0,02? Interpretar resultado.

5.34.

En una muestra aleatoria de 500 clientes de una empresa local se encontró que 340 se habían suscripto al servicio de correo electrónico y 160 al de correo electrónico e Internet.

a) Construir un intervalo de confianza del 95% para la proporción actual de clientes suscriptos al correo electrónico de esa empresa de servicios. Interpretar resultado.

X:Cantidad de clientes suscriptos X~Bernulli (p)

$\widehat{P}$: proporción de clientes suscriptos al correo electrónico.

$$\hat{p} = \frac{340}{500} = 0,68$$

$$\widehat{P} \to N(\mu_{\hat{P}} = p, \sigma_{\hat{P}} = \sqrt{\frac{pq}{n}})$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}; (1-\alpha)$$

$$0,68 - 1,96\sqrt{\frac{0,68\,0,32}{500}} < p < 0,68 + 1,96\sqrt{\frac{0,68\,0,32}{500}}; (0,95)$$

$$(0,6391;0,7209)$$

Tenemos un 95 % de confianza a partir de los datos de muestra que el intervalo (0,6391;0,7209) contiene a la verdadera proporción de clientes suscriptos al correo Electrónico.

Error de estimación

b) ¿Qué puede decir acerca del error cometido en la estimación, con un 95% de confianza?

$$e = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$e = 1{,}96\sqrt{\frac{0{,}68.0{,}32}{500}} = 0{,}04$$

Se tiene un 95% de confianza de que el error máximo de la estimación de la verdadera proporción es del 4 %

c) A partir de los resultados obtenidos de la muestra preliminar ¿qué tan grande se requiere que sea la muestra, si se desea tener una confianza del 95% de que la estimación de p estará dentro de 0,02? Interpretar resultado.

Si el error de estimación debe ser menor a 0,02

$$z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < e$$

$$1{,}96\sqrt{\frac{0{,}68.0{,}32}{n}} < 0{,}02$$

$$n > \frac{1{,}96^2 0{,}68.0{,}32}{0{,}02^2} > 2089{,}83$$

$$n \geq 2090$$

La muestra deberá de ser de 2090* clientes *o superior para tener una confianza de al menor el 95% de que la estimación de la verdadera proporción estará dentro del 2% de error.

Población binomial. Valor de $\hat{p}$ cuando se desconoce totalmente

- d) Asumiendo que no se dispone de una muestra preliminar para obtener información acerca de p ¿qué tan grande se requiere que sea la muestra si se quiere tener una confianza al menos del 95% de que la estimación de p estará dentro del 0,02?

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} < e \rightarrow z^2\ \frac{\hat{p}.\hat{q}}{n} < e^2\ \rightarrow \quad z^2\ \frac{\hat{p}.\hat{q}}{e^2} < n$$

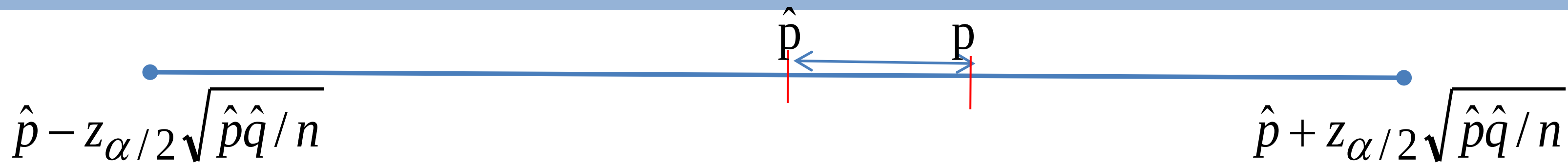
Si $\hat{p} = 0{,}5$

$$n > \frac{z^2}{4.e^2} \qquad n > \frac{1{,}96^2}{4.0{,}02^2} = 2401$$

$$n \geq 2401$$

La muestra deberá de ser 2401 clientes o superior para tener una confianza de al menor el 95% de que la estimación de la verdadera proporción estará dentro del 2% de error.

Error de estimación y tamaño de muestra



$$|\hat{p} - p| \leq e$$

$$\hat{p} - e \leq p \leq \hat{p} + e$$

$$\hat{p} - z_{\alpha/2}\,\hat{p}\hat{q}/n < \mu < \hat{p} + z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n};(1-\alpha)$$

$$e = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$$

ERROR DE ESTIMACIÓN

$$n = \frac{z^2{}_{\alpha/2}\,\hat{p}\hat{q}}{e^2}$$

TAMAÑO DE MUESTRA

$$n = \frac{z^2{}_{\alpha/2}}{4e^2}$$

Si desconocemos p

5.3. *

Se estudia la cantidad de pintura que contienen los envases de un litro compradas a un conocido fabricante local. Por las especificaciones del productor se sabe que la desviación estándar de la cantidad de pintura es igual a 0,02 litros. Al seleccionar una muestra de 50 latas de las mismas características, el contenido promedio de pintura resultó ser igual a 0,995 litros.

a) ¿Es aceptable suponer que el contenido de pintura promedio en las muestras se distribuye normalmente?
b) Estimar el contenido promedio real de la población de pintura en latas de un litro, mediante un intervalo de confianza del 99 %.
c) Con base en estos resultados ¿es posible afirmar que, en promedio, el productor envasa menos de un litro en tales latas? ¿Por qué?
d) ¿Con qué nivel de confianza puede decirse que el contenido medio de pintura de las latas de un litros es de $0{,}996 \pm 0{,}005$?

5.3. *

Se estudia la cantidad de pintura que contienen los envases de un litro compradas a un conocido fabricante local. Por las especificaciones del productor se sabe que la desviación estándar de la cantidad de pintura es igual a 0,02 litros. Al seleccionar una muestra de 50 latas de las mismas características, el contenido promedio de pintura resultó ser igual a 0,995 litros.

a) ¿Es aceptable suponer que el contenido de pintura promedio en las muestras se distribuye normalmente?

X : Cantidad de pintura que contienen los embases de un litro

$$X \sim \text{desconocida}(\sigma = 0{,}02)$$

$\overline{X}$: contenido promedio de las latas de pintura de un litro

$$n = 50 \quad \overline{x} = 0{,}995$$

SI POR EL TEOREMA DEL LIMITE CENTRAL

b) Estimar el contenido promedio real de la población de pintura en latas de un litro, mediante un intervalo de confianza del 99 %.

$$\overline{X} \xrightarrow{T.L.C} N(\mu;\frac{\sigma}{\sqrt{n}})$$

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}};(1-\alpha)$$

$$0{,}995 - 2{,}58\frac{0{,}02}{\sqrt{50}} < \mu < 0{,}995 + 2{,}58\frac{0{,}02}{\sqrt{50}};(0{,}99)$$

$$\left(0{,}988;1{,}002\right)$$

Tenemos un 99 % de confianza a partir de los datos de muestra que el intervalo (0,988;1,002) contiene al verdadero promedio del contenido de las latas.

c) Con base en estos resultados ¿es posible afirmar que, en promedio, el productor envasa menos de un litro en tales latas? ¿Por qué?

Como el intervalo contiene a l 1 podemos asegurar con un 99% de confianza que los tarros en promedio tienen un litro de contenido.

d) ¿Con qué nivel de confianza puede decirse que el contenido medio de pintura de las latas de un litros es de $0{,}996 \pm 0{,}005$?

$$0{,}996 - 0{,}005 < \mu < 0{,}996 - 0{,}005 \xrightarrow{entonces} (0{,}9910; 1{,}001)$$

$$e = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad e = 0{,}005 \xrightarrow{entonces} 0{,}005 = z_{\alpha/2}\frac{0{,}02}{\sqrt{50}}$$

$$z_{\alpha/2} = \frac{0{,}05.\sqrt{50}}{0{,}02} \xrightarrow{entonces} z_{\alpha/2} = 1{,}77$$

$$(1-\alpha) \approx 0{,}92$$

5.4. *

Una empresa eléctrica fabrica focos cuya duración está distribuida normalmente, con desviación estándar de 40 horas. Se ensaya una muestra de 30 focos y se obtiene una media de 780 horas y desviación estándar de 39 horas.

a) Estime la duración media de la población de todos los focos que produce la empresa. Interprete el resultado.

b) ¿Cuántos focos se debe ensayar al nivel de confianza utilizado, si se desea que la estimación puntual esté dentro de las 10 horas de la media real? Interprete el valor numérico obtenido.

5.4. *

Una empresa eléctrica fabrica focos cuya duración está distribuida normalmente, con desviación estándar de 40 horas. Se ensaya una muestra de 30 focos y se obtiene una media de 780 horas y desviación estándar de 39 horas.

a) Estime la duración media de la población de todos los focos que produce la empresa. Interprete el resultado.

$$\mathrm{X}: \text{Duración de los focos}$$

$$\mathrm{X} \sim Normal(\mu; \sigma = 40) \rightarrow \overline{\mathrm{X}} \sim Normal(\mu; \sigma = \frac{40}{\sqrt{30}})$$

$$\overline{X}: \text{duración promedio de los focos}$$

$$\mathrm{n} = 30 \quad \overline{\mathrm{x}} = 780 \quad \mathrm{s} = 39$$

a) Estime la duración media de la población de todos los focos que produce la empresa. Interprete el resultado.

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}};(1-\alpha)$$

$$780 - 1{,}64\frac{40}{\sqrt{30}} < \mu < 780 + 1{,}64\frac{40}{\sqrt{30}};(0{,}90)$$

$$(768{,}02;791{,}98)$$

Tenemos un 95 % de confianza a partir de los datos de muestra que el intervalo (768,02;791.98) contiene al verdadero promedio de la duración de los focos

5.23.

Un fabricante de baterías para automóvil afirma que sus baterías durarán, en promedio, tres años con una varianza de un año. Se prueban cinco de tales baterías y arrojan una duración de 1,9; 2,4; 3,0; 3,5 y 4,2 años. ¿Es válida la afirmación del fabricante en cuanto a la *dispersión* de los tiempos de duración de sus baterías? Realice los cálculos necesarios para justificar su respuesta. Suponga que la duración de las baterías se distribuye normalmente.

$$\text{X : Vida de una bateria} \quad X \sim N(\mu, \sigma)$$

$$n = 5 \quad y \quad s \; = 0{,}9028$$

$$donde \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{\nu=4}$$

$$1 - \alpha = 0{,}95 \Rightarrow \chi^2_{0{,}025} = 0.484 \;\; y \;\; \chi^2_{0{,}975} = 11.14$$

$$\frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} ; 1-\alpha$$

$$\frac{4.0{,}9028^2}{11{,}143} < \sigma^2 < \frac{4.0{,}9028^2}{0{,}484} ; 0{,}95$$

$$(0{,}2926; 6{,}7360)$$

En base a la evidencia muestral podemos decir que es valida la afirmación del comerciante, debido a que el intervalo hallado incluye el valor que asegura el fabricante.

REPASO

1Estimación de la media de una población

Media muestral: $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

	Población normal, cualquier n	Población no necesariamente normal, o desconocida, n≥30 (TLC)
Varianza conocida	$\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$	$\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} -> N(0,1)$
Varianza desconocida	$\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$	$\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} -> N(0,1)$

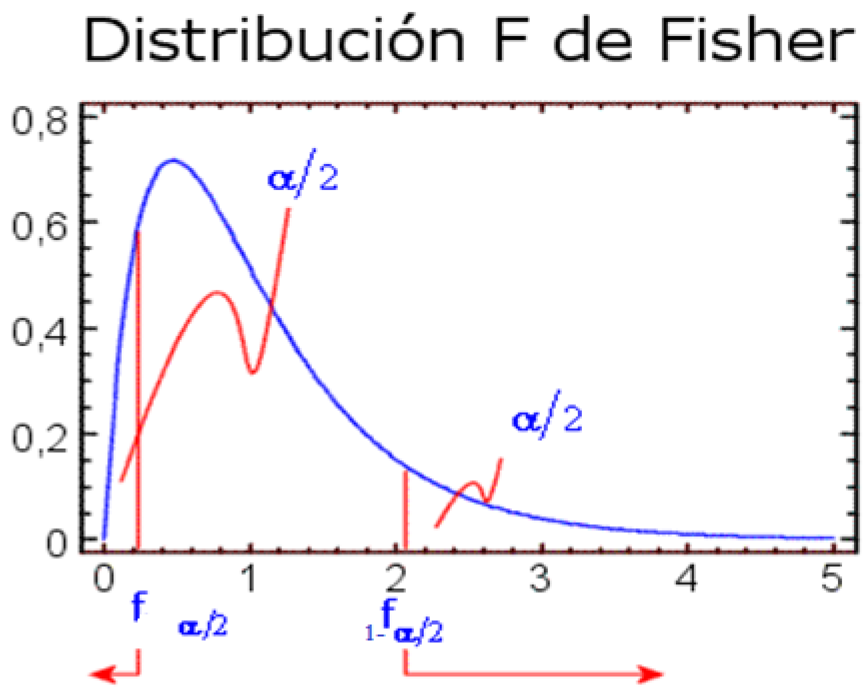
ESTIMACIÓN POR INTERVALOS

PARA DOS POBLACIONES INDEPENDIENTES

POBLACIONES NORMALES
ESTIMACION POR INTERVALOS PARA EL COCIENTE DE VARIANZAS

$$donde \quad F=\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim f_{(\upsilon_1=n_1-1\ y,\ \upsilon_2=n_2-1)}$$

$$P(f_{\alpha/2(\upsilon_1,\upsilon_2)} < F < f_{1-\alpha/2(\upsilon_1,\upsilon_2)}) = 1-\alpha$$



POBLACIONES NORMALES
ESTIMACION POR INTERVALOS PARA EL COCIENTE DE VARIANZAS

$$P\left(f_{\alpha/2(\upsilon_1,\upsilon_2)} < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{1-\alpha/2(\upsilon_1,\upsilon_2)} \right) = 1-\alpha$$

$$P\left(\frac{1}{f_{\alpha/2(\upsilon_1,\upsilon_2)}} > \frac{\sigma_1^2 S_2^2}{\sigma_2^2 S_1^2} > \frac{1}{f_{1-\alpha/2(\upsilon_1,\upsilon_2)}} \right) = 1-\alpha$$

$$P\left(\frac{1}{f_{1-\alpha/2(\upsilon_1,\upsilon_2)}} < \frac{\sigma_1^2 S_2^2}{\sigma_2^2 S_1^2} < \frac{1}{f_{\alpha/2(\upsilon_1,\upsilon_2)}} \right) = 1-\alpha$$

Dos poblaciones normales Estimación por intervalos para el cociente de varianzas.

- Intervalo aleatorio

$$P\left(\frac{1}{f_{1-\alpha/2(\upsilon_1,\upsilon_2)}}\frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_{\alpha/2(\upsilon_1,\upsilon_2)}}\frac{S_1^2}{S_2^2}\right) = 1-\alpha$$

RECORDEMOS LA PROPIEDAD

$$f_{1-\alpha/2(\upsilon_1,\upsilon_2)} = \frac{1}{f_{\alpha/2(\upsilon_2,\upsilon_1)}}$$

POBLACIONES NORMALES
ESTIMACION PARA EL COCIENTE DE VARIANZAS

Si s_1^2 y s_2^2 son las varianzas observadas de muestras aleatorias independientes de tamaño n_1 y n_2, respectivamente, de poblaciones normales, entonces un intervalo de confianza observado, obtenido con una confianza del $(1 - \alpha)100\%$ para σ_1^2/σ_2^2 es:

$$\left(\frac{1}{f_{1-\alpha/2(\upsilon_1,\upsilon_2)}}\frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_{\alpha/2(\upsilon_1,\upsilon_2)}}\frac{s_1^2}{s_2^2}\right);1-\alpha$$

POBLACIONES NORMALES
ESTIMACION PARA EL COCIENTE DE VARIANZAS

- x=c(…….)
- y=c(……)
- var.test (x,y, conf.level=0.98)
- Nivel de confianza del 95% lo hace por default

5.14.

En un experimento se comparó el consumo de combustible para dos tipos de camiones con motor diesel, en condiciones similares. Doce vehículos del Tipo A y diez del Tipo B se probaron a 90 km/h. Si los doce camiones promediaron 16 km/l con desviación 1 km/l y los diez camiones consumieron 11 km/l con desviación 1,8 km/l.

a) Construir un intervalo de confianza del 98 % para el cociente de las varianzas de los recorridos por litro de los vehículos Tipo A y B. Interprete el resultado en su contexto.

b) Determinar un intervalo de confianza del 90 % para la diferencia entre los rendimientos medios de cada tipo de camión, suponiendo que los rendimientos tienen distribución normal y varianzas iguales. Concluya.

Muestreo de dos poblaciones Normales
Distribución de la Diferencia de Medias

1 Caso: Sea Población normal y varianzas conocidas

$$\left.\begin{array}{c} X_1 \sim N(\mu_1, \sigma_1^2) \text{ se toma una muestra de tamaño } n_1 \\ \\ X_2 \sim N(\mu_2, \sigma_2^2) \text{ se toma una muestra de tamaño } n_2 \\ X_1 \text{ Independiente a } X_2 \end{array}\right\} \rightarrow \begin{array}{c} \bar{X}_1 \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n_1}\right) \\ \bar{X}_2 \sim N\left(\mu_2, \dfrac{\sigma_2^2}{n_2}\right) \end{array}$$

$$\left.\begin{array}{c} \bar{X}_1 \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n_1}\right) \\ \bar{X}_2 \sim N\left(\mu_2, \dfrac{\sigma_2^2}{n_2}\right) \end{array}\right\} \rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim Z(0,1)$$

1ER CASO- POBLACIONES NORMALES VARIANZAS CONOCIDAS ESTIMACION POR INTERVALO DE LA DIFERENCIA DE MEDIAS

$$\overline{X}_1 - \overline{X}_2 \sim N(\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2 ; \sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma_1^2 / n_1 + \sigma_2^2 / n_2)$$

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim Z(0,1)$$

$$P(z_1 < Z < z_2) = 1 - \alpha$$

$$P(z_1 < \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n + \sigma_2^2 / n}} < z_2) = 1 - \alpha$$

$$P\left(\left(\overline{X}_1 - \overline{X}_2\right) - z_2 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\overline{X}_1 - \overline{X}_2\right) - z_1 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

1ER CASO- POBLACIONES NORMALES
VARIANZAS CONOCIDAS
ESTIMACION POR INTERVALO DE LA DIFERENCIA DE MEDIAS

Si $\bar{x}_1$ y $\bar{x}_2$ son los valores observados de las medias de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones con varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza observado con una confianza $(1-\alpha)100\%$ para $\mu_1-\mu_2$ es:

$$\left(\left(\bar{x}_1-\bar{x}_2\right)-z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}<\mu_1-\mu_2<\left(\bar{x}_1-\bar{x}_2\right)-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right);(1-\alpha)$$

5.8.

Se realizó un estudio para determinar si un aditivo agregado en la dosificación de hormigones para acelerar el tiempo de fragüe tenía alguna influencia en la resistencia a la compresión del mismo. Una muestra aleatoria de 62 probetas con aditivo ensayadas a los 7 días dan una resistencia media de 70,5 kg/cm^2, con desviación 6,5 kg/cm^2, en tanto que las 60 probetas sin aditivo, ensayadas a la misma edad, dieron una resistencia media de 86,8 kg/cm^2, con desviación 7,2 kg/cm^2. Suponga que de estudios anteriores se conoce la desviación estándar de las poblaciones

y son 5 y 5,5 kg/cm^2 para las probetas con y sin aditivo, respectivamente.

a) A nivel de confianza del 95%, ¿qué puede decir respecto de la influencia del aditivo en la resistencia media del hormigón?
b) ¿Qué conclusión obtendría si desconoce las desviaciones poblacionales, pero las supone iguales?
c) ¿Qué conclusión obtendría si desconoce las desviaciones poblacionales y las supone distintas?

2DO CASO- POBLACIONES NORMALES
VARIANZAS DESCONOCIDAS E IGUALES
ESTIMACION POR INTERVALO DE LA DIFERENCIA DE MEDIAS

Se utilizara la distribución t como en el caso de una sola muestra.

En el caso que se desconozcan σ_1^2 y σ_2^2. Pero se consideran iguales

***Si $\sigma_1^2 = \sigma_2^2 = \sigma^2$**, utilizaremos la variable:*

$$T = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

donde

$$Sp^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

$$P(t_1 < T < t_2) = 1 - \alpha$$

$$P\left(t_1 < \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_2\right) = 1 - \alpha$$

2DO CASO- POBLACIONES NORMALES
VARIANZAS DESCONOCIDAS E IGUALES
ESTIMACION POR INTERVALO DE LA DIFERENCIA DE MEDIAS

Intervalo aleatorio para la diferencia de medias poblacionales:

$$P\left(\left(\overline{X}_1-\overline{X}_2\right)-t_{1-\alpha/2}Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<\left(\overline{X}_1-\overline{X}_2\right)-t_{\alpha/2}Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)=1-\alpha$$

Si $\bar{x}_1$ y $\bar{x}_2$ son las medias observadas de m.a.i. de tamaño n_1 y n_2 de poblaciones normales con varianzas desconocidas σ_1^2 y σ_2^2, y ***Si $\sigma_1^2 = \sigma_2^2$*** *respectivamente, un intervalo de confianza observado obtenido con una confianza del $(1-\alpha)100\%$ para $\mu_1-\mu_2$ es:*

$$\left(\left(\bar{x}_1-\bar{x}_2\right)-t_{1-\alpha/2}sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<\left(\bar{x}_1-\bar{x}_2\right)-t_{\alpha/2}sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right);(1-\alpha)$$

3er CASO- POBLACIONES NORMALES

VARIANZAS DESCONOCIDAS Y DISTINTAS

ESTIMACION POR INTERVALO DE LA DIFERENCIA DE MEDIAS

- *En el caso que se desconozcan σ_1^2 y σ_2^2.* ***Si $\sigma_1^2 \neq \sigma_2^2$****, utilizaremos la variable:*

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_\upsilon$$

$$\upsilon = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left[\left(s_1^2/n_1\right)^2/n_1 - 1 + \left(s_2^2/n_2\right)^2/n_2 - 1\right]}$$

$$P(t_1 < T < t_2) = 1 - \alpha$$

$$P\left(\left(\overline{X}_1 - \overline{X}_2\right) - t_{1-\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\overline{X}_1 - \overline{X}_2\right) - t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = (1 - \alpha)$$

- Intervalo aleatorio de γ%

3er CASO- POBLACIONES NORMALES
VARIANZAS DESCONOCIDAS Y DISTINTAS
ESTIMACION POR INTERVALO DE LA DIFERENCIA DE MEDIAS

- Si $\bar{x}_1$ y $\bar{x}_2$ son la media de dos m.a.i *de tamaño n_1 y n_2,* provenientes de dos poblaciones normales *con varianzas desconocidas σ_1^2 y σ_2^2*, *y* ***Si $\sigma_1^2 \neq \sigma_2^2$***
- *Un intervalo de confianza observado encontrado con (1-α)100% para μ_1-μ_2 es:*

$$\left(\bar{x}_1-\bar{x}_2\right)-t_{1-\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}<\mu_1-\mu_2<\left(\bar{x}_1-\bar{x}_2\right)-t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}};(1-\alpha)$$

$$\upsilon=\left[\frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\left(s_1^2/n_1\right)^2/n_1-1+\left(s_2^2/n_2\right)^2/n_2-1}\right]$$

Intervalo de confianza para la diferencia de media

- Población normal, varianzas desconocidas pero iguales
- t.test(x,y, var.equal=TRUE, conf.level=0.98)
- (var.equal=TRUE) lo hace el software por default

4to CASO POBLACION NO NORMAL O DESCONOCIDA. INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS

Al trabajar con muestras grandes (mayores que 30), nos permitirá aproximar Z a normal por el TLC

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

$$\left(\left(\overline{X}_1 - \overline{X}_2\right) - z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\overline{X}_1 - \overline{X}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right); (1-\alpha)$$

Obtuvimos un intervalo aleatorio, con una confianza de γ%
En el caso que no se conozcan las varianzas poblacionales se estimarán con las varianzas muestrales

4to CASO POBLACION NO NORMAL O DESCONOCIDA. INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS. Ejemplo Bernoulli

- *Interés recae en comparar la proporción de un grupo con otro, en relación con alguna característica en común*
- $\bar{X} - \bar{Y} \xrightarrow[\substack{n\to\infty \\ m\to\infty}]{} \mathrm{N}\left(p_x - p_y, \frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}\right)$
- La varianza es desconocida utilizaremos la estimada
- $Z = \frac{\bar{X}-\bar{Y}-(p_x-p_y)}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}+\frac{\hat{p}_y(1-\hat{p}_y)}{m}}} \xrightarrow[TLC]{} N(0,1)$
- $P\left(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = \gamma = (1-\alpha)$

4to CASO POBLACION NO NORMAL O DESCONOCIDA. INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS. Ejemplo Bernoulli

- $P\left(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = \gamma = (1-\alpha)$

- $P\left(z_{\frac{\alpha}{2}} < \dfrac{\bar{X}-\bar{Y}-(p_x-p_y\ )}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}+\frac{\hat{p}_y(1-\hat{p}_y)}{m}}} < z_{1-\frac{\alpha}{2}}\right) = (1-\alpha)$

- $\left((\bar{X}-\bar{Y}) - z_{1-\frac{\alpha}{2}}.\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}+\frac{\hat{p}_y(1-\hat{p}_y)}{m}};\ (\bar{X}-\bar{Y}) - z_{\frac{\alpha}{2}}.\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}+\frac{\hat{p}_y(1-\hat{p}_y)}{m}}\right); (1-\alpha)$