Estimación por Intervalos

Estadística inferencial

Situación: La variable aleatoria X sigue una distribución paramétrica $X \sim f(x;\theta)$, con $\theta \in \mathbb{R}^k$ desconocido.

Objetivo: Proponer métodos para *estimar* θ.

La clase anterior vimos métodos de estimación puntual: dada una muestra aleatoria de X, vimos algunos métodos para estimar un valor de θ y así obtener una *estimación* $\hat{\theta}$.

Estimación por Intervalos

Cualquier parámetro que quiera estimarse tiene como conjunto de valores posibles un intervalo de n^{o}s reales.
Ejemplos:

1. Tiempo de vida útil (promedio) de cierta lámpara de bajo consumo puede ser cualquier número, digamos, entre 900 y 1100 hrs.
2. Cantidad de clientes que utilizan cierto cajero automático entre las 10 y las 11 de la mañana, puede ser cualquier número natural entre (por ejemplo) 1 y 60.

- Muestras aleatorias $\Rightarrow$ altamente probable que $\hat{\theta} \neq \theta$.
- El estimador puntual no da información de cuán próximo está $\hat{\theta}$ de θ.

Objetivo:Encontrar un intervalo de números reales que con alta probabilidad contenga al verdadero valor del parámetro en estudio.

Intervalo de Confianza para un parámetro poblacional

Intervalo de valores posibles que puede tomar dicho parámetro. Sus extremos inferior y superior son denominados límites de confianza inferior y superior, respectivamente, y son cantidades aleatorias que *dependen de la muestra*.

Se pretende que el intervalo estimado:

- Contenga al verdadero valor del parámetro con una alta probabilidad (nivel de confianza)
- Sea relativamente angosto

Intervalos de confianza

Sean θ_I, θ_S los límites inferior y superior del intervalo de confianza para el parámetro θ ($\theta \in \mathbb{R}$).

Se busca encontrar estimadores $\hat{\Theta}_I, \hat{\Theta}_S$ tales que

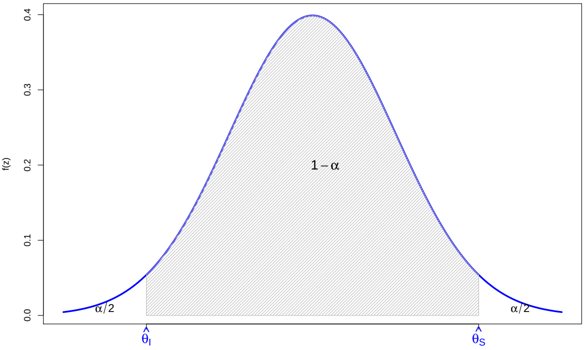
$$P(\hat{\Theta}_I < \theta < \hat{\Theta}_S) = 1 - \alpha$$

$0 < \alpha < 1$

$(1 - \alpha)$ es el nivel de confianza del intervalo.

En tal caso el intervalo $(\hat{\Theta}_I, \hat{\Theta}_S)$ es un intervalo de confianza del $100 \cdot (1 - \alpha)\%$ (bilateral)

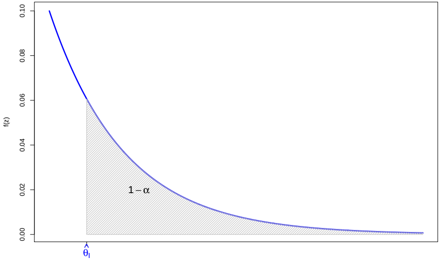
Intervalo de confianza para media poblacional de un distribución simétrica



Intervalos de Confianza Unilaterales

$$P(\hat{\Theta}_I < \theta) = 1 - \alpha \Rightarrow (\hat{\Theta}_I, \infty)$$

$$P(\theta < \hat{\Theta}_S) = 1 - \alpha \Rightarrow (-\infty, \hat{\Theta}_S)$$



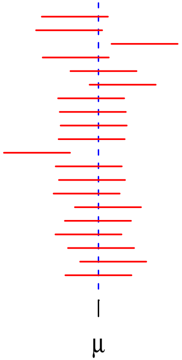
Intervalos de Confianza - Interpretación

Observaciones

- $\hat{\Theta}_I$, $\hat{\Theta}_S$ son estadísticos (funciones de la muestra).
- $P(\hat{\Theta}_I \leq \theta \leq \hat{\Theta}_S) = 1 - \alpha$ se lee como *la probabilidad que el intervalo* ($\hat{\Theta}_I$,$\hat{\Theta}_S$) *contenga al verdadero valor de* θ *es* $1 - \alpha$
- Una vez observada la muestra se calcula $(\hat{\theta}_I, \hat{\theta}_S)$, que es el intervalo "observado". El verdadero valor de θ puede, o no, estar contenido en ese intervalo.
- Si se toman muestras aleatorias de una misma población y con cada una de ellas se calcula un int. de confianza para $\theta \Rightarrow (1 - \alpha)$% de las veces dicho intervalo contendrá el verdadero valor de θ

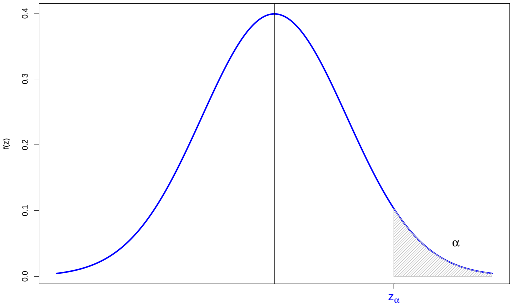
Interpretación

- Objetivo: estimar μ, media poblacional de una v.a $X \sim f(x;\mu)$
- Calcular intervalo de 90% de confianza para μ.
- Se obtuvieron 20 muestras aleatorias de $X \sim f(x;\mu)$, y con ellas se calcularon los 20 intervalos de confianza.



Sea $Z \sim N(0,1)$. Se define z_α a aquel valor que satisface

$$P(Z \geq z_\alpha) = \alpha$$



$1-\alpha$	0.90	0.95	0.99
$\alpha/2$	0.05	0.025	0.005
$z_{\alpha/2}$	1.64	1.96	2.57

Intervalos de confianza para media poblacional, μ

En esta materia derivaremos intervalos de confianza (IC) para la media poblacional ($E(X) = \mu$) en las siguientes situaciones:

1. Población normal, con varianza conocida. $(X \sim \mathcal{N}(\mu, \sigma^2))$
2. Población normal, varianza desconocida, tamaño de muestra n pequeño. ($(X \sim \mathcal{N}(\mu, \sigma^2)$, $n < 30$)
3. Muestras grandes ($n \geq 30$, **independientemente** de la distribución de la variable)
   3.1 varianza conocida
   3.2 varianza desconocida

1. IC - Población normal, σ^2 conocida

Sea $X_1, \ldots, X_n$ muestra aleatoria de una v.a. con distribución $N(\mu, \sigma^2)$, con σ conocida.
Si se quiere encontrar un intervalo del $100 \cdot (1-\alpha)\%$ de confianza para μ, se buscan $\hat{\Theta}_I, \hat{\Theta}_S$ tales que $P(\hat{\Theta}_I < \mu < \hat{\Theta}_S) = 1 - \alpha$.

$$X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, (\sigma/\sqrt{n})^2) \text{ y } \mathrm{Z} = \frac{\bar{\mathrm{X}} - \mu}{\sigma/\sqrt{\mathrm{n}}} \sim \mathrm{N}(0, 1)$$

1. IC media - Población normal, σ^2 conocida

$$
\begin{aligned}
1-\alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\
&= P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < \frac{\sigma}{\sqrt{n}}\cdot z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < \frac{\sigma}{\sqrt{n}}\cdot z_{\alpha/2} - \bar{X}\right) \\
&= P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} + \bar{X} < \mu < \frac{\sigma}{\sqrt{n}}\cdot z_{\alpha/2} + \bar{X}\right)
\end{aligned}
$$

1. IC media - Población normal, σ^2 conocida

$$(\hat{\Theta}_I, \hat{\Theta}_S) = (\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2})$$

es un intervalo del $100.(1-\alpha)\%$ de confianza para μ.

- Es un intervalo aleatorio, centrado en $\bar{X}$ y se extiende $z_{\alpha/2} \cdot \sigma/\sqrt{n}$ a cada lado de $\bar{X}$.
- La longitud del intervalo es $2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$ que NO es aleatoria.

Cálculo IC media - Población normal, σ^2 conocida

Si $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución $N(\mu, \sigma^2)$ con σ conocida, y $x_1, x_2, \ldots, x_n$ es una realización de dicha muestra. Un intervalo de confianza del $100.(1-\alpha)\%$ para μ está dado por

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

Ejemplo 1: Ej 5, Guía TP 10

La duración de la batería de cierto modelo de teléfono móvil se puede aproximar por una distribución normal con una desviación estándar de 5 meses. Se toma una muestra aleatoria simple de 10 baterías y se obtienen las siguientes duraciones (en meses):

$$33, 34, 26, 37, 30, 39, 26, 31, 36, 19$$

Construir un intervalo de 95% de confianza para la duración media de la batería de ese modelo de teléfono móvil.

Elección del tamaño de muestra

- El error de estimación , $E = |\bar{X} - \mu| \leq z_{\alpha/2}\sigma/\sqrt{n}$, con un nivel de confianza del $100 \cdot (1 - \alpha)\%$
- Si la situación (experimento) lo permite, fijado $(1 - \alpha)$, puede elegirse n, el tamaño de la muestra, de manera tal que el error en la estimación no exceda cierta cota E(fijada de antemano). En tal caso

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

- La amplitud del intervalo de confianza es $2 \cdot z_{\alpha/2}\sigma/\sqrt{n}$
- Si se quiere estimar un intervalo de amplitud menor o igual a A, entonces debe tomarse una muestra de tamaño, $n \geq \left(\frac{2z_{\alpha/2}\sigma}{A}\right)^2$

Ejemplo 2: Ej 7, Guía TP 10

El tiempo de conexión a internet de los alumnos de cierta universidad sigue una distribución normal con desviación estándar igual a 15 minutos. Para estimar la media de la población se quiere calcular un intervalo de confianza que tenga una amplitud menor o igual que 6 minutos, con un nivel de confianza del 95%. Determinar cuál es el tamaño mínimo de la muestra que es necesario observar.

Ejemplo 3: Ej 8, Guía TP 10

Se estima que el tiempo de reacción de un conductor ante un obstáculo imprevisto tiene una distribución normal con desviación estándar de 0,05 segundos. Si se quiere conseguir que el error de estimación de la media no supere los 0,01 segundos con un nivel de confianza del 99%,
¿qué tamaño mínimo ha de tener la muestra de tiempos de reacción?

2. IC media - población normal, σ desconocida

Distribución t-student

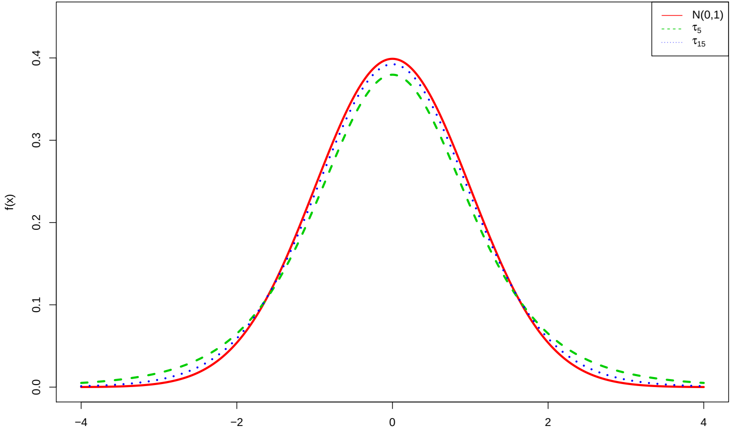
Sea $X_1, \ldots, X_n$ muestra aleatoria de una v.a normal con media μ y desviación estándar σ desconocidas.

Entonces la v.a.

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

tiene distribución t de Student con $n - 1$ grados de libertad.$\bar{X}$ y S son la media muestral y la desviación estándar muestral, respectivamente.

Distribución t

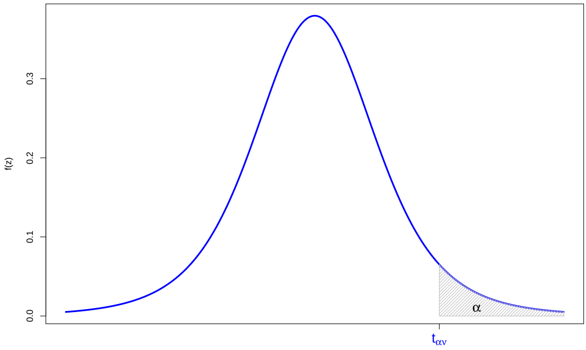


2. IC media - población normal, σ desconocida

De manera similar a lo hecho anteriormente, resulta que un intervalo del $100 \cdot (1-\alpha)\%$ de confianza para la media μ de una población normal, cuando no se conoce σ es

$$\left(\bar{x} - t_{\frac{\alpha}{2},(n-1)} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}/,(n-1)} \cdot \frac{s}{\sqrt{n}}\right)$$

donde $\bar{x}, s$ son la media y desviación estándar muestrales, respectivamente, y $t_{\frac{\alpha}{2},\nu}$ es el valor crítico tal que $P(T > t_{\frac{\alpha}{2},\nu}) = \alpha/2$


f(z)
0.0
0.1
0.2
0.3
α
tαν

Ejemplo 4: Ejercicio 9 Guía TP 10

El número de errores diarios que se cometen al intentar conectar con una determinada red informática se distribuye normalmente con media desconocida. Para intentar conocer dicha media se realiza un M.A.S. de tamaño 10 días; resultando en

$$2;\ 3;\ 4;\ 5;\ 4;\ 3;\ 5;\ 0;\ 0;\ 1$$

errores.

a) Obtenga un intervalo de confianza para la media de errores cometidos diariamente con un nivel de significancia del 1%.

3a). IC media - muestras grandes, σ conocida

Sea $X_1, \ldots, X_n$ muestra aleatoria con media μ y desviación estándar $\sigma < \infty$.

Si n es grande ($n \geq 30$) $\Rightarrow$ Teo. Central del Límite dice que $\bar{X}$ es aproximadamente normal y que $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ es aproximadamente $N(0,1)$. Así

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

Similarmente a lo hecho anteriormente,

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

es un intervalo de confianza del $100.(1-\alpha)\%$ para la media μ cuando el tamaño de la muestra es grande y σ es conocida.

3b. IC media - muestras grandes, σ desconocida

Sea $X_1, \ldots, X_n$ muestra aleatoria con media μ y desviación estándar σ desconocida.

Se usa S como estimación de σ, si n es grande ($n \geq 30$)
$\Rightarrow Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ es aproximadamente $N(0,1)$ y

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

es un intervalo de confianza del $100.(1-\alpha)\%$ para la media μ cuando el tamaño de la muestra es grande y σ NO es conocida.

Ejemplo 5

Un método para resolver la carencia de energía eléctrica requiere de la construcción de plantas eléctricas flotantes unas millas mar adentro. Se necesita una estimación de la densidad del tráfico naval en el área para evitar posibles colisiones de un barco con la planta flotante (aunque anclada).

a) Sea X: número de barcos que pasan por día en un radio de 10 kilómetros de una cierta planta flotante. ¿Qué distribución propondría para X?

b) Si quiere estimar el número promedio de barcos que pasan por día en un radio de 10km de esa planta, ¿cómo lo haría?

Continuación ejemplo 5

c) El número de barcos que pasan en un radio de 10 kilómetros de la ubicación propuesta de la planta eléctrica registrado en 60 días de verano tuvo una media de $\bar{x} = 7.2$. Suponiendo $\sigma = 2$ conocida, calcular un intervalo de confianza del 95% para el número promedio de barcos que pasan dentro de un radio de 10 kilómetros de localización de la planta eléctrica.

d) De una muestra de 90 observaciones diarias de barcos en los meses de junio a agosto se obtuvieron los siguientes datos: $\bar{x} = 4.7, s^2 = 4.9$. Obtener un intervalo de confianza del 95% para para el número promedio de barcos que pasan dentro de un radio de 10 kilómetros de localización de la planta eléctrica en los meses de invierno.

Ejemplo 6

Se hicieron pruebas para estimar el número promedio μ de horas por mes que pasa un niño en edad escolar frente al televisor, obteniéndose dos intervalos de confianza:

$$36.7 \leq \mu \leq 38.5 \quad ; \quad 34.8 \leq \mu \leq 40.4$$

a) ¿Cuál es el valor del promedio muestral de horas mensuales que pasaron los niños frente al televisor?

b) Uno de los intervalos fue calculado con un nivel de confianza del 90% y el otro con el 99%. Si se usó la misma muestra para ambos intervalos, ¿cuál de ellos se calculó con un 90% de confianza? Justificar

Distribución χ^2

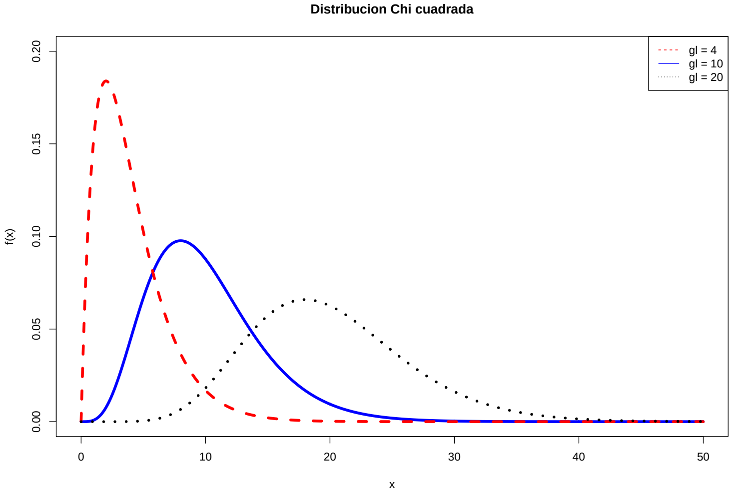
Sea $X_1, \ldots, X_n$ muestra aleatoria de una distribución normal con media μ y varianza σ^2. Entonces $Z_i = \frac{X_i - \mu}{\sigma}$ son v.a. independientes con distribución normal estándar y la v.a definida como

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

tiene distribución χ^2 con n grados de libertad.

Si $Y \sim \chi_n^2 \Rightarrow E(Y) = n, \mathrm{Var}(Y) = 2n$

Distribución χ^2



Distribución χ^2

Sea $X_1, \ldots, X_n$ muestra aleatoria de una distribución normal con media μ y varianza σ^2. Si $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$, entonces

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)S^2}{\sigma^2}$$

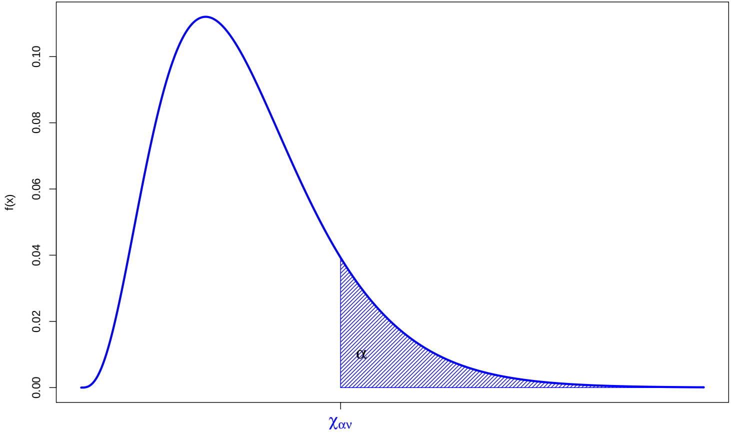
tiene distribución χ^2 con $(n-1)$ grados de libertad. Además $\bar{X}$ y S^2 son v.a. independientes.

Cálculo de Probabilidades

Supongamos $\chi^2 \sim \chi^2_n$. Para calcular probabilidades debe recurrirse a tablas que dan valores aproximados. Fijado $0 < \alpha < 1$ se define el *valor de porcentaje* $\chi^2_{\alpha,n}$ como

$$P(\chi^2 > \chi^2_{\alpha,n}) = \alpha$$

Cálculo de Probabilidades χ^2



Uso de tabla χ^2

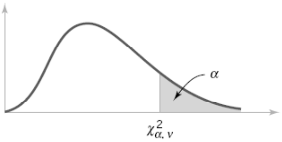


Table III Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80

Uso de tabla χ^2

Ejemplos:

1. Encontrar los valores de los siguientes percentiles: $\chi^2_{0.05,10}, \chi^2_{0.025,5}, \chi^2_{0.95,8}$
2. Si $\chi^2 \sim \chi^2_{11}$, calcular $P(\chi^2 > 17.28)$
3. En R invocar la función pchisq(q, df, lower.tail = FALSE)

Intervalos de Confianza para σ^2 Dist. Normal

Sea $X_1, \ldots, X_n$ muestra aleatoria de una v.a normal con media μ y desviación estándar σ desconocidas.

Entonces la v.a.

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2, \text{ con n} - 1 \text{ g.l}$$

Y un intervalo del $100 \cdot (1-\alpha)\%$ de confianza para σ^2 de una distribución normal es:

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,(n-1)}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,(n-1)}} \right)$$

donde s^2 es la varianza muestral.

Ejemplo 7: Ejercicio 9 Guía TP 10

El número de errores diarios que se cometen al intentar conectar con una determinada red informática se distribuye normalmente con media desconocida. Para intentar conocer dicha media se realiza un M.A.S. de tamaño 10 días; resultando en

$$2; 3; 4; 5; 4; 3; 5; 0; 0; 1$$

errores.

Halle un intervalo de confianza de 99% para la desviación estándar de la población muestreada

Intervalos de confianza para proporciones

Se toma una muestra aleatoria de tamaño n de una población

- Sea $X =$ n^o de individuos de la muestra que pertencen a una clase de interés (solo dos clases)
- $\hat{P} = \frac{X}{n}$ estimador puntual de la proporción p de la población que pertenece a dicha clase.
- Si n es grande ($n \geq 30$) por Teo. DeMoivre -Laplace, $Z = \frac{X-np}{\sqrt{np(1-p)}} = \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}}$ es aproximadamente normal estándar
- $\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$ es un intervalo del $100{\cdot}(1-\alpha)\%$ de confianza para la proporción poblacional p.

Ejemplo 8: Ejercicio 10 Guía TP 10

Una muestra de 100 votantes elegidos al azar entre todos los de un distrito dado, indicó que 55 de ellos estaban a favor de un determinado candidato.
a) Halle un intervalo de confianza para la proporción de todos los votantes que estaban a favor de ese candidato, con un nivel de significación del 5%.
b) En las mismas condiciones del apartado anterior, se realiza la experiencia para conseguir una cota de error del 0.01. ?'Cuál debe ser el tamaño de la muestra?

Situaciones:

1. Un veterinario desea comparar el efecto de dos dietas de engorde en terneros
2. Se desea estudiar los efectos de dos marcas distintas de combustible en la eficiencia operativa de motores de automóviles (k/l).
3. Se quiere comparar la vida útil de pilas Duracell y pilas Eveready
4. Un médico desea comparar el efecto de dos concentraciones diferentes de una droga antidepresiva en la capacidad de atención.

Preguntas:

- ¿Qué similitudes hay con los casos vistos hasta ahora?
- ¿Qué diferencias hay con los casos vistos hasta ahora?

IC diferencia de medias

En algunas aplicaciones interesa estimar la diferencia entre dos parámetros poblacionales, como por ejemplo, la diferencia entre las medias de dos poblaciones distintas.

Suposiciones

- $X_{11}, X_{12}, \ldots X_{1n_1}$, es una muestra aleatoria de tamaño n_1 de la población 1 (X_1).
- $X_{21}, X_{22}, \ldots X_{2n_2}$, es una muestra aleatoria de tamaño n_2 de la población 2 (X_2).
- Las dos poblaciones, X_1 y X_2, son independientes.

IC diferencia de medias - Poblaciones normales, varianzas conocidas

Si $X_{11}, X_{12}, \ldots X_{1n_1}$ y $X_{21}, X_{22}, \ldots X_{2n_2}$ son muestras aleatorias de tamaños n_1 y n_2 de dos poblaciones normales $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ con σ_1 y σ_2 conocidas, entonces

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

y un intervalo bilateral del $100 \cdot (1-\alpha)\%$ de confianza para $\mu_1 - \mu_2$ es

$$(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

donde

$\bar{x}_i$, es la media muestral de la población i, $i = 1, 2$.

IC diferencia de medias- Poblaciones normales, varianzas desconocidas

Si $X_{11}, X_{12}, \ldots X_{1n_1}$ y $X_{21}, X_{22}, \ldots X_{2n_2}$ son muestras aleatorias de tamaños n_1 y n_2 de dos poblaciones normales $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ con σ_1 ,σ_2 desconocidas pero iguales $(\sigma_1 = \sigma_2)$, *entonces*

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}$$

Un intervalo bilateral del $100 \cdot (1-\alpha)\%$ de confianza para $\mu_1 - \mu_2$ es

$\left(\bar{x_1} - \bar{x_2} - t_{\alpha/2,(n_1+n_2-2)} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x_1} - \bar{x_2} + t_{\alpha/2,(n_1+n_2-2)} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$

donde

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

$\bar{x}_i$, s_i^2 son la media y varianza muestrales, respectivamente, de la población i, $i = 1, 2$.

IC Diferencia de Proporciones

Si se toman dos muestras aleatorias de tamaño n_1 y n_2 de dos poblaciones X_1 y X_2, donde X_i cuenta el número de observaciones que pertenecen a la clase de interés en la muestra i, y siempre que n_1 y n_2 sean lo suficientemente grandes, entonces:

$$\hat{P}_1 = X_1/n_1; \quad \hat{P}_2 = X_2/n_2$$

tienen aproximadamente distribución normal, y un intervalo bilateral del $100 \cdot (1-\alpha)\%$ de confianza para $p_1 - p_2$ es

$$\left(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

donde
$\hat{p}_i = x_i/n_i$, es la proporción muestral para la población i, $i = 1, 2$.

Ejemplo 9: Ej. 12 Guía TP 10

La altura media de 1.000 estudiantes de la Carrera de Licenciatura en Sistemas de la Facultad de Ciencias Exactas de la UNNE, en el año 1.990 fue de 168 cm, mientras que en el año 1.995, la altura media de 1.200 estudiantes fue de 169,2 cm. Las desviaciones típicas poblacionales son 21 cm y 25 cm, respectivamente.
Hallar un intervalo de confianza para la diferencia de medias poblacionales, con un nivel de confianza del 95%.

Resolver ejercicios 13,14 y 15 guía TP 10.