Intervalo de Confianza

Definición:

Sea $X_1, X_2, \ldots, X_n$ una m.a de una distribución $F(\theta)$. Un intervalo de confianza de nivel (1-α) o intervalo de 100(1-α)% de confianza, es un intervalo de extremos aleatorios; $g_1(X_1, X_2, \ldots, X_n)$ (extremo inferior) y $g_2(X_1, X_2, \ldots, X_n)$ (extremo superior), que contiene al parámetro θ, con probabilidad 1-α, esto quiere decir que:

$$P(g_1(X_1, X_2, \ldots, X_n) \leq \theta \leq g_2(X_1, X_2, \ldots, X_n)) = 1-\alpha$$

Notación: A dicho intervalo de confianza lo denotamos como

$$IC_{1-\alpha}(\theta) = (g_1(X_1, X_2, \ldots, X_n) \; ; \; g_2(X_1, X_2, \ldots, X_n))$$

¿Cómo construimos un IC?

Utilizamos el método del pivote

Los pasos a seguir son:

1) Se busca una función de la m.a y del parámetro de interés, que llamaremos *función pivote,* y cuya distribución no dependa de ningún parámetro desconocido, donde θ es la única cantidad desconocida. Denotamos a la función pivote como: h($X_1, X_2, \ldots, X_n$)
2) Determinar un par de números reales a y b tales que:

$$P(a<h(X_1, X_2, \ldots, X_n)<b)=1-\alpha \quad (*)$$

3) Siempre que sea posible, despejar los extremos aleatorios de (*) y se obtendrá $IC_{1-\alpha}(\theta)=(g_1(X_1, X_2, \ldots, X_n) \,;\, g_2(X_1, X_2, \ldots, X_n))$

Ejemplo:

Ingenieros industriales realizaron un estudio sobre una muestra de 31 mecanógrafos entrenados para determinar la altura preferida del teclado, dando un promedio de $\bar{x}$ =80cm. Supongamos que la altura está normalmente distribuida con σ =2cm. Obtener un Intervalo de confianza para μ del 95%(1-α), es decir, para la altura preferida promedio verdadera por la población de los mecanógrafos experimentados.

Primero definimos la v.a: X_i: "la altura en cm preferida del teclados del i-ésimo mecanógrafo"

$$X_i \sim N(\mu; 2^2) \text{ i=1,..., 31}$$

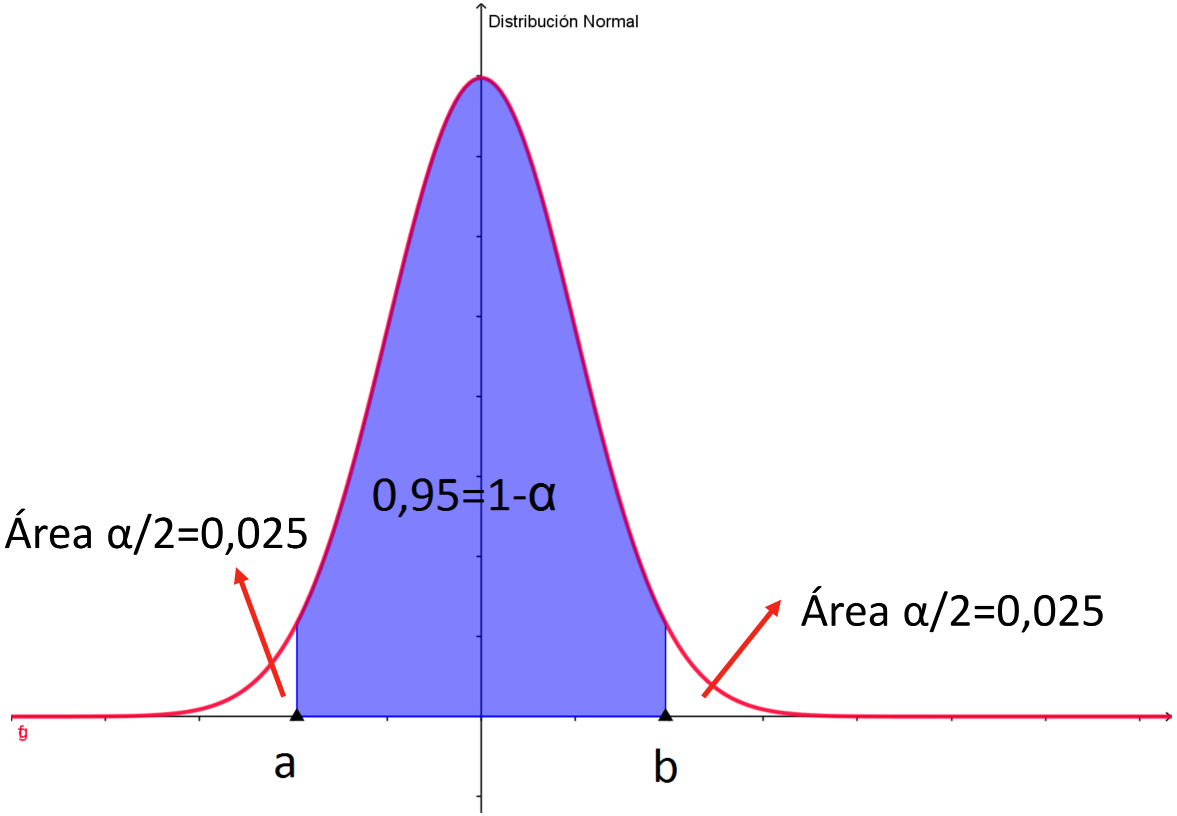
Además $\bar{X} \sim N\left(\mu; \frac{2^2}{31}\right) \Longrightarrow \quad Z = \frac{\bar{X}-\mu}{\sqrt{2^2/_{31}}} \sim N(0;1)$

Notar que **Z** es una función que depende sólo de μ y de la m.a Y su distribución no depende de ningún parámetro desconocido.

1)Entonces basta con tomar como función pivote a Z= h($X_1, X_2, \dots, X_n$)

2)Ahora debemos buscar a y b tales que:

P(a ≤ Z ≤ b)=0,95, Entonces a=-1,96 y b=1,96



Del gráfico vemos que
$\Phi(a) = 0{,}025$, y que
1- $\Phi(a) = 1 - 0{,}025$
Por tabla $\Phi(-1{,}96) = 0{,}025$

$P(-1{,}96 \leq Z \leq 1{,}96)=0{,}95$ (1)

3) despejo µ de (1)

$P\left(-1{,}96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1{,}96\right)=0{,}95$ multiplico por $\sigma/\sqrt{n}$

$P\left(-1{,}96 \cdot \sigma/\sqrt{n} \leq \bar{X}-\mu \leq 1{,}96 \cdot \sigma/\sqrt{n}\right)=0{,}95$ sumo por $-\bar{X}$

$P\left(-1{,}96 \cdot \sigma/\sqrt{n} - \bar{X} \leq -\mu \leq 1{,}96 \cdot \sigma/\sqrt{n} - \bar{X}\right)=0{,}95$ multiplico por (-1)

$P\left(\bar{X} + 1{,}96 \cdot \sigma/\sqrt{n} \geq \mu \geq \bar{X} - 1{,}96 \cdot \sigma/\sqrt{n}\right)=0{,}95$ reordeno

$P\left(\underbrace{\bar{X} - 1{,}96 \cdot \sigma/\sqrt{n}}_{g_1(X_1,X_2,\dots,X_n)} \leq \mu \leq \underbrace{\bar{X} + 1{,}96 \cdot \sigma/\sqrt{n}}_{g_2(X_1,X_2,\dots,X_n)}\right)=0{,}95$

Decimos que con una probabilidad de 0,95 el intervalo aleatorio contiene a μ.

El intervalo obtenido es:

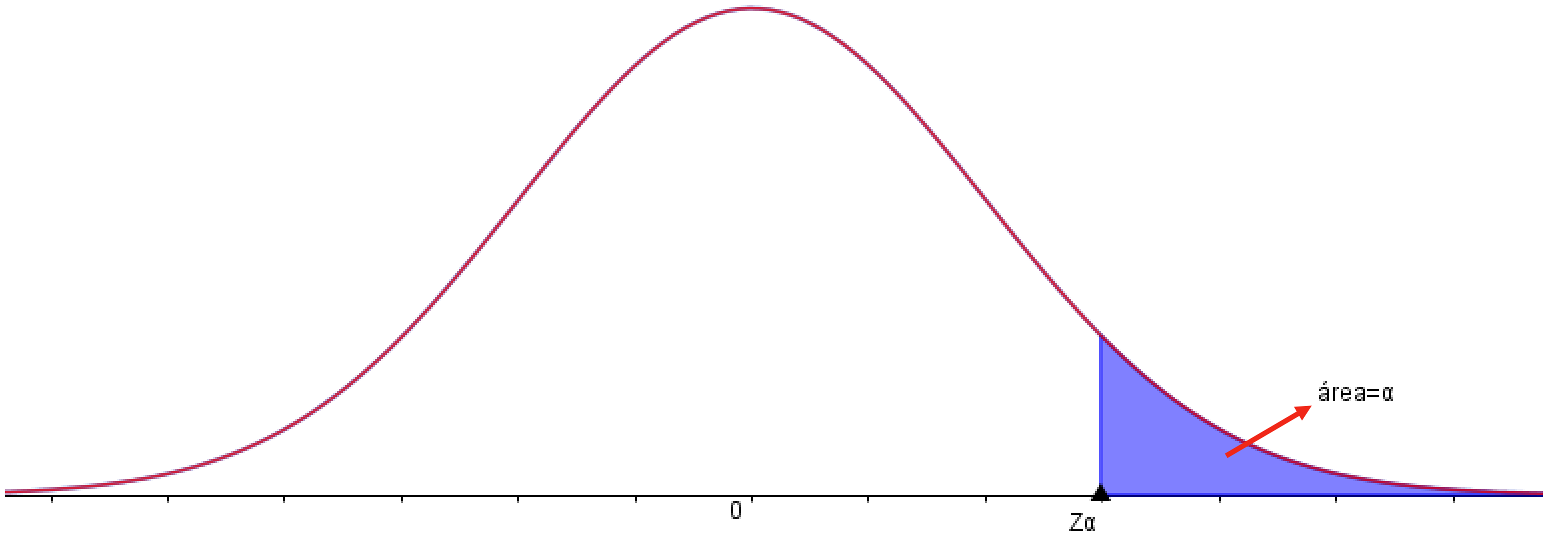
$IC_{0,95}(\mu)=[\bar{X} - 1{,}96 \cdot \sigma/\sqrt{n};\ \bar{X} + 1{,}96 \cdot \sigma/\sqrt{n}]$ este intervalo es ***aleatorio***

El intervalo numérico: $IC_{0,95}(\mu)=[80-1{,}96 \cdot 2/\sqrt{31};80+1{,}96 \cdot 2/\sqrt{31}]=[79{,}3;80{,}7]$

Notar que el intervalo es angosto, se estimó a μ con bastante precisión.

Definición: Sea Z$\sim$ $N(0,1)$ y sea $\alpha\epsilon(0,1)$, $\alpha\epsilon\mathbb{R}$, se define el ***valor crítico*** z_α como el valor que cumple P(Z> z_α)=α

α=1-$\Phi(z_\alpha)$



Observación: Notar que en el ejemplo a= -$z_{\alpha/2} = -z_{0,025}$=-1,96y

b= $z_{\alpha/2}$= $z_{0,025}$=1,96

¿Cómo buscar en la tabla los valores críticos de una distribución Normal?

Por ejemplo: P(Z> $z_{0,025}$)=0,025

$z_{0,025}$=1,96

Tabla III. Valores críticos de la distribución Normal estándar
$P[Z > Z_\alpha] = \int_{Z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp^{-z^2/2} dz = \alpha$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2910	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1597	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0995
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0394	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143

- Entonces se obtiene: P(-$z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$)=1- α

Dando el IC, $IC_{1-\alpha}(\mu)$=[$\bar{X}$ $-z_{\alpha/2}.\,{}^{\sigma}/_{\sqrt{n}}$; $\bar{X}$ $+z_{\alpha/2}.\,{}^{\sigma}/_{\sqrt{n}}$]

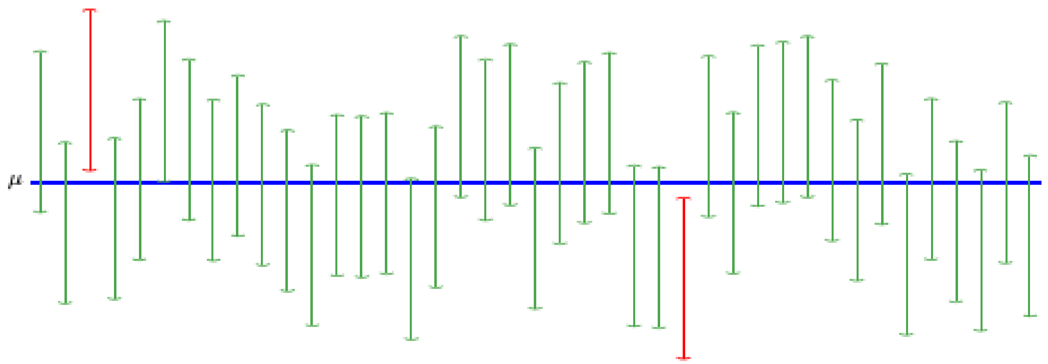
Resumiendo:

Si la m.a $X_1, X_2, \ldots.., X_n$ tiene distribución normal con varianza conocida, la función pivote es Z=$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ y el

$$IC_{1-\alpha}(\mu)=[\bar{X} - z_{\alpha/2}.\,{}^{\sigma}/_{\sqrt{n}};\ \bar{X} + z_{\alpha/2}.\,{}^{\sigma}/_{\sqrt{n}}\ ]$$

Interpretación del nivel de confianza:

Supongamos que se seleccionan muchas muestras de tamaño 31 (por ejemplo 100), por ende se obtendrán diferentes promedios $\bar{x}$ para cada muestra, dando como resultado intervalos numéricos diferente. Lo que se puede afirmar es que aproximadamente el 95% de estos intervalos contendrán al verdadero μ y el 5% no contendrán al verdadero valor de μ. Entonces cuando construimos un solo intervalo de nivel 0,95, podemos tener un 95% de confianza de que ese intervalo sea uno de los que contienen a μ.



Notar que el verdadero valor de μ está sobre la línea azul, los intervalos de color rojo son los que no contienen a μ, los de color verde sí contienen a μ.

Nivel de confianza, precisión y tamaño de la muestra

Lo deseable es que el nivel de confianza 1- α sea lo mayor posible, pero consecuentemente z_α aumenta dando intervalos más anchos.

Por ejemplo si queremos un nivel del 99%, los valores críticos son -2,58 y 2,58, dando un intervalo de confianza de mayor longitud.

Si el intervalo es más ancho, hay mayor probabilidad de que el intervalo contenga a μ, eso se paga con menos precisión.

Del ejemplo tenemos:

$$IC_{0,99}(\mu)=[80-2,58\ .\ ^{2}/_{\sqrt{31}};80+2,58\ .\ ^{2}/_{\sqrt{31}}] =[79,073;80,927]$$

La longitud del intervalo es: 80,927-79,073=1,854

Mientras que a nivel 0,95 la longitud es: 80,7-79,3=1,4

¿Qué deberíamos hacer si queremos tener un nivel de confianza del 99% pero con mayor precisión?

Supongamos que queremos que la longitud no sea mayor que uno

La expresión general para este tipo de IC es: $\boxed{L = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$

Quiero que $L \leq 1$, entonces reemplazando se tiene

$$2x2{,}58\,\frac{2}{\sqrt{n}} \leq 1 \text{ despejo n } \Rightarrow n \geq 106{,}5024$$

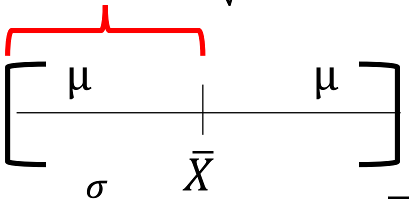
Como n es natural, basta con tomar un $n \geq 107$

En general, a mayor nivel de confianza se tiene menor precisión, una solución posible es aumentar el tamaño de la muestra.

Definición: el ***error de estimación*** ε es la distancia entre un estimador $\hat{\theta}$ y su parámetro θ. Esto es $\left|\hat{\theta} - \theta\right|$= ε y al límite del error de estimación se lo denota con la letra b, por ejemplo: $\left|\hat{\theta} - \theta\right|$≤ b, es decir que b es la máxima distancia que se acepta entre $\hat{\theta}$ y θ.

En el intervalo de confianza, a la mitad del ancho se lo llama límite del error de estimación, es decir que: $|\bar{X} - \mu| \leq L/2$

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



$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$\bar{X}$+ $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

El intervalo de confianza está centrado en $\bar{X}$, además $\bar{X}$ no estará a más de $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ de distancia de μ con (1-α)% de confianza.

Intervalo de confianza para la media de muestras grandes

Si se tiene $X_1, X_2, \ldots, X_n$ una m.a de una población con media μ y desvío $\boldsymbol{\sigma}$ **conocido** siempre que n sea lo suficientemente grande(n ≥ 30).

El TCL nos dice que $\bar{X} \approx N(\mu; {}^{\sigma^2}/_{n})$ con lo cual $Z = \dfrac{\bar{X}-\mu}{\sqrt{{}^{\sigma^2}/_{n}}} \approx N(0;1)$

Tomo a Z como función pivote

Se plantea P(-$z_{\alpha/2} \leq \dfrac{\bar{X}-\mu}{{}^{\sigma}/_{\sqrt{n}}} \leq z_{\alpha/2}$)≅1- α

Despejando μ, se tiene que $\boxed{IC_{1-\alpha}(\mu)=[\bar{X} - z_{\alpha/2} . {}^{\sigma}/_{\sqrt{n}}; \bar{X} + z_{\alpha/2} . {}^{\sigma}/_{\sqrt{n}}]}$ de nivel aproximado

Si $\boldsymbol{\sigma}$ es **desconocido**, puede ser aproximado por la desviación estándar muestral s cuando n ≥ 30 y el IC de nivel aproximado es en este caso:

$$IC_{1-\alpha}(\mu)=[\bar{X} - z_{\alpha/2}.{}^{s}/_{\sqrt{n}}; \bar{X} + z_{\alpha/2}.{}^{s}/_{\sqrt{n}}]$$

El cual se obtiene de realizar los siguientes pasos:

Considerar la función pivote: $Z = \dfrac{\bar{X}-\mu}{\sqrt{{}^{s^2}/_{n}}} \approx N(0;1)$

(Utilizando el lema se Slutzky que permite reemplazar $\boldsymbol{\sigma}$ por s) y de plantear P(-$z_{\alpha/2} \leq \dfrac{\bar{X}-\mu}{{}^{s}/_{\sqrt{n}}} \leq z_{\alpha/2}$)≅1- α y despejar μ

Ejemplo:

Un científico interesado en vigilar contaminantes químicos en alimentos y consecuentemente, la acumulación de contaminantes en la dieta humana, seleccionó una m.a de 50 hombres. Se encontró que el promedio de la ingesta diaria de productos lácteos fue de $\bar{x} = 756\ grs$ por día, con una desviación estándar de $s = 35\ grs$ por día. Use esta información para construir el IC de 95% para la ingesta diaria media verdadera de productos lácteos para hombres.

X_i="la ingesta diaria de productos lácteos en grs del i-ésimo hombre" i=1,..,50

X_i tiene distribución desconocida (n ≥ 30) y hay independencia

Función pivote: $$Z = \frac{\bar{X}-\mu}{\sqrt{s^2/n}} \approx N(0;1)$$

Planteo:

P($-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{s/\sqrt{n}} \leq z_{\alpha/2}$)≅1- α y despejo μ,

P($-z_{\alpha/2}.\, s/\sqrt{n} \leq \frac{\bar{X}-\mu}{\cancel{s/\sqrt{n}}}\, \cancel{s/\sqrt{n}} \leq z_{\frac{\alpha}{2}}\, .\, s/\sqrt{n}$) ≅1- α multiplico por $s/\sqrt{n}$

P($-z_{\alpha/2}.\, s/\sqrt{n} - \bar{X} \leq \cancel{\bar{X}} - \mu \cancel{-\bar{X}} \leq z_{\frac{\alpha}{2}}\, .\, s/\sqrt{n} - \bar{X}$) ≅1- α sumo por $-\bar{X}$

P($+z_{\alpha/2}.\, s/\sqrt{n} + \bar{X} \geq -\mu(-1) \geq -z_{\frac{\alpha}{2}}\, .\, s/\sqrt{n} + \bar{X}$) ≅1- α multiplico por (-1)

Acomodo: P($\bar{X} - z_{\alpha/2}.\, s/\sqrt{n} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}}\, .\, s/\sqrt{n}$) ≅1- α

Por lo tanto: $IC_{1-\alpha}(\mu)$=[$\bar{X} - z_{\alpha/2}.\, s/\sqrt{n}$; $\bar{X} + z_{\alpha/2}.\, s/\sqrt{n}$] nivel aproximado (intervalo aleatorio)

El intervalo numérico es: $IC_{0,95}(\mu)$=[756−1,96. $35/\sqrt{50}$;756+1,96. $35/\sqrt{50}$] =[746,3;765,7] grs por día

IC para la media de una distribución normal con varianza desconocida

Ejemplo:

Un fabricante ha invertido una nueva pólvora que fue probada en ocho proyectiles. Las velocidades resultantes en la boca del cañón, en pies por segundo, fueron las siguientes: 3005; 2925;2935;2965;2995; 3005; 2937;2905

Se desea encontrar un IC de 95% para el verdadero valor de promedio de velocidad μ para proyectiles de este tipo. Suponiendo que las velocidades en la boca del cañón están normalmente distribuidas.

IC para la media de una distribución normal con varianza desconocida

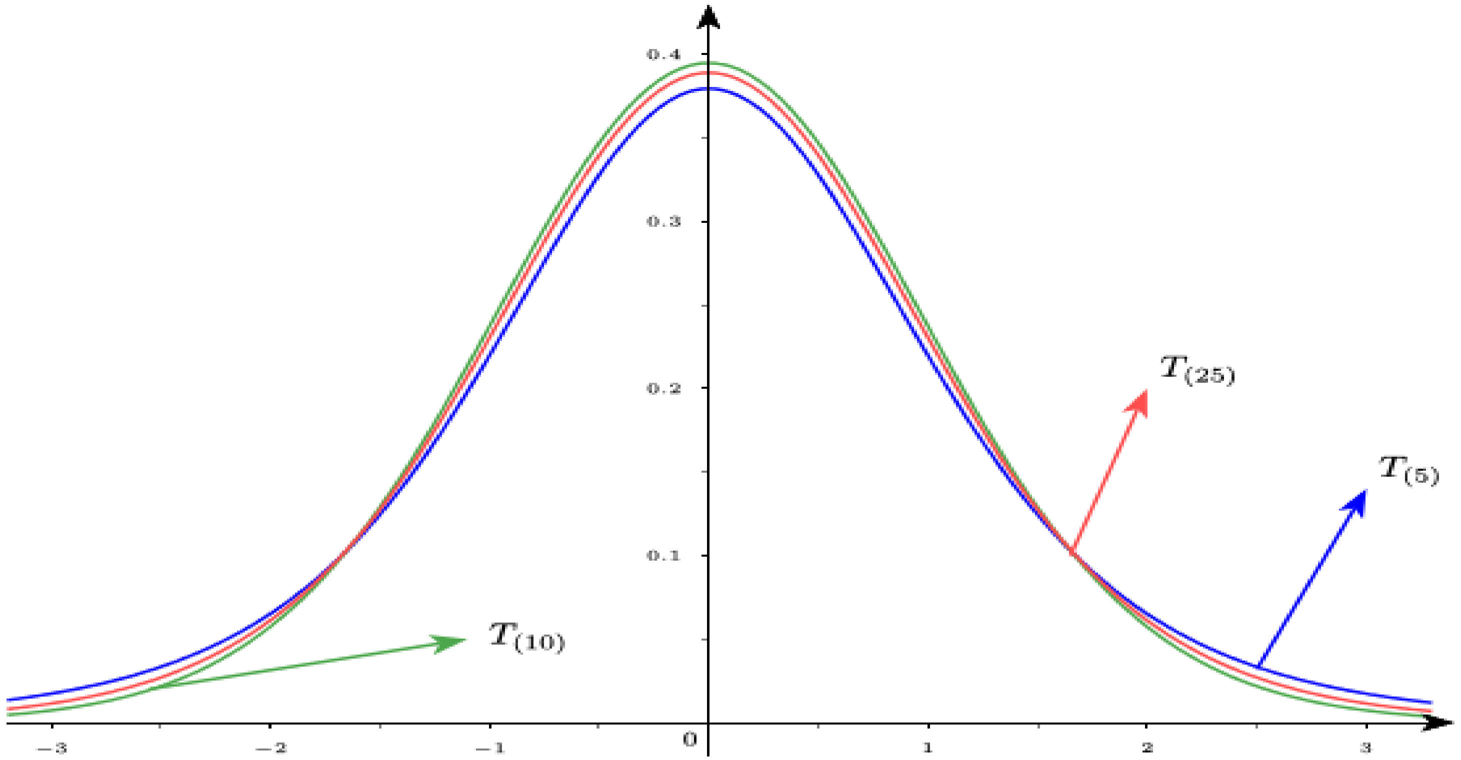
Cuando n es pequeño, debemos conocer la distribución de la m.a de la v.a X, es decir $X \sim N(\mu; \sigma^2)$ con σ desconocido, entonces debemos estimar a σ por S.

En este caso tomamos como función pivote:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \underset{\downarrow}{\sim} T(n-1) \quad \text{siendo gl=n-1}$$

Distribución T-Student

La distribución T-Student es simétrica, con forma de campana con las colas más pesadas que la distribución normal.



- Para obtener el IC debemos plantear:

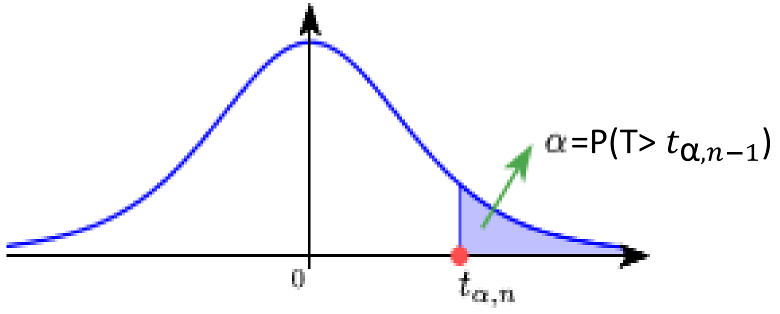
P(-t≤T ≤t)=1-α, (1)   nivel es exacto

siendo P(T>t)= α/2, es decir que t es el valor crítico de la distribución T-Student lo denotamos $t_{\alpha/2,n-1}$.

Idem con P(T<-t).

De (1)

P($-t_{\alpha/2}$ ≤ T ≤ $t_{\alpha/2}$)=1-α despejo μ quedando



$$IC_{1-\alpha}(\mu)=[\bar{X} - t_{\alpha/2}.\, {}^{s}/_{\sqrt{n}};\ \bar{X} + t_{\alpha/2}.\, {}^{s}/_{\sqrt{n}}]$$

Si $X_1, X_2, \dots, X_n$ es una m.a de tamaño n de una v.a X donde X$\sim N(\mu;\sigma^2)$ y σ es desconocido, un intervalo de confianza para μ de nivel 1-α es:

$$IC_{1-\alpha}(\mu)=[\bar{X} - t_{\alpha/2}.{}^{S}/_{\sqrt{n}}; \bar{X} + t_{\alpha/2}.{}^{S}/_{\sqrt{n}}]$$

Seguimos con el ejemplo:

Definimos a X_i como "la velocidad en la baca del cañón del i-ésimo proyectil en pies/seg" siendo $X_i \sim N(\mu;\sigma^2)$ σ desconocido

Función pivote: T=$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim T(n-1)$ gl=n-1=8-1=7

De los datos $\bar{x} = 2959$ y s=39,7 $t_{\alpha/2,n-1}$= $t_{0,025,7}$=2,365

Tabla IV. Valores críticos de la distribución t de Student:
Abcisas $t_{\alpha;\nu}$ que dejan a su derecha un área α en una t con ν grados de libertad.

ν	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781

$IC_{0,95}(\mu)=[\bar{X} - t_{\alpha/2}.{}^{s}/_{\sqrt{n}};\bar{X} + t_{\alpha/2}.{}^{s}/_{\sqrt{n}}]$= reemplazo por los valores de la m.a

$= IC_{0,95}(\mu)=[2959-2,365.{}^{39,7}/_{\sqrt{8}};2959+2,365.{}^{39,7}/_{\sqrt{8}}]$

=[2926,35;2991,7]

IC para una proporción (muestra grande)

Ejemplo:

Una m.a de 985 electores, fueron encuestados durante un maratón telefónico realizado por el partido A. De ellos 592 indicaron que tenían intención de votar por la candidata de dicho partido. Construya el IC de 90% para p, proporción de electores de la población que tiene intención de votar a la candidata del partido A.

¿Se puede decir que la candidata ganará la elección?

(supongamos que para ganar necesita más del 50% de los votos)

Sea p=P(elector tenga intención de voto favorable a la candidata A)

Defino: $X_i = \begin{cases} 1, si\ el\ elector\ i\ tiene\ intención\ de\ voto\ positivo \\ 0, \qquad c.c \end{cases}$

Con i= 1,…, 985, se tiene que $X_i \sim B(1, p)$

Sea X="cantidad de electores entre 985 que votan a la candidata del partido A" siendo X$\sim B(n, p)$

$\sum_{i=1}^{n} X_i$ =X$\approx N(np;$ np(1-p)) np>5y n(1-p)>5

Sabemos que $\hat{p} \approx N(p; \frac{p(1-p)}{n})$ a partir de este estimador considero

$\hat{p}$= $\frac{X}{n}$ = $\bar{X}$ =

$$\frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$ (n suficientemente grande)

recordar que: E($\hat{p}$)=p y V($\hat{p}$)= $\frac{p(1-p)}{n}$

Veamos que: E($\hat{p}$)=p y V($\hat{p}$)= $\frac{p(1-p)}{n}$

E($\hat{p}$)=E($\frac{X}{n}$)=$_{(*)}$ $\frac{1}{n}$ E(X)=$_{(**)}$ $\frac{1}{\cancel{n}}$ $\cancel{n}$p

(*)Linealidad de la esperanza

(**) E(X)=np porque X tiene distribución Binomial

V($\hat{p}$)=V($\frac{X}{n}$)=$_{(*)}$$\frac{1}{n^2}$V(X)=$_{(**)}$ $\frac{1}{\cancel{n}^2}$ $\cancel{n}$p(1-p) = $\frac{p(1-p)}{n}$

(*)Propiedad de la Varianza

(**) V(X)=np(1-p) por tener X distribución Binomial

- Pero para despejar p de la expresión $\frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}}$ es difícil, aparece en el numerador y denominador

Estimo a la V($\hat{P}$)$= \frac{p(1-p)}{n}$ $\Rightarrow \widehat{V(\hat{P})}$=$\frac{\hat{P}(1-\hat{P})}{n}$

La función pivote resultante es: Z= $\frac{\hat{P}-p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}$ $\approx N(0,1)$

(p aparece sólo una vez)

- Para obtener el IC planteo:

P$(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \cong 1-\alpha \Rightarrow$ despejo p

P$\left(-z_{\alpha/2} \leq \frac{\hat{P}-p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \leq z_{\alpha/2}\right) \cong 1-\alpha$

P$\left(-z_{\alpha/2}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq \frac{\hat{P}-p}{\cancel{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}}.\cancel{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \leq z_{\alpha/2}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right) \cong 1-\alpha$ multiplico por $\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$

P$\left(-z_{\alpha/2}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} - \hat{P} \leq \cancel{\hat{p}} - p - \cancel{\hat{p}} \leq z_{\alpha/2}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} - \hat{P}\right) \cong 1-\alpha$ sumo $-\hat{P}$

P$\left(+z_{\alpha/2}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} + \hat{P} \geq -p.(-1) \geq -z_{\alpha/2}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} + \hat{P}\right) \cong 1-\alpha$ multiplico por (-1)

P$\left(\hat{p} - z_{\frac{\alpha}{2}}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq \boldsymbol{p} \leq \hat{P} + z_{\alpha/2}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right) \cong 1-\alpha$ Acomodo

$$\therefore\ IC_{1-\alpha}(p) = \left[\hat{P} - z_{\frac{\alpha}{2}}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\ ;\ \hat{P} + z_{\frac{\alpha}{2}}.\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right]$$ nivel aproximado

En el problema tenemos que $\hat{p} = \frac{592}{985}$ y que $z_{\alpha/2}$= $z_{0,05} = 1{,}645$

$$IC_{0,9}(p) = \left[\hat{p} - z_{\frac{\alpha}{2}} . \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,; \hat{P} + z_{\frac{\alpha}{2}} . \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [0{,}575;\ 0{,}627]$$

Entre el 57,5% y el 62,7% de votantes tienen intención de voto a favor de la candidata del partido A, los extremos son mayores al 50%, con lo cual se puede decir que con un 90% de confianza que la candidata ganará.

Observación 1:

Este procedimiento depende de la aproximación normal a la distribución binomial. Por lo tanto el intervalo se puede utilizar si

n.$\hat{P}$ > 10 y n(1− $\hat{P}$) > 10 , es decir, la muestra debe contener un mínimo de diez éxitos y diez fracasos

Observación 2:

Notar que p∈ $(0,1)$ entonces si los extremos del intervalo sobrepasan esos valores se debe recortar:

Por ejemplo: (-0,3;0,7) ⇒ (0;0,7)

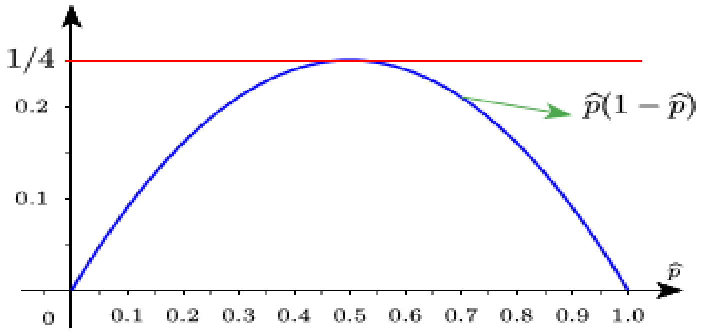
(0,3;1,7) ⇒ (0,3;1)

Si ahora queremos saber el tamaño necesario para un IC del 90% para p si su ancho debe ser cuando mucho 0,1 independientemente de p

En este caso quiero que L=2. $z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0{,}1$

$$\frac{2.z_{\frac{\alpha}{2}} .\sqrt{\hat{p}.(1-\hat{p})}}{0{,}1} \leq \sqrt{n} \Rightarrow \frac{2.z_{\frac{\alpha}{2}} .\sqrt{\hat{p}.(1-\hat{p})}}{0{,}1} \leq_{(*)} \left(\frac{2.z_{\frac{\alpha}{2}} .\sqrt{1/4}}{0{,}1}\right)^2 = \left(\frac{2.z_{\frac{\alpha}{2}} .(\frac{1}{2})}{0{,}1}\right)^2 \leq n$$

Como no se tiene conocimiento de $\hat{p}$, debemos proceder así:



Para cualquier $\hat{p}$ vale $\hat{p}(1-\hat{p}) \leq 1/4$ (*)

$270{,}60 \leq n$, como n es natural, basta con tomar al menos 271 entrevistas($271 \leq n$)

Si queremos ahora saber el tamaño muestral necesario para que un IC del 90% para p, con un ancho menor o igual a 0,1, si la candidata está convencida de que por lo menos el 70% votará por ella.

En este caso tenemos información acerca de $\hat{p}$ nos dice que es ≥ 0,7.

L=2. $z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0{,}1$ despejo n

$$\frac{2.z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{p}.(1-\hat{p})}}{0{,}1} \leq \sqrt{n} \Rightarrow \left(\frac{2.z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{p}.(1-\hat{p})}}{0{,}1}\right)^2 \leq n$$

Usando la información:

$$\left(\frac{2.1{,}645 \cdot \sqrt{0{,}7.(1-0{,}7)}}{0{,}1}\right)^2 \leq n \Rightarrow \quad 227{,}3061 \leq n$$

Basta con tomar al menos 228 entrevistas.