Distribuciones Muestrales

Inferencia Estadística

Normalmente, por distintas razones, es imposible estudiar la totalidad de la población para tomar decisiones u obtener conclusiones. **Por tal razón, para obtener conclusiones sobre la población se recurre a una muestra de la misma.**

Inferencia Estadística

Normalmente, por distintas razones, es imposible estudiar la totalidad de la población para tomar decisiones u obtener conclusiones. **Por tal razón, para obtener conclusiones sobre la población se recurre a una muestra de la misma.**

La inferencia estadística está formada por un conjunto de métodos utilizados para **tomar decisiones u obtener conclusiones sobre una población a partir de la información contenida en una muestra** obtenida en forma aleatoria de la población.

Inferencia Estadística

Normalmente, por distintas razones, es imposible estudiar la totalidad de la población para tomar decisiones u obtener conclusiones. **Por tal razón, para obtener conclusiones sobre la población se recurre a una muestra de la misma.**

La inferencia estadística está formada por un conjunto de métodos utilizados para **tomar decisiones u obtener conclusiones sobre una población a partir de la información contenida en una muestra** obtenida en forma aleatoria de la población.

Cualquier inferencia o conclusión obtenida de la población, necesariamente, estará basada en la información proporcionada por la muestra.

Inferencia Estadística

En muchos casos sabemos o presumimos conocer cómo se distribuye una población.

Sabemos por ejemplo que la población es aproximadamente normal; pero desconocemos la media y la varianza poblacionales.

Sabemos que la variable de interés es binomial pero desconocemos la probabilidad de éxito poblacional o el número de pruebas de Bernoulli.

Sabemos que se trata de un proceso de Poisson pero desconocemos la frecuencia media de ocurrencia por unidad.

Presumimos que la variable es exponencial pero desconocemos el parámetro que define la distribución exponencial poblacional.

La estimación de uno o varios parámetros poblacionales desconocidos es posible construyendo funciones de probabilidad de variables aleatorias muestrales, más conocidos como estimadores muestrales.

Población y Muestra

Población o universo: Es el conjunto de individuos o elementos con características comunes que se desean investigar. Normalmente es demasiado grande para poder abarcarla. **El estudio de toda la población se denomina censo**.

Muestra: es un subconjunto de la población al que tenemos acceso y sobre el que realmente hacemos las observaciones (mediciones). El objetivo principal de la toma de una muestra aleatoria es obtener información sobre los parámetros no conocidos de la población.

La **inferencia estadística** consiste en generalizar las conclusiones extraídas de una muestra sobre la población.

Muestreo Aleatorio Simple (M.A.S.)

Muestreo. Es el proceso para extraer una muestra de una población. Es necesario utilizar ***"muestras representativas"*** del total de la población, es decir, muestras en las que exista alguna garantía de que cualquier elemento de la población queda representado.

En el ***Muestreo Aleatorio Simple***, cada elemento de la población tiene la misma probabilidad de ser elegido para formar parte de la muestra. Cada muestra del mismo tamaño tiene la misma probabilidad de ser seleccionada.

- El muestreo aleatorio simple en poblaciones finitas se realiza "***con sustitución***".
- Este procedimiento **garantiza la independencia de las observaciones**.

Muestreo Aleatorio Simple (M.A.S.)

Formalmente:

Muestra aleatoria simple: Si X es una variable aleatoria con distribución de probabilidades o *fdp* $f(x)$, el conjunto de n observaciones tomadas de la variable X, $X_1, X_2, \ldots, X_n$, con resultados numéricos $x_1, x_2, \ldots, x_n$, es una muestra aleatoria de tamaño n si:

a) Las X_i son variables aleatorias independientes.

b) Cada observación X_i tiene la misma distribución de probabilidades.

El objetivo de tomar una muestra es obtener información sobre los parámetros no conocidos de la población.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Inferencia Estadística

Estimación puntual: Es el valor numérico que toma un estimador. Se calcula con los datos de la muestra, del cual se espera que estime un parámetro poblacional.

Parámetro Poblacional	Estimador	Estimación
Media μ	$\hat{\mu} = \overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$	$\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
Varianza σ^2	$\hat{\sigma}^2 = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$	$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$
Proporción p	$\hat{p} = \dfrac{X}{n} = \dfrac{\text{número éxitos}}{\text{número pruebas}}$	$\hat{p} = \dfrac{x}{n}$
Parámetros poblacionales, estimadores y estimaciones		

Distribuciones de Muestreo.

Las muestras aleatorias obtenidas de una población son, por naturaleza propia, impredecibles. No se debería esperar que dos muestras aleatorias del mismo tamaño y tomadas de la misma población tengan, por ejemplo, la misma media muestral (o que sean completamente parecidas). Por lo tanto es de esperar que cualquier estadístico cambie su valor de una muestra a otra, por ello, se quiere estudiar la distribución de todos los valores posibles de un estadístico.

La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

Inferencia Estadística

Distribuciones de Muestreo.

Cómo los valores de un estadístico, varían de una muestra aleatoria a otra, se lo puede considerar como una **variable aleatoria** con su correspondiente distribución de probabilidades. La distribución de probabilidades de un estadístico se conoce como ***distribuciones de muestreo***.

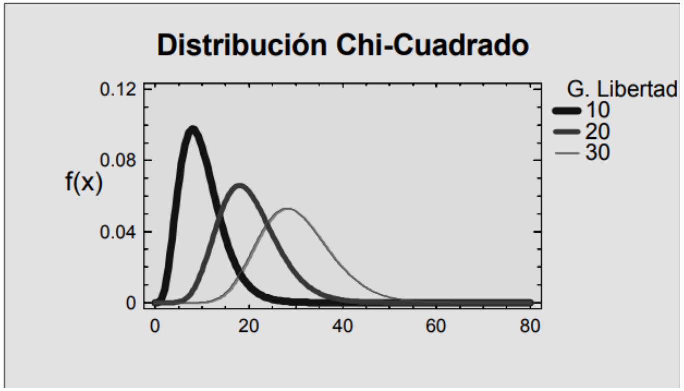
La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

Objetivo: ***Conocer la distribución muestral de los siguientes estadísticos bajo ciertas condiciones.***

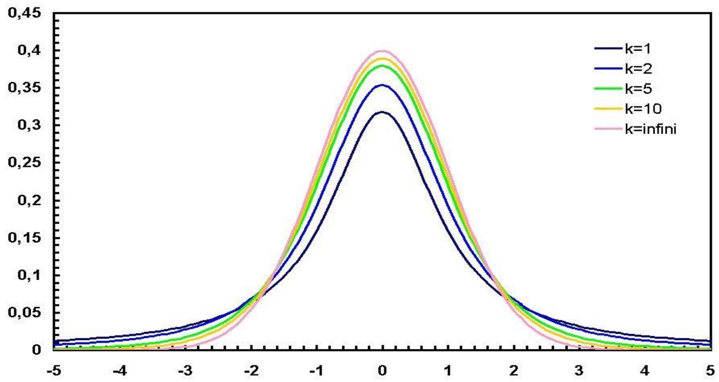
Estimador
$\hat{\mu} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$
$\hat{\sigma}^2 = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
$\hat{p} = \dfrac{X}{n} = \dfrac{\text{númeroéxitos}}{\text{númeropruebas}}$

Antes de continuar veamos algunas distribuciones asociadas al muestreo

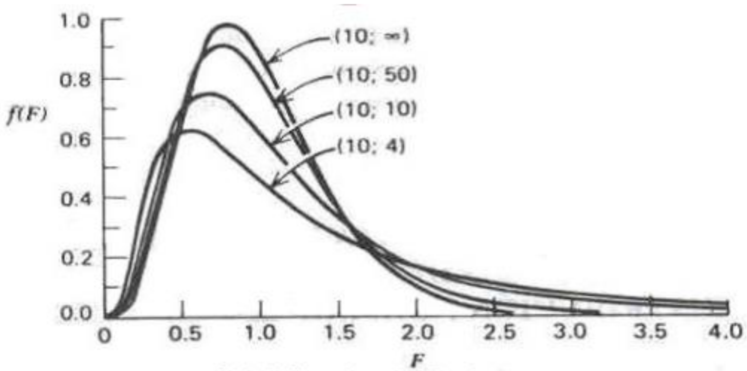
Distribución Chi-Cuadrado



Distribución t-Student



Distribución F de Fisher



Distribución F para varios grados de libertad.

Distribuciones asociadas al muestreo

Distribución Ji-Cuadrado

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias distribuidas normalmente e independientemente con media $\mu = 0$ y varianza $\sigma^2 = 1$. Entonces la variable aleatoria $\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$, tiene una distribución **Ji-Cuadrado** con n grados de libertad, cuya fdp está definida por:

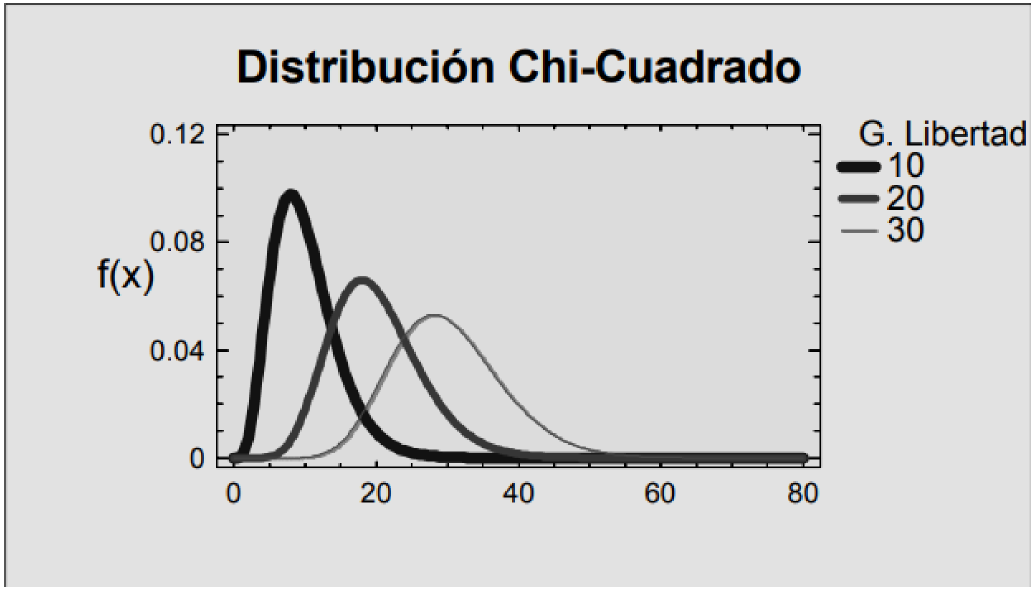
$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Donde Γ es la función Gamma.

En matemáticas, la función gamma denotada por $\Gamma(z)$, extiende el concepto de factorial a los números complejos. Se define como:

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$$

La siguiente figura muestra la gráfica de la función ji-Cuadrado para distintos grados de libertad



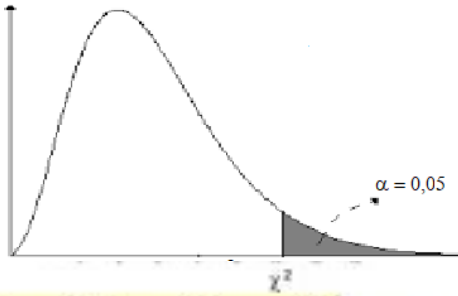
Observación:

- ✓ Si los grados de libertad $n \to \infty$ la forma límite de χ^2 es la distribución normal.
- ✓ La media y la varianza de la distribución ji-cuadrado son $\mu = n$ y $\sigma^2 = 2n$.
- ✓ Calculo de probabilidades a partir de la tabla correspondiente

$$P\left(\chi_n^2 \geq \chi_{(\alpha,n)}^2\right) = \int_{\chi_{(\alpha,n)}^2}^{\infty} f_{\chi_n^2}(x)\,dx = \alpha$$

Ejemplo: uso de tabloa ji-Cuadrado.

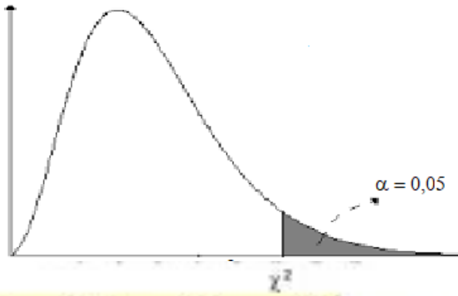
$$P\left(\chi_{10}^{2} > \chi^{2}\right) = 0.05$$



g. d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Ejemplo: uso de tabloa ji-Cuadrado.

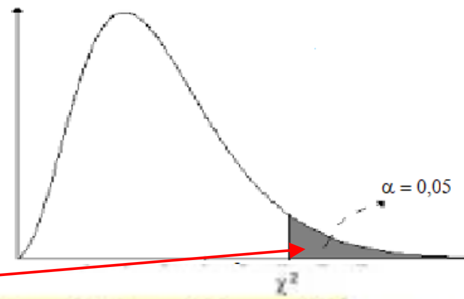
$$P\left(\chi^2_{10} > \chi^2\right) = 0.05$$



g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
$\boxed{10}$	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Ejemplo: uso de tabloa ji-Cuadrado.

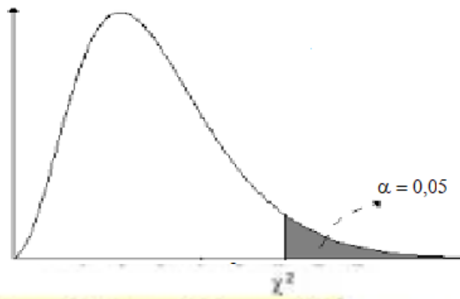
$$P\left(\chi_{10}^{2} > \chi^{2}\right) = 0.05$$



g. d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
$\boxed{10}$	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Ejemplo: uso de tabloa ji-Cuadrado.

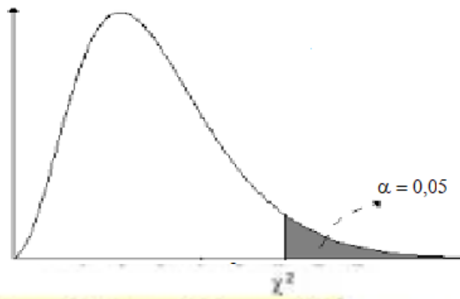
$$P\left(\chi^2_{10} > \chi^2\right) = 0.05$$



g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Ejemplo: uso de tabloa ji-Cuadrado.

$$P\left(\chi_{10}^2 > \chi^2\right) = 0.05 \Rightarrow \chi^2 = 18.307$$



g. d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	$\boxed{0,05}$	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
$\boxed{10}$	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Distribuciones asociadas al muestreo

Distribución T de Student

Si Z es una variable aleatoria con distribución normal $N(0,1)$, V una variable aleatoria con distribución Ji-Cuadrado con n grados de libertad y además Z y V son independientes, entonces la variable aleatoria:

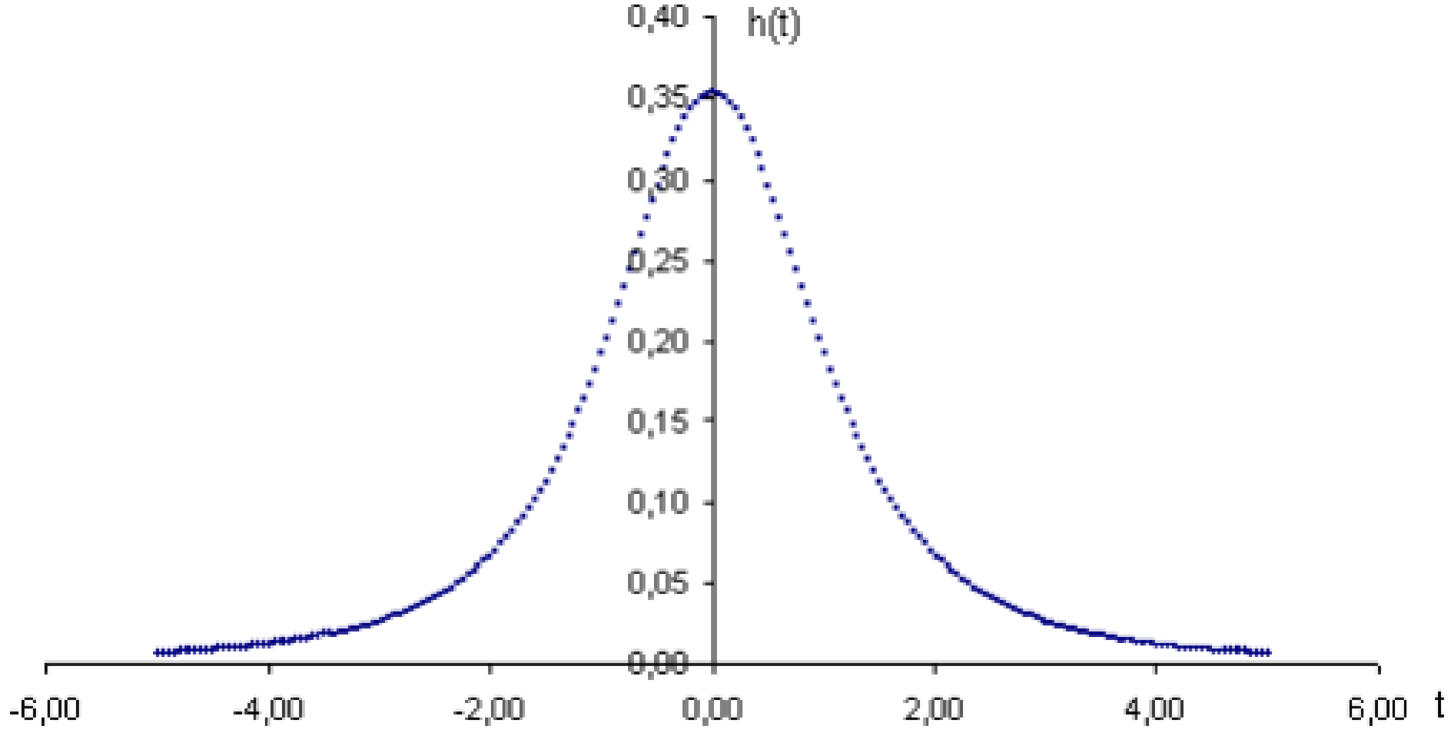
$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n$$

Tiene la siguiente fdp:

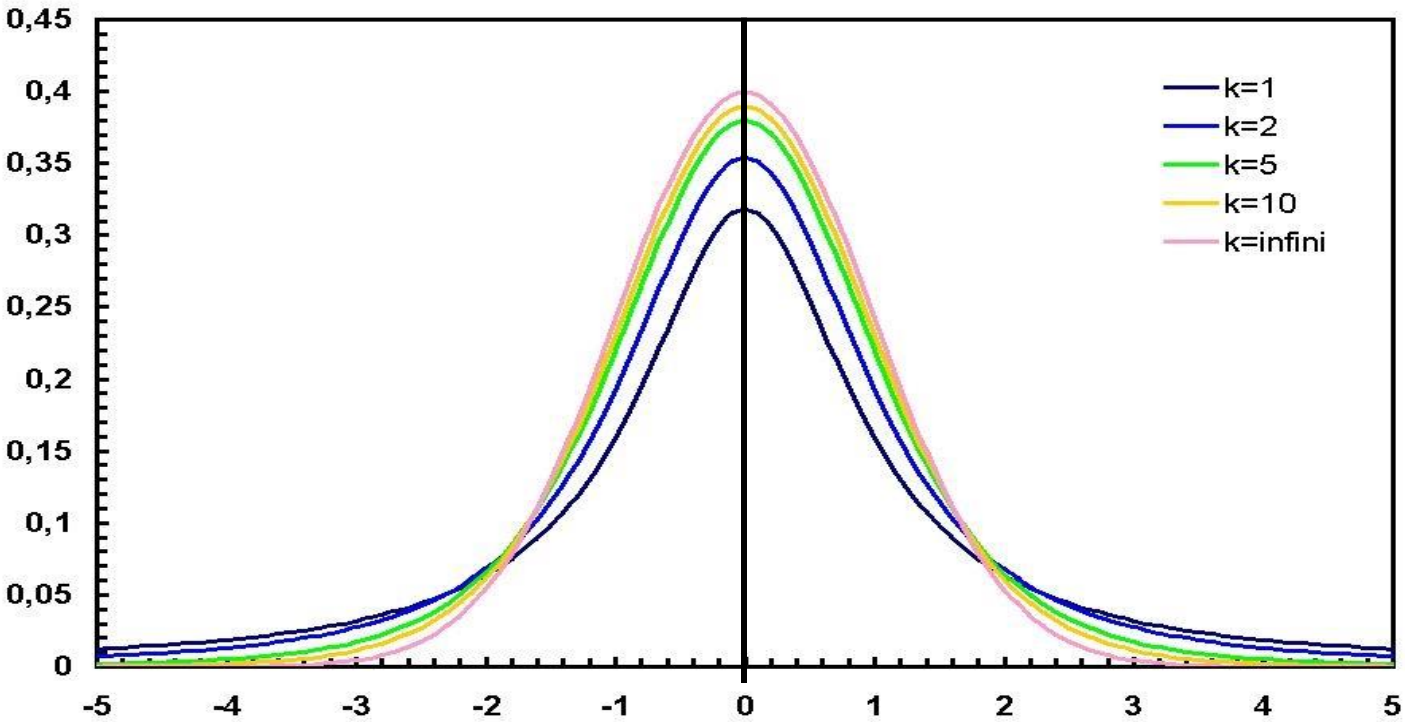
$$f(x) = \frac{\Gamma\left(\frac{(n+1)}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(\frac{x^2}{n}+1\right)^{(n+1)/2}} \qquad -\infty < x < \infty$$

La distribución de *T* se llama ahora la **distribución *t* de Student**. El parámetro n representa el número de *grados de libertad*.

La gráfica de esta función de densidad es simétrica, respecto del eje de ordenadas, con independencia del valor de **k**, y de forma algo semejante a la de una distribución normal:

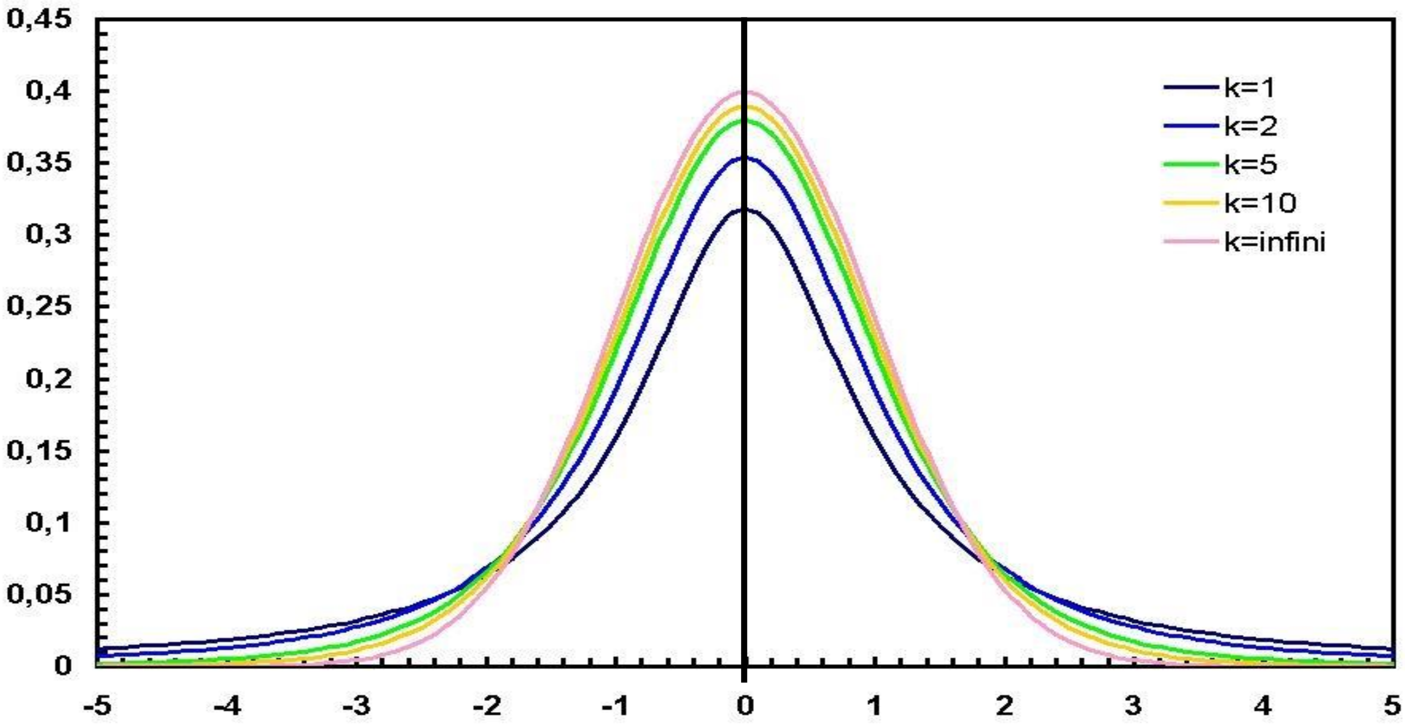


Distribución t de Student con 10 grados de liberta



Propiedades de la distribución t

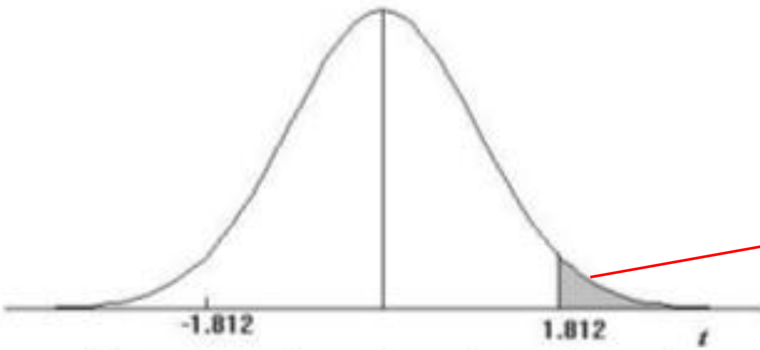
- Cada curva t tiene forma de campana con centro en 0.
- A medida que los grados de libertad (k) aumentan, la dispersión de la curva t correspondiente disminuye.
- A medida que los grados de libertad (k) tienen a infinito, la curva t se aproxima a la curva normal estándar



Observación: A diferencia de la distribución normal que depende de la media y la varianza, la distribución t solo depende de los grados de libertad, del inglés, *degrees of freedom* (df). En otras palabras, controlando los grados de libertad, controlamos la distribución.

Inferencia Estadística

Puntos de porcentaje de la distribución t



Ejemplo

Para $\phi = 10$ grados de libertad:

$P[\, t > 1.812 \,] = 0.05$
$P[\, t < -1.812 \,] = 0.05$

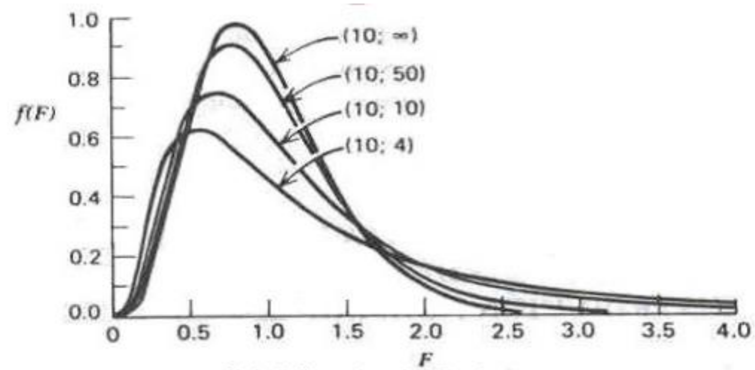
r \ α	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
$\boxed{10}$	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850

Distribuciones asociadas al muestreo

La distribución F de Fisher

Si χ_1^2 y χ_2^2 son v.a. independientes que tienen distribución ji-cuadrado, con n_1 y n_2 grados de libertad, respectivamente, la variable aleatoria definida por:

$$F = \frac{\dfrac{\chi_1^2}{n_1}}{\dfrac{\chi_2^2}{n_2}}$$



Distribución F para varios grados de libertad.

Tiene una distribución F con n_1 y n_2 grados de libertad.

Inferencia Estadística

<u>Distribuciones de Muestreo</u>.

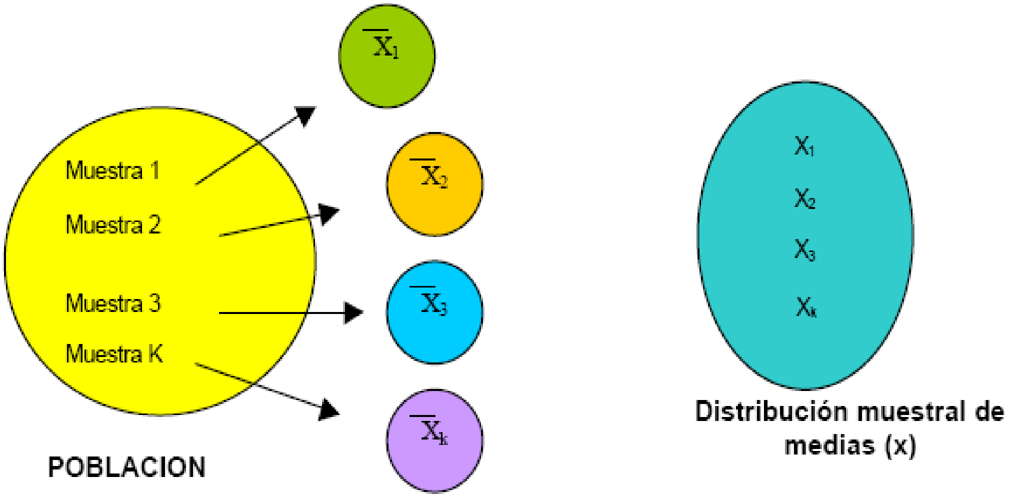
Cómo los valores de un estadístico, varían de una muestra aleatoria a otra, se lo puede considerar como una **variable aleatoria** con su correspondiente distribución de probabilidades. La distribución de probabilidades de un estadístico se conoce como ***distribuciones de muestreo***.

La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

Objetivo: ***Conocer la distribución muestral de los siguientes estadísticos bajo ciertas condiciones.***

Estimador
$\hat{\mu} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$
$\hat{\sigma}^2 = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
$\hat{p} = \dfrac{X}{n} = \dfrac{\text{númeroéxitos}}{\text{númeropruebas}}$

<u>Determinación empírica de la distribución muestral de un estadístico:</u> Para hallar empíricamente la distribución muestral de un estadístico es necesario seleccionar todas las muestras de tamaño ***n*** de dicha población y a partir de dicha información construir la distribución de frecuencia relativa de los valores del estadístico, la cual es considerada como su distribución muestral.



Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que los parámetros poblacionales son: la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que los parámetros poblacionales son: la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Tomamos **muestras de tamaño 2, con reposición**. Cada muestra es de la forma donde: (X_1, X_2)

Consideremos la variable aleatoria: .

$\overline{X}$: promedio de 2 observaciones elegidas al azar de entre las 4

Tabla 2		
Muestra, n=2		
x_1	x_2	Promedio
2	2	2
2	4	3
2	6	4
2	8	5
4	2	3
4	4	4
4	6	5
4	8	6
6	2	4
6	4	5
6	6	6
6	8	7
8	2	5
8	4	6
8	6	7
8	8	8

Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que los parámetros poblacionales son: la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Tomamos **muestras de tamaño 2, con reposición**. Cada muestra es de la forma donde: (X_1, X_2)

Consideremos la variable aleatoria: .

$$\overline{X}: \text{promedio de 2 observaciones elegidas al azar de entre las 4}$$

$$S_{\overline{X}} = \{2,3,4,5,6,7,8\}$$

Tabla 2		
Muestra, n=2		
x_1	x_2	Promedio
2	2	2
2	4	3
2	6	4
2	8	5
4	2	3
4	4	4
4	6	5
4	8	6
6	2	4
6	4	5
6	6	6
6	8	7
8	2	5
8	4	6
8	6	7
8	8	8

Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Tomamos **muestras de tamaño 2, con reposición**. Cada muestra es de la forma donde: (X_1, X_2)

Consideremos la variable aleatoria: .

$\overline{X}$: promedio de 2 observaciones elegidas al azar de entre las 4

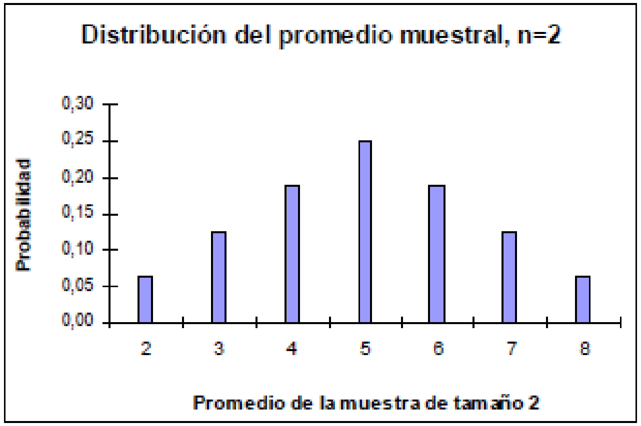
Tabla 2		
Muestra, n=2		
x_1	x_2	Promedio
2	2	2
2	4	3
2	6	4
2	8	5
4	2	3
4	4	4
4	6	5
4	8	6
6	2	4
6	4	5
6	6	6
6	8	7
8	2	5
8	4	6
8	6	7
8	8	8

La distribución de probabilidad es:

$$E\left(\overline{X}\right) = 5$$

$$V\left(\overline{X}\right) = \frac{5}{2}$$

Tabla 3: n=2	
$\overline{X}$	*Probabilidad*
2	0,0625
3	0,1250
4	0,1875
5	0,2500
6	0,1875
7	0,1250
8	0,0625
	1,0000



Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

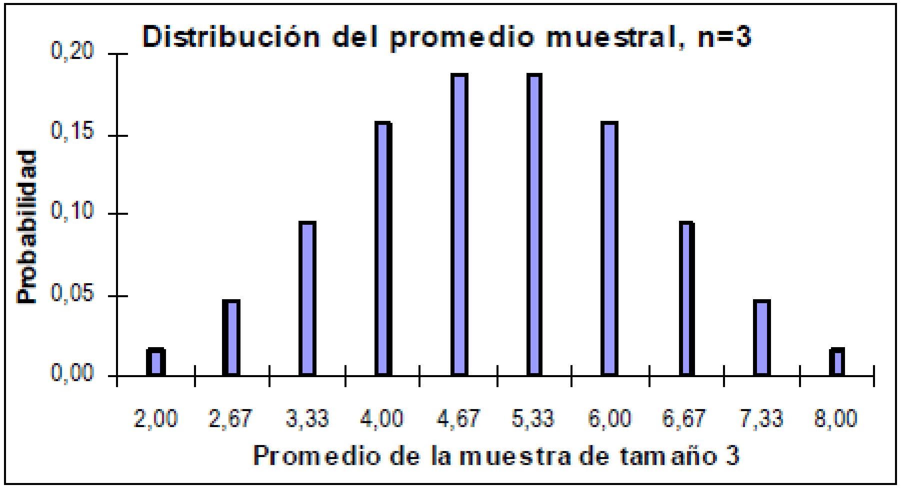
Repitiendo la experiencia, con muestras de tamaño **3**, obtenemos la distribución de los promedios que mostramos a continuación, acompañada de la gráfica:

$n = 3$

Promedio	*Probabilidad*
2,00	0,015625
2,67	0,046875
3,33	0,093750
4,00	0,156250
4,67	0,187500
5,33	0,187500
6,00	0,156250
6,67	0,093750
7,33	0,046875
8,00	0,015625
	1

$$E\left(\overline{X}\right) = 5$$

$$V\left(\overline{X}\right) = \frac{5}{3}$$



Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

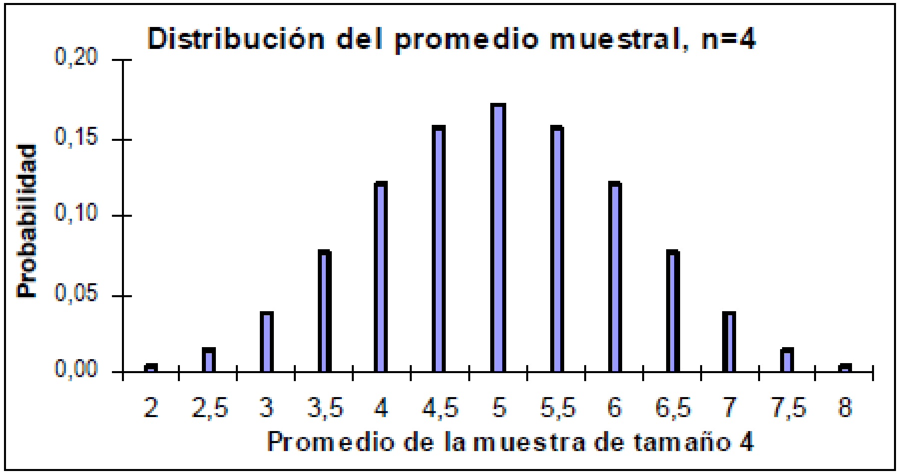
Repitiendo la experiencia, con muestras de tamaño **4**, obtenemos la distribución de los promedios que mostramos a continuación, acompañada de la gráfica:

n = 4

Promedio	*Probabilidad*
2	0,00390625
2,5	0,01562500
3	0,03906250
3,5	0,07812500
4	0,12109375
4,5	0,15625000
5	0,17187500
5,5	0,15625000
6	0,12109375
6,5	0,07812500
7	0,03906250
7,5	0,01562500
8	0,00390625
	1

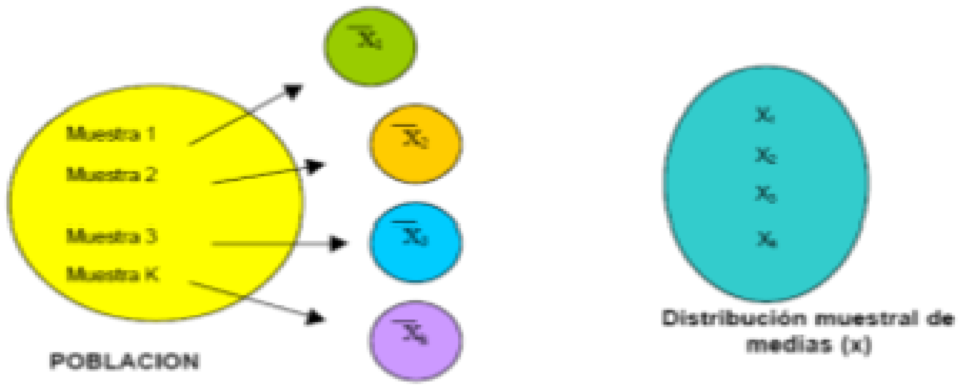
$$E\left(\overline{X}\right)=5$$

$$V\left(\overline{X}\right)=\frac{5}{4}$$



Distribuciones de Muestreo

Distribución de muestreo de la media $\bar{x}$



Si $X_1, X_2, X_3, \ldots, X_n$ son v.a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Sea $\bar{X} = \sum_{i=1}^{n} \frac{x_i}{n}$, *entonces* $\boxed{E(\bar{X}) = \mu ,\ V(\bar{X}) = \frac{\sigma^2}{n} \text{ y } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.}$

$$E(\bar{X}) = E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{\cancel{n}} \cancel{n} \mu = \mu$$

$$V(\bar{X}) = V\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{1}{n^{\cancel{2}}} \cancel{n} \sigma^2 = \frac{\sigma^2}{n}$$

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) **Distribución muestral de la media para una población normal con varianza** σ^2 **conocida.**

b) **Distribución muestral de la media para una población normal con varianza** σ^2 **desconocida.**

Distribuciones de Muestreo

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) Distribución muestral de la media para una población normal con varianza σ^2 conocida.

b) Distribución muestral de la media para una población normal con varianza σ^2 desconocida.

a) <u>Distribución muestral de la media para una población normal con varianza</u> σ^2 <u>conocida.</u>

<u>Teorema</u>: Si $X_1, X_2, X_3, \ldots, X_n$ es una muestra aleatoria extraída de una población que se distribuye normalmente $X \sim N(\mu, \sigma^2)$ entonces la media muestral se distribuye normalmente con $E(\overline{X}) = \mu$ y $V(\overline{X}) = \frac{\sigma^2}{n}$, es decir $\overline{X} - N\left(\mu, \frac{\sigma^2}{n}\right)$.

Sea $\boxed{\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} \ ,}$

$$E(\overline{X}) = E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{\cancel{n}}\cancel{n}\mu = \mu$$

$$V(\overline{X}) = V\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^{\cancel{2}}}\cancel{n}\sigma^2 = \frac{\sigma^2}{n}$$

Cómo la población se distribuye normalmente con varianza σ^2 conocida entonces por el resultado anterior para cualquier tamaño de la muestra $\overline{X} - N\left(\mu, \frac{\sigma^2}{n}\right)$.

Ley de los grandes números

Teorema de Bernoulli generalizado

Sean $x_1, x_2, x_3, \ldots, x_n$ v. a. **independientes idénticamente distribuidas** con esperanza y varianza finita, $E(x_i) = \mu$ y $V(x_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Sea $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$, para todo $\delta > 0$, $\lim_{n \to \infty} \left\{ P\left(\left| \bar{x} - \mu \right| < \delta \right) \right\} = 1$ o $\lim_{n \to \infty} \left\{ P\left(\left| \bar{x} - \mu \right| \geq \delta \right) \right\} = 0$

Interpretación: Pensando $x_1, x_2, x_3, \ldots, x_n$ como medidas independientes de una característica numérica X que producen el promedio $\bar{x}$. Entonces el límite en probabilidad, de la media muestral para $n \to \infty$ es igual a la media poblacional de la que se extrajo la muestra.

En la práctica, esto significa que si a una muestra "grande" le calculamos su media, este valor numérico puede tomarse como un valor muy próximo a la media de la población.

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) **Distribución muestral de la media para una población normal con varianza** σ^2 **conocida.**

b) **Distribución muestral de la media para una población normal con varianza** σ^2 **desconocida.**

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) **Distribución muestral de la media para una población normal con varianza σ^2 conocida.**

b) **Distribución muestral de la media para una población normal con varianza σ^2 desconocida.**

Antes de analizar este punto analicemos como se distribuye la varianza muestral.

Inferencia Estadística

Distribución de la varianza muestral.

Definamos la varianza muestral como: $$S^2 = \sum_{i=1}^{n} \frac{\left(x_i - \overline{x}\right)^2}{n-1}$$

Si la población de la cual se extrae la muestra tiene una distribución normal, la variable

$$\left(\sum_{i=1}^{n} \underbrace{\left(\frac{\left(X_i - \overline{X}\right)}{\sigma} \right)^2}_{N(0,1)} \right) \sim \chi^2_{n-1} \quad \Longrightarrow \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

• La distribución de la varianza muestral no es simétrica: tiene asimetría positiva.

Si se extraen todas las muestras posibles de una población normal y a cada una de ellas se le calcula su varianza, se obtendrá la distribución muestral de la varianza con esta distribución.

Ejemplo: Una máquina de llenado opera con una varianza de 0.83 gr^2. Si se toma una muestra de 15 unidades, ¿cuál es la probabilidad de tener una varianza muestral superior a 1,249 gr^2?

$$P\left(S^2 \geq 1{,}249\right) = P\left(\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}} \geq \frac{14 \times 1{,}249}{0{,}83}\right) = P\left(\chi^2_{14} \geq 21{,}067\right) = 0{,}10$$

g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

b) Distribución muestral de la media para una población normal con varianza σ^2 **desconocida.**

Hasta ahora estábamos admitiendo que se conoce la varianza de la población de la que se extrae la muestra, pero esta no sería la situación general, sino que la mayoría de las veces no conocemos la varianza de la población, entonces cómo se dispone de una muestra aleatoria de tamaño n, podemos, calcular la varianza muestral S^2 para estimar la varianza poblacional σ^2 desconocida.

Cuando la varianza poblacional σ^2 es desconocida la distribución del estadístico $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$

depende del tamaño de la muestra:

a) El tamaño de la muestra es grande $n \geq 30$.

Cuando el tamaño de la muestra es grande, es decir $n \geq 30$, la distribución del estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ sigue siendo aproximadamente $N(0,1)$

b) El tamaño de la muestra es pequeño $n < 30$.

Si el tamaño de la muestra es pequeño, $n < 30$, los valores de la varianza muestral S^2 varían considerablemente de muestra en muestra, pues S^2 disminuye a medida que n aumenta, y la distribución del estadístico ya no sería una distribución normal. Por lo tanto, cuando la varianza poblacional σ^2 es desconocida la distribución del estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$

se ajusta a una **distribución *t* de Student**, con $n-1$ grados de libertad.

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

a) <u>El tamaño de la muestra es grande</u> $n \geq 30$.

Cuando el tamaño de la muestra es grande, es decir $n \geq 30$, la distribución del estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ sigue siendo aproximadamente $N(0,1)$

b) <u>El tamaño de la muestra es pequeño</u> $n < 30$.

Si el tamaño de la muestra es pequeño, $n < 30$, los valores de la varianza muestral S^2 varían considerablemente de muestra en muestra, pues S^2 disminuye a medida que n aumenta, y la distribución del estadístico ya no sería una distribución normal. Por lo tanto, cuando la varianza poblacional σ^2 es desconocida la distribución del estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ se ajusta a una **distribución *t* de Student**, con $n-1$ grados de libertad.

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

Recordemos que si Z es una variable aleatoria con distribución normal $N(0,1)$, V una variable aleatoria con distribución Ji-Cuadrado con n grados de libertad y además Z y V son independientes, entonces la variable aleatoria:

$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n$$

$$t_{n-1} \sim \frac{\frac{\overbrace{\overline{X}-\mu}^{N(0,1)}}{\sigma / \sqrt{n}}}{\sqrt{\left[\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}}\right] \Big/ (n-1)}} =$$

Recordemos que si Z es una variable aleatoria con distribución normal $N(0,1)$, V una variable aleatoria con distribución Ji-Cuadrado con n grados de libertad y además Z y V son independientes, entonces la variable aleatoria:

$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n$$

$$t_{n-1} \sim \frac{\overbrace{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}^{N(0,1)}}{\sqrt{\left[\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}}\right]\Big/(n-1)}} = \frac{\frac{\overline{X}-\mu}{\cancel{\sigma}}\sqrt{n}\,\cancel{\sqrt{n-1}}}{\cancel{\sqrt{n-1}}\;S/\cancel{\sigma}} = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

DISTRIBUCION DE LA PROPORCIÓN MUESTRAL $\overline{p}$

Consideremos una v.a $X \sim B(n, p)$ donde p es la proporción de "éxito" en la población. Para tamaños grandes de n, n>30, la distribución Binomial se aproxima a una distribución normal.

$$X \cong N\left(np, np(1-p)\right)$$

Definamos el estadístico $\hat{p} = \frac{X}{n}$, es el estimador puntual de la proporción poblacional.

La distribución de muestreo de $\hat{p}$ es aproximadamente normal con esperanza p y varianza $\frac{p(1-p)}{n}$ con p no cerca de 0 y 1.

Este resultado se obtiene de la aplicación directa del TLC y el Teorema de Bernouli Generalizado.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Estimación puntual: Es el valor numérico que toma un estimador. Se calcula con los datos de la muestra, del cual se espera que estime un parámetro poblacional.

Parámetro Poblacional	Estimador	Estimación
Media μ	$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Varianza σ^2	$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$
Proporción p	$\hat{p} = \frac{X}{n} = \frac{número éxitos}{número pruebas}$	$\hat{p} = \frac{x}{n}$
Parámetros poblacionales, estimadores y estimaciones		

Estimación de parámetros

Puede haber varios estimadores puntuales diferentes para un mismo parámetro. Por ejemplo, si deseamos estimar la media poblacional, podríamos considerar:

$$\hat{\mu}_1 = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \hat{\mu}_2 = \frac{X_1 + X_n}{2}$$

Teniendo en cuenta esto:

- **¿Qué características queremos que posea un buen estimador?**
- **¿Cómo decimos que un estimador es mejor que otro?**

Estimación de parámetros

Propiedades de los estimadores

Nuestro objetivo ahora será dar algunas propiedades deseables de los estimadores puntuales, con el fin de poder conocer la bondad de los mismos.

Propiedad de insesgadura:

Si tenemos un gran número de muestras de tamaño n y obtenemos el valor del estimador en cada una de ellas, sería deseable que la media de todas estas estimaciones coincidiera con el valor medio de la población . Se dice que un estimador es insesgado si su esperanza matemática coincide con el valor del parámetro a estimar.

Formablemente: Un estimador $\hat{\theta}$ es un **estimador insesgado** del parámetro desconocido θ si $E(\hat{\theta}) = \theta$. Es decir, la media de su distribución muestral es el parámetro.

Ejemplo:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

Estimación de parámetros

Ejemplo:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

Formablemente: Un estimador $\hat{\theta}$ es un **estimador insesgado** del parámetro desconocido θ si $E(\hat{\theta}) = \theta$. Es decir, la media de su distribución muestral es el parámetro.

Ejemplo:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

Ejemplo:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

$$E\left(\overline{X}\right) = E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{\cancel{n}} \cancel{n} \mu = \mu$$

Estimación de parámetros

Veamos ahora que $S^2=\sum_{i=1}^{n}\frac{\left(X_i-\overline{X}\right)^2}{n-1}$ es un estimador insesgado de σ^2, es decir de la varianza poblacional.

$$\boxed{E\left(S^2\right)}=E\left(\sum_{i=1}^{n}\frac{\left(X_i-\overline{X}\right)^2}{n-1}\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i^2-2X_i\overline{X}+\overline{X}^2\right)\right)=$$

$$=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2+\sum_{i=1}^{n}-2X_i\overline{X}+\sum_{i=1}^{n}\overline{X}^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-2\overline{X}\underbrace{\sum_{i=1}^{n}X_i}_{n\overline{X}}+n\overline{X}^2\right)=$$

$$=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-2n\overline{X}\,\overline{X}+n\overline{X}^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-2n\overline{X}^2+n\overline{X}^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)=$$

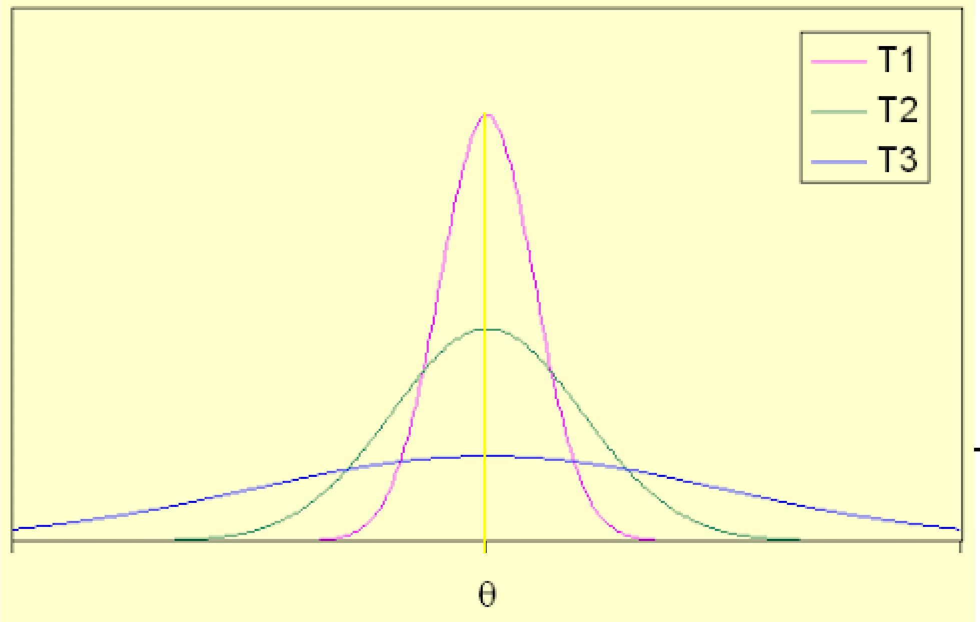
$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}E\left(X_i^2\right)-E\left(n\overline{X}^2\right)\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}E\left(X_i^2\right)-nE\left(\overline{X}^2\right)\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\frac{\sigma^2}{n}\right)\right)=$$

$$=\frac{1}{n-1}\left(\cancel{n\mu^2}+n\sigma^2-\cancel{n\mu^2}-\cancel{n}\left(\frac{\sigma^2}{\cancel{n}}\right)\right)=\frac{1}{n-1}\left(n\sigma^2-\sigma^2\right)=\frac{1}{\cancel{n-1}}\sigma^2\cancel{(n-1)}=\boxed{\sigma^2}$$

Propiedad de eficiencia:

Se dice que los estimadores son eficientes cuando generan una distribución muestral con el mínimo error estándar, es decir, entre dos estimadores insesgados de un parámetro dado es más eficiente el de menor varianza.

Formalmente: Un estimador ***insesgado*** $\hat{\theta}_1$ es más eficiente que otro $\hat{\theta}_2$ Si son insesgados de θ y la varianza de $\hat{\theta}_1$ es menor que la varianza de $\hat{\theta}_2$.



$$E(T_1) = \theta$$
$$E(T_2) = \theta$$
$$E(T_3) = \theta$$

T_1 es el de varianza mínima

Estimador consistente:

Un estimador se dice ***consistente*** cuando su valor tiende hacia el verdadero valor del parámetro a medida que aumenta el tamaño de la muestra. Es decir, la probabilidad de que la estimación sea el verdadero valor del parámetro tiende a 1.

Propiedad de suficiencia:

Se dice de un estimador que es ***suficiente*** cuando es capaz de extraer de los datos toda la información importante sobre el parámetro.

Un estimador es ***suficiente*** si utiliza toda la información de la muestra.

Estimador consistente:

Un estimador se dice ***consistente*** cuando su valor tiende hacia el verdadero valor del parámetro a medida que aumenta el tamaño de la muestra. Es decir, la probabilidad de que la estimación sea el verdadero valor del parámetro tiende a 1.

Propiedad de suficiencia:

Se dice de un estimador que es ***suficiente*** cuando es capaz de extraer de los datos toda la información importante sobre el parámetro.

Un estimador es ***suficiente*** si utiliza toda la información de la muestra.

$$\hat{\mu}_1 = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \hat{\mu}_2 = \frac{X_1 + X_n}{2}$$

Inferencia Estadística

<u>El tamaño de la muestra es grande $n \geq 30$.</u>

Cuando el tamaño de la muestra es grande, es decir $n \geq 30$, la distribución del estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ sigue siendo aproximadamente $N(0,1)$

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

Estimación por Intervalos

Distribuciones de Muestreo.

Distribución de la media muestral proveniente de una distribución normal.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

1) Distribución muestral de la media muestral para una población normal con varianza σ^2 conocida.

$\bar{X} \sim N\left(\mu, \sigma^2/n\right)$	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Distribuciones de Muestreo.

Distribución de la media muestral proveniente de una distribución normal.

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

1) Distribución muestral de la media muestral para una población normal con varianza σ^2 conocida.

$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$	$Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Resultado: Al estudiar la distribución normal se consideraron algunas propiedades que posee dicha distribución, una de ellas era referente a la distribución de una combinación lineal de variables aleatorias normales. Así pues, sabemos que si $X_1, X_2, X_3, \ldots, X_n$ son v.a. independientes distribuidas normalmente $X_i \sim N\left(\mu_i, \sigma_i^2\right)$ $\forall i = 1, \ldots, n$ y si $a_i, \ldots, a_n$ son números reales, entonces la variable aleatoria:

$$X = \sum_{i=1}^{n} a_i X_i$$

Tiene una distribución normal $X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$

Este resultado nos será de utilidad para obtener la distribución de la media muestral,

Estimación por Intervalos

Distribuciones de Muestreo.

Distribución de la media muestral proveniente de una distribución normal.

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

1) Distribución muestral de la media muestral para una población normal con varianza σ^2 conocida.

$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$	$Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

2) Distribución muestral de la media muestral para una población normal con varianza σ^2 desconocida.

a) El tamaño de la muestra es pequeño $n < 30$.		b) El tamaño de la muestra es grande $n \geq 30$.	
$\overline{X} \sim t\left(\mu, S^2/n\right)$	$T = \dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$	$\overline{X} \sim N\left(\mu, S^2/n\right)$	$\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$

Estimación por Intervalos

Distribuciones de Muestreo.

La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

3) **Distribución de la varianza muestral.**

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

Estimación por Intervalos

Distribuciones de Muestreo.

La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

DISTRIBUCION DE LA PROPORCIÓN MUESTRAL $\overline{p}$

Consideremos una v.a $X \sim B(n,p)$ donde p es la proporción de "éxito" en la población. Para tamaños grandes de n, n>30, la distribución Binomial se aproxima a una distribución normal.

$$X \cong N(np, np(1-p))$$

Definamos el estadístico $\hat{p} = \frac{X}{n}$, es el estimador puntual de la proporción poblacional.

La distribución de muestreo de $\hat{p}$ es aproximadamente normal con esperanza p y varianza $\frac{p(1-p)}{n}$, es decir, $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$.

Este resultado se obtiene de la aplicación directa del TLC y el Teorema de Bernouli Generalizado.

Estimación por Intervalos

2022.

Estimación por Intervalos:

En muchas situaciones, una estimación puntual no proporciona información suficiente sobre un parámetro. Por el hecho de ser un solo número no proporciona información sobre la precisión y confiabilidad de la estimación.

Para una muestra de tamaño pequeño, la estimación puntual puede diferenciarse **"bastante"** del parámetro que se estima, es decir, da lugar a errores marcados. Por esta razón hay que utilizar **estimación por intervalos**.

Por ejemplo, si nos interesa estimar la vida media de una lámpara de 75 watts, un solo número puede no tener mucho significado. Un estimación por intervalo podría ser más útil.

Estimación por Intervalos:

Supongamos que la característica numérica $\hat{\theta}$ que surge de los datos de la muestra sirve de estimación del parámetro desconocido θ. Es claro que $\hat{\theta}$ determina con mayor precisión al parámetro θ si:

$$\left|\hat{\theta}-\theta\right|<\delta \quad \forall \delta \to 0$$

Estimación por Intervalos:

Supongamos que la característica numérica $\hat{\theta}$ que surge de los datos de la muestra sirve de estimación del parámetro desconocido θ. Es claro que $\hat{\theta}$ determina con mayor precisión al parámetro θ si:

$$\left|\hat{\theta}-\theta\right|<\delta \quad \forall\delta\to 0$$

Sin embargo, no podemos, en los métodos estadísticos, afirmar categóricamente que la estimación $\hat{\theta}$ satisface la desigualdad $\left|\hat{\theta}-\theta\right|<\delta \quad \forall\delta\to 0$, sólo es posible hablar de:

$$P\left(\left|\hat{\theta}-\theta\right|<\delta\right)=1-\alpha$$

$1-\alpha$ se denomina **nivel de confianza**. En estas circunstancias, α es el llamado **error aleatorio o nivel de significación**.

Estimación por Intervalos:

Supongamos que la característica numérica $\hat{\theta}$ que surge de los datos de la muestra sirve de estimación del parámetro desconocido θ. Es claro que $\hat{\theta}$ determina con mayor precisión al parámetro θ si:

$$\left|\hat{\theta}-\theta\right|<\delta \quad \forall\delta\to 0$$

Sin embargo, no podemos, en los métodos estadísticos, afirmar categóricamente que la estimación $\hat{\theta}$ satisface la desigualdad $\left|\hat{\theta}-\theta\right|<\delta \quad \forall\delta\to 0$, sólo es posible hablar de:

$$P\left(\left|\hat{\theta}-\theta\right|<\delta\right)=1-\alpha$$

$$P\left(\underbrace{\hat{\theta}-\delta}_{L}<\theta<\underbrace{\hat{\theta}+\delta}_{U}\right)=1-\alpha$$

$1-\alpha$ se denomina **nivel de confianza**. En estas circunstancias, α es el llamado **error aleatorio o nivel de significación**.

Estimación por Intervalos:

En la estimación por intervalos se obtienen dos valores ***L*** y ***U*** (un extremo inferior ***L*** y un extremo superior ***U***) que definen un intervalo sobre la recta real, el cual contendrá con "***cierta confianza***" el valor del parámetro desconocido $\boldsymbol{\theta}$. Dado que los valores de ***L*** y ***U*** varían de una muestra a otra debido a la naturaleza aleatoria de los datos ***L*** y ***U*** son en sí mismos estadísticos.

Estimación por Intervalos:

En la estimación por intervalos se obtienen dos valores ***L*** y ***U*** (un extremo inferior ***L*** y un extremo superior ***U***) que definen un intervalo sobre la recta real, el cual contendrá con "***cierta confianza***" el valor del parámetro desconocido ***θ***. Dado que los valores de ***L*** y ***U*** varían de una muestra a otra debido a la naturaleza aleatoria de los datos ***L*** y ***U*** son en sí mismos estadísticos.

Resumiendo, para hallar un intervalo de confianza, debemos hallar los estadísticos L y U tales que:

$$P\left(L \leq \theta \leq U\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$

El nivel de confianza y la amplitud del intervalo varían conjuntamente, de forma que un intervalo más amplio tendrá más probabilidad de acierto (mayor nivel de confianza), mientras que para un intervalo más pequeño, que ofrece una estimación más precisa, aumenta su probabilidad de error. Para la construcción de un determinado intervalo de confianza es necesario conocer la distribución teórica que sigue el parámetro a estimar, θ.

Estimación por Intervalos

Intervalo de confianza para la media poblacional conocida la dispersión poblacional σ.

Si $X \sim N\left(\mu, \sigma^2\right)$ entonces sabemos que $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$ por lo tanto $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\boxed{P\left(L \leq \theta \leq U\right) = \underbrace{1-\alpha}_{\textit{nivel de confianza}}}$$

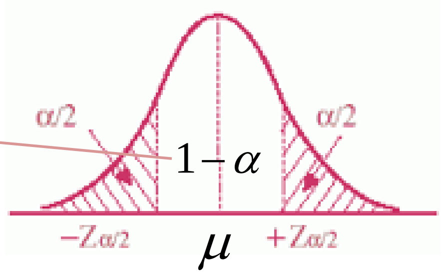
Estimación por Intervalos

Intervalo de confianza para la media poblacional conocida la dispersión poblacional σ.

Si $X \sim N\left(\mu,\sigma^2\right)$ entonces sabemos que $\overline{X} \sim N\left(\mu,\sigma^2/n\right)$ por lo tanto $Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\boxed{P\left(L \le \theta \le U\right) = \underbrace{1-\alpha}_{\textit{nivel de confianza}}}$$

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = \underbrace{1-\alpha}_{\textit{nivel de confianza}}$$



$$\left|\overline{X}-\mu\right| < \delta$$

Donde $Z_{\frac{\alpha}{2}}$ es tal que:

$$P\left(X > Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

Estimación por Intervalos

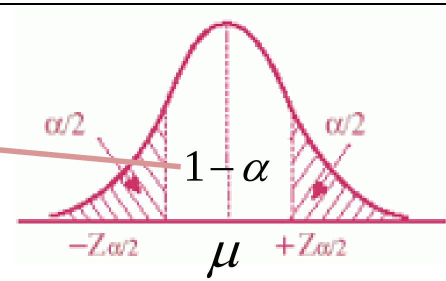
Intervalo de confianza para la media poblacional conocida la dispersión poblacional σ.

Si $X \sim N(\mu, \sigma^2)$ entonces sabemos que $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$ por lo tanto $\boxed{Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}} \sim N(0,1)$

$$\boxed{P\left(L \le \theta \le U\right) = \underbrace{1-\alpha}_{\textit{nivel de confianza}}}$$

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = \underbrace{1-\alpha}_{\textit{nivel de confianza}}$$

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



$$\left|\overline{X} - \mu\right| < \delta$$

Donde $Z_{\frac{\alpha}{2}}$ es tal que:

$$P\left(X > Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

Estimación por Intervalos

<u>Intervalo de confianza para la media poblacional conocida la dispersión poblacional σ.</u>

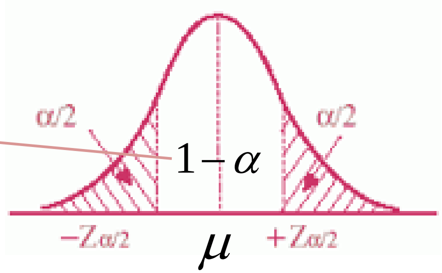
Si $X \sim N(\mu, \sigma^2)$ entonces sabemos que $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$ por lo tanto $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\boxed{P(L \le \theta \le U) = \underbrace{1-\alpha}_{\text{nivel de confianza}}}$$

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$

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

$$P\left(-\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \le -\mu \le -\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$



$$\left|\overline{X} - \mu\right| < \delta$$

Donde $Z_{\frac{\alpha}{2}}$ es tal que:

$$P\left(X > Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

Estimación por Intervalos

<u>Intervalo de confianza para la media poblacional conocida la dispersión poblacional σ.</u>

Si $X \sim N(\mu, \sigma^2)$ entonces sabemos que $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$ por lo tanto $Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

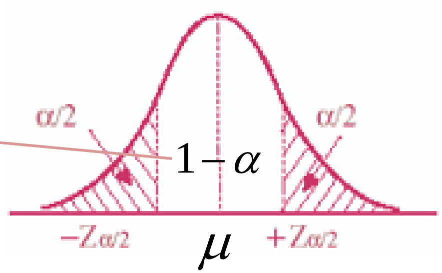
$$\boxed{P(L \le \theta \le U) = \underbrace{1-\alpha}_{\text{nivel de confianza}}}$$

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$

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

$$P\left(-\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \le -\mu \le -\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \le \mu \le \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}\right) = 1-\alpha$$



$$\left|\overline{X} - \mu\right| < \delta$$

Donde $Z_{\frac{\alpha}{2}}$ es tal que:

$$P\left(X > Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

Estimación por Intervalos

<u>Intervalo de confianza para la media poblacional conocida la dispersión poblacional σ.</u>

Si $X \sim N(\mu, \sigma^2)$ entonces sabemos que $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$ por lo tanto $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

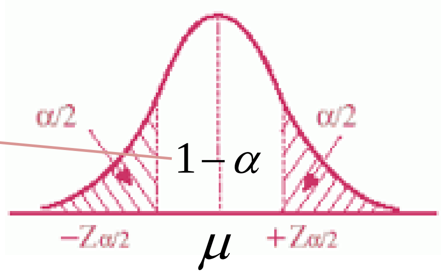
$$\boxed{P(L \le \theta \le U) = \underset{\textit{nivel de confianza}}{1 - \alpha}}$$

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = \underbrace{1-\alpha}_{\textit{nivel de confianza}}$$

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

$$P\left(-\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \le -\mu \le -\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \le \mu \le \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}\right) = 1 - \alpha$$



$$\left|\overline{X} - \mu\right| < \delta$$

Donde $Z_{\frac{\alpha}{2}}$ es tal que:

$$P\left(X > Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

$$\boxed{\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \le \mu \le \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}}$$

Estimación por Intervalos

Ejemplo: Se sabe que la vida media en hs de una lámpara de 75 watts es aproximadamente normal, con dispersión de 25 hs. Una muestra aleatoria de 20 lámparas tiene una vida media de 1014 hs. Construir un intervalo de confianza del 95% respecto de la vida media de las lámparas.

Solución/

Estimación por Intervalos

<u>Ejemplo</u>: Se sabe que la vida media en hs de una lámpara de 75 watts es aproximadamente normal, con dispersión de 25 hs. Una muestra aleatoria de 20 lámparas tiene una vida media de 1014 hs. Construir un intervalo de confianza del 95% respecto de la vida media de las lámparas.

Solución/

Por hipótesis sabemos que:

$$n = 20; \quad \sigma = 25; \quad \overline{X} = 1014; \quad 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \qquad 1 - \frac{\alpha}{2} = 0.975$$

Estimación por Intervalos

<u>Ejemplo</u>: Se sabe que la vida media en hs de una lámpara de 75 watts es aproximadamente normal, con dispersión de 25 hs. Una muestra aleatoria de 20 lámparas tiene una vida media de 1014 hs. Construir un intervalo de confianza del 95% respecto de la vida media de las lámparas.

Solución/

Por hipótesis sabemos que:

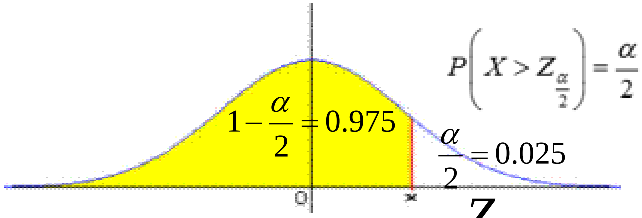
$$n = 20; \quad \sigma = 25; \quad \overline{X} = 1014; \quad 1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \qquad 1-\frac{\alpha}{2} = 0.975$$

Reemplazando en el intervalo que acabamos de construir tenemos que

$$\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \qquad z_{\frac{\alpha}{2}} = 1.96 \qquad P\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

Estimación por Intervalos

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = P(Z \leq z)$$



$Z_{\frac{\alpha}{2}}$

1.96

z	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9278	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9648	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	$\boxed{0,9750}$	0,9756	0,9762	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857

Ejemplo: Se sabe que la vida media en hs de una lámpara de 75 watts es aproximadamente normal, con dispersión de 25 hs. Una muestra aleatoria de 20 lámparas tiene una vida media de 1014 hs. Construir un intervalo de confianza del 95% respecto de la vida media de las lámparas.

Solución/

Por hipótesis sabemos que:

$$n = 20; \quad \sigma = 25; \quad \overline{X} = 1014; \quad 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \qquad 1 - \frac{\alpha}{2} = 0.975$$

Reemplazando en el intervalo que acabamos de construir tenemos que

$$\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \qquad z_{\frac{\alpha}{2}} = 1.96 \qquad P\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

$$1014 - 1.96 \times \frac{25}{\sqrt{20}} \leq \mu \leq 1014 + 1.96 \times \frac{25}{\sqrt{20}} \quad \Rightarrow \quad 1003.043 \leq \mu \leq 1024.956$$

Estimación por Intervalos

Observaciones:

- Sería erróneo escribir $P(1003{,}043 \leq \mu \leq 1024{,}956) = 0{,}95$ ¿por qué?. μ es un parámetro constante, o se encuentra en el intervalo hallado (en tal caso el suceso $1003{,}043 \leq \mu \leq 1024{,}956$ es un suceso cierto y su probabilidad es 1) o bien, dicho parámetro no se encuentra en el intervalo hallado (en tal caso el suceso $1003{,}043 \leq \mu \leq 1024{,}956$ es un suceso imposible y su probabilidad es 0).

- Esto significa que no hay que vincular el nivel de confianza $1-\alpha$ con el parámetro que se estima, ya que está ligada solamente con los límites del intervalo que varían de una muestra a otra.

Estimación por Intervalos

Observaciones:

- Sería erróneo escribir $P(1003{,}043 \le \mu \le 1024{,}956) = 0{,}95$ ¿por qué?. μ es un parámetro constante, o se encuentra en el intervalo hallado (en tal caso el suceso $1003{,}043 \le \mu \le 1024{,}956$ es un suceso cierto y su probabilidad es 1) o bien, dicho parámetro no se encuentra en el intervalo hallado (en tal caso el suceso $1003{,}043 \le \mu \le 1024{,}956$ es un suceso imposible y su probabilidad es 0).
- Esto significa que no hay que vincular el nivel de confianza $1-\alpha$ con el parámetro que se estima, ya que está ligada solamente con los límites del intervalo que varían de una muestra a otra.

Interpretación: aunque nunca sabremos si la media poblacional se encuentra en el intervalo hallado, tendremos la seguridad de que el método utilizado para la obtención de dicho intervalo es confiable el 95 % de las veces, es decir, se puede esperar que contenga a dicho parámetro en el 95 % de las veces.

Estimación por Intervalos

Observaciones:

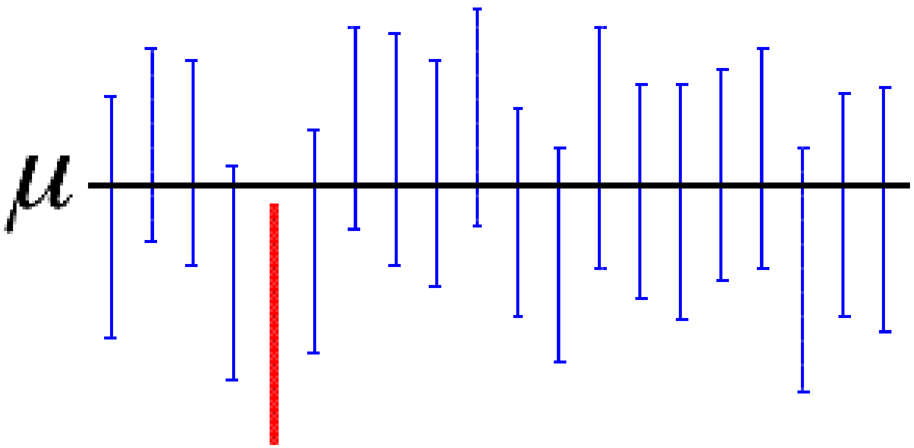
- Sería erróneo escribir $P(1003{,}043 \leq \mu \leq 1024{,}956) = 0{,}95$ ¿por qué?. μ es un parámetro constante, o se encuentra en el intervalo hallado (en tal caso el suceso $1003{,}043 \leq \mu \leq 1024{,}956$ es un suceso cierto y su probabilidad es 1) o bien, dicho parámetro no se encuentra en el intervalo hallado (en tal caso el suceso $1003{,}043 \leq \mu \leq 1024{,}956$ es un suceso imposible y su probabilidad es 0).

- Esto significa que no hay que vincular el nivel de confianza $1-\alpha$ con el parámetro que se estima, ya que está ligada solamente con los límites del intervalo que varían de una muestra a otra.

Interpretación: aunque nunca sabremos si la media poblacional se encuentra en el intervalo hallado, tendremos la seguridad de que el método utilizado para la obtención de dicho intervalo es confiable el 95 % de las veces, es decir, se puede esperar que contenga a dicho parámetro en el 95 % de las veces.

Los niveles de confianza más frecuentemente utilizados son 95%, 99% y 90%.

Estimación por Intervalos



La línea negra horizontal representa el valor fijo de la media desconocida de la población, μ. Los intervalos de confianza azules verticales que se sobreponen a la línea horizontal contienen el valor de la media de la población. El intervalo de confianza rojo que está completamente por debajo de la línea horizontal no lo contiene. Un intervalo de confianza de 95% indica, para este ejemplo gráfico, que 19 de 20 muestras (95%) de la misma población darán intervalos de confianza que contendrán el parámetro de población desconocido.

Intervalo de confianza para la media con varianza poblacional desconocida y $n<30$

Si la población base es normal, la varianza es desconocida y el tamaño de la muestra menor que 30, la media muestral tiene distribución t de Student con $n-1$ grados de libertad. Luego el intervalo de confianza para la media con varianza poblacional desconocida y $n<30$ está dado por:

$$\boxed{\overline{X}-t_{\left(\frac{\alpha}{2};n-1\right)}\times\frac{S}{\sqrt{n}}\leq\mu\leq\overline{X}+t_{\left(\frac{\alpha}{2};n-1\right)}\times\frac{s}{\sqrt{n}}}$$ donde $P\left(T>t_{\left(\frac{\alpha}{2};n-1\right)}\right)=\frac{\alpha}{2}$

Recordemos que bajo las condiciones anteriores:

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

Intervalo de confianza para la media con varianza poblacional desconocida y $n<30$

Si la población base es normal, la varianza es desconocida y el tamaño de la muestra menor que 30, la media muestral tiene distribución t de Student con $n-1$ grados de libertad. Luego el intervalo de confianza para la media con varianza poblacional desconocida y $n<30$ está dado por:

$$\overline{X}-t_{\left(\frac{\alpha}{2};n-1\right)}\times\frac{S}{\sqrt{n}}\leq\mu\leq\overline{X}+t_{\left(\frac{\alpha}{2};n-1\right)}\times\frac{s}{\sqrt{n}} \text{ donde } P\left(T>t_{\left(\frac{\alpha}{2};n-1\right)}\right)=\frac{\alpha}{2}$$

Ejercicio: Deducir la forma del intervalo anterior. Para esto partir de:

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\sim t_{n-1} \qquad P\left(-t_{\left(\frac{\alpha}{2},n-1\right)}\leq\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\leq t_{\left(\frac{\alpha}{2},n-1\right)}\right)=1-\alpha$$

Estimación por Intervalos

Ejemplo: En un estudio hecho para determinar el tiempo medio necesario para el montaje de cierta pieza de una máquina, 25 trabajadores hicieron un promedio de 42.5 minutos con una varianza de 4.1 minutos cuadrados. Si los tiempos de los trabajadores se distribuyen normalmente, hallar un intervalo de confianza al nivel del 99% para el tiempo promedio necesario para el montaje de la máquina.

Estimación por Intervalos

<u>Ejemplo</u>: En un estudio hecho para determinar el tiempo medio necesario para el montaje de cierta pieza de una máquina, 25 trabajadores hicieron un promedio de 42.5 minutos con una varianza de 4.1 minutos cuadrados. Si los tiempos de los trabajadores se distribuyen normalmente, hallar un intervalo de confianza al nivel del 99% para el tiempo promedio necesario para el montaje de la máquina.

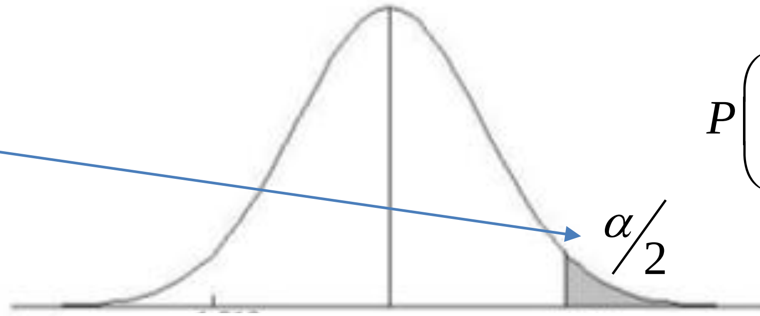
$$n = 25; \quad S^2 = 4.1; \quad \overline{X} = 42.5; \quad 1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$$

$$\overline{X} - t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{s}{\sqrt{n}}$$

Estimación por Intervalos

$$\overline{X} - t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{s}{\sqrt{n}}$$

Puntos de porcentaje de la distribución t



$$P\left(t > t_{\frac{\alpha}{2}}\right) = \alpha/2$$

$$n-1 = 24; \quad S\frac{\alpha}{2} = 0.005$$

$$\pm t_{\alpha/2} =$$

r \ α	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850

Estimación por Intervalos

Intervalo de confianza para la varianza poblacional σ^2:

Si S^2 es la varianza de una muestra aleatoria de tamaño n tomada de una población normal que tiene una varianza σ^2, entonces la variable aleatoria muestral:

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

tiene una distribución ji-cuadrado con $n-1$ grados de libertad.

Estimación por Intervalos

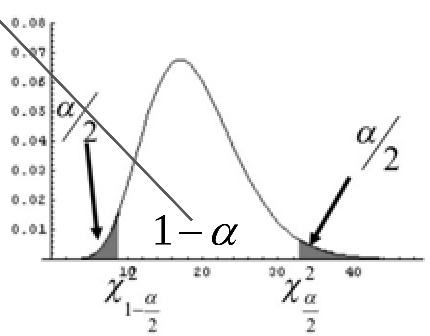
Intervalo de confianza para la varianza poblacional σ^2:

Si S^2 es la varianza de una muestra aleatoria de tamaño n tomada de una población normal que tiene una varianza σ^2, entonces la variable aleatoria muestral:

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

tiene una distribución ji-cuadrado con $n-1$ grados de libertad.

$$P\left(\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\left(\frac{\alpha}{2},n-1\right)}\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$



Estimación por Intervalos

<u>Intervalo de confianza para la varianza poblacional σ^2</u>:

Si S^2 es la varianza de una muestra aleatoria de tamaño n tomada de una población normal que tiene una varianza σ^2, entonces la variable aleatoria muestral:

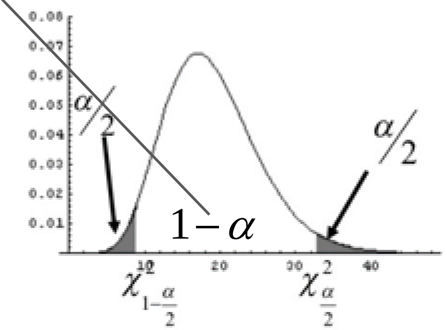
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

tiene una distribución ji-cuadrado con $n-1$ grados de libertad.

$$P\left(\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\left(\frac{\alpha}{2},n-1\right)}\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$

$$P\left(\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\left(\frac{\alpha}{2},n-1\right)}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}\right) = 1-\alpha$$



Intervalo de confianza para la varianza poblacional σ^2:

Si S^2 es la varianza de una muestra aleatoria de tamaño n tomada de una población normal que tiene una varianza σ^2, entonces la variable aleatoria muestral:

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

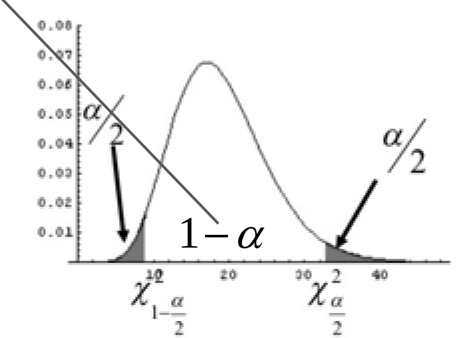
tiene una distribución ji-cuadrado con $n-1$ grados de libertad.

$$P\left(\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\left(\frac{\alpha}{2},n-1\right)}\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$

$$P\left(\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\left(\frac{\alpha}{2},n-1\right)}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}\right) = 1-\alpha$$

$$\boxed{\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}}$$



Ejemplo: Estimar por intervalo, con una confianza del 90 %, la varianza poblacional, si la varianza muestral es de 4,1 en una muestra de tamaño 7.

Estimación por Intervalos

Ejemplo: Estimar por intervalo, con una confianza del 90 %, la varianza poblacional, si la varianza muestral es de 4,1 en una muestra de tamaño 7.

$$S = 4.1, \quad n = 7, \quad 1-\alpha = 0.9 \quad \Rightarrow \quad \alpha = 0.1 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.05$$

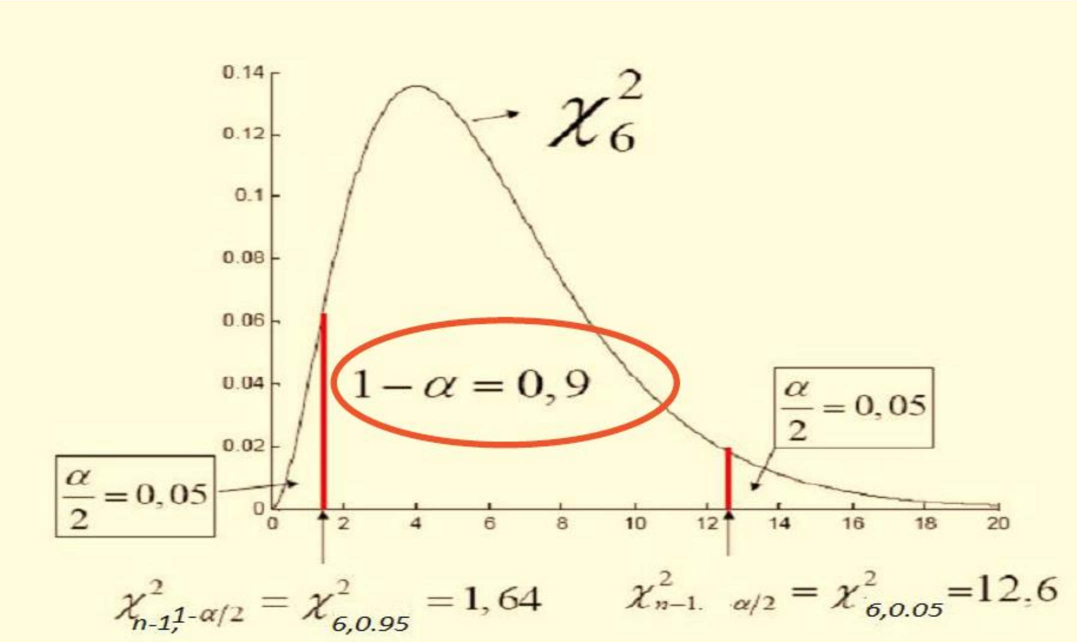
$$\boxed{\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}}$$

Estimación por Intervalos

Ejemplo: Estimar por intervalo, con una confianza del 90 %, la varianza poblacional, si la varianza muestral es de 4,1 en una muestra de tamaño 7.

$$S = 4.1, \quad n = 7, \quad 1-\alpha = 0.9 \quad \Rightarrow \quad \alpha = 0.1 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.05$$

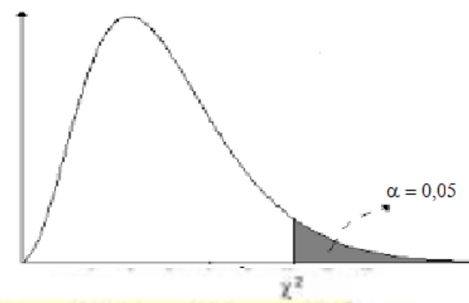
$$\boxed{\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2}, n-1\right)}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2}, n-1\right)}}}$$



$$\chi^2_{(6,0.05)} = 12.592$$

$$\chi^2_{(6,0.95)} = 1.635$$

$$\chi^2_{(6,0.05)} = 12.592 \qquad \chi^2_{(6,0.95)} = 1.635$$



g. d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	$\boxed{0,05}$	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
$\boxed{6}$	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

$$\chi^2_{(6,0.05)} = 12.592 \qquad \chi^2_{(6,0.95)} = 1.635$$

χ^2

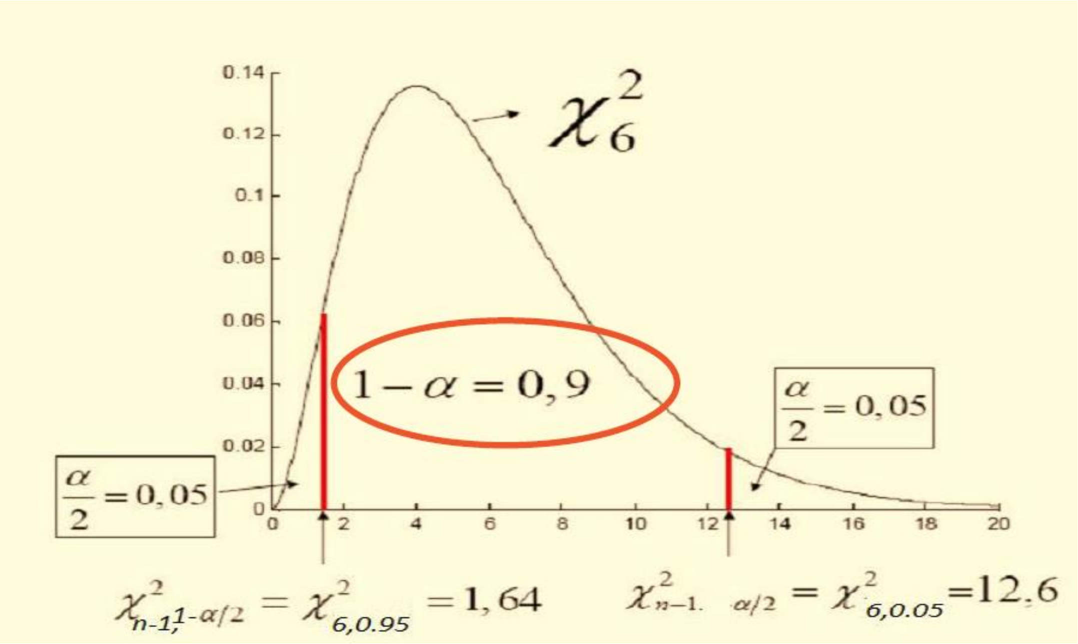
g.d.l	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	0,975	0,98	0,99	0,995	g.d.l
1	0,571	0,455	0,357	0,275	0,206	0,148	0,102	0,064	0,036	0,016	0,004	0,001	0,001	0,000	0,000	1
2	1,597	1,386	1,196	1,022	0,862	0,713	0,575	0,446	0,325	0,211	0,103	0,051	0,040	0,020	0,010	2
3	2,643	2,366	2,109	1,869	1,642	1,424	1,213	1,005	0,798	0,584	0,352	0,216	0,185	0,115	0,072	3
4	3,687	3,357	3,047	2,753	2,470	2,195	1,923	1,649	1,366	1,064	0,711	0,484	0,429	0,297	0,207	4
5	4,728	4,351	3,996	3,655	3,325	3,000	2,675	2,343	1,994	1,610	1,145	0,831	0,752	0,554	0,412	5
$\boxed{6}$	5,765	5,348	4,952	4,570	4,197	3,828	3,455	3,070	2,661	2,204	1,635	1,237	1,134	0,872	0,676	6
7	6,800	6,346	5,913	5,493	5,082	4,671	4,255	3,822	3,358	2,833	2,167	1,690	1,564	1,239	0,989	7
8	7,833	7,344	6,877	6,423	5,975	5,527	5,071	4,594	4,078	3,490	2,733	2,180	2,032	1,646	1,344	8
9	8,863	8,343	7,843	7,357	6,876	6,393	5,899	5,380	4,817	4,168	3,325	2,700	2,532	2,088	1,735	9
10	9,892	9,342	8,812	8,295	7,783	7,267	6,737	6,179	5,570	4,865	3,940	3,247	3,059	2,558	2,156	10
11	10,920	10,341	9,783	9,237	8,695	8,148	7,584	6,989	6,336	5,578	4,575	3,816	3,609	3,053	2,603	11
12	11,946	11,340	10,755	10,182	9,612	9,034	8,438	7,807	7,114	6,304	5,226	4,404	4,178	3,571	3,074	12
13	12,972	12,340	11,729	11,129	10,532	9,926	9,299	8,634	7,901	7,042	5,892	5,009	4,765	4,107	3,565	13
14	13,996	13,339	12,703	12,078	11,455	10,821	10,165	9,467	8,696	7,790	6,571	5,629	5,368	4,660	4,075	14
15	15,020	14,339	13,679	13,030	12,381	11,721	11,037	10,307	9,499	8,547	7,261	6,262	5,985	5,229	4,601	15

Estimación por Intervalos

Ejemplo: Estimar por intervalo con una confianza del 90 % la varianza poblacional, si la varianza muestral es de 4,1 en una muestra de tamaño 7.

$$S = 4.1, \quad n = 7, \quad 1-\alpha = 0.9 \quad \Rightarrow \quad \alpha = 0.1 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.05$$

$$\boxed{\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}}$$



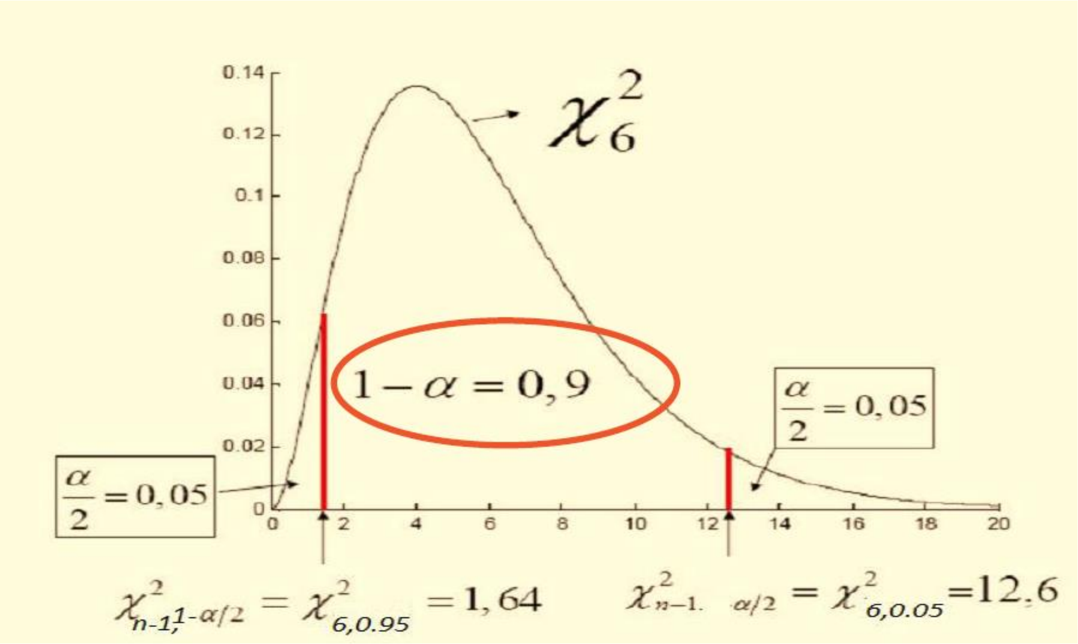
$$\chi^2_{(6,0.05)} = 12.592$$

$$\chi^2_{(6,0.95)} = 1.635$$

$$\frac{6\times 4.1}{12.592} \le \sigma^2 \le \frac{6\times 4.1}{1.635}$$

Ejemplo: Estimar por intervalo con una confianza del 90 % la varianza poblacional, si la varianza muestral es de 4,1 en una muestra de tamaño 7.

$$S^2 = 4.1, \quad n = 7, \quad 1-\alpha = 0.9 \quad \Rightarrow \quad \alpha = 0.1 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.05$$



$$\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}$$

$$\chi^2_{(6,0.05)} = 12.592$$

$$\chi^2_{(6,0.95)} = 1.635$$

$$\frac{6 \times 4.1}{12.592} \leq \sigma^2 \leq \frac{6 \times 4.1}{1.635}$$

$$1.953 \leq \sigma^2 \leq 15,04$$

Intervalo de confianza para la proporción poblacional.

Si $\widehat{p}$ representa la proporción de éxitos en una muestra aleatoria de tamaño n suficientemente grande y $\widehat{q} = 1 - \widehat{p}$, entonces un intervalo de confianza para la proporción poblacional p al nivel de confianza del 100(1-α)% está dado por:

$$\left[\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\left(1-\widehat{p}\right)}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\left(1-\widehat{p}\right)}{n}} \right]$$

en donde $z_{\alpha/2}$ es tal que $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ y la variable aleatoria Z sigue una distribución $N(0,1)$.

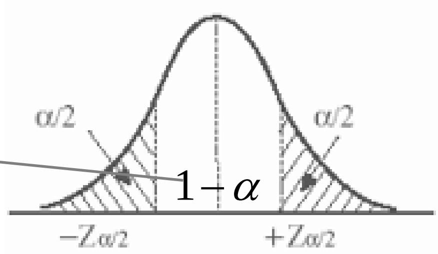
<u>Intervalo de confianza para la proporción poblacional</u>.

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$

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

$$P\left(-\hat{p}-z_{\alpha/2}\times\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \le -p \le -\hat{p}+z_{\alpha/2}\times\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}\right) = 1-\alpha$$

$$P\left(\underbrace{\hat{p}-z_{\alpha/2}\times\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{L} \le p \le \underbrace{\hat{p}+z_{\alpha/2}\times\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{U}\right) = 1-\alpha$$



Donde $Z_{\frac{\alpha}{2}}$ es tal que:

$$P\left(X > Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

Intervalo de confianza para la proporción poblacional.

Ejemplo: En una muestra aleatoria de 75 ejes de árbol, 12 tienen un acabado superficial que es más rugoso que lo permitido por las especificaciones. Una estimación puntual de la proporción de los ejes en la población que excede las especificaciones de rugosidad es $\hat{p} = \frac{X}{n} = \frac{12}{75}$. Construir un intervalo de confianza para p utilizando una confiabilidad del 95%.

Estimación por Intervalos

Intervalo de confianza para la proporción poblacional.

Ejemplo: En una muestra aleatoria de 75 ejes de árbol, 12 tienen un acabado superficial que es más rugoso que lo permitido por las especificaciones. Una estimación puntual de la proporción de los ejes en la población que excede las especificaciones de rugosidad es $\hat{p} = \frac{X}{n} = \frac{12}{75}$. Construir un intervalo de confianza para p utilizando una confiabilidad del 95%.

$$1-\alpha = 0.95 \quad \Rightarrow \quad \alpha = 0.05 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.025 \quad Z_{\frac{\alpha}{2}} = 1.96 \quad \hat{p} = \frac{X}{n} = \frac{12}{75},$$

Intervalo de confianza para la proporción poblacional.

Ejemplo: En una muestra aleatoria de 75 ejes de árbol, 12 tienen un acabado superficial que es más rugoso que lo permitido por las especificaciones. Una estimación puntual de la proporción de los ejes en la población que excede las especificaciones de rugosidad es $\hat{p} = \frac{X}{n} = \frac{12}{75}$. Construir un intervalo de confianza para p utilizando una confiabilidad del 95%.

$$1-\alpha = 0.95 \quad \Rightarrow \quad \alpha = 0.05 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.025 \quad Z_{\frac{\alpha}{2}} = 1.96 \quad \hat{p} = \frac{X}{n} = \frac{12}{75},$$

$$\underbrace{\hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{L} \le p \le \underbrace{\hat{p} + z_{\alpha/2} \times \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{U}$$

Intervalo de confianza para la proporción poblacional.

Ejemplo: En una muestra aleatoria de 75 ejes de árbol, 12 tienen un acabado superficial que es más rugoso que lo permitido por las especificaciones. Una estimación puntual de la proporción de los ejes en la población que excede las especificaciones de rugosidad es $\hat{p} = \frac{X}{n} = \frac{12}{75}$. Construir un intervalo de confianza para p utilizando una confiabilidad del 95%.

$$1-\alpha = 0.95 \quad \Rightarrow \quad \alpha = 0.05 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.025 \quad Z_{\frac{\alpha}{2}} = 1.96 \quad \hat{p} = \frac{X}{n} = \frac{12}{75},$$

$$\underbrace{\hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{L} \le p \le \underbrace{\hat{p} + z_{\alpha/2} \times \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{U}$$

$$\underbrace{\frac{12}{75} - 1.96 \times \sqrt{\frac{\frac{12}{75}\left(1-\frac{12}{75}\right)}{75}}}_{L} \le p \le \underbrace{\frac{12}{75} + 1.96 \times \sqrt{\frac{\frac{12}{75}\left(1-\frac{12}{75}\right)}{75}}}_{U}$$

Intervalo de confianza para la proporción poblacional.

Ejemplo: En una muestra aleatoria de 75 ejes de árbol, 12 tienen un acabado superficial que es más rugoso que lo permitido por las especificaciones. Una estimación puntual de la proporción de los ejes en la población que excede las especificaciones de rugosidad es $\hat{p} = \frac{X}{n} = \frac{12}{75}$. Construir un intervalo de confianza para p utilizando una confiabilidad del 95%.

$$1-\alpha = 0.95 \quad \Rightarrow \quad \alpha = 0.05 \quad \Rightarrow \quad \frac{\alpha}{2} = 0.025 \quad Z_{\frac{\alpha}{2}} = 1.96 \quad \hat{p} = \frac{X}{n} = \frac{12}{75},$$

$$\underbrace{\hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{L} \le p \le \underbrace{\hat{p} + z_{\alpha/2} \times \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}_{U}$$

$$\underbrace{\frac{12}{75} - 1.96 \times \sqrt{\frac{\frac{12}{75}\left(1-\frac{12}{75}\right)}{75}}}_{L} \le p \le \underbrace{\frac{12}{75} + 1.96 \times \sqrt{\frac{\frac{12}{75}\left(1-\frac{12}{75}\right)}{75}}}_{U}$$

$$0.077 \le p \le 0.243$$

Nivel de confianza, precisión, error y tamaño de muestra

1) Si se considera que el ancho del intervalo especifica su precisión, entonces el nivel de confianza está relacionado de manera inversa con la precisión.

$$P\left(L \leq \theta \leq U\right) = \underbrace{1-\alpha}_{\text{nivel de confianza}}$$

El nivel de confianza y la amplitud del intervalo varían conjuntamente, de forma que un intervalo más amplio tendrá más probabilidad de acierto (mayor nivel de confianza), mientras que para un intervalo más pequeño, que ofrece una estimación más precisa, aumenta su probabilidad de error. Para la construcción de un determinado intervalo de confianza es necesario conocer la distribución teórica que sigue el parámetro a estimar, θ.

Nivel de confianza, precisión, error y tamaño de muestra

1) Si se considera que el ancho del intervalo especifica su precisión, entonces el nivel de confianza está relacionado de manera inversa con la precisión.

2) Generalmente $\overline{X}$ no será exactamente igual a μ y entonces se comete un error de muestreo,

$$\varepsilon = \underbrace{\left|\overline{X} - \mu\right|}_{error\ de\ muestreo}$$

Nivel de confianza, precisión, error y tamaño de muestra

1) Si se considera que el ancho del intervalo especifica su precisión, entonces el nivel de confianza está relacionado de manera inversa con la precisión.

2) Generalmente $\overline{X}$ no será exactamente igual a μ y entonces se comete un error de muestreo,

$$\varepsilon = \underbrace{\left|\overline{X} - \mu\right|}_{error\ de\ muestreo}$$

Por lo que vimos anteriormente el intervalo de confianza al nivel del 100(1-α)% para la media poblacional está dado por:

$$\overline{X} - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

Nivel de confianza, precisión, error y tamaño de muestra

1) Si se considera que el ancho del intervalo especifica su precisión, entonces el nivel de confianza está relacionado de manera inversa con la precisión.

2) Generalmente $\overline{X}$ no será exactamente igual a μ y entonces se comete un error de muestreo,

$$\varepsilon = \underbrace{\left|\overline{X} - \mu\right|}_{error\ de\ muestreo}$$

Por lo que vimos anteriormente el intervalo de confianza al nivel del 100(1-α)% para la media poblacional está dado por:

$$\overline{X} - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \left|\overline{X} - \mu\right| \leq z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

Nivel de confianza, precisión, error y tamaño de muestra

1) Si se considera que el ancho del intervalo especifica su precisión, entonces el nivel de confianza está relacionado de manera inversa con la precisión.

2) Pero generalmente $\overline{X}$ no será exactamente igual a μ y entonces se comete un error de muestreo,

$$\varepsilon = \underbrace{\left|\overline{X} - \mu\right|}_{error\ de\ muestreo}$$

Por lo que vimos anteriormente el intervalo de confianza al nivel del 100(1-α)% para la media poblacional está dado por:

$$\overline{X} - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \Rightarrow \left|\overline{X} - \mu\right| \leq z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

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

Determinación del tamaño de muestra

3) De la expresión anterior se puede determinar el tamaño de la muestra que se necesita para lograr una cierta precisión.

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

Despejando el valor de ***n*** tenemos que:

$$n = \left[\frac{z_{\frac{\alpha}{2}} \times \sigma}{\varepsilon}\right]^2$$

Para utilizar esta fórmula, se debe conocer el nivel de confianza 1 – α, el error de muestreo y σ.

Resumiendo

Estimación por Intervalos

1) **Intervalo de confianza para la media poblacional de una población normal con varianza σ^2 conocida.**

$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$	$\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \leq \mu \leq \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}$

Estimación por Intervalos

1) Intervalo de confianza para la media poblacional de una población normal con varianza σ^2 conocida.

$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$	$\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \le \mu \le \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}$

2) Intervalo de confianza para la media poblacional de una población normal varianza σ^2 desconocida.

a) El tamaño de la muestra es pequeño $n < 30$.	**b) El tamaño de la muestra es grande $n \ge 30$.**

Estimación por Intervalos

1) **Intervalo de confianza para la media poblacional de una población normal con varianza σ^2 conocida.**

$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$	$\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \leq \mu \leq \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}$

2) **Intervalo de confianza para la media poblacional de una población normal varianza σ^2 desconocida.**

a) **El tamaño de la muestra es pequeño $n < 30$.**	b) **El tamaño de la muestra es grande $n \geq 30$.**

Estimación por Intervalos

1) **Intervalo de confianza para la media poblacional de una población normal con varianza σ^2 conocida.**

$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$	$\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \leq \mu \leq \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}$

2) **Intervalo de confianza para la media poblacional de una población normal varianza σ^2 desconocida.**

$\boxed{\text{a) El tamaño de la muestra es pequeño } n < 30.}$		b) **El tamaño de la muestra es grande** $n \geq 30$.
$\overline{X} \sim t\left(\mu, S^2/n\right)$	$\boxed{\overline{X} - t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{s}{\sqrt{n}}}$	

Estimación por Intervalos

1) **Intervalo de confianza para la media poblacional de una población normal con varianza σ^2 conocida.**

$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$	$\underbrace{\overline{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{L} \leq \mu \leq \underbrace{\overline{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}_{U}$

2) **Intervalo de confianza para la media poblacional de una población normal varianza σ^2 desconocida.**

a) **El tamaño de la muestra es pequeño $n < 30$.**		b) **El tamaño de la muestra es grande $n \geq 30$.**	
$\overline{X} \sim t\left(\mu, S^2/n\right)$	$\overline{X} - t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\left(\frac{\alpha}{2};n-1\right)} \times \frac{s}{\sqrt{n}}$	$\overline{X} \sim N\left(\mu, S^2/n\right)$	$\boxed{\overline{X} - z_{\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}}$

Estimación por Intervalos

3) Intervalo de confianza para la varianza poblacional.

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$	$\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}$

Estimación por Intervalos

3) Intervalo de confianza para la varianza poblacional.

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$	$\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}$

4) Intervalo de confianza para la proporción

Si se toma una muestra aleatoria de tamaño ***n*** de una gran población donde X observaciones en esta muestra pertenecen a la clase de interés de un suceso A.

$\tilde{p} \sim N\left(p, \frac{(1-p)}{n}\right)$	$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$

Estimación por Intervalos

Intervalo de confianza para la diferencia de medias poblacionales.

a) Si σ es conocida

$$\left(\overline{X}-\overline{Y}\right)-z_{\frac{\alpha}{2}}\times\sqrt{\frac{\sigma^2}{n_X}+\frac{\sigma^2}{n_Y}}\leq\mu_1-\mu_2\leq\left(\overline{X}-\overline{Y}\right)+z_{\frac{\alpha}{2}}\times\sqrt{\frac{\sigma^2}{n_X}+\frac{\sigma^2}{n_Y}}$$

b) Si σ NO es conocida

$$\left(\overline{X}-\overline{Y}\right)-t_{\left(\frac{\alpha}{2},n_X+n_Y-2\right)}\times S_P\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\leq\mu_1-\mu_2\leq\left(\overline{X}-\overline{Y}\right)+t_{\left(\frac{\alpha}{2},n_X+n_Y-2\right)}\times S_P\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}$$

$$\left(S_P\right)^2=\frac{\left[\left(n_x-1\right)S_x^2+\left(n_y-1\right)S_y^2\right]}{n_x+n_y-2}$$