UTN
Facultad Regional Córdoba

TUP

Tecnicatura Universitaria en Programación

ESTADÍSTICA

Unidad Temática N° 5:
Teoría de Muestreo. Estimación

Guía de Estudio
1° Año – 2° Cuatrimestre

Índice

Ejercicio N° 1

Las empresas de cierta región se encuentran clasificadas de acuerdo al nivel de consumo de energía eléctrica:

Estrato	Consumo	Cantidad de empresas	Desviación estándar
I	Bajo	150	20,1
II	Medio	500	8,3
III	Alto	180	3,0
		N= 830	

a) Obtener una muestra de 96 empresas de acuerdo a los tres tipos de afijación conocidos

b) ¿Cuál de los tres criterios es el más adecuado en el caso planteado? ¿Por qué?

Resolución:

Clase de silo	Ni	σi	Ni σi	Afijación Igual	Afijación Proporcional	Afijación Optima
I	150	20,1	150 * 20,1 = 3015	**32**	**17**	**37**
II	500	8,3	500 * 8,3 = 4150	**32**	**58**	**52**
III	180	3,0	180 * 3,0 = 540	**32**	**21**	**7**
	N= 830		∑= 7705	n = 96	n = 96	n = 96

Afijación Igual

$$\boxed{n_1 = n_2 = n_3 = \frac{n}{r} = \frac{96}{3} = 32}$$

Afijación Proporcional

$$\boxed{n_1 = n\frac{N_1}{N} = 96\frac{150}{830} = 17{,}34 \cong 17}$$

$$\boxed{n_2 = n\frac{N_2}{N} = 96\frac{500}{830} = 57{,}83 \cong 58}$$

$$\boxed{n_3 = n\frac{N_3}{N} = 96\frac{180}{830} = 20{,}81 \cong 21}$$

Afijación Optima

$$\boxed{n_1 = 96\frac{3015}{7705} = 37{,}565 \cong 37}$$

$$\boxed{n_2 = 96\frac{4150}{7705} = 51{,}71 \cong 52}$$

$$\boxed{n_3 = 96\frac{540}{7705} = 6{,}72 \cong 7}$$

Ejercicio Nº 2

Un equipo de investigación médica necesita realizar la estimación del peso promedio de los recién nacidos en un hospital con un error máximo tolerado de 0,15 kg. y un riesgo del 10%. Se conoce que la desviación estándar es de 0,5 kg.

a) Calcular el tamaño de muestra requerido para cumplir con los requerimientos de la investigación mencionados. Realizar la correspondiente interpretación.

b) Calcular e interpretar el error máximo tolerado para una muestra de tamaño 81. La desviación estándar y el riesgo son los indicados originalmente.

Respuesta:

a) Datos para calcular el tamaño de la muestra riesgo del 10%: α=0,10

error máximo tolerado: e=0,15

desviación estándar: σ=0,5

En la tabla normal se busca el valor de **z** que corresponde a una probabilidad

1-α/2, previamente debe calcularse dicha probabilidad con la cual se ingresará a la tabla normal:

1 – α/2= 1 – 0,10/2

= 1 – 0,05

1 – α /2 = **0,95**

Posteriormente se busca en la Tabla Normal el valor de **z** que corresponde a la probabilidad **0,95**

z	Probabilidad
1,645	**0,95**

Entonces el valor de z que corresponde a una probabilidad 1– α / 2 es **1,645**

Se reemplazan en la fórmula los datos y el valor de **z** encontrado en la tabla:

$$n = \frac{(z)^2\sigma^2}{e^2}$$

$$= \frac{(1{,}645)^2(0{,}5)^2}{0{,}15^2}$$

n = 5,48

n ≥ 6

La muestra deben incluirse 6 niños.

b) Datos:

riesgo del 10%: **α=0,10**

desviación estándar: σ=0,5

tamaño de la muestra: n=81

Al igual que en el punto anterior α=0,10 por lo tanto en la tabla se obtendrá el mismo valor de z: **z= 1,645**

$$e = \frac{(z)\sigma}{\sqrt{n}}$$

$$= \frac{(1{,}645)0{,}5}{\sqrt{81}}$$

e= 0,091

Interpretación: Diferencia entre el peso promedio en la muestra de 81 recién nacidos y el peso promedio de todos los recién nacidos en un hospital.

Ejercicio N° 3

Se desea estimar la distancia promedio diaria que recorren todos los viajantes de una empresa. Para ello es seleccionada una muestra de 36 viajantes y en dicha muestra se obtuvo una media de 234 km. y una desviación de 9,7 km. Realizar una estimación por intervalos de la **verdadera media** con un nivel de confianza del 95%.

$\boldsymbol{\sigma}$ desconocida y n=36 (n>30)por lo tanto la media muestral tiene distribución normal.

$\hat{S}$**:** Desviación de las distancias recorridas por los 36 viajantes incluidos en la muestra.

Desviación de las distancias recorridas por los 36 viajantes incluidos en la muestra.

$\bar{x}$: Media muestral

$\bar{x}$: distancia promedio recorrida por los 36 viajantes incluidos en la muestra

$\bar{x}$=234 Estimación puntual

Estimación puntual de media: Es el valor de la media en la muestra seleccionada.

$$1-\alpha = \Pr\{\bar{x} - e \leq \mu \leq \bar{x} + e\}$$

$$1-\alpha = \Pr\{\bar{x} - z * \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + z * \frac{\hat{s}}{\sqrt{n}}\}$$

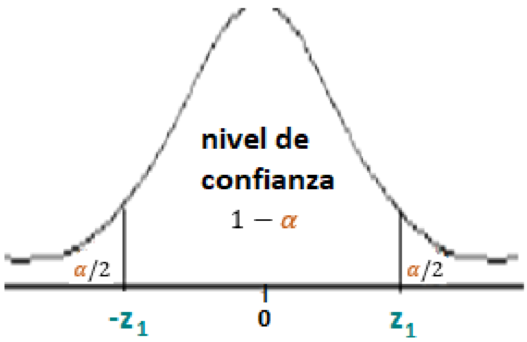
Datos:

n=36

$\overline{x} = 234$

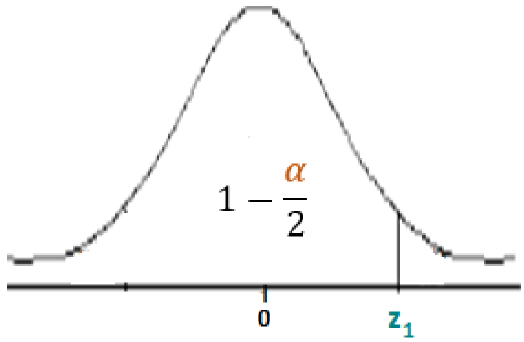
$\hat{S} = 9,7$

Confianza del 95% → 1-α= 0,95 es la probabilidad de que el intervalo obtenido contenga la **media poblacional**



En tabla Normal encontramos la probabilidad:

$P(z < z_1) = 1 - \alpha/2$



1-α= 0,95

α= 1-0,95

riesgo: α= 0,05

1-α/2= 1-0,05/2

1-α/2=0,975 Se ingresa a tabla normal con esta probabilidad para encontrar z

0,14	0,5557	**0,74**	0,7704	**1,34**	0,9099	**1,94**	0,9738	**2,54**	0,9945	**3,14**	0,9992
0,15	0,5596	**0,75**	0,7734	**1,35**	0,9115	**1,95**	0,9744	**2,55**	0,9946	**3,15**	0,9992
0,16	0,5636	**0,76**	0,7764	**1,36**	0,9131	**1,96**	0,9750	**2,56**	0,9948	**3,16**	0,9992
0,17	0,5675	**0,77**	0,7794	**1,37**	0,9147	**1,97**	0,9756	**2,57**	0,9949	**3,17**	0,9992

$P(z < z_1) = 0,975$ entonces $z_1 = 1,96$

Tabla resumida:

1-α = 0,90	α = 0,10	1- α/2= 0,95	z_1= 1,645

$1-\alpha = 0{,}95$	$\alpha = 0{,}05$	$1-\alpha/2 = 0{,}975$	$z_1 = 1{,}96$

$1-\alpha = \Pr\{\bar{x} - e \le \mu \le \bar{x} + e\}$

$234 - 1{,}96 * \frac{9{,}7}{\sqrt{36}} \le \mu \le 234 + 1{,}96 * \frac{9{,}7}{\sqrt{36}}$

$234 - 3{,}17 \le \mu \le 234 + 3{,}17$

$230{,}83 \le \mu \le 237{,}17$

Interpretación del nivel de confianza en el caso planteado:

Es la probabilidad de que el intervalo obtenido contenga la **media poblacional**.

Es la probabilidad de que el intervalo obtenido contenga el verdadera distancia promedio (**distancia promedio diaria que recorren todos los viajantes de una empresa**)

De cada 100 intervalos obtenidos, 95 si contendrán la **distancia promedio de todos los viajantes de una empresa**, y 5 no la contendrán.

El intervalo obtenido puede ser uno de los 95 que si contienen distancia promedio de todos los viajantes de una empresa o uno de 5 que no la contienen.

Ejercicio N° 4

Se desea estimar el consumo promedio diario de energía eléctrica de las empresas de cierta región con un error máximo aceptable de 9 KW., una desviación estándar de 52 KW. y un riesgo del 10%. Determinar el tamaño de muestra adecuado a estos requerimientos.

$$z_1 = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Si se está dispuesto a aceptar que la diferencia | $\bar{x}$ - μ |sea a lo sumo igual a **e**, entonces se reemplaza la diferencia por **e** a la diferencia:

$$z_1 = \frac{e}{\frac{\sigma}{\sqrt{n}}}$$

Si se resuelve el cociente, se obtendrá la siguiente expresión:

$$z_1 = \frac{e\sqrt{n}}{\sigma}$$

Fórmula para calcular el tamaño de la muestra:

Se despeja la fórmula que permite determinar “n” a partir de la expresión correspondiente a z:

$z = \dfrac{e\sqrt{n}}{\sigma}$

Se despeja la incógnita que es n:

$$e\sqrt{n} = z.\boldsymbol{\sigma}$$

$$\sqrt{n} = \frac{z.\sigma}{e}$$

$$n = \frac{z^2.\sigma^2}{e^2}$$

Resolución:

En la tabla normal se busca el valor de **z** que corresponde a una probabilidad 1- α/2, previamente debe calcularse dicha probabilidad:

riesgo del 10%: α =0,10

Calculamos 1 – α/2:

1 – α/2 = 1 – 0,10/2

= 1 – 0,05

1 – α/2 = 0,95

Posteriormente se busca en la Tabla Normal el valor de **z** que corresponde a una probabilidad **0,95**

0,44	0,6700	1,04	0,8508	1,64	0,9495	2,24	0,9875	2,84	0,997
0,45	0,6736	1,05	0,8531	$\boxed{1,65}$	$\boxed{0.9505}$	2,25	0,9878	2,85	0,997
0,46	0,6772	1,06	0,8554	1,66	0,9515	2,26	0,9881	2,86	0,997

Entonces el valor de **z** que corresponde a una probabilidad 1 – α/2 es **1,645**

P(z<1,645)= 0,95 Se reemplaza este valor de **z= 1,645** **en la fórmula para determinar el tamaño de muestra.**

En la fórmula para calcular el tamaño de la muestra también se reemplazan los valores del error y de la desviación que vienen dados como datos:

$$n = \frac{z^2\sigma^2}{e^2}$$

$$= \frac{(1,645)^2(52)^2}{9^2}$$

n = 90,33

El tamaño de muestra siempre se redondea al entero inmediato superior, cualquiera sea la cifra decimal.

n ≥ 91

Ejercicio N° 5

Se desea estimar la distancia promedio diaria que recorren los viajantes de una empresa. con un error máximo aceptable de 7 km., una desviación estándar de 35 km. y un riesgo del 10%. Determinar el tamaño de muestra es el adecuado a estos requerimientos

Realizar una estimación por intervalos de la proporción de empleados que trabajan en las industrias textiles de cierta región y que perciben más de $54000 mensuales con una confianza del 95%

éxito: salario mayor que $54000

x: cantidad de empleados con salarios mayores que $54000

x= 34

$$\hat{p} = \frac{x}{n} = \frac{34}{97}$$

$$\hat{p} = 0,3505$$

Se busca el valor de **z** para una probabilidad 1 – α/2.

Previamente debe calcularse 1 – α/2, para ello se parte del Nivel de Confianza:

Nivel de Confianza: 1 – α = 0,95

Se despeja α:

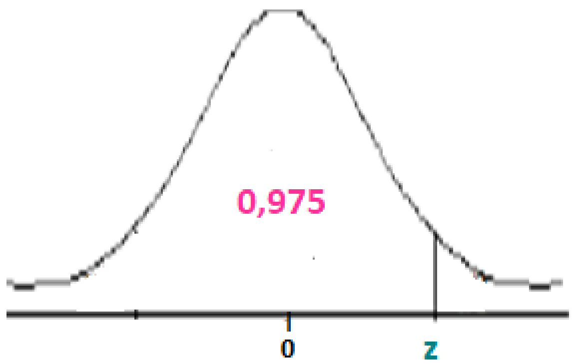
α = 1- 0,95

α = 0,05

1 – α/2= 1 – 0,05/2

= 1 – 0,025

1 – α /2 = **0,975**



Se busca el valor de **z** en la Tabla Normal o en Infostat:

0,13	0,5517	0,73	0,7673	1,33	0,9082	1,93	0,9732	2,53	0,9943	3,13	0,9991
0,14	0,5557	0,74	0,7704	1,34	0,9099	1,94	0,9738	2,54	0,9945	3,14	0,9992
0,15	0,5596	0,75	0,7734	1,35	0,9115	1,95	0,9744	2,55	0,9946	3,15	0,9992
0,16	0,5636	0,76	0,7764	1,36	0,9131	1,96	0,9750	2,56	0,9948	3,16	0,9992
0,17	0,5675	0,77	0,7794	1,37	0,9147	1,97	0,9756	2,57	0,9949	3,17	0,9992
0,18	0,5714	0,78	0,7823	1,38	0,9162	1,98	0,9761	2,58	0,9951	3,18	0,9993
0,19	0,5753	0,79	0,7852	1,39	0,9177	1,99	0,9767	2,59	0,9952	3,19	0,9993
0,20	0,5793	0,80	0,7881	1,40	0,9192	2,00	0,9772	2,60	0,9953	3,20	0,9993

$$1-\alpha=\{\hat{p}-e<P<\hat{p}+e\}$$

$$1-\alpha=\{\hat{p}-z*\sigma_{\hat{p}}<P<\hat{p}+z*\sigma_{\hat{p}}\}$$

$$1-\alpha=\left\{\hat{p}-z*\sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}<P<\hat{p}+z*\sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}\right\}$$

$$0,35-1,96*\sqrt{\frac{0,35*(1-0,35)}{97}}<P<0,35+1,96*\sqrt{\frac{0,35*(1-0,35)}{n}}$$

$$0,35-1,96*0,048<P<0,35+1,96*0,048$$

$$0,35-0,09408<P<0,35+0,09408$$

$$0,25592<\mathrm{P}<0,44408$$

Conclusión:

Con un 95% de confianza se puede afirmar que la proporción de empleados que trabajan en las industrias textiles de cierta región y que perciben más de $54000 se encuentra comprendida entre 0,25592 y 0,44408

De un total de 100 intervalos, 95 si contendrán la verdadera proporción de empleados que trabajan en las industrias textiles de cierta región y que perciben más de $54000, y 5 no la contendrán. El intervalo obtenido puede ser uno de los 95 que si contienen la verdadera proporción o puede ser uno de los 5 que no la contienen.