Facultad de Ingeniería

Departamento de Ciencias Básicas
Probabilidad y Estadística
Apuntes de Clase Semana 12

# APUNTES DE CLASE

30 de enero – 3 de febrero de 2023

- Estas notas de clase son las realizadas en los encuentros sincrónicos.
- Cada vez que se realice un nuevo encuentro el documento se irá retroalimentando.
- Si encuentran algún error por favor háganmelo saber para ir mejorando el documento.
- En algunos casos el documento tendrá información extra que sirva como complemento.

Profesor: Erik Petrovish Navarro Barón

# TAMAÑO DE LA MUESTRA

Es importante saber la cantidad de elementos que deben conformar una muestra a partir de una población, principalmente porque estás muestras deben cumplir ciertas características:

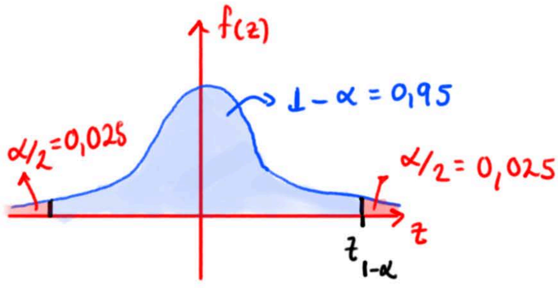
- Representativa: todos y cada uno de los elementos de la población debe tener la misma oportunidad de ser tomado en cuenta para conformar la muestra.
- Adecuada y válida: el error de la muestra debe ser mínimo respecto al de la población.
- Confiabilidad: el tamaño de la muestra debe obtenerse mediante algún proceso matemático que elimine la incidencia del error.

## Cálculo del tamaño de muestras

El cálculo del tamaño de una muestra no es sencillo en términos generales, sin embargo, cualquier formula que se utilice para dicho cálculo debe tener los siguientes tres factores:

- Un porcentaje de confianza con el que se desea generalizar los datos de la muestra a la población. Este porcentaje se representa por $1-\alpha$, y está estrechamente relacionado al concepto de intervalo de confianza. Para calcular la muestra se usa el puntaje $z_{1-\alpha}$. Se acostumbra usar un porcentaje de confianza de $1-\alpha = 95\% = 0{,}95$. En ese caso el valor de $z_{1-\alpha} = 1{,}96$ (este se encuentra haciendo uso de las tablas de puntaje z.

$$1-\alpha = 0{,}95 \rightarrow \alpha = 1 - 0{,}95$$
$$\alpha = 0{,}05$$




$$1-\frac{\alpha}{2} = 0{,}975$$

Usando la tabla del puntaje z obtenemos $\rightarrow z_{1-\alpha} = 1{,}96$

- Un porcentaje de error permitido para aceptar la generalización. Este porcentaje de error lo representamos mediante la letra $\varepsilon$ y esta estrechamente relacionado con las pruebas de hipótesis. Usualmente ese porcentaje de error estará entre 0 y 0,1. Mientras más pequeña sea la muestra más grande será el porcentaje de error.

- Nivel de variabilidad, la probabilidad con la que se presentan los fenómenos de estudio y esos valores son denotados por $p$ y $q = 1-p$, donde $p$ es el porcentaje de confiabilidad. Cuando no se tienen antecedentes sobre la investigación (no hay información previa, o no se pudo aplicar una prueba previa) entonces el valor de variabilidad se considera máximo, este valor se obtiene cuando $p = q = 0{,}5 = \sigma$.

Caso 1: Tamaño de muestra cuando no se conoce N o la población es infinita.

Cuando no se conoce $n$ y las observaciones presentan normalidad el tamaño de la muestra para estimar la medida es:

$$n \geq \frac{p(1-p)\, Z_{1-\alpha}^2}{\varepsilon^2}$$

Ejemplo: Un ingeniero de control de calidad de una linea de producción de envases de cerveza debe tomar una muestra diaria de la línea de producción para inspeccionarla. Decide que su estudio tenga una confianza del 96% y permite un error del 5%. Calcular el tamaño de muestra para inspeccionar la producción

a) Cuando inicia y no tiene información previa.

b) Cuando tiene información previa de varias revisiones diarias con las que se ha obtenido una variabilidad positiva del 0,75.

Tenemos en común que se desea un 96% de confianza. Calcular el $z_{1-\alpha}$, que debe ser central, por lo que realmente se busca $\alpha/2$

$$1-\alpha = 0{,}96 \longrightarrow \alpha = 0{,}04 \longrightarrow \frac{\alpha}{2} = 0{,}02.$$

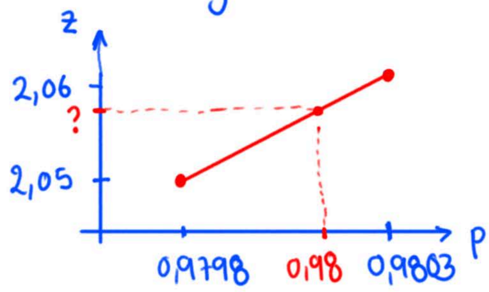
Por lo tanto usaremos $1-\alpha/2 = 0{,}98$.

Buscando en la tabla tenemos los siguientes valores:

$P_1 = 0{,}9798 \longrightarrow z_1 = 2{,}05$

$p = 0{,}98 \longrightarrow z = ?$

$P_2 = 0{,}9803 \longrightarrow z_2 = 2{,}06$




$$z = z_1 + \frac{z_2 - z_1}{P_2 - P_1}(P - P_1)$$

$$z = 2{,}05 + \frac{(2{,}06 - 2{,}05)}{0{,}9803 - 0{,}9798}(P - 0{,}9798) = 2{,}05 + \frac{0{,}01}{0{,}0005}(P - 0{,}9798)$$

$$z = 2{,}05 + \frac{1}{0{,}05} \cdot (0{,}98 - 0{,}9798) = 2{,}05 + \frac{0{,}0002}{0{,}05}$$

$$z = 2{,}05 + 0{,}004 = 2{,}054$$

a) Cuando NO hay conocimiento previo.

$$\left.\begin{array}{l} p = 0{,}5 \\ 1-p = 0{,}5 \\ \varepsilon = 0{,}05 \\ z_{1-\alpha} = 2{,}054 \end{array}\right\} \longrightarrow n \geq \frac{p(1-p)\, z_{1-\alpha}^2}{\varepsilon^2} = \frac{0{,}5 * 0{,}5 * 2{,}054^2}{0{,}05^2}$$

$$n \geq \frac{0{,}25 * 4{,}218916}{0{,}0025} = 421{,}89$$

$$n \geq 422$$

b) Variabilidad $p = 0{,}75$

$$n \geq \frac{0{,}75 * 0{,}25 * 2{,}054^2}{0{,}05^2}$$

$$n \geq \frac{0{,}1875 * 4{,}218916}{0{,}0025} = 316{,}4187 \longrightarrow n \geq 317$$

## Caso 2: Cuando se conoce el tamaño de la población (N)

$$n \geq \frac{N\, p(1-p)\, Z_{1-\alpha}^2}{(N-1)\varepsilon^2 + p(1-p)\, Z_{1-\alpha}^2}$$

**Ejemplo:** Queremos calcular una muestra de votantes de una población de 38,8 millones de habitantes. con un porcentaje de confianza del 97,8% y un error del 2%. Si no se tiene conocimientos previos de las intenciones de voto.

**Solución:** $N = 38'800.000$

$p = 0{,}5$

$1-p = 0{,}5$

$\varepsilon = 0{,}02$

$1-\alpha = 0{,}978$

$\alpha = 1 - 0{,}978 = 0{,}022$

$\frac{\alpha}{2} = 0{,}011$

$1 - \frac{\alpha}{2} = 0{,}989$

Buscando el respectivo Z en la tabla tenemos:

$Z_{1-\alpha} = 2{,}29$

Con esto tenemos:

$$n \geq \frac{N\, p(1-p) \cdot Z_{1-\alpha}^2}{(N-1)\varepsilon^2 + p(1-p)\, Z_{1-\alpha}^2}$$

$$n \geq \frac{(38'800.000)(0{,}5)(0{,}5)(2{,}29)^2}{(38'800.000-1)(0{,}02)^2 + (0{,}5)(0{,}5)(2{,}29)^2}$$

$$n \geq \frac{50'867.770}{15.519{,}9996 + 1{,}311}$$

$$n \geq 3.278$$

## Caso 3: Cuando los datos son cualitativos

Este caso es utilizado en fenómenos sociales dónde se busca estudiar la ausencia o presencia de algún atributo.

$$n \geq \frac{P}{1 + P/N}$$

Donde; $n$: tamaño muestra
$N$: tamaño población

$$P = \frac{p(1-p)}{(Se)^2}$$

$p$: variabilidad positiva (confiabilidad).
$Se$: error estandar

**Ejemplo:** Supongamos que se tiene una población de 2150 personas. que se van a entrevistar para conocer si estan a favor o en contra de un candidato para presidente municipal. Se pretende conocer la aceptación del candidato y es estudio se hara mediante una muestra. ¿Cuál deberá ser el tamaño mínimo de la muestra que cumpla con un error estandar de 0,02 y una confiabilidad de 90%?

**Solución:** $N = 2150$
$se = 0,02$
$p = 0,90$

$$P = \frac{(0,90).(0,10)}{(0,02)^2} = 225$$

Luego,

$$n \geq \frac{P}{1 + P/N}$$

$$n \geq \frac{225}{1 + \frac{225}{2150}} = \frac{225}{1,1}$$

$$n \geq 204,5 \quad ; \quad n \approx 205$$

## Caso 4: Disminución del tamaño de muestra cuando N es pequeña

Cuando se conoce el tamaño de la población y este es pequeño, es posible calcular un tamaño de muestra usando los casos 2 y 3. Sin embargo, al ser la población muy pequeña es posible que la muestra resultante corresponda a un porcentaje muy alto, por lo cual tendra que realizarse una corrección que disminuya el tamaño de la muestra.

$$n = \frac{n^* N}{n^* + N}$$

Con, $n$ = nuevo tamaño de muestra

$n^*$ = tamaño de muestra segun caso 2 ó 3

$N$ = población.

**Ejemplo:** Un ingeniero de control de calidad de cinescopios debe revisar lotes de 200 articulos cada uno. Decide que su estudio tenga una confianza de 95% y permite un error de 5%. Calcule el tamaño de la muestra para inspeccionar un lote cuando va iniciando y nunca ha revisado el lote, es decir, no tiene información previa. Después, disminuya el tamaño de la muestra con la corrección.

**Solución:** Calculemos el tamaño de muestra teniendo:

$N = 200$

$1-\alpha = 95\% = 0,95 \longrightarrow \alpha = 0,05 \longrightarrow \frac{\alpha}{2} = 0,025$

$\varepsilon = 0,05$ Luego $1-\alpha/2 = 0,975$

$p = 0,5$

$1-p = 0,5$

$\downarrow$

$Z_{1-\alpha} = 1,96$

$$n \geq \frac{N\,p(1-p)\,z_{1-\alpha}^{2}}{(N-1)\,\varepsilon^{2} + p(1-p)\,z_{1-\alpha}^{2}}$$

$$n \geq \frac{(200)(0{,}5)(0{,}5)(1{,}96)^{2}}{(200-1)(0{,}05)^{2} + (0{,}5)(0{,}5)(1{,}96)^{2}} = \frac{192{,}08}{0{,}4975 + 0{,}9604}$$

$$n \geq \frac{142{,}08}{1{,}46} \longrightarrow \quad n = 132$$

Haciendo la corrección tomaremos $n^{*} = 132$

$$n \geq \frac{132 * 200}{132 + 200} \approx 80$$