UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD CIENCIAS DE LA SALUD
PROGRAMA CIENCIAS DEL DEPORTE Y LA RECREACIÓN
SEMINARIOS DE INVESTIGACIÓN

3. DE LA MUESTRA A LA POBLACIÓN

Tomado de: ROWNTREE, Derek. Introducción a la estadística: un enfoque no matemático. Bogotá: Norma, 1984.

ESTIMACIONES E INFERENCIAS

La INFERENCIA es el proceso a través del cual se hallan los PARÁMETROS que caracterizan una población, con cierta probabilidad calculada de error, a partir de los ESTADÍSTICOS encontrados de una muestra.

BASE TEÓRICA

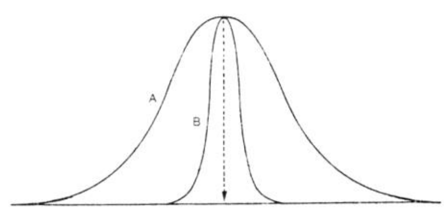
La idea es inferir la DE y media de una población, a partir de la DE y media de una muestra. La primera es sencilla: cuando la muestra tiene 30 ó más elementos, su DE es un estimador adecuado de la DE de la población. La segunda es más compleja y se presenta enseguida una breve explicación sobre su cálculo.

Si obtuviéramos infinidad de muestras de una población, cada una de éstas tendría su media. Al graficar estas medias, obtendríamos una "**curva de distribución de medias muestrales**" con las siguientes características:

- Su media sería la misma de la población, porque algunas medias muestrales será mayores que la de la población y otras serán menores, de forma que su promedio será la media poblacional.
- Su desviación estándar (DE), denominada "error estándar" (EE), sería menor que la poblacional, porque no se tienen en cuenta los datos que están ubicados a los lados de las medias de las muestras.
- Su curva se aproximaría a la curva normal, porque: *"en la población los valores se van volviendo más escasos a medida que son mayores o menores que la media de la población. Por esta razón, al tomar muestras de esta población, tenemos más probabilidad de escoger valores similares a la media de la población. Por consiguiente, serán más numerosas las muestras con valores observados cercanos a la media de la población, que las muestras con muchos valores distantes de la media de la población."* (Rowntree, 1984, 71)

El EE (la DE de una distribución de medias muestrales) es más pequeño:

- Cuanto menor sea la DE de la población: la menor dispersión de los datos se refleja en una menor dispersión de las medias muestrales, por ejemplo, en una población de 0 a 5 años que en una de 0 a 100 años.
- Cuanto mayor sea el tamaño de la muestra. A más grandes las muestras, más se parecerán a la población y, por lo tanto, más cerca estarán sus medias de la media poblacional; pero en forma logarítmica: un gran incremento en el tamaño de la muestra implica una pequeña disminución en el tamaño del EE.



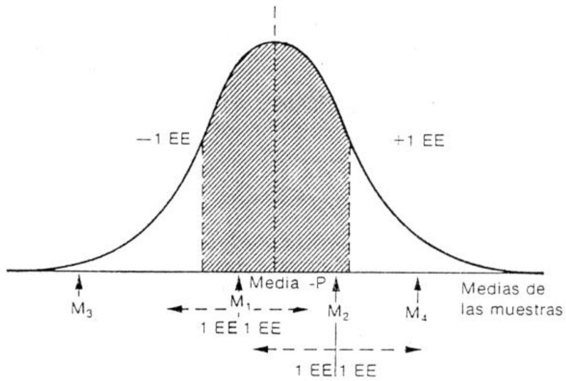
A- Distribución de la población.
B- Distribución de medias muestrales.

- También cuanto mayor sea el porcentaje de población incluida en la muestra, pero éste influye muy poco en el EE. *"La precisión de los datos está determinada por el tamaño de la muestra, por la simple **cantidad** de información disponible, y no por el porcentaje de información"*. (Rowntree, 1984, 75).

CÁLCULO

De acuerdo con lo anterior, el cálculo del EE es el siguiente: $EE = \frac{DE}{\sqrt{n}}$

El rango media poblacional ± 1 EE (µ±1EE) contendrá el 68% de las medias muestrales. Para cada una de éstas, el rango media muestral ± 1 EE ($\bar{x}$±1EE), llamado "intervalo de confianza del 68%", incluirá la media poblacional, tal como se muestra enseguida:



Del mismo modo, el rango media muestral ± 1.96 EE ($\bar{x}$±1.96EE) incluirá a la media poblacional el 95% de las veces; esto es, el intervalo de confianza del 95% va de $\bar{x}$-1.96 EE a $\bar{x}$+1.96EE. **El "intervalo de confianza" es pues el conjunto de valores entre dos límites en el cual, con una determinada probabilidad, está la media poblacional.** La expresión "1.96EE", en forma más genérica: "**ZEE**", se denomina "**error permisible**" ("d"): **es la cantidad de error en la inferencia que el investigador define** y corresponde a la cantidad que se suma o resta de la media de la muestra para determinar los puntos extremos del intervalo de confianza correspondiente (d = ZEE). Un error permisible pequeño requerirá una muestra grande, mientras uno grande requerirá una muestra pequeña

EE de una proporción:

Se calcula multiplicando la proporción de presencia de la variable en la población "p" por la proporción de ausencia de la misma "q" (p+q=1), dividiendo este producto entre el tamaño de la muestra "n" y obteniendo la raíz cuadrada de esta división.	$EEp = \sqrt{\frac{pq}{n}}$

Cálculo de la muestra:

Variable principal numérica:
El error permisible (d) es igual a Z veces el EE

$$d = Z\,EE \quad ==> \quad d = Z\,\frac{DE}{\sqrt{n'}} \quad ==> \quad d^2 = Z^2\,\frac{DE^2}{n'} \quad ==> \quad n' = \frac{Z^2\,DE^2}{d^2}$$

Variable principal cualitativa:

$$d = Z\,EE \quad ==> \quad d = Z\,\sqrt{\frac{pq}{n'}} \quad ==> \quad d^2 = Z^2\,\frac{p\,q}{n'} \quad ==> \quad n' = \frac{Z^2\,p\,q}{d^2}$$

Corrección para Población Finita:

$$n = \frac{n'}{1 + (n' / N)}$$

Si se conoce la población, y la muestra es mayor o igual a 5% de ella ($n \geq 0.05N$ ó $N \leq 20n$).

Ejemplos de cálculo de la muestra:

Variable principal numérica:

Para población infinita. Si se **des**conoce el tamaño de la población o su número es infinito sólo hacer esta:

$$n' = \frac{Z^2\, DE^2}{d^2}$$

Para población finita. Si se conoce el tamaño de la población y n' ≥ 5% de N, luego de la anterior continuar así:

$$n = \frac{n'}{1 + (n' / N)}$$

- n' es el número de la muestra para población **in**finita.
- n es el número de la muestra para población finita.
- N es el número de la población total.
- Z es el valor de Z (usualmente 1.96 para el 95% de confianza)
- DE es la desviación estándar de la población (obtenida de estudios previos o de una muestra piloto de 30 ó más unidades)
- d es el error permisible (usualmente 0.05 para el 5%)

Si **DE** = 0.5, entonces:

$$\mathbf{n'} = \frac{1.96^2 \times 0.5^2}{0.05^2} = 384.16 \approx \mathbf{385}$$

Si **N** = 1000, entonces:

$$\mathbf{n} = \frac{384.16}{1 + (384.16 / 1000)} = 277.54 \approx \mathbf{278}$$

Variable principal cualitativa:

Para población infinita. Si se **des**conoce el tamaño de la población o su número es infinito sólo hacer esta:

$$n' = \frac{Z^2\, p\, q}{d^2}$$

Para población finita. Si se conoce el tamaño de la población y n' ≥ 5% de N, luego de la anterior continuar así:

$$n = \frac{n'}{1 + (n' / N)}$$

- n' es el número de la muestra para población **in**finita.
- n es el número de la muestra para población finita.
- N es el número de la población total.
- Z es el valor de Z (usualmente 1.96 para el 95% de confianza)
- p es la probabilidad de presencia de la característica (si se desconoce poner 0.5, así pq da el máximo valor posible).
- q es la probabilidad de **ausencia** de la característica (como: p + q = 1, si se desconoce poner 0.5)
- d es el error permisible (usualmente 0.05 para el 5%)

$$\mathbf{n'} = \frac{1.96^2 \times 0.5 \times 0.5}{0.05^2} = 384.16 \approx \mathbf{385}$$

Si **N** = 1000, entonces:

$$\mathbf{n} = \frac{384.16}{1 + (384.16 / 1000)} = 277.54 \approx \mathbf{278}$$