APUNTE

MUESTREO

Índice:

Ing. Rogelio Hernán Bello

MUESTREO

OBJETIVO DEL MÉTODO: Inferir datos de la población total en base a una muestra acotada de n integrantes (menor al tamaño de la población en estudio).

Para poder realizar un muestreo, deberemos introducir algunos términos.

1) MEDIA: Dados los n números $a_1, a_2, \ldots, a_n$, la **media aritmética** se define simplemente como:

$$\bar{x} = \frac{\sum_{i=1}^{n} a_i}{n} = \frac{a_1 + \cdots + a_n}{n} \qquad \text{o sea} \qquad \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

expresión mas conocida como "promedio". Nótese que tanto a_i como X_i son los valores que estamos analizando de la población.

2) VARIANZA: La varianza representa la media aritmética de las desviaciones con respecto a la media que son elevadas al cuadrado.

Expresión de la varianza muestral:

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

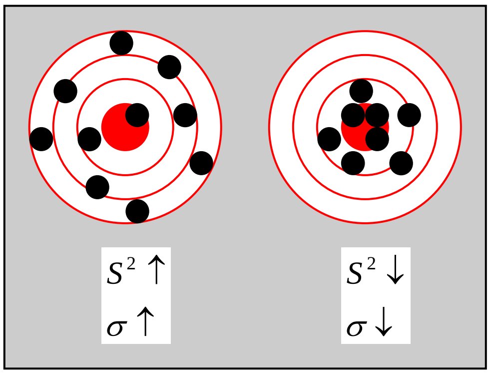
NOTA: En la varianza, los valores están elevados al cuadrado para que los que tienen signo negativo no me reste amplitud de dispersión.

3) DESVÍO: No es otra cosa que la raíz cuadrada de la Varianza y me da una idea de qué grado de dispersión (apartamiento del $\overline{X}$) tienen los valores de la muestra, es decir, qué tanto varían unos de otros.

Expresión de la desviación estándar muestral:

$$\sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

En este sentido el desvío y la varianza pueden verse como los impactos a un blanco de tiro. Cuanto mas dispersos (izq.) mayores ambos, y cuanto mas juntos (der.) a la media aritmética (centro del blanco) menores.



Ejemplo

Aquí se muestra cómo calcular la desviación estándar de un conjunto de datos (muestra de una población). Los datos representan la edad de los miembros de un grupo de niños.

DATOS: **4, 1, 11, 13, 2, 7 años**

1. Calcular el promedio o media aritmética $\overline{x}$.

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$.

En este caso, *N* = 6 porque hay seis datos:

$$x_1 = 4$$
$$x_2 = 1$$
$$x_3 = 11$$
$$x_4 = 13$$
$$x_5 = 2$$
$$x_6 = 7$$

i=número de datos para sacar desviación estándar

$$\overline{x} = \frac{1}{6} \sum_{i=1}^{6} x_i$$
$$\overline{x} = \frac{1}{6} (x_1 + x_2 + x_3 + x_4 + x_5 + x_6)$$
$$\overline{x} = \frac{1}{6} (4 + 1 + 11 + 13 + 2 + 7)$$
$\overline{x} = 6,33$ Este es el promedio.

2. Calcular la desviación estándar σ

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

Sustituyendo *N* - *1* por 5; (6 - 1) queda:

$$\sigma = \sqrt{\frac{1}{5} \sum_{i=1}^{6} (x_i - \overline{x})^2}$$

Sustituyendo $\overline{x}$ por 6,33

$$\sigma = \sqrt{\frac{1}{5} \sum_{i=1}^{6} (x_i - 6,33)^2}$$

$$\sigma = \sqrt{\frac{1}{5}\left[(4-6,33)^2+(1-6,33)^2+(11-6,33)^2+(13-6,33)^2+(2-6,33)^2+(7-6,33)^2\right]}$$

$$\sigma = \sqrt{\frac{1}{5}\left[(-2,33)^2+(-5,33)^2+4,67^2+6,67^2+(-4,33)^2+(0,67^2)\right]}$$

$$\sigma = \sqrt{\frac{1}{5}(5,43+28,4+21,8+44,5+18,7+0,449)}$$

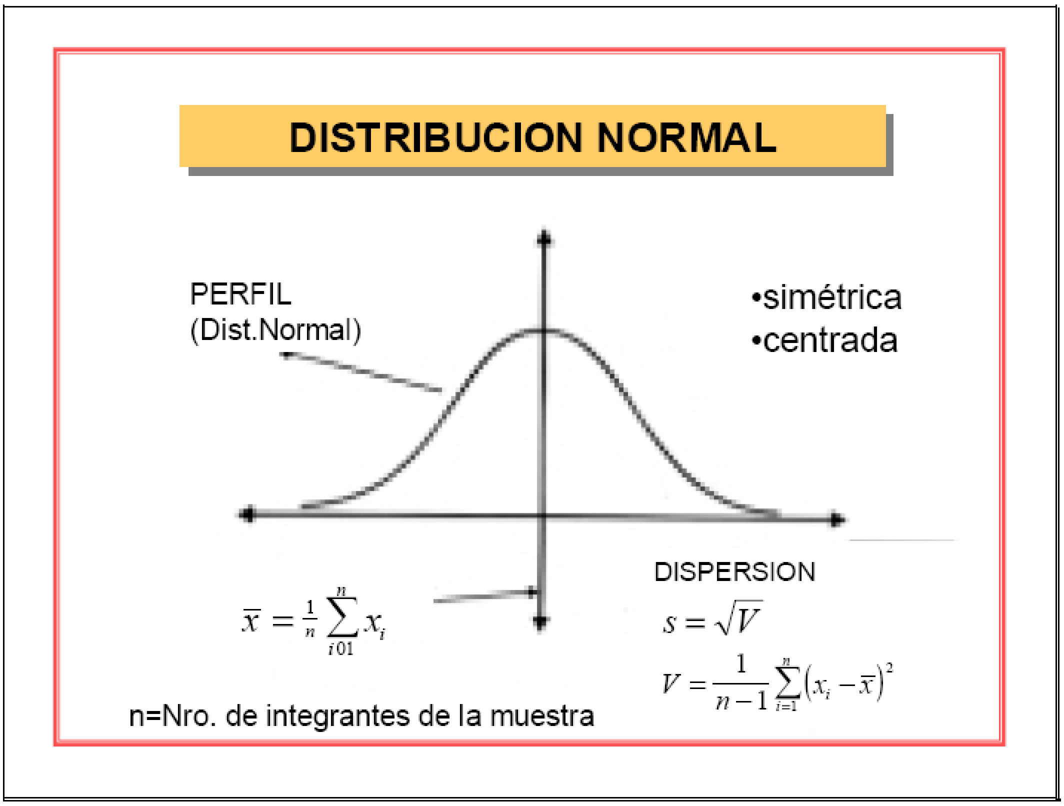
$$\sigma = \sqrt{\frac{119,28}{5}}$$

$$\sigma = \sqrt{23,86}$$

$\sigma = 4,88$ Éste es el valor de la desviación estándar

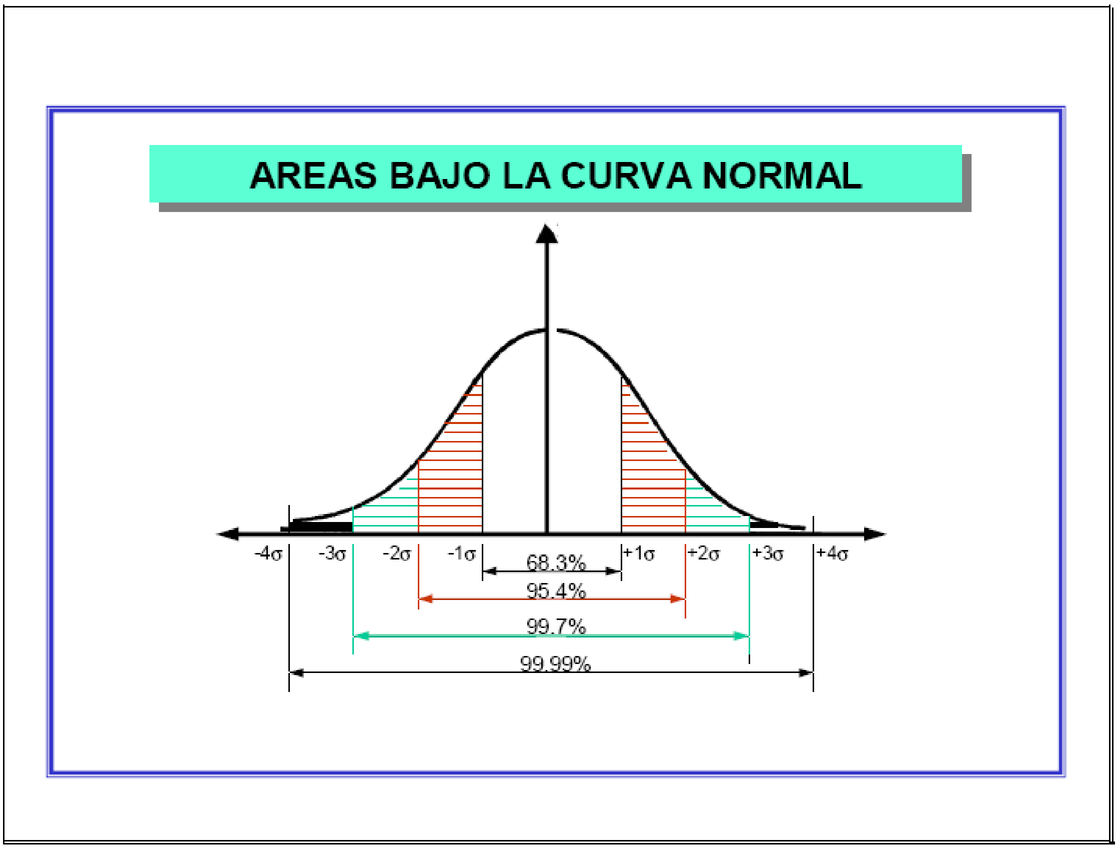
CURVA NORMAL – DISTRIBUCIÓN DE GAUSS (TEÓRICO)

Debo saber analizar e interpretar una campana de Gauss, entender qué me representan los valores de MEDIA Y DESVÌO calculados, interpretar las limitaciones existentes y ver cómo mejorar la precisión de mi análisis.



Con estos datos, puedo decir que (respecto del ejemplo anterior):

- El 68,3% de los niños de la población total tienen 6,33 ± 4,88 años
- El 95,4% de los niños tienen 6,33 ± 9,76 años
- El 99,7% de los niños tienen 6,33 ± 14,64 años
- El 99,99% de los niños tienen 6,33 ± 19,52 años
- El 100% de los niños tienen 6,33 ± ∞ años



OBSERVACIONES:

1. EN ESTE CASO EL GRAFICO ESTA LIMITADO INFERIORMENTE (CORTADO A LA IZQUIERDA A UNA DISTANCIA DE 1,3 x σ = 6,33) YA QUE NO PUEDE HABER NIÑOS DE MENOS DE 0 AÑOS.

2. VER QUE SI YO AUMENTO EL TAMAÑO DE LA MUESTRA (N > 6) PODRÈ MEJORAR LA PRECISIÒN, ES DECIR, DISMINUIR σ , PERO NECESITARÈ CONSEGUIR Y PROCESAR MAS DATOS.

TAMAÑO DE LA MUESTRA

Como se ha visto, la precisión del método esta ligada al tamaño de la muestra analizada.

PARA PENSAR: Si yo analizo el 100% de la población, tengo una certeza absoluta (100% de confianza) en los resultados, lo que equivale a realizar una medición directa.

PLANTEO: Cuando yo debo realizar un muestreo y necesito cierta precisión en los resultados, deberé calcular primeramente el tamaño de la muestra que será necesario.

Vamos a utilizar esta fórmula (extremadamente simplificada) que nos arrojará un resultado n (tamaño de muestra a tomar) que será muy confiable:

$$n = \left(\frac{\sigma \, Z_{1-\frac{\alpha}{2}}}{d} \right)^2$$

Donde:

$Z_{(1\text{-alfa}/2)}$ = coeficiente de la curva Normal relacionado con la confianza de los resultados
$Z_{(1\text{-alfa}/2)}$ = 1,96 (para el 95% de confianza en los resultados)

d = diferencia máxima aceptada entre la media estimada y el valor real buscado

INCONVENIENTE DEL MÉTODO: Requiere un muestreo piloto (previo) para estimar una semilla (valor inicial del Desvío) que es requerido para los cálculos. Anyway, los datos obtenidos del muestreo piloto sirven para el muestreo definitivo (ej.: si tomé 15 muestras piloto para obtener el desvío inicial –para estimar n- y el tamaño de muestras me dio n=46, requiero 31 muestras mas, ya que utilizo las 15 anteriores ya efectuadas).

ACLARACIÓN: el coef. Z será = 1,96 (equivalente al 95% de certeza, es decir, $\overline{X} \pm 2\sigma$ de la campana de Gauss) para todo el curso, invariablemente (siempre tomar Z=1,96)

EJEMPLO: Se desea estimar el peso promedio de los sacos que son llenados por un nuevo instrumento en una industria. Se conoce que el peso de un saco que se llena con este instrumento es una variable aleatoria con distribución normal. Si se supone que la desviación típica del peso es de 0,5 kg. Determine el tamaño de muestra aleatoria necesaria para determinar una probabilidad igual a 0,95 de que el estimado y el parámetro se diferencien modularmente en menos de 0,1 kg.

Solución:

$$\begin{aligned} d &= 0{,}1 \\ \sigma &= 0{,}5 \\ 1-\alpha &= 0{,}95 \\ 1-\alpha/2 &= 0{,}975 \\ Z_{1-\alpha/2} &= 1{,}96 \end{aligned}$$

$$n = \left(\frac{\sigma \, Z_{1-\frac{\alpha}{2}}}{d} \right)^2 = \left(\frac{(0{,}5)(1{,}96)}{0{,}1} \right)^2 = 96{,}4$$

Evidentemente un tamaño de muestra no puede ser fraccionario por lo que se debe aproximar por exceso. El tamaño de muestra sería de 97.