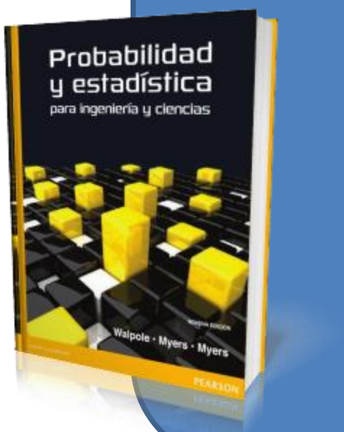


Tema 7: Técnicas y Distribuciones de Muestreo

w – Cap 8

Veremos en este tema algunas técnicas de muestreo de datos y Distribuciones Muestrales.

- Definiciones de Población, Muestra e Inferencia Estadística.
- Muestras, Sesgo, Tipos de Muestreo.
- Estadísticas de Muestreo y Distribuciones Muestrales: Tendencia Central y Variabilidad.
- Teorema del Límite Central.
- Inferencia: Media, Varianza, Desvío Estándar, Diferencia de Media, Cocientes de Varianza.
- Representaciones Gráficas.

- **<u>Población</u>**

  Totalidad de los elementos de la entidad bajo análisis.

  El **numero de observaciones en la población** se define como **el tamaño de la población** y representa a todos los elementos en el ámbito de análisis.

Población

- **Finita.** Ej. Número de alumnos en un aula. Ámbito: Aula.

  Ej. Número de alumnos en una escuela. Ámbito: Escuela.

- **Infinita.** Ej. mediciones diarias de concentración de gases en la atmósfera.

  Ámbito: Atmósfera.

Cada observación de la población toma un valor de la variable aleatoria y su comportamiento puede ser expresado como una función de variable aleatoria:

$B(x; n, p)$ Función de V.A.D.

$N(\mu, \sigma)$ Función de V.A.C.

Cada una de estas funciones de V.A. tiene **parámetros poblacionales que no varían**.

Parámetros de la POBLACIÓN

- **<u>Muestra</u>**

  Es un **subconjunto de la Población**. Una parte de la Población a la cual analizamos para intentar conocer las características de la Población sin tener que observar a todos los elementos de ella.

- **<u>Inferencia Estadística</u>**

  Obtener conclusiones de la población desde el punto de vista estadístico **a partir** del subconjunto de observaciones o **Muestra**.

Si se van a **extraer conclusiones** de una **población** a partir de los valores de una o varias **muestras**, éstas deben **ser representativas de la población**.

Esto significa que la **muestra** debe ser **INSESGADA**

- **<u>Sesgo</u>**

  Se produce por **cuestiones de subjetividad** en la **elección de los elementos** de una o varias **muestras** de la población.

- **<u>Técnicas para obtener muestras representativas de una población</u>**

  - <u>Muestreo Aleatorio Irrestricto</u>

Elección de elementos AL AZAR a partir de la población

Ej: función RND# de una calculadora

Igual probabilidad para todos los elementos de la población

Independencia del observador

  - <u>Muestreo Sistemático</u>

Elección de algunos elementos de la Población en un orden SISTEMÁTICO tomando un PASO FIJO (de tiempo o incremento) donde el primer elemento se selecciona al azar.

Ej: De una lista de alumnos ordenada por abecedario selecciono el primero al azar y luego los demás tomando un paso de 5 alumnos (quinteo).

- Muestreo por Conglomerados

Se opta por este método cuando los dos anteriores no son operativos. Se seleccionan CONGLOMERADOS EN FORMA ALEATORIA y luego se observan todos los elementos de ese conglomerado o se utiliza alguno de los métodos anteriores en ese conglomerado.

Ej: Censo de maestros o alumnos en una provincia. Los alumnos están agrupados en subconjuntos esparcidos geográficamente (localidades de la provincia) y se hace oneroso respetar la selección aleatoria de la población de alumnos para censarlos debido a los gastos de movilidad. En este caso se seleccionan localidades al azar y se censan todo los alumnos de las escuelas de las localidades seleccionadas.

- Muestreo Estratificado

La selección de los individuos o elementos de la muestra se PONDERA POR ESTRATOS PREVIAMENTE DEFINIDOS cuando supongo que la población no tiene una distribución Normal $N(\mu, \sigma)$. Estos estratos se elijen en función de la forma de la $f(x)$ que define el comportamiento de la población. Ej: Estratos que cumplan la condición de tener áreas iguales bajo la $f(x)$.

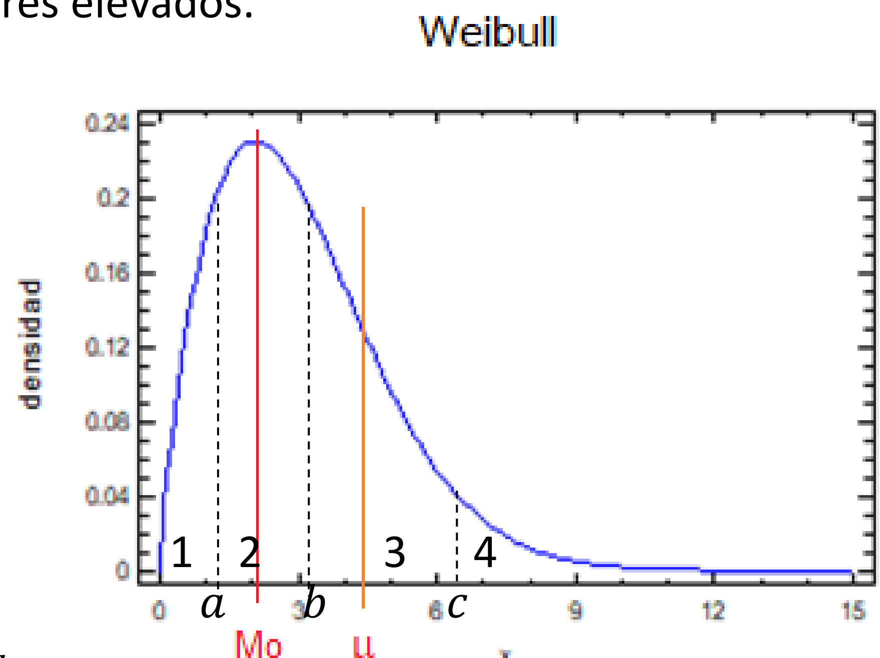
Ej: Velocidad del viento. Se sabe que esta variable no sigue una $\sim N(\mu, \sigma)$ sino una $\sim W(\alpha, \beta)$. Primero se establece la cantidad de estratos a elegir (debe ser mayor a 3) para luego encontrar los límites de cada estrato en relación al área bajo la curva de $\sim W(\alpha, \beta)$. Una vez obtenidos estos límites, se realiza el registro de datos considerando cantidades iguales de mediciones en cada uno de los estratos. Si no lo hago de esta manera corro el riesgo de tener un mayor número de mediciones de valores bajos de velocidad del viento y no tener ninguna medición con valores elevados.

1) Días con baja velocidad

2) Días con Velocidad cercana a Mo

3) Días con velocidad cercana a μ

4) Días con altas velocidades



$$\int_0^\infty f(x)dx = \int_0^a f(x)dx + \int_a^b f(x)dx + \int_b^c f(x)dx + \int_c^\infty f(x)dx$$

Estadísticas de Muestreo

De Tendencia Central y de Dispersión de los parámetros de la muestra (no de los elementos de la muestra como veníamos viendo anteriormente.)

- Tendencia Central: $\mu_{\overline{x}}$; $Mo_{\overline{x}}$; $Me_{\overline{x}}$

- Dispersión: $\sigma_{\overline{x}}$; $\sigma_{\overline{x}}^2$; $CV_{\overline{x}}$; $Q_{1\,\overline{x}}$; $Q_{3\,\overline{x}}$; $R_{\overline{x}}$

Estas Estadísticas de Muestreo se obtiene con ayuda de las Distribuciones Muestrales

Distribuciones Muestrales

Densidad de Probabilidad de los estadísticos muestrales

Permite ajustar funciones de densidad de probabilidad de los datos muestrales y realizar Inferencias acerca de los valores Poblacionales

Distribución Muestral de $\overline{X}$:

Si *x* es una V.A. en una muestra de tamaño *n* entonces:

$$\overline{x_a} = \frac{x_1 + x_2 + \ \dots + \ x_i + \ \dots + \ x_n}{n} \quad \text{de la muestra } a$$

$$\vdots$$

$$\overline{x_j} = \frac{x_1 + x_2 + \ \dots + \ x_j + \ \dots + \ x_n}{n} \quad \text{de la muestra } j$$

$$\vdots$$

$$\overline{x_n} = \frac{x_1 + x_2 + \ \dots + \ x_i + \ \dots + \ x_n}{n} \quad \text{de la muestra } n$$

$$E(\overline{x}) = \mu_{\overline{x}} = \sum_{j=1}^{n} \frac{\overline{x_j}}{n} = \frac{1}{n} \sum_{j=1}^{n} \overline{x_j} \quad \text{pero} \quad \sum_{j=1}^{n} \overline{x_j} = n\mu \Rightarrow E(\overline{x}) = \frac{1}{n} n\mu = \mu$$

$$E(\overline{x}) = \frac{1}{n} n\mu = \mu$$

Que también se puede demostrar a partir de las propiedades de la Esperanza de una función lineal de V.A.

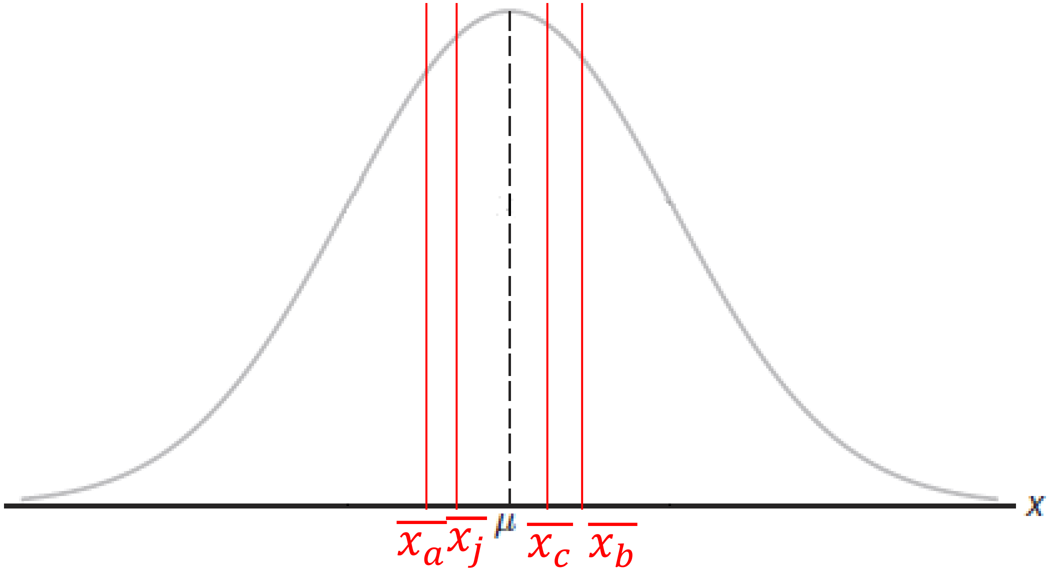
$$E(ax + b) = aE(x) + b$$

Para el caso de: $E(\overline{x}) = E\left(\frac{\sum x}{n}\right)$

$$a = \frac{1}{n}, b = 0$$

$$x = \sum x$$

$$E(\overline{x}) = \frac{1}{n} E\left(\sum x\right) = \frac{1}{n} \sum E(x) = \frac{1}{\cancel{n}} [\cancel{n} E(x)] = E(x) = \mu$$



$$\boxed{E(\overline{x}) = \mu_{\overline{x}} = \mu}$$

Distribución Muestral de $V(\overline{x})$:

Habíamos visto la propiedad de la Varianza de una combinación lineal de V.A. x:

$$V(ax+b) = a^2V(x)$$

Entonces ahora: $V(\overline{x}) = V\left(\frac{1}{n}\sum x\right)$ Como en el caso anterior $a = \frac{1}{n}, b = 0$

$$x = \sum x$$

$$V(\overline{x}) = \frac{1}{n^2}V\left(\sum x\right) = \frac{1}{n^2}\sum V(x) = \frac{1}{n^{\cancel{2}}}[\cancel{n}V(x)] = \frac{1}{n}V(x)$$

$$\boxed{V(\overline{x}) = \frac{1}{n}V(x)}$$

Teorema del Límite Central

Si *x* es una V.A. con $\sim N(\mu, \sigma)$ y $\overline{x}$ es un estadístico de la muestra de tamaño *n*, entonces:

La V. A. normalizada $z = \frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} \quad \rightarrow \quad \sim N(0,1)$ si $n \rightarrow \infty$

Además como $\mu_{\overline{x}} = \mu$ y $\sigma_{\overline{x}} = \sqrt{V(\overline{x})} = \sqrt{\frac{1}{n} V(x)} = \frac{\sigma_x}{\sqrt{n}}$

Entonces: $z = \frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$

Por lo tanto si $n \rightarrow \infty$ $(n > 30)$ entonces por más que la ~ V.A. $x \neq \sim N(\mu, \sigma)$:

$$\boxed{z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \quad \rightarrow \quad \sim N(0,1)}$$