

UT 1: Fundamentos de Métodos Cuantitativos

Estadística Univariada

Estadística descriptiva - Nociones de Población y muestra

Definición: Una población es un conjunto que contiene la totalidad de individuos a ser estudiados por sus características.

Definición: Una muestra es un subconjunto de la población seleccionada mediante alguna técnica de muestreo

Observación

- © En general denotaremos la cantidad de individuos de la población por N y la cantidad de individuos de la muestra por n .
- © En general denotaremos a una variable estadística por X , Y o Z . Si tenemos una muestra de tamaño n , esta será denotada por $X_1, X_2, \dots, X_n$

Recolección de Datos

- © Censar: Proceso de recuento de todos los elementos de la población
 - © Muestrear: Proceso de selección de un subconjunto de la población
 - ▶ Muestreo aleatorio simple
 - ▶ Muestreo aleatorio estratificado
 - ▶ Muestreo aleatorio de conglomerados
 - ▶ Muestreo aleatorio sistemático
-

Tipos de Variables Estadísticas

- © Cualitativas: Expresan cualidades, características, categorías o modalidades
 - © Cuantitativas: Toman como argumento cantidades numéricas
 - ▶ Discretas: El conjunto de valores posibles es finito o infinito numerable.
Ejemplo: X : Número de integrantes de una familia
 - ▶ Continuas: Son aquellas que toman infinitos valores
Ejemplo: Y : Rentabilidad de un fondo mutuo
-

Escala de Variables Estadísticas

- © Nominal: Las realizaciones de la variable estadística son categorías que no tienen un orden.
Ejemplo: X : Color de ojos (Café, verde, azul,...)
 - © Ordinal: Las realizaciones de la variable estadística se pueden ordenar pero no son necesariamente números.
Ejemplo: Y : Calificación de un examen (Reprobado, Aprobado, Distinción)
 - © Intervalar: El recorrido de la variable estadística es un subconjunto de $\mathbb{R}$.
Ejemplo: Ingreso mensual de una compañía (miles de pesos)
-

Organización de la Información

Consideremos una muestra de tamaño n , obtenida desde alguna técnica de muestreo. Supongamos que la variable estadística está definida en k clases denotadas por $C_1, C_2, \dots, C_k$.

Definición: Se llama frecuencia absoluta de la clase C_i al número de individuos que pertenece a dicha clase. Lo denotamos por n_i para $i = 1, \dots, k$.

Observación

$$\textcircled{c} \quad n = \sum_{i=1}^k n_i$$

Definición: Se llama frecuencia relativa de la clase C_i al número de individuos, sobre el total de la muestra, de la clase i -ésima. Lo denotamos por f_i para $i = 1, \dots, k$.

Observación

- © $f_i = \frac{n_i}{n}$ para $i = 1, \dots, k$
 - © $\sum_{i=1}^k f_i = 1$
 - © Las frecuencias definidas tienen sentido en cualquier tipo de escala.
-

Ejemplo

Se registra la información de los grados de Mercalli de 20 sismos ocurridos en la ciudad de Santiago.

V VI VI VII V IV IV III II III
 II II III IV V V VI II I III

Resumen en la siguiente tabla:

Esc. Mercalli	n_i	f_i
I	1	1/20
II	4	4/20
III	4	4/20
IV	3	3/20
V	4	4/20
VI	3	3/20
VII	1	1/20

Supongamos que estamos trabajando en una escala ordinal, es¹⁻⁸ decir, las clases $C_1, \dots, C_k$ están ordenadas.

Definición: Se llama frecuencia absoluta acumulada hasta la clase C_j , al número de individuos que pertenecen a las clases $C_1, C_2, \dots, C_j$. Lo denotamos por N_j con $j = 1, \dots, k$.

Observación

$$\textcircled{C} N_j = \sum_{i=1}^j n_i ; N_k = n$$

Definición: Se llama frecuencia relativa acumulada hasta la clase C_j , al número de individuos, sobre el total de la muestra, que pertenecen a las clases $C_1, C_2, \dots, C_j$. Lo denotamos por F_j con $j = 1, \dots, k$.

Observación

$$\textcircled{C} F_j = \sum_{i=1}^j f_i ; F_k = 1$$

Ejemplo Anterior

Se registra la información de los grados de Mercalli de 20 sismos ocurridos en la ciudad de Santiago.

Esc. Mercalli	n_i	f_i	N_j	F_j
I	1	1/20	1	1/20
II	4	4/20	5	5/20
III	4	4/20	9	9/20
IV	3	3/20	12	12/20
V	4	4/20	16	16/20
VI	3	3/20	19	19/20
VII	1	1/20	20	20/20=1

Ejemplo 1

Supongamos que se desea estudiar las ventas de un producto de una tienda. Se toman los datos de las ventas realizadas en 36 meses. Los datos son los siguientes:

490	650	1800	1240	1200	600	870	890	630	600
1200	1210	650	470	1070	1200	680	600	870	1020
880	1090	1050	1250	690	1210	810	810	890	1250
1200	1000	1000	630	690	1270				

Observación

- 1) La mejor forma de ordenar los datos es de menor a mayor.
- 2) Estos datos se utilizarán como ejemplo de todos los estadísticos que a continuación se presentan

Ventas	m_i	n_i	f_i	N_i	F_i
[469,5-691,3)	580,4	12	12/36	12	12/36
[691,3-913,1)	802,2	7	7/36	19	19/36
[913,1-1134,9)	1024	6	6/36	25	25/36
[1134,9-1356,7)	1245,8	10	10/36	35	35/36
[1356,7-1578,5)	1467,6	0	0	35	35/36
[1578,5-1800,3)	1684,4	1	1/36	36	1

Observación

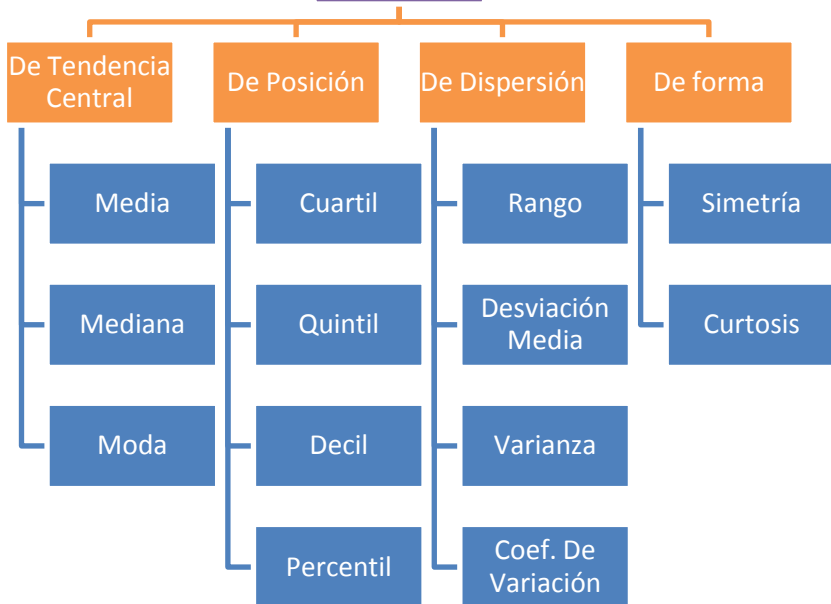
Los datos resumidos en una tabla se conocen como datos agrupados

Medidas Descriptivas

Son valores numéricos calculados a partir de una muestra y nos entregan información contenida en ella. Las medidas descriptivas que veremos son:

1. Tendencia central
 2. Posición
 3. Dispersión
 4. Forma
-

Medidas Descriptivas



Tendencia Central

Estas medidas nos entregan información para resumir los datos y además tienden a ubicarse en el centro del conjunto de datos. Supongamos que tenemos una muestra de tamaño n dada por $X_1, X_2, \dots, X_n$.

1. Promedio o media aritmética: El promedio de un conjunto es el centro de gravedad de dicho conjunto.

- Datos a granel: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- Datos tabulados: $\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i \cdot m_i = \sum_{i=1}^k f_i \cdot m_i$

Ejemplo:

- Datos a granel: $\bar{X} = 935$
- Datos tabulados: $\bar{X} = 912,961$

2. Mediana: Se define como el valor que ocupa el lugar ^{1.1.8}central de los datos, después de ser ordenados de menor a mayor.

- Datos a granel:

$$\text{Me}(X) = \begin{cases} X_{(\frac{n+1}{2})} & , \quad n \text{ impar} \\ 0,5 \cdot [X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}] & , \quad n \text{ par} \end{cases}$$

donde $X_{(i)}$ es el dato que ocupa la posición i -ésima cuando están ordenados de menor a mayor.

- Datos tabulados: Primero se debe identificar la clase mediana, es decir, la clase cuya frecuencia acumulada sea mayor o igual a la mitad de la cantidad de total de datos.

$$\text{Me}(X) = L_{I_e} + \frac{(\frac{n}{2} - N_{e-1})}{n_e} \cdot a_e$$

Ejemplo:

- Datos a granel: $\text{Me}(X) = 890$
 - Datos tabulados: $\text{Me}(X) = 881,41$ (Clase mediana: C_2)
-

3. Moda: Corresponde al valor o clase con la mayor frecuencia del conjunto de datos. Para calcularlo es necesario identificar la clase modal, la cual es la que tiene la mayor frecuencia.

$$\text{Mo}(X) = L_{I_{cmo}} + \left(\frac{n_{cmo} - n_{cmo-1}}{2 \cdot n_{cmo} - n_{cmo-1} - n_{cmo+1}} \right) \cdot a_{cmo}$$

Ejemplo: $\text{Mo}(X) = 626,06$ (Clase modal C_1)

Medias de Posición

Estas medidas están basadas en los cuantiles, los cuales son medidas de un conjunto de datos que se consideran fracciones acumuladas del conjunto de observaciones, cuando son ordenados de menor a mayor.

1. Cuartil (Q): Divide el conjunto de datos en 4 sectores del 25% cada uno.
 2. Quintil (K): Divide el conjunto de datos en 5 sectores del 20% cada uno.
 3. Decil (D): Divide el conjunto de datos en 10 sectores del 10% cada uno.
 4. Percentil (P): Divide el conjunto de datos en 100 sectores del 1% cada uno.
-

© Datos a granel: $P_i = X_{\frac{i(n+1)}{100}} \quad i = 1, \dots, 100$

En el caso de que $\frac{i(n+1)}{100}$ no sea entero, se considerará como percentil el promedio de los dos datos mas cercanos a dicha posición.

Ejemplo:

$$P_{25} = Q_1 = X_{\frac{25(36+1)}{100}} = X_{(9,25)=665}$$

$$P_{75} = Q_3 = X_{\frac{75(36+1)}{100}} = X_{(27,75)=1200}$$

- © Datos tabulados: Primero se debe identificar la clase percentil, la cual es la primera cuya frecuencia relativa acumulada sea mayor o igual a $i/100$, con $i = 1, \dots, 100$

$$P_i = L_{I_{pi}} + \frac{\left(\frac{n \cdot i}{100} - N_{pi-1}\right)}{n_{pi}} \cdot a_{pi}$$

Ejemplo:

$$P_{25} = Q_1 = 635, 85 \text{ (Clase percentil } C_1)$$

$$P_{75} = Q_3 = 1179, 26 \text{ (Clase percentil } C_4)$$

Medidas de Variabilidad

Son medidas que nos muestran o miden la dispersión, precisión o riesgo del conjunto de datos.

1. Rango:

- ▶ Datos a granel: $R = \max\{X_i\}_{i=1}^n - \min\{X_i\}_{i=1}^n$
- ▶ Datos tabulados: $R = L_{S_k} - L_{I_1}$

Ejemplo:

- ▶ Datos a granel: $R = 1330$
- ▶ Datos tabulados: $R = 1330, 8$

Observación

Rango intercuartil: $R / Q = Q_3 - Q_1$

2. Desviación media:

► Datos a granel: $DM = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|$

► Datos tabulados:

$$DM = \frac{1}{n} \sum_{i=1}^k n_i |m_i - \bar{X}| = \sum_{i=1}^k f_i |m_i - \bar{X}|$$

Ejemplo:

► Datos a granel: $DM = 242,5$

► Datos tabulados: $DM = 264,781$

Observación

También se puede definir la desviación mediana y desviación modal, basta reemplazar el promedio por la mediana y la moda, respectivamente.

3. Varianza:

- ▶ Datos a granel: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- ▶ Datos tabulados:

$$S^2 = \frac{1}{n} \sum_{i=1}^k n_i (m_i - \bar{X})^2 = \sum_{i=1}^k f_i (m_i - \bar{X})^2$$

Observación

- ▶ Se puede probar:
 - Datos a granel: $S^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 + n\bar{X}^2$
 - Datos tabulados: $S^2 = \sum_{i=1}^k f_i \cdot m_i^2 - X^2$
- ▶ La raíz cuadrada de la varianza se define como la desviación estándar .

4. Coeficiente de Variación: Este índice representa la variabilidad relativa de una población o muestra. Nos proporciona información sobre la homogeneidad de la muestra.

$$CV = \frac{S}{|X|}$$

Observación

- ▶ El CV nos permite comparar muestras con distintas unidades de medidas.
- ▶ A mayor CV entonces la muestra es menos homogénea.
- ▶ Se puede definir el coeficiente de variación robusta:

$$CV R = \frac{R / Q}{Q_1 + Q_3}$$

Medidas de Forma

- © **Indicadores de Simetría:** Apuntan a la tendencia que tienen los datos para concentrarse en ciertos valores.
- ▶ **Distribución Simétrica:** Los datos están concentrados en los valores centrales de la variable.
 - ▶ **Distribución Asimétrica Positiva: o a la Derecha** Los datos están concentrados en los valores bajos de la variable.
 - ▶ **Distribución Asimétrica Negativa o a la Izquierda:** Los datos están concentrados en los valores altos de la variable.
-

Indicadores de Simetría

1. Indicador de Yule: $I_y = \frac{Q_1 + Q_3 - 2Q_2}{2Q_2}$
2. Indicador de Simetría: $I_s = \frac{Q_1 + Q_3 - 2Q_2}{R / Q}$

3. Coeficiente de Pearson: $A_s = \frac{3(\overline{X} - M_e)}{S}$
4. Coeficiente de simetría de Fisher: $\nu_3 = \frac{m_3}{S^3}$

donde

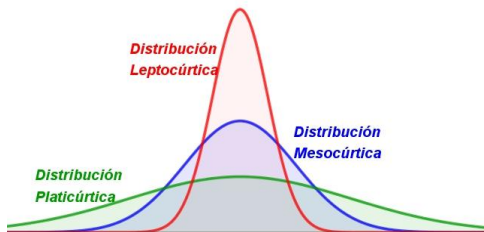
$$m_3 = \begin{cases} \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^3 & ; \text{ granel} \\ \sum_{i=1}^k f_i (m_i - \overline{X})^3 & ; \text{ tabulados} \end{cases}$$

- © Se dice que la muestra es simétrica si los estadísticos de simetría son iguales a cero.
- © Si el estadístico de simetría es positivo, entonces tenemos una asimetría positiva o a la derecha en la muestra.
- © Si el estadístico de simetría es negativo, entonces tenemos una asimetría negativa o a la izquierda en la muestra.



© **Indicadores de Curtosis:** Apuntan al grado de concentración de los datos.

- ▶ **Mesocúrtica:** Variabilidad ideal, se asemeja a una distribución normal.
- ▶ **Leptocúrtica:** Presenta una gran concentración de los datos en torno a la media.
- ▶ **Platicúrtica:** Presenta una concentración menor alrededor a la media.



Indicadores de Curtosis

1. Coeficiente K_2 : $K_2 = \frac{D_9 - D_1}{1,9RIQ} - 1$

2. Coeficiente de Fisher: $\gamma_4 = \frac{m_4}{S^4} - 3$

donde

$$m_4 = \begin{cases} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4 & ; \text{ granel} \\ \sum_{i=1}^k f_i (m_i - \bar{X})^4 & ; \text{ tabulados} \end{cases}$$

Observación

- © Si los coeficientes de curtosis son iguales a cero entonces la variabilidad es mesocurtica.
- © Si los coeficientes de curtosis son positivos entonces la variabilidad es leptocurtica.
- © Si los coeficientes de curtosis son negativos entonces la variabilidad es platicurtica.

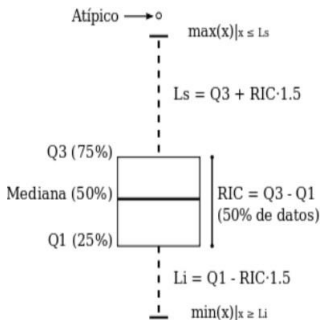
Ejemplo 1 con datos a granel

- © Coeficiente de asimetría de Fisher: $\gamma_3 = 0,4962$.
Entonces la muestra presenta una asimetría positiva o a la derecha.

 - © Coeficiente de Fisher: $\gamma_4 = 0,0495$
Entonces la muestra presenta una variabilidad mesocúrtica.
-

Box-Plot

Es un gráfico que nos permite visualizar la simetría, la existencia de outliers (datos atípicos), la dispersión con respecto a la mediana y el rango de los datos. La construcción del gráfico es la siguiente:



Box-Plot

El valor del bigote inferior se calcula de la siguiente forma:

$$B_I = \max\{L_I, \min\{X_i\}_{i=1}^n\}$$

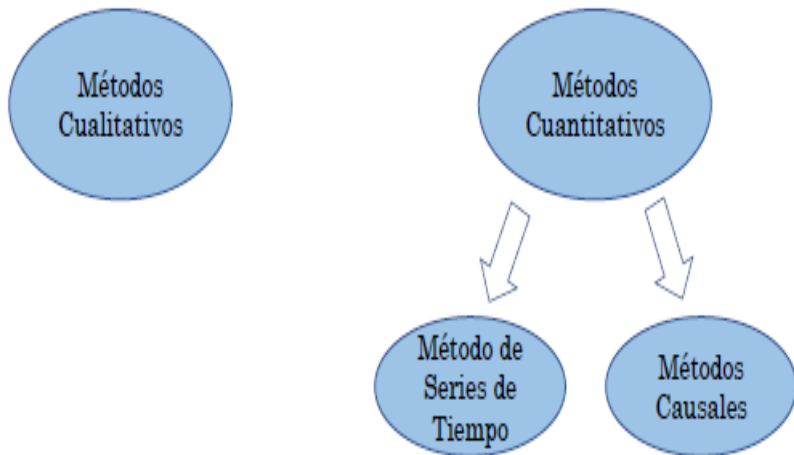
donde: $L_I = Q_1 - 1,5 \cdot RIQ.$

El valor del bigote superior se calcula de la siguiente forma:

$$B_S = \min\{L_S, \max\{X_i\}_{i=1}^n\}$$

donde: $L_S = Q_3 + 1,5 \cdot RIQ.$

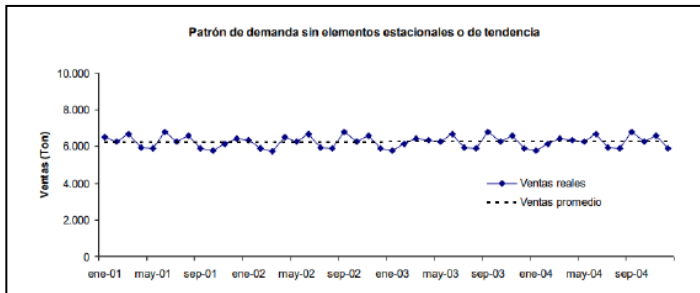
Parte 2



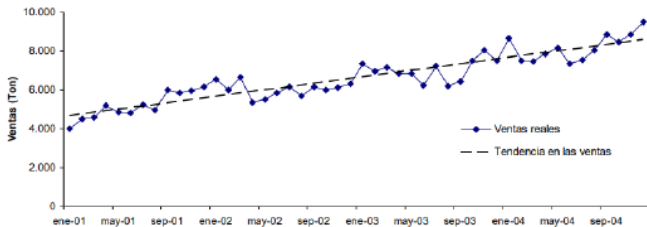
Métodos de Series de Tiempo

Una serie de tiempo está compuesta de cuatro componentes o patrones generales:

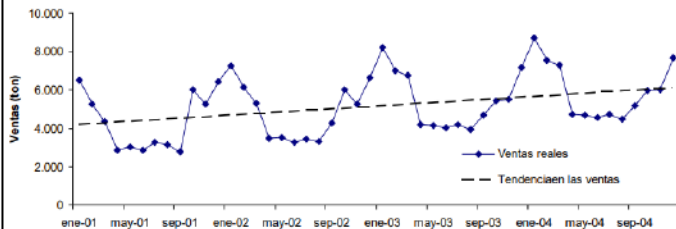
- **Tendencia:** Incremento o decremento sistemático de la media de la serie a través del tiempo.
- **Estacionalidad:** Patrón repetible de incrementos y decrementos de la demanda (horas, días, semanas, meses, temporada).
- **Ciclos:** Pauta de incrementos o decrementos graduales y menos previsibles de la demanda (años, decenios).
- **Variaciones de Azar:** series de variaciones imprevisibles de la demanda.



Patrón de demanda con tendencia creciente

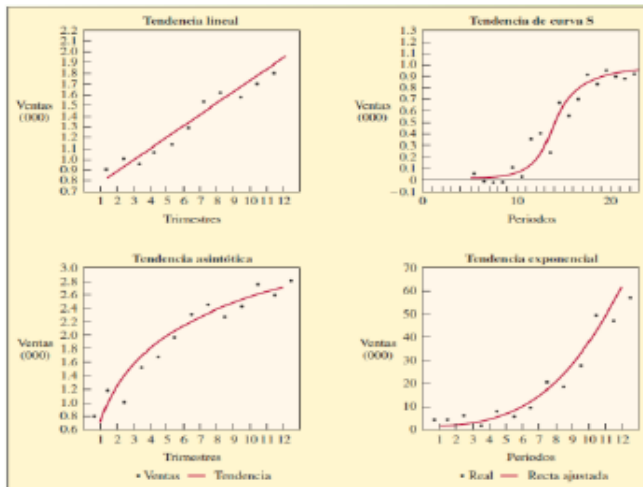


Patrón de demanda con elementos de tendencia y estacionalidad



Métodos de Series de Tiempo

Tendencias Comunes:



Métodos de Series de Tiempo

Los modelos de series de tiempo suponen que la demanda futura es una función del pasado, ignorando otras variables.

Los modelos de series de tiempo más usados son:

- Simplista
- Promedios móviles
- Suavización exponencial
- Suavización exponencial doble
- Descomposición

