

Segunda Clase

Tema

Expresión Resumida de los datos – Medidas de Posición y de Dispersión

Introducción

En esta segunda clase vamos a incorporar algunos indicadores y herramientas que nos permiten una mejor comprensión de los datos obtenidos, con el fin de presentar la información procesada de una forma más sintética y de mayor claridad comprensiva para nuestro demandante, de la investigación claro.

En el siguiente cuadro mostramos una síntesis esquemática de las medidas de posición que vamos a ver, luego haremos lo propio con las medidas de dispersión.



Modo (Mo)

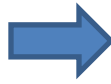
Es el valor de la variable que observa la frecuencia más alta

Formas de cálculo. Como en el resto de las medidas a presentar hay dos posibilidades, a) Frecuencias Simples, b) Frecuencias Agrupadas (intervalos)

a) Frecuencias Simples

Tabla 1: Nivel de estudios terminados

| NET | Casos |
|--------------|------------|
| Primarios | 30 |
| Secundarios | 50 |
| Terciarios | 20 |
| Total | 100 |

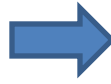


En este Ejemplo el modo pertenece a los alumnos secundarios con 50 casos

b) Frecuencias Agrupadas

Tabla 2: Edades

| Grupo de Edad | Casos |
|------------------|------------|
| 16 a 36 años (Y) | 48 |
| 37 a 55 años (X) | 28 |
| 56 a 74 años (Z) | 18 |
| 75 o más años | 6 |
| Total | 100 |



En las series agrupadas el primer paso consiste en encontrar el intervalo con la mayor frecuencia, que para el caso es el grupo de edad comprendido entre los 16 años y los 36 años

Una vez identificado el intervalo que contiene el valor modal se realiza una interpolación a su interior, con la siguiente formula

$$Mo = Li + i * ((FApost / (FAant + FApost)))$$

Donde



Li=límite inferior de la categoría modal
i=Amplitud de la categoría
FAant=Frecuencia absoluta de la categoría anterior
FApost=Frecuencia absoluta de la categoría posterior

Reemplazando

$$Mo = 16 + 20 * ((28 / (0 + 28)))$$

$$Mo = 16 + 20 = 30$$

$$Mo = 30$$

Mediana (Me)

Se define como el valor de la variable que supera a no más de la mitad de datos y es superado, a su vez, por no más de la mitad de los casos

Formas de Cálculo

Para este caso vamos a tomar como ejemplo uno de los subgrupos de la tabla de ingresos, y vamos a considerar dos posibilidades: a) Una cantidad impar de datos y b) una cantidad par de datos. Estos representan las dos posibilidades que se dan en los casos de tablas ordenadas por frecuencia simple.

Tabla 3: Subgrupo Ingresos

| Nro Reg | Ingresos | Nro Reg | Ingresos |
|---------|----------|---------|----------|
| 1 | 15.000 | 1 | 15.000 |
| 2 | 20.000 | 2 | 20.000 |
| 3 | 50.010 | 3 | 50.010 |
| 4 | 56.000 | 4 | 56.000 |
| 5 | 26.000 | 5 | 26.000 |
| 6 | 30.000 | 6 | 30.000 |
| 7 | 25.000 | 7 | 25.000 |
| 8 | 40.000 | 8 | 40.000 |
| 9 | 31.000 | 9 | 31.000 |
| | | 10 | 28.000 |

a) Frecuencias Simples: Cantidad de Observaciones Impar

| Nro Reg | Ingresos |
|---------|----------|
| 1 | 15.000 |
| 2 | 20.000 |
| 7 | 25.000 |
| 5 | 26.000 |
| 6 | 30.000 |
| 9 | 31.000 |
| 8 | 40.000 |
| 3 | 50.010 |
| 4 | 56.000 |



El primer paso consiste en ordenar de menor a mayor la variable bajo análisis y luego ubicar el caso que esta al medio de la serie

En este caso la mediana corresponde al valor que esta al medio de la distribución es decir el valor de 30.000 pesos que corresponde al registro que hemos denominado 6 (en ese lugar podría estar el nombre de la persona entrevistada, al que le asignamos un número de registro al momento de la carga de datos).

b) Frecuencias Simples: cantidad de observaciones par

| Nro Reg | Ingresos |
|---------|----------|
| 1 | 15.000 |
| 2 | 20.000 |
| 7 | 25.000 |
| 5 | 26.000 |
| 10 | 28.000 |
| 6 | 30.000 |
| 9 | 31.000 |
| 8 | 40.000 |
| 3 | 50.010 |
| 4 | 56.000 |



El primer paso consiste en ordenar de menor a mayor la variable bajo análisis y luego ubicar el valor que divide en dos a la distribución

La mediana se encuentra entre 28000 y 30000, es decir $Me = (28000 + 30000) / 2 = 29000$.

c) Frecuencias Agrupadas: caso de agrupadas pero no por intervalos

Tabla 4: Imagen del Dirigente X

| Imagen | Casos |
|--------------|------------|
| Muy mala | 10 |
| Mala | 40 |
| Regular | 15 |
| Buena | 30 |
| Muy Buena | 5 |
| Total | 100 |

| Imagen | FA | FAA |
|--------------|------------|-----|
| Muy Buena | 5 | 5 |
| Muy mala | 10 | 15 |
| Regular | 15 | 30 |
| Buena | 30 | 60 |
| Mala | 40 | 100 |
| Total | 100 | |



Ordenamos la serie de menor a mayor, calculamos la FAA. Luego para ubicar el valor que esta al medio, dividimos las cantidad total de casos (n) por 2 (la mitad), esto corresponde al valor de 50 ($100 / 2$).

Es decir hasta la categoría **regular** se ubican 30 casos, si la mediana está en el valor 50, significa que no puede estar allí, ahora hasta la categoría **buena** se ubican 60 casos de los 100 entrevistados, lo que significa que allí está el valor de la mediana.

d) Frecuencias Agrupadas por intervalos

| Ingresos | FA |
|-----------------|------------|
| 15.000 – 29.000 | 50 |
| 29.000 – 43.000 | 40 |
| 43.000 – 57.000 | 10 |
| Total | 100 |



En el caso de que las frecuencias estén agrupadas por intervalo, primero identificamos la categoría donde se encuentra la mediana y luego y luego obtenemos el valor de la misma mediante interpolación, con la formula siguiente:

$$Me = Li + i * ((n/2 - FA_{\text{afant}}) / F_{\text{ap}})$$

Donde



$$Me = 15.000 + 14.000 * ((100/2) - 0) / 50$$

$$Me = 29.000$$

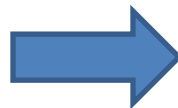
Li = Límite inferior del intervalo que contiene a la mediana
 i = Amplitud del intervalo
 n = cantidad total de casos
 FA_{afant} = frecuencia absoluta acumulada anterior a la que contiene a la mediana
 F_{ap} = frecuencia absoluta propia que contiene a la mediana

Media (\bar{X})

Es el valor promedio del conjunto de datos que observa la variable. Su cálculo solo tiene sentido en el caso de variables cuantitativas

a) Frecuencias Simples

| Nro. Reg | Xi | Ingresos |
|----------|-----|----------|
| 1 | X1 | 15.000 |
| 2 | X2 | 20.000 |
| 3 | X3 | 50.010 |
| 4 | X4 | 56.000 |
| 5 | X5 | 26.000 |
| 6 | X6 | 30.000 |
| 7 | X7 | 25.000 |
| 8 | X8 | 40.000 |
| 9 | X9 | 31.000 |
| 10 | X10 | 28.000 |
| Total | | 321.010 |



El valor promedio se calcula como la suma de todos los valores y se divide por cantidad de casos, es decir

$$\bar{X} = 321.010 / 10 = 32.101.$$

En formulas es

$$\bar{X} = (X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10}) / n$$

$$\bar{X} = \sum Xi / n, \text{ para } i \text{ que va de } 1 \text{ a } 10$$

$$\bar{X} = 32.101$$

b) Frecuencias Agrupadas pero no en intervalos

Para este caso vamos a recurrir a un ejemplo externo a la SP1, dado que no incluimos una variable numérica que pudieramos clasificar de este modo. Supongamos obtenemos el numero de materias aprobadas de los Alumnos de IA al cabo del primer Año y obtenemos la siguiente tabla

| Numero de Materias Aprobadas (X) | Cantidad de casos (FA) | X * FA |
|----------------------------------|------------------------|------------|
| 0 | 5 | 0 |
| 1 | 6 | 6 |
| 2 | 6 | 12 |
| 3 | 3 | 9 |
| 4 | 8 | 32 |
| 5 | 8 | 40 |
| 6 | 10 | 60 |
| 7 | 3 | 21 |
| 8 | 4 | 32 |
| Total | 53 | 212 |

En este caso los datos de la tabla nos indican que la variable 0 se repite 5 veces, la 1 7 veces y así sucesivamente. Por lo que el cálculo ahora será:

$$X = (0*5 + 1*6 + 2*6 + 3*3 + 4*8 + 5*8 + 6*10 + 7*3 + 8*4) / 53 = 212 / 53 = 4$$

 En formulas

$$X = X_0*FA_0 + X_1*FA_1 + X_2*FA_2 + X_3*FA_3 + X_4*FA_4 + X_5*FA_5 + X_6*FA_6 + X_7*FA_7 + X_8*FA_8 =$$

$$X = \sum X_i*FA_i \text{ para } i \text{ que va de } 0 \text{ a } 8$$

c) Frecuencias Agrupadas por intervalos

| Intervalo (X) | FA | MC (Xi) | FA * MC |
|-----------------|-----------|---------|----------------|
| 15.000 a 25.250 | 5 | 20.125 | 60.375 |
| 25.250 a 35.500 | 4 | 30.750 | 121.500 |
| 35.500 a 45.750 | 1 | 40.625 | 40.625 |
| 47.750 a 56.000 | 2 | 50.875 | 101.750 |
| Total | 10 | | 324.250 |

Se resuelve de la misma forma que en el caso anterior, solo que como valor de las X_i , se toma lo que vamos a llamar Marca de Clase (MC) que no es otra cosa que el valor promedio entre los dos valores extremos del intervalo. Por lo que el cálculo ahora será:

$$X = (5*20.125 + 4*30.750 + 1*40.625 + 2*50.875) / 10 = 324.250 / 10 = 32.425$$

 En formulas

$$X = (X_1*FA_1 + X_2*FA_2 + X_3*FA_3 + X_4*FA_4) / n$$

$$X = (\sum X_i*FA_i) / n \text{ para } i \text{ que va de } 1 \text{ a } 4$$

Medidas de Dispersión

Basadas en
la media

Otras

Varianza

Desviación
Estándar

Coefficiente de
Variación

Recorrido

Amplitud
Intercuartilica

Coefficiente de
incertidumbre

Varianza (V)

Es un indicador del grado de dispersión de los valores de la variable respecto a su media.

a) Frecuencia Simple

| Nro. Reg | Xi | Ingreso | $\bar{X} - Xi$ | $(\bar{X} - Xi)^2$ |
|-------------|-----|---------|----------------|--------------------|
| 1 | X1 | 15.000 | -17.101 | 292.444.201 |
| 2 | X2 | 20.000 | -12.101 | 146.434.201 |
| 3 | X3 | 50.010 | 17.909 | 320.732.281 |
| 4 | X4 | 56.000 | 23.899 | 571.162.201 |
| 5 | X5 | 26.000 | -6.101 | 37.222.201 |
| 6 | X6 | 30.000 | -2.101 | 4.414.201 |
| 7 | X7 | 25.000 | -7.101 | 50.424.201 |
| 8 | X8 | 40.000 | 7.899 | 62.394.201 |
| 9 | X9 | 31.000 | -1.101 | 1.212.201 |
| 10 | X10 | 28.000 | -4.101 | 16.818.201 |
| Total | | 321.010 | 0 | 1.503.258.090 |
| \bar{X} | | 32.101 | | 150.325.809 |
| | | Media | | Varianza |

Para calcular la varianza se calcula primero la diferencia entre la media y cada valor de la variable (desvíos), luego se eleva al cuadrado cada una de esas diferencias y se procede a la suma total de estas últimas. Al resultado se lo divide por la cantidad total de casos.

En formula es : $(\bar{X}-X1)^2 + (\bar{X}-X2)^2 + (\bar{X}-X3)^2 + (\bar{X}-X4)^2 + (\bar{X}-X5)^2 + (\bar{X}-X6)^2 + (\bar{X}-X7)^2 + (\bar{X}-X8)^2 + (\bar{X}-X9)^2 + (\bar{X}-X10)^2 / n$

$V(X) = (\sum (\bar{X} - Xi)^2) / n$ para i que va de 1 a 10

b) Frecuencias Agrupadas. Numero de Materias Aprobadas

| X | FA | X * FA | $X - \bar{X}$ | $(X - \bar{X})^2$ | $(X - \bar{X})^2 * FA$ |
|-------|----|--------|---------------|-------------------|------------------------|
| 0 | 5 | 0 | -4 | 16 | 80 |
| 1 | 6 | 6 | -3 | 9 | 54 |
| 2 | 6 | 12 | -2 | 4 | 24 |
| 3 | 3 | 9 | -1 | 1 | 3 |
| 4 | 8 | 32 | 0 | 0 | 0 |
| 5 | 8 | 40 | 1 | 1 | 8 |
| 6 | 10 | 60 | 2 | 4 | 40 |
| 7 | 3 | 21 | 3 | 9 | 27 |
| 8 | 4 | 32 | 4 | 16 | 64 |
| Total | 53 | 212 | | 60 | 300 |
| Media | 4 | | | | |



Paso 1: Diferencia entre el valor de la variable y su media (columna 4)

Paso 2: Elevar al cuadrado esas diferencias (columna 5)

Paso 3: Multiplicar lo obtenido en el paso anterior por su frecuencia absoluta (columna 6)

Paso 4: Sumar el resultado de la columna anterior

Paso 5: Dividir lo anterior por la cantidad de casos

En formula $V(X) = \sum (x - \mu)^2 * FA$

$V(X) = 300 / 53 = 5.67$

c) Frecuencias agrupadas por intervalos

| Li | Ls | FA | MC | MC-μ | (MC-μ) ² | (MC-μ) ² *FA |
|-------|-------|--------------------------------|-----|-------|---------------------|-------------------------|
| 155,5 | 160,5 | 2 | 158 | -16,7 | 278,89 | 557,78 |
| 160,5 | 165,5 | 4 | 163 | -11,7 | 136,89 | 547,56 |
| 165,5 | 170,5 | 9 | 168 | -6,7 | 44,89 | 404,01 |
| 170,5 | 175,5 | 11 | 173 | -1,7 | 2,89 | 31,79 |
| 175,5 | 180,5 | 11 | 178 | 3,3 | 10,89 | 119,79 |
| 180,5 | 185,5 | 8 | 183 | 8,3 | 68,89 | 551,12 |
| 185,5 | 190,5 | 3 | 188 | 13,3 | 176,89 | 530,67 |
| 190,5 | 195,5 | 1 | 193 | 18,3 | 334,89 | 334,89 |
| | | 49 | | | | 3077,61 |
| Media | 174,7 | Var (X) = 3077.61 / 49 = 62,80 | | | | |



Paso 1: Para el cálculo de la varianza primero se obtiene la marca de clase de cada intervalo (MC), que va a ser el valor representativo de X (Columna 4)

Paso 2: Se calcula la diferencia entre cada valor de X (Marca de Clase) y la media de la distribución (columna 5)

Paso 3: Se eleva al cuadrado lo obtenido en la columna anterior (columna 6)

Paso 4: Se multiplica lo obtenido en la columna anterior por la FA (columna 7)

Paso 5: Se suman los valores de la columna anterior y se los divide por la cantidad de casos

En formulas

$$\text{Var (X)} = (\sum (X_i - \mu)^2 * \text{FA}) / n$$