

III Medio Electivo

Cálculo de las medidas de dispersión para datos agrupados

Media, desviación media, varianza, desviación típica o estándar y coeficiente de variación (CV)

Profesores: Francisco Rozas – Nestor Albano
rozasfrancisco@Gmail.com - nalbano1@Gmail.com

Unidad 1

Objetivo: Construir tablas de frecuencia para datos dados, determinar intervalos de clase y calcular medidas de dispersión (media, desviación media, varianza, desviación típica y coeficiente de variación).

Habilidades: -Reconocer
-Aplicar
-Resolver

Construcción de tablas de frecuencia para datos agrupados

Construcción de tablas de frecuencia.

Datos agrupados.

Cuando el tamaño de la muestra es considerable o grande y los datos numéricos son muy diversos ($n > 15$), conviene agrupar los datos de tal manera que permita establecer patrones, tendencias o regularidades de los valores observados. De esta manera podemos condensar y ordenar los datos tabulando las frecuencias asociadas a ciertos intervalos de los valores observados.

Intervalos de Clase: Son los intervalos en los que se agrupan y ordenan los valores observados. Cada uno de estos intervalos está delimitado (acotado) por dos valores extremos que les llamamos límites.

Pasos a seguir para construir intervalos de clase.

- ## 1. Determinar la cantidad de intervalos apropiada

Una forma de determinar la cantidad de intervalo que se necesitan para construir la tabla es usando la regla de Sturges:

$K = 1 + 3.3 \log n$ Donde: k = número de intervalos
 Log = logaritmo en base 10
 n = número de datos.

Otra regla usada para datos < 50 es $K = \sqrt{n}$

El número de intervalos determinado mediante cualquier regla se aproxima al valor entero más cercano

2.- Calcular el rango de los datos.

El rango se representa por la letra R y se calcula como: $R = \text{dato mayor} - \text{dato menor}$

3.- Obtención de la amplitud o anchura que tendrá cada intervalo.

Se encuentra dividiendo el rango por el número de intervalos regularmente es de 5 a 6. Se representa con la letra A de tal manera que $A = R/k$

4.- Construcción de los intervalos.

Los intervalos de clase son conjuntos numéricos y deben ser excluyentes y exhaustivos; es decir, si un dato pertenece a un intervalo determinado, ya no podrá pertenecer a otro, esto quiere decir excluyentes y además todos y cada uno de los datos deberá estar contenido en alguno de los intervalos, esto les da el valor de exhaustivos.

Se construyen intervalos cerrados por la izquierda y abiertos por la derecha; esto se simboliza a través del uso de corchetes y paréntesis respectivamente. El último intervalo será cerrado por ambos extremos.

El primer intervalo se construye de la siguiente manera: Habrá de iniciar con el dato menor, el cual será el extremo inferior del intervalo; el otro extremo se obtiene de la suma del dato menor y la amplitud, con este mismo valor iniciamos el segundo intervalo, del cual el segundo extremo se encuentra sumando al valor anterior la amplitud y este proceso se repite sistemáticamente hasta completar el total de intervalos indicado por la regla elegida, por ejemplo la de Sturges.

Los valores extremos o límites de intervalo.

Límite inferior: Es el valor menor de cada intervalo, se denota por Li

Límite superior: Es el número mayor de cada intervalo, se denota por Ls.

Marca de Clase (MC) de cada intervalo: Se refiere al Punto Medio del intervalo y a través de él representaremos a todo el intervalo y una de las maneras de calcularla es promediando los valores límite de cada intervalo, su

fórmula es: $Mc = \frac{Li + Ls}{2}$

Ejemplo: Un grupo de investigadores pertenecientes a la secretaría de seguridad pública, tomó una muestra aleatoria de las velocidades (km/h) registradas por 30 vehículos en el trayecto Santiago a Rancagua, con el fin de establecer nuevos límites máximos de velocidad para una carretera. La muestra arrojó los datos siguientes:

90, 99, 104, 99, 119, 98, 95, 112, 95, 120, 100, 90, 116, 96, 114, 108, 98, 118, 100, 106, 114, 100, 112, 106, 100, 115, 111, 105, 114, 97

Toda vez que se tienen los datos, se recomienda ordenarlos de menor a mayor o viceversa

90, 90, 95, 95, 96, 97, 98, 98, 99, 99, 100, 100, 100, 100, 104, 105, 106, 106, 108, 111, 112, 112, 114, 114, 114, 115, 116, 118, 119, 120

1º obtendremos el número de intervalos que vamos a utilizar, para lo cual empleamos la Regla de Sturges:

$$K = 1 + 3.3 * \log(30) = 1 + 3.3 * (1.4771212547) = 1 + 4.87 = 5.87 \approx 6$$

2º calculamos el rango de variación,

$$R = 120 - 90 = 30$$

3º obtenemos la amplitud de cada intervalo de clase como sigue:

$$Ac = \frac{R}{K} = \frac{30}{6} = 5$$

4º construimos los intervalos: el primero de ellos inicia con 90 que es el extremo inferior que, sumado a 5 obtenemos 95, que será el extremo superior; este extremo será el inferior del segundo intervalo; y al sumar nuevamente la amplitud tendremos 100 que será el extremo superior y así sucesivamente hasta completar los 6 intervalos., que se muestran enseguida:

[90 – 95), [95 – 100), [100 – 105), [105 – 110), [110 – 115) y [115 – 120]

Los corchetes expresan que el valor extremo se incluye en el intervalo y los paréntesis dan a entender que el valor extremo del intervalo no se incluye en el.

Construyamos la tabla.

Intervalos de clases	Frecuencia absoluta f_i	Frecuencia absoluta acumulada F_{ac}	Marca de clase x_i
[90 - 95)	2	2	92,5
[95 - 100)	8	10	97,5
[100 - 105)	5	15	102,5
[105 - 110)	4	19	107,5
[110 - 115)	6	25	112,5
[115 - 120]	5	30	117,5
Total	30		

Cálculo de medidas de dispersión

Para calcular las medidas de dispersión usaremos las siguientes fórmulas:

$$\text{Media } (\bar{X}) = \frac{\sum x_i \cdot f_i}{n} \quad \begin{array}{l} \text{con } x_i: \text{ marca de clase,} \\ f_i: \text{ frecuencia absoluta.} \end{array}$$

$$\text{Desviación media (DM)} = \frac{\sum |x_i - \bar{X}| \cdot f_i}{n} \quad \text{con } \bar{X}: \text{ Media}$$

$$\text{Varianza } (\sigma^2) = \frac{\sum |x_i - \bar{X}|^2 \cdot f_i}{n} \quad \begin{array}{l} \text{con } x_i: \text{ marca de clase} \\ \bar{X}: \text{ Media} \end{array}$$

$$\text{También se puede usar } \sigma^2 = \frac{\sum x_i^2 \cdot f_i}{n} - \bar{X}^2$$

$$\text{Desviación típica } (\sigma) = \sqrt{\sigma^2} \quad \text{Raíz cuadrada de la varianza}$$

$$\text{Coeficiente de variación (CV)} = \frac{\sigma}{\bar{X}}$$

Calculemos las medidas de dispersión completando la tabla de frecuencias con las columnas que se necesitan de acuerdo a las fórmulas que usaremos en el ejemplo anterior.

Intervalos de clases	Frecuencia absoluta f_i	Frecuencia absoluta acumulada F_{ac}	Marca de clase x_i	$x_i * f_i$	$ x_i - X $	$ x_i - X * f_i$	$ x_i - X ^2$	$ x_i - X ^2 * f_i$
[90 - 95)	2	2	92,5	185	13.16	26.32	173.1856	346.3712
[95 - 100)	8	10	97,5	780	8.16	65.28	66.5856	532.6848
[100 - 105)	5	15	102,5	512.5	3.16	15.8	9.9856	49.928
[105 - 110)	4	19	107,5	430	1.84	7.36	3.3856	13.5424
[110 - 115)	6	25	112,5	675	6.84	41.04	46.7856	280.7136
[115 - 120)	5	30	117,5	587.5	11.84	59.2	140.1856	700.928
Total (Σ)	30			3170		215		1924.168

$$\text{Media } (X) = \frac{\sum x_i * f_i}{n} = \frac{3170}{30} = 105,66 \text{ km/h} \quad (\text{DM}) = \frac{\sum |x_i - X| * f_i}{n} = \frac{215}{30} = 7.16 \text{ km/h}$$

$$\sigma^2 = \frac{\sum |x_i - X|^2 * f_i}{n} = \frac{1924,168}{30} = 64,13 \quad ; \quad \sigma = \sqrt{\sigma^2} = \sqrt{64,13} = 8,01 \text{ km/h} \quad ; \quad \text{CV} = \frac{8,01}{105,66} = 0,0758 * 100 = 7,58\%$$

Usando el mismo método, resuelve los siguientes ejercicios:

- 1) Los resultados de un test de aptitud tomado a un grupo de 100 personas se volcaron en la siguiente tabla: Complete la tabla y calcule : Media, DM, varianza, desviación típica, y CV.

Intervalos de clase	Frecuencia f_i							
[0,5 - 5,5)	7							
[5,5 - 10,5)	12							
[10,5 - 15,5)	21							
[15,5 - 20,5)	32							
[20,5 - 25,5]	28							
Total								

- 2) Los datos que se dan a continuación corresponden a los pesos en Kg. de ochenta personas: determine cantidad de intervalos; construya una tabla de frecuencia y calcule todas la medidas de dispersión.

60; 66; 77; 70; 66; 68; 57; 70; 66; 52; 75; 65; 69; 71; 58; 66; 67; 74; 61; 63; 69; 80; 59; 66; 70; 67; 78; 75; 64; 71; 81; 62; 64; 69; 68; 72; 83; 56; 65; 74; 67; 54; 65; 65; 69; 61; 67; 73; 57; 62; 67; 68; 63; 67; 71; 68; 76; 61; 62; 63; 76; 61; 67; 67; 64; 72; 64; 73; 79; 58; 67; 71; 68; 59; 69; 70; 66; 62; 63; 66;

<https://www.youtube.com/watch?v=AbN977Xd96k>

Soluciones:

$$1) \text{ Media} = X = \frac{\sum x_i \cdot f_i}{n} = \frac{1610}{100} = 16,1$$

$$DM = \frac{\sum |x_i - X| \cdot f_i}{n} = \frac{508}{100} = 5,8$$

$$\sigma^2 = \frac{\sum |x_i - X|^2 \cdot f_i}{n} = \frac{3639}{100} = 36,39$$

$$\sigma = \sqrt{36,39} = 6,03$$

$$CV = \frac{\sigma}{X} \cdot 100 = \frac{6,03}{16,1} \cdot 100 = 37,45 \%$$

$$2) \text{ Media} = X = \frac{\sum x_i \cdot f_i}{n} = \frac{5396}{80} = 67,45 \text{ kg}$$

$$DM = \frac{\sum |x_i - X| \cdot f_i}{n} = \frac{378,5}{80} = 4,73 \text{ kg}$$

$$\sigma^2 = \frac{\sum |x_i - X|^2 \cdot f_i}{n} = \frac{2951,8}{80} = 36,8975$$

$$\sigma = \sqrt{36,8975} = 6,07 \text{ kg}$$

$$CV = \frac{\sigma}{X} \cdot 100 = \frac{6,07}{67,45} \cdot 100 = 8,99 \%$$

¿Qué aprendiste hoy?

¿Te fue fácil resolver los ejercicios?

¿Tuviste alguna dificultad?