

SESIÓN 5

RELACIÓN ENTRE LOS VALORES NUMÉRICOS

I. CONTENIDOS:

1. Relación entre valores numéricos.
2. Cálculo de media, mediana y moda en datos agrupados y no agrupados.
3. La media, mediana y moda en variable nominal y ordinal.

II. OBJETIVOS:

Al término de la Sesión, el alumno:

- Razonará la importancia de las medidas de tendencia central en el ámbito profesional.

III. PROBLEMATIZACIÓN:

Comenta las preguntas con tu Asesor y selecciona las ideas más significativas.

- ¿Es posible realizar los cálculos de las medidas de tendencia central de la misma forma para todos los tipos de variables y cualquier tipo de tabla estadística?
- ¿Qué significan los valores numéricos de media mediana y moda bajo un caso práctico?

IV. TEXTO INFORMATIVO-FORMATIVO:

1.1. Relación entre valores numéricos

La distribución de las categorías de una variable se puede presentar con una coincidencia de sus medidas de tendencia central, si es el caso, se dice que es una distribución simétrica. La mediana queda, por definición, entre la media y la moda; cuando las medidas de tendencia central no son coincidentes, se dice entonces que la distribución tiene un sesgo.

El sesgo es positivo si la media es mayor que las otras medidas de tendencia central, en el histograma pareciera que hay un corte en el lado izquierdo, es decir hay menos categorías que en el lado derecho de la mediana. Por otra parte, la distribución tiene sesgo negativo si la media es menor que la moda y la mediana, la distribución tiene menos categorías en la parte derecha de la mediana que en la parte izquierda. Para las distribuciones asimétricas, las que tienen sesgo, el sintetizador más adecuado es la mediana.

2.1. Cálculo de media, mediana y moda en datos agrupados y no agrupados

DATOS NO AGRUPADOS

Cuando las observaciones de las categorías de una variable **no** están integradas en una tabla de distribución de frecuencias con intervalos, entonces se dice que la distribución es con datos no agrupados. La manera de calcular las medidas de tendencia central en datos no agrupados ya se explicó en la clase anterior, sin embargo, no está de más otro ejemplo:

Ejemplo 1: Se hizo una encuesta entre los niños de una escuela primaria federal en el Estado de Jalisco sobre las horas que estaban frente a la televisión cada día. Calcule la media, mediana y moda con la información recabada.

Horas frente a la televisión	Total 930
1	99
2	293
3	359
4	121
5	58

Solución: Ahora etiquetemos las categorías con X, luego se elabora la columna de fX , multiplicando las categorías por las frecuencias, la columna de frecuencia acumulada (fa) se construye como ya se ha explicado antes:

X	f	fX	fa
1	99	99	99
2	293	586	392
3	359	1077	751
4	121	484	872
5	58	290	930
SUMA	930	2536	

La media se calcula con la fórmula: $\bar{X} = \frac{\sum fX}{\sum f} = \frac{2536}{930} = 2.7$

La media es 2.7 horas. Significa que en promedio estos niños ven 2.7 horas la televisión cada día. La posición de la mediana se calcula sumado 1 a $\sum f$ y el resultado se divide entre dos:

$$\frac{930 + 1}{2} = 465.5$$

Luego buscamos el valor inmediatamente mayor a este en la columna de frecuencias acumuladas. Resulta ser: 751

Ubicamos entonces el número de horas que corresponde a esa frecuencia acumulada y esta es la mediana. La mediana es: 3 horas. Significa que por lo menos la mitad de los niños de esta escuela ven la televisión 3 horas cada día.

La moda es el dato de mayor frecuencia. Buscamos en la segunda columna (la de frecuencias) el valor mayor (359) y ubicamos el dato que corresponde en la primera columna (3), esta es la moda. La moda es 3 horas. Significa que el caso más típico observado fue de los niños que ven la televisión 3 horas al día.

DATOS AGRUPADOS

Cuando los datos de la variable están agrupados, es posible también calcular las medidas de tendencia central de manera aproximada, hay que recordar que al construir una distribución de frecuencias se pierden muchos detalles y esta es la razón de que el resultado sea aproximado.

Para calcular las medidas de tendencia central en datos agrupados, se utilizan las siguientes fórmulas, para la media:

$$\bar{X} = \frac{\sum fX}{\sum f}$$

Dónde f es la frecuencia de cada intervalo de clase, X es la marca de clase de cada intervalo.

Para la mediana

$$Me = LRI + \left[\frac{\frac{N}{2} - \sum faa}{f} \right] a$$

Dónde N es el número total de datos de la distribución, LRI es el límite real inferior del intervalo dónde se localiza la mediana, faa es la frecuencia acumulada del intervalo anterior al que contiene a la mediana, f la frecuencia del intervalo dónde está la mediana y, a la amplitud de clase de los intervalos. Para localizar el intervalo dónde se encuentra la mediana, se suma uno al total de datos y el resultado se divide entre dos, luego se busca en la columna de frecuencias

acumuladas el valor inmediato superior, una vez localizado este valor, se señala el intervalo al que pertenece pues es dónde se encuentra la mediana. La Moda es la marca de clase del intervalo de mayor frecuencia.

Ejemplo 2: trescientas tres probetas extensibles de cierto barniz o laca quebradiza (empleada en el análisis experimental de esfuerzos) produjeron los valores siguientes de resistencia después de 7 días:

Intervalo de resistencia KN/m ²	Número de piezas de prueba
200 – 230	7
230 – 260	30
260 – 290	50
290 – 320	77
320 – 350	53
350 – 380	40
380 – 410	35
410 – 440	6
440 – 470	3
470 – 500	2

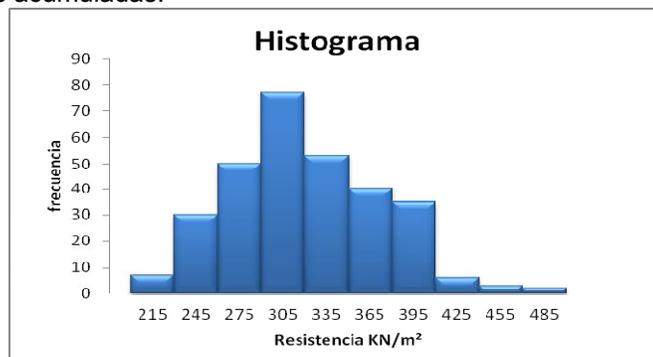
Trace el histograma, de frecuencias acumuladas, calcule la media, mediana y moda.

Solución: Debemos primero construir la tabla de frecuencias, recordando que la marca de clase se calcula sumando los límites de cada intervalo y dividiendo el resultado entre dos:

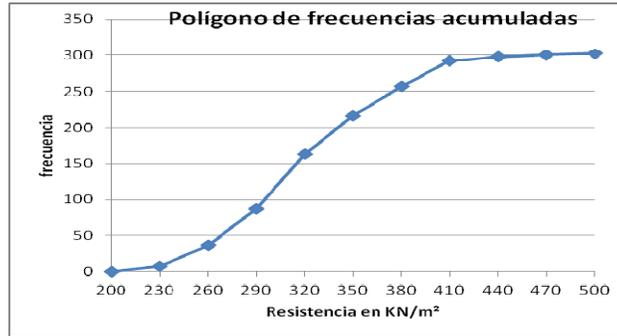
Intervalo	Marca de clase (X)	f	fX	f.a.
200 – 230	215	7	1505	7
230 – 260	245	30	7350	37
260 – 290	275	50	13750	87
290 – 320	305	77	23485	164
320 – 350	335	53	17755	217
350 – 380	365	40	14600	257
380 – 410	395	35	13825	292
410 – 440	425	6	2550	298
440 – 470	455	3	1365	301
470 – 500	485	2	970	303

303 97155

Utilizando los procedimientos antes vistos podemos graficar la distribución en un histograma y polígono de frecuencias acumuladas:



En el histograma de la distribución, la marca de clase de cada intervalo debe ponerse como centro de cada rectángulo cuya base representa la amplitud de clase de cada intervalo y la altura la frecuencia de dicho intervalo.



El gráfico de frecuencias acumuladas nos permite conocer la cantidad de piezas que tienen una resistencia “menor que” un valor de los que se representan en el eje horizontal. En el eje horizontal se disponen los límites comenzando por el LRI del primer intervalo y terminando con el LRS del último intervalo, la primera frecuencia es cero y las demás son las de la columna de frecuencias acumuladas. A este gráfico también se le llama Ojiva.

La media se calcula con la fórmula y sustituyendo adecuadamente:

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{97155}{303} = 320.6$$

Significa que el promedio de resistencia de las probetas es de 320.6 KN/m²

El intervalo donde se localiza la mediana se calcula como antes se mencionó

$$\frac{303 + 1}{2} = 152$$

Luego se busca el valor inmediatamente mayor en la columna de frecuencias acumuladas en la tabla, para encontrar el intervalo donde se encuentra la mediana:

Intervalo	Marca de clase (X)	f	fX	f.a.
200 – 230	215	7	1505	7
230 – 260	245	30	7350	37
260 – 290	275	50	13750	87
290 – 320	305	77	23485	164
320 – 350	335	53	17755	217
350 – 380	365	40	14600	257
380 – 410	395	35	13825	292
410 – 440	425	6	2550	298
440 – 470	455	3	1365	301
470 – 500	485	2	970	303
			303	97155

De este cálculo se derivan los valores de LRI = 290, $\sum faa = 87$, $f = 77$, $a = 30$.

Sustituyendo en la fórmula

$$= 290 + \left[\frac{\frac{303}{2} - 87}{77} \right] 30 = 290 + \left[\frac{64.5}{77} \right] 30 = 290 + 25.1 = 315.1$$

La mediana es 315.1 KN/m². Significa que por lo menos la mitad de las probetas tienen una resistencia menor a 315.1 KN/m².

Como sabemos la moda es la marca de clase del intervalo con mayor frecuencia:

- La mayor frecuencia es: 77 y la marca de clase de este intervalo es 305.
- La moda es 305. Significa que la resistencia más común de las probetas debe ser 305 KN/m².

Si graficamos las medidas de tendencia central en el histograma con líneas verticales podemos observar cómo se encuentran en la distribución:

3.1. La media, mediana y moda en variable nominal y ordinal

En las variables nominal y ordinal no se puede calcular la media, pues implica la sumatoria y no se pueden sumar categorías que no son números. La mediana sólo se puede calcular en la variable ordinal.

Ejemplo 3 Para el ejemplo antes visto de las evaluaciones de un curso de capacitación a un grupo de trabajadores. Calcula la moda y la mediana

Evaluación	Frecuencia	Frecuencia acumulada
	25	%
Excelente	2	2
Muy bien	4	6
Bien	7	13
Regular	9	22
No aprobado	3	25

Tal como ya hemos visto, la moda es la categoría que presenta mayor frecuencia. En este caso es “Regular”, pues su frecuencia es la mayor (9). Significa que la evaluación más común de los trabajadores en su curso fue “Regular”

Por otra parte, la mediana se determina a partir del valor central. Para determinarla, se suma el número uno al total de datos y el resultado se divide entre dos:

$$\frac{25+1}{2} = 13$$

Luego, se busca en las frecuencias acumuladas el valor inmediato mayor. En este caso es el mismo trece pues coinciden. El dato número 13 que nos dio la operación anterior está en la categoría “Bien”. Significa que por lo menos la mitad de los trabajadores tienen una evaluación de “Bien” o más

V. ESTRATEGIAS CENTRADAS EN EL APRENDIZAJE:

Calcula e interpreta las medidas de tendencia central. Especifica el tipo de sesgo y el sintetizador más adecuado. Construye también el histograma y la ojiva o gráfico de frecuencias acumuladas.

Los siguientes datos son el peso en miligramos de tabletas bufferin.

672.2	655.7	684.7	663.2	666.7	672.8	673.9	670.0
667.4	667.0	662.7	670.4	680.2	669.4	659.7	680.0
665.2	672.6	655.1	677.3	659.7	673.2	662.3	663.7
671.3	665.5	659.5	652.2	674.3	670.3	660.0	672.6
679.2	662.5	661.3	671.7	658.3	680.1	673.5	671.9
667.0	672.2	669.1	682.2	667.9	663.4	681.5	657.3
657.5	672.9	660.5	657.8	672.6	672.3	673.9	677.6
662.1	665.6	654.2	672.6	679.9	671.8	652.7	682.3
669.8	662.2	663.6	667.9	658.1	653.4	661.6	671.7
670.7	664.8	658.1	670.4	680.0	663.2	672.7	668.5