

MÉTODOS ESTADÍSTICOS BÁSICOS

GRADO II



CONTENIDOS

- Referentes de aprendizaje
- Población, Muestra, Variable
- Medidas de tendencia central
- Medidas de dispersión
- Distribución de frecuencias
- Datos NO agrupados y agrupados
- Probabilidad
- Combinaciones y Permutaciones

■ METODOS DE APRENDIZAJE

Estudio de un problema o pregunta de interés a la comunidad de la institución y/o ciudadanía.

■ Estándares

- Usar comprensivamente algunas medidas de centralización, localización, dispersión y correlación.
- Diseñar experimentos aleatorios para estudiar un problema o pregunta.

REFERENTES DE APRENDIZAJE

Propósitos

- Reconocer la importancia del análisis estadístico, para manejar apropiadamente los conceptos e interpretaciones de medidas y representaciones de conjuntos de datos, y así aplicar los conocimientos adquiridos para la resolución de problemas de situaciones cotidianas.

Desempeños

- Comprende la importancia del análisis estadístico para la resolución de problemas de situaciones cotidianas, aplicando en forma acertada los conocimientos estadísticos para la interpretación e inferencia de tablas, gráficos y conjuntos de datos.

MEDIDAS DE TENDENCIA CENTRAL

Medidas básicas

- Son medidas estadísticas que proponen "resumir" en un solo valor la tendencia de un conjunto de datos o aspecto particular.
- **Media.** La medida numérica central obtenida por la suma total de mediciones de la muestra dividida por el total de datos.
- **Moda.** La medida más frecuente (repetida) de la muestra.
- **Mediana.** La medida posicionada en el centro de la muestra, luego que los datos han sido ordenados (alfa-numéricamente).

Notación y fórmulas

- Descripción:

$$\bar{X} \equiv \mu = \frac{1}{N} \sum_{i=1}^N x_i$$

$$M_o = \{\text{Dato(s) más repetido(s)}\}$$

$$M_e = \{\text{Ordenar datos y obtener el central}\}$$

MEDIDAS DE DISPERSIÓN

Dispersión

- Las medidas de dispersión son números (positivos), que informan sobre cuánto se alejan los datos de la medida central.
- Desviación: $x - \bar{X}$
- Cuando los datos son iguales este valor es cero.
- Cuando los datos se vuelven más diversos este se incrementa.
- Determinación requiere cálculo de la media (obligatorio).

Clases

- Desviación media

$$\begin{aligned} D_m &= \frac{\text{Suma valor absoluto de desviaciones}}{\text{Número total datos}} \\ &= \frac{\sum |x - \bar{X}|}{n} \end{aligned}$$

- Desviación estándar

$$\begin{aligned} \sigma &= \sqrt{\frac{\text{Suma cuadrado de desviaciones}}{\text{Número total datos}}} \\ &= \sqrt{\frac{\sum (x - \bar{X})^2}{n}} \end{aligned}$$

DISTRIBUCIÓN DE FRECUENCIA

Definición

- Método de recolección de datos según su magnitud y que indican el número de observaciones en cada una de ellas.
- Frecuencia (repetición)
- Clases:
 - Frecuencia absoluta (conteo)
 - Frecuencia relativa (porcentual respecto a total de datos)
 - Frecuencia acumulada (suma consecutiva dato anterior).
 - Frecuencia acumulada relativa
- Gráficos
 - Histograma; Polígono de frecuencias

Ejemplo

- Se indagó por el número de hermanos que tienen los 25 alumnos de un curso de grado sexto. Los resultados se presentan en la siguiente tabla.

No de hermanos	No de alumnos	% de alumnos	Frecuencia Acumulada
	Frecuencia absoluta	Frecuencia relativa	
0	3	12%	
1	8	32%	
2	6	24%	
3	4	16%	
4	2	8%	
5	2	8%	

ANÁLISIS ESTADÍSTICO DE DATOS

NO agrupados

- Independiente de la Población (pequeño o grande).
- La muestra es menor a 20 - 30 elementos.
- La información está en bruto (no necesita clasificación).
- El manejo o tratamiento puede ser manual (a suela!).



Agrupados

- Independiente de la Población (pequeño o grande).
- La muestra es mayor a 30 elementos.
- La información debe ser refinada: es aconsejable agrupar los datos en clases o categorías.
- El manejo o tratamiento debe ser metódico (fórmulas, tablas) y numérico (calculadora, programas informáticos).

A photograph of a rectangular piece of paper. The top half of the paper is covered with a dense, random pattern of small black dots. The bottom half of the paper has the handwritten formula $P_v = n R T$ in black ink.

- Objetivo: resumir la información en tendencias numéricas (media, desviación, gráficos, ...).

ANÁLISIS EN DATOS AGRUPADOS

Definición: objeto

- Los elementos (datos) son clasificados por grupos o intervalos según su frecuencia.
- Con datos agrupados se pierde detalle de la información original, pero se reduce la información numérica.
- La información (muestra) es resumida en una distribución de frecuencia (tablas, gráficos).

Clasificación: intervalo de clase

- En esencia, localizar y contar un dato dentro de los valores numéricos de un intervalo.

ANÁLISIS EN DATOS AGRUPADOS

Definición: objeto

- Los elementos (datos) son clasificados por grupos o intervalos según su frecuencia.
- Con datos agrupados se pierde detalle de la información original, pero se reduce la información numérica.
- La información (muestra) es resumida en una distribución de frecuencia (tablas, gráficos).

Clasificación: intervalo de clase

- En esencia, localizar y contar un dato dentro de los valores numéricos de un intervalo.

Ejemplo ilustrativo

■ Consumo bebidas en el PAE

Se necesita realizar un estudio sobre el número de bebidas (jugos, lácteos) que consume la comunidad educativa del colegio en un mes. De la población del colegio se elige una muestra de 100 personas.



ANÁLISIS EN DATOS AGRUPADOS

Distribución de frecuencia

Tabla de frecuencia		
Intervalo de clase	Frecuencia Absoluta	Frecuencia acumulada
Bebidas consumidas	Número de estudiantes	Número de estudiantes
I – 10		
11 - 20		
21 - 30		
31 - 40		
Total		

Ejemplo ilustrativo

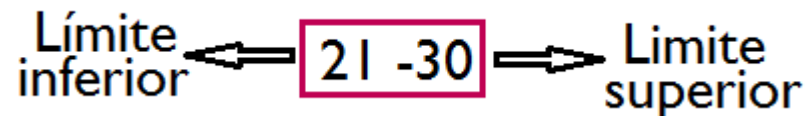
■ Consumo bebidas en el PAE

31	18	10	27	14	31	18	10	27	14
12	24	16	28	20	12	24	16	28	20
13	30	22	9	26	13	30	22	9	26
19	11	23	15	32	19	11	23	15	32
25	17	29	21	8	25	17	29	21	8
31	18	10	27	14	31	18	10	27	14
12	24	16	28	20	12	24	16	28	20
13	30	22	9	26	13	30	22	9	26
19	11	23	15	32	19	11	23	15	32
25	17	29	21	8	25	17	29	21	8

Oiga! No olvide que son 4 columnas. Si! Ud. expositor!

ANÁLISIS EN DATOS AGRUPADOS

■ Intervalo de clase



■ Marca de clase (Mc)

Promedio de los límites de clase; es la medida representativa del intervalo.

■ Ancho de clase real (C)

Diferencia entre dos marcas de clase contiguas; puede ser (des)igual.

Con las marcas de clase y frecuencias relativas se calculan las medidas estadísticas.

Ejemplo ilustrativo

Tabla de frecuencia		
Intervalo de clase	Frecuencia Absoluta (f_i)	Marca de clase (Mc)
1 - 10		
11 - 20		
21 - 30		
31 - 40		
Total		

ANÁLISIS EN DATOS AGRUPADOS

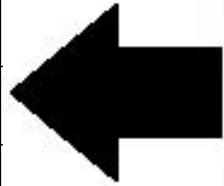
Cálculo de intervalos de clase

Tabla de frecuencia		
Intervalo	f_i	Mc
I – 10		
II - 20		
2I - 30		
3I - 40		
Total		

ANÁLISIS EN DATOS AGRUPADOS

Cálculo de intervalos de clase

Tabla de f
Intervalo
I – 10
II - 20
2I - 30
3I - 40
Total



ANÁLISIS EN DATOS AGRUPADOS

Cálculo de intervalos de clase

■ Rango (R)

Muestra el ancho numérico de variación de la muestra. Diferencia del dato máximo con el dato mínimo.

■ Número de clases (k)

Número de intervalos de agrupación; depende del total de datos (N).

- Elección empírica (justificada).
- Fórmula de Sturges (redondeada a un entero*):

$$k = 1 + \frac{\ln N}{\ln 2}$$

- Regla de Rice (de arroz!) (redondeada a un entero*):

$$k = 2\sqrt[3]{N}$$

■ Ancho de clase (C)

División del rango entre el número de clases:
 $C=R/k$.

■ Formar las clases

Poner como límite inferior del primer intervalo un valor un poco menor que el dato menor hallado en la muestra y posteriormente se suma a este el valor C; para hallar el límite superior del primer intervalo se “disminuye un poco” el límite inferior del siguiente intervalo (1 unidad, 1 décima, etc.; puede disminuirse 1 en la última cifra significativa).

Se procede a obtener los límites del intervalo siguiente y así sucesivamente. 14

ANÁLISIS EN DATOS AGRUPADOS

Cálculo de medidas estadísticas

■ Notación - símbolos

La evaluación de las medidas, requiere una notación adecuada para el uso de fórmulas.

Estas fórmulas provienen de conceptos elementales de datos NO agrupados y algunos argumentos geométricos implementados sobre un histograma.

Símbolo	Descripción
x_i	(Léase x sub i) Cualquier marca de clase en la muestra
N	Total de elementos
f_i	Frecuencia absoluta
h_i	Frecuencia acumulada
C	Ancho de clase real
Δ	Resta o diferencia de valores
Σ_i	Sume los elementos desde 1ro. hasta el último

ANÁLISIS EN DATOS AGRUPADOS

Cálculo de medidas estadísticas

- Media (promedio) aritmética

El valor típico o representativo del conjunto de datos.

$$\bar{X} = \frac{\sum_i f_i x_i}{N}$$

Símbolo	Descripción
x_i	Marca de clase
f_i	Frecuencia absoluta
N	Total de elementos

Ejemplo. Media en el consumo de bebidas PAE

Tabla de frecuencia			
Intervalo	f_i	x_i	$f_i x_i$
1 - 10	12	5,5	
11 - 20	40	15,5	
21 - 30	40	25,5	
31 - 40	8	35,5	
Total	100		

ANÁLISIS EN DATOS AGRUPADOS

Cálculo de medidas estadísticas

■ Mediana

Es el valor (eje horizontal) que divide un histograma en dos partes de igual área.

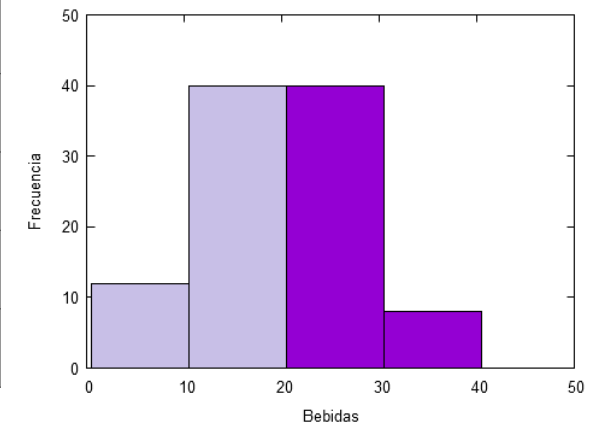
$$M_e = L_{\text{inf}} + \frac{\frac{N}{2} - h_{\text{ant}}}{f_{\text{med}}} C$$

Símbolo	Descripción
L_{inf}	Límite inferior (real) del intervalo mediano
N	Número total datos
h_{ant}	Frecuencia acumulada intervalo anterior al mediano
f_{med}	Frecuencia absoluta del intervalo mediano
C	Ancho de clase real

Ejemplo. Mediana en el consumo de bebidas PAE

■ Aquí, $N/2=50$

Tabla de frecuencia			
Intervalo	f_i	xi	hi
1 - 10	12	5,5	12
11 - 20	40	15,5	52
21 - 30	40	25,5	92
31 - 40	8	35,5	100
Total			



ANÁLISIS EN DATOS AGRUPADOS

Cálculo de medidas estadísticas

■ Moda

El valor más común de la muestra; el pico máximo de un polígono de frecuencia.

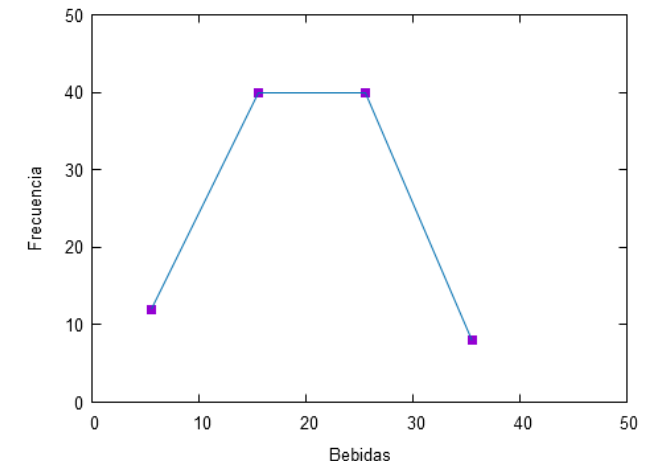
$$M_o = L_{\text{inf}} + \frac{\Delta_{\text{ant}}}{\Delta_{\text{ant}} + \Delta_{\text{post}}} C$$

Símbolo	Descripción
L_{inf}	Límite inferior (real) del intervalo modal
Δ_{ant}	Diferencia de frecuencia entre el intervalo modal y frecuencia del intervalo anterior
Δ_{post}	Diferencia de frecuencia entre el intervalo modal y frecuencia del intervalo posterior
C	Ancho de clase real

Ejemplo. Moda en el consumo de bebidas PAE

Tabla de frecuencia		
Intervalo	f_i	xi
1 - 10	12	5,5
11 - 20	40	15,5
21 - 30	40	25,5
31 - 40	8	35,5
Total	100	

■ La distribución es bimodal.



ANÁLISIS EN DATOS AGRUPADOS

Cálculo de medidas estadísticas

■ Desviación estándar (típica)

El valor numérico (estadístico) de la dispersión de los datos respecto a la media.

$$\sigma = \sqrt{\frac{\sum_i f_i x_i^2}{N} - \bar{X}^2}$$

Símbolo	Descripción
x_i	Marca de clase
f_i	Frecuencia absoluta
N	Total de elementos
\bar{X}	Media

Ejemplo. Desviación estándar en el consumo de bebidas PAE

Tabla de frecuencia			
Intervalo	f_i	x_i	$f_i x_i^2$
1 - 10	12	5,5	
11 - 20	40	15,5	
21 - 30	40	25,5	
31 - 40	8	35,5	
Total	100		

PROBABILIDAD

- **Suceso (A).** Un evento posible, real, cuantificable.
- **Espacio de Muestras (S).** El conjunto de TODOS modos en que se desarrolla un suceso.
- **Concepto de Probabilidad (P).** Medida de oportunidad de que ocurra un suceso respecto de un espacio de muestras.
- **Cálculo de una probabilidad (Regla de Laplace).**
Asumiendo sucesos equiprobables, la probabilidad de ocurrencia de un suceso se obtiene dividiendo el número de modos que forman el suceso A entre el número de modos del espacio muestral.

**¿CUÁL ES LA PROBABILIDAD
DE SACAR UNA BOLA AZUL?**



$$P(A) = \frac{N_A}{N_S}, \quad 0 \leq P(A) \leq 1$$

$$P(A') = 1 - P(A)$$

- Expresada en forma decimal o porcentual (x100).

PERMUTACIONES Y COMBINACIONES

- **Factorial.** $n! = 1.2.3 \dots n$
- **Permutación.** El número de grupos eligiendo k objetos de n objetos, teniendo en cuenta el orden (pueden repetirse).

$${}_nP_k = \frac{n!}{(n - k)!}$$

- **Combinaciones.** El número de grupos eligiendo k objetos de n objetos, sin tener cuenta el orden.

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n - k)!}$$

- **Principio fundamental del conteo.** “Contad y Multiplicad”. Si una cosa se puede hacer de n_1 modos, una segunda cosa de n_2 modos, ..., una k -ésima de n_k modos, entonces hay $n_1 n_2 \dots n_k$ modos de hacer las k cosas.
- **Diagrama de árbol.**

EJEMPLOS

- **Ejemplo 1.** ¿Cuántos números de 3 cifras se pueden formar con los números 4,5,6,7,8,9 sin repetir cifra? ¿Cuántas combinaciones hay?
- **Ejemplo 2.** En un restaurante local, las alas son la especialidad. El restaurante ofrece 3 tamaños de alas, 4 niveles de calor y salsa para untar ranchera o BBQ. ¿Cuántas órdenes diferentes son posibles?

APLICACIONES: DESCUBRIENDO PARTÍCULAS!

- El bosón de Higgs.
- Propuesta desde 1964 (teoría).
- Encargada de “dar masa” a las partículas.
- Cuánto pesa? $\text{masa} = X$



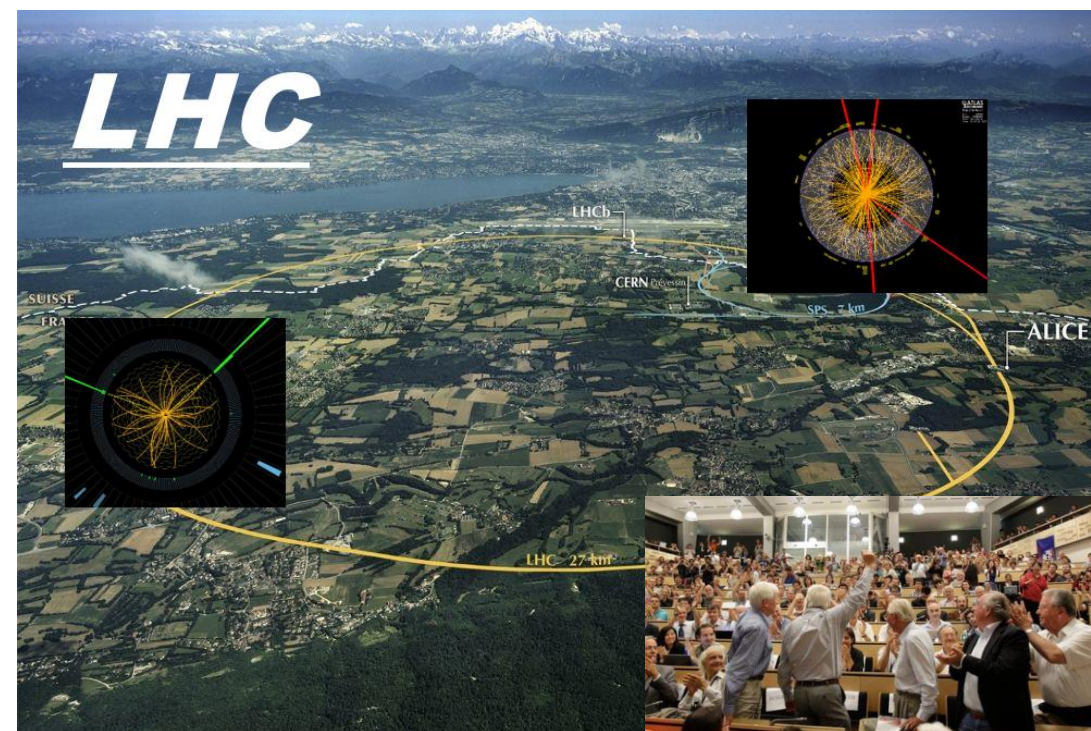
APLICACIONES: DESCUBRIENDO PARTÍCULAS!

- En física, una medida (estadística) se muestra como,

$$\bar{X} \pm \sigma$$

siempre y cuando el 99.999% de los datos experimentales se encuentren dentro del rango establecido.

- En el 2011, se detectó la partícula en un rango 115-140 GeV/c² con desviación estándar de 2.9, ocupando 99.998% del rango. NO HUBO CONFIRMACIÓN!
- En el 2012, nuevas revisiones estadísticas dejaron un rango de 122.5-127 GeV/c² y una desviación estándar de 0.5, dejando 99.999%. HUBO CONFIRMACIÓN!



- Y EN EL 2013 HUBO PREMIO NOBEL!

$$m_H = 125.3 \pm 0.6 \text{ GeV}/c^2$$

ACTIVIDAD 4 – EN CONSTRUCCIÓN

Realizar tabla de distribución de frecuencias completa y polígono de frecuencias para la situación del consumo de bebidas en el PAE

1. Usando la fórmula de Sturges.
2. Usando la regla de Rice.
3. Mencione las diferencias entre los métodos de clasificación.



31	18	10	27	14	31	18	10	27	14
12	24	16	28	20	12	24	16	28	20
13	30	22	9	26	13	30	22	9	26
19	11	23	15	32	19	11	23	15	32
25	17	29	21	8	25	17	29	21	8
31	18	10	27	14	31	18	10	27	14
12	24	16	28	20	12	24	16	28	20
13	30	22	9	26	13	30	22	9	26
19	11	23	15	32	19	11	23	15	32
25	17	29	21	8	25	17	29	21	8

REFERENCIAS

- [1] Blog de Abogados de El Salvador. (2015). *Cómo trabajar con tablas de datos agrupados en Estadística*. Recuperado el, 13 de febrero de 2019 de <http://blogdeabogadosdeelsalvador.com/2015/02/21/como-trabajar-con-tablas-de-datos-agrupados-en-estadistica/>
- [2] Murray R. Spiegel. (2000). *Estadística*. Madrid: McGraw-Hill.
- [3] Portal Educativo. (2012). *Media, moda y mediana para datos agrupados*. Recuperado el, 13 de febrero de 2019 de <https://www.portaleducativo.net/octavo-basico/792/Media-moda-y-mediana-para-datos-agrupados>
- [4] *Tratamiento Para Datos Agrupados*. Recuperado el, 13 de febrero de 2019 de http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Tratamiento%20para%20Datos%20Agrupados.htm

TAREAS

- Tarea 1
- ¿Qué es la frecuencia acumulada relativa?
- ¿Qué es el límite real superior/inferior de una clase?
- Tarea 2
- ¿Qué es la varianza?