

Programa de entrenamiento en epidemiología de campo

Módulo 2

Medidas de tendencia
central

Unidad 2. Parte 1



Objetivos de aprendizaje

- Explicar y calcular las siguientes medidas de ubicación central y dispersión: moda, mediana, media y rango.

Medidas de tendencia central

La medición de la ubicación central es un valor único que se supone que es un buen resumen de una distribución completa de datos. Revisaremos en esta unidad tres medidas: moda, mediana y media aritmética (promedio)

Moda

Definición: Valor que ocurre con mayor frecuencia en un conjunto de datos

Para identificar la moda:

1 Crear tabla de
distribución de frecuencia

2 Identifique el valor que
ocurre con más
frecuencia

La tabla 1 corresponde al periodo de incubación (días) de una muestra de 19 pacientes que desarrollaron la Enfermedad por el Virus del Ébola en África Occidental en 2014.

Tabla 1

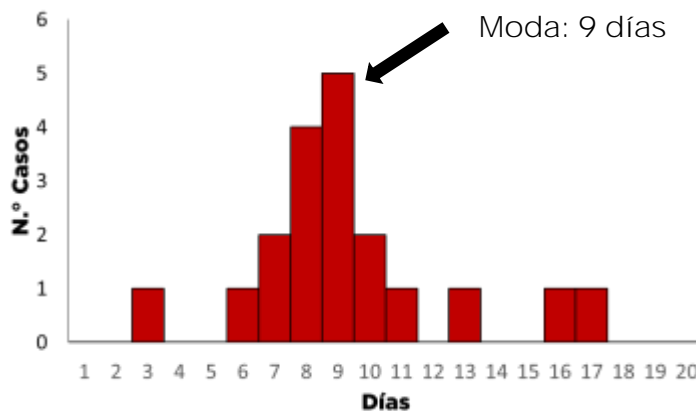
Obs	Días
1	9
2	8
3	16
4	9
5	10
6	7
7	9
8	9
9	6
10	3
11	11
12	8
13	17
14	9
15	8
16	13
17	8
18	10
19	7

Distribución de frecuencia por periodo de incubación (días)

Días	Frecuencia
3	1
4	0
5	0
6	1
7	2
8	4
9	5
10	2
11	1
12	0
13	1
14	0
15	0
16	1
17	1
Total	19

En la tabla de la derecha se ha realizado la distribución de frecuencia de los días de incubación de Ébola de los 19 casos, evidenciando que 9 días es el período de incubación más común (ocurre 5 veces), por lo tanto, la moda de esta distribución es 9.

También podemos realizar un histograma a partir de la tabla 1, e identificar el valor mas frecuente, como se ve en el siguiente gráfico:



Moda: Propiedades y usos

- Responde a la pregunta ¿Cuál es el grupo más común?
- Es la medida más sencilla de la ubicación central para comprender, explicar e identificar.
- En un conjunto de datos puede haber más de una moda, o puede que no exista moda
- La moda puede no ser "central", por lo cual esta afectada por los valores extremos o atípicos.

Mediana

Definición:

Valor medio; valor que divide la distribución en dos partes iguales

- El 50% de las observaciones son iguales o están por debajo de la mediana
- El 50% de las observaciones son iguales o están por encima de la mediana

Para calcular la mediana, siga estos pasos

- 1 Organice las observaciones en orden
- 2 Encuentre la posición media como $(n + 1) / 2$.
n: corresponde al número de registros u observaciones.
- 3 Identifique el valor en el medio

Siguiendo con el ejemplo del conjunto de datos disponibles del periodo de incubación (días) de 19 casos que desarrollaron la enfermedad de Ébola, calcularemos la mediana siguiendo los tres pasos descritos:



1. Organizar los datos de mayor a menor (columna de días de incubación)

Obs.	Días
1	3
2	6
3	7
4	7
5	8
6	8
7	8
8	8
9	9
10	9
11	9
12	9
13	9
14	10
15	10
16	11
17	13
18	16
19	17

9 observaciones iguales o por encima de la mediana

2. Encontrar la posición media = $\frac{(n+1)}{2}$
 $\frac{(19+1)}{2} = 10$

Mediana = 9

3. Mediana = valor en la posición 10° = 9

9 observaciones iguales o por debajo de la mediana

Observen que en este ejemplo el número de observaciones o datos es impar (19 registros), y la mediana corresponde al valor medio del conjunto de datos, ubicamos en la tabla la décima posición y el valor que encontramos allí (columna “días”) es la mediana, que puede interpretarse de la siguiente manera:

“El 50% de los casos de Ébola tuvo un periodo de incubación por debajo de 9 días y otro 50% por encima de 9 días”

Ahora, revisemos como calcular la mediana cuando el número de registros o conjuntos de datos es par, para ello agregaremos a la tabla un registro mas:



- Organizar los datos de mayor a menor (columna de días de incubación)

Obs.	Días
1	3
2	6
3	7
4	7
5	8
6	8
7	8
8	8
9	9
10	9
11	9
12	9
13	9
14	10
15	10
16	11
17	13
18	16
19	17
20	49

- Encontrar la posición media = $(n+1)/2$
 $(20+1)/2 = 10,5$

Mediana = 9

- Mediana = valor en la posición 10° y 11° = $(9+9)/2 = 9$

Cuando el número de observaciones es par, se usa la misma forma pero ahora se obtiene un promedio entre los dos valores de las observaciones que flanquean la posición en que cayó la mediana, en este caso la media de las observaciones 10 y 11 es 9.

Mediana: propiedades y usos

- Buena medida descriptiva
- Medida de elección para los datos que no están distribuidos simétricamente
- Se enfoca en el centro de los datos, así que no está afectada por algunos valores extremos (atípicos)

Media aritmética

La medida de tendencia central que más oímos es el promedio o media aritmética. Es el valor típico de la serie y se calcula sumando los valores de las observaciones en la serie que se divide por el número de observaciones

Los pasos para calcular la media son:

1. Sumar todos los valores

2. Dividir la suma por la cantidad de observaciones (n)

Obs.	Días
1	3
2	6
3	7
4	7
5	8
6	8
7	8
8	8
9	9
10	9
11	9
12	9
13	9
14	10
15	10
16	11
17	13
18	16
19	17

Suma 177

Suma todos los valores: 177

$n=19$

$$\frac{177}{9} = 9,3$$

Interpretación: El promedio de días de incubación del virus del Ébola es 9,3 días



Media: propiedades y usos

- Probablemente es la mejor medida conocida de tendencia central
- Usa todos los datos
- Afectada por los valores extremos (atípicos)
- Ideal para datos simétricamente distribuidos

¿Qué medidas son “sensibles” a los valores atípicos?

Veamos en el siguiente ejemplo que pasa con la moda, mediana y media cuando tenemos un valor extremo o atípico

Obs	Datos originales Días	Datos con valor atípico Días
1	3	3
2	6	6
3	7	7
4	7	7
5	8	8
6	8	8
7	8	8
8	8	8
9	9	9
10	9	9
11	9	9
12	9	9
13	9	9
14	10	10
15	10	10
16	11	11
17	13	13
18	16	16
19	17	56
Suma	177	216

Medida	Datos originales	19.º obs = 56 días
Moda	9	9
Mediana	9	9
Media	9,3	11,4



Podemos observar que para la columna de datos originales la moda y la mediana son igual a 9, y la media es 9.3 días.

En la columna de datos con valores atípicos, hemos cambiado el número de días para el caso 19° suponiendo que era de alguien que dijo que asistió a un funeral y participó en el lavado del cuerpo 8 semanas antes y negó cualquier contacto conocido desde entonces. Eso puede ser verdad o no, pero esos son los datos con los que cuenta. Veamos cómo cambiar el número 17 por el 56 (8 semanas x 7 días / semanas) afecta las tres mediciones.

¿Cuál es la nueva moda?

- Sigue siendo 9.

¿Cuál es la nueva mediana?

- Sigue siendo 9.

¿Cuál es la nueva media? ¿Cómo lo calcularíamos?

- 216 dividido por 19, la nueva media es 11.4

Entonces, con los nuevos resultados se observa que la media aritmética o promedio vario de 9.3 a 11.4 días, a esto nos referimos cuando decimos que esta medición es sensible a valores extremos o atípicos

Medida de dispersión: Rango

Definición:

Diferencia entre el mayor y el menor valor de una distribución de datos.

Para identificar el rango:

1. Clasificar datos o crear distribución de frecuencia
2. Encontrar valores mínimos y máximos

El rango no es una medición de la ubicación central. Es una medición de la dispersión que es bien conocida y lo que comúnmente se usa con la mediana. El rango es simplemente toda la dispersión de datos, desde lo más pequeño hasta lo más grande.



Obs.	Días
1	3
2	6
3	7
4	7
5	8
6	8
7	8
8	8
9	9
10	9
11	9
12	9
13	9
14	10
15	10
16	11
17	13
18	16
19	17

Valor mínimo = 3

Valor máximo = 17

$$\text{Rango} = 17 - 3 = 14$$

Calculemos el rango. En la serie de períodos de incubación de Ébola, es 17-3 o 14 días.

Resumen

Las medidas de tendencia central son varias: medida única que representa toda una distribución

- Moda: valor más común
- Mediana: valor central, es la opción más segura
Use una mediana con rango
- Media: valor promedio
Usa todos los datos, sensible a valores atípicos
Media preferida para datos simétricos, que no es común

La medida de dispersión: el rango describe la variabilidad de los datos

Bibliografía

1. Centers for Disease Control and Prevention (CDC). Summarizing Data. Lesson 1.05 FETP Frontline. In 2020.

