

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

Marzo 2021



4.- Los datos

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Una variable se puede definir como: *“una propiedad con respecto a la cual los individuos de una población difieren de algún modo verificable”*. Si en todos los individuos de la población la propiedad no difiere, no es de interés estadístico. Puede ser de interés la longitud, el peso, el contenido en ácidos grasos, el color etc. En cambio no son de interés, la presencia de vértebras en mamíferos o el número de patas de un ave.

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Hay muchas formas de clasificar las variables y los datos.

Las variables pueden clasificarse en tres grandes grupos:

- ▶ **atributos:** no pueden medirse ni ordenarse. Se deben expresar cualitativamente, como son los colores.
- ▶ **variables clasificables en rangos:** no pueden medirse pero sí pueden clasificarse en orden dependiendo de su magnitud.
- ▶ **variables medibles:** son aquellas cuyos valores pueden expresarse de forma numéricamente ordenada.

Pueden ser de dos clases:

- ▶ - **variables discretas** o merísticas, que sólo tienen valores numéricos fijos, sin posibilidad de valores intermedios.
- ▶ - **variables continuas**, que pueden tomar un número infinito de valores entre dos puntos determinados.

Escalas de medida

Para cada tipo de variable se usará una escala de medida adecuada. En el caso de variables cualitativas o atributos se deberá emplear la escala de medida nominal, que simplemente proporciona características cualitativas de la variable. En las variables clasificables en rangos, la escala de medida a usar será la ordinal, y las variables medibles (continuas o discretas) se medirán en escala de intervalos o de razones (también llamada de cocientes).

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión




5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Escalas de medida:

1.- Escala nominal		Débiles	
2.- Escala ordinal			
3.- Escala de intervalo		Fuentes	
4.- Escala de razón			

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Cualidades para cada tipo de escala:

1. Medida de datos no ordenados, ej.: Macho, Hembra etc. se pueden asignar claves numéricas 1/2.
2. No se pueden asignar a números reales pero sí a clases que tengan un orden o rango, ej.: picoplancton, nanoplancton, microplancton o análisis granulométrico.
3. Podemos asignar un valor numérico concreto a cada intervalo pero no existe un cero absoluto. Ej.: tiempo.
4. Además de lo anterior, en la escala de razón o cocientes existe un cero absoluto, ej.: altura, peso.

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Las escalas fuertes, utilizan prácticamente las mismas técnicas estadísticas.

4.- Los datos

Escalas de
medidaTipos de datos
y su
preparaciónExactitud,
precisión5.- Estadística
descriptivaMedidas de
centralizaciónMedidas de
dispersiónMedidas de
forma

Escala de medida	Índice de Tendencia Central	Transformación	Test
Nominal	Moda	Biunívocas	Igualdad
Ordinal	Mediana/Moda	Monótonas	Orden creciente
Intervalos	Media/Mediana/Moda	Lineales positivas	Igualdad de diferencias
Razones	Media/Mediana/Moda	Lineales positivas	Igualdad de razones

La moda es un índice de tendencia débil. La mediana es un índice de tendencia central muy bueno (mejor que la moda) y además es más “robusto” que la media.

Tipos de datos y su preparación

Una posible clasificación de los diferentes tipos de datos, podría ser:

Tipos de datos			
Doble estado	Presencia/Ausencia		
	Excluyentes		
Multiestados	Cualitativos	sin secuencia lógica (nominal)	
		con secuencia lógica (ordinal)	
	Cuantitativos	discontinuos	(intervalos/razones)
		continuos	

4.- Los datos

Escala de
medidaTipos de datos
y su
preparaciónExactitud,
precisión5.- Estadística
descriptivaMedidas de
centralizaciónMedidas de
dispersiónMedidas de
forma

Aunque existen más tipos de clasificaciones, en esencia sólo existen cuatro tipos de datos asociados las cuatro escalas de medida, es decir:

- ▶ **Nominales**
- ▶ **Ordinales**
- ▶ **Intervalos**
- ▶ **Razones o Cocientes.**

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

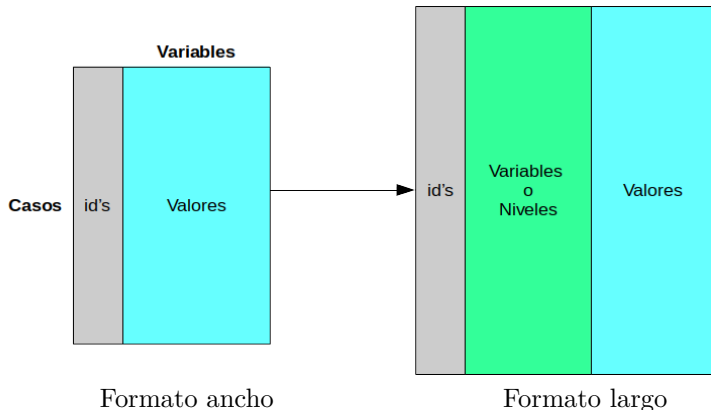
Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Preparación, manipulación y buenas prácticas sobre los datos:

A veces es necesario tener los “dataframes” en formato ancho y largo. El formato ancho, tiene una columna para cada variable; mientras que en el formato largo, cada fila es una combinación única de variable de identificación.



4.- Los datos

Escala de
medidaTipos de datos
y su
preparaciónExactitud,
precisión5.- Estadística
descriptivaMedidas de
centralizaciónMedidas de
dispersiónMedidas de
forma

Decálogo sobre los datos:

1. Tener identificadores para los casos.
2. No truncar, ni redondear nunca.
3. Prever suficientes columnas.
4. Uso del punto decimal. No usar la coma (delimitador en ASCII) y bien indentados.
5. Asignar valores a los datos que faltan. Decidir una codificación que sea clara para los “missing values”(NA o mv).
6. Mejor números que letras, pero no ceros (en algunas técnicas se consideran NA). Conviene codificar (de forma identificable) los datos no numéricos.
7. No bajar la escala de medida aunque se pueda hacer, ni por supuesto subirla. Los datos deben de estar con toda su significación (precisión) y en escala de razón si es posible.
8. Las variables pueden agruparse de modo significativo.
9. Dos o más variables se pueden combinar sin perder información.
10. Si se transforman, no perder la información original. Pensar bien en las implicaciones sobre la alteración y nueva significación de la información transformada.

Exactitud, precisión

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Exactitud: es la proximidad de un valor medido o calculado al valor real.

Precisión: es la proximidad de dos medidas repetidas de la misma cantidad. A menos que el instrumento de medida este sesgado, la precisión conducirá a la exactitud.

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

La mayoría de las variables continuas son aproximadas, el valor exacto de la medida individual es desconocido. La última cifra implicará precisión, los límites entre los cuales creemos que se encuentra la verdadera medida. Por ej. una medida de 13.3, implica que la verdadera longitud se encuentra entre 13.25 y 13.35. Entre estos límites implícitos, no sabemos donde se encuentra la longitud real. Si quisiésemos decir que el valor es 13.30 deberíamos haber añadido esa cifra significativa, estableciéndose los límites implícitos entre 13.305 y 13.295. Si este no es el propósito no se debería añadir la última cifra significativa.

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Cuando se quiere reducir cifras significativas se realiza el redondeo:

Las reglas son muy “sencillas’’: el número a redondear no se cambia si va seguido de uno menor que 5 y sí se cambia (aumentando en uno) si va seguido de un número mayor de 5 ó 5 seguido de otros números distintos de cero. Si el número es 5 solo o seguido de ceros se cambia, aumentando en uno si el anterior es impar, pero no se cambia si es par.

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

5.- Estadística descriptiva

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Como se mencionó anteriormente, la estadística descriptiva realiza estimaciones de los estadísticos de la **muestra**.

No hace inferencias sobre parámetros de la población, simplemente describe y resume la información proporcionada por los datos.

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

Existen principalmente tres tipos de estadísticos¹
descriptivos:

Medidas de centralización

También conocidas como medidas de localización:

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

- **Media aritmética:** actúa como centro geométrico o de gravedad del conjunto de puntos, y se define como:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Existen otras medias tales como la geométrica, o la armónica, que se adaptan mejor a datos transformados (logarítmica o inversamente).

- **Mediana:** es un valor tal que, ordenados en magnitud los datos, el 50 % es menor que ella y el 50 % es mayor. Al ordenar los datos sin agrupar, la mediana es el valor central.
- **Moda,** es el valor más frecuente en los datos.

Medidas de dispersión

Fundamentos
en Estadística

V. Trujillo

GRC-MERVEX
(CO de Vigo.
IEO)

4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística
descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma

- **Desviación típica o estándar:** es un promedio de las desviaciones de las puntuaciones (directas) con respecto a su media (puntuación diferencial). Se elevan al cuadrado las desviaciones para hacerlas positivas y que no se anulen al sumarlas. El cuadrado de la desviación típica se denomina varianza, s^2 :

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2)$$

- **MEDA:** medida de dispersión que se asocia a la mediana (Med). Es la mediana de las desviaciones absolutas. Tiene la ventaja de no verse afectada por datos extremos. Es por lo tanto una medida robusta o resistente:

$$MEDA_{\bar{x}} = \text{mediana}|x_i - \bar{x}| \quad (3)$$

- **Rango o Recorrido:** diferencia entre el valor máximo y mínimo.
- **Cuantil:** valor que separa en un tanto por ciento dado los datos: cuartiles(4) 25 %, quintiles(5) 20 %, deciles(10) 10 % y percentiles(100) 1 %. En general, cuantiles de orden p (Kendall, 1940) etc.

- **Coefficiente de asimetría:** indica como de simétricos son los datos respecto a la media. Es adimensional:

$$CA_{\bar{x}} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{ns^3} \quad (4)$$

Otra medida de asimetría menos utilizada es respecto a la mediana:

$$CA_{mediana} = \frac{\bar{x} - mediana}{s} \quad (5)$$

- **Coefficiente de apuntamiento:** indica la forma de la distribución. Se define como el momento de orden 4 respecto a la media dividido por la desviación típica elevada a cuatro. También es denominado índice de curtosis.

$$CAp_{\bar{x}} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{ns^4} \quad (6)$$



4.- Los datos

Escalas de
medida

Tipos de datos
y su
preparación

Exactitud,
precisión

5.- Estadística descriptiva

Medidas de
centralización

Medidas de
dispersión

Medidas de
forma