

Estadística Descriptiva

Datos, Bases de Datos y Manejo de Bases
de Datos con Excel y SPSS

Rodrigo Asun Inostroza

Tipos de Datos

- **Primarios:**

- Son aquellos que **el equipo investigador produce** al realizar su investigación.
- Tienen la ventaja que son producidos con las delimitaciones teóricas y metodológicas de la investigación, por lo que son muy **pertinentes** a ella.
- Son **caros** de producir: en tiempo, esfuerzo y dinero.

- **Secundarios:**

- Son aquellos que se **encuentran ya producidos** por otras investigaciones o por procesos externos a una investigación.
- Tienen la desventaja que no siempre se ajustan a los límites teóricos y metodológicos deseados por la investigación.
- Usualmente hay que limpiarlos de sesgos o errores
- Son **baratos**: ya están producidos.

¿Qué es una Base de Datos?

- Conjunto de **números ordenados**... usualmente en filas y columnas.
- **Base de datos plana**: sujetos en las filas, variables (propiedades de los sujetos) en las columnas.

A	B	C	D	E	F	G	H	I	J	K	
1	ID Cliente	Zona	País	Tipo de producto	Canal de venta	Prioridad	Fecha pedido	ID Pedido	Fecha envío	Unidades	Precio Unitario
2	C2421	Europa	United Kingdom	Snacks	Offline	Crítica	12/10/2020	242113196	30/11/2020	5530	152,58 €
3	C1908	Europa	Malta	Cárnicos	Online	Alta	26/01/2020	190800607	28/01/2020	994	421,89 €
4	C7652	Australia y Oceanía	Marshall Islands	Cereales	Online	Crítica	09/11/2020	765228068	21/11/2020	6845	205,70 €
5	C2326	África	Iran	Frutas	Offline	Baja	21/08/2021	232631909	02/10/2021	9806	9,33 €
6	C5305	Centroamérica y Caribe	Guatemala	Alimento infantil	Offline	Media	30/09/2022	530560958	12/11/2022	3633	255,28 €
7	C5168	Centroamérica y Caribe	Grenada	Bebida	Offline	Baja	21/01/2022	516876542	21/02/2022	4110	47,45 €
8	C9197	Australia y Oceanía	Fiji	Snacks	Offline	Media	17/02/2022	919752490	27/02/2022	4056	152,58 €
9	C2876	África	Tunisia	Cosméticos	Offline	Media	21/03/2020	287675130	07/05/2020	8319	437,20 €
10	C8394	Centroamérica y Caribe	Grenada	Cuidado personal	Offline	Crítica	13/02/2021	839443290	01/04/2021	8779	81,73 €
11	C8141	Australia y Oceanía	Australia	Material de oficina	Online	Media	15/06/2021	814168298	12/07/2021	9347	651,21 €
12	C3212	Norteamérica	Greenland	Cosméticos	Online	Alta	15/07/2022	321273982	30/08/2022	966	437,20 €
13	C8904	África	Angola	Snacks	Online	Crítica	13/05/2020	890496671	05/06/2020	6609	152,58 €
14	C5218	África	Zambia	Bebida	Offline	Crítica	06/05/2021	521885192	22/05/2021	6281	47,45 €
15	C4358	África	Kenya	Verduras	Offline	Alta	07/11/2020	435800874	25/12/2020	2018	154,06 €
16	C1229	África	Azerbaijan	Frutas	Offline	Crítica	20/10/2021	122917544	17/11/2021	2888	9,33 €
17	C4942	África	Mozambique	Cuidado personal	Offline	Crítica	21/10/2021	494221532	18/11/2021	9989	81,73 €
18	C7310	Australia y Oceanía	Federated States of Micronesia	Cosméticos	Offline	Crítica	19/04/2020	731011664	30/04/2020	1451	437,20 €
19	C5348	Centroamérica y Caribe	Dominican Republic	Snacks	Offline	Crítica	03/11/2022	534899270	04/11/2022	7436	152,58 €
20	C5778	África	Libya	Material de oficina	Online	Alta	10/07/2022	577808177	22/07/2022	5135	651,21 €
21	C2519	Asia	Uzbekistan	Material de oficina	Online	Baja	01/06/2021	251974713	21/06/2021	3772	651,21 €
22	C8199	Asia	Sri Lanka	Cuidado personal	Offline	Media	28/07/2020	819947707	05/09/2020	9602	81,73 €

- Base de datos **tridimensional**: tercera dimensión es agrupamiento de casos:
 - Ejemplo REDATAM de la CEPAL: agrupa personas en comunas.

Uso de Excel para manejo y análisis de datos

- Digital (variables de respuesta simple y múltiple)
- Exportar / importar
- Crear nuevas variables
- Recodificar
- Tendencia central / dispersión / asociación.
- Gráficos básicos: frecuencias / dispersión.

Tipo de variables

1.- Depende del **tipo de datos que tengamos**, no de como lo llame un software. **Todos se pueden digitar con números.**

2.- Tiene consecuencias en los **análisis permitidos y efectivos** para realizar con cada nivel de medición.

- **Nivel Nominal:**

- Sólo importa la propiedad de distinción. No hay orden.

- **Variables ordinales:**

- Importa la distinción y el orden.
- Ej: posiciones en una carrera, ítems tipo likert.
- En general usa valores enteros a distancias regulares

- **Variables Intervalares / de Razón:**

- No es productivo distinguir ambos tipos, porque no tiene consecuencias en estadística social.
- Usa puntajes / puede usar decimales.
- Más de 7 (y sobre todo más de 10) ordenes se considera intervalar en términos prácticos

Se les denomina
“Cualitativas”

Se les denomina
“Cuantitativas”
o “Numéricas”

Recomendaciones para la digitación de variables según su tipo

- **Variables Dicotómicas:**
 - Usar siempre 0 / 1 (variable Dummy).
- **Variables ordinales:**
 - Usar siempre valores de 1 al rango máximo.
- **Variables Intervalares / de Razón:**
 - Escribir el número.
- **Preguntas de respuesta múltiple:**
 - Transformar en serie de Dummies (0/1) (presencia / ausencia)

Pregunta de respuesta múltiple y su digitación

P34.- ¿Qué tipo de películas te gustan? (marca todas las que quieras)

1. De terror.

☒ Animadas.

☒ Del oeste

4. De Fantasía.

☒ De ciencia ficción.

6. Cine arte.

7. Ninguna.

	P34.1	P34.2	P34.3	P34.4	P34.5	P34.6	P34.7
Sujeto 125	0	1	1	0	1	0	0

Uso de SPSS para manejo y análisis de datos

- Importar / exportar bases de datos.
- Etiquetar variables y valores de variables.
- Crear nuevas variables
- Recodificar
- Asignar valores perdidos
- Crear variables dummy (0/1)
- Limpiar bases de datos:
 - Frecuencias cada variable.
 - Quitar valores fuera de rango.
 - Quitar valores extremos (si corresponde).
 - Eliminar variables irrelevantes y anonimizar.
 - Crear base de datos recortada para estudios particulares.

FIN