

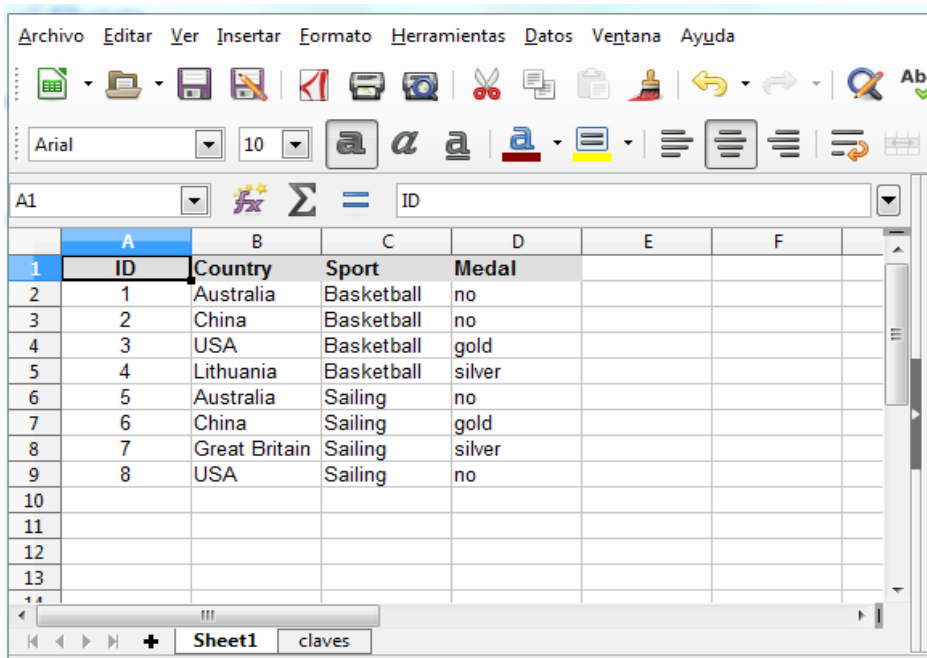
MINERIA DE DATOS

aurora-mareviv

1. Limpieza de datos

1.a ¿Cómo hago con mis datos en Excel?

- Copia de seguridad!
- Primera columna con ID anónimo (alfanumérico).
- Confidencialidad. Etica.
- Ordena para ver datos raros.
- Filtra para ver categorías.
- Documenta (nueva pestaña con claves).



The screenshot shows the Microsoft Excel interface. The menu bar includes Archivo, Editar, Ver, Insertar, Formato, Herramientas, Datos, Ventana, and Ayuda. The toolbar contains various icons for file operations, editing, and formatting. The formula bar shows 'ID' in cell A1. The worksheet grid displays the following data:

	A	B	C	D	E	F
1	ID	Country	Sport	Medal		
2	1	Australia	Basketball	no		
3	2	China	Basketball	no		
4	3	USA	Basketball	gold		
5	4	Lithuania	Basketball	silver		
6	5	Australia	Sailing	no		
7	6	China	Sailing	gold		
8	7	Great Britain	Sailing	silver		
9	8	USA	Sailing	no		
10						
11						
12						
13						

The status bar at the bottom shows 'Sheet1' and 'claves'.

1.b Explora tus datos

- Dos tipos de datos:
 - Discretos.
 - Numéricos.
- Se procesan por separado!
- Se pueden analizar por separado, o conjuntamente.
 - ¿Cuáles de mis variables son continuas?.
 - ¿Cuáles no? Examina todas ellas.

2. Resume tus datos

Datos discretos

Factor	Valor	Frecuencia	Porcentaje %
Factor A	0	312	78.4
	1	86	21.6

Datos discretos: Tablas cruzadas

	Factor B=0	Factor B=1
Factor A=0	206	71
Factor A=1	106	15

Datos Numéricos

Variable	Media	Mediana	Desv. Standard	IQR
Edad (años)	68	71	11.4	[62 - 76]
Temperatura (°C)	36.8	36	1.5	[35.5 - 38]

Resumen tests

Test	Variable 1	Variable 2	Paramétrica?	Gráfico
Chi-cuadrado	Factor	Factor		Barras
Mc Nemar	Factor	Factor (before & after)		Barras
T-student	Númerica	Factor (2 categ.)	Sí	Boxplot/histograma
Wilcoxon	Númerica	Factor (2 categ.)		Boxplot/histograma
ANOVA	Númerica	Factor (>2 categ.)	Sí	Boxplot/histog/lineales
Kruskal-Wallis	Númerica	Factor (>2 categ.)		Boxplot/histog/lineales
Regresión lineal	Númerica	Númerica	Sí	Gráf. dispersión
Spearman/Kendall	Númerica	Númerica		Gráf. dispersión

3. Dibuja

- Dibujar histogramas:
 - <http://www.shodor.org/interactivate/activities/Histogram/>
- Dibujar gráficos de barras:
 - <http://www.socscistatistics.com/descriptive/bar/Default.aspx>
 - <https://www.easycalculation.com/graphs/bar-graph.php>

4. Analiza

- SERGAS - EPIDAT:
 - <http://www.sergas.es/Saude-publica/EPIDAT>
- Calculadora Chi-cuadrado y Fisher:
 - <http://graphpad.com/quickcalcs/contingency1.cfm>
 - <http://vassarstats.net/newcs.html>
- Calculadora McNemar:
 - <http://graphpad.com/quickcalcs/mcNemar1/>
- Calculadora t Student:
 - <http://www.evanmiller.org/ab-testing/t-test.html>
 - <http://www.physics.csbsju.edu/stats/t-test.html>
- Calculadora Wilcoxon/Mann-Whitney:
 - <http://www.socscistatistics.com/tests/signedranks/>
 - <https://ccb-compute2.cs.uni-saarland.de/wtest/?id=www/www-ccb/html/software/wtest/>
- Calculadora ANOVA:
 - <http://vassarstats.net/anova1u.html>
- Calculadora Kruskal-Wallis:
 - <http://vassarstats.net/kw3.html>
- Teoría estadística:
 - <http://www.fisterra.com/formacion/metodologia-investigacion/>
 - <http://graphpad.com/guides/prism/7/statistics/>