

Raccolta dei dati

Lo strumento più utilizzato per la raccolta dati sono i **fogli di calcolo** (es. **MS Excel**), che possono essere direttamente letti da **Jamovi**.

Accorgimenti importanti:

- Le righe corrispondono all'unità di osservazione
 - Tipicamente le persone, ma non sempre (es. misurazioni multiple per persona)
- Le colonne corrispondono alle variabili
 - Stabilire fin da subito le variabili importanti per lo studio
- Le colonne sono **omogenee**, ossia contengono variabili dello stesso tipo (=scala di misura)
- Riportare la legenda per l'interpretazione delle variabili

Raccolta dei dati

	A	B	C	D	E	F	G
1	Paziente	Sesso	Altezza	BMI	rs642742		
2	001	F	1.64	19,3	AA		
3	002	f	1.55	25,5	AB		affetti
4	3	M	1.74	28,7	ab		sani
5	004	M	1.92	28,9	BB		
6	005	M	1.83	26,8	AA		
7							
8	N=5						

Errori frequenti:

- Utilizzo di separatori diversi per i numeri decimali
- Valori categoriali scritti in modo diverso
- Forme non tabellari
- “Uso improprio” dei colori
- Codifica valori mancanti: le celle vanno lasciate vuote

Tipi di file

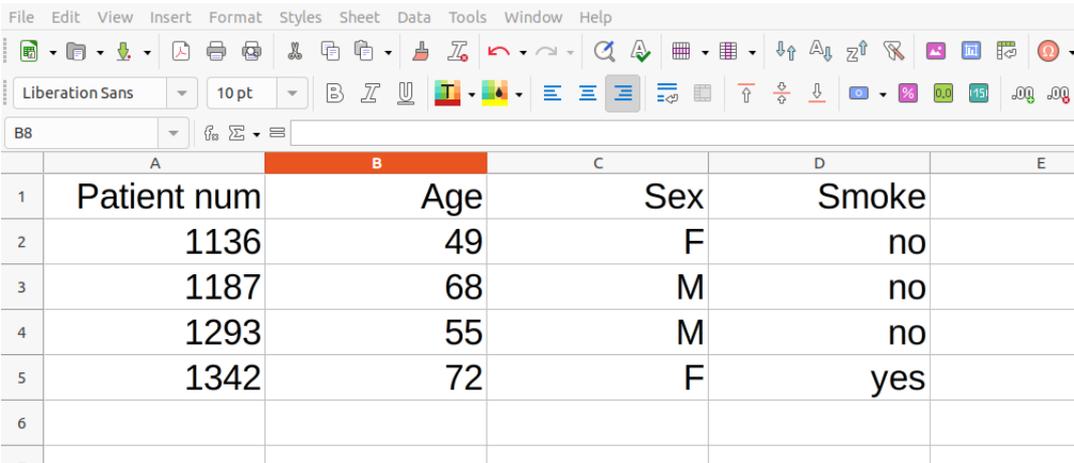
Tratteremo dati in forma tabellare. Caricheremo i dati in Jamovi principalmente dai seguenti tipi di file:

- .xlsx → fogli di calcolo (es. MS Excel)
- .csv → “*Comma Separated Values*”, file di testo che contengono dati tabellari dove le colonne sono definite dal separatore “,” o “;”

Tipi di file

Foglio di calcolo

CSV



The screenshot shows a spreadsheet application interface. The menu bar includes File, Edit, View, Insert, Format, Styles, Sheet, Data, Tools, Window, and Help. The toolbar contains various icons for file operations, editing, and formatting. The active cell is B8. The spreadsheet contains the following data:

	A	B	C	D	E
1	Patient num	Age	Sex	Smoke	
2	1136	49	F	no	
3	1187	68	M	no	
4	1293	55	M	no	
5	1342	72	F	yes	
6					
7					

```
Patient num, Age, Sex, Smoke
1136, 49, F, no
1187, 68, M, no
1293, 55, M, no
1342, 72, F, yes
```



Dataset: *diabetes*

Contiene i dati di un campione di pazienti diabetici, tra cui età, sesso, body mass index, pressione sanguigna, parametri ematologici.

Fonte: Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani (2004) "*Least Angle Regression*".

Il file **.xlsx** include:

- Tabella dei dati
- Legenda



Dataset: *diabetes*

Leggere il file in Jamovi

- Visualizzare la tabella di dati
- Esaminare la lista delle variabili
- Ottenere media e mediana per BP
- Ottenere media e mediana per le prime 3 variabili
- Ottenere informazioni riassuntive rapide su tutte le variabili del dataset
- Ricodificare SEX (assegnando nomi ai livelli, es . 1=M, 2=F)



Dataset: *diabetes*

Creare delle nuove categorie in base al **BMI**:

≤ 18.5 sottopeso

(18.5, 25] normopeso

(25, 30] sovrappeso

> 30 obeso

Ottenere la distribuzione di frequenze



Dataset: *diabetes*

Il rapporto **LDL/HDL** può essere un indicatore del rischio per malattie cardiovascolari:

LDL/HDL < 3 : non a rischio

LDL/HDL \geq 3 : a rischio

Assegnare gli individui a una delle due categorie e calcolare la distribuzione di frequenza

Dataset: *diabetes*

- Rappresentare le classi di BMI con un grafico a barre
- Creare un istogramma per la variabile l'età
- Creare grafico a dispersione per BMI e BP