

## **Lezione 4**

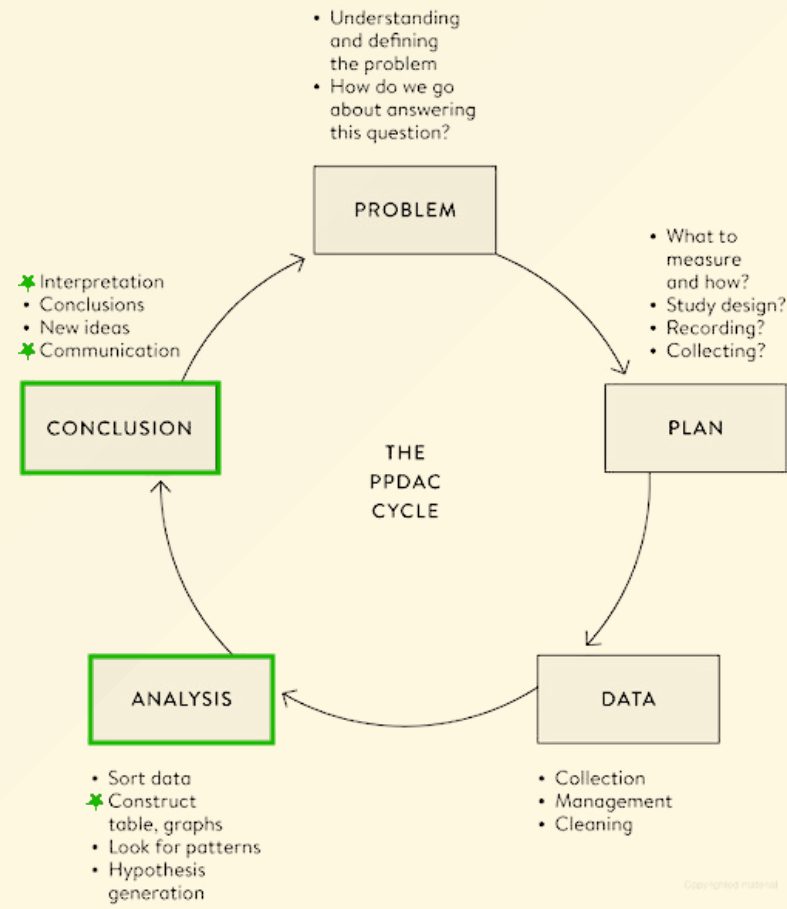
# **La statistica descrittiva**

### **(Parte I: Le variabili categoriche)**

# Obiettivi di apprendimento

- Saper costruire e interpretare tabelle di frequenza e contingenza
- Saper visualizzare dati categorici
- Saper interpretare tabelle e figure in articoli scientifici

# Le fasi della ricerca



# I tipi di dato

? Di che tipo sono i dati  
in questa tabella?

Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

**Table 1** Patient characteristics.

All cohorts	
<b>N (pre-treatment)</b>	
<b>N (follow-up)</b>	
<b>Sex</b>	<i>Binario</i>
Male	
Female	
<b>Age (years)</b>	<i>Numerico</i>
<b>BMI (kg/m<sup>2</sup>)</b>	<i>Continuo</i>
<b>BRAF mutant</b>	<i>Binario</i>
<b>LDH (<math>\leq</math>ULN)</b>	<i>Binario</i>
<b>Metastatic stage</b>	<i>Ordinale</i>
Stage III unresectable	
M1a	
M1b	
M1c	
M1d	
<b>ECOG performance status</b>	<i>Ordinale</i>
0	
1	
2	
3	
<b>ICI therapy</b>	<i>Multi-categoria</i>
Ipilimumab	
Pembrolizumab	
Nivolumab	
Ipilimumab + Nivolumab	

# Perché ci interessa?

**Table 1** Patient characteristics. Categorical variables are presented as number (percentage). Continuous variables are presented as mean  $\pm$  standard deviation.

Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

**Table 1** Patient characteristics.

	All cohorts
<b>N (pre-treatment)</b>	88
<b>N (follow-up)</b>	66
<b>Sex</b>	
Male	57 (64.8%)
Female	31 (35.2%)
<b>Age (years)</b>	60.5 $\pm$ 15.0
<b>BMI (kg/m<sup>2</sup>)</b>	28.0 $\pm$ 5.4
<b>BRAF mutant</b>	40 (45.5%)
<b>LDH (<math>\leq</math>ULN)</b>	58 (65.9%)
<b>Metastatic stage</b>	
Stage III unresectable	2 (2.3%)
M1a	14 (15.9%)
M1b	17 (19.3%)
M1c	32 (36.4%)
M1d	23 (26.1%)
<b>ECOG performance status</b>	
0	47 (53.4%)
1	31 (35.2%)
2	8 (9.1%)
3	2 (2.3%)
<b>ICI therapy</b>	
Ipilimumab	1 (1.1%)
Pembrolizumab	20 (22.7%)
Nivolumab	30 (34.1%)
Ipilimumab + Nivolumab	37 (42.0%)

# Variabili categoriche

## Tabella di frequenza

- frequenza assoluta (numero)
- frequenza relativa (percentuale)

Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

Table 1 Patient characteristics.	
	All cohorts
<b>N (pre-treatment)</b>	88
<b>Sex</b>	
Male	57 (64.8%)
Female	31 (35.2%)
<b>Metastatic stage</b>	
Stage III unresectable	2 (2.3%)
M1a	14 (15.9%)
M1b	17 (19.3%)
M1c	32 (36.4%)
M1d	23 (26.1%)
<b>ECOG performance status</b>	
0	47 (53.4%)
1	31 (35.2%)
2	8 (9.1%)
3	2 (2.3%)
<b>ICI therapy</b>	
Ipilimumab	1 (1.1%)
Pembrolizumab	20 (22.7%)
Nivolumab	30 (34.1%)
Ipilimumab + Nivolumab	37 (42.0%)

# Esercizio #1

? Quale sesso è predominante?

**Table 1** Patient characteristics.

	All cohorts
<b>N (pre-treatment)</b>	88
<b>Sex</b>	
Male	57 (64.8%)
Female	31 (35.2%)

Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

# Esercizio #1 -- Soluzione

? Quale sesso è predominante?

**Table 1** Patient characteristics.

	All cohorts
<b>N (pre-treatment)</b>	88
<b>Sex</b>	
Male	57 (64.8%)
Female	31 (35.2%)

Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3



# Tabelle di contingenza

**Study Objectives:** Persistent insomnia, although very common in general practice, often proves problematic to manage. This study investigates the clinical effectiveness and the feasibility of applying cognitive behavior therapy (CBT) methods for insomnia in primary care.

**Design:** Pragmatic randomized controlled trial of CBT versus treatment as usual.

**Setting:** General medical practice.

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

## Esercizio #2

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi			
Femmine			
Totale			

## Esercizio #2 -- Soluzione

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi			
Femmine	72	65	
Totale	107	94	201

## Esercizio #2 -- Soluzione

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi	35	29	
Femmine	72	65	
Totale	107	94	201

## Esercizio #2 -- Soluzione

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi	35	29	64
Femmine	72	65	137
Totale	107	94	201

Abbiamo completato la nostra Tabella di Contingenza (con frequenze assolute)

## Esercizio #2 -- Soluzione

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi	35 (35/107)	29 (29/94)	64 (64/201)
Femmine	72 (72/107)	65 (65/94)	137 (64/201)
Totale	107	94	201

Opzione 1: andiamo a dividere "per colonne", ovvero a calcolare la percentuale di maschi e di femmine in ciascun braccio dell'esperimento

## Esercizio #2 -- Soluzione

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi	35 (32.7%)	29 (30.9%)	64 (31.8%)
Femmine	72 (67.3%)	65 (69.1%)	137 (68.2%)
Totale	107	94	201

Opzione 1: andiamo a dividere "per colonne", ovvero a calcolare la percentuale di maschi e di femmine in ciascun braccio dell'esperimento

## Esercizio #2 -- Soluzione

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi	35 (35/64)	29 (29/64)	64
Femmine	72 (72/137)	65 (65/137)	137
Totale	107 (107/201)	94 (94/201)	201

Opzione 2: andiamo a dividere "per righe", ovvero a calcolare la percentuale di maschi (o femmine) assegnate a ciascun braccio dell'esperimento



## Esercizio #2 -- Soluzione

- ? Completa la tabella con le corrette frequenze assolute e relative usando le informazioni contenute nell'abstract

**Participants:** Two hundred one adults (mean age, 54 years) randomly assigned to receive CBT (n = 107; 72 women) or treatment as usual (n = 94; 65 women).

Sesso/N	CBT	Standard	Totale
Maschi	35 (54.7%)	29 (45.3%)	64
Femmine	72 (52.6%)	65 (47.4%)	137
Totale	107 (53.2%)	94 (46.8%)	201

Opzione 2: andiamo a dividere "per righe", ovvero a calcolare la percentuale di maschi (o femmine) assegnate a ciascun braccio dell'esperimento

## Esercizio #3

- ? La proporzione di uomini e donne nel campione è simile?  
a) Vero    b) Falso    c) Servono più elementi per decidere
- ? Confrontando i due bracci, la proporzione di uomini e donne è simile?  
a) Vero    b) Falso    c) Servono più elementi per decidere

Sesso/N	CBT	Standard	Totale
Maschi	35 (32.7%)	29 (30.9%)	64 (31.8%)
Femmine	72 (67.3%)	65 (69.1%)	137 (68.2%)
Totale	107	94	201

## Esercizio #3 -- Soluzione

- ? La proporzione di uomini e donne nel campione è simile?  
a) Vero    b) Falso ☒    c) Servono più elementi per decidere
- ? Confrontando i due bracci, la proporzione di uomini e donne è simile?  
a) Vero    b) Falso    c) Servono più elementi per decidere

Sesso/N	CBT	Standard	Totale
Maschi	35 (32.7%)	29 (30.9%)	64 (31.8%)
Femmine	72 (67.3%)	65 (69.1%)	137 (68.2%1)
Totale	107	94	201

## Esercizio #3 -- Soluzione

- ? La proporzione di uomini e donne nel campione è simile?  
a) Vero    b) Falso ☒    c) Servono più elementi per decidere
- ? Confrontando i due bracci, la proporzione di uomini e donne è simile?  
a) Vero ☒    b) Falso    c) Servono più elementi per decidere

Sesso/N	CBT	Standard	Totale
Maschi	35 (32.7%)	29 (30.9%)	64 (31.8%)
Femmine	72 (67.3%)	65 (69.1%)	137 (68.2%1)
Totale	107	94	201

## Esercizio #4

- ? Di 100 studenti di statistica medica, 40 sono sonnolenti (20 dei quali sono interessati alla materia), e 60 sono svegli (40 dei quali interessati alla materia)

Completa la tabella con le corrette frequenze assolute

Interessato/Sonnolento	Sì	No	Totale
Sì			
No			
Totale			

## Esercizio #4 -- Soluzione

- ? Di 100 studenti di statistica medica, 40 sono sonnolenti (20 dei quali sono interessati alla materia), e 60 sono svegli (40 dei quali interessati alla materia)

Completa la tabella con le corrette frequenze assolute

Interessato/Sonnolento	Sì	No	Totale
Sì	20	40	
No			
Totale	40	60	100

## Esercizio #4 -- Soluzione

- ? Di 100 studenti di statistica medica, 40 sono sonnolenti (20 dei quali sono interessati alla materia), e 60 sono svegli (40 dei quali interessati alla materia)

Completa la tabella con le corrette frequenze assolute

Interessato/Sonnolento	Sì	No	Totale
Sì	20	40	
No	20	20	
Totale	40	60	100

## Esercizio #4 -- Soluzione

- ? Di 100 studenti di statistica medica, 40 sono sonnolenti (20 dei quali sono interessati alla materia), e 60 sono svegli (40 dei quali interessati alla materia)

Completa la tabella con le corrette frequenze assolute

Interessato/Sonnolento	Sì	No	Totale
Sì	20	40	60
No	20	20	40
Totale	40	60	100



## Esercizio #5

- ? Qual è la percentuale di sonnolenti/svegli nei due gruppi di chi è interessato o meno nella materia?

Interessato/Sonnolento	Sì	No	Totale
Sì	20	40	60
No	20	20	40
Totale	40	60	100

## Esercizio #5 -- Soluzione

- ? Qual è la percentuale di sonnolenti/svegli nei due gruppi di chi è interessato o meno nella materia?

Interessato/Sonnolento	Sì	No	Totale
Sì	20 (20/60)	40 (40/60)	60
No	20 (20/40)	20 (20/40)	40
Totale	40	60	100

## Esercizio #5 -- Soluzione

? Qual è la percentuale di sonnolenti/svegli nei due gruppi di chi è interessato o meno nella materia?

Interessato/Sonnolento	Sì	No	Totale
Sì	20 (33.3%)	40 (66.7%)	60
No	20 (50%)	20 (50%)	40
Totale	40	60	100

## Esercizio #6

- ? Qual è la percentuale di interessati o meno alla materia nei due gruppi di sonnolenti/svegli?

Interessato/Sonnolento	Sì	No	Totale
Sì	20	40	60
No	20	20	40
Totale	40	60	100

## Esercizio #6 -- Soluzione

- ? Qual è la percentuale di interessati o meno alla materia nei due gruppi di sonnolenti/svegli?

Interessato/Sonnolento	Sì	No	Totale
Sì	20 (20/40)	40 (40/60)	60
No	20 (20/40)	20 (20/60)	40
Totale	40	60	100

## Esercizio #6 -- Soluzione

- ? Qual è la percentuale di interessati o meno alla materia nei due gruppi di sonnolenti/svegli?

Interessato/Sonnolento	Sì	No	Totale
Sì	20 (50%)	40 (66.7%)	60
No	20 (50%)	20 (33.3%)	40
Totale	40	60	100

# Discretizzazione

- Da una serie di valori continui a intervalli (o classi)
- La scelta del numero di classi e degli estremi è arbitraria
- Le classi devono essere mutualmente esclusive
- Riportate anche le classi nulle (frequenza zero)


$$\text{IMC (BMI)} = \frac{\text{peso}_{\text{kg}}}{\text{altezza}_m^2}$$

IMC (BMI)	Classificazione
< 18,5	Sottopeso
18,5 – 24,9	Normopeso
25,0 – 29,9	Sovrappeso
30,0 – 34,9	Obesità 1° grado
35,0 – 39,9	Obesità 2° grado
> 39,9	Obesità 3° grado

# Discretizzazione

**Table 1** Patient characteristics. Categorical variables are presented as number (percentage). Continuous variables are presented as mean  $\pm$  standard deviation.

Table 1 Patient characteristics.	
	All cohorts
N (pre-treatment)	88
LDH ( $\leq$ ULN)	58 (65.9%)

Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

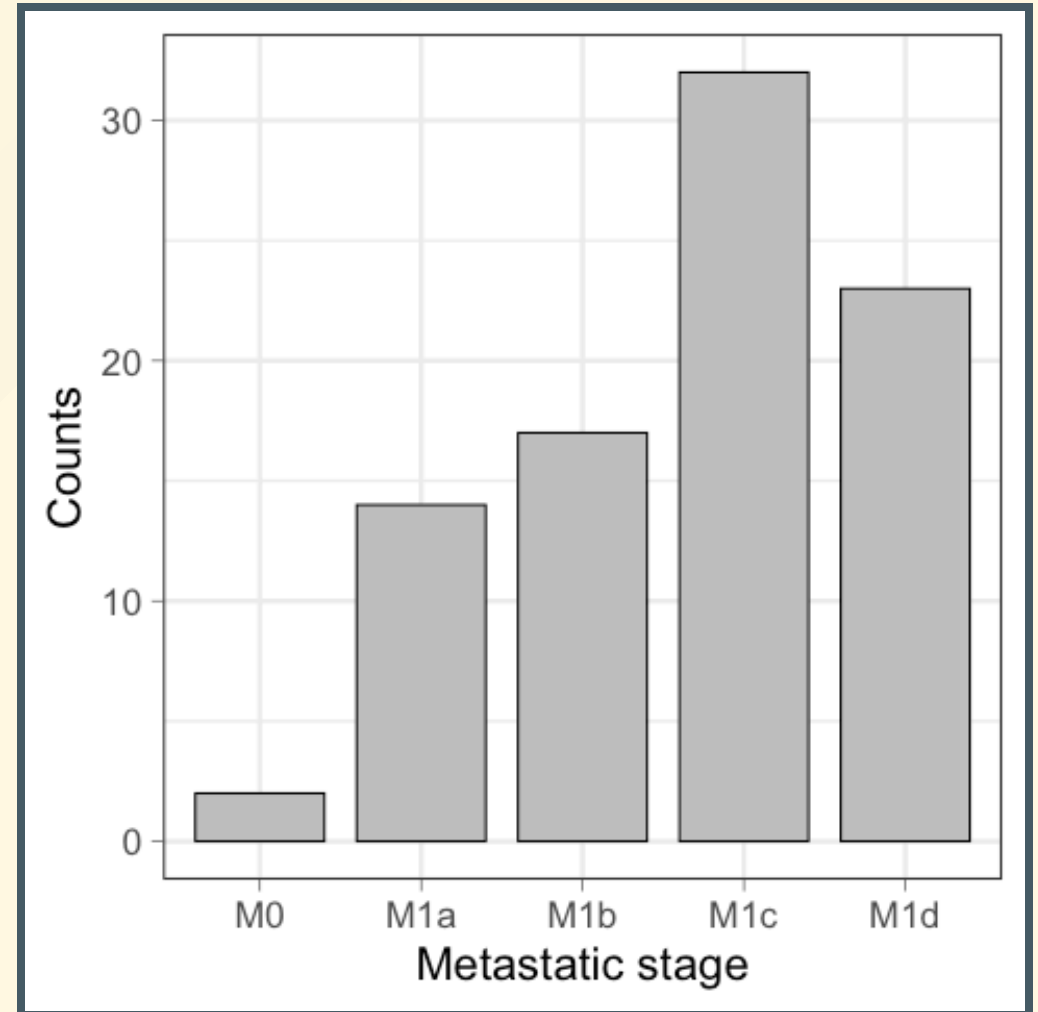


# **La visualizzazione dei dati categorici**

# Bar chart

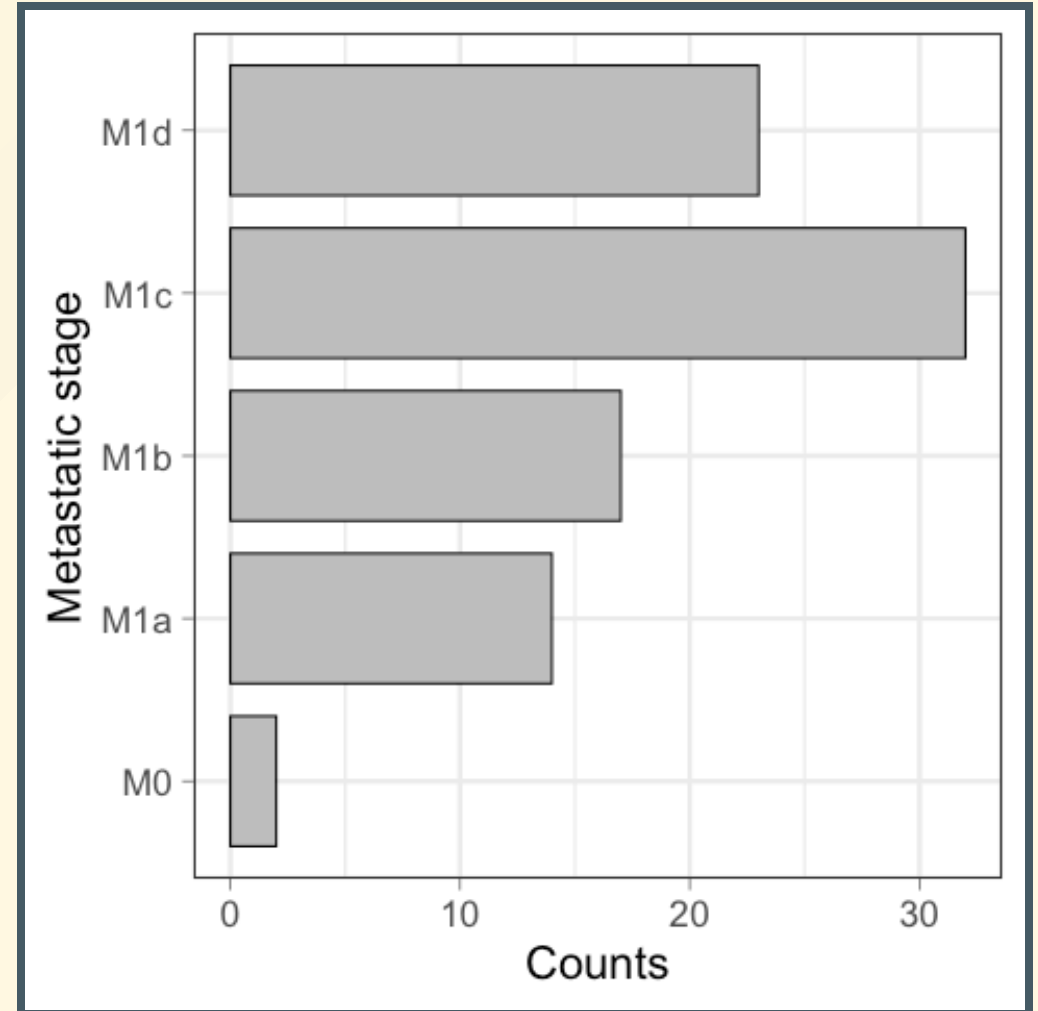
- frequenza assoluta
- frequenza relativa

Visconti A., *et al.*, Total serum *N*-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3



# Horizontal bar chart

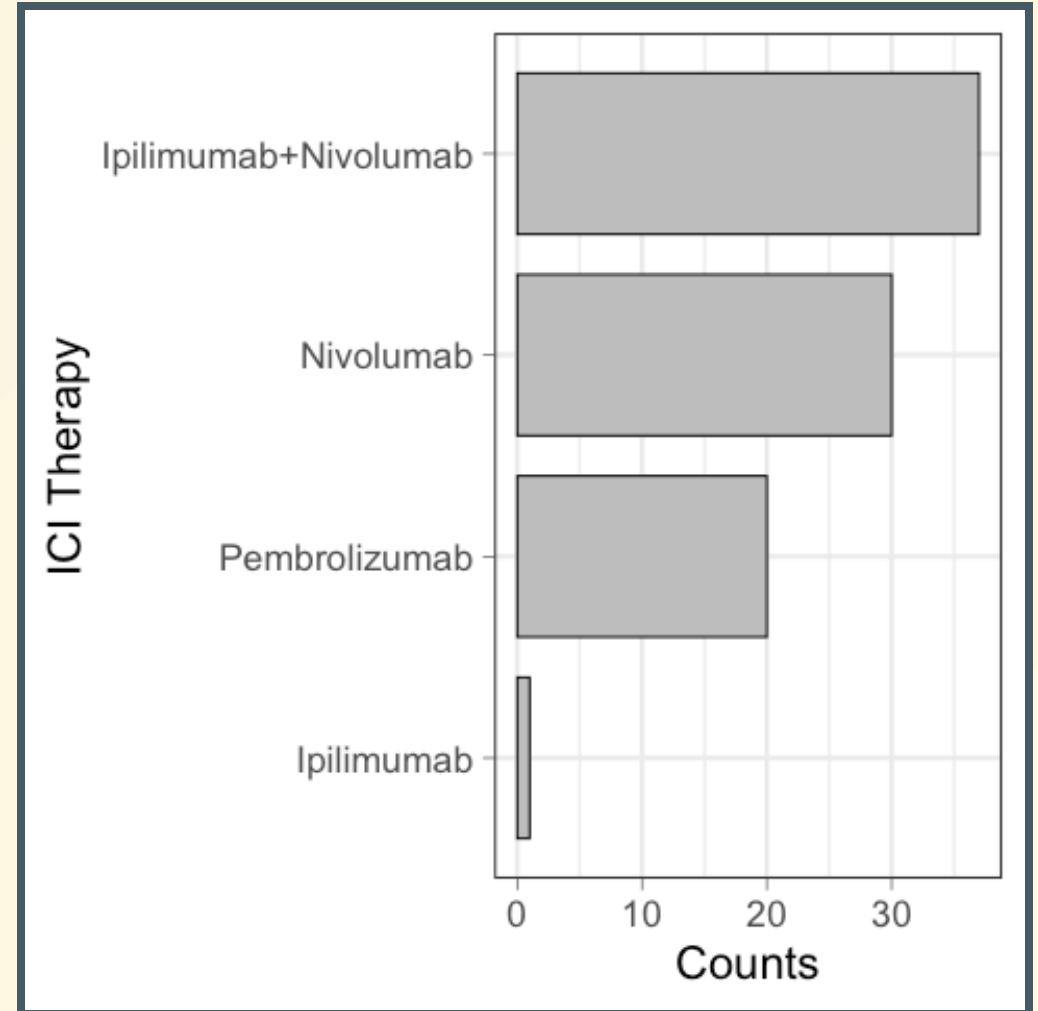
- frequenza assoluta
- frequenza relativa



Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

# Horizontal bar chart

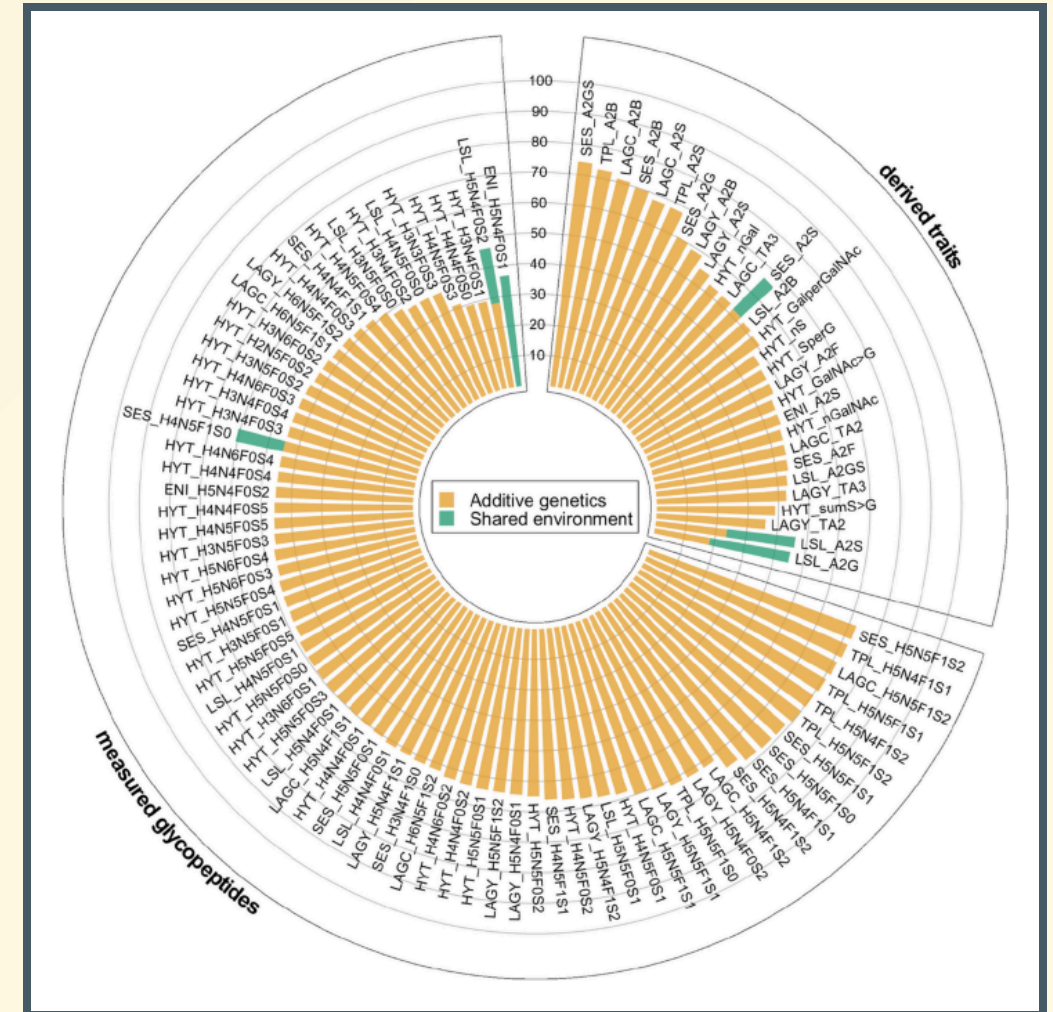
- frequenza assoluta
- frequenza relativa



Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

# Circular bar chart

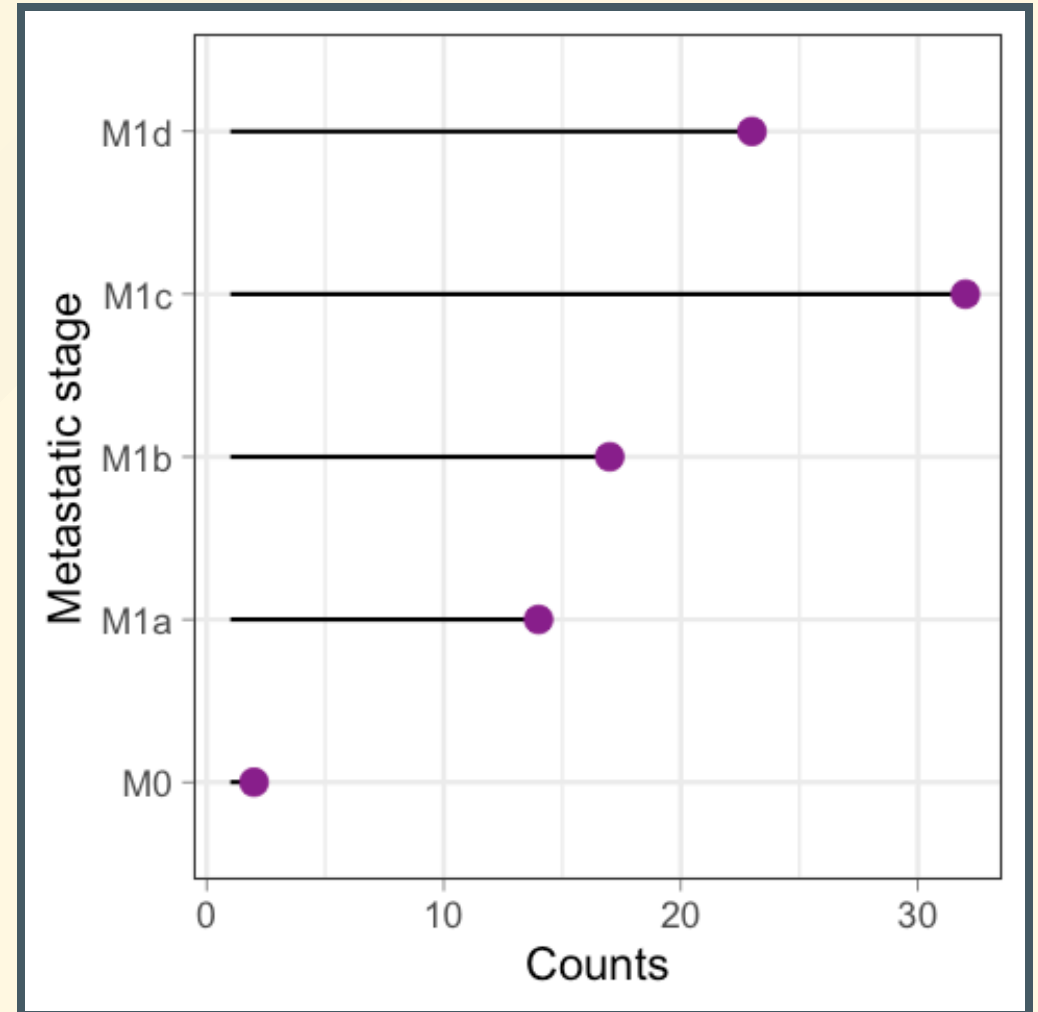
- frequenza assoluta
- frequenza relativa



Visconti, A., et al.. *The genetics and epidemiology of N- and O-immunoglobulin A glycomics.*, 2024, doi:10.1186/s13073-024-01369-6

# Lollipop chart

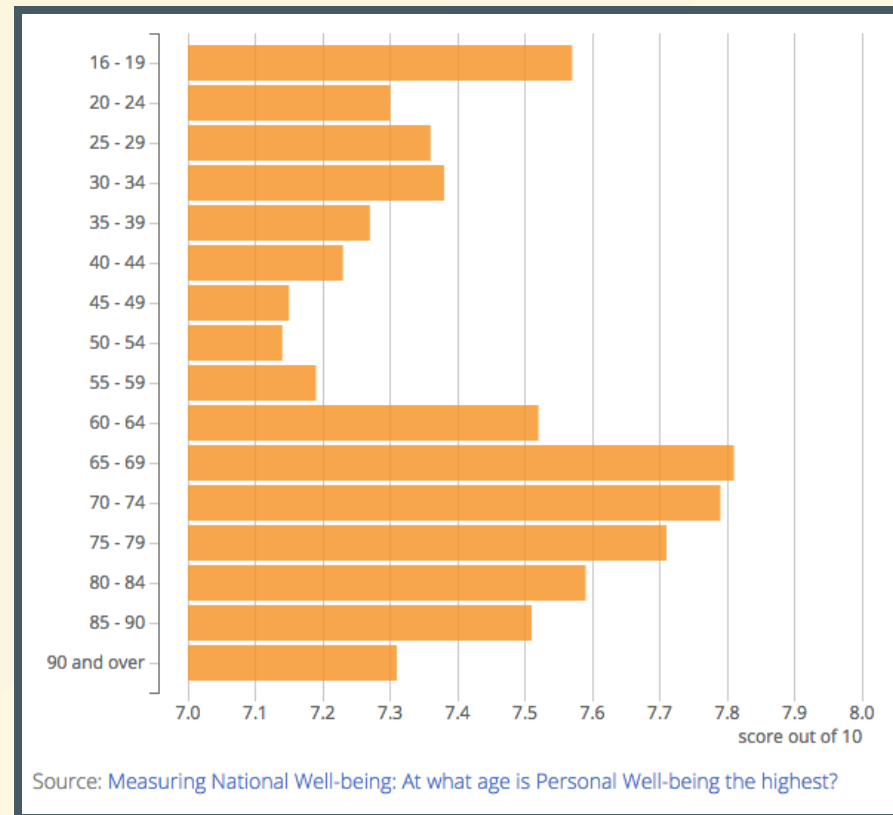
- frequenza assoluta
- frequenza relativa



Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3

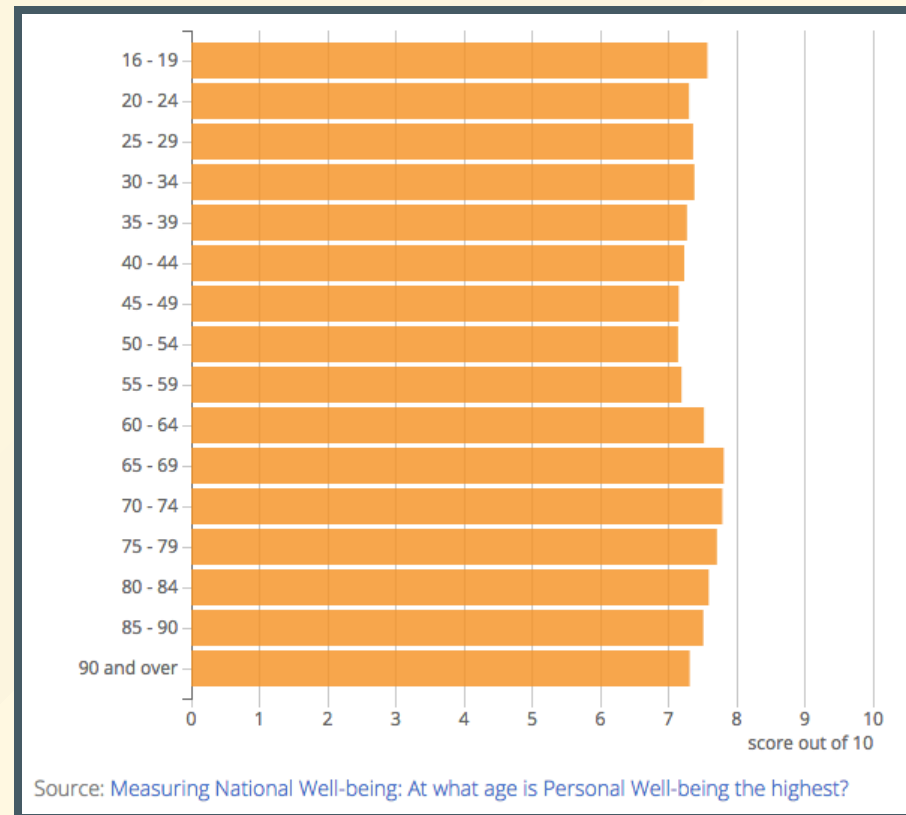
# Cosa ci dice questo grafico?

- Da 1 a 10, quanto eri felice ieri?



# Cosa ci dice questo grafico?

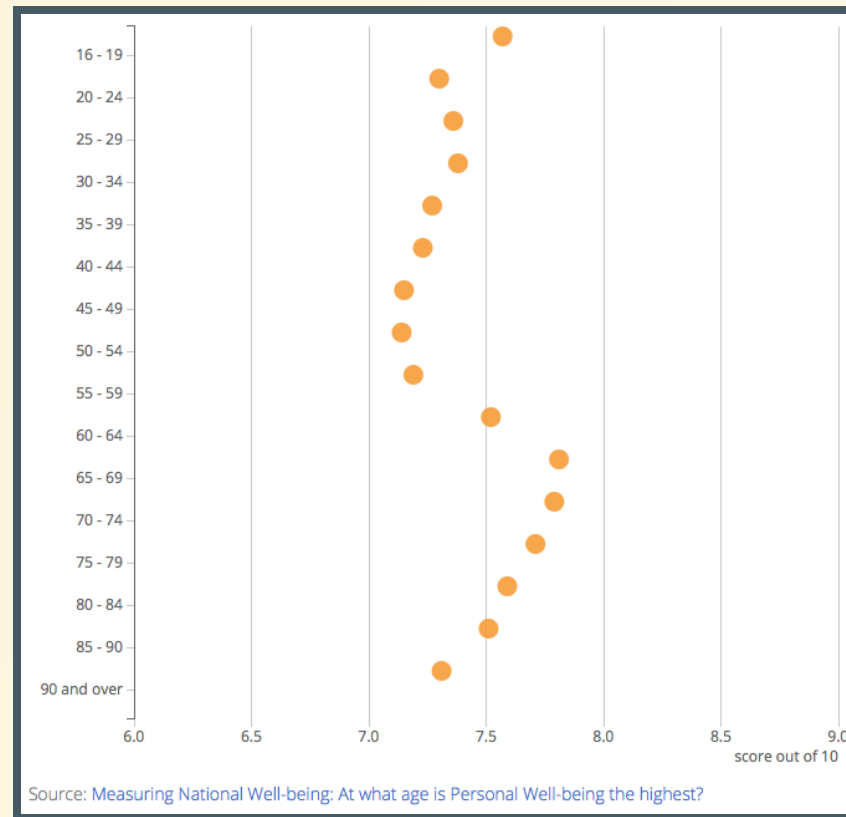
- Da 1 a 10, quanto eri felice ieri?





# Cosa ci dice questo grafico?

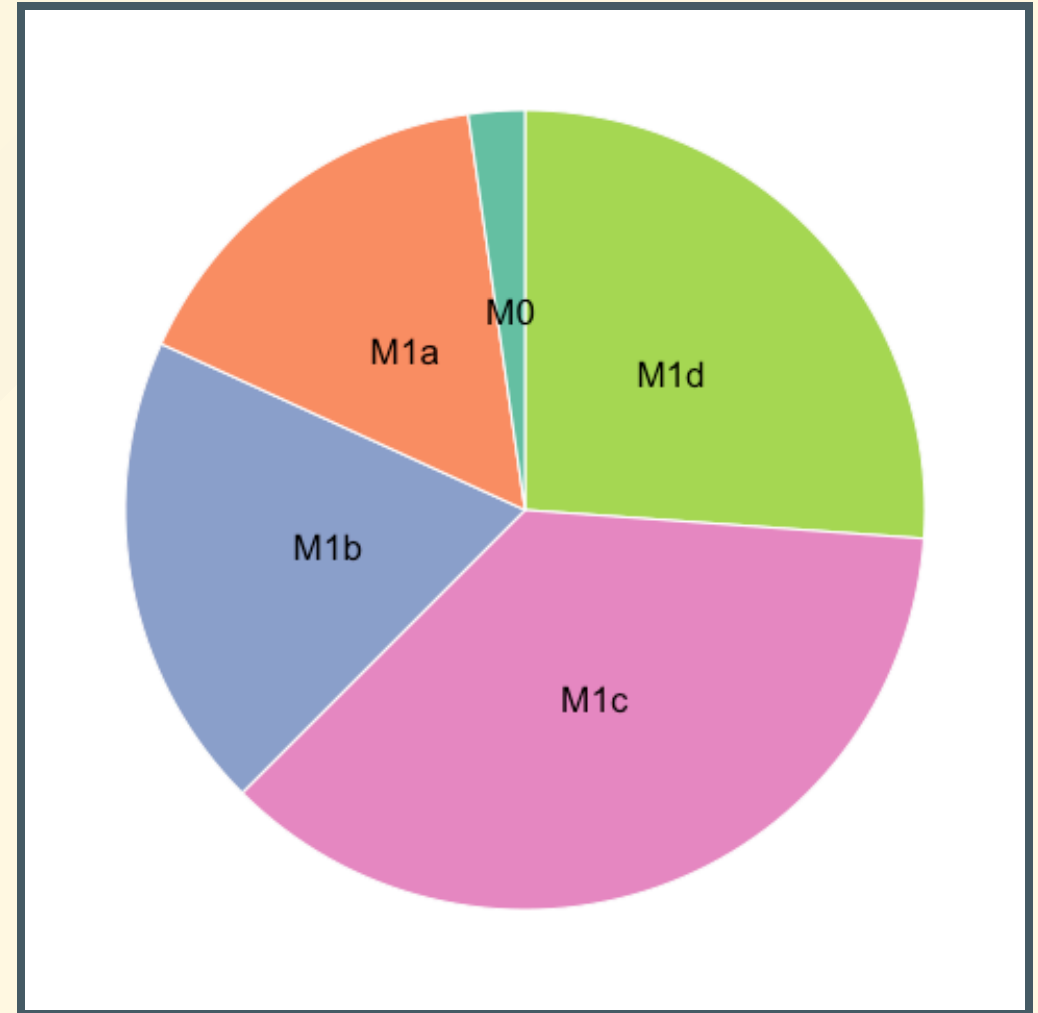
- Da 1 a 10, quanto eri felice ieri?



# Pie chart

- frequenza relativa

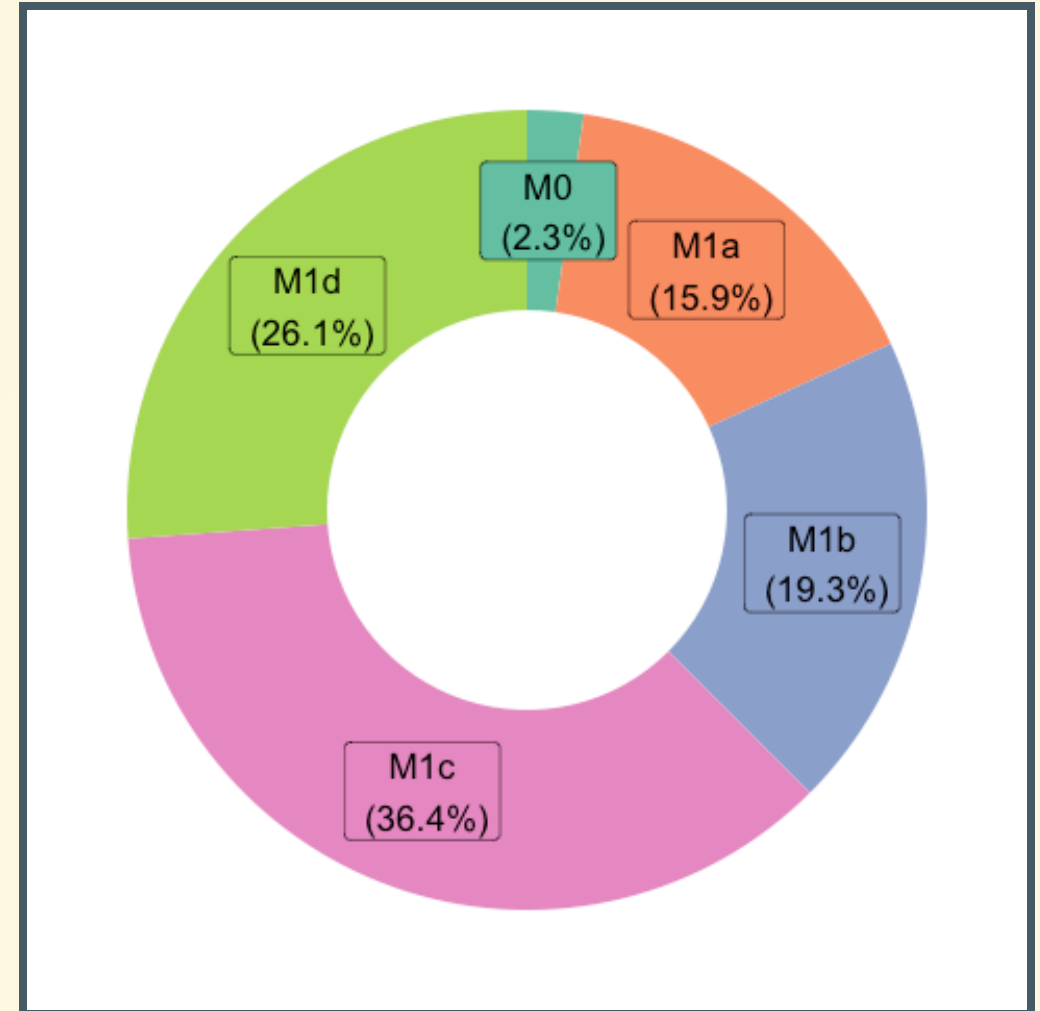
Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3



# Donut chart

- frequenza relativa

Visconti A., et al., *Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma*, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3



# Waffle chart

- frequenza relativa

*Visconti A., et al., Total serum N-glycans associate with response to immune checkpoint inhibition therapy and survival in patients with advanced melanoma, BMC Cancer, 2023 doi:10.1186/s12885-023-10511-3*



# Infographics



Spiegelhalter, D., *The Art of Statistics: Learning From Data*, Pelican, 2019

# Cosa abbiamo imparato in questa lezione?

- Le variabili categoriche possono essere rappresentate con le loro frequenze, assolute o relative
- Tabelle di frequenze e di contingenza ci aiutano a paragonare, a colpo d'occhio, due o più variabili
- Variabili numeriche possono essere trasformate in variabili categoriche attraverso la loro "discretizzazione"
- Le variabili categoriche possono essere rappresentate graficamente in diversi modi, ma bar chart sono da preferire, seguiti da waffle chart