



## Spring Meeting Giovani Ricercatori



SID



SIIA



SIMI



SIPREC



SISA

# Come approcciarsi ai test statistici: building the foundation

Coordinatore: Mario Luca Morieri (Padova)

Intervengono: Federica Galimberti (Milano) e Davide Bernasconi (Milano)

# Agenda

---

**WORKSHOP (parte 1) - Come approcciarsi ai test statistici: building the foundation**

**WORKSHOP (parte 2) - Come approcciarsi ai test statistici: road to Mendelian randomization**

**WORKSHOP (parte 3): Come approcciarsi ai test statistici: road to Propensity Score**

## **Obiettivi:**

- **Favorire il networking intra e intersocietario**
- **Riconoscere gli elementi chiave per usare e interpretare correttamente un test statistico (W1)**
- **Esempi specifici**
  - **Randomizzazione mendeliana (W2)**
  - **Propensity Score (W3)**

# Conosciamoci

---

**Dove lavorate?**

**Di cosa vi occupate ? (argomento principale del vostro abstract)**

**Che livello di esperienza avete?**

# Premessa: associazione vs predizione

---

L'applicazione dei metodi statistici nella ricerca clinica/biologica può avere sostanzialmente due scopi:

## **1. Verificare e quantificare l'associazione tra uno o più fattori di esposizione (variabili indipendenti) e un outcome (variabile dipendente).**

- Spesso questo significa in pratica verificare e quantificare la **differenza** (es. tra medie o tra rischi) **esistente tra diversi gruppi**.
- Esistono inoltre metodi per fare in modo che questa associazione possa essere interpretata in modo **causale**.

## **2. Fornire una predizione dell'outcome specifica per ogni soggetto sulla base del valore delle variabili indipendenti**

- Questo task è solitamente più complesso (fattori associati all'outcome non sempre consentono di ottenere una buona predizione per tutti i soggetti) e più oneroso (dati)
- La performance predittiva dipende molto anche dal modello (regressione, metodi ML)

**In questo workshop ci focalizzeremo principalmente sulla situazione 1.**

# Scelta (e applicazione) del test statistico più appropriato

**Quali elementi chiave valutate per scegliere il test statistico più appropriato?**

- 1. Quesito scientifico**
  - 2. Definizione ipotesi nulla e alternativa**
  - 3. Identificazione variabili di interesse (dipendente e indipendente)**
  - 4. Definizione del disegno dello studio**
  - 5. Fattori confondenti, modulatori o mediatori**
- **Scelta del test statistico e interpretazione risultati (errore alpha e beta)**

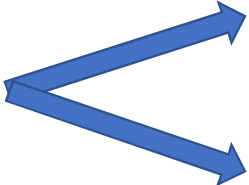
**Quale è il quesito scientifico del vostro studio?**

# Ipotesi nulla e alternativa e test di ipotesi

---

**Ipotesi nulla ( $H_0$ ):** è un'affermazione riguardo alla **popolazione** che si assume essere vera fino a che non ci sia una prova evidente del contrario (status quo, mancanza di effetto etc.)

**Ipotesi alterantiva ( $H_a$ ):** è un'affermazione riguardo alla **popolazione** che è contraria all'ipotesi nulla e che viene accettata solo nel caso in cui **i dati osservati siano non compatibili con  $H_0$**

Test di ipotesi  $\rightarrow$  decisione fra  $H_0$  e  $H_a$  

1. Rifiutare  $H_0$  (e quindi accettare  $H_a$ )
2. Non rifiutare  $H_0$

**P-value (p):** probabilità di trovare nel campione una differenza uguale o più estrema di quella osservata, dato che nella popolazione la differenza non esiste

**Quale è l'ipotesi nulla (e alternativa) del vostro studio?**

# Interpretazione del risultato, errore tipo I (alpha) e II (beta)

---

**Criterio di decisione:** se  $p < \alpha \rightarrow$  rifiuto  $H_0$   
se  $p \geq \alpha \rightarrow$  non rifiuto  $H_0$

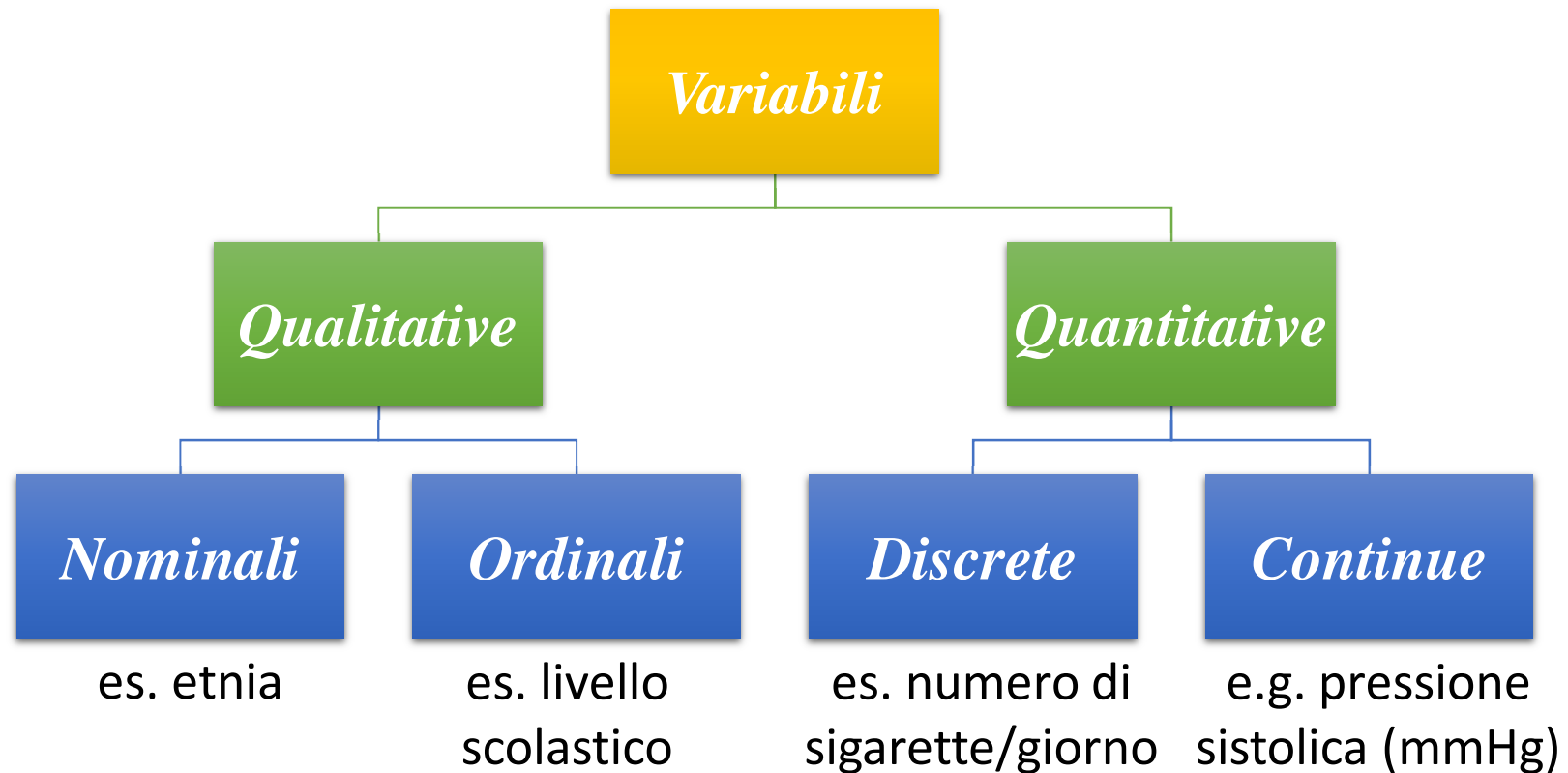
Che probabilità ho di sbagliare o di decidere correttamente?

## Decisione statistica

	Rifiuto $H_0$	Non rifiuto $H_0$
Realtà	<u>Corretta (<math>1-\beta</math>)</u> Associazione osservata quando esiste ( <b>power</b> )	<u>Errore tipo II (<math>\beta</math>)</u> Associazione non osservata, ma in realtà esiste
	<u>Errore tipo I (<math>\alpha</math>)</u> Associazione osservata, ma in realtà non esiste	<u>Corretta (<math>1-\alpha</math>)</u> Associazione non osservata quando non esiste

# Identificazione e tipologia di variabili dipendenti e indipendenti

---



**Quali sono e di che tipo sono le variabili dipendenti e indipendenti?**

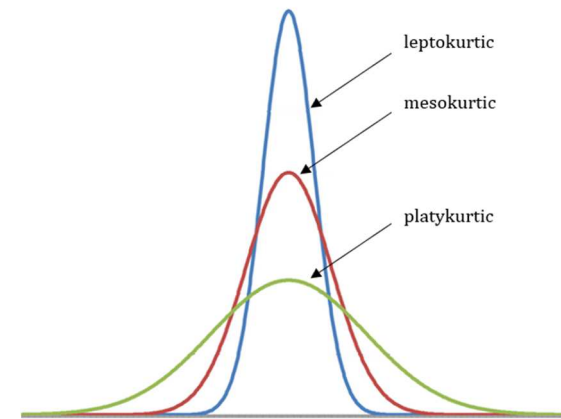
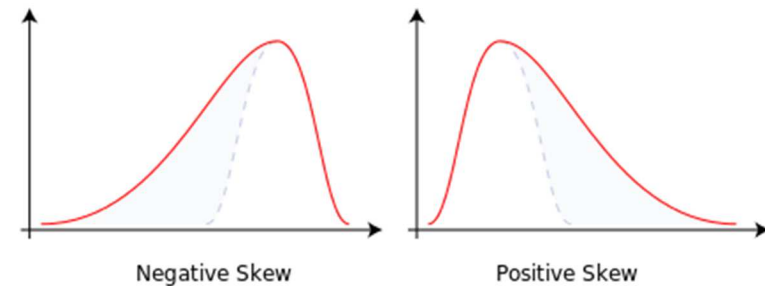


# Come si valuta la normalità di una distribuzione?

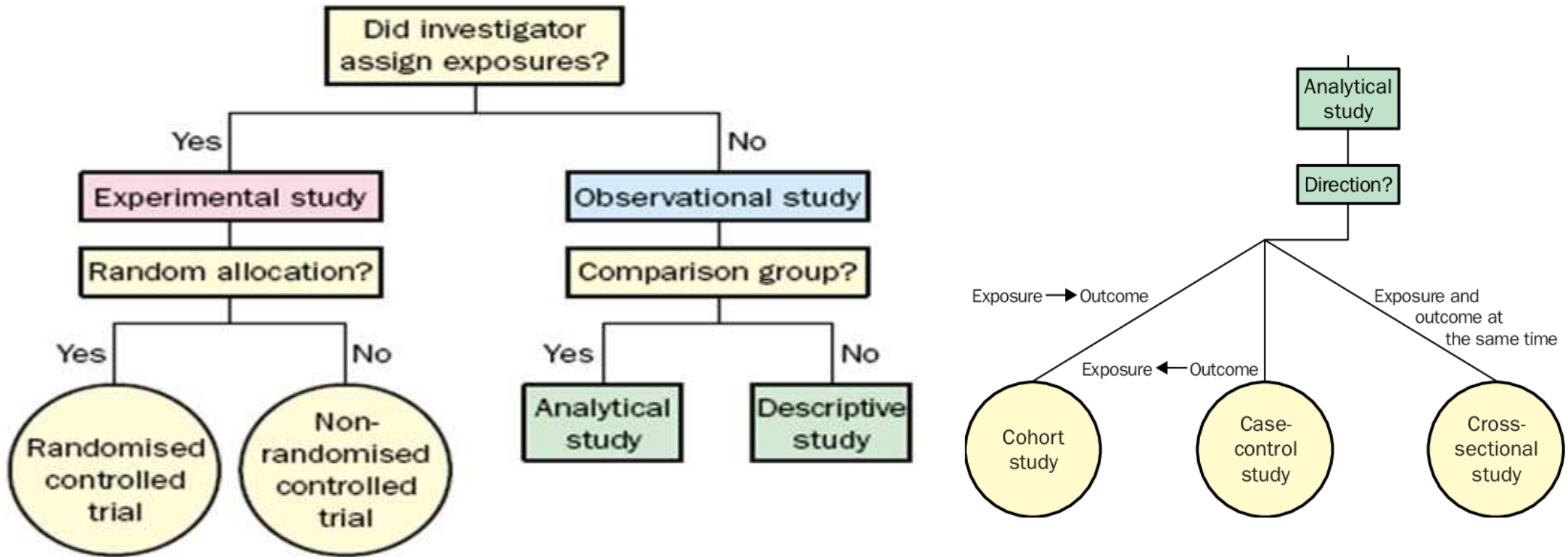
---

*Per le variabili continue*

- **Approccio grafico**
- **Skeweness (simmetria) e Kurtosis**
- **Test di Kolmogorov-Smirnov**



# Valutazione Disegno dello studio



**Quale è il disegno dello studio?**

# Scegliere il test statistico corretto

Nature of Independent Variables	<a href="#">Nature of Dependent Variable(s)*</a>	Test(s)	How to SAS	How to Stata	How to SPSS	How to R
0 IVs (1 population)	interval & normal	one-sample t-test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	ordinal or interval	one-sample median	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical (2 categories)	binomial test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical	Chi-square goodness-of-fit	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
1 IV with 2 levels (independent groups)	interval & normal	2 independent sample t-test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	ordinal or interval	Wilcoxon-Mann Whitney test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical	Chi-square test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		Fisher's exact test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
1 IV with 2 or more levels (independent groups)	interval & normal	one-way ANOVA	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	ordinal or interval	Kruskal Wallis	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical	Chi-square test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
1 IV with 2 levels (dependent/matched groups)	interval & normal	paired t-test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	ordinal or interval	Wilcoxon signed ranks test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical	McNemar	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
1 IV with 2 or more levels (dependent/matched groups)	interval & normal	one-way repeated measures ANOVA	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	ordinal or interval	Friedman test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical (2 categories)	repeated measures logistic regression	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>

# Scegliere il test statistico corretto

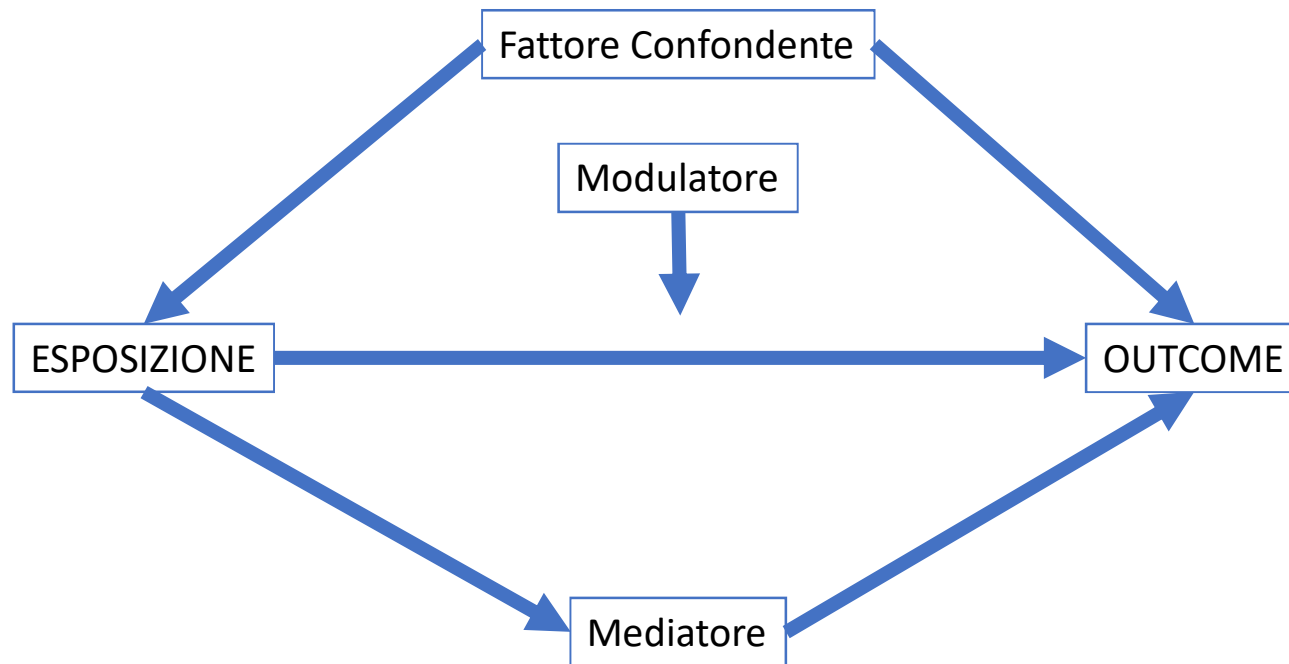
Nature of Independent Variables	<a href="#">Nature of Dependent Variable(s)*</a>	Test(s)	How to SAS	How to Stata	How to SPSS	How to R
2 or more IVs (independent groups)	interval & normal	factorial ANOVA	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	ordinal or interval	ordered logistic regression	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical (2 categories)	factorial logistic regression	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
1 interval IV	interval & normal	correlation	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	interval & normal	simple linear regression	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	ordinal or interval	non-parametric correlation	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical	simple logistic regression	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
1 or more interval IVs and/or 1 or more categorical IVs	interval & normal	multiple regression	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		analysis of covariance	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
	categorical	multiple logistic regression	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		discriminant analysis	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>

<https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/>

# Fattori Confondenti/Modulatori/Mediatori

---

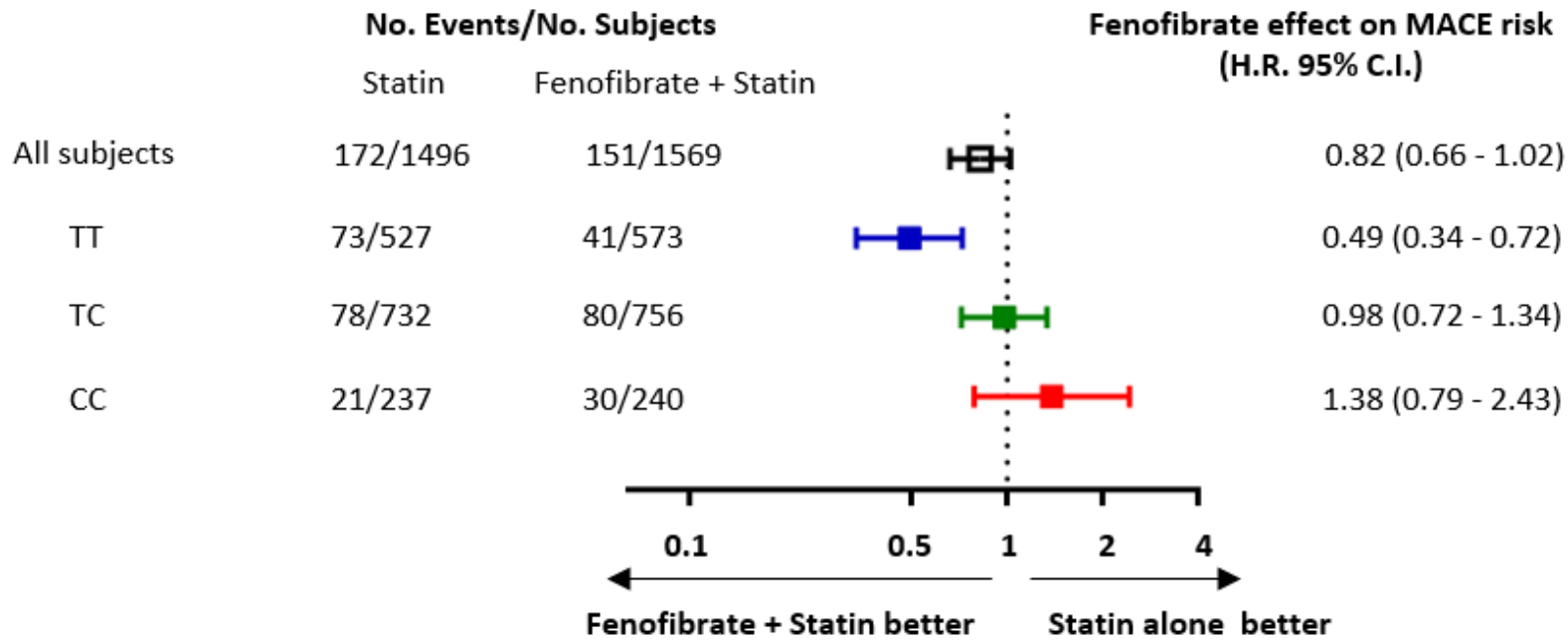
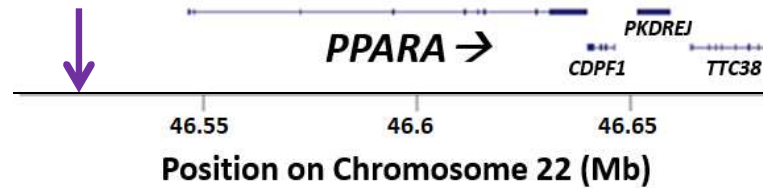
1. Fattore Confondente
2. Modulatore
3. Mediatore



# Modulatori

*Quesito: identificazione di modulatori genetici dell'efficacia del fenofibrato*

Non-coding SNP
rs6008845
Reference Allele T
Alternative Allele C
C Allele Frequency 40%



**P Value for Interaction**

P= 4x10<sup>-4</sup>

# Interazione

---

$$Y = Int + B_1 Farmaco + B_2 SNP + \underline{B_3 "SNP * Farmaco"}$$

# Interpretazione del risultato, errore tipo I (alpha) e II (beta)

---

**Criterio di decisione:** se  $p < \alpha \rightarrow$  rifiuto  $H_0$   
se  $p \geq \alpha \rightarrow$  non rifiuto  $H_0$

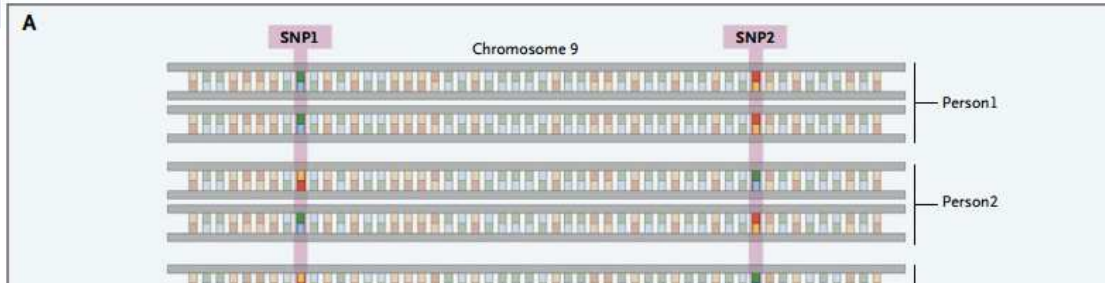
Che probabilità ho di sbagliare o di decidere correttamente?

## Decisione statistica

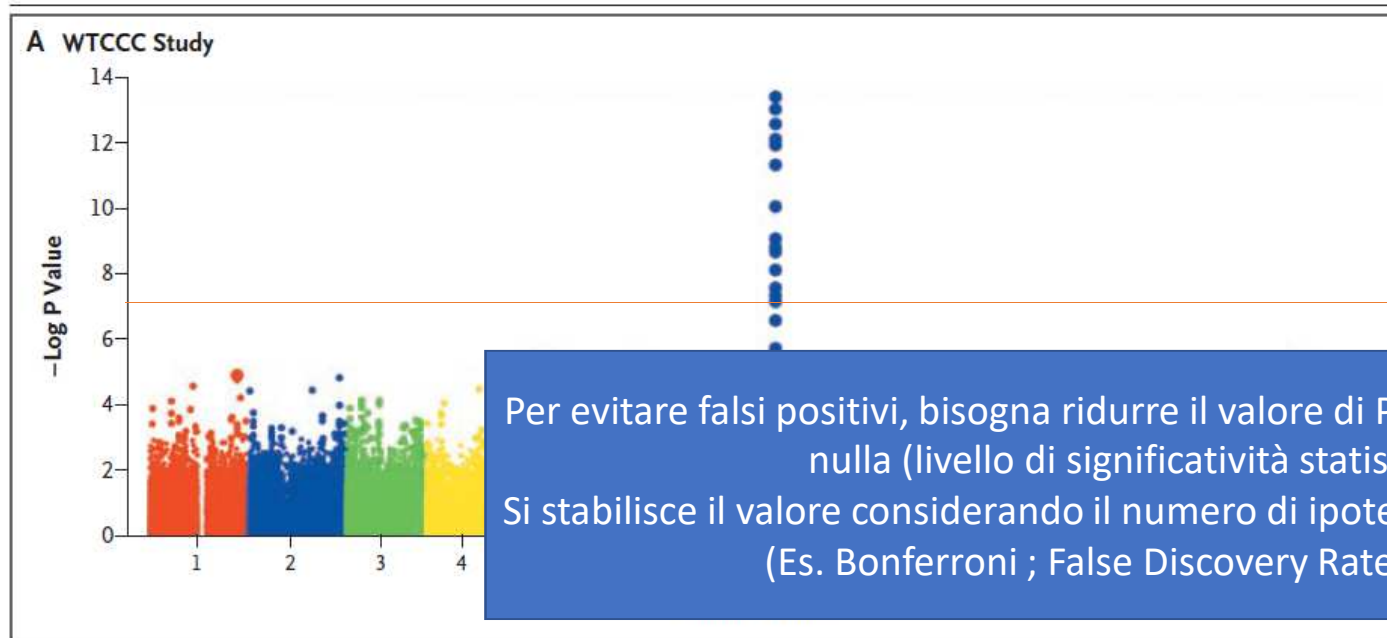
	Rifiuto $H_0$	Non rifiuto $H_0$
Realtà	<u>Corretta (<math>1-\beta</math>)</u> Associazione osservata quando esiste ( <b>power</b> )	<u>Errore tipo II (<math>\beta</math>)</u> Associazione non osservata, ma in realtà esiste
	<u>Errore tipo I (<math>\alpha</math>)</u> Associazione osservata, ma in realtà non esiste	<u>Corretta (<math>1-\alpha</math>)</u> Associazione non osservata quando non esiste



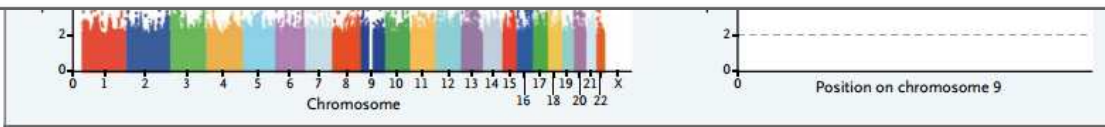
# Esempio possibile inflazione errore alpha – studi genomici



Single Nucleotide Polymorphism - SNP



$P = 5 \times 10^{-8}$



Manolio, NEJM 2010;  
Samani NEJM 2007