



Corso di Laurea di Scienze Forestali ed Ambientali

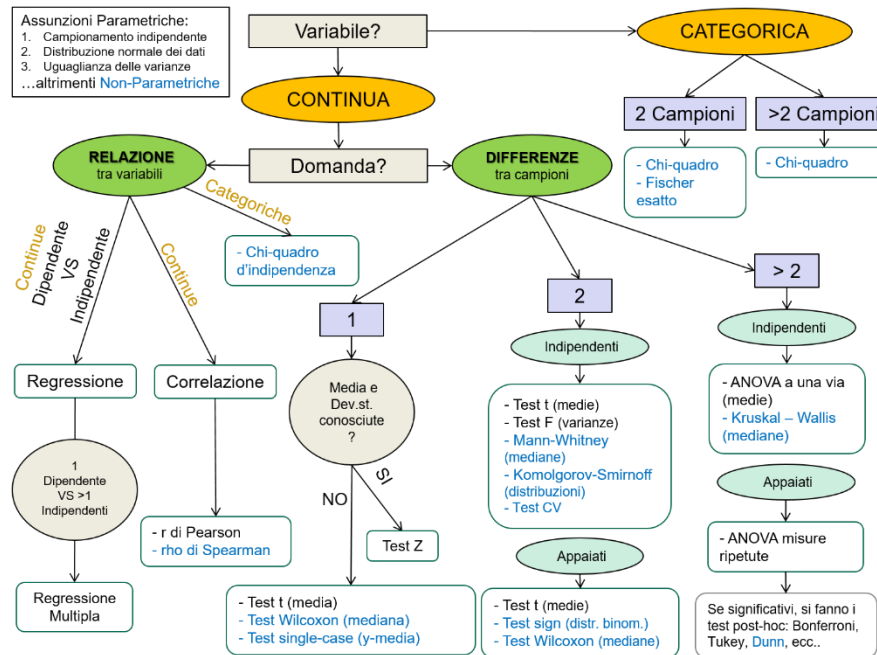
Ecologia e Statistica

per l'ambiente



a.a. 2020-2021 - MATTEO GARBARINO - matteo.garbarino@unito.it

TEST DATI CATEGORICI



TEST DELLE IPOTESI

I test delle ipotesi cercano di rispondere alla domanda «esiste una **differenza** significativa tra i campioni?»

Vengono classificati in **test parametrici** se si verificano le seguenti assunzioni:

1. Il campionamento è indipendente;
2. La distribuzione dei dati è normale;
3. Sussiste l'uguaglianza delle varianze

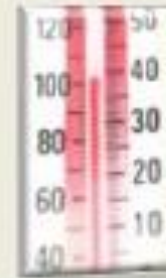
...altrimenti ci si rivolge a **test non-parametrici**. I test parametrici esistono solo per variabili continue, mentre i non parametrici permettono di testare variabili continue, categoriche o nominali.

Test Parametrici

I metodi parametrici vengono utilizzati quando utilizziamo le **STATISTICHE CAMPIONARIE** per rappresentare i **PARAMETRI DI POPOLAZIONE** quando la distribuzione è **NORMALE** e le variabili sono **SCALARI**.



Normal Distribution



Scaled Data

A zero point

Equal Intervals

Test t e ANOVA (PARAMETRICI) sono basati su alcune assunzioni...

1. **Variabili continue** o almeno misurate in un intervallo (es. non conosco il valore assoluto, ma posso quantificare le differenze fra due valori)

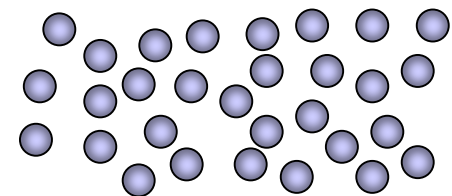
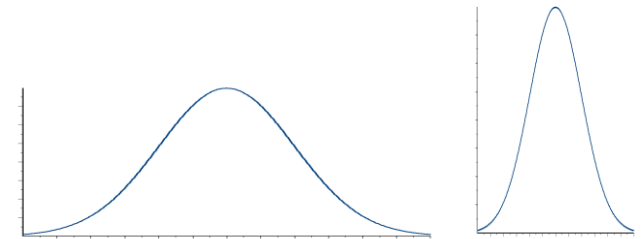
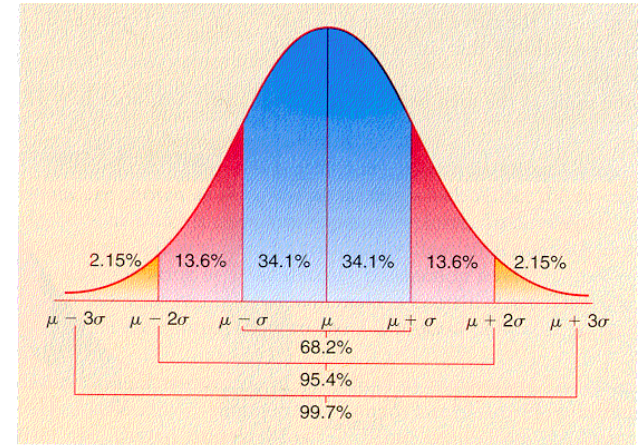
2. **Indipendenza fra media e varianza** (l'errore di misura casuale = indipendente dal valore misurato)

3. **Variabili** distribuite in modo (approssimativamente) **normale**

4. **Omogeneità delle varianze**

5. **I risultati** ottenuti con l'analisi del campione **si applicano alla popolazione**

6. **Dimensione campione** $(n) > 10$ (meglio se ≥ 30)



Test Non-parametrici

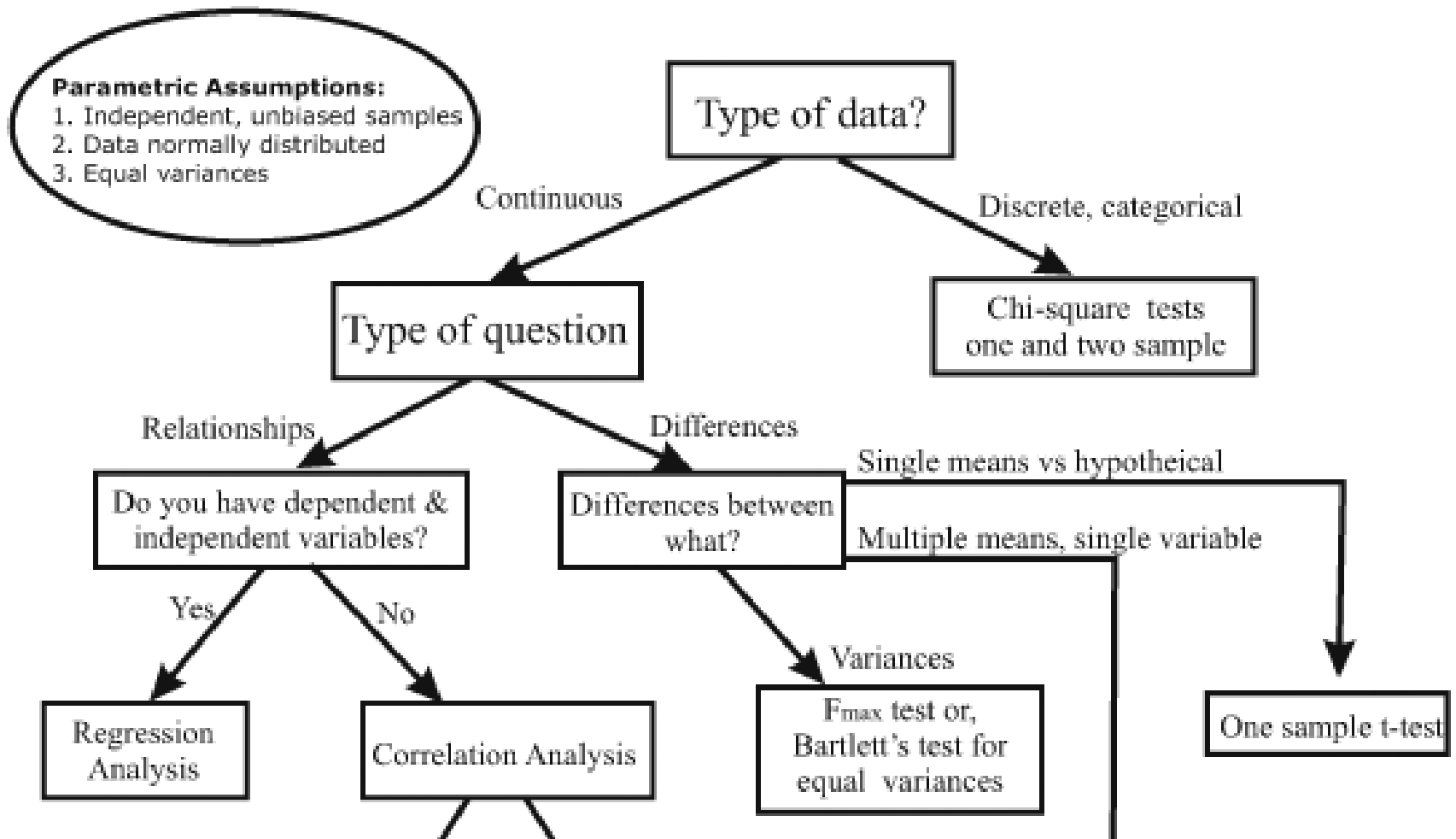
- Questi test si impiegano quando almeno una delle assunzioni alla base del test t di Student o dell'ANOVA è violata.
- Sono chiamati “non-parametrici” perchè essi non implicano la stima di parametri statistici (media, deviazione standard, varianza, etc.).

Ne esistono almeno due grandi categorie:

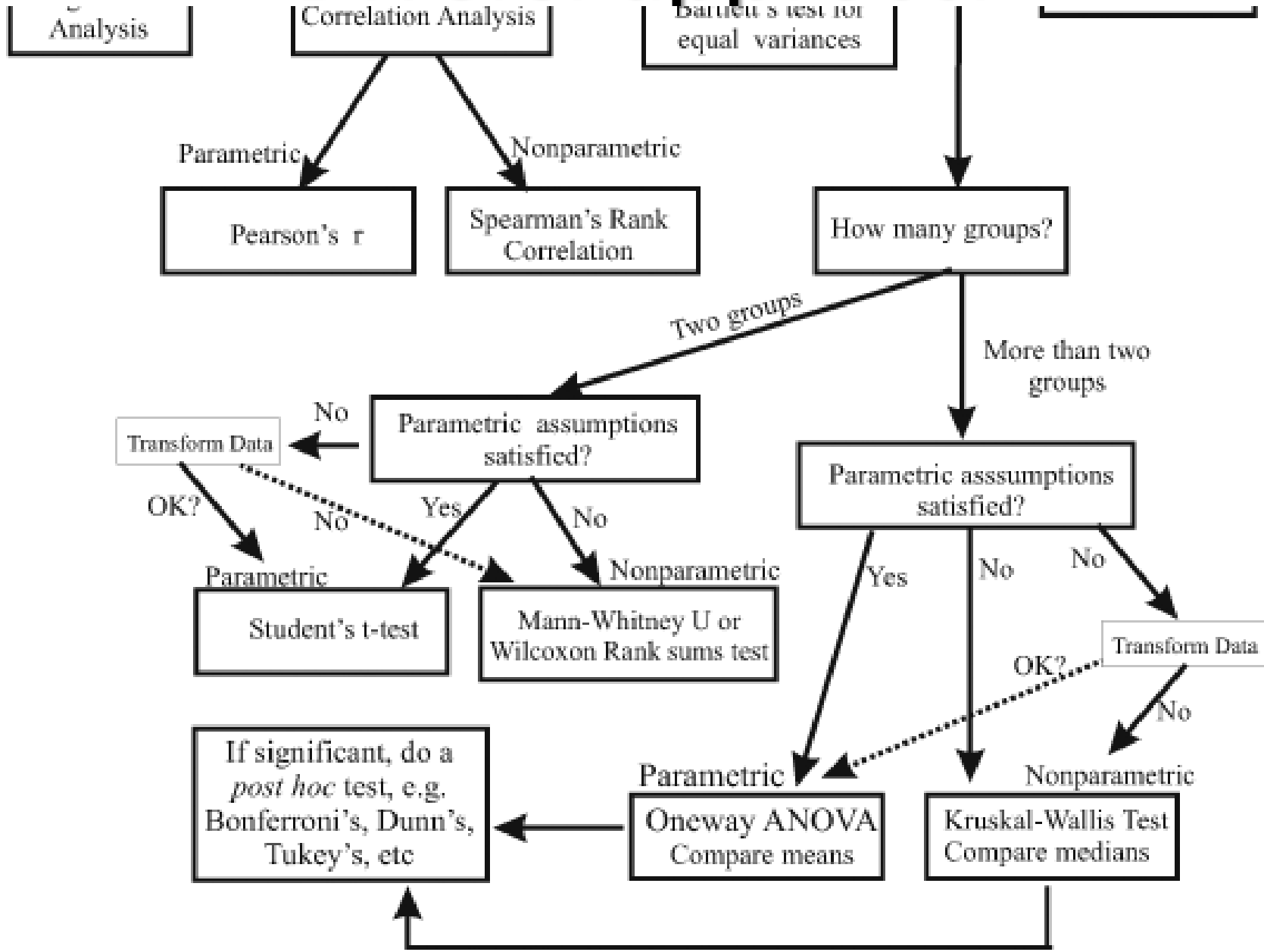
- 1) Test di conformità (confronto fra valori osservati e valori attesi opportunamente calcolati)
- 2) Test equivalenti di test parametrici

Diversi approcci

Flow Chart for Selecting Commonly Used Statistical Tests

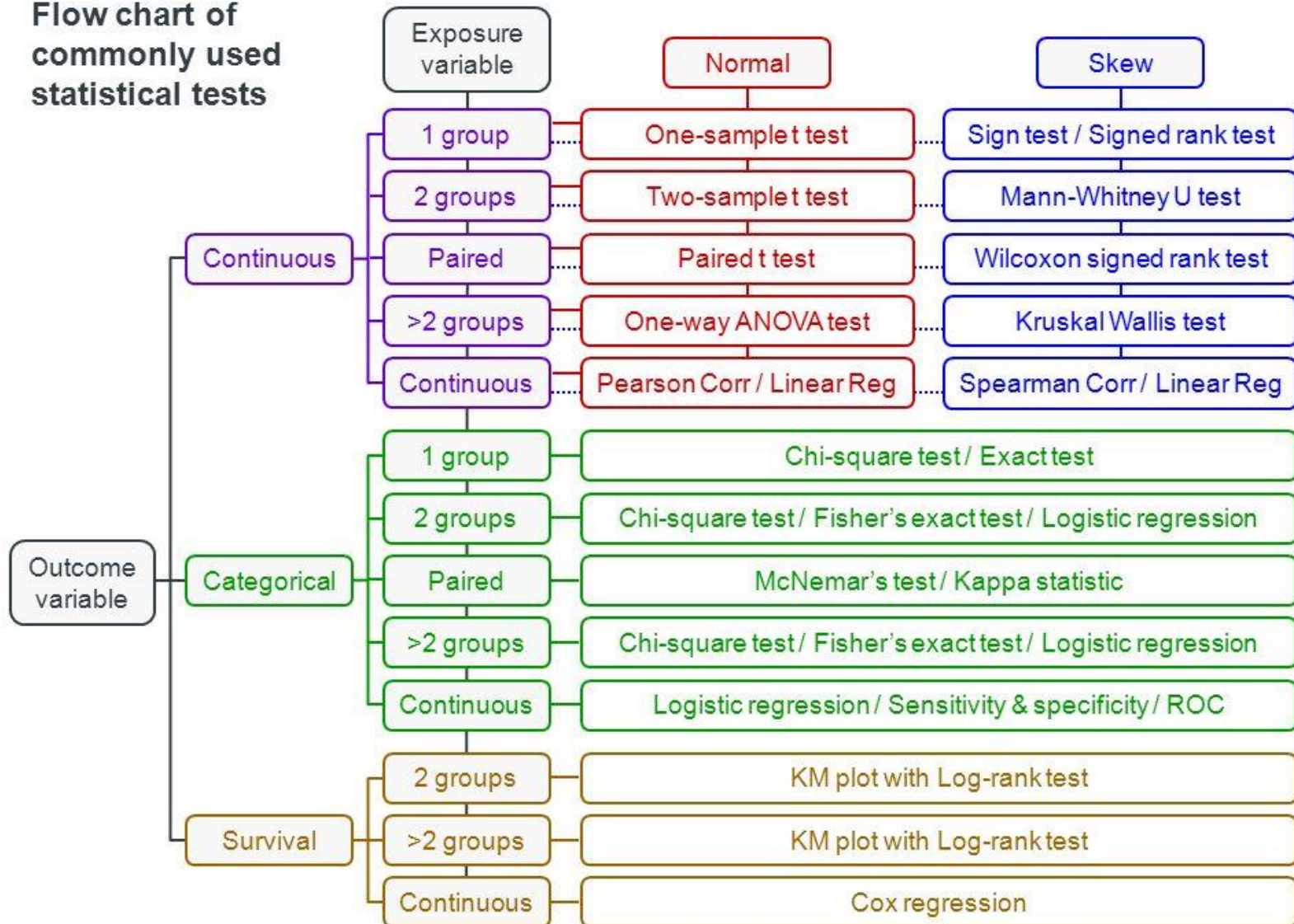


Diversi approcci

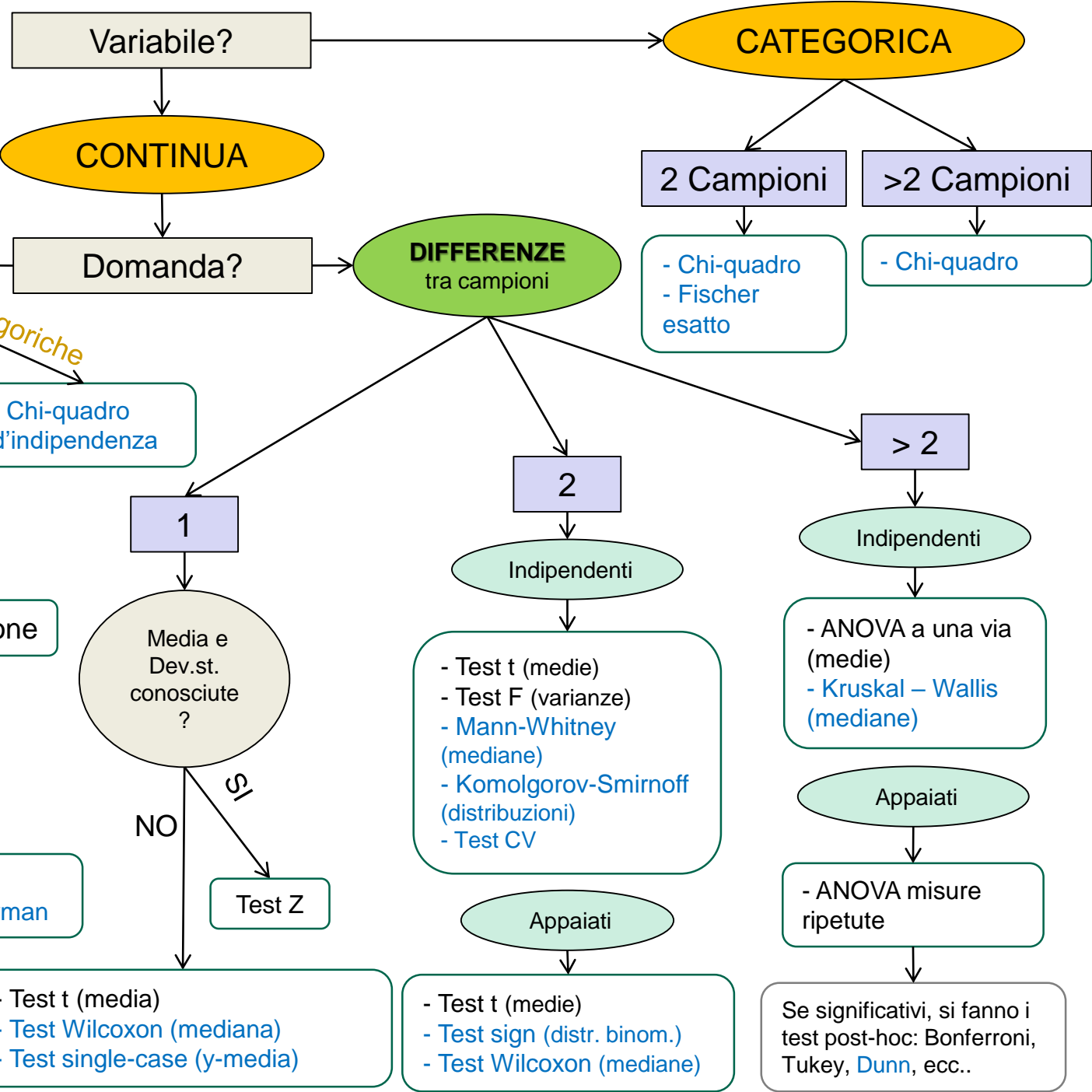


Diversi approcci

Flow chart of commonly used statistical tests



Assunzioni Parametriche:
 1. Campionamento indipendente
 2. Distribuzione normale dei dati
 3. Uguaglianza delle varianze
 ...altrimenti **Non-Parametriche**



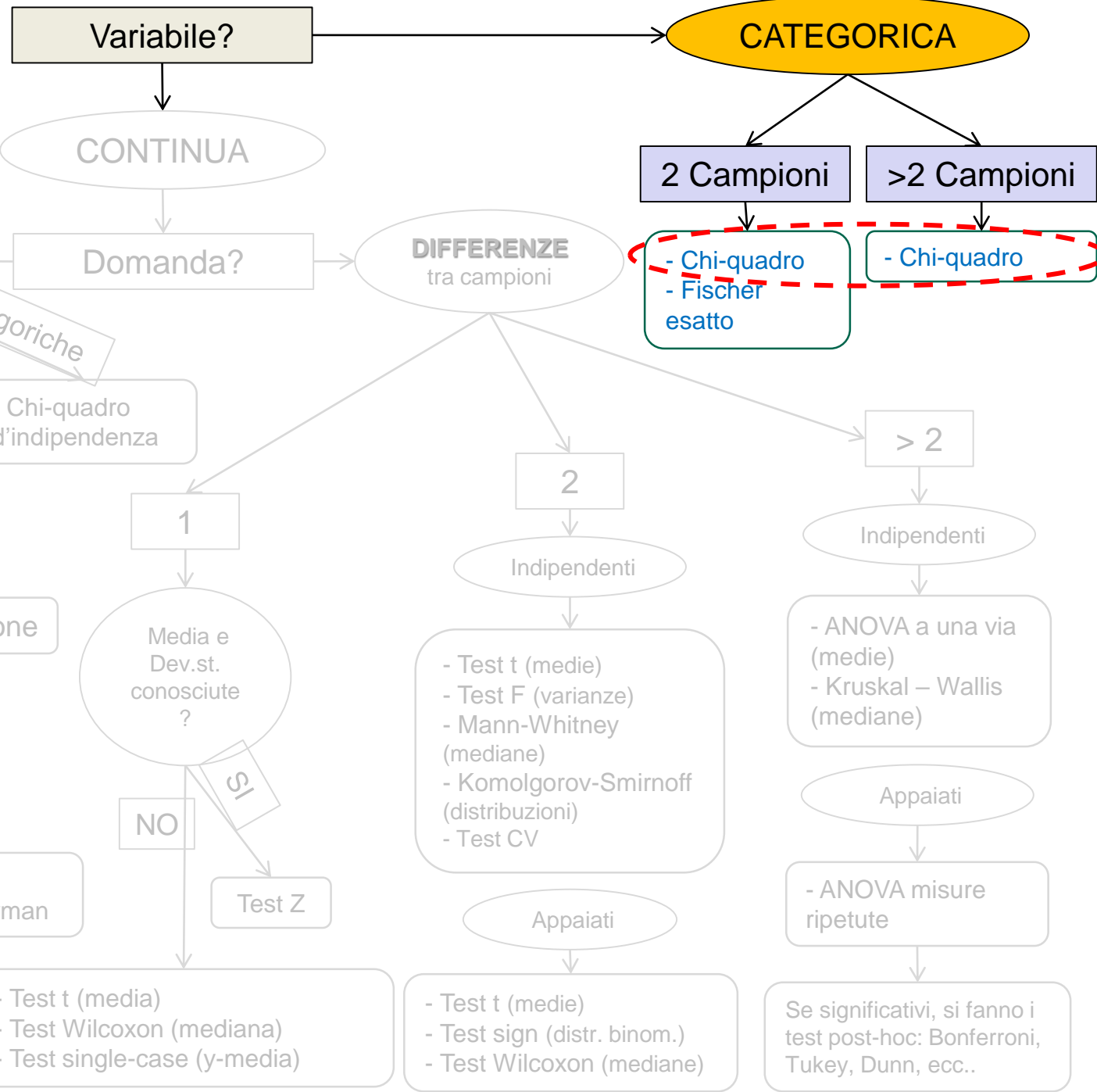
Continue Dipendente VS Indipendente

Continue

Categoriche

Se significativi, si fanno i test post-hoc: Bonferroni, Tukey, Dunn, ecc..

Assunzioni Parametriche:
 1. Campionamento indipendente
 2. Distribuzione normale dei dati
 3. Uguaglianza delle varianze
 ...altrimenti **Non-Parametriche**



Test per dati categorici

Test del Chi-quadro (χ^2)

Esistono due tipi di test del chi-quadro ed entrambi utilizzano la statistica e la distribuzione del chi-quadro (χ^2) per obiettivi differenti:

- **Test χ^2 di goodness-of-fit (GoF)** del chi-quadro, stabilisce se una variabile categorica (dati osservati) è estratta da una distribuzione teorica (dati attesi).
- **Test χ^2 di indipendenza o tabella di contingenza** confronta due variabili «nominali» in una tabella di contingenza per stabilire se le 2 variabili sono corrispondenti. È una generalizzazione del test precedente che testa se due variabili categoriche hanno una relazione o meno.
 - Un valore molto basso di χ^2 indica che **i dati osservati corrispondono (fittano) molto bene con i dati attesi. Relazione forte tra le due variabili categoriche o nominali.**
 - Un valore molto grande di χ^2 indica che non esiste una relazione significativa tra variabile osservata e attesa.

Test del Chi-quadro - GoF

- H_0 : Ipotesi nulla che indica che una distribuzione in classi osservata corrisponde ad una distribuzione attesa.
- H_1 : L'ipotesi alternativa indica che la distribuzione osservata è differente da quella attesa.

Condizione di validità: il campione deve essere sufficientemente grande e non devono esistere classi con numerosità campionaria inferiore a 5.

Per campioni di media misura ($40 < n < 200$) è prevista la correzione di Yates

Test del Chi-quadro - **GoF**

$$\chi_v^2 = \sum_{i=1}^n \frac{(fo_i - fa_i)^2}{fa_i}$$

fo = frequenze osservate

fa = frequenze attese

n = numero di classi

v = numero di gradi di libertà:

Il numero di gradi di libertà corrisponde al numero di classi - 1

Correzione di Yates

$$\chi^2 = \sum_{i=1}^k \frac{(|o_i - e_i| - 0,5)^2}{e_i}$$

Dove O = osservate

E = attese

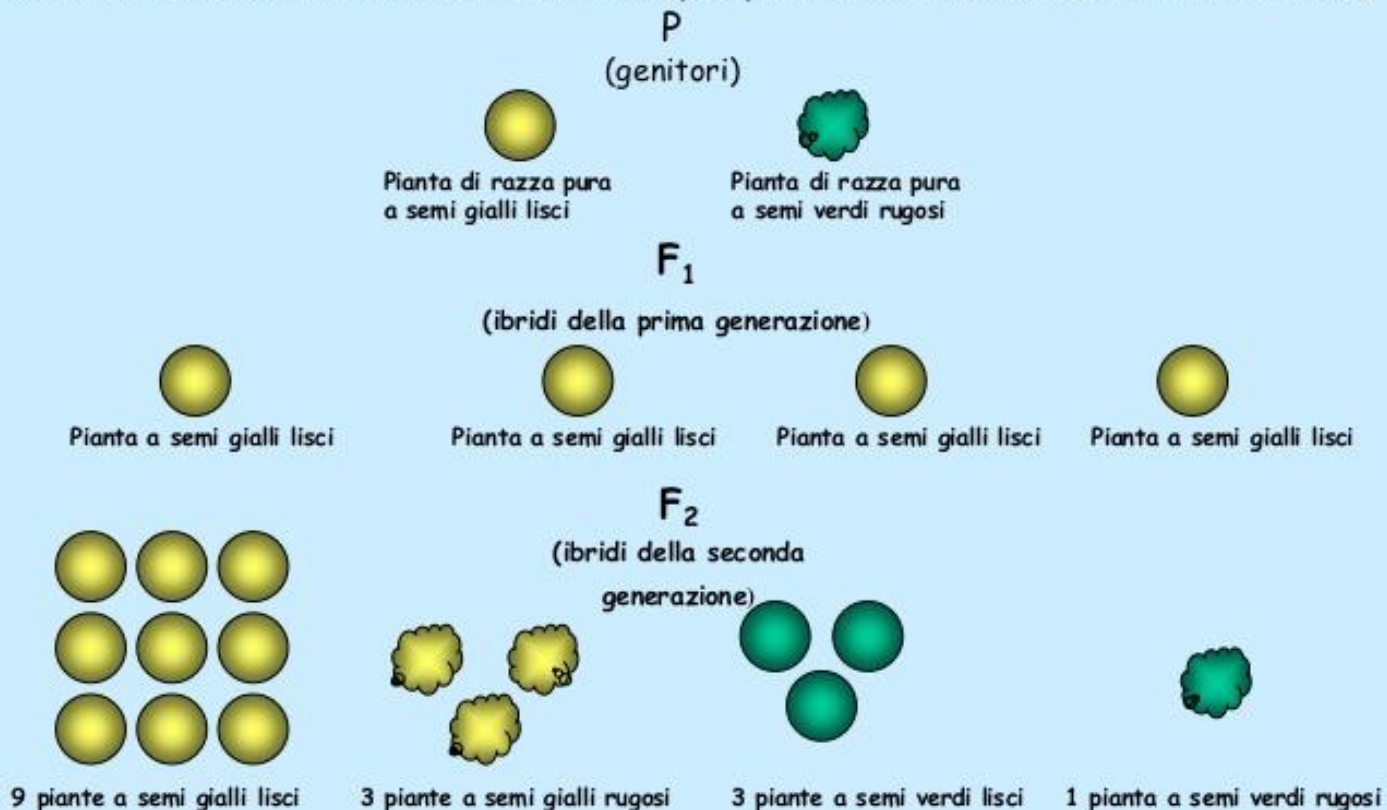
Test del Chi-quadro (χ^2)

Goodness-of-fit

Frequenze attese secondo la teoria

Terza legge di Mendel

Mendel incrociò individui che differivano per più caratteri (colore e forma del seme)



Test del Chi-quadro (χ^2)

Goodness-of-fit

Segregazione di un diibrido	Frequenze	
	Osservate	Attese
Gialli – Lisci	<i>315</i>	$9/16 = 312,75$
Gialli – Rugosi	<i>101</i>	$3/16 = 104,25$
Verdi – Lisci	<i>108</i>	$3/16 = 104,25$
Verdi – Rugosi	<i>32</i>	$1/16 = 34,75$
<i>Totale</i>	<i>556</i>	<i>556,00</i>

Testare l'ipotesi nulla secondo cui la distribuzione delle frequenze osservate può essere considerata uguale a quella delle frequenze attese. Vedi esercizi su Excel

Test del Chi-quadro (χ^2)

Indipendenza – tabelle di contingenza 2 x 2

Nel confronto tra due differenti campioni con risposte alternative di tipo binario, la tabella costruita con i dati sperimentali è chiamata tabella 2 x 2.

Il test chi quadrato permette di verificare se le proporzioni di successi e di insuccessi nei due gruppi sono indipendenti dal trattamento al quale sono sottoposti oppure se esiste associazione tra essi.

Test del Chi-quadro (χ^2)

Indipendenza – tabelle di contingenza 2 x 2

	Cavie		
	Morte	Sopravv.	Totale
Tossico A	22	48	70
Tossico B	24	26	50
Totale	46	74	120

Testare l'ipotesi nulla secondo cui la distribuzione dei morti e dei sopravvissuti nelle due classi è uguale e quindi i due tossici hanno manifestato lo stesso effetto. Complessivamente la tabella ha 1 solo grado di libertà, poiché fissati i totali è possibile fissare solo un numero mentre gli altri derivano di conseguenza

Test del Chi-quadro (χ^2)

Indipendenza – tabelle di contingenza 2 x 2

- Creare una tabella 2 x 2 con le frequenze attese proporzionali ai totali marginali.
- Calcolare il valore del chi quadrato con la formula già vista.
$$\chi_v^2 = \sum_{i=1}^n \frac{(fo_i - fa_i)^2}{fa_i}$$
- Il numero dei gradi di libertà è 1.
- Verificare la probabilità del chi quadrato con i suoi gradi di libertà.

Test del Chi-quadro (χ^2)

Indipendenza – tabelle di contingenza 2 x 2

Frequenze attese = Tot.colonna*Tot.riga/TotaleGenerale

OSS	Morte	Vive	TOT	ATT	Morte	Vive	TOT
Tossico A	22	48	70	Tossico A	26.83333	43.16667	70
Tossico B	24	26	50	Tossico B	19.16667	30.83333	50
TOT	46	74	120	TOT	46	74	120

Calcolo del Chi² →
$$\chi_v^2 = \sum_{i=1}^n \frac{(fo_i - fa_i)^2}{fa_i}$$

Calcolo di p su excel → = distrib.chi(X²;gdl=1)

Se $p < 0.05$ rifiuto l'ipotesi nulla (nessun effetto del tossico)

Quindi con una probabilità alta posso dire che il tossico ha un effetto significativo sulla sopravvivenza delle cavie.

Test del Chi-quadro (χ^2)

Indipendenza – tabelle M x N

Creare una tabella M x N con le frequenze attese proporzionali ai totali marginali.

Calcolare il valore del chi quadrato con la formula già vista.

Calcolare il numero di gradi di libertà con la formula

$$\text{gdl} = (M - 1) \times (N - 1)$$

Verificare la probabilità del chi quadrato con i suoi gradi di libertà.

Test del Chi-quadro (χ^2)

Indipendenza – tabelle M x N

	Specie A	Specie B	Specie C	Specie D	Specie E	Totale
Coltivazione I	12	8	21	5	4	50
Coltivazione II	15	10	5	20	8	58
Coltivazione III	9	6	10	17	11	53
Coltivazione IV	23	12	12	31	17	95
Totale	59	36	48	73	40	256

Si vuole verificare se esiste associazione tra tipo di coltivazione del terreno e presenza di alcune specie d'insetti. In 4 diversi appezzamenti di terreno, con coltivazioni differenti, è stata contata la presenza di 5 specie differenti di insetti, secondo le frequenze riportate nella tabella.

Testare l'ipotesi nulla secondo cui la distribuzione delle differenti specie di insetti è la stessa nelle 4 coltivazioni.

Test del Chi-quadro (χ^2)

Indipendenza – tabelle $M \times N$

Interpretazione dei valori χ^2 e del suo p

In generale se ottengo un valore alto di chi-quadrato e quindi un valore basso di p devo rifiutare l'ipotesi nulla e quindi ritenere attendibile l'ipotesi alternativa secondo cui esiste una relazione tra le variabili categoriche analizzate.