Alma Mater Studiorum
Università di Bologna

scuola di scienze
Corso di Laurea Magistrale in Matematica

I test d'ipotesi e la loro declinazione in ambito medico

Tesi di Laurea in Statistica inferenziale

Relatore:
Prof.
PAOLO NEGRINI

Correlatore:
Prof.
ADA DORMI

Presentata da:
FRANCESCA GIOSA

Sessione I
Anno Accademico 2017-2018

Indice

Introduzione

Una domanda di rilevante interesse in ambito medico e in particolare nell'ambito della ricerca medica è:

- un determinato farmaco è efficace per curare una malattia?

E ancora:

- avendo a disposizione un certo numero di farmaci quale di questi è più efficace?

E' chiaro quanto sia importante conoscere l'efficacia di un farmaco prima di somministrarlo, in quanto se si pensa che il farmaco sia efficace quando non lo è, si rischia non solo di non guarire il paziente, ma anche di far subentrare altre problematiche dovute al farmaco inutile. Allo stesso tempo sarebbe grave anche non rendersi conto dell'efficacia di un farmaco che potrebbe invece curare un paziente.
La statistica medica e in particolare lo strumento statistico comunemente chiamato *Test d'ipotesi* si occupano di rispondere a queste domande e di limitare il più possibile le problematiche esposte, dovute agli errori che si possono commettere.
Lo scopo di questa tesi è spiegare l'utilizzo dei test d'ipotesi e in particolare di analizzare il confronto tra due e più popolazioni (che si traduce ad esempio nel saggiare la differenza tra 4 farmaci) ponendo attenzione alle varie possibilità di impostazione del problema a seconda delle ipotesi che vengono di volta in volta soddisfatte.
Infatti a seconda del disegno sperimentale del problema che si vuole trattare si fa un uso diverso della stessa tecnica. La maggior parte degli errori nella letteratura biomedica riguarda errori basilari nel disegno sperimentale, come per esempio una errata procedura di randomizzazione o l'omissione del gruppo di controllo, oppure l'utilizzo non appropriato del *test t* per confronti multipli.
Pertanto è bene fare attenzione alle diverse possibilità e capire quali sono le più adatte nei vari casi. Per fare ciò la presente tesi è stata suddivisa in quattro capitoli così articolati: nel primo capitolo, tratto da D.Piccolo [5], si riassumono le nozioni fondamentali di statistica che sono utilizzate nel seguito. Successivamente nel secondo capitolo, tratto da P. Baldi [2], si spiegano i Test d'ipotesi nel caso generale e nel caso del confronto di due popolazioni Gaussiane (confronto sia di media che di varianza). Nel terzo capitolo, tratto da Wayne W. Daniel [3] e dal sito di Franco Anzani e Maria Pia D'Ambrosio [1], si studia l'*analisi della varianza di tipo parametrico*, ovvero il confronto tra medie di più di due popolazioni Gaussiane, analizzata in tutti i suoi aspetti e nelle varie possibilità. Infine nel quarto ed ultimo capitolo, tratto sia da Wayne W. Daniel [3] che da Stanton A. Glantz [4], si analizza l'*analisi della varianza non parametrica*, ovvero il confronto di più di due popolazioni nel caso in cui le distribuzioni non si possano considerare di tipo Gaussiano.
Il percorso seguito è basato sulla importanza della *adeguatezza del modello utilizzato* e della verifica delle ipotesi che permettono di utilizzare un metodo di risoluzione piuttosto che un altro. Nella tesi sono esposti con dimostrazione i vari metodi e concetti utilizzati in modo

da avvalorare l'importanza delle ipotesi necessarie e sufficienti per l'utilizzo dei vari test. Alla fine del terzo capitolo presentiamo un esempio tratto da una situazione reale, con dati forniti dall'istituto per lo scompenso cardiaco di Bologna. Con questo esempio si vuole mettere in evidenza l'importanza della *adeguatezza del modello* e dei problemi pratici in cui si può incorrere.

Capitolo 1

Richiami ad alcune nozioni di statistica

Lo scopo di questo capitolo è ricordare (senza dimostrazione) alcuni risultati di statistica, che saranno utili nei capitoli successivi.

Definizione 1.1 (**Variabile aleatoria**). *Dato uno spazio di probabilità* $(\Omega, \mathcal{A}, p)$ *dove* Ω *è lo spazio degli eventi,* $\mathcal{A} \subset P(\Omega)$ *è una sigma algebra su* Ω *e* p *è una misura di probabilità, una variabile aleatoria è una funzione misurabile:*

$$X : \Omega \longrightarrow \mathbb{R}^n.$$

Definizione 1.2 (**Funzione di ripartizione**). *Data una variabile aleatoria* X*, si definisce funzione di ripartizione di* X *la funzione:*

$$F_X(x) : \mathbb{R} \longrightarrow [0, 1]$$

tale che:

$$F_X(x) = P(X \leq x),$$

in altre parole è la funzione che ad ogni elemento x *associa la probabilità dell'evento: "la variabile aleatoria* X *assume valori minori uguali di* x*".*
La funzione di ripartizione $F_X(x)$ *è una funzione non decrescente, continua a destra e tale che:*

$$\lim_{x \to +\infty} F_X(x) = 1 \quad e \quad \lim_{x \to -\infty} F_X(x) = 0.$$

- *Se* X *è una variabile aleatoria discreta, che assume valori in* $E = x_1, ...x_n..$ *insieme numerabile, allora definendo* $p_i(x) = P(X = x_i)$ *si ha che:*

$$F_X(x) = \sum_{x_i \leq x} p(x_i).$$

- *Se* X *è una variabile aleatoria assolutamente continua, la funzione di ripartizione di* X *è:*

$$F_X(x) = \int_{-\infty}^{x} f(u)\, du,$$

dove $f(x)$ *è la funzione densità di* X*.*

La funzione densità è l'analogo della funzione di probabilità nel caso continuo. Infatti:

$$f(x) : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \to \lim_{dx \to 0} \left[\frac{P(x < X \leq x + dx)}{dx} \right],$$

tale che:

$$f(x) \geq 0 \quad e \quad \int_{-\infty}^{+\infty} f(x)\, dx = 1$$

Definizione 1.3 (**Valore atteso**). *Data una variabile aleatoria X e uno spazio di probabilità $(\Omega, \mathcal{A}, p)$, si definisce valore atteso della variabile aleatoria X o speranza matematica la quantità:*

$$E[X] = \int_{\Omega} X(\omega)\, dp(\omega)$$

- *Se la variabile aleatoria X è discreta e ammette funzione di probabilità p_i, si ha:*

$$E[X] = \sum_{i=1}^{+\infty} x_i p_i.$$

- *Se la variabile aleatoria X è assolutamente continua e ammette funzione densità $f(x)$ si ha:*

$$E[X] = \int_{-\infty}^{+\infty} x f(x)\, dx.$$

Definizione 1.4 (**Varianza**). *La varianza di una variabile aleatoria X è una funzione che fornisce una misura della variabilità dei valori assunti dalla variabile stessa. Nello specifico è la misura di quanto essi si discostino quadraticamente dal valore atteso. La varianza di una variabile aleatoria X è genericamente indicata con il simbolo σ_X^2, dove il quadrato sta ad indicare che è una quantità sempre positiva ed è definita come segue:*

$$\sigma_X^2 = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2.$$

Definizione 1.5 (**Modello statistico**). *Si definisce modello statistico una famiglia di spazi di probabilità $\left(\Omega, \mathcal{A}, P^\theta\right)$ dipendenti da un parametro $\theta \in \Theta$.*

Definizione 1.6 (**Osservazione**). *Si definisce osservazione di una variabile aleatoria X definita su Ω, un vettore $(X_1 ... X_n)$ di variabili aleatorie definite su Ω indipendenti tra loro e aventi stessa distribuzione di X.*

Definizione 1.7 (**Statistica**). *Data una osservazione $(X_1 .... X_n)$ di una variabile aleatoria, una statistica è una funzione dell'osservazione:*

$$T = t(X_1 .... X_n).$$

Definizione 1.8 (**Stimatore**). *Sia $\psi(\theta)$ una funzione del parametro θ, che si vuole stimare; uno stimatore è una statistica $T = t(X_1 .... X_n)$ che assume valori nel codominio di ψ.*

Definizione 1.9 (**Stimatore corretto o non distorto**). *Uno stimatore $T = t(X_1 ... X_n)$ si dice corretto se per ogni $\theta \in \Theta$:*

$$E^\theta[\psi(\theta)] = E^\theta[T],$$

dove $E^\theta[T]$ indica il valore atteso calcolato rispetto alla probabilità P^θ, dipendente dal parametro.

Proposizione 1.10 (**Stimatore corretto per la media**). *sia $X = (X_1....X_n)$ un campione di n variabili aleatorie indipendenti ed equidistribuite, con speranza matematica finita per ogni $\theta \in \Theta$, allora:*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

è uno stimatore corretto per la media di ogni X_i.

Proposizione 1.11 (**Stimatore corretto per la varianza**). *sia $X = (X_1....X_n)$ un campione di n variabili aleatorie indipendenti ed equidistribuite, con media e varianza finite per ogni $\theta \in \Theta$, allora:*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

è uno stimatore corretto per la varianza di ogni X_i.

1.0.1 Richiami a distribuzioni di probabilità continue

Definiamo ora alcune distribuzione di probabilità continue che saranno utilizzate nei prossimi capitoli:

Definizione 1.12 (**Distribuzione normale**). *La distribuzione normale, o di Gauss (o gaussiana) è una distribuzione di probabilità continua che trova applicazione in molteplici situazioni. Dipende da due parametri: la media μ e la varianza σ^2 ed è generalmente indicata con il simbolo:*

$$N(\mu, \sigma^2)$$

e la densità è data dalla formula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad con\ x \in \mathbb{R}.$$

Definizione 1.13 (**Distribuzione normale standard**). *La distribuzione normale standard è una normale con media $\mu = 0$ e varianza $\sigma^2 = 1$:*

$$Z \sim N(0, 1),$$

è molto utile nelle applicazioni e se una variabile aleatoria X ha distribuzione normale $X \sim N(\mu, \sigma^2)$, allora:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

La densità di Z è evidentemente:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x^2}{2}} \quad con\ x \in \mathbb{R},$$

ed è rappresentata in figura 1.1

Definizione 1.14 (**Distribuzione Gamma**). *La distribuzione Gamma è una distribuzione di probabilità continua, che comprende, come casi particolari, le distribuzioni esponenziale e chi quadrato. Assume valori reali positivi e dipende da due parametri, di solito indicati con le lettere α e λ. Indicata con $\Gamma(.)$ la funzione di Eulero:*

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^x \, dx,$$

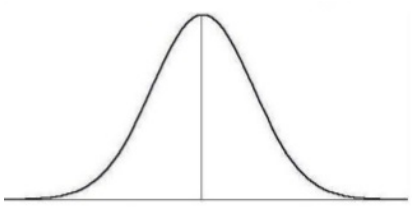


Figura 1.1: Densità normale standard

si dice che X ha distribuzione gamma con parametri (α, λ) e si scrive :

$$X \sim \Gamma(\alpha, \lambda),$$

se la densità di probabilità di X è:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp^{-\lambda x}, \quad x > 0.$$

Definizione 1.15 (**Distribuzione Chi quadro**). *la distribuzione χ^2 è la distribuzione di probabilità della somma dei quadrati di variabili aleatorie indipendenti normalizzate:*

$$\chi^2(n) = \sum_{i=1}^{n} X_i \quad con \quad X_i \sim N(0,1),$$

n è detto numero di gradi di libertà. La distribuzione χ^2 è un caso particolare della distribuzione Γ, infatti:

$$\chi^2(n) = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right),$$

pertanto la sua densità è definita solo per valori positivi e per $n \geq 3$ ha il seguente aspetto qualitativo 1.2:

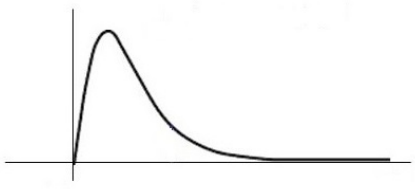


Figura 1.2: Densità Chi quadro a n gradi di libertà

Una importante proprietà della distribuzione χ^2 è l'additività: la somma di variabili aleatorie indipendenti con distribuzioni:

$$\chi^2(k_1), \ldots, \chi^2(k_n),$$

è una variabile aleatoria con distribuzione:

$$\chi^2(k_1 + \ldots + k_n).$$

Definizione 1.16 (**Distribuzione t-student**). *Una variabile aleatoria T ha distribuzione t-student a n gradi di libertà, cioè $T \sim t(n)$ se:*

$$T = \sqrt{n}\frac{X}{\sqrt{Y}} \quad con \quad X \sim N(0,1) \quad e \quad Y \sim \chi^2(n).$$

La densità della variabile T é:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

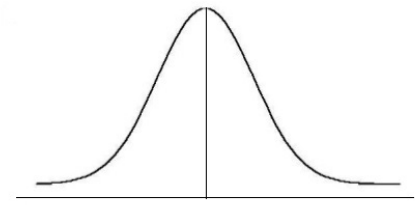


Figura 1.3: Densità t-student a n gradi di libertà

Come si vede in figura 1.3, la densità è una funzione pari, tende a zero per $|x| \to +\infty$ e ha il suo massimo in zero, pertanto il suo andamento è simile a quello della densità di una normale standard, ma decresce a zero più lentamente di quest'ultima, come si vede in figura 1.4. E' bene osservare, però, che tanto più n è grande, tanto più la densità è simile a una Gaussiana:

$$t(n) \longrightarrow N(0,1) \quad per \quad n \to +\infty,$$

si ha una convergenza puntuale per la successione in n:

$$\lim_{n\to\infty}\left(\left(1+\frac{x^2}{n}\right)^n\right)^{-\frac{n+1}{2}\frac{1}{n}} = \exp^{-\frac{x^2}{n}}.$$

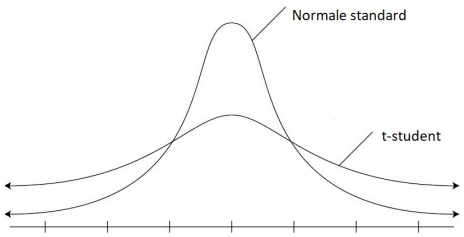


Figura 1.4: Densità t-student a n gradi di libertà enormale standard

Definizione 1.17 (**Distribuzione di Fisher**). *La distribuzione di Fisher-Snedecor è una distribuzione di probabilità continua che regola il rapporto "riscalato" tra due variabili aleatorie che seguono due distribuzioni χ^2; $F(n,m)$ ha distribuzione di Fisher con n, m gradi di libertà se:*

$$F(n,m) = \frac{Z_n}{Z_m}\frac{m}{n} \quad con \quad Z_n \sim \chi^2(n) \quad e \quad Z_m \sim \chi^2(m),$$

essa assume solo valori positivi e ha per densità:

$$f(x) = \frac{\Gamma\left(\frac{n}{2}+\frac{m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} (m+nx)^{-\frac{1}{2}(n+m)} \quad per \quad x > 0$$

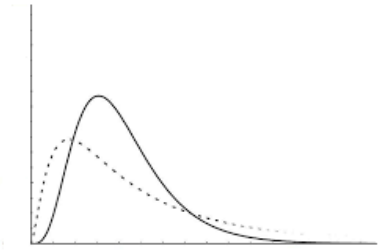


Figura 1.5: Densità di Fisher

Teorema 1.18 (**teorema di Cochran**). *Sia $X = (X_1.....X_n)$ una variabile aleatoria con distribuzione $N(0, I)$ e $X_i \sim N(0, 1)$ indipendenti. Siano $E_1.....E_k$ sottospazi vettoriali di $\mathbb{R}^n$ a due a due ortogonali e tali che $dim(E_i) = n_i$.*
Allora $P_{E_i}(X)$, ovvero le proiezioni di X sui sottospazi E_i sono indipendenti a due a due e

$$\|P_{E_i}(X)\|^2 \sim \chi^2(n_i).$$

Corollario 1.19. *Siano $Z_1.....Z_n$ tali che $Z_i \sim N(\mu, \sigma^2)$ e indipendenti a due a due.*
Allora poste:

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

$$S_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2,$$

si ha che:

- *$\bar{Z}$ e S_z^2 sono due variabili aleatorie indipendenti e*
- *$\frac{n-1}{\sigma^2}S_z^2 \sim \chi^2(n-1)$ e $\sqrt{n}\frac{\bar{Z}-\mu}{S_z} \sim t(n-1)$.*

Capitolo 2

Test d'ipotesi

Lo scopo di questo capitolo è quello di ricordare il concetto di test d'ipotesi e i passi fondamentali da compiere per giungere a una tesi "valida".

Il test d'ipotesi è lo strumento statistico che si utilizza quando si vuole rispondere "si" o "no" a una domanda di rilevante interesse a proposito di un dato problema. Nell'ambito medico una domanda interessante potrebbe essere:

- Un farmaco di nuova concezione è davvero efficace per curare una determinata malattia?

Tale domanda non può avere una risposta certa. Infatti risulta impossibile testare il farmaco sulla totalità dei pazienti e constatare che per tutti o per nessuno il farmaco è efficace. L'unica cosa possibile è affidarsi alla statistica e quindi all'analisi di un campione di pazienti su cui sperimentare il farmaco. Successivamente, in base ai risultati dell'esperimento si dà una risposta positiva o negativa alla domanda posta, nella consapevolezza della presenza di un errore a probabilità non nulla riguardo alla tesi raggiunta.

2.1 Come formulare un test d'ipotesi

Come abbiamo brevemente accennato lo scopo di un test d'ipotesi è di "raggiungere" una conclusione riguardo un parametro della popolazione, esaminando un campione estratto da quella popolazione. I più consueti test statistici si riferiscono a valori di parametri scalari e si svolgono seguendo i seguenti tre semplici passi:

1. **Formulazione dell'ipotesi**

   Nella verifica delle ipotesi vi sono due ipotesi statistiche che devono essere esplicitamente formulate. La prima è la cosiddetta *ipotesi nulla* che è l'ipotesi che deve essere testata e che solitamente vuole essere screditata; è ciò che si oppone a quanto il ricercatore sta cercando di affermare. Questa viene indicata con il simbolo H_0. Se il procedimento di verifica porta al rifiuto di tale ipotesi, diremo che i dati non sono compatibili con l'ipotesi nulla, ma lo sono con qualche altra ipotesi. Quest'ultima è detta *ipotesi alternativa* e viene rappresentata dal simbolo H_A. A questo punto si stabilisce una partizione dello spazio dei parametri:

   $$\Theta = \Theta_{\mathbf{A}} \bigcup \Theta_{\mathbf{H}}$$

   e si usa dire: *l'ipotesi è vera* se $\theta \in \Theta_{\mathbf{H}}$, mentre, *l'ipotesi è falsa* se $\theta \in \Theta_{\mathbf{A}}$.

2. **Modello statistico**

   Si stabilisce una statistica adeguata, detta $T(X)$, funzione delle osservazioni, per effettuare l'esperimento.

3. **Regola di decisione**

   Si stabilisce una regola di decisione che, in base ai valori di $T(X)$ porti a respingere l'ipotesi oppure a non respingerla. Per fare questo si definisce un sottoinsieme D (regione di rigetto) dell'insieme dei valori che $T(X)$ può assumere. Se l'osservazione $T(X)$ assumerà un valore in D respingeremo l'ipotesi.

Pertanto, per eseguire un test d'ipotesi è fondamentale scegliere una consona statistica test e una opportuna regione di rigetto. E' bene osservare che qualunque sia la regione di rigetto D c'è sempre una probabilità non nulla di ottenere un'osservazione $T(X) \in D$ anche se l'ipotesi nulla è vera, cioè anche se $\theta \in \Theta_0$. Affinchè il test sia valido l'unica cosa che si può fare è limitare la probabilità di errore. A questo proposito distinguiamo due tipi di errori che si possono commettere:

Definizione 2.1 (**Errore di prima specie**). *L'errore di prima specie è l'errore che si commette quando si respinge una ipotesi quando essa è vera. L'errore di prima specie si compie quindi se:*

$$\theta_0 \in \Theta_H \quad e \quad T(X) \in D.$$

La probabilità di commettere tale errore viene indicata con

$$\Pi(\theta_0) = P^{\theta_0}(T(X) \in D) \quad con \quad \theta_0 \in \Theta_H$$

Definizione 2.2 (**Errore di seconda specie**). *Si chiama errore di seconda specie il non respingere l'ipotesi quando essa è falsa. Si compie un errore di seconda specie se:*

$$\theta_0 \notin \Theta_H \quad e \quad T(X) \notin D.$$

Per chiarire meglio queste definizioni, mostriamo graficamente (figura 2.1) la probabilità di compiere un errore di prima specie (che chiameremo α) e la probabilità di compiere un errore di seconda specie (che chiameremo β), nel caso in cui il problema sia quello di decidere se una variabile aleatoria $X \sim N(\mu, 1)$ ha distribuzione:

$$X \sim N(\mu_0, 1) \quad H_0 : \mu = \mu_0,$$

oppure:

$$X \sim N(\mu_A, 1) \quad H_A : \mu = \mu_A$$

Nelle applicazioni si ritiene generalmente più grave l'errore di prima specie. Torniamo all'esempio iniziale sul controllo dell'efficacia di un farmaco. In questo caso la sperimentazione viene svolta somministrando il farmaco a un gruppo di pazienti, a fronte di un altro gruppo trattato con placebo. L'ipotesi nulla corrisponde all' inefficacia del farmaco; in seguito all'osservazione si spera di rifiutarla, ossia di attribuire un beneficio terapeutico al farmaco sperimentale. *l'errore di prima specie* consiste in questo caso nell'attribuire al farmaco proprietà terapeutiche che esso in realtà non possiede. Questo risulta evidentemente più grave di quello di *seconda specie*, che consiste invece nel giudicare inefficace un farmaco che funziona (in questo caso si continuerà la sperimentazione, mentre nel primo caso si rischia di diffondere un farmaco inefficace e che potrebbe portare ad altre problematiche).

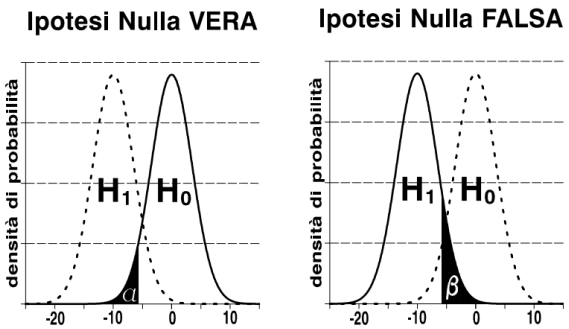


Figura 2.1: Errore di prima α e di seconda specie β

Lo scopo è quindi quello è quello di limitare il più possibile *l'errore di prima specie*. A questo proposito, si deve osservare che il valore "vero" del parametro θ_0 è sconosciuto: il suo valore è proprio ciò su cui si indaga; quindi la probabilità sopra indicata per l'errore di prima specie non è disponibile in modo esplicito. Per questa ragione serve definire un'altra grandezza per stimare almeno la probabilità dell'errore di prima specie, quello a cui si dedica maggiore attenzione; si tratta del *livello del test statistico*.

Definizione 2.3 (**Livello del test statistico**). *Si definisce livello di un test statistico con regione critica D il numero:*

$$\alpha = \sup_{\theta \in \Theta_H} P^{\theta}(T(X) \in D),$$

cioè l'estremo superiore delle probabilità dell'errore di prima specie.

Poiché si vorrebbe evitare l'errore di prima specie, è opportuno che α sia piccolo e i valori più comuni nella letteratura sono 10%, 5%, 1%. Una volta stabilito il livello del test e la statistica opportuna, si determina la regione di rigetto e quindi la validità o meno della ipotesi statistica formulata. inoltre per avere maggiori informazioni sulla *significatività del test* è bene riportare anche il valore di p.

Definizione 2.4 (**Livello di significatività osservato**). *Il livello di significatività del test (valore di p), relativo al valore sperimentale è la probabilità di ottenere, quando H_0 è vera, un valore della statistica del test uguale o maggiore (nel verso appropriato che porta ad H_A) di quello realmente calcolato. Il valore di p o p value, può essere anche definito come il valore più piccolo di α per cui l'ipotesi nulla può essere rifiutata:*

$$p = inf\left\{\alpha; \quad T_{oss} \in D_{\alpha}\right\}.$$

Una regola generale nei test d'ipotesi è che: *se il valore di p è minore o uguale di α, rifiutiamo l'ipotesi nulla. Se il valore di p è maggiore di α, non rifiutiamo l'ipotesi nulla.* Inoltre riportare i valori di p alla fine di un'indagine statistica, dà maggiori informazioni al ricercatore rispetto all'affermazione "l'ipotesi nulla è rifiutata ad un livello di significatività del 5%". infatti, se ad esempio il valore calcolato di p è 0,03 sappiamo che avremmo potuto scegliere un valore di α fino a 0,03 ed essere ancora in condizioni di rifiutare l'ipotesi nulla. Un altro concetto di rilevante importanza nei test statistici è la *potenza del test*:

Definizione 2.5. *La potenza di un test statistico è la probabilità che un test statistico ha di falsificare l'ipotesi nulla quando l'ipotesi nulla è effettivamente falsa. Quindi la* ***potenza*** *è definita come:*

$$potenza = 1 - \beta,$$

ovvero 1− *P(errore di seconda specie).*
In altre parole, la Potenza di un test è la sua capacità di cogliere delle differenze, quando queste differenze esistono.

I test d'ipotesi che si possono effettuare sono di diverso tipo, ma in questo capitolo ci concentreremo solo su alcuni di essi che saranno utili per entrare nel merito dello studio successivo.

2.2 Confronto tra medie di due popolazioni Gaussiane

Nel caso del confronto tra medie di popolazioni Gaussiane possiamo trovarci di fronte a due tipi di problemi diversi. Enunciamo due esempi per poi apprezzarne le differenze:

Problema 2.6 (**Problema** 1)**.** *Alcuni ricercatori sono interessati a conoscere se i dati da essi raccolti consentono di rilevare una differenza fra i livelli medi di acido urico sierico di soggetti normali e soggetti con sindrome di Down. I dati sono costituiti dai valori di acido urico sierico presi da* 12 *individui con sindrome di Down e da* 15 *individui normali. Le medie dei dati sono* $\bar{X} = 4{,}5\frac{mg}{100ml}$ *e* $\bar{Y} = 3{,}4\frac{mg}{100ml}$

Problema 2.7 (**Problema** 2)**.** *E' stata condotta una ricerca per studiare le funzioni della cistifellea in pazienti con reflusso esofageo , prima e dopo un intervento chirurgico per fermare il reflusso. Gli autori della ricerca hanno misurato la funzionalità della cistifellea (**GBEF**) prima e dopo l'intervento di fondoplicazione. L'obiettivo era di aumentare il **GBEF** con tale intervento. Si vuole sapere se dai dati raccolti si evincono informazioni sufficienti da poter affermare che l'intervento è risultato determinante nell'aumentare le funzioni **GBEF**.*

La domanda a cui si vuole rispondere in entrambi i problemi è: **si può concludere che ci sia una differenza tra i livelli medi di due popolazioni Gaussiane?**, ma con una sostanziale differenza:

- Nel primo problema le due popolazioni sono costituite da osservazioni (individui) diverse, casuali e tutte indipendenti tra loro;
- Nel secondo problema gli individui osservati sono gli stessi, *prima* e *dopo* un trattamento, pertanto i due campioni non sono indipendenti tra loro, ma vengono detti *campioni appaiati.*

2.2.1 Campioni indipendenti

Consideriamo due variabili aleatorie con distribuzione normale e delle osservazioni tutte indipendenti di tali variabili:

$$X \sim N(\mu_1, \sigma^2) \qquad X_1.....X_n \quad i.i.d$$

$$Y \sim N(\mu_2, \sigma^2) \qquad Y_1.....Y_m \quad i.i.d$$

Osserviamo che nel caso del nostro problema iniziale $n = 12$ e $m = 15$, quindi le X_i rappresentano le osservazioni di acido urico sierico sui pazienti con sindrome di Down, mentre le Y_i, sui pazienti normali. Per il trattamento di questo problema conviene supporre che le varianze delle due popolazioni siano le medesime(o quanto meno che abbiano lo stesso ordine di grandezza). Questa assunzione appare ragionevole in numerosi casi applicativi. L'ipotesi nulla che vogliamo testare è

$$H_0 : \mu_1 = \mu_2$$

mentre l'alternativa è

$$H_A : \mu_1 \neq \mu_2$$

Consideriamo il caso in cui la varianza non sia nota e pertanto il primo passo per costruire una statistica test è cercare un consultivo per quest'ultima. Per prima cosa ricordiamo che per il *Corollario del teorema di Cochran* 1.19:

$$\frac{1}{\sigma^2}\sum_{i=0}^{n}(X_i - \bar{X})^2 = \frac{n-1}{\sigma^2}S_x^2 \quad \sim \chi^2(n-1)$$

e

$$\frac{1}{\sigma^2}\sum_{i=0}^{m}(Y_i - \bar{Y})^2 = \frac{m-1}{\sigma^2}S_y^2 \quad \sim \chi^2(m-1)$$

Ora consideriamo

$$S_{tot}^2 = \frac{1}{n+m-2}\left(\sum_{i=0}^{n}(X_i - \bar{X})^2 + \sum_{i=0}^{m}(Y_i - \bar{Y})^2\right)$$

che è tale che, moltiplicando a destra e sinistra per σ^2:

$$\frac{n+m-2}{\sigma^2}S_{tot}^2 \quad \sim \chi^2(n+m-2)$$

Ora ricordiamo che una variabile aleatoria ha distribuzione t-student 1.16 se è data da :

$$T = \sqrt{r}\frac{X}{\sqrt{Y}} \quad con \quad X \sim N(0,1) \quad e \quad Y \sim \chi^2(r)$$

Poichè;

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2\right)$$

usando la variabile aleatoria standardizzata 1.13 e considerando la validità dell'ipotesi H_0 : $\mu_1 = \mu_2$ si ha:

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \quad \sim N(0,1)$$

da cui :

$$T(n+m-2) \quad \sim \sqrt{n+m-2}\frac{\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\frac{n+m-2}{\sigma^2}S_{tot}^2}}$$

e semplificando opportunamente si trova la statistica test distribuita come una t-student a $n+m-2$ gradi di libertà:

$$T = \frac{\bar{X} - \bar{Y}}{S_{tot}\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Una volta definita la statistica e impostato il livello del test α, non resta che stabilire la regola di decisione e quindi la regione di rigetto D dell'ipotesi nulla. Poichè l'ipotesi nulla è $\mu_1 = \mu_2$ e gli stimatori corretti della media sono $\bar{X}$ e $\bar{Y}$, risulta naturale rifiutare l'ipotesi quando $T = \frac{\bar{X}-\bar{Y}}{S_{tot}\sqrt{\frac{1}{n}+\frac{1}{m}}}$ assume valori sufficientemente lontani dallo zero, cioè quando:

$$T \in]-\infty, -t_{1-\frac{\alpha}{2}}[\quad \cup \quad]t_{1-\frac{\alpha}{2}}, \infty[$$

con

$$P(-t_{1-\frac{\alpha}{2}} < T < t_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Come si vede in figura 2.2.
Calcolare il *livello di significatività del test* significa poi, imporre l'uguaglianza:

$$T_{oss} = t_{1-\frac{p}{2}},$$

dove p è l'incognita, e quindi bisogna calcolare:

$$1 - \frac{p}{2} = F_T\left(T_{oss}\right),$$

con $F_T(x)$ funzione di ripartizione della variabile aleatoria T, i cui valori si ottengono dalle opportune tavole della distribuzione t-student, in alcuni casi operando con interpolazione lineare. Osserviamo che, nel caso in cui la varianza sia nota, il problema che si pone è ancora

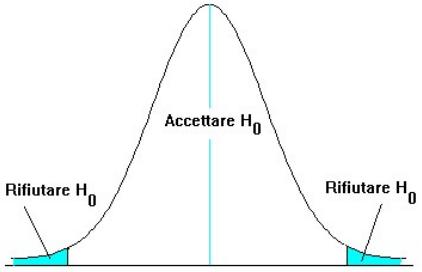


Figura 2.2: Test t-student bilaterale

più elementare in quanto la statistica è semplicemente data da:

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \quad \sim N(0,1).$$

2.2.2 Campioni appaiati

Un metodo frequentemente usato per accettare l'efficacia di un trattamento nelle procedure sperimentali è quello di fare uso di osservazioni legate tra di loro e provenienti da campioni non indipendenti. Lo scopo dei test appaiati è quello di ridurre al minimo la variabilità che può intervenire da fonti esterne. Pensiamo al nostro esempio iniziale in cui uno stesso campione di pazienti viene osservato prima e dopo un trattamento, per capire l'efficacia di quest'ultimo. Se invece avessimo osservato due campioni di pazienti diversi, uno trattato con l'intervento e l'altro con placebo, sarebbe stato più difficile raggiungere una conclusione, perchè ad esempio il trattamento sarebbe potuto sembrare efficace sul gruppo di pazienti scelto, ma unicamente per il fatto che questi avevano delle caratteristiche iniziali migliori dei pazienti trattati con placebo.

Consideriamo due variabili aleatorie distribuite normalmente:

$$X \sim N(\mu_1, \sigma^2) \qquad X_1.....X_n \quad i.i.d$$

$$Y \sim N(\mu_2, \sigma^2) \qquad Y_1.....Y_n \quad i.i.d$$

che rappresentato cosa accade *prima* e *dopo* il trattamento sugli stessi pazienti, infatti il numero di osservazioni è lo stesso per entrambe.

Vogliamo ancora testare l'ipotesi nulla:

$$H_0 : \mu_1 = \mu_2$$

Costruiamo la statistica test opportuna e quindi la regola di decisione. Dato che i campioni sono appaiati è naturale considerare gli:

$$Z_i = X_i - Y_i \quad per \quad i = 1....n$$

ovvero la differenza tra il valore osservato prima e dopo il trattamento sullo stesso individuo. Le Z_i sono distribuite normalmente, in quanto differenze tra osservazioni di variabili aleatorie normali, pertanto è possibile riferirsi al *Corollario del teorema di Cochran* 1.19 per avere la statistica:

$$T = \sqrt{n}\frac{\bar{Z}}{S_z} \quad \sim t(n-1),$$

assumendo la validità dell'ipotesi nulla e definendo $S_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$, lo stimatore corretto per la varianza di Z. A questo punto il ragionamento è esattamente analogo al precedente e la regione di rigetto dell'ipotesi nulla la medesima:

$$D =]-\infty, -t_{1-\frac{\alpha}{2}}[\quad \cup \quad]t_{1-\frac{\alpha}{2}}, \infty[.$$

E' bene osservare che in questo caso sarebbe più interessante testare una ipotesi unilaterale del tipo:

$$H_0 : \mu_1 > \mu_2$$

che corrisponde al chiedersi se l'intervento ha effettivamente avuto successo, nel nostro caso, se la cistifellea ha aumentato il suo funzionamento e quindi vorremmo poter rifiutare l'ipotesi H_0. Nulla cambia nel problema, se non la regione di rigetto che in questo caso sarà:

$$D =]-\infty, -t_{1-\alpha}[$$

come si vede in figura 2.3, poichè in questo caso si rigetta l'ipotesi quando $\mu_1 - \mu_2 < k$.
Si può sempre ragionare in questo modo, sia su ipotesi bilaterali che unilaterali, facendo attenzione a considerare l'opportuna regione di rigetto.

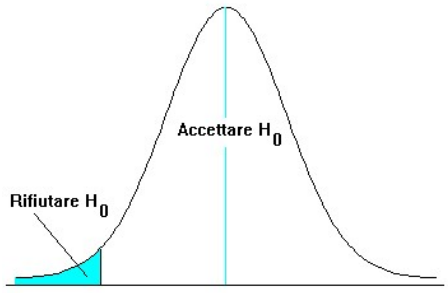


Figura 2.3: Test t-student unilaterale

2.3 Confronto tra varianze di due popolazioni Gaussiane

Fino ad ora abbiamo supposto che le varianze delle popolazioni Gaussiane fossero pressochè uguali e calcolabili attraverso gli opportuni stimatori dati dal campione. Potrebbe succedere, però, che due metodi di analisi chimica forniscano gli stessi risultati in media, ma che i risultati prodotti con un metodo siano più variabili di quelli prodotti con un altro. Vogliamo quindi trovare un modo per determinare la probabilità che ciò sia vero. Questa è una verifica molto importante perché, come abbiamo visto, anche per svolgere i test riguardanti il confronto tra le medie è spesso necessario potere supporre che le varianze siano uguali. Un esempio di problema che si può porre è il seguente:

Problema 2.8. *Un test per misurare il livello di ansia è stato somministrato ad un campione di maschi e ad un campione di femmine, poco prima di sottoporsi alla stessa operazione chirurgica. Le dimensioni campionarie e le varianze calcolate dai valori sono:*

$$Maschi \quad n = 16 \quad S_x^2 = 150$$

$$Femmine \quad m = 21 \quad S_y^2 = 275$$

Questi risultati forniscono degli elementi per concludere che nella popolazione rappresentante i valori ottenuti nelle donne sono più variabili di quelli negli uomini?

Questo tipo di problema si affronta in generale nel modo seguente:
consideriamo due variabili gaussiane indipendenti:

$$X \sim N(\mu_1, \sigma_1^2) \qquad X_1.....X_n \quad i.i.d$$

e

$$Y \sim N(\mu_2, \sigma_2^2) \qquad Y_1.....Y_m \quad i.i.d,$$

l'ipotesi nulla da sottoporre a verifica statistica è:

$$H_0 : \sigma_1 = \sigma_2,$$

contro l'ipotesi alternativa:

$$H_A : \sigma_1 \neq \sigma_2.$$

Consideriamo gli stimatori S_x^2 e S_y^2 e ricordiamo che per il *corollario del teorema di Cochran* 1.19;

$$W_1 = \frac{n-1}{\sigma^2} S_x^2 \quad \sim \chi^2(n-1) \quad e \quad W_2 = \frac{m-1}{\sigma_2^2} S_y^2 \quad \sim \chi^2(m-1).$$

Ora una una variabile aleatoria ha distribuzione di *Fisher* 1.17 con (n, m) gradi di libertà se è del tipo:

$$F = \frac{Z_n}{Z_m} \cdot \frac{m}{n},$$

con $Z_n \sim \chi^2(n)$ e $Z_m \sim \chi^2(m)$ e indipendenti tra loro.
Quindi:

$$\frac{W_1}{W_2} \cdot \frac{m-1}{n-1} = \frac{S_x^2}{\sigma_1^2} : \frac{S_y^2}{\sigma_2^2} \sim F(n-1, m-1)$$

In particolare se l'ipotesi nulla è vera $H_0 : \sigma_1 = \sigma_2$ e pertanto si ha come statistica:

$$F = \frac{S_x^2}{S_y^2} \quad \sim F(n-1, m-1)$$

In questo caso si accetta l'ipotesi (considerando sempre un test a livello α) quando F è non lontano dal valore 1 e si rifiuta altrimenti,

$$P(F_{\frac{\alpha}{2}} < F < F_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Quindi la regione di rigetto dell'ipotesi è:

$$R = \mathbb{R}^+ - [F_{\frac{\alpha}{2}}, F_{1-\frac{\alpha}{2}}]$$

come si vede in figura 2.4.
Osserviamo che in questo problema la regione di rigetto comprende sia valori sufficientemente più piccoli di uno, e quindi vicini allo zero (in quanto la densità assume solo valori positivi), che valori sufficientemente maggiori di uno, poichè l'ipotesi viene accettata quando i due stimatori delle varianze assumono circa lo stesso valore e quindi solo quando il loro rapporto è un valore prossimo a uno.

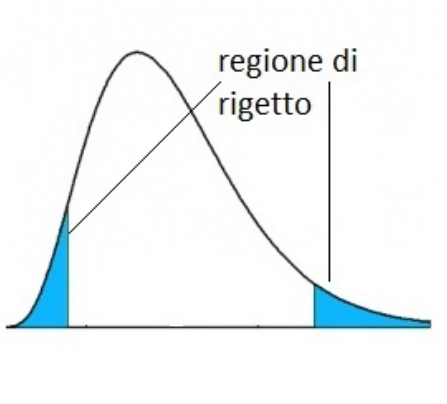


Figura 2.4: Test di Fisher bilaterale

Anche in questo caso il tipo di problema che si può presentare è quello di saggiare l'ipotesi unilaterale:

$$H_0 : \sigma_1^2 < \sigma_2^2,$$

in questo caso non cambierà nulla se non la regione di rigetto:

$$D =]F_{1-\alpha}, \infty[$$

Capitolo 3

Analisi della varianza di tipo parametrico

Lo scopo di questo capitolo è trovare dei metodi per saggiare l'ipotesi di uguaglianza tra le medie di un certo numero di popolazioni Gaussiane. Nel capitolo precedente abbiamo visto come saggiare l'ipotesi tra due popolazioni utilizzando il test t-student, ora la domanda è:

- se dobbiamo confrontare 5 popolazioni Gaussiane come ci comportiamo?.

La prima idea potrebbe essere quella di usare un test t-student per ogni coppia di popolazioni, ma questo non sarebbe efficace, vediamo perchè:

1. per prima cosa osserviamo che dovremmo effettuare un test t per ogni coppia di medie e quindi nel nostro caso: $\binom{5}{2} = 10$ test, più in generale se le popolazioni da confrontare sono k dovremmo effettuare $\binom{k}{2} = \frac{k(k-1)}{2}$ test (il che potrebbe diventare estremamente lungo).

2. ora se pensiamo di fissare un livello $\alpha = 0{,}05$ per ogni test t che eseguiamo, osserviamo che la probabilità di compiere $r = 0$ errori su tutta la serie di test è:

   $$P(r = 0) = (0{,}95)^{10} = 0{,}5987,$$

   e quindi la probabilità di compiere almeno un errore su tutta la serie di test, ovvero la probabilità di rifiutare l'ipotesi di $k = 5$ medie uguali, quando è vera è:

   $$1 - (0{,}95)^{10} = 0{,}4013,$$

   il che implicherebbe un errore di prima specie sull'esperimento globale superiore al 40%, che non può essere accettabile.

3. Inoltre bisogna tenere presente che il problema sarebbe ancora più complicato poichè non è detto che tre o più test t-student basati sugli stessi dati siano indipendenti l'uno dall'altro.

Pertanto è chiaro che bisogna delineare altri metodi per saggiare la differenza tra le medie di più di due campioni.

3.1 L'analisi della varianza ANOVA ad una via

L'analisi della varianza ad una via (in inglese: Analysis of variance, abbreviata con l'acronimo ANOVA) è una semplice estensione a tre o più campioni del test t-student applicato a due popolazioni Gaussiane. Questo tipo di analisi è utilizzata per testare le differenze tra medie di più popolazioni e per fare questo si prendono in considerazione le rispettive varianze. Il principio alla base di questo test è quello di stabilire se due o più medie campionarie possono derivare da popolazioni che hanno la stessa media parametrica o più brevemente se si può concludere che i dati analizzati provengano dalla stessa popolazione. Viene chiamato test ad una via perchè l'analisi viene fatta rispetto ad una sola fonte di variazione o *fattore* con k diversi livelli del fattore o più comunemente chiamati **trattamenti**.
La situazione tipica dell'analisi della varianza ad una via, è data quando si vuole saggiare l'ipotesi nulla che tre o più trattamenti danno luogo allo stesso risultato. L'unica variabile in gioco è quindi il trattamento.
A seconda del disegno sperimentale del problema che si vuole trattare si fa un uso diverso della stessa tecnica della analisi della varianza ad una via. La maggior parte degli errori nella letteratura biomedica riguarda errori basilari nel disegno sperimentale, come per esempio una errata procedura di randomizzazione o l'omissione del gruppo di controllo, oppure l'utilizzo non appropriato del test t per confronti multipli. Pertanto è bene fare attenzione alle diverse possibilità e capire quali sono le più adatte nei vari casi.

3.1.1 Disegno sperimentale completamente randomizzato

Si parla di *disegno sperimentale completamente randomizzato* quando ai soggetti in studio vengono assegnati i trattamenti in modo totalmente casuale.
Quindi, se ad esempio vogliamo confrontare quattro farmaci (A,B,C,D) e abbiamo a disposizione 16 pazienti, dobbiamo somministrare il farmaco A a un numero n_1 di pazienti scelti in modo casuale dai 16 disponibili, il farmaco B a un n_2 di pazienti scelti in modo casuale dai 16 disponibili e così per i farmaci C (a n_3 pazienti) e D (a n_4 pazienti), in modo che $n_1 + n_2 + n_3 + n_4 = 16$, dove gli n_j in generale non sono uguali tra loro. Lo scopo del test è quello di poter dire se, in base ai dati ottenuti, i farmaci hanno tutti lo stesso effetto, ovvero se le medie delle rispettive popolazioni sono tutte uguali tra loro.
Questo disegno sperimentale è il più semplice, ma è conveniente solo quando il materiale utilizzato è altamente omogeneo. Ad esempio, in un esperimento di laboratorio per valutare l'effetto di k farmaci somministrati a N cavie, per ottenere la maggior potenza del test si richiede che esse siano tutte dello stesso ceppo (quindi che abbiano gli stessi genitori), abbiano la stessa età (quindi siano della stessa nidiata), lo stesso peso, il medesimo sesso e in generale siano identiche per tutti quei fattori che si ritiene influenzino il valore che verrà misurato. Solamente in queste condizioni è credibile che:

- le differenze tra le medie siano imputabili solamente ai differenti effetti dei farmaci,
- alla fine dell'esperimento la varianza d'errore sarà minima.

Altrimenti si devono utilizzare altri disegni sperimentali che verranno trattati in seguito, in questo capitolo.

Costruiamo un modello generale facendo alcune ipotesi restrittive. In queste ipotesi, da un punto di vista inferenziale, si usa parlare di *modello ad effetti fissi*.

- Rappresentiamo le osservazioni e le medie campionarie delle k popolazioni nel seguente modo:

Trattamento	1	2	...	k
	x_{11}	x_{12}	...	x_{1k}
	x_{21}	x_{22}	...	x_{2k}
	x_{31}	x_{32}	...	x_{3k}
	...	...	...	...
	...	...	...	...
	...	...	...	...
	...	...	...	...
	$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_k k}$
Totale	$T_{.1}$	$T_{.2}$	...	$T_{.k}$
Media	$\bar{X}_{.1}$	$\bar{X}_{.2}$	...	$\bar{X}_{.k}$

Dove:

x_{ij}= è la i-esima osservazione del j-esimo trattamento con $i = 1.....n_j$ e $j = 1.....k$,

$T_{.j} = \sum_{i=1}^{n_j} x_{ij}$ è il totale del j-esimo trattamento,

$\bar{X}_{.j} = \frac{T_{.j}}{n_j}$ è la media del j-esimo trattamento,

$T_{..} = \sum_{j=1}^{k} T_{.j} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} T_{.j}$ è il totale di tutte le osservazioni,

$\bar{X}_{..} = \frac{T_{..}}{N}$ con $N = \sum_{j=1}^{k} n_j$.

- Scomponiamo una qualunque osservazione del nostro insieme dei dati nel modo seguente:

$$x_{ij} = \mu + \tau_j + e_{ij} \quad per \quad i = 1...n_j \quad e \quad j = 1....k,$$

i termini sono definiti in questo modo:

1. $\mu = \sum_{j=1}^{k} \mu_j$ è la *media generale*, ovvero la media di tutte le k medie delle popolazioni,
2. $\tau_j = \mu - \mu_j$ è l'*effetto del trattamento*, ovvero la differenza tra la media della popolazione j-esima e la media generale, nel modello ad effetti fissi τ_j sono delle costanti e non delle variabili aleatorie,
3. $e_{ij} = x_{ij} - \mu_j$ è il *termine errore*, cioè l'ammontare per cui una misurazione individuale x_{ij} differisce dalla media della popolazione a cui appartiene.

In questo modo si ha:

$$x_{ij} = \mu + (\mu_j - \mu) + (x_{ij} - \mu_j)$$

E' bene notare che, ciò che è di nostro interesse è *l'effetto del trattamento*, in poche parole affinchè tutti i trattamenti abbiano lo stesso effetto, ovvero tutte le popolazioni abbiano la stessa media si vuole $\tau_j = 0$:

$$x_{ij} = \mu + e_{ij},$$

cioè:

$$x_{ij} = \mu + (x_{ij} - \mu), \quad con \quad \mu = \frac{1}{k} \sum_{j=1}^{k} \mu, \quad ovvero \quad \mu = \mu_j \quad \forall j = 1....k$$

- Le assunzioni per il modello sono:
  - i k *trattamenti* costituiscono k campioni indipendenti, ognuno dei quali proviene da una specifica popolazione distribuita normalmente:

$$X_j \sim N(\mu_j, \sigma_j^2) \quad per \quad j = 1...k,$$

    notiamo che nell'esempio dei farmaci $k = 4$.
  - Le varianze sono tutte uguali:

$$\sigma_1^2 = \sigma_2^2 = ...\sigma_k^2 = \sigma^2.$$

  - $\sum_{j=1}^{k} \tau_j = 0$;
  - $E[e_{ij}] = 0$, cioè le variabili aleatorie e_{ij} hanno media zero, infatti $E[x_{ij}] = \mu_j$.
  - $var(e_{ij}) = var(x_{ij})$, dato che le e_{ij}, come variabili aleatorie, differiscono dalle x_{ij} solo di una costante.
  - Le e_{ij} sono distribuite normalmente e indipendentemente.

- Facciamo l'ipotesi nulla che tutti i trattamenti medi, ovvero tutte le popolazioni sono uguali:

$$H_0 : \quad \mu_1 = \mu_2.... = \mu_k$$

  mentre l'alternativa è:

$$H_A : \quad non \quad tutte \quad le \quad \mu_j \quad sono \quad uguali.$$

  Inoltre se tutte le medie sono uguali, ogni effetto del trattamento è uguale a zero, quindi le ipotesi posso anche essere riscritte nel modo seguente:

$$H_0 : \quad \tau_j = 0 \quad per \quad j = 1...k$$

  mentre l'alternativa è:

$$H_A : \quad non \quad tutte \quad le \quad \tau_j \quad sono \quad uguali \quad a \quad zero.$$

  Se le popolazioni sono tutte distribuite normalmente, con varianze uguali e H_0 è vera, allora le distribuzioni sono identiche e i grafici si sovrappongono tutti uno sull'altro, come si vede in figura 3.1.

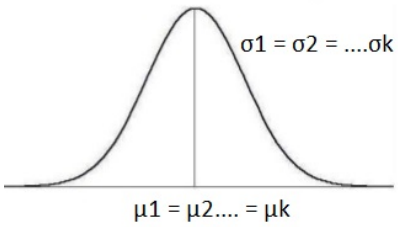


Figura 3.1: Grafico delle popolazioni quando H_0 è vera

Quando invece H_0 è falsa, o alcune delle popolazioni hanno media diversa dalle altre, oppure sono tutte diverse tra loro, come in figura 3.2.

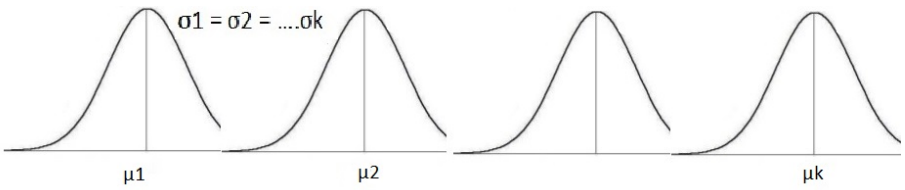


Figura 3.2: Grafico delle popolazioni quando H_0 è falsa

- Costruiamo una opportuna *statistica test*:
  supponiamo vera l'ipotesi H_0 e costruiamo due stimatori per la varianza σ^2, per fare ciò definiamo:

  1. **la somma dei quadrati all'interno dei gruppi**, che corrisponde alla somma delle deviazioni al quadrato di ogni osservazione dalla propria media, questa quantità è talvolta chiamata anche *somma dei quadrati degli errori* ed è abbreviata con l'acronimo **SSW**:
  $$SSW = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{X}_{.j})^2,$$
  questa quantità fornisce il numeratore della **prima stima di** σ^2 infatti:
  $$\frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{X}_{.j})^2}{n_j - 1}$$
  fornisce uno stimatore corretto della varianza della popolazione alla quale il campione appartiene (1.11). Nell'ipotesi che le varianze delle popolazioni siano le stesse, possiamo aggregare le k stime per ottenere:
  $$MSW = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{X}_{.j})^2}{\sum_{j=1}^{k} n_j - 1},$$
  questa è la prima stima di σ^2 che può essere chiamata **varianza all'interno dei gruppi** o in inglese: **varianza within**, in letteratura abbreviata con l'acronimo MSW.
  2. **La somma dei quadrati tra i gruppi** che è la somma delle deviazioni al quadrato della media del gruppo dalla media generale moltiplicata per la dimensione del gruppo, questa quantità è anche denominata **SSA**:
  $$SSA = \sum_{j=1}^{k} n_j (\bar{X}_{.j} - \bar{X}_{..})^2,$$
  questa quantità, nell'ipotesi che tutte le osservazioni x_{ij} provengano dalla stessa popolazione, fornisce il numeratore della **seconda stima di** σ^2, detta anche **varianza tra i gruppi**, in inglese **varianza between**, in letteratura denominata con l'acronimo MSA. Per ricavarla ricordiamo che in generale:
  $$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} n\sigma^2$$

otteniamo:

$$\sigma^2 = n\sigma^2_{\bar{x}}.$$

Ora per $\sigma^2_{\bar{x}_{..}}$ nell'ipotesi H_0 si ha:

$$Var(\bar{X}_{..}) = Var\left(\frac{1}{\sum_{j=1}^{k} n_j}\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}\right) = \frac{1}{\left(\sum_{j=1}^{k} n_j\right)^2}\sum_{j=1}^{k} n_j\sigma^2,$$

allora:

$$Var(\bar{X}_{..}) = \frac{\sigma^2}{\sum_{j=1}^{k} n_j}$$

e dato che una stima corretta per $Var(\bar{X}_{..})$ è:

$$\frac{\sum_{j=1}^{k}(\bar{X}_{.j} - \bar{X}_{..})^2}{k-1},$$

si ottiene che uno stimatore corretto per σ^2 è:

$$MSA = \frac{1}{k-1}\sum_{j=1}^{k} n_j(\bar{X}_{.j} - \bar{X}_{..})^2.$$

Osserviamo che i due stimatori trovati per σ^2 nell'ipotesi H_0 corrispondono alla scomposizione dello stimatore canonico di σ^2:

$$S^2 = \frac{1}{N-1}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{X}_{..})^2,$$

dove N corrisponde alla dimensione della totalità del campione, come se le osservazioni provenissero tutte dalla medesima popolazione, ovvero $N = \sum_{j=1}^{k} n_j$. Infatti, chiamando SST il numeratore di S^2 e quindi **la somma totale dei quadrati** (la somma dei quadrati delle deviazioni delle singole osservazioni dalla media generale su tutte le osservazioni prese nell'insieme), ovvero:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{X}_{..})^2$$

si può facilmente dimostrare che:

$$SST = SSW + SSA$$

cioè:

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{X}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{k} n_j(\bar{X}_{.j} - \bar{X}_{..})^2.$$

Per dimostrare l'uguaglianza svolgiamo i calcoli al secondo membro:

$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{k} n_j(\bar{X}_{.j} - \bar{X}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}^2 + \bar{X}_{.j}^2 - 2x_{ij}\bar{X}_{.j}) + \sum_{j=1}^{k} n_j(\bar{X}_{.j}^2 + \bar{X}_{..}^2 - 2\bar{X}_{.j}\bar{X}_{..})$$

$$= \sum_{j=1}^{k} \left(\sum_{i=1}^{n_j} x_{ij}^2 + \underline{n_j \bar{X}_{\cdot j}^2} - \underline{2n_j \bar{X}_{\cdot j}^2} \right) + \sum_{j=1}^{k} \left(\underline{n_j \bar{X}_{\cdot j}^2} + n_j \bar{X}_{\cdot\cdot}^2 - 2n_j \bar{X}_{\cdot j} \bar{X}_{\cdot\cdot} \right)$$

$$= \sum_{j=1}^{k} \left(\sum_{i=1}^{n_j} x_{ij}^2 + n_j \bar{X}_{\cdot\cdot}^2 - 2n_j \bar{X}_{\cdot j} \bar{X}_{\cdot\cdot} \right)$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij}^2 + \bar{X}_{\cdot\cdot}^2 - 2x_{ij} \bar{X}_{\cdot\cdot} \right) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{X}_{\cdot\cdot})^2$$

Da cui la tesi.
Dobbiamo renderci conto (questo è il punto fondamentale della analisi), come abbiamo già osservato, che la *varianza tra i gruppi* (MSA) fornisce una stima corretta per σ^2, solo quando H_0 è vera (oltre all'ipotesi sulla uguaglianza tra le varianze delle popolazioni) e quindi in questa ipotesi dovremmo aspettarci che le due stime per σ^2 risultino abbastanza vicine, in caso di ipotesi nulla H_0 falsa invece, dovremmo aspettarci che la *varianza tra i gruppi* (MSA) risulti più grande della *varianza all'interno dei gruppi* (MSW) in quanto viene calcolata attraverso le deviazioni al quadrato delle medie campionarie dalla media totale.
A questo punto è chiaro che saggiare l'ipotesi H_0 si riduce a testare il rapporto tra le due stime di σ^2 e quindi la statistica da utilizzare è il test di Fisher (utilizzato nella sezione: **Confronto tra varianze di due popolazioni Gaussiane**).
Costruiamo il test di Fisher. Per il corollario del teorema di Cochran (1.19) abbiamo che:

$$\frac{S^2(N-1)}{\sigma^2} \sim \chi^2(N-1),$$

inoltre è facile dimostrare che:

$$\frac{MSW(N-k)}{\sigma^2} \sim \chi^2(N-k),$$

infatti:

$$MSW = \frac{1}{N-k} \left(\sum_{i=1}^{n_1} (x_{i1} - \bar{X}_{\cdot 1})^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{X}_{\cdot 2})^2 + ... \sum_{i=1}^{n_k} (x_{ik} - \bar{X}_{\cdot k})^2 \right).$$

Allora moltiplicando per $N-k$ e dividendo per σ^2 otteniamo:

$$\frac{MSW(N-k)}{\sigma^2} = \frac{S_1^2(n_1-1)}{\sigma^2} + \frac{S_2^2(n_2-1)}{\sigma^2} + ... \frac{S_k^2(n_k-1)}{\sigma^2}$$

quindi usando il corollario del teorema di Cochran (1.19) e l'additività delle χ^2 (1.2) abbiamo:

$$\frac{MSW(N-k)}{\sigma^2} = \chi^2(n_1-1) + \chi^2(n_2-1) + ... \chi^2(n_k-1) \sim \chi^2(N-k).$$

Ora, come abbiamo già osservato:

$$SST = SSW + SSA,$$

quindi:

$$S^2(N-1) = MSW(N-k) + MSA(k-1),$$

e dividendo per σ^2:

$$\frac{S^2(N-1)}{\sigma^2} = \frac{MSW(N-k)}{\sigma^2} + \frac{MSA(k-1)}{\sigma^2},$$

da cui:

$$\frac{MSA(k-1)}{\sigma^2} \sim \chi^2(k-1).$$

Pertanto ricordando che una variabile aleatoria ha distribuzione di Fisher (1.17) se è il rapporto di due χ^2 moltiplicate per i rispettivi gradi di libertà abbiamo che:

$$T = \frac{\frac{MSA(k-1)}{\sigma^2}}{\frac{MSW(N-k)}{\sigma^2}} \cdot \frac{(N-k)}{(k-1)},$$

da cui:

$$T = \frac{MSA}{MSW} \sim F(k-1, N-k),$$

La statistica T ha distribuzione di Fisher con $k-1$ gradi di libertà al numeratore e $\sum_{j=1}^{k}(n_j - 1) = N - k$ al denominatore.

- Stabiliamo in fine la regola di decisione e riportiamo il livello di significatività del test p. La regola di decisione corrisponderà a rifiutare l'ipotesi H_0 quando il valore sperimentale di T assumerà valore nella regione di rigetto illustrata in figura 3.3:

$$D =]F_{1-\alpha}, +\infty[\quad e \quad p \quad t.c \quad T_{oss} = F_{1-p}$$

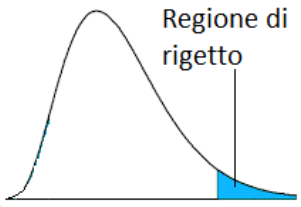


Figura 3.3: Regione di rigetto di H_0

ovvero quando il valore di T è sufficientemente maggiore di uno, in modo da poter attribuire la differenza osservata tra le due stime di σ^2 a qualcosa di diverso dalla casualità del campionamento. E' lecito chiedersi:

– Siamo sicuri che rifiutare l'ipotesi sulle varianze significa rifiutare l'ipotesi di medie uguali nelle popolazioni?

La risposta è sì, in quanto un valore elevato di T deriva dal fatto che *la varianza tra i gruppi* (MSA) è considerevolmente più grande della *varianza all'interno dei gruppi* (MSW) e poichè la varianza between (calcolata come stimatore di σ^2 supponendo vera l'ipotesi sulla uguaglianza tra le medie) è basata sulla dispersione delle medie campionarie rispetto alla loro media, questa quantità sarà grande quando c'è una grande discrepanza tra le dimensioni delle medie campionarie, da cui la tesi.

Riportiamo ora un esempio di problema, svolto utilizzando il software Mathematica.

Problema 3.1. *Alcuni tipi di carne, come quella di cervo e di scoiattolo grigio, sono mangiati dalle famiglie, per motivi di salute, culturali o personali. Uno studio condotto da David Holben ha valutato il contenuto di selenio nella carne di cervo (in inglese Vension) e di scoiattolo grigio (in inglese Squirrel) in una regione degli Stati Uniti con bassi livelli di selenio. Tali valori sono stati comparati con quelli presenti nella carne di manzo (in inglese Beef) prodotta dentro e al di fuori della stessa regione. Vogliamo sapere se i livelli di selenio sono differenti nei quattro tipi di carne.*

Utilizzando la funzione ANOVA in Mathematica e riportando i dati in modo corretto è possibile ottenere i risultati necessari per rispondere al problema.
Nel problema seguente abbiamo $k = 4$ livelli del trattamento, che corrispondono ai tipi di carne, $n_{tot} = 144$ osservazioni e fissiamo un livello $\alpha = 0{,}01$.
Le osservazioni vanno riportate in una lista di liste, dove il primo termine corrisponde al livello del trattamento. Abbiamo, quindi riportato i dati relativi alle osservazioni sul contenuto di selenio in $\frac{\mu g}{100gr}$ nei quattro differenti tipi di carne ,(cervo=1, scoiattolo=2, manzo regionale=3, manzo non regionale=4).
Inoltre abbiamo chiesto al programma di riportare il valore esatto del quantile $F_{1-\alpha}(k-1, n_{tot}-k)$, che è il valore con cui dobbiamo confrontare il valore di $T = \frac{MSA}{MSW}$ sperimentale, che in Mathematica è chiamato $FRatio$.
Utilizzando poi il comando ANOVA sulla lista di dati, Mathematica calcola:

- i gradi di libertà DF;
- la somma dei quadrati $SumOfSq$ (SSA, SSW e SST)
- la media quadratica $MeanSq$ (MSA e MSW);
- il rapporto di varianze $FRatio$;
- il valore di p;
- la media campionaria totale e su ogni livello del trattamento, $CellMeans$.

Il trattamento è chiamato $Model$ e i livelli sono indicati inserendo una parentesi quadra che riporta il livello corrispondente.
Riportiamo ora sia l'input che l'output di Mathematica relativo al seguente problema:

Needs["ANOVA"]

dataset1 = {{1, 26.72}, {1, 28.52}, {1, 29.71}, {1, 26.95}, {1, 10.97}, {1, 21.97}, {1, 14.35}, {1, 32.21},
{1, 19.19}, {1, 30.92}, {1, 10.42}, {1, 35.49}, {1, 36.84}, {1, 25.03}, {1, 33.59}, {1, 14.86},
{1, 16.47}, {1, 25.19}, {1, 37.45}, {1, 45.08}, {1, 25.22}, {1, 22.11}, {1, 33.01}, {1, 31.20},
{1, 26.50}, {1, 32.77}, {1, 8.70}, {1, 25.90}, {1, 29.80}, {1, 37.63}, {1, 33.74}, {1, 18.02},
{1, 22.27}, {1, 26.10}, {1, 20.89}, {1, 29.44}, {1, 21.69}, {1, 21.49}, {1, 18.11}, {1, 31.50},
{1, 27.36}, {1, 21.33}, {2, 37.42}, {2, 56.46}, {2, 51.91}, {2, 62.73}, {2, 4.55}, {2, 39.17},

{2, 38.44}, {2, 40.92}, {2, 58.93}, {2, 61.88}, {2, 49.54}, {2, 64.35}, {2, 82.49}, {2, 38.54},
{2, 39.53}, {2, 37.57}, {2, 25.71}, {2, 23.97}, {2, 13.82}, {2, 42.21}, {2, 35.88}, {2, 10.54},
{2, 27.97}, {2, 41.89}, {2, 23.94}, {2, 49.81}, {2, 30.71}, {2, 50.00}, {2, 87.50}, {2, 68.99},
{3, 11.23}, {3, 29.63}, {3, 20.42}, {3, 10.12}, {3, 39.91}, {3, 32.66}, {3, 38.38}, {3, 36.21},
{3, 16.39}, {3, 27.44}, {3, 17.29}, {3, 56.20}, {3, 28.94}, {3, 20.11}, {3, 25.35}, {3, 15.82},
{3, 27.74}{3, 22.35}, {3, 34.78}, {3, 35.09}, {3, 32.60}, {3, 37.03}, {3, 27.00}, {3, 44.20},
{3, 13.09}, {3, 33.03}, {3, 9.69}, {3, 32.45}, {3, 37.38}, {3, 34.91}, {3, 21.77}, {3, 31.62},
{3, 32.63}, {3, 30.31}, {3, 46.16}, {3, 56.61}, {3, 24.47}, {3, 29.39}, {3, 40.71}, {3, 18.52},
{3, 27.80}, {3, 19.49}, {3, 27.99}, {3, 22.36}, {3, 22.68}, {3, 26.52}, {3, 46.01}, {3, 38.04},
{3, 30.88}, {3, 30.04}, {3, 25.91}, {3, 18.54}, {3, 25.51}, {4, 44.33}, {4, 76.86}, {4, 4.45},
{4, 55.01}, {4, 58.21}, {4, 74.72}, {4, 11.82}, {4, 139.09}, {4, 69.01}, {4, 94.61}, {4, 48.35},
{4, 37.65}, {4, 66.36}, {4, 72.48}, {4, 87.09}, {4, 26.34}, {4, 71.24}, {4, 90.38}, {4, 50.86}};

$k = 4$;

ntot = 144;

alfa = 0.01;

quant = Quantile[FRatioDistribution[$k - 1$, ntot − k], 1 − alfa];

Print[“La soglia in valore assoluto della regione critica per T=FRatio è”];

Print[quant]

ANOVA[dataset1]

La soglia in valore assoluto della regione critica per T=FRatio è

3.92462

ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
Model	3	21261.9	7087.3	27.0001	7.692037481713481*^-14
Error	140	36748.9	262.492		
Total	143	58010.8			

CellMeans →

All	35.4468
Model[1]	25.874
Model[2]	43.2457
Model[3]	29.083
Model[4]	62.0453

Analizziamo ora i risultati:

1. supponendo $\alpha = 0{,}01$ il valore critico è 3.92462, la regola di decisione è quindi quella di rifiutare H_0 quando il rapporto tra varianze $FRatio$ rilascia un valore maggiore o uguale di questo,

2. il valore di $FRatio$ calcolato da Mathematica è $27,00$, che risulta maggiore del valore critico, quindi rifiutiamo H_0,

3. poichè abbiamo rifiutato H_0 concludiamo che l'ipotesi alternativa è vera e quindi che i quattro tipi di carne non hanno tutti il medesimo contenuto di selenio,

4. poichè $27 > 3.92462$ si ha che $p < 0{,}01$ e il suo valore è riportato in tabella.

Osservazione 3.2 (**Attenzione al termine *trattamento***). I *trattamenti* non sempre vanno intesi nel senso comune del termine. Infatti il termine "trattamento" usato nei disegni sperimentali è piuttosto generico. Ad esempio, potremmo essere interessati a studiare la risposta allo stesso trattamento (nel senso comune del termine) di diverse razze animali. In tal caso i *trattamenti* del disegno sperimentale sono le razze animali.

Osservazione 3.3 (**il test t-student è un'analisi della varianza**). Si può facilmente osservare che il test $t-student$ utilizzato nel capitolo precedente è un'analisi della varianza ad una via con disegno sperimentale completamente randomizzato per due popolazioni gaussiane indipendenti.
Dimostriamo, infatti, che nel caso del confronto delle medie di di due popolazioni gaussiane X e Y di ampiezza rispettivamente n ed m, la statistica usata per il test di Fisher è uguale al quadrato della statistica usata nel test t-student e cioè che:

$$T_F = \frac{MSA}{MSW} = T_t^2 = \left[\frac{\bar{X} - \bar{Y}}{S_{tot}\sqrt{\frac{1}{n} + \frac{1}{m}}}\right]^2,$$

se ciò è vero i due test sono equivalenti, o in altri termini il test t è semplicemente un caso speciale di analisi della varianza applicata a due gruppi.
Infatti se $T_t \sim t(n+m-2)$ allora $T_t^2 \sim F(1, n+m-2)$, in quanto per come definite le distribuzioni di Fisher (1.17) e t-student (1.16) si ha che:

$$T_t \sim t(n+m-2) \quad se \quad T = \sqrt{(n+m-2)}\frac{X}{\sqrt{Y}} \quad con \quad X \sim N(0,1) \quad e \quad Y \sim \chi^2(n+m-2)$$

da cui:

$$T_t^2 = (n+m-2)\frac{X^2}{Y} \quad con \quad X^2 \sim \chi^2(1) \quad e \quad Y \sim \chi^2(n+m-2)$$

e pertanto:

$$T_t^2 \sim F(1, n+m-2).$$

Quindi non resta che dimostrare che $T_F = T_t^2$.
Consideriamo il caso in cui $n = m$ e costruiamo:

$$T_F = \frac{MSA}{MSW}$$

- $\bar{X}_{..} = \frac{1}{2}\left(\bar{X} + \bar{Y}\right)$,
- $MSA = n\left[\bar{X} - \frac{1}{2}\left(\bar{X} + \bar{Y}\right)\right]^2 + n\left[\bar{Y} - \frac{1}{2}\left(\bar{X} + \bar{Y}\right)\right]^2 = n\left(\frac{1}{2}\bar{X} - \frac{1}{2}\bar{Y}\right)^2 + n\left(\frac{1}{2}\bar{Y} - \frac{1}{2}\bar{X}\right)^2 = (2n)\left[\frac{1}{2}\left(\bar{X} - \bar{Y}\right)\right]^2 = \frac{n}{2}\left(\bar{X} - \bar{Y}\right)^2$

- $MSW = \frac{1}{2n-2}\left[\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] = S_{tot}^2$.

Allora

$$T_F = \frac{\frac{n}{2}\left(\bar{X} - \bar{Y}\right)^2}{S_{tot}^2} = \left[\frac{\left(\bar{X} - \bar{Y}\right)}{\sqrt{\frac{2}{n}}S_{tot}}\right]^2 = T_t^2.$$

Nel caso in cui $n \neq m$ la dimostrazione è analoga, l'unica differenza è che la media generale sarà la media pesata:

$$\bar{X}_{..} = \frac{n\bar{X} + m\bar{Y}}{n+m}.$$

In conclusione il test t e l'analisi della varianza ad una via con disegno sperimentale completamente randomizzato, sono solo due procedure diverse relative allo stesso test per due gruppi. Naturalmente se ci sono più di due gruppi non è possibile utilizzare l'analisi della varianza sotto forma di un test t, ma si deve usare la forma generale sviluppata in questo capitolo.

Riportiamo ora un esempio (in cui $n \neq m$) che ci permetterà di vedere sperimentalmente che il test t-student e l'ANOVA ad una via su due popolazioni gaussiane indipendenti, danno lo stesso risultato. Analizziamo il seguente problema:

Problema 3.4. *E' stata studiata la morfologia cranio-facciale di* 26 *uomini con la sindrome delle apnee durante il sonno (OSAS) (cioè delle ostruzioni delle vie aeree superiori durante il sonno) e di* 37 *uomini senza sindrome (Non-OSAS). Una delle variabili di interesse era la distanza dal punto più supero-anteriore del corpo dell'osso ioide al piano orizzontale di Francoforte (il piano immaginario che connette il bordo superiore del meato acustico esterno e il bordo infraorbitario) misurato in mm. Si può dire da questi dati che le due popolazioni campionate differiscono rispetto alla distanza media dal corpo dell'osso ioide?*

Fissiamo un livello $\alpha = 0{,}05$ e risolviamo il seguente problema con il software Mathematica, prima utilizzando un test t-student e poi un ANOVA ad una via.
Riportiamo sia l'input che l'output di Mathematica:

data1 = {96.80, 100.70, 94.55, 99.65, 109.15, 102.75, 97.70, 92.10, 91.90, 89.50, 97.00, 101.00,

97.70, 97.00, 94.55, 106.45, 94.55, 94.05, 89.45, 89.85, 98.20, 88.25, 92.60, 98.25,

90.85, 95.25, 88.80, 101.40, 90.55, 109.80, 88.95, 101.05, 92.60, 97.00, 91.95, 88.95, 95.75};

data2 = {105.95, 114.90, 110.35, 123.10, 119.30, 110.00, 98.95, 114.20, 108.95, 105.05, 114.90, 114.35,

112.25, 106.15, 102.60, 102.40, 105.05, 112.65, 128.95, 117.70, 113.70, 116.30, 108.75, 113.30, 106.00, 101.75};

alfa = 0.05; n = 37; m = 26;

s1 = Variance[data1];

s2 = Variance[data2];

stot = $\frac{1}{n+m-2}$((n − 1)s1 + (m − 1)s2);

MX = Mean[data1];

MY = Mean[data2];

$T = \frac{\text{MX}-\text{MY}}{\sqrt{\text{stot}\left(\frac{1}{n}+\frac{1}{m}\right)}}$;
Tf = T^2;

quant = Quantile $\left[\text{StudentTDistribution}[n+m-2], 1-\frac{\text{alfa}}{2}\right]$;

$p = 2(1 - \text{CDF}[\text{StudentTDistribution}[n+m-2], -T])$;

Print["la media di X vale MX"];

Print[MX];

Print["la media di Y vale MY"];

Print[MY];

Print["la statistica test T vale"];

Print[T];

Print["La soglia in valore assoluto della regione critica per T è"];

Print[quant]

Print["Il livello di significatività p è"];

Print[p]

Print["la statistica test al quadrato Tf vale"];

Print[Tf];

Print["la soglia della regione critica per Tf vale"];

Print $\left[\text{quant}^2\right]$;

la media di X vale MX
95.8541
la media di Y vale MY
111.06
la statistica test T vale
−9.60486
La soglia in valore assoluto della regione critica per T è
1.99962
Il livello di significatività p è
8.149037000748649*^-14
la statistica test al quadrato Tf vale
92.2534
la soglia della regione critica per Tf vale

3.99849

```
Needs["ANOVA"]
ddata = {{1, 96.80}, {1, 100.70}, {1, 94.55}, {1, 99.65}, {1, 109.15}, {1, 102.75}, {1, 97.70}, {1, 92.10},
{1, 91.90}, {1, 89.50}, {1, 97.00}, {1, 97.70}, {1, 97.00}, {1, 94.55}, {1, 106.45}, {1, 94.55},
{1, 94.05}, {1, 89.45}, {1, 89.85}, {1, 98.20}, {1, 101.00}, {1, 88.25}, {1, 92.60}, {1, 98.25},
{1, 90.85}, {1, 95.25}, {1, 88.80}, {1, 101.40}, {1, 90.55}, {1, 109.80}, {1, 88.95}, {1, 101.05},
{1, 92.60}, {1, 97.00}, {1, 91.95}, {1, 88.95}, {1, 95.75}, {2, 105.95}, {2, 114.90}, {2, 110.35},
{2, 123.10}, {2, 119.30}, {2, 110.00}, {2, 98.95}, {2, 114.20}, {2, 108.95}, {2, 105.05}, {2, 114.90},
{2, 114.35}, {2, 112.25}, {2, 106.15}, {2, 102.60}, {2, 102.40}, {2, 105.05}, {2, 112.65}, {2, 128.95},
{2, 117.70}, {2, 113.70}, {2, 116.30}, {2, 108.75}, {2, 113.30}, {2, 106.00}, {2, 101.75}};

k = 2;
ntot = 63;
alfa = 0.05;
quant = Quantile[FRatioDistribution[k − 1, ntot − k], 1 − alfa];
Print["La soglia in valore assoluto della regione critica per T=Fratio è"];
Print[quant]
ANOVA[ddata]
```

La soglia in valore assoluto della regione critica per T=Fratio è
3.99849

		DF	SumOfSq	MeanSq	FRatio	PValue
ANOVA →	Model	1	3530.53	3530.53	92.2534	8.14349397063001*^-14
	Error	61	2334.46	38.2699		
	Total	62	5864.99			

CellMeans →	All	102.129
	Model[1]	95.8541
	Model[2]	111.06

Dai risultati ottenuti osserviamo che:

1. il valore d T per il test di student in valore assoluto è $T = 9.60486$ e la soglia della regione critica al livello $\alpha = 0{,}05$ vale: $t_{1-\frac{\alpha}{2}} = 1.99962$, quindi dato che $T > t_{1-\frac{\alpha}{2}}$ si

rifiuta l'ipotesi di uguaglianza delle medie delle due popolazioni Gaussiane,

2. il valore di $FRatio$ per il test ANOVA è $FRatio = 92.2534$ e la soglia della regione critica al livello $\alpha = 0{,}05$ vale $F_{1-\alpha} = 3.99849$, quindi dato che $Fratio > F_{1-\alpha}$ si rifiuta l'ipotesi di uguaglianza delle medie sui due livelli del trattamento,

3. $T^2 = T_f = FRatio$ e $(t_{1-\frac{\alpha}{2}})^2 = F_{1-\alpha}$,

4. infine i valori di p sono uguali nei rispettivi test e $p < 0{,}05$.

Da questo esempio è chiaro che nel caso di un trattamento con due livelli i test sono esattamente identici.

3.1.2 Verifica adeguatezza del modello, analisi dei residui

Il modello ad effetti fissi che abbiamo utilizzato prevede:

$$x_{ij} = \mu + \tau_j - e_{ij},$$

dove $e_{ij} = x_{ij} - \mu_j$ rappresenta il *termine errore* ed è una variabile aleatoria che indica la distanza tra un'osservazione generica e la previsione del modello.
Il controllo delle assunzioni del modello attraverso l'analisi dei residui è una fase importante affinchè il test ANOVA abbia validità.
Le assunzioni sono riassunte in $e_{ij} \sim N(0, \sigma^2)$, ovvero:

1. normalità;

2. varianza costante (omoschedasticità);

3. indipendenza,

del termine di errore casuale e_{ij} (non osservabile) che compare nel modello di rappresentazione dei dati. *I residui* sono definiti come:

$$\bar{e}_{ij} = x_{ij} - \bar{X}_{.j},$$

ovvero come la differenza tra dato osservato e valore previsto dal modello.
Se gli errori rispettassero le assunzioni, i residui (da considerare delle realizzazioni della variabile casuale e_{ij}) "erediterebbero" le stesse caratteristiche. Definiamo quindi i *residui standardizzati*:

$$d_{ij} = \frac{\bar{e}_{ij}}{\sqrt{MSW}},$$

dato che MSW è uno stimatore corretto di σ^2, si dovrebbe avere:

$$d_{ij} \sim N(0, 1)$$

e quindi nel 95% dei casi:

$$-1{,}96 < d_{ij} < 1{,}96.$$

Ai fini di effettuare una analisi corretta è bene verificare che ciò sia verificato, in quanto valori di $|d_{ij}| > 2$ potrebbero suggerire che le assunzione sugli errori sono state violate e quindi che il modello utilizzato non è coerente.

3.1.3 Confronti multipli a posteriori

Ricordiamo che quando l'analisi della varianza conduce al rifiuto dell'ipotesi nulla di nessuna differenza tra le medie delle popolazioni, non significa che le medie sono tutte significativamente diverse l'una dall'altra ma, piuttosto, che c'è almeno una coppia di medie la cui differenza è statisticamente significativa. Quindi viene naturale chiedersi quali sono quelle coppie di medie che possono considerarsi diverse. Abbiamo già osservato, all'inizio del capitolo, che testare la differenza tra le medie di tutte le coppie di trattamenti con un test t-student non sarebbe efficace per diversi motivi. Negli ultimi anni sono però state proposte diverse tecniche per fare confronti multipli. Per individuare i test appropriati occorre distinguere se i confronti sono stabiliti a priori oppure se non sono prestabiliti, ma il loro interesse si manifesti dopo aver osservato il risultato della sperimentazione, cioè a posteriori. Questa precisazione potrebbe risultare di non immediata comprensione, in effetti ci si potrebbe chiedere:

- cosa cambia, data una certa differenza tra due medie, se il loro confronto era prestabilito oppure no?

Il punto cruciale sta nel fatto che se il confronto era prestabilito, è chiaro che lo sperimentatore si aspetta di trovare a priori un differenza tra i trattamenti utilizzati e quindi si vuole solo verificare se questa differenza è sufficientemente grande da essere imputata ad un effetto dovuto al trattamento e non solo a fattori casuali. Il test da utilizzare è quindi un test "poco conservativo", che ci consente di respingere l'ipotesi nulla, perché siamo abbastanza fiduciosi di non commetter un errore di primo tipo (respingere un ipotesi nulla che in effetti poteva essere vera). Nella pratica sperimentale però è raro riuscire a pianificare a priori i confronti, soprattutto quando questi sono intrinsecamente equivalenti. Pertanto eventuali confronti interessanti nascono dai risultati dell'esperimento. Ad esempio, nell'esperimento in cui vogliamo saggiare l'efficacia di differenti farmaci, ma non sappiamo quale sia il più valido, non possiamo programmare a priori dei confronti. In questo caso è più conveniente basarsi su quanto suggeriscono i dati sperimentali e verificare se il farmaco risultato più produttivo è significativamente migliore di quello che lo segue nella scala di merito. Il test da utilizzare dovrà essere pertanto più conservativo perché ci sono discrete probabilità di commettere un errore di primo tipo. I *confronti a posteriori* vengono detti anche *confronti non prestabiliti o non pianificati* e in inglese *post-hoc comparison.*
In alcuni testi i confronti a posteriori sono presentati come sostitutivi all'analisi della varianza, ma per un principio di cautela, molti autori suggeriscono di utilizzarli solo dopo che l'analisi della varianza ha permesso di rifiutare l'ipotesi H_0.
Per effettuare questi confronti sono stati proposti diversi metodi, che come impostazione logica derivano dal test t-student per ipotesi bilaterali e intervalli di fiducia. La scelta del test più adeguato dipende da tre problemi tra loro collegati e che hanno determinato soluzioni differenti e quindi proposte di test differenti:

1. la stima di α per ogni confronto, in modo che la probabilità totale dell'errore di prima specie non superi il livello prestabilito;

2. il numero totale p dei confronti che si devono effettuare;

3. il calcolo di un intervallo di fiducia valido per tutti i confronti.

Il problema più discusso è stato il primo, riguardante l'errore di primo tipo. Per capire meglio di cosa si tratta distinguiamo:

- **il comparisonwise**, collegato alla potenza del singolo test, corrisponde alla probabilità di commettere un errore di tipo uno nei singoli confronti e che all'aumentare del numero di confronti, contribuisce a far aumentare l'errore sull'esperimento complessivo, come mostrato all'inizio del capitolo;
- **l'experimentwise**, collegato al principio di cautela o protezione di tutta la serie di test, corrisponde alla probabilità d'errore su tutto l'esperimento

La ricerca del difficile equilibrio tra potenza del test e protezione, per il quale non è ancora stata trovata una soluzione universalmente condivisa, ha portato alla formulazione di diversi metodi basati su logiche differenti e che quindi portano anche a risultati divergenti, l'unica cosa su cui si resta concordi per un fatto di protezione è quella di far sempre precedere un *test ANOVA* ai confronti multipli. Le diverse procedure si basano tutte essenzialmente sul test t, ma includendo correzioni opportune per il fatto che vengono confrontate più medie, in questa tesi approfondiremo solo tre dei diversi test per confronti multipli a posteriori che sono presenti in letteratura:

1. Test LSD (Least Significant Difference) di Fisher
2. Test t con correzione di Bonferroni
3. Test HSD (Honestly Significant Difference) di Tukey e l'estensione di Kramer

Analizziamo ora i diversi test.

3.1.4 Il test LSD di Fisher e la correzione di Bonferroni

Il metodo LSD di **Fisher** è il primo metodo di confronto multiplo, sviluppato dallo stesso inventore dell'ANOVA. Esso consiste nel calcolare una differenza minima (LSD) che deve essere oltrepassata perché una differenza tra medie possa essere considerata significativa.
Dati due gruppi i e j, la differenza minima richiesta perché lo scarto tra le loro medie possa essere considerato significativo è data da:

$$LSD = t_{1-\frac{\alpha}{2}} \sqrt{MSW \left(\frac{1}{n_i} + \frac{1}{n_j} \right)},$$

ovvero l'unica differenza dal test t classico è che nella proposta di Fisher il test è costruito utilizzando lo stimatore della varianza entro i gruppi (MSW) calcolato nel test ANOVA:

$$T = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MSW \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}},$$

in questo modo i gradi di libertà del test t sono quelli corrispondenti a MSW e quindi il numero totale delle osservazioni meno il numero di trattamenti, piuttosto che $n_i + n_j - 2$ e quindi sono in numero superiore a quelli di un test t classico.
Nel metodo LSD ciò che è "buona prassi" per gli altri post-hoc, ovvero anteporre una analisi ANOVA prima di effettuare i confronti multipli, diventa una condizione imprescindibile. Infatti l'unica logica seguita è che, se il valore F è significativo, allora il rischio di incappare in un errore di prima specie è più bassa del normale, perché sappiamo già che H_0 non è valida, ma questa logica è molto controversa perchè non prende in considerazione il problema della relazione tra *experimentwise* e *comparisonwise*. Per questo motivo oggi, il metodo LSD è

ormai poco utilizzato e si preferiscono altri approcci, come la correzione di Bonferroni.
Il matematico italiano **Bonferroni** si occupò di migliorare il test LSD di Fisher proponendo una procedura che teneva conto dell'aumento dell'errore di prima specie all'aumentare del numero di confronti.
A Bonferroni è attribuita la relazione sulla disuguaglianza (in molti testi chiamata **disuguaglianza di Boole**) tra probabilità che afferma che:

Definizione 3.5. *Se $A_1, A_2...A_n$ sono eventi compatibili (il verificarsi di uno non esclude il verificarsi degli altri e possono verificarsi contemporaneamente), la probabilità che almeno uno di essi si verifichi è minore o al più uguale alla somma delle probabilità che ciascuno di essi ha di verificarsi, indipendentemente dagli altri. Essa può essere scritta come:*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq P(A_1) + ... + P(A_n)$$

Secondo questa disuguaglianza la stima di α_T sulla serie di test (*experimentwise*), quando il numero di confronti da effettuare è p e il livello dei test è α è :

$$\alpha_T < p\alpha,$$

quindi:

$$\frac{\alpha_T}{p} < \alpha.$$

Questo significa che se per esempio $p = 3$ e se la probabilità totale α_T di commettere un errore di prima specie non deve essere superiore a 0,05, la probabilità α di ogni singolo confronto deve essere minore di 0,0166 ($\frac{0.05}{3}$).
In realtà però, la relazione corretta tra α e α_T non è lineare, ma esponenziale. Per una stima più accurata del *comparisonwise* sulla base dell'*experimentwise* è utile osservare che:

$$\alpha_T = 1 - (1 - \alpha)^p.$$

Quindi se per esempio, $\alpha_T = 0{,}05$ e $p = 5$, si ha che α non è 0,01 ($\frac{0{,}05}{5}$), ma è:

$$\alpha = 1 - 0{,}095^{\frac{1}{5}} = 0{,}01021,$$

quindi c'è una differenza rispetto alla stima di Bonferroni del 2,1%, che tende ad aumentare quando il numero di confronti aumenta.
Bonferroni propose di utilizzare ciascun test t usando il valore critico corrispondente a $\frac{\alpha_T}{p}$, ma questa procedura funziona abbastanza bene quando ci sono pochi gruppi da confrontare, come abbiamo visto quando il numero di confronti cresce l'errore percentuale aumenta e oltre a 10 confronti il valore di t richiesto per concludere che esiste una differenza diventa molto più grande del necessario e quindi il metodo proposto da Bonferroni diventa troppo cautelativo. A questo proposito ci sono altri test che sono meno prudenti, come il test di Holm o il test di Tukey che verrà discusso in seguito, ma tutte le procedure sono simili al test t di Bonferroni.
Anche nella proposta di Bonferroni il test t che si utilizza è costruito utilizzando lo stimatore della varianza entro i gruppi (MSW) calcolato nel test ANOVA, ovvero:

$$T = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MSW\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}},$$

Questo permette di rendere il test un pò meno cautelativo, in quanto il valore critico di t decresce quando i gradi di libertà aumentano. Quindi è possibile rilevare una differenza, con un grado prefissato di fiducia, anche in presenza di minore differenze assolute fra le medie.
Come abbiamo già osservato il problema della correzione di Bonferroni è che può far calare troppo il valore α se i confronti sono tanti. Se ci sono più di 10 confronti, α può scendere tanto da rischiare di prendere per non significativi dei risultati che invece dovrebbero esserlo, ovvero aumenta troppo l'errore di seconda specie e il test perde potenza. Quindi quando i confronti sono più di dieci, si preferisce utilizzare altri tipi di test come il test di Tukey.

3.1.5 Il test HSD di Tukey e l'estensione di Kramer

Il test di Tukey del 1953 è chiamato *wholly significant difference test* (differenza interamente significativa), perchè pone attenzione *all'experimentwise*. Come proposto dallo stesso autore, viene generalmente chiamato *honestly significant difference*, da cui l'acronimo HSD, perchè vuole essere un compromesso "onesto" tra *experimentwise* e *comparisonwise*. Una volta rigettata H_0, si intende eseguire un test statistico confrontando tutte le possibili combinazioni di medie $\binom{k}{2}$ dei trattamenti:

$$H_0 : \mu_i = \mu_j$$
$$H_A : \mu_i \neq \mu_j$$

per ogni coppia (i, j).
Tukey ha proposto una procedura per questo test delle ipotesi, la cui significatività complessiva è pari proprio all'experimentwise α, nel caso in cui le dimensioni del campione siano uguali per tutti i trattamenti.
Costruiamo ora il test:

- si fa riferimento alla distribuzione della *statistica "studentizzata" di intervallo*:

$$Q = \frac{\bar{X}_{max} - \bar{X}_{min}}{\sqrt{\frac{MSW}{n}}} \sim q(k, N-k),$$

  dove:

  - $q(k, N-k)$ è la distribuzione della statistica "studentizzata" di Tukey, dove k è il numero di gruppi (livelli del trattamento) e $N-k$ è il numero di gradi di libertà di MSW;
  - n è il numero di osservazioni di ogni livello del trattamento;
  - $\bar{X}_{max}$ e $\bar{X}_{min}$ sono rispettivamente la massima e la minima media campionaria sui k livelli del trattamento;
  - MSW è l'errore o la media quadratica entro i gruppi, stimatore corretto della varianza all'interno dei gruppi calcolato nel test ANOVA (quindi invece di considerare lo stimatore per la varianza totale sui due singoli gruppi confrontati, come si farebbe in un test t-student, si utilizza lo stimatore MSW su tutti i k gruppi).

- E' bene notare che per il confronto simultaneo tra le $\binom{k}{2}$ coppie di medie, il livello di significatività è costruito sul caso peggiore, cioè sulla differenza massima data da $\bar{X}_{max} - \bar{X}_{min}$, per questo si ritiene che il test di Tukey fornisca una probabilità *experimentwise* appropriata per il complesso dei confronti.

- Definiamo $q_\alpha(k, N-k)$ i valori critici della statistica (che si possono trovare in tabelle disponibili in letteratura), dove:

  - α è il livello di significatività del test,
  - k è il numero di levelli del trattamento presi in considerazione,
  - $N-k$ è il numero di gradi di libertà di MSW.

- Costruiamo ora la regione di rigetto dell'ipotesi H_0 al livello α. E' chiaro che rifiuteremo l'ipotesi quando la differenza $\bar{X}_{max} - \bar{X}_{min}$ è sufficientemente lontana dallo zero, ovvero se:

$$P\left(Q > q_\alpha(k, N-k)\right) = \alpha$$

che è equivalente a:

$$P\left(\bar{X}_{max} - \bar{X}_{min} > q_\alpha(k, N-k)\sqrt{\frac{SSW}{n}}\right) = \alpha$$

e quindi la regione di rigetto è:

$$D = \left]q_\alpha(k, N-k)\sqrt{\frac{SSW}{n}}, +\infty\right[.$$

Quindi nel caso in cui il numero di osservazioni di ogni trattamento è sempre uguale ad n il test consiste nel confrontare tutte le differenze tra le medie delle popolazioni con il valore detto HSD:

$$HSD = q_\alpha(k, N-k)\sqrt{\frac{SSW}{n}},$$

e ogni differenza che conduce ad un valore assoluto che supera HSD viene dichiarata significativa.
Nel caso in cui il numero di osservazioni è n_j, diverso per ogni trattamento, il valore di confronto è diverso per ogni differenza. Questo è stato trovato da Kramer estendendo il lavoro di Tukey ed è:

$$HSD^* = q_\alpha(k, N-k)\sqrt{\frac{SSW}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}.$$

Riprendiamo ora il problema analizzato in precedenza (3.1) e effettuiamo un test di Bonferroni e un test di Tukey-Kramer (in quanto i livelli dei trattamenti non hanno tutti la stessa dimensione). Utilizziamo sempre il software Mathematica, che per prima cosa effettua una analisi della varianza e solo in caso di rifiuto dell'ipotesi H_0 permette di effettuare un test $post-hoc$.

Riportiamo ora l'input e l'output di Mathematica:

ANOVA[dataset1, PostTests → Bonferroni, CellMeans → False]

$$\left\{\begin{array}{lllllll} & & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\ \text{ANOVA} \to & \text{Model} & 3 & 21261.9 & 7087.3 & 27.0001 & 7.692037481713481\text{*^-}14 \\ & \text{Error} & 140 & 36748.9 & 262.492 & & \\ & \text{Total} & 143 & 58010.8 & & & \end{array}\right\}$$

$$\left\{\text{PostTests} \to \left\{\text{Model} \to \quad \text{Bonferroni} \quad \{\{1,2\},\{2,3\},\{1,4\},\{2,4\},\{3,4\}\} \right\}\right\}$$

ANOVA[dataset1, PostTests → Tukey, CellMeans → False]

$$\left\{\begin{array}{lllllll} & & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\ \text{ANOVA} \to & \text{Model} & 3 & 21261.9 & 7087.3 & 27.0001 & 7.692037481713481\text{*^-}14 \\ & \text{Error} & 140 & 36748.9 & 262.492 & & \\ & \text{Total} & 143 & 58010.8 & & & \end{array}\right\}$$

$$\left\{\text{PostTests} \to \left\{\text{Model} \to \quad \text{Tukey} \quad \{\{1,2\},\{2,3\},\{1,4\},\{2,4\},\{3,4\}\} \right\}\right\}$$

Le coppie riportate dal test rappresentano le coppie di medie per le quali si rifiuta l'ipotesi H_0 di uguaglianza. Quindi poichè il test viene fatto su $\binom{4}{2} = 6$ coppie, abbiamo che per 5 di esse si conclude che si ha una differenza significativa, mentre sulla coppia mancante, $(1,3)$, non si riscontra una differenza significativa e quindi si può accettare l'ipotesi di uguaglianza. Inoltre osserviamo che i risultati ottenuti con il test di Tukey sono gli stessi di quelli ottenuti con il test di Bonferroni in quanto il numero di confronti è minore di 10.

3.1.6 Il disegno sperimentale a blocchi completamente randomizzato

Nella sezione precedente abbiamo visto come usare il *disegno sperimentale completamente randomizzato* nel caso in cui le unità trattate sono omogenee. Ora però se ci troviamo di fronte a un problema come il seguente:

Problema 3.6. *Un fisioterapista desidera confrontare tre metodi di insegnamento per l'uso di una protesi per alcuni pazienti. Suppone però che il tasso di apprendimento sia diverso a seconda dell'età dei pazienti e vuole approntare un disegno sperimentale in modo tale da poter eliminare l'effetto età.*

E' chiaro che in questo caso il *disegno sperimentale completamente randomizzato* non sarebbe efficace in quanto implicherebbe di scegliere casualmente un campione di pazienti al quale somministrare il primo metodo di insegnamento, un altro a cui somministrare il secondo e così via e in questo modo non si potrebbe eliminare l'effetto dell'età.
In questo caso infatti si deve utilizzare il *disegno sperimentale a blocchi completamente randomizzato*, che fu sviluppato da R.A Fisher nel 1925 ed è, tra tutti i disegni sperimentali, il più usato.
Il *blocco* è un fattore di disturbo noto e controllabile (quasi sempre di tipo qualitativo) che molto probabilmente produce sulla risposta un effetto, che non interessa però allo sperimentatore. Tuttavia la variabilità che trasmette alla risposta deve essere minimizzata. Per limitare il fattore di disturbo si usa il disegno sperimentale a blocchi completamente randomizzato che è un disegno in cui le unità sperimentali, alle quali sono applicati i trattamenti (i metodi di insegnamento nel nostro esempio), sono suddivise in gruppi omogenei chiamati *blocchi* (che corrispondono alle età nel nostro esempio), in modo tale che il numero delle unità sperimentali in un blocco è uguale al numero (o a qualche suo multiplo) dei trattamenti in studio. I trattamenti vengono poi assegnati a caso alle unità sperimentali all'interno di ogni blocco, in modo che ogni trattamento è presente in ogni blocco e che ogni blocco contiene tutti i trattamenti.
L'obiettivo del disegno sperimentale a blocchi completamente randomizzato è proprio quello di isolare e di sollevare dalla componente errore la variazione attribuibile ai blocchi, garantendo che nelle medie dei trattamenti non è presente alcun effetto dovuto ai blocchi. L'abilità nel formare blocchi omogenei dipende dalla conoscenza del ricercatore del materiale degli esperimenti. Ad esempio negli esperimenti sugli animali si pensa che differenti razze rispondano in maniera diversa allo stesso trattamento , e quindi la razza degli animali potrebbe essere usata come l'elemento per la formazione dei blocchi.
In conclusione il disegno sperimentale a blocchi completamente randomizzato si utilizza quando esiste un fattore sub-sperimentale che è causa di un'alta variabilità nelle risposte; come l'età infantile, adulta oppure anziana, tra pazienti ai quali sia stato somministrato lo stesso insegnamento. In questi casi è utile ridurre la variabilità non controllata o varianza d'errore, se si vuole aumentare la probabilità che il test di confronto tra medie del fattore sperimentale (gli insegnamenti nel nostro esempio) risulti significativo.
La tecnica che analizza i dati di un disegno sperimentale a blocchi completamente randomizzato viene denominata *analisi della varianza a una via con blocco*, alcuni autori parlano di *analisi della varianza a due vie* in quanto una osservazione viene classificata rispetto a due criteri: il blocco e il trattamento, ma ricordiamo che il nostro interesse è posto unicamente sul *trattamento*, quindi su un unico *fattore*, pertanto preferiamo fare rientrare questo caso nella analisi della varianza ad una via. Costruiamo ora un modello generale facendo alcune ipotesi restrittive.

- Rappresentiamo i dati nel seguente modo:

Trattamenti						
Blocchi	1	2	...	k	Totale	Media
1	x_{11}	x_{12}	...	x_{1k}	$T_{1.}$	$\bar{X}_{1.}$
2	x_{21}	x_{22}	...	x_{2k}	$T_{2.}$	$\bar{X}_{2.}$
3	x_{31}	x_{32}	...	x_{3k}	$T_{3.}$	$\bar{X}_{3.}$
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
n	x_{n1}	x_{n2}	...	x_{nk}	$T_{n.}$	$\bar{X}_{n.}$
Totale	$T_{.1}$	$T_{.2}$	...	$T_{.k}$	$T_{..}$	
Media	$\bar{X}_{.1}$	$\bar{X}_{.2}$	...	$\bar{X}_{.k}$		$\bar{X}_{..}$

Analizziamo i dati:

x_{ij} rappresenta una generica osservazione, l'indice i-esimo è relativo al blocco e l'indice j-esimo al trattamento e ogni blocco è soggetto a tutti i trattamenti.

$T_{i.} = \sum_{j=1}^{k} x_{ij}$ è il totale del blocco i-esimo;

$\bar{X}_{i.} = \frac{1}{k}\sum_{j=1}^{k} x_{ij} = \frac{T_{i.}}{k}$ è la media del blocco i-esimo;

$T_{.j} = \sum_{i=1}^{n} x_{ij}$ è il totale del trattamento j-esimo;

$\bar{X}_{.j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} = \frac{T_{.j}}{n}$ è la media del trattamento j-esimo;

$T_{..} = \sum_{j=1}^{k} T_{.j} = \sum_{i=1}^{n} T_{i.}$ è il totale generale;

$\bar{X}_{..} = \frac{1}{n}\sum_{i=1}^{n} \bar{X}_{i.} = \frac{1}{k}\sum_{j=1}^{k} \bar{X}_{.j}$ è la media generale che si può ottenere sia sommando per riga che per colonna.

- Scomponiamo una qualunque osservazione del nostro insieme di dati nel modo seguente:

$$x_{ij} = \mu + \beta_i + \tau_j + e_{ij}$$

$$i = 1, 2, ...n \quad e \quad j = 1, 2....k,$$

in questo modello:

1. x_{ij} è un valore proveniente dalla intera popolazione,
2. μ è una costante incognita,
3. β_i rappresenta *l'effetto del blocco* ovvero $\beta_i = \mu_i - \mu$,

4. τ_j rappresenta *l'effetto del trattamento*, ovvero $\tau_j = \mu_j - \mu$,
5. e_{ij} è la componente residua che rappresenta tutte le fonti di variabilità, eccetto quella dovuta al trattamento e al blocco, ovvero $e_{ij} = x_{ij} - \mu_{ij}$.

Osserviamo che nel modello costruito:

$$x_{ij} = \mu + (\mu_i - \mu) + (\mu_j - \mu) + (x_{ij} - \mu_{ij}),$$

si ha che:

$$\mu = \mu_i + \mu_j - \mu_{ij}.$$

inoltre nella ipotesi per cui *i trattamenti* hanno tutti lo stesso effetto, ovvero non producono medie significativamente differenti si vuole che $\tau_j = 0$, ovvero:

$$x_{ij} = \mu + (\mu_i - \mu) + (x_{ij} - \mu_{ij}), \quad con \quad \mu_{ij} = \mu_i,$$

in altre parole questo significa che la media di ogni variabile aleatoria x_{ij} dipende solo dal blocco a cui appartiene e non dal trattamento a cui è sottoposta e quindi i trattamenti sono tra loro "equivalenti".

- Le assunzioni per il *modello ad effetti fissi* sono:
  - Ogni x_{ij} osservato costituisce un campione casuale semplice di dimensione 1 proveniente da una delle kn popolazioni.
  - Ognuna di queste kn popolazioni è distribuita normalmente con media μ_{ij} e varianza costante σ^2.
  - Gli e_{ij} sono variabili aleatorie indipendenti e normalmente distribuite con media 0 e varianza σ^2
  - I τ_j e β_i sono un insieme di costanti fissate.
  - Gli effetti del trattamento e del blocco sono additivi :

    $$\sum_{j=1}^{k} \tau_j = \sum_{i=1}^{n} \beta_i,$$

    questa assunzione può essere interpretata come assenza di *interazione* tra i trattamenti e i blocchi. In altre parole, una particolare combinazione blocco-trattamento non produce un effetto che è maggiore o minore della somma della somma dei singoli effetti individuali.

- Facciamo l'ipotesi nulla che tutti i trattamenti medi siano uguali, ovvero che le x_{ij} siano osservazioni di n blocchi omogenei sullo stesso trattamento:

$$H_0 : \tau_j = 0 \quad j = 1, 2...k$$

contro l'alternativa:

$$H_A : non \quad tutti \quad i \quad \tau_j = 0.$$

L'ipotesi nulla corrisponde così a $\mu_{ij} = \mu_i$, ovvero: $\mu_{i1} = \mu_{i2} = ... = \mu_{ik}$, cioè eliminando la variabilità dovuta ai blocchi, le medie rispetto ai trattamenti sono le stesse. Un test d'ipotesi relativo agli effetti dei blocchi non viene usualmente fatto sotto l'assunzione di un modello ad effetti fissi, in quanto l'interesse primario è *l'effetto del trattamento* e i blocchi servono in genere per dare uno strumento per eliminare un fattore di variabilità estranea, inoltre i blocchi sono ottenuti in maniera del tutto non casuale.

- Costruiamo ora la statistica test opportuna.
  Definiamo:

  – la devianza totale:

  $$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \bar{X}_{..})^2$$

  e ricordando che nel modello ad effetti fissi:

  $$x_{ij} = \mu + \beta_i + \tau_j + e_{ij},$$

  – la devianza attribuibile ai blocchi:

  $$SSBl = \sum_{j=1}^{k}\sum_{i=1}^{n}(\bar{X}_{i.} - \bar{X}_{..})^2, \quad da \quad \beta_i = \mu_i - \mu,$$

  – la devianza attribuibile ai trattamenti:

  $$SSTr = \sum_{j=1}^{k}\sum_{i=1}^{n}(\bar{X}_{j.} - \bar{X}_{..})^2, \quad da \quad \tau_j = \mu_j - \mu,$$

  – la devianza attribuibile ad un altro tipo di errore:

  $$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2, \quad da \quad e_{ij} = x_{ij} - \mu_i - \mu_j + \mu,$$

  Si può dimostrare che:

  $$SST = SSBl + SSTr + SSE,$$

  in quanto:

  $$\sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \bar{X}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n}\left[(\bar{X}_{i.} - \bar{X}_{..}) + (\bar{X}_{.j} - \bar{X}_{..}) + (x_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})\right]^2$$

  $$= \sum_{j=1}^{k}\sum_{i=1}^{n}(\bar{X}_{i.} - \bar{X}_{..})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n}(\bar{X}_{j.} - \bar{X}_{..})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2,$$

  dato che i doppi prodotti si annullano.
  Inoltre i gradi di libertà per ognuna delle rispettive somme sono:

  $$kn - 1 = (n-1) + (k-1) + (n-1)(k-1),$$

  dove il numero di gradi di libertà per SSE può essere ottenuto per sottrazione:

  $$(kn-1) - (n-1) - (k-1) = (n-1)(k-1).$$

  Inoltre nel caso in cui si potesse ipotizzare che tutte le osservazioni x_{ij} provenissero dalla stessa popolazione gaussiana, ovvero sia che l'effetto dei trattamenti che dei blocchi fosse nullo, avremmo che ognuna di queste somme divisa per i rispettivi gradi di libertà sarebbe uno stimatore corretto della varianza σ^2 e la dimostrazione sarebbe equivalente a quella condotta per trovare stimatori corretti di σ^2 nel test *ANOVA ad*

una via della sezione precedente.
Nel caso in questione, invece, assumendo l'ipotesi nulla H_0 per la quale gli effetti dei trattamenti sono nulli abbiamo che:

$$MSTr = \frac{SSTr}{k-1} \quad e \quad MSE = \frac{SSE}{(n-1)(k-1)}$$

sono stimatori corretti per σ^2. A questo punto, ragionando in modo analogo al caso di un singolo fattore, è chiaro che saggiare l'ipotesi nulla equivale a saggiare il rapporto tra i due stimatori per la varianza σ^2.

- Costruiamo la regola di decisione:
  quando l'ipotesi nulla è vera abbiamo che:

$$T = \frac{MSTr}{MSE} \sim F((k-1),(n-1)(k-1)),$$

pertanto accetteremo H_0 quando il rapporto T è minore di $F_{1-\alpha}$, ovvero quando le due quantità sono effettive stime di σ^2.
Se invece il valore sperimentale T assume valore in:

$$D =]F_{1-\alpha}, +\infty[$$

rifiuteremo l'ipotesi in quanto ciò implica che $MSTr$ assuma un valore superiore a MSE e quindi che la devianza attribuibile ai trattamenti sia non trascurabile.
Per completare l'analisi statistica è bene riportare anche il valore di p; p tale che:

$$T_{oss} = F_{1-p}.$$

Analizziamo ora il problema prima descritto (3.6) facendo uso del software Mathematica.
In questo caso i dati vanno riportati in una lista di liste dove il primo e secondo termine indicano rispettivamente livello del trattamento e blocco, che Mathematica denota come $factor1$ e $factor2$. Inoltre, dato che il disegno sperimentale prevede due fattori (trattamento e blocco) il comando ANOVA necessita sia della lista di dati, ma anche del modello e della opzione, che in questo caso è solo $\{factor1, factor2\}$ in quanto non si ha interazione tra i fattori. L'output di Mathematica rilascia sia il valore $FRatio$ sui trattamenti che sui blocchi, anche se noi non siamo interessati a questo in quanto stiamo operando una analisi della varianza ad una via, in cui la nostra unica variabile di interesse è il trattamento. Sappiamo già che i blocchi sono diversi tra loro e sono stati costruiti opportunamente.
Fissiamo ora un livello $\alpha = 0{,}05$, chiamiamo $k = 3$ il numero di livelli del trattamento e $n = 5$ il numero di blocchi e riportiamo sia l'input che l'output di Mathematica:

```
Needs["ANOVA"]

dataset2 = {{1, 1, 7}, {1, 2, 8}, {1, 3, 9}, {1, 4, 10}, {1, 5, 11}, {2, 1, 9}, {2, 2, 9}, {2, 3, 9}, {2, 4, 9}, {2, 5, 12},
{3, 1, 10}, {3, 2, 10}, {3, 3, 12}, {3, 4, 12}, {3, 5, 14}};

k = 3;

n = 5;
```

alfa = 0.05;

quant = Quantile[FRatioDistribution[$k-1,(n-1)(k-1)$], 1 − alfa];

Print["La soglia in valore assoluto della regione critica per T=FRatio sui trattamenti è"];

Print[quant]

ANOVA[dataset2, {factor1, factor2}, {factor1, factor2}]

La soglia in valore assoluto della regione critica per T=FRatio sui trattamenti è

4.45897

		DF	SumOfSq	MeanSq	FRatio	PValue
ANOVA →	factor1	2	18.5333	9.26667	21.3846	0.000616535
	factor2	4	24.9333	6.23333	14.3846	0.00100166
	Error	8	3.46667	0.433333		
	Total	14	46.9333			

CellMeans →	All	10.0667
	factor1[1]	9.
	factor1[2]	9.6
	factor1[3]	11.6
	factor2[1]	8.66667
	factor2[2]	9.
	factor2[3]	10.
	factor2[4]	10.3333
	factor2[5]	12.3333

Osserviamo che il valore $FRatio$ sui trattamenti vale : $FRAtio = 21.3846$ e il valore soglia $F_{1-\alpha}$ con rispettivi gradi di libertà vale: $F_{1-\alpha} = 4.45897$, quindi poichè $FRatio > F_{1-\alpha}$ rifiutiamo l'ipotesi nulla e quindi ammettiamo che non tutti gli effetti dei trattamenti sono uguali a zero, o in maniera equivalente che non tutte le medie dei gruppi sono uguali. Inoltre osserviamo che in questo caso $p < 0{,}005$.
Una volta rifiutata l'ipotesi nulla, ci si può chiedere quali siano effettivamente i metodi di insegnamento che differiscono tra loro. Si può quindi effettuare un test di Tukey o indifferentemente un test di Bonferroni in quanto il nostro interesse è posto sui livelli del trattamento e quindi il numero di coppie da confrontare è $\binom{3}{2} = 3 < 10$.
Riportiamo l'input e l'output di Mathematica sia per il test di Tukey che per il test di Bonferroni:

ANOVA[dataset2, {factor1, factor2}, {factor1, factor2}, PostTests → Tukey, CellMeans → False]

$$\left\{\begin{array}{llllll} & & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\ & \text{factor1} & 2 & 18.5333 & 9.26667 & 21.3846 & 0.000616535 \\ \text{ANOVA} \rightarrow & \text{factor2} & 4 & 24.9333 & 6.23333 & 14.3846 & 0.00100166 \\ & \text{Error} & 8 & 3.46667 & 0.433333 & & \\ & \text{Total} & 14 & 46.9333 & & & \end{array}\right\}$$

$$\left\{\text{PostTests} \rightarrow \left\{\text{factor1} \rightarrow \text{Tukey} \quad \{\{1,3\},\{2,3\}\}\right\}\left\{\text{factor2} \rightarrow \text{Tukey} \quad \{\{1,5\},\{2,5\},\{3,5\},\{4,5\}\}\right\}\right\}$$

ANOVA[dataset2, {factor1, factor2}, {factor1, factor2}, PostTests → Bonferroni, CellMeans → False]

$$\left\{\begin{array}{llllll} & & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\ & \text{factor1} & 2 & 18.5333 & 9.26667 & 21.3846 & 0.000616535 \\ \text{ANOVA} \rightarrow & \text{factor2} & 4 & 24.9333 & 6.23333 & 14.3846 & 0.00100166 \\ & \text{Error} & 8 & 3.46667 & 0.433333 & & \\ & \text{Total} & 14 & 46.9333 & & & \end{array}\right\}$$

$$\left\{\text{PostTests} \rightarrow \left\{\text{factor1} \rightarrow \text{Bonferroni} \quad \{\{1,3\},\{2,3\}\} \text{ , factor2} \rightarrow \text{Bonferroni} \quad \{\{1,5\},\{2,5\},\{3,5\}\}\right\}\right\}$$

E' bene osservare che richiedendo un test post-hoc Mathematica calcola prima i risultati del test ANOVA, che abbiamo già analizzato e solo successivamente opera con il test di Tukey o di Bonferroni. Inoltre il test *post-hoc* richiesto opera sia sui trattamenti ($factor1$), che sui blocchi ($factor2$) riportando le coppie di livelli che differiscono tra loro, ovvero quelle per le quali si può rifiutare l'ipotesi nulla di uguaglianza. Nonostante il nostro interesse riguardi solo i livelli del trattamento osserviamo che nel caso dei blocchi il test di Bonferroni è meno efficiente del test di Tukey in quanto non considera come significativa la differenza $\{4,5\}$, che invece è contemplata nel test di Tukey, questo perchè il numero di confronti in questo caso è pari a $\binom{5}{2} = 10$ e quindi il test di Bonferroni perde potenza e diventa troppo cautelativo.

3.1.7 Il disegno per misure ripetute ad una via

Il *disegno sperimentale per misure ripetute ad una via* si può considerare come un caso particolare del *disegno sperimentale a blocchi completamente randomizzato*. E' uno dei disegni sperimentali più usati nelle scienze medico-sanitarie e generalizza il confronto tra le medie di due popolazioni Gaussiane nel caso di campioni appaiati.
Infatti, il disegno sperimentale per misure ripetute è un disegno in cui le misure di una variabile di interesse sono fatte su uno stesso soggetto in due o più occasioni. Ovvero preso un campione di n pazienti, ognuno di questi viene sottoposto a tutti i k trattamenti in esame. Come nel caso di due popolazioni Gaussiane in cui, ad esempio, ogni paziente viene osservato *prima* e *dopo* la somministrazione di un farmaco.
Il motivo principale dell'uso del disegno sperimentale per misure ripetute è dato dalla volontà di controllare la variabilità tra i soggetti; infatti utilizzando gli stessi soggetti su ogni trattamento la variabilità diminuisce.
Questo tipo di disegno sperimentale è molto utilizzato quando si è interessati a vedere ri-

sposte ripetute nel tempo. Inoltre è molto impiegato quando si hanno a disposizione pochi soggetti (o è costoso reperirli). Infatti è bene notare che in un disegno sperimentale per misure ripetute è necessaria una dimensione campionaria inferiore rispetto ad un disegno sperimentale con una sola rilevazione. Se ad esempio abbiamo 5 trattamenti o 5 istanti temporali e in ognuno di essi desideriamo avere 10 osservazioni, se il disegno sperimentale fosse ad un unica rilevazione, come quello completamente randomizzato, oltre a rinunciare a controllare la variabilità tra i soggetti, necessiteremmo di 50 pazienti, mentre nel disegno per misure ripetute sono sufficienti 10 pazienti.
Questo tipo di disegno sperimentale presenta anche alcuni svantaggi, ai quali bisogna fare attenzione e che talvolta portano a preferire disegni sperimentali diversi, come quello completamente randomizzato (quando le unità sperimentali sono omogenee).
Uno dei più rilevanti difetti è l'*effetto trascinamento*; quando si devono valutare più trattamenti, il ricercatore deve verificare che la risposta di un soggetto ad un trattamento non rifletta la risposta data al trattamento precedente.
Un altro possibile problema è l'*effetto posizione*; la risposta di un soggetto a un trattamento che è messo in ultima posizione, potrebbe essere diversa da quella data, nel caso in cui questo fosse stato messo in prima posizione. Se, ad esempio, pensiamo a studi in cui è necessaria la partecipazione fisica dei soggetti, è possibile che l'entusiasmo iniziale si trasformi in disinteresse alla fine dell'esperimento. Una maniera per ovviare a questo problema è quella di randomizzare la sequenza con cui vengono somministrati i trattamenti, indipendentemente dai soggetti. Quindi, quando il processo di randomizzazione è indipendente per ogni soggetto, è preferibile che l'ordine con il quale i soggetti vengono sottoposti ai trattamenti sia casuale.
Una volta analizzato il disegno sperimentale, costruiamo il modello. Abbiamo k trattamenti e n pazienti e vogliamo eliminare la variabilità dovuta ai soggetti, pertanto consideriamo l'*analisi della varianza ad una via con blocco* dove i blocchi in questo caso sono i pazienti. Possiamo rappresentare i dati nel modo seguente:

Trattamenti						
Blocchi	1	2	...	k	Totale	Media
1	x_{11}	x_{12}	...	x_{1k}	$T_{1.}$	$\bar{X}_{1.}$
2	x_{21}	x_{22}	...	x_{2k}	$T_{2.}$	$\bar{X}_{2.}$
3	x_{31}	x_{32}	...	x_{3k}	$T_{3.}$	$\bar{X}_{3.}$
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
n	x_{n1}	x_{n2}	...	x_{nk}	$T_{n.}$	$\bar{X}_{n.}$
Totale	$T_{.1}$	$T_{.2}$	...	$T_{.k}$	$T_{..}$	
Media	$\bar{X}_{.1}$	$\bar{X}_{.2}$	...	$\bar{X}_{.k}$		$\bar{X}_{..}$

dove i dati hanno lo stesso significato di quelli studiati nell'*analisi della varianza ad una via con blocco* con la differenza che l'indice di riga rappresenta il particolare paziente. Cioè il primo blocco è costituito dal primo paziente che viene sottoposto a tutti i k trattamenti, il secondo blocco è costituito da secondo paziente e così via. Ma nulla cambia nella costruzione del modello (*modello additivo ad effetti fissi*). La scomposizione di una qualsiasi osservazione

è quindi:

$$x_{ij} = \mu + \beta_i + \tau_j + e_{ij}$$
$$i = 1, 2...n \quad e \quad j = 1, 2, ...k$$

dove β_i è l'effetto dovuto ai soggetti.
Le assunzioni previste per il *disegno sperimentale additivo ad effetti fissi* sono:

1. ogni osservazione x_{ij} è un campione casuale semplice di dimensione 1 proveniente da kn popolazioni;
2. le kn popolazioni hanno medie differenti, ma tutte la medesima varianza;
3. i k trattamenti sono fissi, o meglio sono gli unici trattamenti di interesse statistico per l'esperimento;
4. non c'è alcuna interazione tra i trattamenti e i soggetti, ovvero l'effetto dei trattamenti e l'effetto dei soggetti sono additivi.

Una volta riconosciuto questo modello, come quello del disegno sperimentale a blocchi completamente randomizzato, trattato nel paragrafo precedente, dove i blocchi sono i soggetti è chiaro che la procedura della verifica delle ipotesi è uguale a quella già discussa.

Riportiamo ora un esempio di problema che fa uso di questo disegno sperimentale:

Problema 3.7. *E' stato condotto uno studio in cui si sono esaminati soggetti con dolore lombare cronico non specifico. In questo studio,* 18 *soggetti hanno ricevuto una rudimentale manipolazione osteopatica e hanno compilato un questionario di valutazione della funzionalità fisica all'inizio e dopo* 1, 3 *e* 6 *mesi. Valori più alti indicano una migliore funzionalità fisica. Lo scopo dell'esperimento era quello di determinare se i soggetti manifestassero miglioramenti nel tempo, anche miglioramenti minimi. Desideriamo quindi sapere se c'è differenza nelle medie dei valori nei quattro istanti di tempo.*

Per risolvere il problema facciamo uso del software Mathematica, allo stesso modo di come abbiamo fatto nel caso di un disegno sperimentale a blocchi completamente randomizzato, con la sola differenza che in questo caso i blocchi, ovvero i livelli del $factor2$ sono $n = 18$ in quanto corrispondono ai soggetti in studio e $k = 4$, ovvero i quattro istanti temporali che corrispondono ai livelli del trattamento.
Riportiamo l'input e l'output di Mathematica:

Needs["ANOVA"]

dataset3 = {{1, 1, 80}, {1, 2, 95}, {1, 3, 65}, {1, 4, 50}, {1, 5, 60}, {1, 6, 70}, {1, 7, 80}, {1, 8, 70}, {1, 9, 80},
{1, 10, 65}, {1, 11, 60}, {1, 12, 50}, {1, 13, 50}, {1, 14, 85}, {1, 15, 50}, {1, 16, 15}, {1, 17, 10}, {1, 18, 80},
{2, 1, 60}, {2, 2, 90}, {2, 3, 55}, {2, 4, 45}, {2, 5, 75}, {2, 6, 70}, {2, 7, 80}, {2, 8, 60}, {2, 9, 80},
{2, 10, 30}, {2, 11, 70}, {2, 12, 50}, {2, 13, 65}, {2, 14, 45}, {2, 15, 65}, {2, 16, 30}, {2, 17, 15}, {2, 18, 85},
{3, 1, 95}, {3, 2, 95}, {3, 3, 50}, {3, 4, 70}, {3, 5, 80}, {3, 6, 75}, {3, 7, 85}, {3, 8, 75}, {3, 9, 70},
{3, 10, 45}, {3, 11, 95}, {3, 12, 70}, {3, 13, 80}, {3, 14, 85}, {3, 15, 90}, {3, 16, 20}, {3, 17, 55}, {3, 18, 90},
{4, 1, 100}, {4, 2, 95}, {4, 3, 45}, {4, 4, 70}, {4, 5, 85}, {4, 6, 70}, {4, 7, 80}, {4, 8, 65}, {4, 9, 65},
{4, 10, 60}, {4, 11, 80}, {4, 12, 60}, {4, 13, 65}, {4, 14, 80}, {4, 15, 70}, {4, 16, 25}, {4, 17, 75}, {4, 18, 70}};

$k = 4$;

$n = 18$;

alfa = 0.05;

quant = Quantile[FRatioDistribution[$k-1, (n-1)(k-1)$], 1 − alfa];

Print["La soglia in valore assoluto della regione critica per T=FRatio sui trattamenti è"];

Print[quant]

ANOVA[dataset3, {factor1, factor2}, {factor1, factor2}]

La soglia in valore assoluto della regione critica per T=FRatio sui trattamenti è 2.78623

		DF	SumOfSq	MeanSq	FRatio	PValue
ANOVA →	factor1	3	2395.83	798.611	5.50084	0.00237118
	factor2	17	20237.5	1190.44	8.19977	2.1787542102330235*^-9
	Error	51	7404.17	145.18		
	Total	71	30037.5			

CellMeans →	All	66.25
	factor1[1]	61.9444
	factor1[2]	59.4444
	factor1[3]	73.6111
	factor1[4]	70.
	factor2[1]	83.75
	factor2[2]	93.75
	factor2[3]	53.75
	factor2[4]	58.75
	factor2[5]	75.
	factor2[6]	71.25
	factor2[7]	81.25
	factor2[8]	67.5
	factor2[9]	73.75
	factor2[10]	50.
	factor2[11]	76.25
	factor2[12]	57.5
	factor2[13]	65.
	factor2[14]	73.75
	factor2[15]	68.75
	factor2[16]	22.5
	factor2[17]	38.75
	factor2[18]	81.25

Osserviamo che, poichè il valore $FRatio = 5.50084$ è superiore al valore soglia $F_{1-\alpha} = 2.78623$, rifiutiamo l'ipotesi nulla, ovvero concludiamo che esiste una differenza tra le medie delle quattro popolazioni. Inoltre il valore di p è $p < 0{,}005$.

Una volta rifiutata l'ipotesi nulla, è lecito chiedersi quali coppie di medie sono effettivamente

diverse tra loro. Per rispondere a questa domanda effettuiamo un test di Tukey o un test di Bonferroni, in quanto i livelli del trattamento sono $k = 4$, pertanto le coppie a confronto sono $\binom{4}{2} = 6$ e quindi in numero minore di 10. Osserviamo però che nel confronto sul $factor2$, che corrisponde ai blocchi, il test di Bonferroni rischia di essere troppo cautelativo e di non rilevare come significative delle differenze che invece lo sono, in quanto in questo caso il numero di confronti da effettuare è $\binom{18}{2}$ e quindi di molto superiore a 10. Sebbene sia utile osservare ciò, per rendersi conto della differenza tra il test di Tukey e il test di Bonferroni, in questo caso non è rilevante in quanto il nostro interesse è posto unicamente sui livelli del trattamento, poichè stiamo effettuando una analisi della varianza ad una via.
Riportiamo sia l'input che l'output di Mathematica, ricordando che il nostro interesse è solo sui livelli del trattamento ($factor1$) e non sugli individui, che corrispondono al blocco ($factor2$).

ANOVA[dataset3, {factor1, factor2}, {factor1, factor2}, PostTests → Tukey, CellMeans → False]

$$
\left\{
\begin{array}{lllllll}
 & & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\
 & \text{factor1} & 3 & 2395.83 & 798.611 & 5.50084 & 0.00237118 \\
\text{ANOVA} \to & \text{factor2} & 17 & 20237.5 & 1190.44 & 8.19977 & 2.1787542102330235\text{*\^{}-9} \\
 & \text{Error} & 51 & 7404.17 & 145.18 & & \\
 & \text{Total} & 71 & 30037.5 & & &
\end{array}
\right\}
$$

$$
\left\{\text{PostTests} \to \left\{\text{factor1} \to \quad \text{Tukey} \quad \{\{1,3\},\{2,3\}\} \right\}\right\}
$$

$$
\left\{
\begin{array}{l}
\text{PostTests} \to \left\{
\begin{array}{ll}
 & \text{Tukey} \\
 & \{\{2,3\},\{2,4\},\{1,10\},\{2,10\},\{2,12\},\{1,16\},\{2,16\},\{4,16\}, \\
\text{factor2} \to & \{5,16\},\{6,16\},\{7,16\},\{8,16\},\{9,16\},\{11,16\},\{12,16\}, \\
 & \{13,16\},\{14,16\},\{15,16\},\{1,17\},\{2,17\},\{5,17\},\{6,17\},\{7,17\},\{9,17\}, \\
 & \{11,17\},\{14,17\},\{16,18\},\{17,18\}\}
\end{array}
\right\}
\end{array}
\right\}
$$

ANOVA[dataset3, {factor1, factor2}, {factor1, factor2}, PostTests → Bonferroni, CellMeans → False]

$$
\left\{
\begin{array}{llllllll}
 & & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} & \\
 & \text{factor1} & 3 & 2395.83 & 798.611 & 5.50084 & 0.00237118 & \\
\text{ANOVA} \rightarrow & \text{factor2} & 17 & 20237.5 & 1190.44 & 8.19977 & 2.1787542102330235\text{*\^{}-9} & \\
 & \text{Error} & 51 & 7404.17 & 145.18 & & & \\
 & \text{Total} & 71 & 30037.5 & & & &
\end{array}
\right\}
$$

$$
\left\{ \text{PostTests} \rightarrow \left\{ \text{factor1} \rightarrow \text{Bonferroni} \quad \{\{1,3\},\{2,3\}\} \right\} \right\}
$$

$$
\left\{ \text{PostTests} \rightarrow \left\{ \text{factor2} \rightarrow \begin{array}{l} \text{Bonferroni} \\ \{\{2,3\},\{2,4\},\{1,10\},\{2,10\},\{2,12\},\{1,16\},\{2,16\},\{4,16\}, \\ \{5,16\},\{6,16\},\{7,16\},\{8,16\},\{9,16\},\{11,16\},\{12,16\},\{13,16\}, \\ \{14,16\},\{15,16\},\{1,17\},\{2,17\},\{5,17\},\{7,17\},\{9,17\},\{11,17\}, \\ \{14,17\},\{16,18\},\{17,18\}\} \end{array} \right\} \right\}
$$

Osservazione 3.8. Come abbiamo già brevemente accennato, il test ANOVA ad una via per misure ripetute, nel caso di due soli livelli di un trattamento, coincide con il test t-student sulla differenza tra medie per due popolazioni Gaussiane con campioni appaiati. Per convincerci di ciò dimostriamo che:

$$T_F = T^2,$$

cioè che la statistica usata nel test di Fisher è uguale al quadrato della statistica utilizzata nel test t-student. Se proviamo ciò abbiamo la tesi, in quanto abbiamo già osservato (nel caso di campiono non appaiati e ANOVA con disegno sperimentale completamente randomizzato) che: $t^2(n-1) = F(1, n-1)$ cioè che il quadrato di una distribuzione t-student con rispettivi gradi di libertà è uguale a una distribuzione di Fisher con un grado di libertà al numeratore e gli stessi gradi di libertà al denominatore della distribuzione t-student corrispondente.
Dimostriamo allora che:

$$\frac{MSTr}{MSE} = \left(\sqrt{n}\frac{\bar{Z}}{S_Z}\right)^2$$

In questo caso abbiamo due livelli del trattamento, che corrispondono alle due popolazioni Gaussiane e n blocchi che corrispondono ai pazienti. La tabella delle osservazioni può essere riassunta così:

Trattamenti			
Blocchi	1	2	Media
1	x_1	y_1	$\frac{x_1+y_1}{2}$
2	x_2	y_2	$\frac{x_2+y_2}{2}$
3	x_3	y_3	$\frac{x_3+y_3}{2}$
...	...	...	...
...	...	...	...
...	...	...	...
...	...	...	...
n	x_n	y_n	$\frac{x_n+y_n}{2}$
Media	$\bar{X}$	$\bar{Y}$	$\frac{\bar{X}+\bar{Y}}{2}$

Inoltre abbiamo che:

1. $z_i = x_i - y_i$
2. $\bar{Z} = \bar{X} - \bar{Y}$.

A questo punto calcoliamo

$$T_f = \frac{MSTr}{MSE} = \frac{SSTr}{\frac{SSE}{n-1}},$$

Nel nostro caso:

$$SSTr = n\left(\bar{X} - \frac{\left(\bar{X}+\bar{Y}\right)}{2}\right)^2 + n\left(\bar{Y} - \frac{\left(\bar{X}+\bar{Y}\right)}{2}\right)^2$$

$$= \frac{n}{4}\left(\bar{X}-\bar{Y}\right)^2 + \frac{n}{4}\left(\bar{Y}-\bar{X}\right)^2 = \frac{n}{2}\left(\bar{X}-\bar{Y}\right)^2 = \frac{n}{2}\bar{Z}^2;$$

$$SSE = \sum_{i=1}^{n}\left(x_i - \frac{x_i+y_i}{2} - \bar{X} + \frac{\bar{X}+\bar{Y}}{2}\right)^2 + \sum_{i=1}^{n}\left(y_i - \frac{x_i+y_i}{2} - \bar{Y} + \frac{\bar{X}+\bar{Y}}{2}\right)^2$$

$$= \frac{1}{4}\sum_{i=1}^{n}\left((x_i-y_i) - \left(\bar{X}-\bar{Y}\right)\right)^2 + \frac{1}{4}\sum_{i=1}^{n}\left(-(x_i-y_i) + \left(\bar{X}-\bar{Y}\right)\right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{n}\left((x_i-y_i) - \left(\bar{X}-\bar{Y}\right)\right)^2$$

Pertanto si ha che:

$$MSE = \frac{SSE}{n-1} = \frac{1}{2}S_Z^2$$

In conclusione:

$$T_F = n\frac{\bar{Z}^2}{S_Z^2} = T^2,$$

come si voleva dimostrare.

Proponiamo ora un esempio che ci permetterà di verificare sperimentalmente quanto abbiamo dimostrato. Utilizzando il software Mathematica, risolviamo quindi il seguente problema sia con l'ANOVA ad una via che con un test t-student per campioni appaiati:

Problema 3.9. *L'obiettivo di uno studio era di valutare l'effetto della terapia con interluchina-2 (IL-2) somministrata ad intermittenza per aumentare in modo significativo il numero di linfociti T CD4 combinata con una terapia intervirale altamente attiva (Highly Active Antiretroviral Therapy, HAART). La seguente tabella mostra il numero di cellule T CD4 all'inizio e dopo 12 mesi di terapia HAART con IL-2. Questi dati, indicano ad un livello $\alpha = 0{,}05$, un cambiamento significativo nel numero di cellule T CD4?*

Pazienti	**T CD4** $\times 10^6$ **prima**	**T CD4** $\times 10^6$ **dopo**
1	173	257
2	58	108
3	103	315
4	181	362
5	105	141
6	301	549
7	169	369

datax $= \{173., 58., 103., 181., 105., 301., 169.\}$;

datay $= \{257., 108., 315., 362., 141., 549., 369.\}$;

alfa $= 0.05$; $n = 7$;

dataz $=$ **datax** $-$ **datay**;

sz $=$ **Variance[dataz]**;

MX $=$ **Mean[datax]**;

MY $=$ **Mean[datay]**;

MZ $=$ **Mean[dataz]**;

$T = \frac{\sqrt{n}\,\mathrm{MZ}}{\sqrt{\mathrm{sz}}}$;

Tf $= T^2$;

quant $=$ **Quantile** $\left[\mathrm{StudentTDistribution}[n-1], 1 - \frac{\mathrm{alfa}}{2}\right]$;

$p = 2(1 - \mathrm{CDF}[\mathrm{StudentTDistribution}[n-1], -T])$;

Print["la media di X vale MX"**]**;

Print[MX];

Print["la media di Y vale MY"**]**;

Print[MY];

Print[“la statistica test T vale”];

Print[*T*];

Print[“La soglia in valore assoluto della regione critica per T è”];

Print[quant]

Print[“Il livello di significatività p è”];

Print[*p*];

Print[“la statistica test al quadrato Tf vale”];

Print[Tf];

Print[“la soglia della regione critica per Tf vale”];

Print [quant^2] ;

la media di X vale MX
155.714
la media di Y vale MY
300.143
la statistica test T vale
-4.46001
La soglia in valore assoluto della regione critica per T è
2.44691
Il livello di significatività p è
0.00428297
la statistica test al quadrato Tf vale
19.8917
la soglia della regione critica per Tf vale
5.98738

Needs[“ANOVA”]

dataset = {{1, 1, 173}, {1, 2, 58}, {1, 3, 103}, {1, 4, 181}, {1, 5, 105}, {1, 6, 301},

{1, 7, 169}, {2, 1, 257}, {2, 2, 108}, {2, 3, 315}, {2, 4, 362}, {2, 5, 141},

{2, 6, 549}, {2, 7, 369}};

***k* = 2;**

***n* = 7;**

alfa = 0.05;

quant = Quantile[FRatioDistribution[*k* − 1, (*n* − 1)(*k* − 1)], 1 − alfa];

Print[

"La soglia in valore assoluto della regione critica per T=FRatio sui trattamenti è"];

Print[quant]

ANOVA[dataset, {factor1, factor2}, {factor1, factor2}]

La soglia in valore assoluto della regione critica per T=FRatio sui trattamenti è 5.98738

ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
factor1	1	73008.6	73008.6	19.8917	0.00428297
factor2	6	149924.	24987.4	6.80798	0.0171788
Error	6	22021.9	3670.31		
Total	13	244955.			

CellMeans →

All	227.929
factor1[1]	155.714
factor1[2]	300.143
factor2[1]	215.
factor2[2]	83.
factor2[3]	209.
factor2[4]	271.5
factor2[5]	123.
factor2[6]	425.
factor2[7]	269.

Da questo esempio è evidente che i due test sono identici, in quanto $T^2 = T_F = FRatio$ e la soglia della regione critica per $Fratio$ è esattamente il doppio della soglia della regione critica nel test t-student.
Inoltre possiamo concludere che, poichè $FRatio = 19.8917$ è maggiore del valore soglia $F_{1-\alpha} = 5.98738$, o analogamente che poichè il valore assoluto di $T = 4.46001$ è maggiore del valore soglia $t_{1-\frac{\alpha}{2}} = 2.44691$, rifiutiamo l'ipotesi nulla di uguaglianza tra le medie dei livelli dei trattamenti. Pertanto possiamo concludere che ad un livello α questi dati indicano un cambiamento significativo nel numero di cellule T CD4. Inoltre poichè il valore di p è $p = 0.00428297$ saremo giunti alla stessa conclusione anche con un livello di $\alpha < 0{,}005$, quindi il test è molto significativo.

3.2 L'analisi della varianza ANOVA a due vie

Nei disegni sperimentali, che abbiamo considerato fino a questo punto, abbiamo fissato l'attenzione sugli effetti di una solo variabile: *il trattamento*. Invece, capita spesso di essere

interessati all'effetto simultaneo di due o più variabili, queste variabili di interesse vengono chiamate *fattori*. Quando l'interesse è posto su due o più fattori si parla di *analisi della varianza a due vie*.
Le diverse categorie dei fattori sono denominate *livelli* e il tipo di disegno sperimentale è *l'esperimento fattoriale*. Per esempio se abbiamo intenzione di studiare l'effetto sui tempi di reazione di tre dosaggi di un farmaco e vogliamo che nello studio siano inclusi due gruppi distinti di pazienti; uno con età inferiore ai 65 anni e l'altro con età maggiore o uguale a 65 anni, è chiaro che dobbiamo utilizzare un esperimento fattoriale. In questo caso abbiamo due fattori: il fattore farmaco che si presenta su 3 livelli e il fattore età, che si presenta su due livelli. In generale, diciamo che il fattore A si presenta con a livelli e il fattore B con b livelli. Inoltre in un esperimento fattoriale potremmo essere interessati, non solo allo studio degli effetti dei fattori presi in maniera individuale, ma anche, all'*interazione* tra i fattori.

Definizione 3.10 (**interazione**). *Diciamo che c'è interazione tra due fattori se un cambiamento in uno dei due fattori produce un cambiamento ad un livello dell'altro fattore diverso da quello prodotto negli altri livelli dello stesso fattore.*

Un esperimento fattoriale può essere studiato attraverso disegni sperimentali già discussi.

3.2.1 Il disegno sperimentale completamente randomizzato a due fattori

Pensiamo di trovarci di fronte a un problema del seguente tipo:

Problema 3.11. *In uno studio sulla durata delle visite domiciliari da parte di infermerie pubbliche, è stato costruito un registro in cui veniva annotata l'età dell'infermiere e e la malattia del paziente. I ricercatori vorrebbero avere una risposta, da questi dati, alle seguenti domande:*

1. *Può esserci differenza nella durata media delle visite domiciliari a seconda delle età delle infermiere?*

2. *Può il tipo di paziente avere effetto sulla durata media della visita?*

3. *C'è interazione tra l'età dell'infermiera e il tipo di paziente?*

Questo problema prevede 3 domande e quindi 3 ipotesi da saggiare e due tipi di trattamenti: l'età delle infermiere e il tipo di malattia del paziente.
Una prima idea per risolvere questo problema potrebbe essere quella di condurre due test diversi ognuno dei quali analizzi singolarmente uno dei due fattori, ma ciò sarebbe svantaggioso in quanto richiederebbe molto tempo e non permetterebbe di studiare l'interazione tra i fattori.
Per questo si preferisce utilizzare l'esperimento fattoriale che permette di rispondere a tutte le domande analizzando tutte le osservazioni in una sola volta. Il disegno sperimentale più comune e relativo al nostro problema è quello *completamente randomizzato*, in quanto i livelli dei trattamenti in studio sono associati in modo totalmente casuale ai pazienti.

- Riassumiamo i dati del disegno sperimentale completamente randomizzato a due fattori nel modo seguente:

	Fattore B					
Fattore A	1	2	...	b	Totali	Medie
1	x_{111} $\vdots$ x_{11n}	x_{121} $\vdots$ x_{12n}	... $\vdots$ ...	x_{1b1} $\vdots$ x_{1bn}	$T_{1..}$	$\bar{X}_{1..}$
2	x_{211} $\vdots$ x_{21n}	x_{221} $\vdots$ x_{22n}	... $\vdots$ ...	x_{2b1} $\vdots$ x_{2bn}	$T_{2..}$	$\bar{X}_{2..}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	x_{a11} $\vdots$ x_{a1n}	x_{a21} $\vdots$ x_{a2n}	... $\vdots$ ...	x_{ab1} $\vdots$ x_{abn}	$T_{a..}$	$\bar{X}_{a..}$
Totali	$T_{.1.}$	$T_{.2.}$	...	$T_{.b.}$	$T_{...}$	
Medie	$\bar{X}_{.1.}$	$\bar{X}_{.2.}$	...	$\bar{X}_{.b.}$		$\bar{X}_{...}$

Abbiamo a livelli del fattore A, b livelli del fattore B e n osservazioni per ogni combinazione dei livelli (nel nostro studio consideriamo il caso in cui il numero delle osservazioni in ogni cella è uguale. Quando ciò non accade l'analisi diventa più complicata e il disegno viene definito *non bilanciato*). Il pedice i rappresenta il livello del fattore A, il pedice j il livello del fattore B e il pedice k l'ordine all'interno di ogni cella. Possiamo ragionare come se ognuna delle ab combinazioni dei livelli del fattore A con i livelli del fattore B rappresentasse un trattamento. In questo modo abbiamo ab trattamenti, ognuno con n osservazioni e pertanto il totale dei dati è abn.
Analizziamo i dati riportati in tabella:

$T_{i..} = \sum_{k=1}^{n} \sum_{j=1}^{b} x_{ijk}$ è il totale del trattamento i-esimo;

$\bar{X}_{i..} = \frac{1}{nb} \sum_{k=1}^{n} \sum_{j=1}^{b} x_{ijk} = \frac{T_{i..}}{nb}$ è la media del trattamento i-esimo;

$T_{.j.} = \sum_{k=1}^{n} \sum_{i=1}^{a} x_{ijk}$ è il totale del trattamento j-esimo;

$\bar{X}_{.j.} = \frac{1}{na} \sum_{k=1}^{n} \sum_{i=1}^{a} x_{ijk} = \frac{T_{.j.}}{na}$ è la media del trattamento j-esimo;

$T_{...} = \sum_{j=1}^{b} T_{.j.} = \sum_{i=1}^{a} T_{i..}$ è il totale generale;

$\bar{X}_{...} = \frac{1}{a} \sum_{i=1}^{a} \bar{X}_{i..} = \frac{1}{b} \sum_{j=1}^{b} \bar{X}_{.j.}$ è la media generale che si può ottenere sia sommando per riga che per colonna.
Aggiungiamo anche il totale e la media della cella j-esima:

$T_{ij.} = \sum_{k=1}^{n} x_{ijk}$

$\bar{X}_{ij.} = \frac{T_{ij.}}{n}$.

La tabella utilizzata può anche essere considerata come una generalizzazione di quella usata nel *disegno a blocchi completamente randomizzato*, quando si considera la prima osservazione di ogni cella come appartenente al blocco relativo. E' bene notare che a differenza del disegno sperimentale a blocchi completamente randomizzato, l'esperimento fattoriale necessita almeno di due osservazioni per ogni cella, perchè lo sperimentatore potrebbe essere interessato alle interazioni.
Osserviamo che in caso di assenza di interazione tra i fattori e nel caso in cui siamo interessati a saggiare l'ipotesi nulla su uno solo dei fattori, in quanto l'altro si considera come *blocco* e non come variabile di interesse, l'esperimento e la discussione risulta analoga al *test ANOVA ad una via con blocco*.
Inoltre nel caso in cui si possa assumere assenza di interazione tra i fattori, ma si vogliono considerare entrambi come variabili, il modello utilizzato è analogo a quello che stiamo trattando, ma senza l'effetto interazione e per ogni cella è sufficiente considerare una sola osservazione.

- Scomponiamo ora una qualsiasi osservazione x_{ijk} secondo un *modello ad effetti fissi*:

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ij}$$

$$i = 1, 2, ...a; \quad j = 1, 2, ...b; \quad k = 1, 2, ...n$$

Dove:

1. μ è una costante incognita;
2. $\alpha_i = \mu_i - \mu$ rappresenta l'effetto del fattore A;
3. $\beta_j = \mu_j - \mu$ rappresenta l'effetto del fattore B;
4. $(\alpha\beta)_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$ rappresenta l'effetto dell'interazione tra A e B;
5. $e_{ijk} = x_{ijk} - \mu_{ij}$ rappresenta l'errore sperimentale.

E' facile mostrare che in questo modo si ha una identità:

$$x_{ijk} = \mu + (\mu_i - \mu) + (\mu_j - \mu) + (\mu_{ij} - \mu_i - \mu_j + \mu) + (x_{ijk} - \mu_{ij}),$$

osserviamo che se non c'è interazione tra i trattamenti, ovvero $(\alpha\beta)_{ij} = 0$ allora $\mu = \mu_i + \mu_j - \mu_{ij}$ come nel caso del *disegno a blocchi completamente randomizzato* in cui si assume assenza di interazione tra blocco e trattamento.

- Le assunzioni per questo modello sono:
  - le osservazioni in ognuna delle ab celle costituiscono un campione casuale indipendente di dimensione n estratto da una popolazione definita attraverso una particolare combinazione dei livelli dei due fattori.
  - Tutte le ab popolazioni sono distribuite normalmente.
  - Le popolazioni hanno tutte la medesima varianza.

- Le ipotesi nulle da saggiare sono tre:

1. $H_0 : \alpha_i = 0 \qquad i = 1, 2, ...a$
   H_A : non tutti gli $\alpha_i = 0$

2. $H_0 : \beta_j = 0 \qquad j = 1, 2, ...b$
   H_A : non tutti i $\beta_j = 0$

3. $H_0 : (\alpha\beta)_{ij} = 0 \qquad i = 1, 2, ...a; \quad j = 1, 2, ...b$
   H_A : non tutti gli $(\alpha\beta)_{ij} = 0$

- Costruiamo ora una statistica test opportuna:
  - analogamente a quanto fatto per un disegno sperimentale completamente randomizzato a un fattore, possiamo ragionare in questo modo: abbiamo nab osservazioni provenienti da ab popolazioni distribuite normalmente, con stessa varianza. Allora se tutte le tre ipotesi sono vere, possiamo definire due diversi stimatori corretti per la varianza:
    1. Il primo stimatore è quello per la varianza all'interno dei gruppi (che nel test ad una via avevamo chiamato MSW), che ora chiamiamo MSE:

$$MSE = \frac{1}{ab(n-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(x_{ijk} - \bar{X}_{ij.}\right)^2 .$$

    2. Il secondo stimatore è quello per la varianza tra i gruppi (che avevamo nominato MSA), che ora nominiamo $MSTr$:

$$MSTr = \frac{1}{ab-1} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(\bar{X}_{ij.} - \bar{X}_{...}\right)^2 .$$

    Inoltre lo stimatore corretto "canonico" della varianza è:

$$MST = \frac{1}{nab-1} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(x_{ijk} - \bar{X}_{...}\right)^2 .$$

    Ora come abbiamo già visto nel caso di un disegno sperimentale completamente randomizzato, si può dimostrare che:

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(x_{ijk} - \bar{X}_{...}\right)^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(\bar{X}_{ij.} - \bar{X}_{...}\right)^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(x_{ijk} - \bar{X}_{ij.}\right)^2 ,$$

    ovvero:

$$SST = SSTr + SSE.$$

    Da qui si costruisce la statistica:

$$T = \frac{MSTr}{MSE} \sim F\left((ab-1), (abn-ab)\right).$$

  - Osserviamo che dalla scomposizione del modello:

$$x_{ijk} = \mu + (\mu_i - \mu) + (\mu_j - \mu) + (\mu_{ij} - \mu_i - \mu_j + \mu) + (x_{ijk} - \mu_{ij}),$$

si ha che:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(x_{ijk}-\bar{X}_{...}\right)^2 =$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left[\left(\bar{X}_{i..}-\bar{X}_{...}\right)+\left(\bar{X}_{.j.}-\bar{X}_{...}\right)+\left(\bar{X}_{ij.}-\bar{X}_{i..}-\bar{X}_{.j.}+\bar{X}_{...}\right)+\left(x_{ijk}-\bar{X}_{ij.}\right)\right]^2 =$$

E poichè i doppi prodotti si annullano abbiamo che:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(x_{ijk}-\bar{X}_{...}\right)^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\bar{X}_{i..}-\bar{X}_{...}\right)^2$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\bar{X}_{.j.}-\bar{X}_{...}\right)^2$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\bar{X}_{ij.}-\bar{X}_{i..}-\bar{X}_{.j.}+\bar{X}_{...}\right)^2$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(x_{ijk}-\bar{X}_{ij.}\right)^2.$$

Ovvero:

$$SST = SSA + SSB + SSAB + SSE,$$

quindi abbiamo scomposto $SSTr$:

$$SSTr = SSA + SSB + SSAB$$

i gradi di libertà delle rispettive somme sono:

$$ab - 1 = (a-1) + (b-1) + (a-1)(b-1).$$

Questa decomposizione è quella più utile perchè permette di saggiare solo alcune delle tre ipotesi, oppure tutte e tre ma in momenti distinti.
Seguendo questo procedimento le statistiche da utilizzare, a seconda della ipotesi che si vuole saggiare sono:

$$T_1 = \frac{MSA}{MSE} \sim F\left((a-1),(abn-ab)\right),$$

$$T_2 = \frac{MSB}{MSE} \sim F((b-1),(abn-ab)),$$

$$T_3 = \frac{MSAB}{MSE} \sim F((a-1)(b-1),(abn-ab)).$$

Queste seguono una distribuzione di Fisher con rispettivi gradi di libertà in quanto assumendo le ipotesi vere, MSA, MSB e $MSAB$ sono stimatori corretti per la varianza σ^2.

- Prima di procedere al test statistico, il ricercatore deve decidere quali delle ipotesi saggiare. Si scelgono le ipotesi e si procede in modo usuale a un livello di significatività α. Quando si decidono di saggiare tutte e tre le ipotesi la situazione è complicata dal fatto che i tre test non sono indipendenti da un punto di vista probabilistico. Se α è il livello di significatività associato ad un test unico comprendente le tre ipotesi e α_1, α_2, α_3 i livelli di significatività associati, rispettivamente alle ipotesi 1, 2 e 3 si ha che:

  $$\alpha < 1 - (1-\alpha_1)(1-\alpha_2)(1-\alpha_3).$$

  Se ad esempio $\alpha_1 = \alpha_2 = \alpha_3 = 0{,}05$ si ha che $\alpha < 0{,}143$. Ciò significa che la probabilità di rifiutare una delle 3 ipotesi quando le tre ipotesi sono tutte vere è minore di 0,143. E' bene tenere conto di ciò quando si devono interpretare i risultati.

- La regola di decisione è quella già nota: si rifiuta la rispettiva H_0 quando il valore della statistica T_i corrispondente è maggiore o uguale del valore critico $F_{1-\alpha}$ con i dovuti gradi di libertà e con i livelli di significatività prescelti. Se rifiutiamo H_0 concludiamo che H_A è vera, se non possiamo rifiutare H_0 concludiamo che H_0 potrebbe essere vera. In conclusione la regione di rigetto delle ipotesi con rispettivi gradi di libertà e livello di significatività è:

  $$D =]F_{1-\alpha}, +\infty[.$$

  Inoltre è bene calcolare il valore di p tale che:

  $$T_{oss} = F_{1-p}.$$

Rispondiamo ora al problema posto inizialmente (3.11), utilizzando il software Mathematica. Per prima cosa rappresentiamo i dati in una lista di liste in cui il primo termine corrisponde al livello del primo trattamento (in Mathematica $factor1$) e il secondo termine indica il livello del secondo trattamento (in Mathematica $factor2$). Indichiamo poi con $a = 4$ e $b = 4$ rispettivamente il numero di livelli del primo fattore (età della infermiera: $20-29$, $30-39$, $40-49$, 50 e oltre) e del secondo fattore (tipo di paziente: cardiopatico, con cancro, con malattie celebrovascolari e con tubercolosi), indichiamo invece con $n = 5$ il numero di osservazioni per ogni coppia di livelli dei trattamenti.
Osserviamo che in questo caso il comando ANOVA richiede come modello $\{factor1, factor2, All\}$, che indica che il modello prevede interazione tra i fattori (trattamenti), All significa che tutte le possibili interazioni sono ammesse.
Riportiamo ora l'input e l'output di Mathematica.

Needs["ANOVA"]

dataset4 = {{1, 1, 20}, {1, 1, 25}, {1, 1, 22}, {1, 1, 27}, {1, 1, 21}, {1, 2, 30}, {1, 2, 45}, {1, 2, 30}, {1, 2, 35},

{1, 2, 36}, {1, 3, 31}, {1, 3, 30}, {1, 3, 40}, {1, 3, 35}, {1, 3, 30}, {1, 4, 20}, {1, 4, 21}, {1, 4, 20},

{1, 4, 20}, {1, 4, 19}, {2, 1, 25}, {2, 1, 30}, {2, 1, 29}, {2, 1, 28}, {2, 1, 30}, {2, 2, 30}, {2, 2, 29},

{2, 2, 31}, {2, 2, 30}, {2, 2, 30}, {2, 3, 32}, {2, 3, 35}, {2, 3, 30}, {2, 3, 40}, {2, 3, 30}, {2, 4, 23},

{2, 4, 25}, {2, 4, 28}, {2, 4, 30}, {2, 4, 31}, {3, 1, 24}, {3, 1, 28}, {3, 1, 24}, {3, 1, 25}, {3, 1, 30},

{3, 2, 39}, {3, 2, 42}, {3, 2, 36}, {3, 2, 42}, {3, 2, 40}, {3, 3, 41}, {3, 3, 45}, {3, 3, 40}, {3, 3, 40},

{3, 3, 35}, {3, 4, 24}, {3, 4, 25}, {3, 4, 30}, {3, 4, 26}, {3, 4, 23}, {4, 1, 28}, {4, 1, 31}, {4, 1, 26},

```
{4,1,29},{4,1,32},{4,2,40},{4,2,45},{4,2,50},{4,2,45},{4,2,60},{4,3,42},{4,3,50},
{4,3,40},{4,3,55},{4,3,45},{4,4,29},{4,4,30},{4,4,28},{4,4,27},{4,4,30}};

a = 4;
b = 4;
n = 5;
ab = 16;
alfa = 0.05;
quant1 = Quantile[FRatioDistribution[b − 1, ab(n − 1)], 1 − alfa];
quant2 = Quantile[FRatioDistribution[a − 1, ab(n − 1)], 1 − alfa];
quant3 = Quantile[FRatioDistribution[(a − 1)(b − 1), ab(n − 1)], 1 − alfa];
Print["La soglia in valore assoluto della regione critica per T=FRatio sul fattore1 è"];
Print[quant1]
Print["La soglia in valore assoluto della regione critica per T=FRatio sul fattore2 è"];
Print[quant2]
Print["La soglia in valore assoluto della regione critica per T=FRatio sull'interazione fattore1fattore2 è"];
Print[quant3]
ANOVA[dataset4, {factor1, factor2, All}, {factor1, factor2}]
```

La soglia in valore assoluto della regione critica per T=FRatio sul fattore1 è

2.74819

La soglia in valore assoluto della regione critica per T=FRatio sul fattore2 è

2.74819

La soglia in valore assoluto della regione critica per T=FRatio sull'interazione fattore1fattore2 è

2.02979

ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
factor1	3	1201.05	400.35	27.2695	1.763398992480387*^-11
factor2	3	2992.45	997.483	67.9427	7.23086200053274*^-20
factor1factor2	9	608.45	67.6056	4.60489	0.000104683
Error	64	939.6	14.6813		
Total	79	5741.55			

CellMeans →	All	32.175
	factor1[1]	27.85
	factor1[2]	29.8
	factor1[3]	32.95
	factor1[4]	38.1
	factor2[1]	26.7
	factor2[2]	38.25
	factor2[3]	38.3
	factor2[4]	25.45
	factor1[1]factor2[1]	23.
	factor1[1]factor2[2]	35.2
	factor1[1]factor2[3]	33.2
	factor1[1]factor2[4]	20.
	factor1[2]factor2[1]	28.4
	factor1[2]factor2[2]	30.
	factor1[2]factor2[3]	33.4
	factor1[2]factor2[4]	27.4
	factor1[3]factor2[1]	26.2
	factor1[3]factor2[2]	39.8
	factor1[3]factor2[3]	40.2
	factor1[3]factor2[4]	25.6
	factor1[4]factor2[1]	29.2
	factor1[4]factor2[2]	48.
	factor1[4]factor2[3]	46.4
	factor1[4]factor2[4]	28.8

Osserviamo che i tre valori di $FRatio$ sono maggiori dei rispettivi valori critici e quindi rifiutiamo tutte le tre ipotesi nulle. Ovvero concludiamo che:

1. Ci sono differenze tra i livelli del primo trattamento, ovvero in media ci sono differenze nella durata delle visite domiciliari a seconda dell'età dell'infermiera;

2. ci sono differenze tra i livelli del secondo trattamento, ovvero in media ci sono differenze nella durata delle visite domiciliari a seconda del tipo di paziente;

3. i due fattori interagiscono, ovvero differenti combinazioni dei livelli dei due fattori producono effetti diversi.

Osserviamo inoltre che i valori di p sono molto bassi e questo implica che avremmo rifiutato le ipotesi anche a livelli α molto più piccoli di 0,05, come invece abbiamo richiesto.
Infine, è bene dire che quando l'ipotesi di nessuna interazione viene rifiutata, l'interesse nei confronti dei livelli dei fattori diventa, in genere, subordinata agli effetti delle interazioni. In altre parole, l'interesse prevalente è rivolto a capire quali combinazioni di livelli sono significativamente diversi tra loro e questo può essere fatto attraverso confronti multipli a posteriori, facendosi prima una idea grossolana osservando i risultati riportati in tabella $CellMeans$.
Vogliamo quindi rispondere alla domanda:

- Sapendo che i due fattori interagiscono, quali differenti combinazioni dei livelli dei due fattori producono effetti differenziati?

Per rispondere a questa domanda possiamo effettuare un test di Tukey sulle interazioni, per fare questo però, dobbiamo ragionare come se avessimo $ab = 16$ livelli di un unico trattamento, (in quanto, per come abbiamo definito il test di Tukey opera su un trattamento alla volta) ognuno con n osservazioni ed effettuare un test ANOVA ad una via a cui segue un test di Tukey sulle $\binom{16}{2}$ coppie di livelli, che corrispondono a tutte le coppie di possibili combinazioni di livelli dei due fattori (è bene osservare che in questo caso preferiamo un test di Tukey al test di Bonferroni in quanto le coppie da confrontare sono in numero ben superiore a 10).

Riportiamo ora l'input e l'output di Mathematica attraverso il quale potremmo rispondere alla nostra domanda. Osserviamo che abbiamo riscritto la lista di liste che descrive le osservazioni inserendo al primo termine il livello considerato (tra gli ab totali) del trattamento, ad esempio 1.1 indica che consideriamo il livello che prende in esame le osservazioni sulla durata delle visite domiciliari nel caso di infermiere di età $20-29$ e pazienti cardiopatici. Abbiamo anche chiesto di vedere i valori di $CellMeans$ per mostrare, come è ovvio, che sono uguali a quelli precedentemente ottenuti sulla coppia di fattori.

$$\mathbf{dataset4 = \{\{1.1,20\},\{1.1,25\},\{1.1,22\},\{1.1,27\},\{1.1,21\},\{1.2,30\},\{1.2,45\},\{1.2,30\},\{1.2,35\},\{1.2,36\},}$$
$$\mathbf{\{1.3,31\},\{1.3,30\},\{1.3,40\},\{1.3,35\},\{1.3,30\},\{1.4,20\},\{1.4,21\},\{1.4,20\},\{1.4,20\},\{1.4,19\},\{2.1,25\},}$$
$$\mathbf{\{2.1,30\},\{2.1,29\},\{2.1,28\},\{2.1,30\},\{2.2,30\},\{2.2,29\},\{2.2,31\},\{2.2,30\},\{2.2,30\},\{2.3,32\},\{2.3,35\},}$$
$$\mathbf{\{2.3,30\},\{2.3,40\},\{2.3,30\},\{2.4,23\},\{2.4,25\},\{2.4,28\},\{2.4,30\},\{2.4,31\},\{3.1,24\},\{3.1,28\},\{3.1,24\},}$$
$$\mathbf{\{3.1,25\},\{3.1,30\},\{3.2,39\},\{3.2,42\},\{3.2,36\},\{3.2,42\},\{3.2,40\},\{3.3,41\},\{3.3,45\},\{3.3,40\},\{3.3,40\},}$$
$$\mathbf{\{3.3,35\},\{3.4,24\},\{3.4,25\},\{3.4,30\},\{3.4,26\},\{3.4,23\},\{4.1,28\},\{4.1,31\},\{4.1,26\},\{4.1,29\},\{4.1,32\},}$$

{4.2, 40}, {4.2, 45}, {4.2, 50}, {4.2, 45}, {4.2, 60}, {4.3, 42}, {4.3, 50}, {4.3, 40}, {4.3, 55}, {4.3, 45}, {4.4, 29}, {4.4, 30}, {4.4, 28}, {4.4, 27}, {4.4, 30}};

ANOVA[dataset4, PostTests → Tukey, CellMeans → True]

ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
Model	15	4801.95	320.13	21.8054	1.5068550061778187*^-19
Error	64	939.6	14.6813		
Total	79	5741.55			

CellMeans →

All	32.175
Model[1.1]	23.
Model[1.2]	35.2
Model[1.3]	33.2
Model[1.4]	20.
Model[2.1]	28.4
Model[2.2]	30.
Model[2.3]	33.4
Model[2.4]	27.4
Model[3.1]	26.2
Model[3.2]	39.8
Model[3.3]	40.2
Model[3.4]	25.6
Model[4.1]	29.2
Model[4.2]	48.
Model[4.3]	46.4
Model[4.4]	28.8

$$\left\{\begin{array}{ll} & \text{Tukey} \\ & \{\{1.1,1.2\},\{1.1,1.3\},\{1.2,1.4\},\{1.3,1.4\},\{1.4,2.2\},\{1.1,2.3\},\{1.4,2.3\},\{1.2,3.1\}, \\ & \{1.1,3.2\},\{1.4,3.2\},\{2.1,3.2\},\{2.2,3.2\},\{2.4,3.2\},\{3.1,3.2\},\{1.1,3.3\},\{1.4,3.3\}, \\ \text{PostTests} \rightarrow & \{2.1,3.3\},\{2.2,3.3\},\{2.4,3.3\},\{3.1,3.3\},\{1.2,3.4\},\{3.2,3.4\},\{3.3,3.4\},\{1.4,4.1\}, \\ & \{3.2,4.1\},\{3.3,4.1\},\{1.1,4.2\},\{1.2,4.2\},\{1.3,4.2\},\{1.4,4.2\},\{2.1,4.2\},\{2.2,4.2\}, \\ & \{2.3,4.2\},\{2.4,4.2\},\{3.1,4.2\},\{3.4,4.2\},\{4.1,4.2\},\{1.1,4.3\},\{1.2,4.3\},\{1.3,4.3\}, \\ & \{1.4,4.3\},\{2.1,4.3\},\{2.2,4.3\},\{2.3,4.3\},\{2.4,4.3\},\{3.1,4.3\},\{3.4,4.3\},\{4.1,4.3\}, \\ & \{1.4,4.4\},\{3.2,4.4\},\{3.3,4.4\},\{4.2,4.4\},\{4.3,4.4\}\} \end{array}\right\}$$

Possiamo ora concludere che le coppie di combinazioni di livelli riportate nella cella *PostTests*, sono quelle che differiscono significativamente.

3.2.2 Esempio

Per capire meglio l'utilizzo della *analisi ANOVA* e del disegno sperimentale corretto da utilizzare, consideriamo un esempio di problema tratto da una situazione reale. I dati analizzati provengono da un centro per lo scompenso cardiaco di Bologna. Il problema può essere presentato nel modo seguente:

Problema 3.12. *Avendo a disposizione la classificazione clinica NYHA (New York Heart Association) dello scompenso cardiaco, che indica quattro classi funzionali in rapporto alle attività che il paziente riesce ad effettuare in base alla dispnea:*

NYHA	***Sintomi***
I	*Sintomi di scompenso, ma senza conseguenti limitazioni dell'attività fisica ordinaria.*
II	*Il paziente sta bene a riposo ma l'attività fisica ordinaria causa la comparsa di sintomi.*
III	*Compaiono sintomi anche per attività fisiche inferiori all'ordinario ma sta bene a riposo.*
IV	*Il paziente non riesce a svolgere alcuna attività e ha sintomi anche a riposo.*

si vuole capire se vi è una relazione effettiva tra questa classificazione e alcuni parametri fisiologico-biologici dei pazienti.

Per affrontare questo problema si è scelto di studiare alcuni dei parametri fisiologico-biologici che generalmente si utilizzano per caratterizzare i pazienti in scompenso cardiaco:

1. *FE*: frazione di eiezione, indice percentuale di volume di sangue espulso dal ventricolo sinistro in sistole, in rapporto al volume ventricolare in telediastole, indicato dalla formula:
$$FE = \frac{V_{td} - V_{ts}}{V_{td}};$$

2. *BNP*: peptide natriuretico di tipo B misurato in $\frac{pg}{mL}$, è un ormone secreto a livello cardiaco dai ventricoli in risposta a stimoli volumetrici o pressioni sulla parete;

3. *PRO-BNP*: è il frammento N-terminale del peptide precursore del BNP.

Per procedere con lo studio abbiamo a disposizione 50 pazienti diversi, di cui:

1. $n_1 = 27$ appartenenti alla classe 1 di *NYHA*;

2. $n_2 = 17$ appartenenti alla classe 2 di *NYHA*;

3. $n_3 = 6$ appartenenti alla classe 3 di *NYHA*;

4. un solo paziente appartenente alla classe 4 e pertanto lo studio è stato condotto solo sulle prime tre classi funzionali.

Per ognuno di questi pazienti abbiamo a disposizione i valori di: FE, BNP e $PRO-BNP$ riportati in tabella:

ID	NYHA	FE	BNP	PRO-BNP
1	1	70	121, 61	457, 9
2	1	65	142, 54	173
3	1	50	140, 3	
4	1	70	100, 48	406, 3
5	1	72	179, 16	281, 2
6	1	50	216, 78	1055
7	1	65	174, 72	981, 2
8	1	45	31, 76	705, 2
9	1	60	92, 94	572, 7
10	1	50	240, 12	1768
11	1	50	32, 64	229, 1
12	1	50	161, 45	1351
13	1	47	32, 5	96, 67
14	1	65	73, 25	
15	1	35	128, 87	634, 6
16	1	76	91, 26	1732
17	1	44	186, 07	1604
18	1	70	84, 71	636, 6
19	1	49	16, 1	216, 7
20	1	61	176, 09	4142
21	1	60	173, 19	957, 9
22	1	71	201, 71	1161
23	1	58	50, 31	970, 2
24	1	65	93, 58	469, 3
25	1	68	39, 43	460, 9
26	1	49	23, 5	237, 1
27	1	60	167, 58	1257
28	2		72	309, 2
29	2	42	315, 56	7116
30	2	30	280, 21	3902
31	2		163, 73	1578
32	2	44	279, 76	2657
33	2	71	237, 67	2471
34	2	30	241, 51	4322
35	2	87	235	1046
36	2		142, 47	551, 4
37	2	50	180, 78	2558
38	2	41	200, 55	967, 8
49	2	51	241, 14	2641
40	2	60	355, 17	1893
41	2	33	236, 24	
42	2		145, 34	2233
43	2	50	212, 73	
44	2	46	198, 98	925, 3
45	3	58	518, 92	119, 2
46	3	51	3006, 5	12995
47	3	30	413, 85	2596
48	3	46	365, 32	2836
49	3	32	573, 55	12732
50	3	23	3416, 8	14537

Per prima cosa cerchiamo di capire come affrontare il problema. La richiesta è di capire se vi è una relazione tra i livelli di $NYHA$ e i parametri analizzati, pertanto lo scopo del problema è quello di concludere:

1. se i tre livelli di *NYHA* differiscono per il parametro *FE* e in particolari quali combinazioni dei tre differiscono effettivamente per questo parametro;
2. se i tre livelli di *NYHA* differiscono per il parametro *BNP* e in particolari quali combinazioni dei tre differiscono effettivamente per questo parametro;
3. se i tre livelli di *NYHA* differiscono per il parametro *PRO-BNP* e in particolari quali combinazioni dei tre differiscono effettivamente per questo parametro.

Pertanto sembra necessario procedere utilizzando tre differenti analisi della varianza ad una via con disegno sperimentale completamente randomizzato considerando le tre classi funzionali di *NYHA* come i tre livelli di un trattamento e distinguendo le osservazioni in questo modo: i valori di *FE* per la prima analisi, i valori di *BNP* per la seconda analisi e i valori di *PRO-BNP* per la terza analisi. A seguito di ciò sarà fondamentale l'utilizzo di test *post-hoc* per capire effettivamente quali gruppi di pazienti differiscono e quindi poter determinare il tipo di relazione che intercorre tra i livelli di $NYHA$ e i parametri studiati.
Prima di procedere con l'ANOVA è però necessario verificare che le ipotesi di normalità e omoshedasticità siano verificate. Quindi dobbiamo poter supporre che le popolazioni da cui provengono i dati sono *prossochè* delle Normali, *pressochè* con stessa varianza. Poichè abbiamo tre parametri (FE, BNP, $PRO - BNP$) che vorremmo trattare singolarmente sui tre livelli di $NYHA$ con una analisi della varianza, dobbiamo poter supporre che:

1. le osservazioni sul parametro FE dei tre livelli di $NYHA$ provengono *pressochè* da tre popolazioni Gaussiane con stessa varianza;
2. le osservazioni sul parametro BNP dei tre livelli di $NYHA$ provengono *pressochè* da tre popolazioni Gaussiane con stessa varianza;
3. le osservazioni sul parametro $PRO - BNP$ dei tre livelli di $NYHA$ provengono *pressochè* da tre popolazioni Gaussiane con stessa varianza.

Dato che le osservazioni (distinte per parametro) su ogni livello non sono in numero sufficientemente grande da poter supporre a priori di lavorare con delle normali (solo quelle relative al livello 1 lo sarebbero in quanto sono circa 30), possiamo procedere con il cosiddetto **grafico quantile-quantile o q-q plot**.
Il **q-q plot** è la rappresentazione grafica dei quantili di una distribuzione. Esso confronta la distribuzione cumulata della variabile osservata con la distribuzione cumulata della normale associata. Vediamo come viene costruito il grafico:
supponiamo di avere dei dati provenienti da una popolazione incognita Y e di voler stabilire se $Y \sim N(\mu, \sigma^2)$. Per prima cosa ordiniamo i dati in modo crescente e li supponiamo tutti distinti, a questo punto associamo alle osservazioni la rispettiva frequenza e la somma delle frequenze $\sum_{j=1}^{i} f_j$, come in tabella:

Y	Fr	$\sum_{j=1}^{i} f_j$
y_1	f_1	f_1
y_2	f_2	$f_1 + f_2$
y_3	f_3	$f_1 + f_2 + f_3$
$\dots$	$\dots$	$\dots$
$\dots$	$\dots$	$\dots$
y_n	f_n	$f_1 + f_2 + \dots f_n$

A questo punto è chiaro che:

$$P\left(Y \leq y_i\right) = \frac{f_1 + ... + f_i}{f_1 + .. + f_n},$$

e poichè supponiamo che le osservazioni siano tutte distinte::

$$P\left(Y \leq y_i\right) = \frac{i}{n},$$

ovvero y_i è il quantile di livello $\frac{i}{n}$.
A questo punto chiamiamo z_i il quantile di pari livello per una $N(0,1)$:

$$z_i = \Phi^{-1}\left(\frac{i}{n}\right).$$

Ora se fosse $Y \sim N(\mu, \sigma^2)$ si avrebbe:

$$\frac{i}{n} = P\left(\frac{Y-\mu}{\sigma} \leq \frac{y_i - \mu}{\sigma}\right) = \Phi\left(\frac{y_i - \mu}{\sigma}\right),$$

ovvero:

$$z_i = \frac{y_i - \mu}{\sigma}$$

e quindi se Y fosse distribuita normalmente le osservazioni y_i dovrebbero essere ottenute come:

$$y_i = \sigma z_i + \mu.$$

Il grafico **q-q plot** è un grafico (z, x) in cui viene rappresentata sia la retta:

$$X = \sigma Z + \mu,$$

dove le x_i sono le osservazioni che Y avrebbe se fosse una $N(\mu, \sigma^2)$ e gli z_i sono i quantili di livello $\frac{i}{n}$, sia le osservazioni y_i di Y.

Quindi per poter concludere che Y sia distribuita *pressochè* normalmente è necessario che le osservazioni si distribuiscano intorno alla retta dei quantili della normale.

E' bene osservare che in letteratura vi sono molte variazioni della più classica plotting position (ovvero l'operazione che permette di trovare una stima del valore teorico di $F(y_i)$) da noi utilizzata: $\frac{i}{n}$. La più comune è quella di *Hazen* che corrisponde a $\frac{i-0,5}{n}$.

Una volta capito come viene costruito il *grafico quantile-quantile* possiamo utilizzarlo nel nostro caso specifico delle 6 distribuzioni relative ai 3 livelli per ogni parametro, per poter concludere la loro normalità.

Osserviamo i grafici prodotti con Mathematica:

data1F = {70, 65, 50, 70, 72, 50, 65, 45, 60, 50, 50, 50, 47, 65, 35, 76, 44, 70, 49, 61, 60, 71, 58, 65, 68, 49, 60};

data2F = {42, 30, 71, 44, 30, 87, 50, 41, 51, 60, 33, 46, 50};

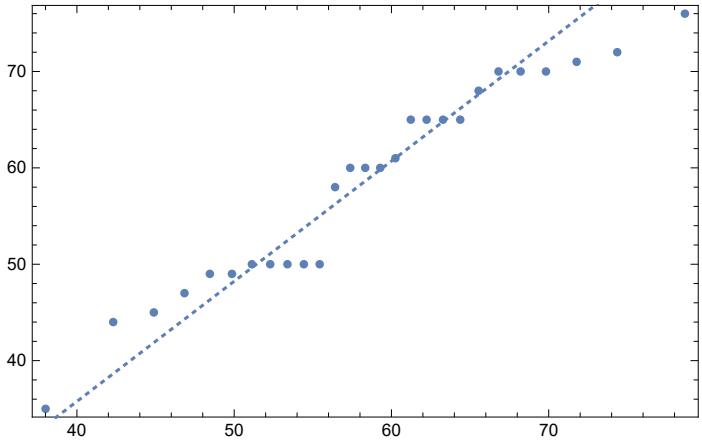
```
data3F = {51, 30, 46, 32, 23, 58};
T1 = Sort[data1F];
T2 = Sort[data2F];
T3 = Sort[data3F];
Print["Livello 1 per FE"]
QuantilePlot[T1]
Print["Livello 2 per FE"]
QuantilePlot[T2]
Print["Livello 3 per FE"]
QuantilePlot[T3]
```

```
data1B = {121.61, 142.54, 140.3, 100.48, 179.16, 216.78, 174.72, 31.76, 92.94, 240.12, 32.64, 161.45,
32.5, 73.25, 128.87, 91.26, 186.07, 173.19, 84.71, 16.1, 176.09, 201.71, 50.31, 93.58, 39.43, 23.5, 167.58};
data2B = {72, 315.56, 280.21, 163.73, 279.76, 237.67, 241.51, 235, 142.47, 180.78, 200.55, 241.14, 355.17,
236.24, 145.34, 212.73, 198.98};
data3B = {518.92, 3006.5, 413.85, 365.32, 573.55, 3416.8};
F1 = Sort[data1B];
F2 = Sort[data2B];
F3 = Sort[data3B];
Print["Livello 1 per BNP"]
QuantilePlot[F1]
Print["Livello 2 per BNP"]
QuantilePlot[F2]
Print["Livello 3 per BNP"]
QuantilePlot[F3]
```

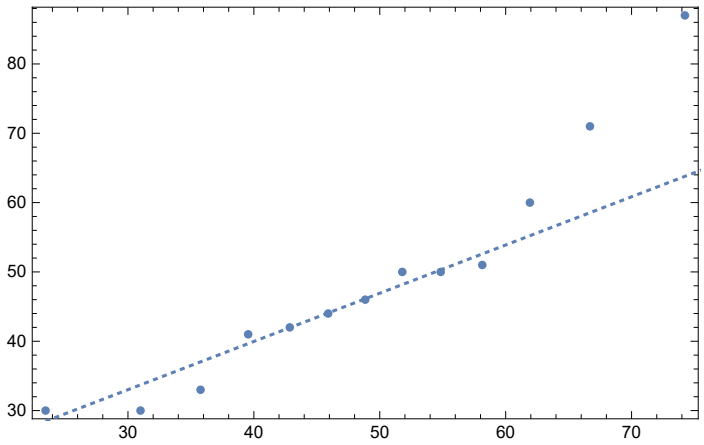
```
data1P = {457.9, 173, 406.3, 281.2, 981.2, 705.2, 572.7, 1768, 229.1, 1351, 1055, 4142, 96.67,
634.6, 1732, 1604, 636.6, 216.7, 957.9, 1161, 970.2, 469.3, 460.9, 237.1};
```

```
data2P = {1257, 309.2, 7116, 3902, 1578, 2657, 2471, 4322, 1046, 551.4, 2558, 967.8, 2641,
1893, 925.3, 2233};
data3P = {119.2, 12995, 2596, 2836, 12732, 14537};
G1 = Sort[data1P];
G2 = Sort[data2P];
G3 = Sort[data3P];
Print["Livello 1 per PRO-BNP"]
QuantilePlot[G1]
Print["Livello 2 per PRO-BNP"]
QuantilePlot[G2]
Print["Livello 3 per PRO-BNP"]
QuantilePlot[G3]
```

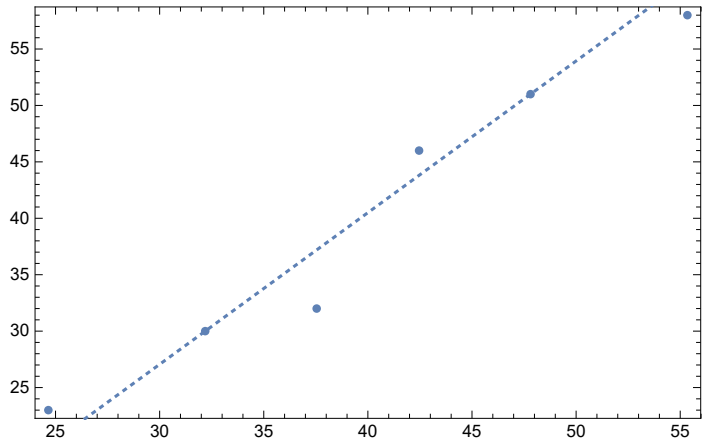
Livello 1 per FE



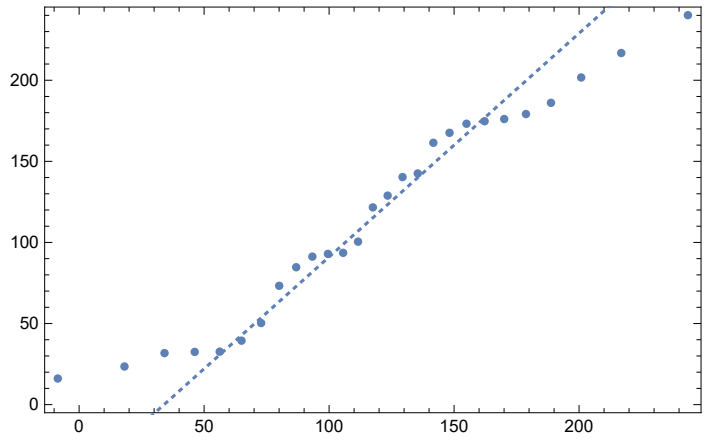
Livello 2 per FE



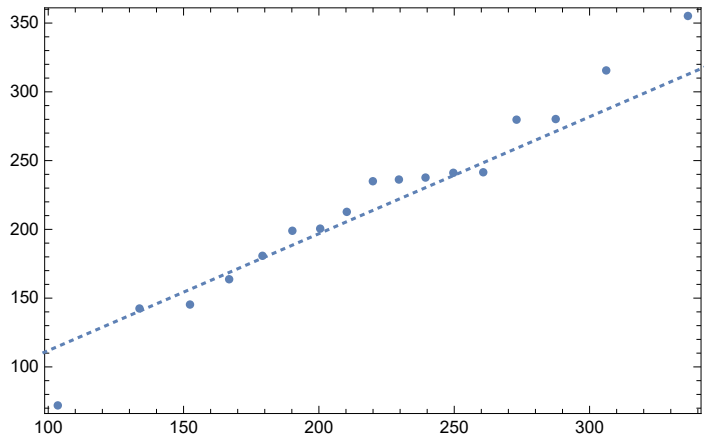
Livello 3 per FE



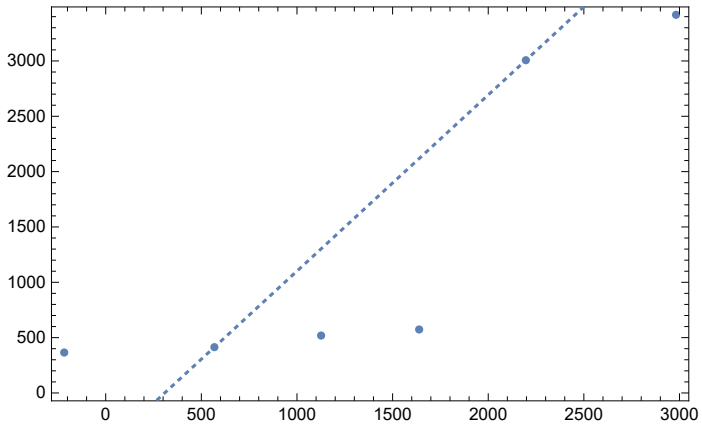
Livello 1 per BNP



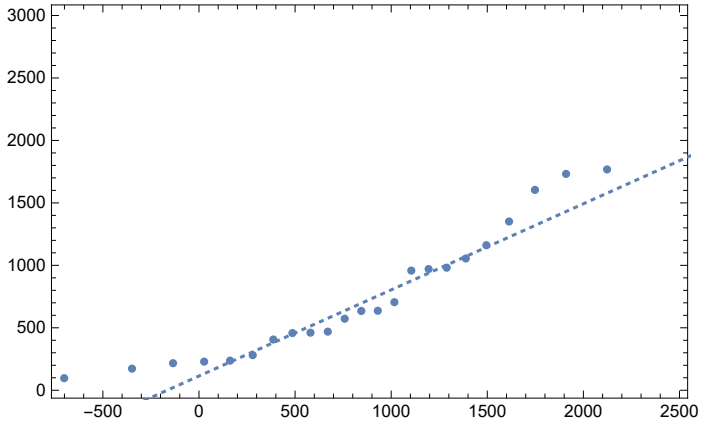
Livello 2 per BNP



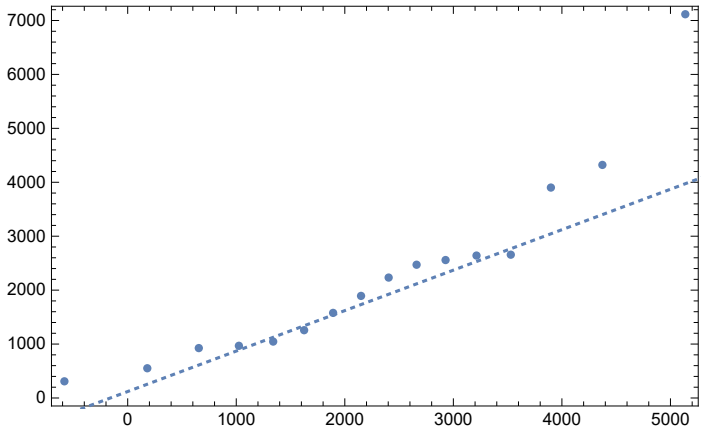
Livello 3 per BNP



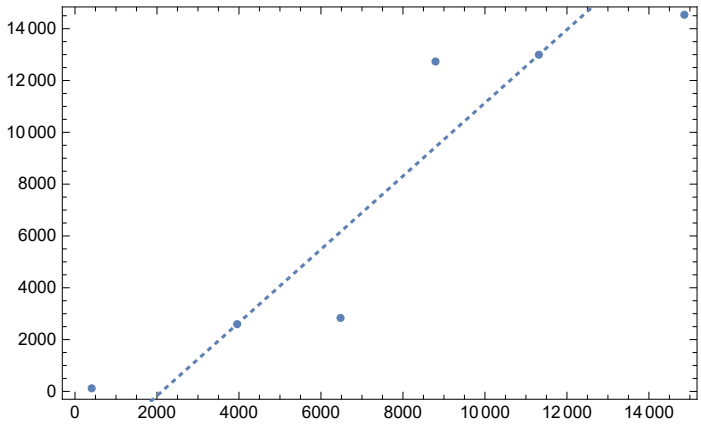
Livello 1 per PRO-BNP



Livello 2 per PRO-BNP



Livello 3 per PRO-BNP



Dai grafici risulta che i dati relativi ai livelli 1 e 2 di tutti e tre i parametri si distribuiscono in modo sufficientemente significativo intorno alla retta dei quantili della normale, in modo da poter concludere che le popolazioni da cui provengono sono *preossochè*Gaussiane. Per quanto riguarda il livello 3 l'evidenza della normalità è molto minore, soprattutto per quanto riguarda i valori del parametro BNP e questo è dovuto alla scarsità di dati disponibili. Comunque, nonostante alcuni dati si discostino considerevolmente dalla retta dei quantili, in tutti e tre i grafici sul livello 3, almeno 2 dati sui 6 totali appartengono esattamente alla retta dei quantili. Questo, unito al fatto che i dati relativi ai livelli 1 e 2 provengono da popolazioni Gaussiane, ci permette di poter supporre che la normalità delle popolazioni in esame è verificata.

A questo punto non resta che verificare l'omoschedasticità delle distribuzioni normali e per farlo possiamo ricorrere alla **analisi dei residui** descritta nel paragrafo (3.1.2).

Riportiamo i risultati ottenuti con Mathematica:

data1F = {70., 65., 50., 70., 72., 50., 65., 45., 60., 50., 50., 50., 47.,

65., 35., 76., 44., 70., 49., 61., 60., 71., 58., 65., 68., 49., 60.};

data2F = {42., 30., 71., 44., 30., 87., 50., 41., 51., 60., 33., 46., 50.};

data3F = {51., 30., 46., 32., 23., 58.};

M1 = Mean[data1F];

M2 = Mean[data2F];

M3 = Mean[data3F];

e1 = data1F − M1;

```
e2 = data2F − M2;

e3 = data3F − M3;

MSW = "305.189";

d1 = e1/√MSW;

d2 = e2/√MSW;

d3 = e3/√MSW;

Print["Lista dei residui standardizzati per livello 1 di FE"]

Print[d1]

Print["Lista dei residui standardizzati per livello 2 di FE"]

Print[d2]

Print["Lista dei residui standardizzati per livello 3 di FE"]

Print[d3]
```

Lista dei residui standardizzati per livello 1 di FE
$\{0.667825, 0.381614, -0.477018, 0.667825, 0.782309, -0.477018, 0.381614, -0.763228, 0.0954035, -0.477018, -0.477018, -0.477018, -0.648744, 0.381614, -1.33565, 1.01128, -0.82047, 0.667825, -0.53426, 0.152646, 0.0954035, 0.725067, -0.0190807, 0.381614, 0.55334, -0.53426, 0.0954035\}$
Lista dei residui standardizzati per livello 2 di FE
$\{-0.391888, -1.07879, 1.26813, -0.277404, -1.07879, 2.18401, 0.0660486, -0.44913, 0.123291, 0.63847, -0.907067, -0.16292, 0.0660486\}$
Lista dei residui standardizzati per livello 3 di FE
$\{0.629663, -0.572421, 0.343453, -0.457937, -0.973116, 1.03036\}$

```
data1B = {121.61, 142.54, 140.3, 100.48, 179.16, 216.78, 174.72, 31.76, 92.94, 240.12, 32.64, 161.45, 32.5,

73.25, 128.87, 91.26, 186.07, 173.19, 84.71, 16.1, 176.09, 201.71, 50.31, 93.58, 39.43, 23.5, 167.58};

data2B = {72, 315.56, 280.21, 163.73, 279.76, 237.67, 241.51, 235, 142.47, 180.78, 200.55, 241.14, 355.17,

236.24, 145.34, 212.73, 198.98};

data3B = {518.92, 3006.5, 413.85, 365.32, 573.55, 3416.8};

N1 = Mean[data1B];

N2 = Mean[data2B];

N3 = Mean[data3B];

p1 = data1B − N1;

p2 = data2B − N2;

p3 = data3B − N3;
```

```
MSW1 = "219935.";
t1 = p1/√MSW1;
t2 = p2/√MSW1;
t3 = p3/√MSW1;
Print["Lista dei residui standardizzati per livello 1 di BNP"]
Print[t1]
Print["Lista dei residui standardizzati per livello 2 di BNP"]
Print[t2]
Print["Lista dei residui standardizzati per livello 3 di BNP"]
Print[t3]
```

Lista dei residui standardizzati per livello 1 di BNP
{0.00875199, 0.0533815, 0.0486051, −0.0363039, 0.131467, 0.211685, 0.122,
− 0.182837, −0.0523816, 0.261453, −0.180961, 0.0937037, −0.181259, −0.094367, 0.0242326,
− 0.0559639, 0.146201, 0.118737, −0.0699306, −0.216229, 0.124921, 0.179551, −0.143282,
− 0.051017, −0.166482, −0.20045, 0.106775}
Lista dei residui standardizzati per livello 2 di BNP
{−0.315438, 0.20391, 0.128533, −0.11984, 0.127573, 0.0378236, 0.0460117, 0.0321303,
−0.165173, −0.0834841, −0.0413281, 0.0452228, 0.288371, 0.0347744, −0.159054, −0.0153565, −0.0446759}
Lista dei residui standardizzati per livello 3 di BNP
{−1.84141, 3.46291, −2.06545, −2.16893, −1.72492, 4.3378}

```
data1P = {457.9, 173, 406.3, 281.2, 981.2, 705.2, 572.7, 1768, 229.1, 1351, 1055, 4142, 96.67,
634.6, 1732, 1604, 636.6, 216.7, 957.9, 1161, 970.2, 469.3, 460.9, 237.1};
data2P = {1257, 309.2, 7116, 3902, 1578, 2657, 2471, 4322, 1046, 551.4, 2558, 967.8,
2641, 1893, 925.3, 2233};
data3P = {119.2, 12995, 2596, 2836, 12732, 14537};
R1 = Mean[data1P];
R2 = Mean[data2P];
R3 = Mean[data3P];
q1 = data1P − R1;
q2 = data2P − R2;
```

q3 = data3P − R3;

MSW2 = "6.21555" × 10^"6";

z1 = $\frac{q1}{\sqrt{MSW2}}$;

z2 = $\frac{q2}{\sqrt{MSW2}}$;

z3 = $\frac{q3}{\sqrt{MSW2}}$;

Print["Lista dei residui standardizzati per livello 1 di PRO-BNP"]

Print[z1]

Print["Lista dei residui standardizzati per livello 2 di PRO-BNP"]

Print[z2]

Print["Lista dei residui standardizzati per livello 3 di PRO-BNP"]

Print[z3]

Lista dei residui standardizzati per livello 1 di PRO-BNP
$\{-0.172308, -0.286584, -0.193005, -0.243184, 0.0375909, -0.0731146, -0.126261, 0.353182,$
$-0.264082, 0.18592, 0.0671926, 1.30541, -0.3172, -0.101433, 0.338742, 0.2874,$
$-0.100631, -0.269055, 0.0282451, 0.10971, 0.0331787, -0.167736, -0.171105, -0.260873\}$
Lista dei residui standardizzati per livello 2 di PRO-BNP
$\{-0.409021, -0.789191, 1.94106, 0.651907, -0.280266, 0.152528, 0.0779226, 0.820372,$
$-0.493655, -0.692042, 0.112819, -0.525021, 0.146111, -0.153917, -0.542069, -0.0175409\}$
Lista dei residui standardizzati per livello 3 di PRO-BNP
$\{-3.01499, 2.14959, -2.02153, -1.92526, 2.04409, 2.76809\}$

Analizzando le liste dei residui standardizzati è facile verificare che per i livelli 1 e 2 di tutti e tre i parametri i valori sono tali che $|d_{ij}| < 2$ e questo assicura che nel 95% dei casi le ipotesi di omoschedasticità sono verificate. Per quanto riguarda il livello 3: per il parametro FE le ipotesi sono verificate, mentre per i parametri BNP e $PRO-BNP$ si trovano alcuni valori tali che $|d_{ij}| > 2$, ma come prima osservato questo può essere associato alla scarsità di dati disponibili e poichè nel caso del parametro FE le ipotesi sono verificate anche per il livello 3, possiamo concludere che in generale anche l'ipotesi di omoschedasticità è sufficientemente verificata.
Una volta fatte queste verifiche non resta che procedere con le tre distinte analisi *ANOVA* e relativi test *post-hoc*.
L'ipotesi H_0 su ogni analisi della varianza è:

- H_0: le osservazioni relative al parametro analizzato sui tre livelli di *NYHA* provengono tutte dalla stessa popolazione gaussiana;

o equivalentemente:

- H_0: i tre livelli di *NYHA* non presentano alcuna differenza rispetto al parametro analizzato.

Il test *post-hoc* che utilizzeremo è quello di Bonferroni in quanto il numero di confronti da effettuare per ogni analisi è sufficientemente basso: $\binom{3}{2} = 3$.

Riportiamo ora i risultati ottenuti con il software Mathematica:

Needs["ANOVA"]

ddata1 = {{1.1, 70}, {1.1, 65}, {1.1, 50}, {1.1, 70}, {1.1, 72}, {1.1, 50}, {1.1, 65}, {1.1, 45}, {1.1, 60},
{1.1, 50}, {1.1, 50}, {1.1, 50}, {1.1, 47}, {1.1, 65}, {1.1, 35}, {1.1, 76}, {1.1, 44}, {1.1, 70}, {1.1, 49},
{1.1, 61}, {1.1, 60}, {1.1, 71}, {1.1, 58}, {1.1, 65}, {1.1, 68}, {1.1, 49}, {1.1, 60}, {2.1, 42}, {2.1, 30},
{2.1, 71}, {2.1, 44}, {2.1, 30}, {2.1, 87}, {2.1, 50}, {2.1, 41}, {2.1, 51}, {2.1, 60}, {2.1, 33}, {2.1, 46},
{2.1, 50}, {3.1, 51}, {3.1, 30}, {3.1, 46}, {3.1, 32}, {3.1, 23}, {3.1, 58}};

ddata2 = {{1.2, 121.61}, {1.2, 142.54}, {1.2, 140.3}, {1.2, 100.48}, {1.2, 179.16}, {1.2, 216.78}, {1.2, 174.72},
{1.2, 31.76}, {1.2, 92.94}, {1.2, 240.12}, {1.2, 32.64}, {1.2, 161.45}, {1.2, 32.5}, {1.2, 73.25},
{1.2, 128.87}, {1.2, 91.26}, {1.2, 186.07}, {1.2, 173.19}, {1.2, 84.71}, {1.2, 16.1}, {1.2, 176.09},
{1.2, 201.71}, {1.2, 50.31}, {1.2, 93.58}, {1.2, 39.43}, {1.2, 23.5}, {1.2, 167.58}, {2.2, 72}, {2.2, 315.56},
{2.2, 280.21}, {2.2, 163.73}, {2.2, 279.76},
{2.2, 237.67}, {2.2, 241.51}, {2.2, 235}, {2.2, 142.47}, {2.2, 180.78}, {2.2, 200.55}, {2.2, 241.14},
{2.2, 355.17}, {2.2, 236.24}, {2.2, 145.34}, {2.2, 212.73}, {2.2, 198.98}, {3.2, 518.92}, {3.2, 3006.5},
{3.2, 413.85}, {3.2, 365.32}, {3.2, 573.55}, {3.2, 3416.8}};

ddata3 = {{1.3, 457.9}, {1.3, 173}, {1.3, 406.3}, {1.3, 281.2}, {1.3, 981.2}, {1.3, 705.2}, {1.3, 572.7},
{1.3, 1768}, {1.3, 229.1}, {1.3, 1351}, {1.3, 1055}, {1.3, 4142}, {1.3, 96.67}, {1.3, 634.6}, {1.3, 1732},
{1.3, 1604}, {1.3, 636.6}, {1.3, 216.7}, {1.3, 957.9}, {1.3, 1161}, {1.3, 970.2}, {1.3, 469.3}, {1.3, 460.9},
{1.3, 237.1}, {1.3, 1257},
{2.3, 309.2}, {2.3, 7116}, {2.3, 3902}, {2.3, 1578}, {2.3, 2657}, {2.3, 2471}, {2.3, 4322},
{2.3, 1046}, {2.3, 551.4}, {2.3, 2558}, {2.3, 967.8}, {2.3, 2641}, {2.3, 1893}, {2.3, 925.3}, {2.3, 2233},
{3.3, 119.2}, {3.3, 12995}, {3.3, 2596}, {3.3, 2836}, {3.3, 12732}, {3.3, 14537}};

$k = 3$;

```
alfa = 0.05;
quant = Quantile[FRatioDistribution[k − 1, 50 − k], 1 − alfa];
Print["La soglia in valore assoluto della regione critica per T=FRatio è"];
Print[quant]
ANOVA[ddata1, PostTests → Bonferroni, CellMeans → True]
ANOVA[ddata2, PostTests → Bonferroni, CellMeans → True]
ANOVA[ddata3, PostTests → Bonferroni, CellMeans → True]
```

La soglia in valore assoluto della regione critica per T=FRatio è

3.19506

$$\left\{\text{ANOVA} \to \begin{array}{lllllll} & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\ \text{Model} & 2 & 2003.18 & 1001.59 & 6.11272 & 0.00460873 \\ \text{Error} & 43 & 7045.69 & 163.853 & & \\ \text{Total} & 45 & 9048.87 & & & \end{array}\right\}$$

$$\left\{\text{CellMeans} \to \begin{array}{ll} \text{All} & 53.2609 \\ \text{Model}[1.1] & 58.3333 \\ \text{Model}[2.1] & 48.8462 \\ \text{Model}[3.1] & 40. \end{array}, \text{PostTests} \to \left\{\text{Model} \to \text{Bonferroni} \quad \{1.1, 3.1\}\right\}\right\}$$

$$\left\{\text{ANOVA} \to \begin{array}{lllllll} & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\ \text{Model} & 2 & 8.03806 \times 10^6 & 4.01903 \times 10^6 & 18.2737 & 1.34548662285103\text{*^-6} \\ \text{Error} & 47 & 1.0337 \times 10^7 & 219935. & & \\ \text{Total} & 49 & 1.8375 \times 10^7 & & & \end{array}\right\}$$

$$\left\{\text{CellMeans} \to \begin{array}{ll} \text{All} & 304.129 \\ \text{Model}[1.2] & 117.506 \\ \text{Model}[2.2] & 219.932 \\ \text{Model}[3.2] & 1382.49 \end{array}, \text{PostTests} \to \left\{\text{Model} \to \text{Bonferroni} \quad \{\{1.2, 3.2\}, \{2.2, 3.2\}\}\right\}\right\}$$

$$\left\{\text{ANOVA} \rightarrow \begin{array}{lllllll} & \text{DF} & \text{SumOfSq} & \text{MeanSq} & \text{FRatio} & \text{PValue} \\ \text{Model} & 2 & 2.1959 \times 10^8 & 1.09795 \times 10^8 & 17.6645 & 2.513185491885743\text{*\^{}-6} \\ \text{Error} & 43 & 2.67269 \times 10^8 & 6.21555 \times 10^6 & & \\ \text{Total} & 45 & 4.86858 \times 10^8 & & & \end{array}\right\}$$

$$\left\{\text{CellMeans} \rightarrow \begin{array}{ll} \text{All} & 2250.92 \\ \text{Model[1.3]} & 902.263 \\ \text{Model[2.3]} & 2344.71 \\ \text{Model[3.3]} & 7635.87 \end{array}, \text{PostTests} \rightarrow \left\{\text{Model} \rightarrow \text{Bonferroni} \quad \{\{1.3, 3.3\}, \{2.3, 3.3\}\}\right\}\right\}$$

Osservando i risultati possiamo concludere che l'ipotesi H_0 di uguaglianza tra i livelli di *NYHA* è da rifiutare su tutti e tre i parametri analizzati, inoltre grazie al test di Bonferroni concludiamo che:

- il livello 1 e il livello 3 di *NYHA* differiscono significativamente per i valori di *FE*;
- il livello 1 e il livello 3 e il livello 2 e 3 di *NYHA* differiscono significativamente per i valori di *BNP*;
- il livello 1 e il livello 3 e il livello 2 e 3 di *NYHA* differiscono significativamente per i valori di *PRO-BNP*.

Ora osservando anche i valori medi riportati in *CellMeans*, possiamo rispondere alla domanda iniziale del problema, ovvero possiamo dire se c'è una relazione tra i livelli di $NYHA$ e i parametri biologici studiati e che tipo di relazione si ha:
in base alla nostra analisi e ai dati disponibili possiamo concludere (ad un livello di significatività $\alpha = 0{,}05$) che:

- i pazienti di livello 1 di $NYHA$ differiscono significativamente dai pazienti di livello 3 per tutti e tre i parametri studiati e in particolare passando dal livello 1 al livello 3 il valore medio di FE diminuisce significativamente, mentre i valori medi di BNP e $PRO - BNP$ aumentano significativamente;
- i pazienti di livello 2 e 3 differiscono significativamente per i valori dei parametri di BNP e di $PRO - BNP$ e in particolare all'aumentare del livello di $NYHA$ il valore di tali parametri aumenta;
- dalla nostra analisi non risulta che vi sia una differenza significativa tra i pazienti di livello 1 e di livello 2 relativamente ai parametri studiati, quindi possiamo concludere che vi è una assenza di relazione tra questi due livelli e i parametri fisiologico- biologici di FE, BNP e $PRO - BNP$.

In conclusione dal nostro studio risulta che la classificazione clinica (basata sui sintomi) $NYHA$ abbia una effettiva tendenza alla progressività (all'aumentare del livello aumenta la gravità del paziente) legata ai parametri fisiologico-biologici analizzati. Ma dai nostri dati risulta anche che non c'è una differenza significativa legata ai parametri fisiologico-biologici tra i pazienti di livello 1 e livello 2 e quindi si può ipotizzare che la classificazione in base alla dispnea di livello 1 e 2 non sia veramente "significativa" ed efficace.
Le conclusioni appena tratte si possono considerare statisticamente veritiere, in quanto lo

studio condotto è uno studio di tipo retrospettivo, ovvero tale che avendo a disposizione un certo numero di pazienti e dati vuole dare una risposta ad un certo livello di significatività. Pertanto, nonostante il numero di campioni a disposizione sia relativamente basso (sono però questi i numeri con cui si lavora generalmente a causa delle disponibilità ridotte dei pazienti con le caratteristiche volute), si può concludere che lo studio è efficace in quanto le ipotesi di normalità e omoshedasticità sono pressochè verificate e questo è sufficiente per poter concludere che i dati analizzati descrivono significativamente l'intera popolazione dei pazienti cardiopatici sui tre livelli di $NYHA$.

Capitolo 4

Analisi della varianza non parametrica

Nel capitolo precedente abbiamo visto che tutte le forme di analisi della varianza, inclusi i test t, sono basate sul presupposto che le osservazioni provengano da popolazioni Gaussiane nelle quali le varianze sono pressochè identiche. Queste condizioni sono spesso soddisfatte in misura sufficiente (come nell'esempio del capitolo precedente) e quindi l'analisi della varianza è un procedimento statistico di estrema utilità, ma talvolta capita che i dati sperimentali non siano compatibili con queste condizioni preliminari. A volte ci si trova anche di fronte a problemi nei quali le osservazioni sono espresse su una scala ordinale e non ad intervalli. E' chiaro che questi problemi non possono essere trattati con l'analisi della varianza descritta precedentemente. Infatti lo scopo di questo capitolo è illustrare metodologie analoghe all'*analisi della varianza*, che però non richiedono informazioni sulla natura della popolazione statistica da cui sono state tratte le osservazioni. In particolare tali metodologie sono basate sui *ranghi* delle osservazioni, cioè sul numero d'ordine, invece che sui valori delle osservazioni. Il termine *rango* traduce l'inglese *rank* che significa posizione in graduatoria/classifica/ordine crescente.
Presenteremo in questo capitolo alcuni di questi test, che non essendo basati sui parametri della popolazione d'origine, sono definiti metodi *non parametrici* o *liberi da distribuzione*. L'unica condizione posta per questi test è che le distribuzioni relative ai diversi trattamenti siano di forma analoga, senza alcuna limitazione rispetto al tipo di distribuzione.
E' bene dire che quando le osservazioni sono tratte da popolazioni che non sono normalmente distribuite, i metodi non parametrici sono, non solo più attendili, ma anche più potenti dei metodi parametrici. Mentre quando le osservazioni sono tratte da popolazioni normalmente distribuite, i metodi non parametrici hanno una potenza pari al 95% degli analoghi metodi parametrici. E' anche vero, però, che nella pratica non è sempre immediato decidere se un campione di dati si possa far provenire da una popolazione Gaussiana o meno, soprattutto nel caso di campioni di piccole dimensioni. Ci sono alcune strategie empiriche che possono aiutare nella decisione, come *l'analisi dei residui* o il *grafico quantile-quantile* che rende lineari le osservazioni distribuite normalmente e quindi esaminando la bontà di adattamento della retta, si può valutare quanto le osservazioni siano compatibili con una distribuzione normale (altrimenti si può fare una analisi di tipo descrittivo utilizzando un *grafico a scatola* o un *istogramma*). Nessuno di questi metodi, però, è sempre in grado di orientare in una direzione precisa, in questi casi la scelta del tipo di approccio è personale e molto discussa.
Noi operiamo ritenendo che se non si hanno elementi sufficienti per affermare che i dati non appartengono ad una popolazione gaussiana, è più opportuno utilizzare i metodi parametrici che sono più potenti e di uso più comune e scegliamo i metodi non parametrici solo quando

siamo certi di non poter usare quelli parametrici.

4.1 Il test di Kruskal-Wallis

Il test di Kruskal-Wallis, che prende il nome dai suoi autori William Kruskal e W. Allen Wallis è la procedura analoga alla analisi della varianza ad una via con disegno sperimentale completamente randomizzato, nel caso non parametrico. Questo test è una immediata generalizzazione del test di Mann-Whitney per la somma dei ranghi, che si applica nel caso di due soli gruppi a confronto. Quest'ultimo non verrà trattato nella seguente tesi, ma si può facilmente ricavare come caso particolare del test di Kruskal-Wallis.
Lo scopo del test di Kruskal-Wallis è quello di saggiare l'ipotesi H_0 di uguaglianza tra le mediane di popolazioni aventi distribuzione non gaussiana, contro l'alternativa H_A che almeno una delle popolazioni in studio tende a manifestare valori più elevati rispetto ad almeno una delle altre popolazioni.
Per rispondere a questa domanda il test di Kruskal-Wallis fa uso dei *ranghi*:

- per prima cosa si considerano tutte le osservazioni, indipendentemente dal gruppo i al quale le osservazioni stesse appartengono;

- si attribuiscono i *ranghi*, cominciando da 1 che viene associato all'osservazione che assume il valore più basso, fino ad $N = n_1 + ... + n_k$ (cioè il totale di tutte le osservazioni prese contemporaneamente) che viene associato al valore più alto. Se due osservazioni assumono lo stesso valore, ad esse viene attribuito lo stesso rango pari alla media dei ranghi che avrebbero avuto in caso non fossero state identiche;

- successivamente si calcola la somma dei ranghi, che denotiamo R_i in ciascun gruppo, la media dei ranghi in ciascun gruppo: $\bar{R}_i = \frac{R_i}{n_i}$ e la media generale $\bar{R} = \frac{1}{N}\sum_{i=1}^{k} R_i$.

A questo punto ragioniamo sui *ranghi* e assumiamo l'ipotesi nulla H_0 che i trattamenti non abbiano alcun effetto, in questa ipotesi ranghi grandi e piccoli dovrebbero essere equamente distribuiti tra i diversi gruppi e quindi il rango medio di ciascun gruppo $\bar{R}_i$ dovrebbe essere vicino alla media di tutti i ranghi $\bar{R}$. Sfruttando questa osservazione costruiamo un test opportuno.
Ipotizziamo di avere k gruppi ognuno con n_i, $i = 1....k$ osservazioni e che il totale sia $N = n_1 + .. + n_k$; la media generale dei ranghi è la media dei primi N numeri interi:

$$\bar{R} = \frac{1 + 2 + 3... + N}{N} = \frac{N + 1}{2},$$

infatti anche in caso di osservazioni ripetute i ranghi sono assegnati in modo tale da coprire tutti i numeri interi da 1 a N. A questo punto consideriamo la somma dei quadrati delle differenze tra il rango medio di ciascun gruppo e il rango medio complessivo ponderato sulle dimensioni di ciascun gruppo:

$$D = \sum_{i=1}^{k} n_i(\bar{R}_i - \bar{R})^2,$$

questa quantità presenta una perfetta analogia con la quantità SSA, ovvero la somma dei quadrati tra i gruppi utilizzata nella costruzione della statistica per l'ANOVA ad una via. Essa può essere utilizzata come indicatore della variabilità tra valori osservati e valori attesi nel caso in cui l'ipotesi che il trattamento non abbia alcun effetto sia vera. Infatti in caso di

ipotesi H_0 vera $\frac{D}{k-1}$ è uno stimatore corretto della varianza della variabile aleatoria relativa ai ranghi. A questo punto consideriamo la quantità:

$$S = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij}-\bar{R}\right)^2}{N-1},$$

questa quantità presenta anch'essa una perfetta analogia con la quantità MST della analisi della varianza. In particolare quest'ultima per N sufficientemente grande si può considerare uguale alla varianza dei ranghi σ^2, infatti sviluppando il quadrato si ha che:

$$S = \frac{1}{N-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij}^2+\bar{R}^2-2r_{ij}\bar{R}\right)\right)$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} r_{ij}^2 + \sum_{i=1}^{k} n_i\bar{R}^2 - 2\bar{R}\sum_{i=1}^{k}\sum_{j=1}^{n_i} r_{ij}\right)$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} r_{ij}^2 - N\bar{R}^2\right)$$

$$= \frac{1}{N-1}\left(\frac{1}{6}N(N+1)(2N+1) - N\left(\frac{N+1}{2}\right)^2\right)$$

$$= \frac{1}{N-1}\left(\frac{N(N+1)(N-1)}{12}\right) = \frac{N(N+1)}{12},$$

mentre la varianza esatta relativa ai ranghi è:

$$\sigma^2 = E[r^2] - E[r]^2 = \frac{1}{N}\sum_{i=1}^{N} i^2 - \frac{1}{N}\sum_{i=1}^{N} i = \frac{(N-1)(N+1)}{12}.$$

A questo punto consideriamo la statistica:

$$H = \frac{D}{S},$$

che corrisponde a:

$$H = (N-1)\frac{\sum_{i=1}^{k} n_i(\bar{R}_i-\bar{R})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij}-\bar{R}\right)^2},$$

e quindi:

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k} n_i\left(\bar{R}_i-\bar{R}\right)^2,$$

o più comunemente scritta esplicitando $\bar{R} = \frac{N+1}{2}$ e ottenendo:

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1).$$

La statistica H è costruita in modo molto simile a una variabile aleatoria $\chi^2(k-1)$, infatti possiamo dire che per N sufficientemente grande:

$$H \sim \frac{MSA(k-1)}{\sigma^2},$$

che nell'ipotesi H_0 di uguaglianza tra i trattamenti ha una distribuzione $\chi^2(k-1)$.
L'esatta distribuzione di H può essere calcolata elencando tutte le combinazioni possibili dei ranghi assegnati ai gruppi, ma le possibilità sono talmente numerose che in molti casi la tabella che ne deriverebbe non potrebbe rappresentarle tutte. Quindi il modo corretto di ragionare è proprio quello di osservare che se le dimensioni del campione non sono troppo ridotte, la distribuzione $\chi^2(k-1)$ (dove k è il numero dei trattamenti) è una buona approssimazione della distribuzione di H. Di conseguenza possiamo verificare l'ipotesi che il trattamento non abbia effetto confrontando il valore di H con i valori critici del χ^2, esattamente come abbiamo fatto nel caso della ANOVA con i valori critici della distribuzione di Fisher. In questo caso la regione di rigetto dell'ipotesi nulla sarà data da:

$$\left]\chi^2_{1-\alpha}(k-1), +\infty\right[,$$

infatti se il valore di H è sufficientemente più grande di 1 possiamo concludere che i ranghi medi $\bar{R}_i$ di ogni gruppo si discostano dalla media generale dei ranghi $\bar{R}$, ovvero la quantità MSA che in questo caso è data da $\frac{D}{k-1}$ non si può considerare come uno stimatore corretto per la varianza dei ranghi e pertanto si può concludere a un livello α di significatività fissato che le osservazioni non provengono tutte dalla medesima popolazione, qualsiasi sia la loro distribuzione.
Questa approssimazione è valida, in esperimenti con tre gruppi, quando ciascun gruppo di trattamento contiene almeno 5 elementi e in esperimenti con 4 gruppi o più, quando nell'intero studio sono coinvolti più di 10 individui. Per studi su campioni più limitati è necessario consultare una tabella che riporti la distribuzione esatta di H.

4.2 Il test di Friedman

Il *test di Friedman* è l'analogo non parametrico della ANOVA ad una via con disegno sperimentale per misure ripetute quando i trattamenti in studio sono più di due (se i trattamenti sono esattamente due si utilizza il *test di Wilcoxon* che non verrà trattato in questa tesi, ma che si può facilmente ricavare come caso particolare di quest'ultimo). Come abbiamo già visto nel caso della analisi della varianza parametrica, in questo disegno degli esperimenti ciascun soggetto è sottoposto a tutti i trattamenti in questione e quindi lo stesso campione di soggetti è sottoposto a tutti i diversi trattamenti da analizzare. Questo disegno sperimentale è molto importante poichè comporta una diminuzione dell'incertezza legata alla variabilità delle risposte tra individui diversi e fornisce un test più sensibile per conoscere l'effetto dei trattamenti.
Il *test di Friedman* viene utilizzato quando le condizioni di applicabilità della ANOVA non sono soddisfatte, infatti per poter utilizzare quest'ultimo, non è necessario che le osservazioni siano distribuite normalmente ed è basato sui *ranghi*. La logica di questo test è esattamente identica a quella del test di *Kruskal-Wallis*, con la semplice differenza che in questo caso si attribuisce il rango alle risposte di ciascun soggetto ai trattamenti senza tenere conto degli altri soggetti. In altre parole, poichè ogni soggetto viene sottoposto a tutti i k trattamenti, qualsiasi sia il numero N di osservazioni complessive, il numero di ranghi utilizzati è pari al numero di trattamenti.
In questo modo se l'ipotesi nulla che i trattamenti non hanno alcun effetto è corretta, allora, in ciascun soggetto i ranghi saranno distribuiti in modo casuale e le somme dei ranghi per ogni trattamento assumeranno valori simili.
Vediamo un esempio di una tabella dei ranghi nel caso di 5 soggetti, sottoposti ciascuno a 4 trattamenti:

	Trattamenti			
Soggetti in studio	1	2	3	4
1	1	2	3	4
2	4	1	2	3
3	3	4	1	2
4	2	3	4	1
5	1	4	3	2
Somma dei ranghi	11	14	13	12

A ciascun trattamento sono attribuiti i ranghi: 1, 2, 3, 4 e ad ogni soggetto è attribuito un rango per ogni trattamento. L'ultima riga riporta le somme dei ranghi per tutti i soggetti sottoposti a a ogni trattamento. Osserviamo che in questo esempio i ranghi sono distribuiti in modo casuale sui trattamenti e che le somme sono tutte simili a 12.5, che è la media dei ranghi: $\frac{(1+2+3+4)}{4} = 2.5$, moltiplicata per il numero di soggetti in studio (5). Questo induce a pensare che l'ipotesi di nessuna differenza tra i trattamenti sia verificata, ovvero non induce a ritenere che uno dei trattamenti eserciti un effetto sui soggetti in studio.
Ora se invece consideriamo una tabella come la seguente:

	Trattamenti			
Soggetti in studio	1	2	3	4
1	4	1	2	3
2	4	2	3	1
3	4	1	2	3
4	4	1	2	3
5	4	2	1	3
Somma dei ranghi	20	7	10	13

è evidente che il primo trattamento induce "sempre" , in tutti i soggetti, la risposta maggiore, il secondo "sempre" la risposta minore, il terzo e quarto "sempre" risposte intermedie, con la risposta del terzo maggiore di quella del quarto. Inoltre in questo caso c'è una grande variabilità nelle somme dei ranghi delle diverse colonne e alcune sono molto più grandi e altre molto più piccole di 12.5. Questo indica chiaramente che i trattamenti esercitano un effetto sulle variabili studiate e quindi implica il rifiuto della ipotesi nulla.
Una volta capito ciò, e sufficiente rendersi conto che l'unica differenza tra il test di Kruskal-Wallis e quest'ultimo sta nell'assegnazione dei ranghi e nel calcolo del rango medio che è dato da:

$$\frac{1+2+3....+k}{k} = \frac{k+1}{2}.$$

Utilizzando questa semplice accortezza il test è formalizzato in maniera esattamente identica a quello di Kruskal-Wallis e la statistica test che ne deriva è:

$$K = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1),$$

dove n è il numero di soggetti in studio, k il numero di trattamenti e $\bar{R}_i$ la media dei ranghi di ogni gruppo. Inoltre quando ci sono 3 (o di più) trattamenti e più di 9 soggetti totali e 4 (o più) trattamenti con più di 4 soggetti ciascuno, si ha che $K \sim \chi^2(k-1)$.

4.3 Esempio

Il problema che vogliamo analizzare è il seguente:

Problema 4.1. *L'ozono O_3 si forma da O_2 in presenza di NO_2 e di radiazione solare. A concentrazioni elevate l'ozono causa congestione polmonare, il limite di accettabilità in Italia è fissato dalla legge in $200gm^3(0{,}1ppm)$. Durante una giornata estiva, in quattro zone di una città (A, B, C, D) si sono rilevate le concentrazioni di O_3 riportate in tabella.*

A	***B***	***C***	***D***
150	120	210	195
140	115	190	210
145	30	185	220
160	155	170	205
165	130		175
170			430
125			480

Esiste una differenza significativa tra le medie della concentrazione di O_3 nelle quattro zone?

Per prima cosa bisogna osservare che i valori di concentrazione di una sostanza nell'aria hanno generalmente valori anomali a causa delle correnti e della disposizione delle fonti. Quindi in questo esempio sono ignote le caratteristiche statistiche della popolazione da cui sono estratti i dati campionari. Inoltre è facile osservare che nel gruppo D, la presenza del valore 430 e 480 determinano una varianza sensibilmente maggiore rispetto agli altri gruppi e questo sembra sottolineare che le ipotesi di omoschedasticità e di normalità necessarie per l'utilizzo della ANOVA non sono verificate.
Per verificare ciò è anche possibile fare una analisi descrittiva utilizzando uno *smooth histogram* e un *box plot*, che danno i seguenti risultati:

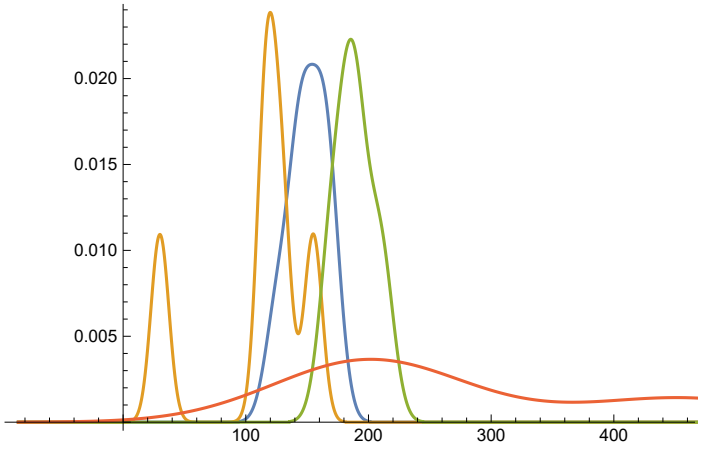
$\mathbf{data1 = \{150, 140, 145, 160, 165, 170, 125\};}$

$\mathbf{data2 = \{120, 115, 30, 155, 130\};}$

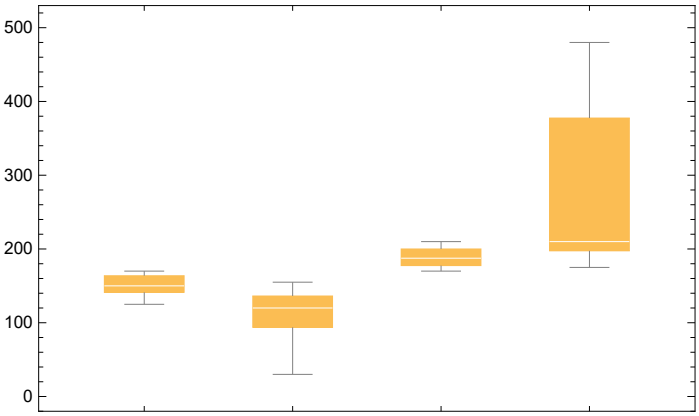
$\mathbf{data3 = \{210, 190, 185, 170\};}$

$\mathbf{data4 = \{195, 210, 220, 205, 175, 430, 480\};}$

$\mathbf{SmoothHistogram[\{data1, data2, data3, data4\}]}$



BoxWhiskerChart[{data1, data2, data3, data4}]



Nonostante le ipotesi non sembrano significativamente soddisfatte, i dati a disposizione sono molto pochi e non possiamo affermare con certezza che non provengono da popolazioni normali (i grafici quantile-quantile in questo caso non darebbero risposte soddisfacenti in nessun senso), quindi in questo caso è consigliabile operare con entrambi i metodi: sia l'analisi della varianza parametrica che non parametrica e poi confrontare i risultati ottenuti.

1. Per prima cosa utilizziamo il test di *Kruskal-Wallis* e saggiamo l'ipotesi relativa alle mediane delle concentrazioni di ozono. Quindi per poter procedere i valori devono essere sostituiti dal loro rango, calcolato su tutte le osservazioni dei k gruppi a confronto. Da essi, si calcola la somma dei ranghi (R_i) ed il numero di osservazioni (n_i) di ogni gruppo o campione:

Gruppi	**A**	**B**	**C**	**D**
	8	3	20	17
	6	2	16	19
	7	1	15	21
	10	9	13	18
	11	5		14
	12			22
	4			23
R_i	58	20	64	134
n_i	7	5	4	7

A questo punto calcoliamo il valore del test H con $N = 23$ e $k = 4$:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1),$$

e quindi:

$$H = \frac{12}{(23)(24)} \left(\frac{58^2}{7} + \frac{20^2}{5} + \frac{64^2}{4} + \frac{134^2}{7} \right) - 3(23+1),$$

da cui si ottiene:

$$H = 18{,}2112.$$

In questo caso si può supporre che nell'ipotesi H_0 di uguaglianza tra i gruppi $H \sim \chi^2(3)$, in quanto abbiamo a disposizione un numero sufficientemente "grande" di osservazioni (4 gruppi per un totale di più di 10 osservazioni), quindi dobbiamo confrontare il valore sperimentale di H con il corrispondente valore di $\chi^2_{1-\alpha}$.
La tabella dei valori critici con $3gdl$ riporta:

- 7,82 alla probabilità $\alpha = 0.05$,
- 11,34 alla probabilità $\alpha = 0.01$,
- 16,27 alla probabilità $\alpha = 0.001$.

Pertanto possiamo concludere che si può rifiutare l'ipotesi nulla, con una probabilità di commettere un errore di prima specie inferiore a 0.001 e quindi possiamo dire che esiste una differenza significativa tra le mediane dei valori di O_3 nei tre gruppi ad un livello di significatività $\alpha = 0.001$.

2. Procediamo ora saggiando l'ipotesi relativa alle medie delle concentrazioni di ozono, utilizzando una analisi della varianza parametrica e supponendo quindi che le ipotesi di normalità e omoschedasticità siano pressochè verificate.

   Riportiamo i risultati ottenuti con Mathematica:

   Needs["ANOVA"]

   data = {{1, 150}, {2, 120}, {3, 210}, {4, 195}, {1, 140}, {2, 115}, {3, 190}, {4, 210},
   {1, 145}, {2, 30}, {3, 185}, {4, 220}, {1, 160}, {2, 155}, {3, 170}, {4, 205}, {1, 165},
   {2, 130}, {4, 175}, {1, 170}, {4, 430}, {1, 125}, {4, 480}};

   k = 4;

   ntot = 23;

   alfa1 = 0.01;

   alfa2 = 0.05;

   alfa3 = 0.001;

   quant1 = Quantile[FRatioDistribution[k − 1, ntot − k], 1 − alfa1];

quant2 = Quantile[FRatioDistribution[$k-1$, ntot $-$ k], 1 $-$ alfa2];

quant3 = Quantile[FRatioDistribution[$k-1$, ntot $-$ k], 1 $-$ alfa3];

Print["La soglia in valore assoluto della regione critica per T=FRatio ad alpha=0.01 è"];

Print[quant1]

Print["La soglia in valore assoluto della regione critica per T=FRatio ad alpha=0.05 è"];

Print[quant2]

Print["La soglia in valore assoluto della regione critica per T=FRatio ad alpha=0.001 è"];

Print[quant3]

ANOVA[data]

La soglia in valore assoluto della regione critica per T=FRatio ad alpha=0.01 è
5.01029
La soglia in valore assoluto della regione critica per T=FRatio ad alpha=0.05 è
3.12735
La soglia in valore assoluto della regione critica per T=FRatio ad alpha=0.001 è
8.27993

ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
Model	3	91306.7	30435.6	5.46441	0.00702246
Error	19	105826.	5569.78		
Total	22	197133.			

Dai risultati ottenuti è evidente che l'ipotesi nulla di uguaglianza tra le medie di concentrazioni di ozono è da rifiutare ad un livello di significatività $\alpha = 0.01$, ma non ad $\alpha = 0.001$, come invece risultava dal test di Kruskal Wallis.
Il confronto tra il metodo non parametrico e quello parametrico, permette di concludere con certezza che ad un livello di significatività $\alpha = 0.01$ si ha una differenza tra le medie di concentrazione di ozono nelle 4 diverse zone.

Bibliografia

[1] Franco Anzani e Maria Pia D'Ambrosio, *Confronti multipli a posteriori o post hoc*, SixSigmaln Team, http://www.sixsigmain.it/ebook/Capu10-5.html

[2] Paolo Baldi, *Calcolo delle probabilità e statistica*, Milano, McGraw-Hill, 1998

[3] Wayne W. Daniel, *Biostatistica*, Napoli, EdiSES s.r.l, 1996

[4] Stanton A. Glantz, *Statistica per Discipline Biomediche*, Milano, McGraw-Hill, 2003

[5] Domenico Piccolo, *Statistica*, Bologna, Il Mulino, 2010