



ISTITUTO SUPERIORE DI SANITA'

Disegni sperimentali e analisi della varianza

A cura di F. Chiarotti e M. Puopolo

ISTITUTO SUPERIORE DI SANITA'

Disegni sperimentali e analisi della varianza

A cura di Flavia Chiarotti e Maria Puopolo

Laboratorio di Fisiopatologia di Organo e di Sistema

Istituto Superiore di Sanità

Disegni sperimentali e analisi della varianza

A cura di Flavia Chiarotti e Maria Puopolo

1995, ii, 102 p. Rapporti ISTISAN 95/23

Manuale sintetico per l'analisi dei dati sperimentali, raccolti secondo disegni di diversa complessità, applicando metodi di analisi della varianza parametrica e non parametrica. Vengono esaminati i disegni completamente randomizzati a uno o più fattori, i disegni a blocchi randomizzati a uno o più fattori e i disegni *split-plot* o a misure ripetute. Per ogni schema sperimentale, viene presentato il modello statistico corrispondente e la tabella dell'analisi della varianza che ne consegue.

Parole chiave: Biostatistica, Confronti multipli, Metodi non parametrici, Metodi parametrici.

Istituto Superiore di Sanità

Experimental designs and analysis of variance

Edited by Flavia Chiarotti and Maria Puopolo

1995, ii, 102 p. Rapporti ISTISAN 95/23 (in Italian)

A condensed textbook for the analysis of experimental data, collected according to experimental designs with different degrees of complexity, using parametric and non parametric analysis of variance. Completely randomized designs including one or more factors, randomized block designs with one or more factors and split-plot or repeated measures designs are examined. For each experimental design the statistical model and the analysis of variance table are presented.

Key words: Biostatistics, Multiple comparisons, Non parametric methods, Parametric methods.

Gli autori hanno contribuito per le seguenti parti:

Giorgio Bignami (a)	articolo introduttivo
Maria Patrizia Carrieri (b)	paragrafi 3.1., 3.2., 3.3., 3.4., 3.5.
Flavia Chiarotti (a)	paragrafi 2.5., 2.6., 3.7., 4.1., 4.2., 4.3., 4.4., 4.5., 4.6., 4.7.
Francesca Menniti-Ippolito (c)	capitolo 1; paragrafi 2.1., 2.2., 2.4.
Maria Puopolo (a)	paragrafi 2.3., 2.7., 3.6., 3.8., 4.8.

(a) Laboratorio di Fisiopatologia di Organo e di Sistema

(b) Servizio Elaborazione Dati

(c) Laboratorio di Epidemiologia e Biostatistica

INDICE

<i>AMARCORD DELL'EVOLUZIONE DELLA BIOSTATISTICA PER DISEGNI SPERIMENTALI NELL'ISTITUTO SUPERIORE DI SANITA' (1960-1995)</i>		1
1.	INTRODUZIONE AL DISEGNO SPERIMENTALE	9
2.	DISEGNO COMPLETAMENTE RANDOMIZZATO	11
2.1.	DEFINIZIONE	11
2.2.	UN SOLO FATTORE A <i>p</i> LIVELLI	13
2.2.1.	Modello sperimentale	13
2.2.2.	Scomposizione della devianza	15
2.2.3.	Gradi di libertà	17
2.2.4.	Analisi della varianza	19
2.2.5.	Assunti per l'applicazione del test F	22
2.3.	UN SOLO FATTORE A 2 LIVELLI	22
2.3.1.	Test t di Student per gruppi indipendenti	22
2.3.2.	Assunti per l'applicazione del test t	24
2.4.	DISEGNO FATTORIALE	25
2.4.1.	Modello sperimentale e scomposizione della devianza	25
2.4.2.	Analisi della varianza	26
2.4.3.	Rappresentazione grafica dell'interazione	28
2.5.	CONFRONTI ORTOGONALI	29
2.6.	CONFRONTI MULTIPLI	31
2.6.1.	Un solo fattore a <i>p</i> livelli	31
2.6.2.	Disegno fattoriale	34
2.7.	STATISTICA NON PARAMETRICA	35
2.7.1.	Un solo fattore a <i>p</i> livelli	36
2.7.2.	Un solo fattore a 2 livelli	38
2.7.3.	Disegno fattoriale	39
2.7.4.	Confronti multipli	42
3.	DISEGNO A BLOCCHI RANDOMIZZATI	43
3.1.	DEFINIZIONE	43
3.2.	UN SOLO FATTORE A <i>p</i> LIVELLI CON UNA SOLA OSSERVAZIONE PER CELLA	45
3.2.1.	Modello sperimentale e scomposizione della devianza	45
3.2.2.	Analisi della varianza	47
3.2.3.	Assunti per l'applicazione del test F	50
3.3.	UN SOLO FATTORE A 2 LIVELLI	52

3.3.1.	Test t di Student per gruppi appaiati	52
3.3.2.	Assunti per l'applicazione del test t	55
3.4.	UN SOLO FATTORE A p LIVELLI CON PIÙ DI UNA OSSERVAZIONE PER CELLA	55
3.5.	DATI MANCANTI	58
3.6.	DISEGNO FATTORIALE	60
3.6.1.	Modello sperimentale e scomposizione della devianza	60
3.6.2.	Analisi della varianza	62
3.7.	CONFRONTI MULTIPLI	66
3.7.1.	Un solo fattore a p livelli	66
3.7.2.	Disegno fattoriale	67
3.8.	STATISTICA NON PARAMETRICA	67
3.8.1.	Un solo fattore a p livelli	67
3.8.2.	Un solo fattore a 2 livelli	71
3.8.3.	Disegno fattoriale	73
3.8.4.	Confronti multipli	76
4.	DISEGNO MISTO: SPLIT-PLOT E MISURE RIPETUTE	77
4.1.	DEFINIZIONE	77
4.2.	UN FATTORE TRA SOGGETTI E UN FATTORE ENTRO SOGGETTI	78
4.2.1.	Disegno sperimentale	78
4.2.2.	Modello sperimentale	79
4.2.3.	Analisi della varianza	80
4.3.	DUE FATTORI TRA SOGGETTI E DUE FATTORI ENTRO SOGGETTI	82
4.3.1.	Disegno sperimentale	82
4.3.2.	Modello sperimentale	83
4.3.3.	Analisi della varianza	85
4.4.	UN FATTORE TRA PLOT, UN FATTORE ENTRO PLOT E UN FATTORE ENTRO SOGGETTI	88
4.4.1.	Disegno sperimentale	88
4.4.2.	Modello sperimentale	89
4.4.3.	Analisi della varianza	91
4.5.	ESERCIZIO	94
4.6.	ASSUNTI PER L'APPLICAZIONE DEL TEST F	99
4.7.	CONFRONTI MULTIPLI	99
4.8.	STATISTICA NON PARAMETRICA	100
	BIBLIOGRAFIA	102

AMARCORD DELL'EVOLUZIONE DELLA BIOSTATISTICA PER DISEGNI SPERIMENTALI NELL'ISTITUTO SUPERIORE DI SANITA' (1960-1995)

(Giorgio Bignami)

There are three kinds of lies: lies, damned lies, and statistics
(attribuita a Benjamin Disraeli)

"Vi sono tre tipi di menzogne: le menzogne, le maledette menzogne, e le statistiche". Molta acqua è passata sotto i ponti da quando Mark Twain [1] attribuì allo statista inglese Benjamin Disraeli, ricompensato dalla regina Vittoria col titolo di 1° Conte di Beaconsfield per la sua opera politica e culturale, questa celebre battuta di protesta contro le manipolazioni dei numeri fatte per servire scopi poco nobili nella gestione della cosa pubblica. "Statistica" deriva appunto da stato e prima che i tedeschi coniassero il termine, l'uso di dati statistici a fini di governo andava sotto denominazioni tanto significative quanto fantasiose, come per esempio "aritmetica politica".

Ma venendo a tempi più recenti e agli aspetti più strettamente scientifici, vanno soprattutto ricordate in questa sede le difficoltà che si sono incontrate in molte istituzioni italiane nei tentativi di superare quelle diffidenze e resistenze che per decenni si sono opposte allo sviluppo di una coscienza o consapevolezza statistica nei ricercatori impegnati in diversi settori della biomedicina e della sanità.

Sarebbe ingenuo e antistorico tentare di nascondere che tali diffidenze e resistenze non erano del tutto immotivate. E' vero, infatti, che l'applicazione dei metodi statistici - il "far girare i mulini delle preghiere" (per realizzare ad ogni costo il $p \leq 0.05$), secondo la battuta che correva nelle istituzioni americane ai tempi dell'ultima velocissima generazione di calcolatrici meccaniche - troppo spesso è servita a coprire le magagne di un esperimento mal disegnato, di un metodo mal costruito, di una conduzione approssimativa delle prove, di un arrampicarsi sugli specchi per convincere se stessi e gli altri della bontà di un risultato incerto.

D'altra parte, molti malintesi sono derivati dal fatto che di una valutazione statistica sofisticata hanno mediamente meno bisogno le discipline cosiddette dure, nelle quali è mediamente maggiore il controllo sulle fonti di errore, che non le discipline cosiddette morbide (*soft*), in particolare quelle comportamentali e sociali. In queste seconde, infatti, non solo i segnali vanno spesso faticosamente estratti da un "rumore" di fondo

che può essere assordante, ma può anche essere ardua la separazione tra ruolo causale e ruolo di confondimento delle diverse variabili. Pertanto, nell'eterna diatriba tra cultori delle discipline dure e cultori delle discipline morbide, i primi hanno spesso ridicolizzato i secondi per la loro necessità di appoggiarsi sulla validazione statistica, mentre i secondi hanno spesso scatenato rappresaglie evidenziando la latitanza o rozzezza delle verifiche statistiche nel lavoro dei primi.

Superate queste *impasse* di sapore guareschiano, è pienamente evidente oggi il valore di una cultura statistica la quale, di certo, esige di esprimersi in diversi modi e con diversa complessità nei diversi settori, ma che in ogni settore può dare un contributo significativo sia alla buona riuscita del lavoro di ricerca, sia alla efficacia di una moltitudine di applicazioni, in particolare in quel campo biomedico e sanitario che qui più direttamente interessa.

Primi sviluppi della biostatistica nell'Istituto Superiore di Sanità (ISS)

La storia degli sviluppi della statistica nell'ISS è assai interessante a fronte dei problemi cui si è appena accennato, risalendo a circa due decenni prima della fondazione del Laboratorio di epidemiologia e biostatistica. Infatti, già nei tardi anni cinquanta, l'apertura nei Laboratori di chimica terapeutica di nuovi campi di ricerca, come quello della farmacologia del comportamento, fece nascere l'esigenza di trattare in maniera appropriata dati "morbidi" che presentavano una variabilità assai più elevata di quelli "duri" ottenuti con metodi fisiologici e biochimici. Questi primi sviluppi furono caratterizzati da grossolani errori, errori inevitabili in quanto a quel tempo comunissimi (e ancor oggi non tanto infrequenti): per esempio, l'impiego indiscriminato, per confronti tra tre o più gruppi sperimentali, di test parametrici o non parametrici disegnati per il confronto tra due campioni della stessa popolazione.

L'analisi della varianza - parametrica o non parametrica che fosse - già da tempo correntemente adoperata in altre istituzioni (ma quasi esclusivamente all'estero, con la notevole eccezione del gruppo fondato da Giulio Maccacaro nell'Istituto di biometria e statistica medica dell'Università di Milano), era soltanto un nome di specie botanica esotica e inaccessibile. Per esempio, dovendo affrontare il problema di una variabilità nella acquisizione di risposte condizionate di evitamento la quale appariva a prima vista minore tra ratti della stessa lettiera che non tra animali di lettiera diverse, "reinventammo" una sorta di analisi della varianza rozza e primitiva. Questa consisteva nel confrontare graficamente la dispersione delle medie delle singole lettiera con quella delle medie di "false lettiera" costituite assegnando a sorte a ciascuna un animale proveniente da ognuna delle vere lettiera.

Vale la pena di spiegare perché fosse così importante, a quel punto delle nostre ricerche, trovare sostegni a un'ipotesi come quella dell'esistenza di un "effetto lettiera":

infatti, tale spiegazione consente di chiarire meglio il ruolo della statistica non solo nell'analisi dei dati e quindi nella validazione dei risultati (termine, *per incidens*, assolutamente improprio per indicare un calcolo della probabilità che sia stata falsificata l'ipotesi nulla), ma anche nelle valutazioni di probabilità di successo (o meglio, di rischio di insuccesso) di strategie alternative nelle successive tappe di un itinerario di ricerca.

Un primo problema, un problema eminentemente pratico, era quello di decidere se valesse o meno la pena di sostenere gli oneri aggiuntivi della procedura *split-litter* nelle esperienze psicofarmacologiche. (La procedura *split-litter* è quella che assegna in maniera casuale ma bilanciata i soggetti di ciascuna lettiera - tecnicamente, un blocco - ai diversi livelli o combinazioni di livelli di uno o più fattori fissi; in linguaggio statistico, si tratta di un modello misto nel quale uno o più fattori fissi sono incrociati con le lettieri trattate come fattore *random* o casuale).

Il secondo problema era di decidere se fosse opportuno avviare un programma pluriennale di selezione genetica bidirezionale per alte e basse prestazioni nell'apprendimento di una risposta attiva di evitamento. Infatti, se da un lato un effetto lettiera può essere dovuto a cause diverse, sia genetiche che ambientali (i cosiddetti fattori materni pre- e postnatali), dall'altro lato l'assenza di detto effetto rende meno verosimile l'ipotesi di un ruolo dei fattori genetici, quindi meno probabile il successo di una esperienza di selezione inevitabilmente assai impegnativa.

Alla fine "decidemmo" che l'effetto lettiera probabilmente c'era. Perciò adottammo la procedura *split-litter* con buoni risultati (nel senso di una riduzione della varianza residua mediamente più importante della perdita di gradi di libertà, quindi di un aumento di potenza dei test F riguardanti gli effetti dei trattamenti), oltre al realizzo di un maggior rispetto degli assunti sia sostanziali che statistico-formali. Il controllo statistico sull'effetto lettiera era anche importante in quanto la produzione di lettieri di ratti tutti dello stesso sesso, formate scartando alla nascita gli animali di sesso opposto nelle lettieri più numerose, date le limitazioni logistiche e di personale nell'ISS, dovette essere commissionata a un allevatore commerciale: e questi sapeva dei nostri metodi di controllo statistico sui rapporti fra variazioni entro e tra lettieri, quindi era costretto a seguire rigorosamente le regole che gli avevamo dettate, pena la sospensione degli ordinativi.

Inoltre avviammo subito l'esperienza di selezione, utilizzando come generazione parentale proprio i soggetti delle lettieri le cui prestazioni erano state in precedenza sottoposte alla già citata pseudoanalisi della varianza. La selezione portò nel giro di poche generazioni a differenze notevolissime di prestazione e così vennero fondati i ceppi *Roman High Avoidance* e *Roman Low Avoidance* [2] (termini non coniatati da noi, ma da un illustre padrino a quell'epoca massimo esperto della materia, il professor Peter Broadhurst [3]), ceppi che oggi si usano in vari laboratori di diversi paesi per una

vasta gamma di ricerche comportamentali, farmacologiche, fisiologico-biochimiche e neuroendocrinologiche. (Il fatto che già nel 1964 i ceppi si siano dovuti trasferire altrove e non siano più ritornati in ISS, è una storia che non compete a questa sede).

A distanza di diversi decenni, echeggia ancora nelle mie orecchie l'affettuosa risata del professor James (Jim) McGaugh - un esperto americano di psicologia sperimentale animale venuto da noi per una serie di ricerche sulla farmacologia dell'apprendimento, in collaborazione con il gruppo diretto dal professor Daniel Bovet - quando gli mostrammo i nostri dati sulle differenze entro e tra lettieri e il modo in cui li avevamo trattati (o meglio, maltrattati). Destreggiandosi tra le pagine di certi preziosi manuali che si era portato dalla California, ci fece vedere con quanta semplicità ed efficacia gli stessi dati potessero essere sottoposti a vere e proprie analisi della varianza, anche se a quei tempi gli schemi di calcolo per i disegni a modello misto con un fattore casuale come la lettiera ancora soffrivano di alcune imperfezioni.

Successivamente un altro importante aiuto venne dal Laboratorio di fisica, allora diretto dal professor Mario Ageno, il quale convinse un esperto di programmazione, il dottor Paolo Lanciani, a mettere a punto un *software* dedicato al disegno sperimentale da noi più comunemente adoperato, quello appunto che prevedeva i trattamenti come fattori fissi e le lettieri come blocchi, anche se ancora non trattava in maniera appropriata né la differenza tra fattori fissi e fattori casuali né le misure ripetute entro soggetti [4]. Malgrado tale aiuto, eravamo destinati a passare ancora parecchi anni in una condizione di sostanziale isolamento. Solo molto più tardi, infatti, comincio gradualmente ad estendersi l'impiego sistematico dei metodi statistici in vari settori biomedici e sanitari, in particolare l'epidemiologia (che fino a tempi relativamente recenti ha seguito a farsi con carta, penna e rudimentali calcolatrici da tavolo e portatili). Inoltre era assai frustrante la disponibilità in ISS di elaboratori sempre più potenti, che tuttavia non eravamo in grado di programmare per i nostri usi e consumi, a fronte di una situazione del mercato dei *software* ancora assai arretrata e di una nostra impreparazione a destreggiarci tra detti *software*.

Un ulteriore aiuto ci venne dal già citato gruppo milanese dell'Istituto di biometria, e in particolare dal professor Giulio Maccacaro e dal professor Ettore Marubini, quando questi incominciarono a distribuire le schede per la effettuazione di vari test statistici sul primo elaboratore elettronico da tavolo programmabile, la famosa e farraginoso Olivetti 101. Molte di queste schede, tuttavia, non erano direttamente utilizzabili per i nostri fini e quindi andavano faticosamente ritoccate e integrate, spesso con una proliferazione dei moduli da utilizzare in cascata per la esecuzione di una singola analisi, quindi con costi assai elevati di ogni pur minimo errore dell'operatore (una sorta di moderno gioco dell'oca).

Il lavoro dell'ultimo decennio

Molti problemi comunque - e non solo i nostri, ma anche quelli di diversi altri settori dell'ISS - non si avviarono a soluzione sinché nel neo-istituito Laboratorio di fisiopatologia di organo e di sistema, definito nella revisione dell'organigramma effettuata nel 1982 e affidato alla direzione del professor Amilcare Carpi de Resmini, non si prese una decisione da molti criticata, da alcuni addirittura accolta con sorrisi di compatimento: la decisione cioè di impegnare uno dei posti disponibili di ricercatore (pochi e preziosi) per acquisire una vera e propria competenza di biostatistica nel campo dei modelli sperimentali; e questo, tenuto conto del fatto che in nessuno dei settori più forti e consolidati dell'Istituto si era mai sentita l'esigenza di muovere un tal passo.

Non si possono qui analizzare in dettaglio gli sviluppi consentiti dalla disponibilità del ricercatore biostatistico nel nostro laboratorio (se non altro poiché non si può utilizzare questo spazio per tessere il panegirico del direttore di corso). In sostanza, la collaborazione stretta tra detto ricercatore e gli esperti di varie discipline - prima nel laboratorio, poi in misura via via crescente anche in altre sedi interne ed esterne all'ISS - ha consentito, in prima battuta, di realizzare una applicazione corretta, efficace ed economica di un'ampia gamma di metodi statistici già validati, sfruttando tra l'altro al meglio la crescente disponibilità di *software* sempre più sofisticati e maneggevoli per elaboratori di diverse dimensioni.

Questa attività ebbe anche ricadute esterne di notevole rilievo. Si incominciò infatti a inviare a riviste *leader* lavori di ontogenesi e teratologia comportamentale con una corretta trattazione dei rapporti tra fattori fissi e fattori casuali nei disegni a modello misto, cosa che creava un certo imbarazzo in qualche esperto della materia chiamato a far da *referee*. Il successo non agevole di questi primi confronti ci stimolò a "impuntarci" nella revisione dei lavori di altre parti, cioè a subordinare il parere favorevole a una rianalisi dei dati secondo modelli appropriati, talora sostenendo estenuanti confronti epistolari, attraverso i curatori delle riviste, con gli autori e i loro consulenti convinti della bontà delle tecniche adottate. Tale strategia, come è stato riconosciuto da più parti, ha contribuito in misura sostanziale all'elevamento degli standard di analisi dei dati che si è verificato sulla scena esterna nella seconda metà degli anni 80 (per una ricapitolazione del sistema di regole di cui si sta parlando v. [5]).

La piena immersione del ricercatore biostatistico nelle problematiche dei disegni sperimentali ha consentito, in parallelo, di identificare le possibilità e i limiti di metodi (soprattutto nel campo della statistica non parametrica) intrinsecamente inadeguati per la applicazione a disegni a modello misto con variabili dipendenti dicotomiche o ordinali (cioè non assoggettabili ad analisi della varianza di tipo parametrico), procedendo alla messa a punto e alla verifica di versioni modificate applicabili a situazioni in precedenza inaccessibili [6]. Ha consentito inoltre, grazie a un lavoro condotto su matrici di dati sperimentali di natura volta per volta diversa, di verificare l'efficacia di metodi di

analisi multivariata sia a fini di convalida dei risultati di analisi univariate reiterate su più variabili dipendenti (e pertanto soggetti alla riserva che discende dalla proliferazione dei test), sia al fine di fornire descrizioni più fedeli e complete dei fenomeni osservati (convalide e descrizioni che hanno trovato spazio su riviste come *Pharmacology Biochemistry and Behavior* [7] e *Behavioral Neuroscience* [8]).

Infine queste operazioni, nel volgere di circa un decennio, hanno anche consentito di conseguire un risultato che non è misurabile sul metro dei lavori volta per volta pubblicati o delle attività di consulenza e formazione volta per volta espletate: cioè hanno portato all'avvicinamento e alla sostanziale amalgamazione di due culture statistiche da sempre separate pur appartenendo alla stessa area della biomedicina e della sanità, quella dei "grandi numeri" sviluppata dalla moderna epidemiologia e quella dei "piccoli numeri" riguardante i modelli sperimentali. I metodi delle due aree restano di necessità diversi; tuttavia il lavoro condotto grazie al sempre più fitto va-e-vieni tra di esse attraverso le varie collaborazioni, ha contribuito a creare e a diffondere una consapevolezza dell'intima natura dei principi-base, della necessità di un rigoroso rispetto degli assunti logico-teorici comuni a tutte le applicazioni della statistica, anche da parte di chi non deve occuparsi né dell'algebra né della gestione del calcolo.

Il lavoro di preparazione di questo corso, quindi, esprime dei valori storici e culturali che sono almeno altrettanto importanti quanto gli aspetti strettamente tecnico-scientifici selezionati per una iniziativa monotematica di formazione. La varietà dei modelli di analisi della varianza applicabili alle situazioni più eterogenee è anche quella che consente di stabilire legami operativi e culturali sempre più saldi tra soggetti con vocazioni e compiti molto diversi, dal ricercatore che lavora su modelli sperimentali a quello che svolge un'attività di sperimentazione clinica intrecciata con quella medico-professionale; dai responsabili delle attività di un laboratorio nazionale di riferimento a quelli che nelle singole situazioni di servizio debbono affrontare quotidiani problemi di controllo di qualità.

Per ciascuna di queste figure vale sempre il vecchio avvertimento - *Lies, damned lies, and statistics* - ; ma il tempo della statistica strumentalmente e comodamente selvaggia è ormai definitivamente superato sul piano scientifico-metodologico, così come sono caduti gli alibi per indulgere in tale *modus operandi*, grazie ai progressi sul piano operativo, che si tratti di disponibilità di consulenze, di macchine o di *software*. Il problema è ora di diffondere il più ampiamente possibile le conoscenze e il *know-how* che si sono sviluppati, per allineare a uno standard elevato tutte le attività svolte in campo biomedico e sanitario, attraverso efficaci iniziative di formazione, di collaborazione e di verifica sui risultati conseguiti dalle diverse parti.

1. Partington, A. (Ed). 1992. *The Oxford Dictionary of Quotations*. 4^a edizione. Oxford University Press, Oxford. p. 249.
2. Bignami, G. 1965. Selection for high rates and low rates of avoidance conditioning in the rat. *Anim. Behav.* **13**: 221-227.
3. Broadhurst, P.L. & Bignami, G. 1965. Correlative effects of psychogenetic selection: a study of the Roman High and Low Avoidance strains of rats. *Behav. Res. Ther.* **2**: 273-280.
4. Lanciani, P. & Bignami, G. 1965. L'impiego dell'elaboratore elettronico nell'analisi statistica di dati relativi ad esperienze di psicofarmacologia. *Ann. Ist. Super. Sanità* **1**: 412-423.
5. Chiarotti, F., Alleva, E. & Bignami, G. 1987. Problems of test choice and data analysis in behavioral teratology: the case of prenatal benzodiazepines. *Neurotoxicol. Teratol.* **9**: 179-186.
6. Cirulli, F. & Alleva, E. 1994. Effects of repeated administrations of EGF and TGF-alpha on mouse neurobehavioral development. *Neurotoxicology* **15**: 819-825.
7. Loggi, G., Laviola, G., Alleva, E. & Chiarotti, F. 1991. Morphine effects on mouse locomotor/exploratory activity: test dependency, test reliability, uni- and multi-variate analyses. *Pharmacol. Biochem. Behav.* **38**: 817-822.
8. Laviola, G., Dell'Omo, G., Chiarotti, F. & Bignami, G. 1994. d-Amphetamine conditioned place preference in developing mice: relations with changes in activity and stereotypies. *Behav. Neurosci.* **108**: 514-524.

1. INTRODUZIONE AL DISEGNO SPERIMENTALE (Francesca Menniti-Ippolito)

La definizione del disegno sperimentale, cioè delle modalità secondo cui un esperimento deve essere condotto, implica:

- a) la formulazione di un'ipotesi statistica;
- b) la determinazione delle condizioni sperimentali e delle condizioni "estranee" da controllare nel corso dell'esperimento;
- c) la determinazione del numero di unità statistiche sperimentali necessarie a raggiungere risultati validi e la scelta della popolazione da cui campionarle;
- d) la definizione di procedure per l'assegnazione delle condizioni sperimentali ai soggetti selezionati;
- e) la scelta dei metodi statistici di analisi dei dati più adeguati.

Un esperimento consiste nell'intervento su una o più variabili da parte dello sperimentatore allo scopo di determinare un effetto di questo intervento su un'altra variabile.

A questo scopo definiamo come **VARIABILE INDIPENDENTE** (o **FATTORE**), la variabile sotto il controllo dello sperimentatore, sulla quale viene effettuato l'intervento; e come **VARIABILE DIPENDENTE** (o **VARIABILE RISPOSTA**), la variabile che riflette l'effetto del trattamento.

Tratteremo di seguito di esperimenti con una sola variabile dipendente e una o più variabili indipendenti.

Le variabili indipendenti possono essere quantitative (per esempio, diversi dosaggi di una stessa sostanza) o qualitative (diverse sostanze).

Le **VARIABILI DI DISTURBO** sono le fonti di variabilità indesiderata dell'esperimento che possono influire sulla variabile dipendente, "disturbando" la misura dell'effetto del trattamento. In pratica sono costituite dalla variabilità intrinseca dei soggetti, dovuta a variabili precedenti all'esperimento o non controllate durante l'esperimento, come ad esempio il sesso, il peso, l'età delle unità sperimentali.

Le variabili di disturbo debbono essere controllate quando possibile nel corso dell'esperimento o durante l'analisi statistica dei dati, per evitare di attribuire al

trattamento un effetto che, invece, è dovuto in parte o in tutto a una di queste variabili.

Viene definita come varianza residua (o dell'errore sperimentale) quella parte della varianza della variabile dipendente che è attribuibile a fonti estranee.

Il controllo delle variabili di disturbo può essere condotto in fase di disegno dell'esperimento tramite:

- a) la **RESTRIZIONE**, che consiste nell'includere nell'esperimento solo soggetti con la stessa modalità della variabile "di disturbo" (per esempio, soggetti con lo stesso sesso, la stessa età o lo stesso peso);
- b) la **RANDOMIZZAZIONE**, cioè l'assegnazione dei soggetti ai trattamenti in modo casuale. La randomizzazione minimizza gli errori sistematici di selezione. Si può supporre, infatti, che le variabili "di disturbo" note e non note siano distribuite in modo casuale nei diversi trattamenti, e quindi non influenzino i trattamenti;
- c) l'**INCLUSIONE** della variabile di disturbo come uno dei fattori dell'esperimento, per esempio costruendo degli strati (o blocchi) in cui i soggetti siano omogenei rispetto alla variabile "di disturbo".

In fase di analisi statistica dei dati il controllo delle variabili "di disturbo" può essere effettuato con i metodi di analisi della covarianza.

2. DISEGNO COMPLETAMENTE RANDOMIZZATO (Francesca Menniti-Ippolito, Maria Puopolo, Flavia Chiarotti)

2.1. DEFINIZIONE

E' il disegno sperimentale più semplice dal punto di vista dell'assegnazione dei diversi trattamenti alle unità sperimentali e per quanto riguarda l'analisi statistica.

Nel disegno sperimentale completamente randomizzato i soggetti vengono assegnati casualmente ai diversi trattamenti. Non è necessario attribuire a ogni trattamento un uguale numero di unità sperimentali anche se le numerosità uguali (bilanciamento) semplificano poi l'analisi (il metodo è più robusto contro violazioni dell'assunto di omogeneità della varianza quando il disegno è bilanciato).

Quando il fattore di cui si vuole valutare l'effetto sulla variabile risposta è un fattore per il quale lo sperimentatore stabilisce sia il numero che il tipo di livelli (l'esperimento contiene, quindi, tutti i livelli di interesse del fattore in esame), esso viene detto **FATTORE AD EFFETTI FISSI**. Il modello statistico che ne consegue viene detto **MODELLO AD EFFETTI FISSI**.

In tale modello l'unica componente ad effetti casuali è l'errore o, più propriamente, la componente residua, che esprime l'effetto del caso e delle variabili che non è possibile controllare nel corso dell'esperimento.

ESEMPIO

Obiettivo: confronto tra i diversi trattamenti per smettere di fumare.

Esperimento: 45 fumatori sottoposti a un trattamento per smettere di fumare.

Variabile indipendente ---> Tipo di terapia (trattamento)

Variabile dipendente ---> Numero di sigarette fumate sei mesi dopo la fine del trattamento (risposta)

Vengono sperimentati tre trattamenti per smettere di fumare:

A_1 : terapia comportamentale

A_2 : ipnosi

A_3 : farmaco

Il fattore A (trattamento per smettere di fumare) viene detto ad effetti fissi se i tre livelli considerati (terapia comportamentale, ipnosi e farmaco) esauriscono l'interesse dello

sperimentatore.

In generale:

A_i livello i -mo del fattore A , con $i=1, \dots, 3$

x_{ij} numero di sigarette fumate dal soggetto j -mo sottoposto al trattamento i -mo

L'ipotesi da verificare è

$$H_0: \mu_1 = \mu_2 = \mu_3$$

dove gli μ_i per $i=1,2,3$ rappresentano il consumo medio di sigarette nei soggetti sottoposti ai tre trattamenti.

Si vuole cioè verificare se il consumo medio di sigarette dei soggetti sottoposti ai tre diversi trattamenti non vari.

Assegniamo i 45 fumatori ai tre trattamenti, 15 per ogni terapia.

A_1	A_2	A_3
$x_{1,1} = 2$	$x_{2,1} = 8$	$x_{3,1} = 3$
$x_{1,2} = 0$	$x_{2,2} = 10$	$x_{3,2} = 9$
$x_{1,3} = 3$	$x_{2,3} = 5$	$x_{3,3} = 8$
...	...	...
...	...	...
$x_{1,15} = 4$	$x_{2,15} = 6$	$x_{3,15} = 9$
$\bar{x}_{1.} = 3$	$\bar{x}_{2.} = 7$	$\bar{x}_{3.} = 8$

e con media generale $\bar{x}_{..} = 6$.

Quindi, per esempio, $x_{2,3}$ misura:

- effetti del trattamento A_2 ;
- effetti specifici al soggetto 3 in A_2 ;
- effetti attribuibili alle fluttuazioni casuali nel comportamento del soggetto 3;
- effetti attribuibili alle condizioni ambientali.

Formalmente possiamo esprimere un'equazione del modello sperimentale, in cui ciascuno dei 45 valori misurati è la somma:

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \text{dove}$$

- μ media generale (costante, riflette il livello di base di sigarette fumate);
- α_i effetto del livello i del fattore A, uguale a $\mu_i - \mu$ (costante all'interno di ogni trattamento);
- ϵ_{ij} residuo associato a x_{ij} , uguale a $x_{ij} - \mu - \alpha_i$.

Il residuo comprende tutti gli effetti non dovuti al trattamento.

Lo sforzo dello sperimentatore sta proprio nel minimizzare l'errore, usando un disegno appropriato e opportuni controlli sperimentali. Un modo è quello di isolare le fonti di variazione che sarebbero altrimenti incluse nel residuo (per esempio, ripartendo le unità sperimentali in blocchi).

2.2. UN SOLO FATTORE A p LIVELLI

2.2.1. Modello sperimentale

Ogni valore ottenuto in un esperimento è composto da tutte le fonti di variabilità che lo hanno determinato e che sono:

- a) effetto medio generale;
- b) effetto del trattamento (livello del fattore);
- c) effetto proprio del soggetto che ha ricevuto quel trattamento;
- d) effetto attribuibile ad altre variabili che non è possibile controllare (per esempio, condizioni ambientali).

Quindi per confrontare le medie di campioni sottoposti a diversi livelli di un trattamento e misurare l'effetto del trattamento è necessario tener conto di tutte queste fonti di variabilità.

Il modello sperimentale che ne deriva è

$$x_{il} = \mu + \alpha_i + \epsilon_{il} = \mu_i + \epsilon_{il} \quad (i=1,\dots,p; l=1,\dots,n)$$

I valori dei parametri μ e α_i e della varianza di ϵ_{il} sono sconosciuti, ma possono essere stimati dai dati campionari:

componente del modello

stima campionaria

μ livello di base

$\bar{x}_{..}$ media generale

α_i variazione sul livello di base attribuibile al trattamento i-mo

$\bar{x}_{i.} - \bar{x}_{..}$ $\bar{x}_{i.}$ media delle osservazioni nel trattamento i-mo

ϵ_{il} somma degli effetti dei fattori non controllabili per il soggetto l-mo nel trattamento i-mo

$\text{Var}(x_{il} - \bar{x}_{i.})$ varianza del residuo

Le assunzioni alla base del modello ad effetti fissi nel disegno completamente randomizzato sono:

- a) l'equazione del modello $x_{il} = \mu + \alpha_i + \epsilon_{il}$ ($i = 1, \dots, p; l = 1, \dots, n$) esprime la somma di tutte le fonti di variazione che influenzano x_{il} ;
- b) l'esperimento contiene tutti i livelli del fattore di interesse. Inoltre, $\alpha_i = \mu_i - \mu$, è un effetto costante per tutte le osservazioni relative al trattamento i-mo e quindi la $\sum \alpha_i = \sum (\mu_i - \mu) = 0$;
- c) i residui ϵ_{il} :
 - 1) sono indipendenti tra loro;
 - 2) hanno media = 0;
 - 3) hanno varianza = σ^2_{ϵ} ;
 - 4) sono distribuiti normalmente per ogni trattamento.

L'ipotesi che si vuole verificare è

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ non c'è né aumento né diminuzione rispetto al livello di base in nessuno dei gruppi sottoposti ai diversi livelli del fattore

ossia

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

il livello medio della risposta non varia nei gruppi che sperimentano i diversi livelli del fattore.

Per confrontare le medie di gruppi sottoposti a diversi livelli di un fattore (per valutare cioè l'effetto del fattore) è necessario tenere conto di tutte le fonti di variabilità. L'ANALISI DELLA VARIANZA (ANOVA) consiste nel confronto fra le diverse medie.

2.2.2. Scomposizione della devianza

La VARIANZA TOTALE degli np valori in un esperimento è

$$\frac{\sum_{i=1}^p \sum_{l=1}^n (x_{il} - \bar{x}_{..})^2}{np}$$

La varianza totale può essere scomposta nelle sue componenti che dipendono dal tipo di disegno sperimentale e, quindi, dal modello statistico che ne consegue.

In realtà, ciò che viene scomposto è la DEVIANZA TOTALE

$$\sum_{i=1}^p \sum_{l=1}^n (x_{il} - \bar{x}_{..})^2$$

Aggiungendo e sottraendo $\bar{x}_{i.}$ da $(x_{il} - \bar{x}_{..})$ si ottiene

$$x_{il} - \bar{x}_{..} = (\bar{x}_{i.} - \bar{x}_{..}) + (x_{il} - \bar{x}_{i.})$$

ed elevando al quadrato i due membri dell'equazione si ha

$$(x_{il} - \bar{x}_{..})^2 = [(\bar{x}_{i.} - \bar{x}_{..}) + (x_{il} - \bar{x}_{i.})]^2$$

Sommando gli np valori si ha

$$\begin{aligned}
\sum_{i=1}^p \sum_{l=1}^n (x_{il} - \bar{x}_{..})^2 &= \sum_{i=1}^p \sum_{l=1}^n [(\bar{x}_{i.} - \bar{x}_{..}) + (x_{il} - \bar{x}_{i.})]^2 = \\
&= \sum_{i=1}^p \sum_{l=1}^n [(\bar{x}_{i.} - \bar{x}_{..})^2 + 2(\bar{x}_{i.} - \bar{x}_{..})(x_{il} - \bar{x}_{i.}) + (x_{il} - \bar{x}_{i.})^2] = \\
&= \sum_{i=1}^p \sum_{l=1}^n (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_{i=1}^p \sum_{l=1}^n 2(\bar{x}_{i.} - \bar{x}_{..})(x_{il} - \bar{x}_{i.}) + \sum_{i=1}^p \sum_{l=1}^n (x_{il} - \bar{x}_{i.})^2
\end{aligned}$$

In particolare, si ha

$$\sum_{i=1}^p \sum_{l=1}^n 2(\bar{x}_{i.} - \bar{x}_{..})(x_{il} - \bar{x}_{i.}) = 2 \sum_{i=1}^p (\bar{x}_{i.} - \bar{x}_{..}) \sum_{l=1}^n (x_{il} - \bar{x}_{i.})$$

Poiché la somma degli scarti dalla media è uguale a zero, si ha

$$\sum_{i=1}^p \sum_{l=1}^n 2(\bar{x}_{i.} - \bar{x}_{..})(x_{il} - \bar{x}_{i.}) = 0$$

da cui segue che

$$\sum_{i=1}^p \sum_{l=1}^n (x_{il} - \bar{x}_{..})^2 = n \sum_{i=1}^p (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_{i=1}^p \sum_{l=1}^n (x_{il} - \bar{x}_{i.})^2$$

DEVIANZA TOTALE = DEVIANZA TRA GRUPPI + DEVIANZA ENTRO GRUPPI

Usando la notazione anglosassone questa relazione può essere espressa come

$$SS_{Tot} = SS_{Between} + SS_{Within}$$

dove SS è l'acronimo di *Sum of Squares*.

In particolare, nel disegno completamente randomizzato con un solo fattore ad effetti fissi la devianza *Between* coincide con la devianza tra le medie nei livelli del fattore,

e la devianza *Within* coincide con la devianza residua. Abbiamo cioè $SS_{\text{Between}} = SS_A$ (avendo indicato il fattore con A), e $SS_{\text{Within}} = SS_{\text{Res}}$.

La devianza tra i gruppi rappresenta quindi la variazione casuale più l'effetto del fattore in studio, mentre la devianza nei gruppi rappresenta la sola variazione casuale.

2.2.3. Gradi di libertà

Definiamo i gradi di libertà come il numero di osservazioni con valori che possono essere assegnati arbitrariamente.

I gradi di libertà (*degrees of freedom* = df) associati alla devianza tra i gruppi

$$n \sum_{i=1}^p (\bar{x}_i - \bar{x}_{..})^2$$

sono $p-1$ (numero dei livelli del fattore meno 1).

Per esempio, $p=3$ medie campionarie, ognuna basata su n osservazioni, sono legate alla media generale dall'equazione

$$\frac{\bar{x}_{1.} + \bar{x}_{2.} + \bar{x}_{3.}}{3} = \bar{x}_{..}$$

Se $\bar{x}_{..} = 6$, $\bar{x}_{1.} = 3$ e $\bar{x}_{2.} = 7$, allora è $\bar{x}_{3.} = 8$. Infatti,

$$\frac{(3) + (7) + 8}{(3)} = (6)$$

Allo stesso modo, se fosse $\bar{x}_{1.} = 5$ e $\bar{x}_{2.} = 1$, allora $\bar{x}_{3.}$ dovrebbe essere uguale a 12, perché

$$\frac{(5) + (1) + 12}{(3)} = (6)$$

Dato il valore della media generale, è possibile assegnare un qualunque valore a due delle tre medie, e il valore della terza media risulterà determinato di conseguenza.

Allo stesso modo i df della devianza nei gruppi sono $p(n-1)$.

La devianza nei gruppi è

$$\sum_{i=1}^p \sum_{l=1}^n (x_{il} - \bar{x}_i)^2$$

Se $p=3$ e $n=8$ per ognuno dei tre trattamenti dovremo calcolare 8 deviazioni

$$(x_{i1} - \bar{x}_i)^2 + (x_{i2} - \bar{x}_i)^2 + \dots + (x_{i8} - \bar{x}_i)^2$$

Ogni valore relativo al trattamento i-mo è in relazione con la i-ma media

$$\frac{x_{i1} + x_{i2} + \dots + x_{i8}}{8} = \bar{x}_i$$

Quindi, come sopra, 7 punteggi possono prendere qualsiasi valore, mentre l'ottavo è determinato dal fatto che la somma dei punteggi diviso per 8 deve essere uguale a $\bar{x}_i$.

Quindi ci sono $n-1 = 8-1 = 7$ df, di conseguenza il numero totale di df è $p(n-1)$.

Se le unità sperimentali (n_i) sono diverse per ogni trattamento il numero di df della devianza nei gruppi è

$$(n_1-1) + (n_2-1) + \dots + (n_p-1) = N-p$$

dove N è il numero totale delle osservazioni.

Il numero di gradi di libertà della devianza totale è

$$(N-p) + (p-1) = N-1$$

Le varianze delle varie componenti si ottengono dividendo le devianze per i relativi df

Varianza = Devianza/df.

Usando la notazione anglosassone abbiamo

$$MS = SS/df$$

dove MS è l'acronimo di *Mean Square*.

In particolare, abbiamo:

$$MS_{Tot} = SS_{Tot}/(np-1); \quad MS_A = SS_A/(p-1); \quad MS_{Res} = SS_{Res}/[p(n-1)].$$

2.2.4. Analisi della Varianza

Per verificare l'ipotesi sulle medie

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_p \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2 \neq \dots \neq \mu_p \quad \text{oppure}$$

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0 \quad \text{vs} \quad H_1 : \alpha_i \neq 0, \text{ per qualche } i$$

si usa il test F di Fisher

$$F = \frac{MS_A}{MS_{Res}} \sim F_{p-1, p(n-1)}$$

Si può riassumere quanto detto in precedenza nella seguente tabella dell'ANOVA:

Fonte di variazione	SS	df	MS	F
Tra gruppi = Fattore A	SS_A	$p-1$	$MS_A = SS_A/(p-1)$	MS_A/MS_{Res}
Nei gruppi = Residuo	SS_{Res}	$p(n-1)$	$MS_{Res} = SS_{Res}/[p(n-1)]$	
Totale	SS_{Tot}	$np-1$		

I valori attesi delle varianze sono:

$$E(MS_A) = \sigma_e^2 + n \sigma_\alpha^2$$

$$E(MS_{Res}) = \sigma_e^2$$

Quando l'ipotesi nulla è vera si ha:

$$E(MS_A) = \sigma_e^2$$

Di conseguenza, sotto l'ipotesi nulla la F diventa il rapporto di due stime della varianza ciascuna delle quali ha lo stesso valore atteso σ_e^2 . Pertanto, valori della $F \approx 1$ portano a non rifiutare l'ipotesi nulla, mentre valori della F sufficientemente maggiori di 1 portano a rifiutare l'ipotesi nulla, e ad affermare che le differenze osservate tra le medie non sono plausibilmente dovute al solo effetto del caso.

ESEMPIO

32 soggetti sono stati divisi casualmente in 4 gruppi di 8 soggetti ciascuno.

Il fattore in studio è la privazione di sonno. I livelli di privazione di sonno considerati sono 4: 12, 24, 36 e 48 ore.

La variabile risposta è una misura della fermezza della mano.

	A ₁ (12)	A ₂ (24)	A ₃ (36)	A ₄ (48)
	3	4	7	7
	6	5	8	8
	3	4	7	9
	3	3	6	8
	1	2	5	10
	2	3	6	10
	2	4	5	9
	2	3	6	11
$\bar{x}_i$	2.75	3.50	6.25	9.00
σ_i	1.49	0.93	1.04	1.31
$\sum_{l=1, \dots, n} x_{il}$	22	28	50	72

$$\sigma_i = \sqrt{\frac{\sum_{l=1}^n x_{il}^2 - \frac{\left[\sum_{l=1}^n x_{il} \right]^2}{n}}{n-1}}$$

$$\sum_{i=1}^p \sum_{l=1}^n x_{il} = 3+6+3+\dots+10+9+11 = 172$$

$$\frac{(\sum_{i=1}^p \sum_{l=1}^n x_{il})^2}{np} = [X] = \frac{172^2}{8 \cdot 4} = 924.5$$

$$\sum_{i=1}^p \sum_{l=1}^n x_{il}^2 = [AS] = 3^2+6^2+\dots+10^2+9^2+11^2 = 1160$$

$$\sum_{i=1}^p \frac{(\sum_{l=1}^n x_{il})^2}{n} = [A] = \frac{22^2+28^2+50^2+72^2}{8} = 1119$$

$$SS_{TOT} = [AS] - [X] = 1160-924.5 = 235.5$$

$$SS_A = [A] - [X] = 1119-924.5 = 194.5$$

$$SS_{RES} = [AS] - [A] = 1160-1119 = 41$$

Ne risulta la seguente tabella dell'ANOVA:

Fonte di variazione	SS	df	MS	F	p
Tra gruppi	194.5	p-1=3	64.833	44.28	< 0.0001
Nei gruppi	41.0	p(n-1)=28	1.464		
Totale	235.5	np-1=31			

Si può concludere che il caso non può essere considerato una spiegazione plausibile delle differenze osservate nelle medie della fermezza della mano, e che tali differenze (statisticamente significative) sono presumibilmente dovute al diverso livello di privazione del sonno. I confronti multipli servono ad indagare quali medie in realtà differiscano tra loro.

2.2.5. Assunti per l'applicazione del test F

- a) Le osservazioni provengono da popolazioni distribuite secondo una distribuzione normale (assunto di NORMALITÀ);
- b) Le osservazioni sono campioni casuali della popolazione con varianze omogenee (assunto di OMOSCHEDASTICITÀ);
- c) Il numeratore e il denominatore del rapporto F sono indipendenti.

2.3. UN SOLO FATTORE A 2 LIVELLI

2.3.1. Test t di Student per gruppi indipendenti

Supponiamo di voler confrontare le medie dei valori osservati su soggetti indipendenti, appartenenti a due diversi gruppi di trattamento. Si può utilizzare il test t di Student per gruppi indipendenti, il cui risultato coincide con quello del test F mostrato in precedenza.

Siano n_1 e n_2 le numerosità dei gruppi sottoposti ai trattamenti A_1 e A_2 , rispettivamente. Siano inoltre $\bar{x}_1$ e $\bar{x}_2$ le medie delle risposte nei due gruppi.

Siano l'ipotesi nulla e l'ipotesi alternativa

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Per l'applicazione del test t di Student, si costruisce la statistica del test

$$t = \frac{\bar{x}_{1.} - \bar{x}_{2.}}{\sqrt{\frac{\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{n_1+n_2-2}$$

Nel caso particolare in cui sia $n_1 = n_2 = n$, si ha

$$t = \frac{\bar{x}_{1.} - \bar{x}_{2.}}{\sqrt{\frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 + \sum_{j=1}^n (x_{2j} - \bar{x}_2)^2}{n(n-1)}}} \sim t_{2(n-1)}$$

Si noti che

$$t_{n_1+n_2-2}^2 = F_{1, n_1+n_2-2}$$

ESEMPIO

Consideriamo l'esempio di cui al punto 2.2.4., esaminando solo i due livelli di privazione di sonno di 12 e 48 ore.

Abbiamo

	12 ore	48 ore
$\bar{n}_i$	8	8
$\bar{x}_i$	2.75	9.00
SS_i	15.50	12.00

Abbiamo

$$t = \frac{6.25}{\sqrt{\frac{15.5 + 12.00}{8 + 8 - 2} \left(\frac{1}{8} + \frac{1}{8} \right)}} = \frac{6.25}{\sqrt{\frac{15.5 + 12.00}{8 \cdot 7}}} = \frac{6.25}{\sqrt{0.49}} = \frac{6.25}{0.70} = 8.92 \sim t_{14}$$

Il valore della t è significativo allo 0.0001, e quindi i due trattamenti influenzano in modo significativamente diverso la fermezza della mano.

Lo stesso disegno, come abbiamo detto, può essere analizzato con l'analisi della varianza per gruppi randomizzati, ottenendo

$$\text{Termine di correzione} = 94^2/16 = 552.25$$

$$SS_{\text{Tot}} = (3^2 + 6^2 + \dots + 9^2 + 11^2) - 552.25 = 736 - 552.25 = 183.75$$

$$SS_A = (22^2 + 72^2)/8 - 552.25 = 708.5 - 552.25 = 156.25$$

$$SS_{\text{Res}} = (3^2 + 6^2 + \dots + 9^2 + 11^2) - (22^2 + 72^2)/8 = 736 - 708.5 = 27.5$$

Ne risulta la seguente tabella dell'ANOVA:

Fonte di variazione	SS	df	MS	F	p
Tra gruppi	156.25	1	156.25	79.55	<0.0001
Nei gruppi	27.50	14	1.96		

Come si può notare, il valore della $F_{1,p(n-1)}$ è identico al valore della $t_{p(n-1)}$ elevato al quadrato

$$8.92^2 = 79.55$$

2.3.2. Assunti per l'applicazione del test t

Per l'applicazione del test t di Student valgono gli stessi assunti di NORMALITÀ,

OMOGENEITÀ DELLA VARIANZA e INDIPENDENZA richiesti per l'applicazione del test F di Fisher. Inoltre, anche nel caso della t di Student il bilanciamento dell'esperimento rende il test più robusto nei confronti della violazione degli assunti.

2.4. DISEGNO FATTORIALE

2.4.1. Modello sperimentale e scomposizione della devianza

Quando i trattamenti di interesse sono tutte le possibili diverse combinazioni di livelli di più fattori, ognuno presente con un solo livello in ogni combinazione, si ha un disegno fattoriale completo. Il disegno fattoriale permette la valutazione simultanea di 2 o più fattori.

Consideriamo un fattore A con p livelli ($A_1, \dots, A_p$) e un fattore B con q livelli ($B_1, \dots, B_q$).

A_1			...	A_p		
B_1	...	B_q		B_1	...	B_q
X_{111}	...	X_{1q1}	...	X_{p11}	...	X_{pq1}
X_{112}	...	X_{1q2}	...	X_{p12}	...	X_{pq2}
X_{113}	...	X_{1q3}	...	X_{p13}	...	X_{pq3}
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
X_{11n}	...	X_{1qn}	...	X_{p1n}	...	X_{pqn}

Ognuno di questi valori può essere espresso come

$$x_{ijl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijl} \quad \text{con } i=1,\dots,p; j=1,\dots,q; l=1,\dots,n$$

dove

μ media generale (che può essere stimata dai dati campionari con $\bar{x}_{\dots}$);

α_i effetto del livello i-mo del fattore A, uguale a $\mu_{i.} - \mu$ (stimabile con $\bar{x}_{i.} - \bar{x}_{\dots}$);

β_j effetto del livello j-mo del fattore B, uguale a $\mu_{.j} - \mu$ (stimabile con $\bar{x}_{.j} - \bar{x}_{\dots}$);