

7. Approfondimenti sull'ANOVA

La descrizione matematica delle tecniche presentate in questo capitolo esula dai limiti propri di questo testo. Tuttavia tali tecniche sono di grande importanza ed in letteratura vi si fa frequentemente ricorso. La scelta di questa introduzione è di fornire quelle informazioni qualitative che possano permettere al lettore non tanto di elaborare direttamente i test, quanto piuttosto di comprenderne il significato se incontrati nella lettura di studi e pubblicazioni. Con questa scelta non si esclude fra l'altro che, con l'utilizzo di software statistico, anche il lettore possa sperimentare in proprio le tecniche presentate, ed utilizzarle nel corso di proprie sperimentazioni o ricerche.

Indicazioni dettagliate per tutte le tecniche di calcolo citate in questo capitolo si possono trovare in R. R. Sokal e F. J. Rohlf, 1982. In particolare per le tecniche non parametriche si può consultare il classico S. Siegel e N. J. Castellan Jr, 1956.

7.1. Confronti fra i gruppi nell'ANOVA a singolo criterio di classificazione

7.1.1. Confronti appaiati non pianificati

In § 6.3.3. abbiamo formalizzato le ipotesi in opposizione in una ANOVA:

- H_0 : I gruppi sono tratti da popolazioni aventi tutte la stessa media parametrica μ ;
- H_1 : I gruppi sono tratti da popolazioni di cui *almeno due* hanno media parametrica diversa fra loro.

Abbiamo ripetutamente osservato (in chiusura del § 6.3.3., in chiusura di § 6.3.4. commentando il risultato del Box 6.2 e in chiusura di § 6.4.2. commentando il risultato del Box 6.3) che una ipotesi alternativa come la H_1 comporta la necessità di ulteriori indagini in caso di significatività di F .

Un possibile tipo di indagine supplementare consiste nel sottoporre a test in modo sistematico la differenza di ogni gruppo con ciascun altro. Dunque si dà corso a un esame sistematico di tutte le possibili coppie di gruppi; per questo motivo le tecniche di questo tipo sono note come *confronti appaiati* (non pianificati).

Citiamo quattro di queste tecniche; ognuna di esse ha un ambito di applicazione elettivo; non è tuttavia scorretto, nei casi dai confini incerti, eseguire l'analisi con tutte le tecniche e scegliere quella più sensibile. Il responso del test può essere sia numerico (in genere in forma tabellare) che grafico.

<i>Tecniche</i>	<i>Ambito di applicazione elettivo</i>
metodo T (*)	Dimensioni uguali del campione
metodo T' (*)	Dimensioni simili dei campioni
metodo $GT2$ (*)	Dimensioni differenti dei campioni
metodo <i>Tukey – Kramer</i>	Dimensioni differenti dei campioni
(*) Con esito sia tabellare che grafico	

Tab. 7.1

Cominciamo ad esemplificare partendo dall'ANOVA sviluppata nel Box 6.2.

Abbiamo quattro insiemi di dati, relativi rispettivamente al gruppo di controllo C e a tre gruppi sperimentali S_1 , S_2 e S_3 . La significatività di F dimostra la differenza fra almeno due dei gruppi. Quali? Come si articolano effettivamente le differenze?

Sottoponendo i dati ad una analisi ulteriore col metodo T (le dimensioni dei gruppi sono uguali) otteniamo un responso di questo tipo:

S_1			
S_2		*	
S_3			
	C	S_1	S_2

Tab. 7.2

Ogni cella della griglia corrisponde ad un confronto appaiato, i cui termini si leggono nelle intestazioni di riga e colonna. Valgono le consuete convenzioni circa l'uso degli asterischi (in questo caso per non appesantire la tabella la mancanza di significatività è segnalata da una cella vuota). Dunque dalla Tab. 7.2 ricaviamo immediatamente l'informazione che i gruppi S_1 e S_2 sono significativamente differenti a livello 0.05.

Una informazione ancora più ricca può talvolta essere desunta da una lettura dei grafici eventualmente prodotti dalle prime tre procedure di Tab. 7.1. Nel nostro esempio abbiamo quello di Fig. 7.1.

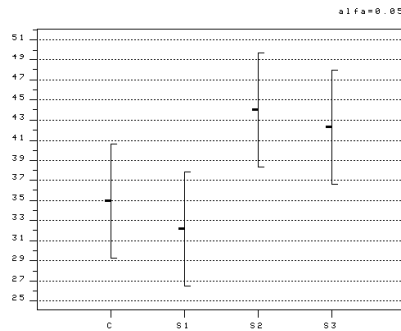


Fig. 7.1

Questo tipo di grafico è simile a quelli esemplificati nel Box 1.1 Parte c. Anche in questo caso abbiamo i dati dei vari gruppi sintetizzati da graffe, in cui la tacca centrale rappresenta la media. Tuttavia nel grafico del Box 1.1 l'ampiezza delle graffe corrispondeva alla deviazione standard del gruppo, mentre in questo caso corrisponde ad una statistica differente (sempre legata indirettamente alla dispersione dei dati) che è più *sensibile* alle differenze fra i gruppi. In pratica la lettura del grafico di Fig. 7.1 avviene in questo modo: se fra le graffe di due gruppi non vi è una regione di sovrapposizione allora vi è differenza significativa, altrimenti no. Anche in questo caso gli unici due gruppi che presentano una differenza significativa sono S_1 e S_2 . Il grafico in Fig. 7.1 contiene una informazione più ricca rispetto a quella della laconica Tab. 7.2, ma ne ripareremo in § 7. 1. 2.

Esemplifichiamo ora con i dati del problema del Box 6.3; l'unica procedura sufficientemente sensibile per l'individuazione delle differenze presenti nei gruppi è quella di *Tukey – Kramer*, e dà un esito sintetizzato in Tab. 7.3.: l'unica differenza significativa (a livello 0.05) è fra i gruppi *C* ed *E*.

B				
C				
D				
E			*	
	A	B	C	D

Tab. 7.3

Tuttavia anche dalla procedura grafica connessa alla procedura *GT2* otteniamo un risultato molto vicino a quello di Tab.7.3 (salvo una piccola zona di sovrapposizione fra le graffe corrispondenti ai gruppi *C* ed *E*, come mostra Fig. 7.2).

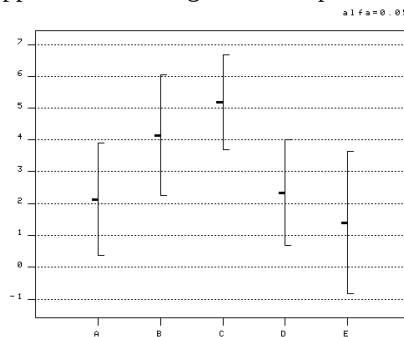


Fig. 7.2

Questo supplemento di indagine chiarisce dunque che è dimostrato un differente atteggiamento nei confronti delle assenze fra gli studenti provenienti dalle scuole *C* ed *E*.

7.1.2. Confronti multipli non pianificati

Talvolta i dati sembrano suggerire una suddivisione dei gruppi in super-gruppi omogenei. Ad esempio Fig. 7.1 sembra suggerire una omogeneità di risultati fra i gruppi *C* ed S_1 in opposizione a quelli dei gruppi S_2 e S_3 a loro volta abbastanza omogenei.

In queste situazioni potremmo desiderare di sottoporre a test la significatività della differenza fra i vari super-gruppi; nel nostro caso vorremmo sottoporre a test la differenza: $\{C, S_1\}$ vs. $\{S_2, S_3\}$.

Esistono numerosissime procedure che svolgono questo lavoro, ciascuna delle quali ha un proprio ambito di applicazione elettivo. In questa nota sommaria ne citeremo solo una, nota sotto il nome di procedura *SS-STP*. Essa ricombina le somme dei quadrati *SS* dei vari gruppi e le sottopone ad un opportuno test.

Nel caso specifico dell'esempio, otteniamo fra i due super-gruppi una differenza significativa a livello 0.05.

Una conclusione (affrettata) potrebbe dunque essere questa:

le condizioni sperimentali a cui è stato sottoposto il gruppo S_1 si sono rivelate inefficaci, mentre sostanzialmente equivalenti sono quelle dei gruppi S_2 e S_3 .

Tuttavia una tale conclusione non tiene conto del fatto che l'unica differenza significativa fra i gruppi è quella fra S_1 e S_2 .

Questo suggerisce il sospetto che la differenza fra i super-gruppi non sia dovuta tanto alla bontà dei trattamenti S_2, S_3 , rispetto al controllo C , quanto piuttosto agli esiti assai scarsi del gruppo S_1 . Il sospetto è confermato dal fatto che sottoponendo a test la differenza $\{C\}$ vs. $\{S_2, S_3\}$ essa non risulta significativa. Ciò permette di concludere che la sperimentazione non ha prodotto risultati significativi.

7. 2. Due modelli di ANOVA

7. 2. 1. ANOVA di modello I e II

Torniamo un'altra volta al problema analizzato nel Box 6.3, relativo alle assenze di alunni provenienti da differenti scuole. Il quesito può essere posto in due modi che differiscono sottilmente.

- La provenienza dalle scuole $A, B, C, \dots$ produce diversi gradi di affezione alle lezioni ?
- La provenienza da scuole differenti può produrre differenti gradi di affezione alle lezioni?

Nel primo caso noi sospettiamo che in ben determinate scuole vi sia un orientamento educativo generale che produce un grado di affezione alle lezioni differente da quello di ben precise altre scuole con orientamenti educativi differenti. Insomma, ipotizziamo che il grado di affezione scolastica sia il risultato diretto di un determinato orientamento educativo. In termini generali ipotizziamo che l'affezione sia un *effetto fisso* dovuto ad un ben preciso *trattamento*. Con l'espressione *effetto fisso* intendiamo che si tratta di una conseguenza sistematica di un orientamento educativo che indichiamo col termine generale di *trattamento*. In questo primo caso, per condurre l'indagine che ci interessa dobbiamo scegliere con cura le scuole che supponiamo produrre studenti affezionati e quelle che invece supponiamo produrre studenti disaffezionati. In questo caso la variabilità fra i gruppi verrà considerata un *effetto dovuto ad un trattamento*.

Il secondo modo di porre il quesito è più generale: si ipotizza cioè che scuole differenti possano produrre, per una serie concomitante di fattori imprecisati e incontrollabili, differenti gradi di affezione scolastica; però, esattamente, non siamo in grado di prevedere scuola per scuola quale sia il differente grado di affezione prodotto. In questa situazione si parla di *effetti casuali*, in opposizione agli *effetti fissi* del caso precedente. Ovviamente se vogliamo indagare su effetti casuali l'indagine va condotta campionando in modo casuale gli istituti di provenienza (se non è possibile un'indagine esaustiva). Le scuole estratte casualmente non rappresentano esattamente se stesse e la propria impostazione educativa, ma campioni dei differenti e casuali modi possibili di condizionare il grado di affezione alle lezioni degli studenti. In questo caso la variabilità fra i gruppi non è un effetto dovuto a un trattamento, ma una semplice *componente aggiunta di varianza fra i gruppi*, di cui non siamo in grado di precisare esattamente le cause, se non riferendoci in modo generico a differenze ambientali imprecisate che inevitabilmente vengono a formarsi.

A seconda del tipo di impostazione si parla di ANOVA di modello I o II. Più precisamente:

- Se si ipotizza che eventuali differenze fra i gruppi siano effetti fissi dovuti ad un trattamento si parla di ANOVA di modello I;
- Se si ipotizza che eventuali differenze fra i gruppi siano dovute ad effetti casuali si parla di ANOVA di modello II.

7. 2. 2. Struttura della variazione nei due modelli; stima della componente aggiunta di varianza

Ogni singola variata trattata nel corso di un'analisi della varianza può considerarsi decomposta in una somma di tre componenti:

- una deriva dall'appartenenza della variata alla grande popolazione da cui sono stati tratti i vari gruppi; tale componente è espressa dalla la media di questa popolazione;
- una seconda componente deriva dall'appartenenza della variata al proprio gruppo; essa è caratterizzato dalla deviazione rispetto alla media generale tipica del gruppo di appartenenza;
- la terza componente deriva dalla variabilità individuale all'interno del gruppo di appartenenza, ed è espressa dalla deviazione individuale

Per formalizzare quanto detto, consideriamo una singola variata Y , diciamo la j – esima del gruppo i – esimo, ed indichiamola con Y_{ij} . Questa sarà data dalla somma algebrica delle tre componenti elencate sopra:

$$\begin{aligned} Y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \\ Y_{ij} &= \mu + A_i + \varepsilon_{ij} \end{aligned} \quad (7.1)$$

La prima delle (7.1) si riferisce ad un'ANOVA di modello I, mentre la seconda ad un'ANOVA di modello II.

Nelle (7.1) μ è la media della popolazione da cui sono estratti i gruppi, α_i è la deviazione della media del gruppo i – esimo dalla media generale (nel caso di effetto fisso), A_i è la deviazione della media del gruppo i – esimo dalla media generale (nel caso di effetto casuale), mentre ε_{ij} è la deviazione individuale dell'elemento j – esimo del gruppo i – esimo. Sia α che A che ε possono essere deviazioni sia positive che negative; ecco perché si è parlato di somma algebrica.

I calcoli necessari all'elaborazione dell'ANOVA dei due modelli sono in linea di massima uguali (almeno per tutto ciò che abbiamo menzionato fino a qui). L'attenzione va posta principalmente nelle operazioni di campionamento (come è stato sommariamente esemplificato in § 7. 2. 1.) e in modo particolare nell'interpretazione dei risultati.

Nel caso di una ANOVA di modello II i software statistici riportano in genere una interessante stima s_A^2 della componente aggiunta di varianza, sia in termini assoluti che percentuali. Ad esempio nel caso dell'indagine sulle assenze abbiamo:

$$s_A^2 = 1.970$$

Percentuale di variazione fra i gruppi: 25.3%

Percentuale di variazione all'interno dei gruppi: 74.7%

Si presti attenzione al fatto che in genere il software statistico fornisce queste indicazioni *comunque*, a prescindere cioè dal modello effettivo di ANOVA (I o II), lasciando all'utente la responsabilità di decidere in merito all'appropriatezza della statistica nel contesto.

7. 3. ANOVA a più criteri di classificazione o fattoriali

L'ANOVA fattoriale è la tecnica statistica adeguata per i disegni fattoriali descritti in § 4.4. della Parte metodologica. Quanto segue presuppone la conoscenza di tali disegni e del connesso concetto di interazione (§ 4.4.1.).

7. 3. 1. ANOVA a due vie con o senza replicazione

Se i fattori del disegno sperimentale sono due la corrispondente analisi della varianza si dice *a due vie*.

Supponiamo che per il primo fattore si presentino n modalità e per il secondo m modalità. I dati raccolti nella sperimentazione possono essere raccolti in una griglia con n colonne ed m righe.

	1	2	3	...	n
1				...	
2				...	
3				...	
...				...	
m				...	

In ciascuna delle celle possono immaginarsi raccolti i dati relativi all'incrocio della particolare modalità del primo fattore con la particolare modalità del secondo. Se ad esempio in una indagine sulla riuscita scolastica i fattori indagati sono il sesso (due modalità: M e F) e area geografica di appartenenza (quattro fattori: Nord, Centro, Sud, Isole), nella cella alla colonna M e alla riga Centro saranno ranno idealmente raccolti i dati relativi alla riuscita scolastica dei maschi del centro.

Se per ogni cella abbiamo un certo numero di singole variare Y siamo in presenza di un disegno *con replicazione*. Se invece per ogni cella c'è a disposizione una singola variata Y abbiamo un disegno *senza replicazione*.

7. 3. 2. Fonti di variazione in un'ANOVA fattoriale con replicazione

Cominciamo a parlare dei disegni fattoriali con replicazione.

Sebbene i calcoli necessari per condurre a termine l'analisi della varianza fattoriale siano piuttosto complessi, tuttavia la loro logica non è dissimile da quella dell'ANOVA a singolo criterio di classificazione. Anche in questo caso la somma dei quadrati totale SS_t ed i gradi di libertà totali v_t vengono partiti in modo da scomporre la variabilità totale fra le diverse fonti di variazione.

In un disegno a due fattori le fonti di variazione sono:

- il primo fattore (in quanto si suppone che se il fattore è efficace determina differenze, e quindi variazioni);
- il secondo fattore (analogo motivo)
- l'interazione fra i due fattori (nel senso che le diverse combinazioni dei due fattori possono produrre effetti sinergici diversificati)
- la variabilità individuale o *errore*

Così come nell'ANOVA ad un fattore la varianza fra gruppi è sottoposta a test in rapporto alla varianza all'interno dei gruppi, anche nell'ANOVA fattoriale ogni livello di variazione è sottoposto a test in rapporto all'appropriato livello precedente.

La tavola riassuntiva di un'ANOVA a due vie si presenta come Tab. 7.4

<i>Fonti di Variazione</i>	<i>v</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Fattore 1	...	...	...	...
Fattore 2	...	...	...	...
Interazione (Fattore 1 x Fattore 2)	...	...	...	...
Varianza errore	...	...	...	/
Totale	...	...		

Tab. 7.4

mentre nel caso di tre fattori la tavola riassuntiva è come Tab. 7.5

<i>Fonti di Variazione</i>	<i>v</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Fattore 1	...	...	...	...
Fattore 2	...	...	...	...
Fattore 3	...	...	...	...
Fattore 1 x Fattore 2	...	...	...	...
Fattore 1 x Fattore 3	...	...	...	...
Fattore 2 x Fattore 3	...	...	...	...
Fattore 1 x Fattore 2 x Fattore 3	...	...	...	...
Varianza errore	...	...	...	/
Totale	...	...		

Tab. 7.5

Come si vede sono presenti tutte le interazioni di ordine 2 (cioè fra due fattori) e quella di ordine 3.

Per ciascuno dei livelli di variazione è valutata la significatività della corrispondente statistica *F*. Le relative modalità di calcolo sono condizionate dal tipo di ANOVA, che può essere di modello I se per tutti i fattori si ha effetto fisso, di modello II se per tutti i fattori l'effetto è casuale e di modello misto se sono presenti tutti e due i tipi di effetto.

Questa precisazione vale a sottolineare il fatto che anche utilizzando un software statistico è necessario specificare per ogni fattore il tipo di effetto.

7.3.3. Fonti di variazione in un'ANOVA fattoriale senza replicazione

Introduciamo l'argomento con un esempio.

Si vuole verificare se fra diversi insegnanti della stessa disciplina c'è accordo nella valutazione di differenti tipologie di errori; ad esempio, in un compito di algebra: errore di trascrizione da un passaggio all'altro; errore di calcolo; errore di segno non assimilabile ad errore di calcolo; errore di algebra (e qui potremmo sbizzarrirci...) e così via.

Si prepara un compito con tanti esercizi quanti sono i tipi di errore che si vogliono osservare; tutti gli esercizi saranno uguali (ad esempio, se gli errori sono 6: sei equazioni di pari livello di difficoltà, oppure sei espressioni, oppure sei sistemi...). Il compito non viene effettivamente svolto da uno studente, ma dal gestore del disegno sperimentale che ha cura di inserire deliberatamente in ogni esercizio un solo errore, di tipo differente da esercizio a esercizio.

L'elaborato viene poi fatto correggere da più docenti sulla base di una griglia di valutazione che prevede lo stesso punteggio (poniamo 10 punti) per ciascun esercizio. Raccolte le correzioni, i punteggi per ciascun esercizio di ciascun docente sono inseriti in una tabella in cui le colonne rappresentano i docenti, e le righe gli esercizi, o meglio, i tipi di errore. In ogni cella avremo un solo dato: si tratta dunque di un disegno fattoriale senza replicazione.

Rispetto ai disegni con replicazione, in quelli senza replicazione si assume che l'interazione fra i fattori sia non significativa. L'esito di un'ANOVA fattoriale senza replicazione viene dunque sintetizzato come in Tab. 7.6

<i>Fonti di Variazione</i>	<i>v</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Righe	...	...	...	...
Colonne	...	...	...	...
Varianza residua	...	...	...	/
Totale	...	...		

Tab. 7.6

I disegni senza replicazione hanno il pregio di richiedere minori costi in senso lato: nel caso dell'esempio soprattutto in termini di energie umane; quindi tali disegni possono essere utilizzati con maggior facilità. Tuttavia nel caso dell'esempio precedente con pochi sforzi in più è possibile realizzare un corrispondente disegno fattoriale con replicazione. Il vantaggio che deriva dalla replicazione è quello di poter sottoporre a test anche l'interazione.

L'ANOVA senza replicazione è la tecnica statistica appropriata per i disegni a blocchi randomizzati completi. In tal caso la tabella riassuntiva è del tipo di Tab. 7.7.

Fonti di Variazione	ν	SS	MS	F
Fra blocchi	...	...	...	...
Fra trattamenti	...	...	...	...
Varianza errore	...	...	...	/
Totale	...	...		

Tab. 7.7

Prima di concludere è opportuno rilevare che nel caso particolare in cui le colonne siano due l'ANOVA fattoriale senza replicazione coincide col t – test per dati appaiati illustrato in § 5. 5. 3. e § 5. 5. 4. e si dimostra che fra le corrispondenti statistiche t ed F vale una relazione analoga a quella citata in § 6. 4. 3.

7. 4. ANOVA gerarchica o nidificata

Si tratta della tecnica statistica idonea al trattamento dei dati sperimentali raccolti attraverso un disegno gerarchico (vedi § 4.5. della Parte metodologica).

Prima di descrivere il tipo di risposte che può fornire una analisi della varianza di tipo gerarchico, occorre precisare un'assunzione fondamentale di questa tecnica.

L'ANOVA gerarchica esige che per ciascun livello di annidamento (tranne al più il primo) gli effetti siano non fissi ma casuali. Ciò implica una organizzazione dell'esperimento conforme a questo vincolo.

Riprendiamo l'esempio fornito in § 2. 5. 2. della Parte metodologica.

Parliamo delle due griglie che ogni docente deve preparare: si è già detto che ciascun docente *non* deve preparare due possibili e differenti griglie (che prevedono un effetto *fisso* dovuto appunto alla programmata diversità delle due griglie), ma semplicemente deve ripetere in momenti differenti ed in modo indipendente l'operazione di stesura della griglia dei punteggi. Le due griglie devono cioè esprimere la variabilità *casuale* cui è effettivamente soggetto il docente nel momento in cui prepara una griglia di valutazione. Analogo ragionamento va fatto per tutti gli altri livelli di annidamento.

Il primo livello, nell'esempio di § 2. 5. 2. È quello relativo alla variabile *docente*, può invece avere effetti sia casuali che fissi. Dunque nell'esempio i docenti potranno essere sia scelti in base a loro supposti differenti metodi valutativi (effetto fisso), o casualmente, a rappresentare la variabilità casuale che caratterizza la distribuzione dei metodi valutativi (effetto casuale).

Nel caso che il primo fattore sia ad effetti casuali (come tutti i successivi) si tratta di un'ANOVA pura di modello II, mentre nel caso il primo livello sia ad effetti fissi si tratta di un'ANOVA di modello misto.

Veniamo alle informazioni che dà una tale analisi. Essa scompone la varianza totale in una serie di componenti dovute a ciascuno dei livelli di annidamento elaborati. La varianza di ogni livello è sottoposta a test di significatività rispetto a quella del livello precedente. Per ogni livello l'analisi offre una stima percentuale della varianza totale. Il prospetto finale è come in Tab. 7.8.

Fonti di Variazione	ν	SS	MS	F	Componenti di varianza	
					Mod. Puro	Mod. Misto
Fra docenti	...	...	...	...	... %	/
Fra griglie all'interno dei docenti	...	...	...	...	... %	... %
Fra correzioni all'interno di griglie	...	...	...	...	... %	... %
Fra valutazioni all'interno delle correzioni	...	...	...	...	... %	... %
Fra elaborati all'interno delle valutazioni	...	...	...	/	... %	... %
Totale	...	...				

Tab. 7.8

A questo punto i docenti hanno un quadro preciso che permetterà di individuare quali siano i passaggi in cui i punteggi subiscono la maggior variabilità e potranno intervenire di conseguenza. Per il primo livello è poi possibile operare una analisi comparata dei gruppi come quella descritta in § 7.1.1. oppure in § 7.1.2.

7. 5. Assunzioni dell'ANOVA e test non parametrici sostitutivi

7. 5. 1. Le assunzioni dell'ANOVA

Presentando la statistica F e la sua distribuzione abbiamo visto che alla base di tutto stanno alcune assunzioni che è bene richiamare esplicitamente:

- i campioni devono essere casuali e indipendenti (*casualità e indipendenza*)
- la popolazione da cui sono estratti deve essere distribuita normalmente (*normalità*)
- le varianze delle sottopopolazioni da cui sono tratti i gruppi devono essere omogenee (*omoschedasticità*)

In difetto di una o più di queste ipotesi il risultato può essere più o meno inaffidabile.

A proposito del primo punto non c'è molto da aggiungere, se non l'obbligo di rispettare tali vincoli nel campionamento, ma questa è ordinaria amministrazione.

Riguardo al difetto di normalità occorre precisare che l'ANOVA è un metodo abbastanza *robusto*: in termine è tecnico ed esprime (approssimativamente parlando) la capacità del test di mantenere la propria validità in difetto delle assunzioni

su cui è basato. L'ANOVA non patisce eccessivamente i difetti di normalità, in quanto lavora sulle medie, le quali tendono ad una distribuzione normale anche in difetto di normalità della popolazione sottostante. Solo forti asimmetrie condizionano negativamente l'esito del test. Pertanto basta sincerarsi che i dati raccolti non siano fortemente asimmetrici (ma anche in questo caso esistono alcune tecniche di trasformazione dei dati che li *forzano* alla simmetria). Per quanto riguarda l'omoschedasticità esistono test per l'omogeneità delle varianze fra i quali uno noto come test di Bartlett. Non è escluso l'utilizzo del test $F_{\max}$ presentato in § 6. 2. 2., per quanto tale metodo rientri nel novero di quelli che gli statistici qualificano come *quick and dirty* (veloci ma sporchi).

In sintesi possiamo dire che prima di eseguire l'ANOVA occorre sincerarsi che le assunzioni siano rispettate in misura adeguata. Tab. 7.9 riassume i controlli che è bene effettuare:

Assunzione

Assunzione	Tecnica di controllo
Casualità e indipendenza	Controllo della tecnica di campionamento
Normalità	Test per l'asimmetria dei dati
Omoschedasticità	Test di Bartlett o $F_{\max}$

Tab. 7.9

7. 5. 2. Test approssimato di Games e Howell

Quando vi è un difetto di omoschedasticità esiste una specifica procedura di calcolo nota come test approssimato di Games e Howell che permette di ottenere confronti appaiati non pianificati fra a gruppi di n_i elementi, analogamente e quanto abbiamo visto in § 7. 1. 1.

7. 5. 3. Alternative non parametriche all'ANOVA

Tutti i test presentati fino ad ora sono detti *parametrici*, in quanto l'ipotesi nulla H_0 riguarda direttamente i vari parametri di una popolazione (fino ad ora essenzialmente μ e σ). I test parametrici sono in genere basati su una serie di assunzioni; ad esempio in § 7. 5. 1. abbiamo richiamato quali siano per l'ANOVA e come effettuare i necessari controlli. In particolare tutti i test presentati fino a questo punto ipotizzano almeno che la distribuzione soggiacente sia normale. Questo vincolo il più delle volte non è eccessivamente restrittivo ed i test basati sulla statistica t ed F sono abbastanza robusti.

Vi sono tuttavia delle circostanze in cui la violazione delle assunzioni è tale da compromettere drasticamente l'attendibilità del test. In queste circostanze sono utilizzabili alcuni metodi alternativi detti *non parametrici* (in quanto l'ipotesi nulla non riguarda direttamente i parametri della popolazione ma solo la distribuzione delle variate Y) o *liberi da distribuzione* (in quanto non è richiesto che la popolazione soggiacente abbia una data distribuzione, come ad esempio quella normale).

Questi metodi sono in genere molto semplici e quindi sono spesso utilizzati. Tuttavia occorre precisare che nel caso le assunzioni dei test parametrici siano soddisfatte, anche solo parzialmente, questi sono più potenti dei corrispondenti non parametrici.

La precisazione necessita di una spiegazione almeno intuitiva.

Il fatto è che quasi tutti i test non parametrici sono adatti per dati su scala ordinale (vedi § 1. 1. 2.), mentre i test parametrici trattano dati quantitativi su scala almeno ad intervalli (vedi § 1. 1. 3.) se non a rapporti (vedi § 1. 1. 4.). Se abbiamo dati su scala ad intervalli (come ad esempio i punteggi grezzi di una prova oggettiva) e vogliamo trattarli con un test non parametrico, dobbiamo sostituire ad ogni dato il suo numero d'ordine in una ipotetica graduatoria che li comprenda tutti (per riportarli ad una scala ordinale). Il numero d'ordine si chiama *rango*. Ebbene, sostituire un dato con il suo rango comporta una *degradazione dell'informazione* contenuta nel dato stesso. Il test non parametrico dunque lavora su dati che hanno subito in qualche misura un degrado; la sua potenza è dunque necessariamente inferiore a quella del corrispondente test non parametrico che tratta il dato integro.

Dunque, i test non parametrici costituiscono un'alternativa all'uso dei test parametrici nel caso di dati su scala almeno ad intervalli che tuttavia non ne soddisfino le assunzioni; è altresì scontato, ma è opportuno dirlo esplicitamente, che i test non parametrici sono *l'unica possibilità* nel caso di dati che già in partenza sono su scala ordinale (come ad esempio i voti scolastici).

Tab. 7.10 elenca varie alternative non parametriche ai test parametrici specificati.

Test parametrico	Test non parametrico corrispondente
ANOVA a singolo criterio di classificazione e t -test per gruppi indipendenti	Test di Kruskal–Wallis o U -test di Mann–Witney
Procedura SS - STP per confronti multipli	Procedura STP non parametrica per confronti multipli
t -test per dati appaiati	Test di Wilcoxon per ranghi con segno
ANOVA a due vie senza replicazione e t -test per dati appaiati	Test di Friedman

Tab. 7.10