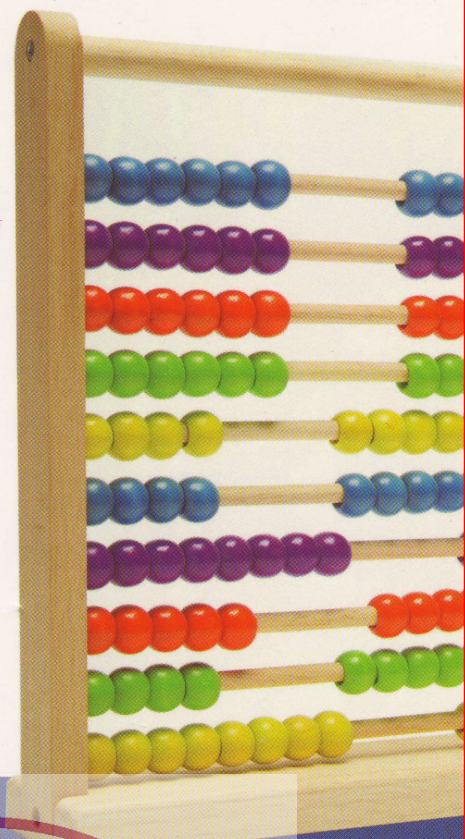


Analisi statistica con Excel



Michael Middleton

SCOPRIRE

le funzionalità statistiche
del foglio elettronico Excel

REALIZZARE

analisi dei dati, grafici,
test statistici e simulazioni



Fondamenti e metodi per l'analisi empirica nelle scienze sociali
Corso di laurea in Teorie, culture e tecniche per il Servizio sociale
Università degli Studi di Macerata

Nel booksite abbinato a questo libro
www.apogeeonline.com/libri/88-503-2475-8/scheda
i file Excel utilizzati negli esempi

APOGEO

Analisi della varianza

Questo capitolo descrive tre strumenti per l'analisi della varianza e uno per verificare l'uguaglianza delle varianze di due popolazioni. Tutti questi metodi utilizzano la distribuzione F per la verifica delle ipotesi.

13.1 Analisi della varianza a un fattore

L'analisi della varianza a un fattore può essere utilizzata per verificare l'ipotesi nulla che le medie di due o più popolazioni siano uguali. La verifica presuppone che i valori della popolazione abbiano una distribuzione normale, le varianze siano uguali e i campioni casuali siano indipendenti. Questo tipo di test è appropriato per i disegni completamente randomizzati.

Esempio 13.1 (Adattato da Cryer, pag. 530) Una casa automobilistica impiega alcuni rappresentanti di vendita per contattare telefonicamente i concessionari. L'obiettivo è confrontare l'efficacia di quattro differenti piani di telefonate. Vengono scelti a caso trentadue rappresentanti dall'ufficio vendite e vengono assegnati a caso ai quattro piani di telefonate, otto per piano. I rappresentanti seguono i loro piani per sei mesi e in questo periodo le loro vendite vengono registrate. La Figura 13.1 mostra i dati. Ci chiediamo se i dati avvalorano l'ipotesi che determinati piani di telefonate favoriscano il raggiungimento di un livello medio di vendite superiore nel lungo periodo.

	A	B	C	D
1	Piano A	Piano B	Piano C	Piano D
2	36	39	44	31
3	40	45	43	43
4	32	54	38	46
5	44	53	40	43
6	35	46	41	36
7	41	42	35	49
8	44	35	37	46
9	42	39	37	48

Figura 13.1 Dati per l'analisi della varianza a un fattore.

M. Middleton, Analisi statistica con Excel, Milano, Apogeo, 2004-2006

Teoricamente, i livelli di vendita nel lungo periodo per i piani di telefonate sono i valori della popolazione e l'ipotesi nulla è che le quattro popolazioni abbiano la stessa media. L'ipotesi alternativa è che almeno una media abbia un valore diverso. I dati nella Figura 13.1 sono campioni casuali di otto valori per ogni popolazione. I seguenti passaggi descrivono come realizzare un grafico con i valori delle vendite e le medie dei quattro piani di telefonate. Successivamente verrà descritta la procedura per l'analisi della varianza.

1. Digitate in un foglio di lavoro le etichette e i dati illustrati nella Figura 13.1.
2. Selezionate la cella A10 e digitate la formula **=MEDIA(A2:A9)**. Il risultato è 39,25. Selezionate la cella A10; fate clic sulla maniglia di riempimento (posta nell'angolo in basso a destra della cella) e trascinatela fino alla cella D10. I risultati sono illustrati nella Figura 13.2 nelle colonne A:D.
3. Digitate le etichette **Piano** nella cella F1 e **Vendite** nella cella G1. Selezionate F1:G1 e fate clic sul pulsante **Allinea a destra**.

	A	B	C	D	E	F	G
1	Piano A	Piano B	Piano C	Piano D		Piano	Vendite
2	36	39	44	31		1	36
3	40	45	43	43		1	40
4	32	54	38	46		1	32
5	44	53	40	43		1	44
6	35	46	41	36		1	35
7	41	42	35	49		1	41
8	44	35	37	46		1	44
9	42	39	37	48		1	42
10	39,25	44,125	39,375	42,75		2	39
11						2	45
12						2	54
13						2	53
14						2	46
15						2	42
16						2	35
17						2	39
18						3	44
19						3	43
20						3	38
21						3	40
22						3	41
23						3	35
24						3	37
25						3	37
26						4	31
27						4	43
28						4	46
29						4	43
30						4	36
31						4	49
32						4	46
33						4	48

Figura 13.2 Disposizione dei dati per preparare il grafico.

M. Middleton, Analisi statistica con Excel, Milano, Apogeo, 2004-2006

4. Digitate il valore 1 nella cella F2; fate clic sulla maniglia di riempimento e trascinatela fino alla cella F9.
5. Analogamente al passaggio precedente, digitate il valore 2 nella cella F10 e copiatelo nelle celle F11:F17. Digitate il valore 3 nella cella F18 e copiatelo nelle celle F19:F25. Infine, digitate il valore 4 nella cella F26 e copiatelo nelle celle F27:F33.
6. Selezionate i dati del Piano A nelle celle A2:A9. Fate clic con il pulsante destro del mouse e scegliete **Copia** dal menu di scelta rapida. Selezionate la cella G2. Fate clic con il pulsante destro del mouse e scegliete **Incolla** dal menu di scelta rapida.
7. Ripetete la procedura del passaggio precedente tre volte per copiare i dati delle celle B2:B9, C2:C9 e D2:D9, rispettivamente, nelle celle G10, G18 e G26. La Figura 13.2 mostra il risultato.
8. Selezionate le celle F1:G33 e fate clic sul pulsante **Creazione guidata Grafico**.
9. Nel primo passaggio della **Creazione guidata Grafico** scegliete il tipo di grafico **Dispers. (XY)** nella scheda **Tipi standard**. Fate clic sul pulsante **Avanti**.
10. Nel secondo passaggio (Dati di origine del grafico), fate clic sul pulsante **Aggiungi** nella scheda **Serie**. Fate clic nella casella **Nome** e digitate **Media**. Premete due volte il tasto **Tab** in modo che venga selezionato il contenuto della casella **Valori Y**. Selezionate le celle A10:D10 dal foglio di lavoro. Fate clic su **Avanti** (se i valori di X non vengono specificati per una serie di dati XY, Excel tratta i valori di X come numeri interi positivi).
11. Nel passaggio 3 (Opzioni del grafico), digitate **Esperimento dei piani di telefonate** nella casella **Titolo del grafico** della scheda **Titoli**; digitate **Trattamento** nella casella **Asse dei valori (X)** e **Vendite** nella casella **Asse dei valori (Y)**. Passate alla scheda **Griglia** e disattivate tutte le caselle di selezione. Fate clic su **Fine**.
12. Trascinate le maniglie di dimensionamento del grafico in modo che il grafico sia approssimativamente largo 8 colonne e alto 18 righe. Riducete la dimensione dei caratteri a 10 punti per il titolo del grafico, i titoli degli assi, gli assi dei valori e la legenda.
13. Fate doppio clic sull'asse verticale. Nella scheda **Scala** della finestra di dialogo **Formato asse** disattivate le caselle di selezione **Minimo** e **Massimo** e inserite rispettivamente i valori 25 e 60; fate clic su **OK**.
14. Fate doppio clic sull'asse orizzontale. Nella scheda **Scala** della finestra di dialogo **Formato asse** disattivate le caselle di selezione **Minimo** e **Massimo** e inserite rispettivamente i valori 0 e 5. Nella scheda **Motivo**, fate clic su **Assente** per il **Segno di graduazione principale** e per il **Segno di graduazione secondario** e su **Assenti** per le **Etichette di graduazione**. Fate clic su **OK**.
15. Per lasciare spazio alle etichette sull'asse orizzontale, selezionate il **Titolo dell'asse dei valori (X)** (**Trattamento**), fate clic sull'estremità sinistra del testo e premete **Invio**.

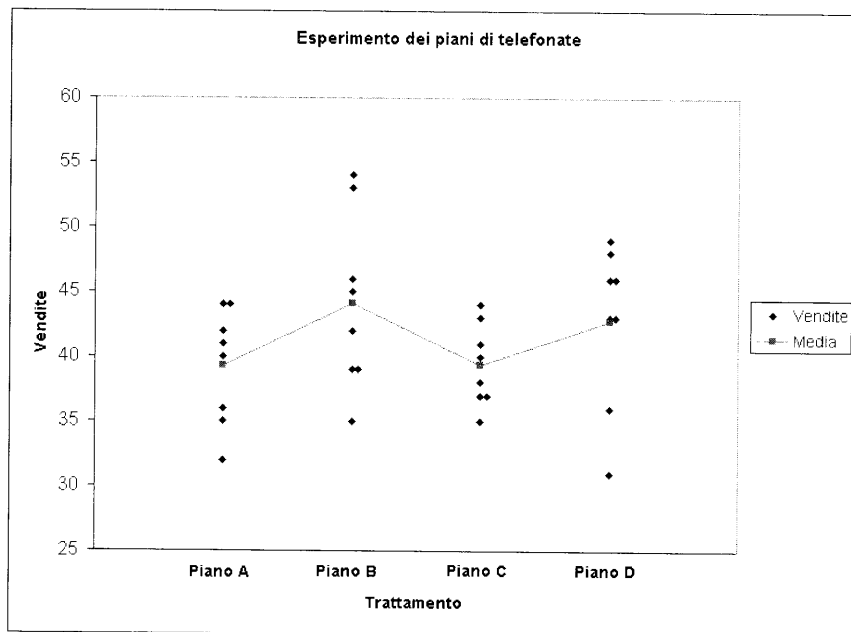


Figura 13.3 Grafico delle vendite e delle medie.

16. Fate clic all'interno del grafico (l'area del grafico), in modo che non vi siano elementi selezionati. Fate clic nella barra della formula e digitate **Piano A**. Dopo avere premuto *Invio*, sul grafico apparirà il testo. Spostate il puntatore del mouse vicino al bordo del testo finché il puntatore non assumerà la forma di quattro frecce; fate clic e trascinate il testo in corrispondenza della serie di dati più a sinistra. Facoltativamente, utilizzate la barra di formattazione per visualizzare il testo in grassetto e ridurre il corpo a 10 punti.
17. Ripetete le operazioni del passaggio precedente per i piani B, C e D.
18. Fate doppio clic su uno dei punti **Media** nel grafico. Nella scheda **Motivo** della finestra di dialogo **Formato serie dati**, selezionate **Automatica** nel riquadro **Linea** e fate clic su **OK**.
19. Nel Piano A il valore 44 compare due volte. Per rendere visibile la frequenza multipla di un valore, sostituite 1,05 a 1 nella cella F8. Analogamente, aggiungete 0,05 ai numeri delle celle F17, F25, F29 e F32. La Figura 13.3 mostra il risultato.

Per avviare l'analisi della varianza, inserite i dati in un foglio di lavoro come illustra la Figura 13.1 (oppure copiate i dati precedentemente inseriti in un nuovo foglio). Scegliete **Analisi dati** dal menu **Strumenti**. Nella finestra di dialogo **Analisi dati** fate doppio clic su **Analisi varianza: ad un fattore** nell'elenco degli **Strumenti di analisi**. Riempite i campi della finestra di dialogo come mostra la Figura 13.4; fate clic su **OK**.

Il risultato è illustrato nella Figura 13.5 nelle celle F1:L16. Formattate i decimali nelle celle I5:J8, G13:G16 e I13:L14 come in figura e inserite una nota nella cella L11 per segnalare che il valore di Alfa immesso nella finestra di dialogo è 0,05.

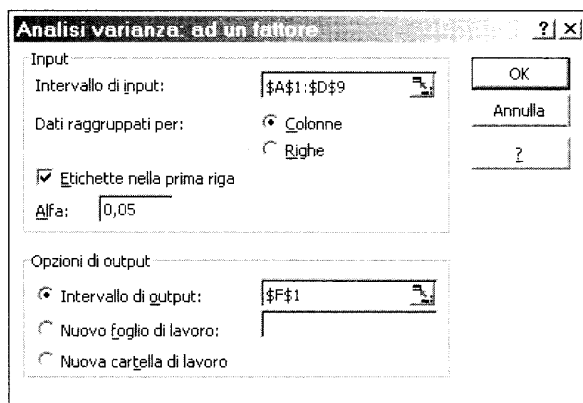


Figura 13.4 La finestra di dialogo Analisi varianza: ad un fattore.

La tabella ANALISI VARIANZA nella Figura 13.5 è una sintesi dei calcoli per la verifica delle ipotesi. La variazione totale dei 32 valori è calcolata utilizzando la somma dei quadrati (SQ) ed è scomposta in variazione tra gruppi e nei gruppi. Le medie dei quadrati (MQ) sono due stime della varianza della popolazione, calcolate dividendo SQ per il corrispondente grado di libertà (gdl). La statistica F nella cella J13 è il rapporto delle due stime della varianza.

Se l'ipotesi nulla è vera (le medie delle quattro popolazioni coincidono), le due stime della varianza dovrebbero essere approssimativamente uguali e il rapporto F prossimo a 1. Se una o più medie sono significativamente diverse dalle altre, la variazione fra gruppi sarà grande e il rapporto F sarà significativamente maggiore di 1. L'analisi della varianza è un test a una coda.

Approccio di reporting: il livello di significatività a una coda è 0,194 e indica che è abbastanza facile ottenere questi risultati quando le medie della popolazione sono uguali.

	F	G	H	I	J	K	L
1	Analisi varianza: ad un fattore						
2							
3	RIEPILOGO						
4	Gruppi	Conteggio	Somma	Media	Varianza		
5	Piano A	8	314	39,250	19,643		
6	Piano B	8	353	44,125	45,839		
7	Piano C	8	315	39,375	9,982		
8	Piano D	8	342	42,750	38,786		
9							
10							
11	ANALISI VARIANZA						Alfa=0,05
12	Origine della variazione	SQ	gdl	MQ	F	Valore di significatività	F crit
13	Tra gruppi	143,75	3	47,917	1,678	0,194	2,947
14	In gruppi	799,75	28	28,563			
15							
16	Totale	943,5	31				

Figura 13.5 Risultato dell'analisi della varianza a un fattore.

Approccio decisionale: utilizzando un livello di significatività del 5%, il valore 1,678 della statistica F non supera il valore critico F (2,947). Di conseguenza, non possiamo rifiutare l'ipotesi nulla di uguaglianza tra le medie della popolazione. In conclusione, non sembra che i diversi piani di telefonate abbiano un effetto significativo sulle vendite.

13.2 Analisi della varianza a due fattori senza replica

Nel Capitolo 11 abbiamo esaminato i test per le medie di due popolazioni con campioni indipendenti, dove possono esservi variazioni significative all'interno di ogni campione, e un test per le medie con campioni accoppiati, in cui l'accoppiamento o il raggruppamento consentono di effettuare test più efficaci. Uno schema analogo può essere utilizzato quando ci sono tre o più popolazioni o trattamenti da confrontare. L'idea generale consiste nell'usare un'altra caratteristica per classificare le osservazioni in gruppi o blocchi che avranno una variazione relativamente piccola prima di essere assegnati casualmente ai trattamenti. Questo approccio è chiamato *disegno a blocchi randomizzati*.

Esempio 13.2 (Adattato da Cryer, pag. 544) L'esperimento dei piani di telefonate è stato modificato utilizzando i vecchi dati di vendite (non riportati qui) per calcolare il livello medio mensile delle vendite di ogni rappresentante. I quattro rappresentanti con il livello medio di vendita più basso vengono posti nel primo blocco; i quattro rappresentanti con il livello medio immediatamente superiore vengono posti nel secondo blocco e così via. In ogni blocco i quattro rappresentanti sono assegnati casualmente ai quattro piani, uno per piano. Le vendite per una nuova analisi di un semestre sono riportate nella Figura 13.6. I dati avvalorano l'ipotesi che uno dei piani di telefonate faciliti il raggiungimento di un livello medio di vendite più alto nel lungo periodo?

Per effettuare l'analisi della varianza, inserite i dati in un foglio di lavoro come indica la Figura 13.6. Selezionate **Analisi dati** dal menu **Strumenti**. Una volta aperta la finestra di dialogo **Analisi dati**, fate doppio clic su **Analisi varianza: a due fattori senza replica** nella lista degli **Strumenti di analisi**. Inserite i riferimenti delle celle che contengono i dati da analizzare nei campi della finestra di dialogo come mostra la Figura 13.7; fate clic su **OK**.

	A	B	C	D	E
1		Piano A	Piano B	Piano C	Piano D
2	Blocco 1	27	31	24	27
3	Blocco 2	38	40	32	37
4	Blocco 3	35	39	34	39
5	Blocco 4	39	41	34	36
6	Blocco 5	36	43	37	39
7	Blocco 6	38	42	36	42
8	Blocco 7	41	46	42	42
9	Blocco 8	44	50	42	45

Figura 13.6 Dati per l'analisi della varianza a due fattori senza replica.

M. Middleton, Analisi statistica con Excel, Milano, Apogeo, 2004-2006

Analisi varianza: a due fattori senza replica [?] [X]

Input

Intervallo di input: [F12]

☒ Etichette

Alfa:

OK

Annulla

?

Opzioni di output

☒ Intervallo di output: [F12]

☐ Nuovo foglio di lavoro:

☐ Nuova cartella di lavoro

Figura 13.7 La finestra di dialogo per l'analisi della varianza a due fattori senza replica.

Il risultato è riportato nelle celle G1:M25 (Figura 13.8). Formattate i decimali nelle celle J4:K16, H21:H25, J21:K23 e M21:M22 come indica la Figura 13.8 e inserite una nota nella cella M19 per segnalare che il valore di Alfa immesso nella finestra di dialogo è 0,05.

La tabella ANALISI VARIANZA nella Figura 13.8 è una sintesi dei calcoli per la verifica delle ipotesi. La variazione totale dei 32 valori è attribuita alla variazione dovuta al fattore Righe (blocchi), alla variazione dovuta al fattore Colonne (trattamenti o piani di telefonate) e all'Errore (variazione che non può essere attribuita ai trattamenti o ai blocchi). L'obiettivo del test è trovare le differenze dovute ai piani delle telefonate, indicati dal fattore Colonne, quindi nella cella K22 abbiamo posto il valore (26,460) della statistica F che sintetizza il test.

	F	G	H	I	J	K	L	M
1		Analisi varianza: a due fattori senza replica						
2								
3		RIEPILOGO	Conteggio	Somma	Media	Varianza		
4		Blocco 1	4	109	27,250	8,250		
5		Blocco 2	4	147	36,750	11,583		
6		Blocco 3	4	147	36,750	6,917		
7		Blocco 4	4	150	37,500	9,667		
8		Blocco 5	4	155	38,750	9,583		
9		Blocco 6	4	158	39,500	9,000		
10		Blocco 7	4	171	42,750	4,917		
11		Blocco 8	4	181	45,250	11,583		
12								
13		Piano A	8	298	37,250	25,071		
14		Piano B	8	332	41,500	30,571		
15		Piano C	8	281	35,125	33,554		
16		Piano D	8	307	38,375	29,696		
17								
18								
19		ANALISI VARIANZA						Alfa=0,05
20		<i>Origine della variazione</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Valore di significatività</i>	<i>F crit</i>
21		Righe	787,375	7	112,482	52,638	6,7E-12	2,488
22		Colonne	169,625	3	56,542	26,460	2,5E-07	3,072
23		Errore	44,875	21	2,137			
24								
25		Totale	1001,875	31				

Figura 13.8 Risultato dell'analisi della varianza a due fattori senza replica.

Se l'ipotesi nulla di medie uguali per i piani delle telefonate è vera, il rapporto F dovrebbe essere prossimo a 1. Se una o più medie della popolazione sono significativamente diverse dalle altre, il rapporto F sarà significativamente maggiore di 1. Approccio di reporting: il livello di significatività a una coda è $2,5E-07$ o $0,00000025$ e indica che è estremamente difficile ottenere questi risultati quando le medie della popolazione sono uguali.

Approccio decisionale: utilizzando un livello di significatività del 5%, il valore 26,460 della statistica F eccede il valore critico della coda destra (3,072). Di conseguenza, possiamo rifiutare l'ipotesi nulla di uguaglianza delle medie della popolazione. In conclusione, la vendita media di uno o più piani di telefonate è significativamente diversa dalle altre.

13.3 Analisi della varianza a due fattori con replica

Come precedentemente descritto, l'analisi della varianza a un fattore è una verifica di ipotesi per l'uguaglianza delle medie dei trattamenti (popolazione); l'analisi della varianza a due fattori senza replica è anch'essa una verifica dell'uguaglianza delle medie dei trattamenti, che si serve della suddivisione in blocchi per tenere in considerazione variazioni non legate ai trattamenti. Questo paragrafo descrive l'analisi della varianza a due fattori con replica, che può essere utilizzata per valutare gli effetti di due trattamenti. Questa analisi è appropriata a un esperimento fattoriale, qui limitato a due fattori, in un disegno completamente randomizzato.

Esempio 13.3 (Adattato da Cryer, pag. 553) Un esperimento studia l'impatto di due fattori (tempo di cottura e temperatura) sul gusto di una torta. La variabile risposta è il gusto della torta ed è misurata dai voti dati da esperti assaggiatori. Il fattore tempo di cottura ha tre livelli: Breve (10% meno del tempo indicato nella confezione), Medio (il tempo indicato nella confezione) e Lungo (10% più del tempo indicato). Il fattore temperatura ha anch'esso tre livelli: Bassa (10% meno della temperatura indicata nella confezione), Media (la temperatura indicata) e Alta (10% più della temperatura indicata). Pertanto, ci sono nove possibili combinazioni dei tempi di cottura e delle temperature. Tre assaggiatori valutano le torte cotte a una determinata combinazione di tempo e temperatura; ci sono quindi tre repliche. Le torte cotte sono assegnate casualmente agli assaggiatori, che danno un voto in base a una scala a 7 punti, da 0 (molto al di sotto della media) a 6 (molto al di sopra della media). I voti espressi sono riportati nella Figura 13.9.

	A	B	C	D
1		Temp. Bassa	Temp. Media	Temp. Alta
2	Tempo Breve	0	0	4
3		0	2	5
4		3	4	6
5	Tempo Medio	2	3	1
6		3	6	2
7		4	6	3
8	Tempo Lungo	4	1	0
9		5	3	1
10		6	5	2

Figura 13.9 Dati per l'analisi della varianza a due fattori con replica.

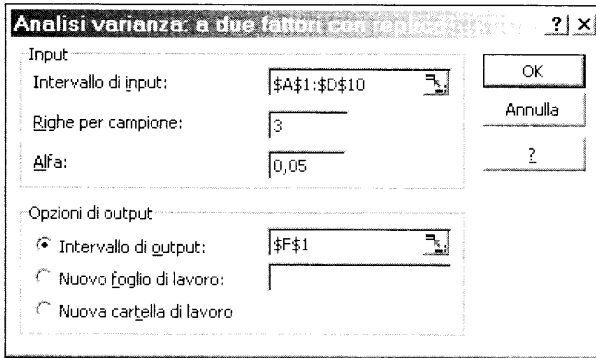
M. Middleton, Analisi statistica con Excel, Milano, Apogeo, 2004-2006

Figura 13.10 La finestra di dialogo per l'analisi della varianza a due fattori con replica.

Per effettuare l'analisi della varianza, inserite i dati in un foglio di lavoro come indica la Figura 13.9. Selezionate **Analisi dati** dal menu **Strumenti**. Una volta aperta la finestra di dialogo **Analisi dati**, fate doppio clic su **Analisi varianza: a due fattori con replica** nella lista degli **Strumenti di analisi**. Inserite i riferimenti delle celle che contengono i dati da analizzare nei campi della finestra di dialogo come mostra la Figura 13.10; fate clic su **OK**.

Il risultato è riportato nelle celle F1:L36 (Figura 13.11). Inserite una nota nella cella L29 per segnalare che il valore di Alfa immesso nella finestra di dialogo è 0,05. La tabella ANALISI VARIANZA nella Figura 13.11 è una sintesi dei calcoli per la verifica delle ipotesi. La variazione totale dei 27 valori è attribuita alla variazione Campione (il fattore Riga, tempo di cottura), alla variazione del fattore Colonne (temperatura), alla variazione dovuta all'Interazione (particolari combinazioni di tempo e temperatura) e alla variazione di questi tre fattori all'interno (In) di ogni combinazione.

Ci sono tre differenti test delle ipotesi che riguardano gli effetti sul gusto: il tempo di cottura (Campione), la temperatura (Colonne) e le combinazioni (Interazione). In ogni test, se l'ipotesi nulla è vera, il rapporto F dovrebbe essere prossimo a 1. Se un particolare effetto è diverso da zero, il corrispondente rapporto F sarà significativamente maggiore di 1.

Approccio di reporting: i valori di significatività a una coda sono approssimativamente 0,63, 0,63 e 0,001 e indicano che è estremamente facile ottenere i voti campione se gli effetti del tempo di cottura e della temperatura sono nulli e che è estremamente difficile ottenere questi risultati se gli effetti d'interazione sono nulli. Approccio decisionale: utilizzando un livello di significatività del 5%, i valori della statistica F per il tempo di cottura e la temperatura sono entrambi approssimativamente pari a 0,47, che è minore del valore critico della coda destra (3,55456). Pertanto non possiamo rifiutare l'ipotesi nulla di assenza di effetti per il tempo e la temperatura. Il valore della statistica F per l'interazione è approssimativamente 7,34 ed eccede il valore critico della coda destra (2,92775). Di conseguenza, possiamo rifiutare l'ipotesi nulla di assenza di effetti dovuti all'interazione.

M. Middleton, Analisi statistica con Excel, Milano, Apogeo, 2004-2006

E	F	G	H	I	J	K	L
1	Analisi varianza: a due fattori con replica						
2							
3	RIEPILOGO Temp. Bassa Temp. Media Temp. Alta Totale						
4	<i>Tempo Breve</i>						
5	Conteggio	3	3	3	9		
6	Somma	3	6	15	24		
7	Media	1	2	5	2,666667		
8	Varianza	3	4	1	5,25		
9							
10	<i>Tempo Medio</i>						
11	Conteggio	3	3	3	9		
12	Somma	9	15	6	30		
13	Media	3	5	2	3,333333		
14	Varianza	1	3	1	3		
15							
16	<i>Tempo Lungo</i>						
17	Conteggio	3	3	3	9		
18	Somma	15	9	3	27		
19	Media	5	3	1	3		
20	Varianza	1	4	1	4,5		
21							
22	<i>Totale</i>						
23	Conteggio	9	9	9			
24	Somma	27	30	24			
25	Media	3	3,333333	2,666667			
26	Varianza	4,25	4,5	4			
27							
28							
29	ANALISI VARIANZA						Alfa=0,05
30	Origine della variazione	SQ	gdl	MQ	F	Valore di significatività	F crit
31	Campione	2	2	1	0,473684	0,630249	3,55456
32	Colonne	2	2	1	0,473684	0,630249	3,55456
33	Interazione	62	4	15,5	7,342105	0,001087	2,92775
34	In	38	18	2,111111			
35							
36	Totale	104	26				

Figura 13.11 Risultato dell'analisi della varianza a due fattori con replica.

Grafico delle interazioni

Un grafico delle interazioni è un utile aiuto visivo per identificare gli effetti delle interazioni. La seguente procedura descrive come costruire lo schema delle interazioni delle medie illustrato nella Figura 13.12.

1. Inserite le etichette nelle celle M1, M4:M6 e N2:P3 come mostrato nella Figura 13.12. Digitate **Livello della temperatura** nella cella N2; selezionate le celle N2:P2, scegliete **Formato | Celle | Allineamento | Orizzontale | Centra nelle colonne**; fate clic su **OK**. Inserite i valori degli indici delle medie nelle celle N4:P6 o copiate questi valori dal risultato dell'analisi della varianza precedente (celle G7:I7, G13:I13 e G19:I19).
2. Selezionate le celle M3:P6 e fate clic sul pulsante **Creazione guidata Grafico**.
3. Nel passaggio 1 (Tipo di grafico) della creazione guidata del grafico, scegliete **Linee** come tipo di grafico nella scheda **Tipi standard**. Fate clic su **Avanti**. Nel passaggio 2 (Dati di origine del grafico), fate clic su **Avanti**. Nel passaggio 3 (Opzioni del grafico), disattivate tutte le caselle di selezione nella scheda **Griglia**.

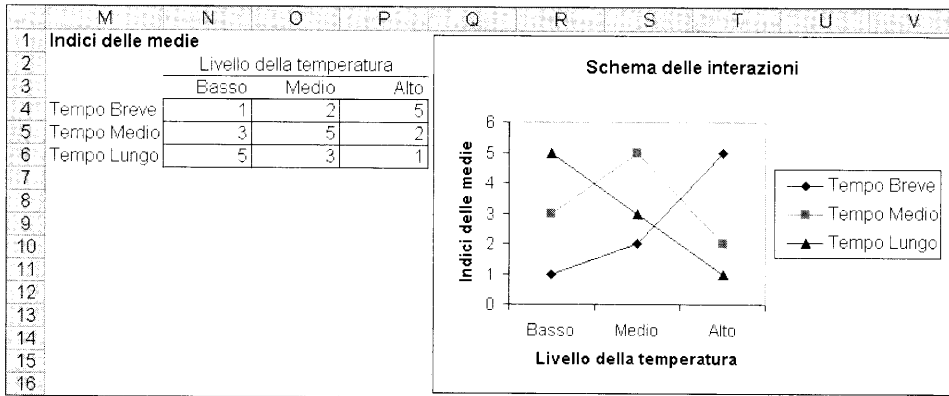


Figura 13.12 Grafico delle interazioni delle medie.

Nella scheda **Titoli**, inserite i titoli del grafico e degli assi come indica la Figura 13.12; fate clic su **Fine**. Il grafico risultante è illustrato nella Figura 13.12.

Nel grafico delle interazioni mostrato nella Figura 13.12 le linee si incrociano, indicando che sussiste un'interazione tra i due fattori tempo di cottura e temperatura. Se non vi fosse stato alcun effetto di interazione, le linee sarebbero state parallele.

Effetti Riga, Colonna e Interazione

L'indice della media per una particolare combinazione di tempo di cottura e temperatura può essere espresso come segue:

Media Cella = Media Totale + Effetto Riga + Effetto Colonna + Effetto Interazione

I seguenti passaggi descrivono come calcolare gli effetti Riga, Colonna e Interazione mostrati nella Figura 13.13.

1. Inserite le etichette nelle celle A1:A10 e B2:F3 in un nuovo foglio di lavoro. Digitate **Livello della temperatura** nella cella B2; selezionate le celle B2:D2, scegliete **Formato | Celle | Allineamento | Orizzontale | Centra nelle colonne**; fate clic su **OK**. Inserite i valori degli indici delle medie nelle celle B4:D6 o copiate questi valori dal risultato dell'analisi della varianza.
2. Digitate la formula **=MEDIA(B4:B6)** nella cella B7. Selezionate la cella B7, fate clic sulla maniglia di riempimento nell'angolo in basso a destra della cella e trascinatela fino alla cella D7.
3. Digitate la formula **=MEDIA(B4:D4)** nella cella E4. Selezionate la cella E4, fate clic sulla maniglia di riempimento e trascinatela fino alla cella E7.
4. Digitate la formula **=B7-\$E\$7** nella cella B8. Selezionate la cella B8, fate clic sulla maniglia di riempimento e trascinatela fino alla cella D8.
5. Digitate la formula **=E4-\$E\$7** nella cella F4. Selezionate la cella F4, fate clic sulla maniglia di riempimento e trascinatela fino alla cella F6.

	A	B	C	D	E	F
1	Indici delle medie					
2		Livello della temperatura				
3		Basso	Medio	Alto	Media Riga	Effetto Riga
4	Tempo Breve	1	2	5	2,67	-0,33
5	Tempo Medio	3	5	2	3,33	0,33
6	Tempo Lungo	5	3	1	3,00	0,00
7	Media Colonna	3,00	3,33	2,67	3,00	
8	Effetto Colonna	0,00	0,33	-0,33		
9						
10	Effetto Interazione					
11		Livello della temperatura				
12		Basso	Medio	Alto		
13	Tempo Breve	-1,67	-1,00	2,67		
14	Tempo Medio	-0,33	1,33	-1,00		
15	Tempo Lungo	2,00	-0,33	-1,67		

Figura 13.13 Effetti Riga, Colonna e Interazione.

- Formattate i risultati delle formule con due cifre decimali. I risultati sono riportati nelle prime otto righe della Figura 13.13.
- Selezionate le celle A2:D6 e fate clic sul pulsante **Copia**. Selezionate la cella A11 e fate clic sul pulsante **Incolla**.
- Digitate la formula $=B4-B\$8-\$F4-\$E\$7$ nella cella B13. Selezionate la cella B13, fate clic sulla maniglia di riempimento e trascinatela fino alla cella D13. Selezionate le celle B13:D13, fate clic sulla maniglia di riempimento e trascinatela fino alla cella D15. Il risultato è illustrato nella Figura 13.13.

La cella E7 nella Figura 13.13 indica che la media totale dei 27 indici di gusto è uguale a 3,00. Le medie delle righe e delle colonne sono prossime alla media totale; ciò significa che gli effetti riga e colonna stimati sono prossimi a zero. La verifica delle ipotesi dell'analisi della varianza ha confermato che questi due effetti non sono significativamente diversi da zero.

D'altro canto, facendo riferimento alla parte inferiore della Figura 13.13, molti tra gli effetti di interazione sono sostanzialmente diversi da zero, specialmente se confrontati con l'entità degli effetti Riga e Colonna. La verifica delle ipotesi dell'analisi della varianza ha dimostrato che gli effetti dell'interazione sono significativamente diversi da zero.

13.4 Test F a due campioni per varianze

La distribuzione F può essere utilizzata per la verifica dell'ipotesi nulla che due popolazioni abbiano varianze uguali. Questo test è utile per stabilire se un successivo test *t* per l'uguaglianza delle medie debba assumere varianze uguali o diverse. Il test suppone che campioni indipendenti e casuali siano estratti da popolazioni con distribuzioni normali; viene eseguito calcolando il rapporto fra le varianze dei due campioni. Se le popolazioni presentano la stessa varianza, ci si aspetta che il rapporto fra le varianze sia prossimo a uno.

	A	B	C	D	E	F	G	H
1		Schema A	Schema B		Test F a due campioni per varianze			
2		8,2	9,5					
3		5,3	8,3					
4		6,5	7,5					
5		5,1	10,9					
6		9,7	11,3					
7		10,8	9,3					
8			8,8					
9			8					
10								
11								
12								

	Schema A	Schema B
Media	7,6	9,2
Varianza	5,552	1,814
Osservazioni	6	8
gdl	5	7
F	3,060	
P(F<=f) una coda	0,089	
F critico una coda	3,972	

(Alfa inserito nella finestra di dialogo = 0,05)

Figura 13.14 Risultato di un test F a due campioni per varianze.

È più appropriato un test a due code nel caso in cui l'ipotesi alternativa è semplicemente che le varianze delle due popolazioni siano diverse. Se il rapporto fra le due varianze è molto diverso da 1 - troppo piccolo (prossimo a zero) o troppo grande - possiamo rifiutare l'ipotesi nulla. Se l'ipotesi alternativa è che una varianza sia significativamente più grande dell'altra, allora è preferibile utilizzare un test a una coda. Per ottenere come output il valore corretto di "F critico una coda", il valore di Alfa da immettere nella finestra di dialogo **Test F a due campioni per varianze** deve corrispondere alla probabilità in una singola coda della regione di rifiuto. Per esempio, per un test a una coda con un livello di significatività del 5%, digitate 0,05; per un test a due code con un livello di significatività del 5%, digitate 0,025.

La Figura 13.14 illustra i dati dell'Esempio 11.2, che corrispondono ai tempi di assemblaggio, espressi in minuti, di 14 operai scelti a caso per due diversi schemi di lavorazione (sequenza delle operazioni svolte).

Dopo aver immesso i dati nelle colonne B e C, scegliete **Analisi dati** dal menu **Strumenti**. Sullo schermo apparirà la finestra di dialogo **Analisi dati**; fate doppio clic su **Test F a due campioni per varianze** nella lista degli **Strumenti di analisi**. Inserite i riferimenti delle celle che contengono i dati da analizzare negli appositi campi della finestra di dialogo, come illustra la Figura 13.15; fate clic su **OK**. Il risultato è illustrato nella Figura 13.14 nelle celle E1:G10. Formattate le cifre decimali

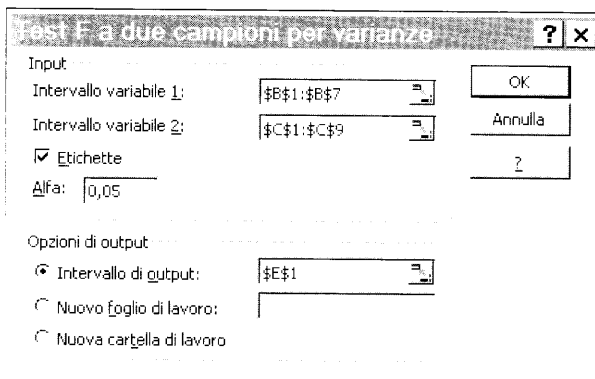


Figura 13.15 La finestra di dialogo per il test F a due campioni per varianze.

nelle celle F5:G5 e F8:F10 come indica la figura. Inserite una nota nella cella E12 per segnalare che il valore di Alfa a una coda immesso nella finestra di dialogo è 0,05. Il valore ottenuto per la statistica F ($5,552/1,814 = 3,060$) è riportato nella cella F8 della Figura 13.14. Il valore 0,089 nella cella F9 indica la probabilità della coda destra di ottenere un valore maggiore o uguale a quello della statistica F. L'etichetta della cella E9 è sbagliata e dovrebbe essere corretta con "P(F>=f) una coda", specialmente se i risultati dell'analisi devono essere presentati ad altre persone. Per una statistica F con gdl (grado di libertà) pari a 5 a numeratore e 7 a denominatore, $P(F \geq 3,060) = 0,089$. Questa probabilità può anche essere ottenuta usando la funzione DISTRIB.F(3,06;5;7), che ha la seguente sintassi:

DISTRIB.F(x;grad_libertà1;grad_libertà2)

Il valore 3,972 nella cella F10, etichettata "F critico una coda" nella cella E10, corrisponde alla probabilità Alfa 0,05 specificata nella finestra di dialogo. Questo valore di F può anche essere ottenuto usando la funzione INV.F(0,05;5;7), che ha la seguente sintassi:

INV.F(probabilità;grad_libertà1;grad_libertà2)

Approccio di reporting: il valore di significatività (*p*-value) a due code è 0,178 e indica che è estremamente facile ottenere queste varianze campionarie, quando le varianze delle popolazioni sono uguali.

Approccio decisionale: utilizzando un livello di significatività del 10% (5% per ogni coda), il valore calcolato (3,060) della statistica F non supera il valore critico 3,972 della coda destra. Di conseguenza, non possiamo rifiutare l'ipotesi nulla di uguaglianza delle varianze.

Esercizi

1. (Adattato da Cryer, pag. 540) Per verificare l'effetto dell'altezza degli scaffali sulle vendite delle torte, un'azienda dolciaria ha condotto un esperimento con la collaborazione di nove supermercati. Le posizioni delle torte (i trattamenti) negli scaffali sono tre: Bassa, Altezza occhi e Alta. A tre supermercati vengono casualmente assegnati i vari tipi di posizionamento delle torte; la seguente tabella riporta le unità vendute in un mese.

<i>Posizione negli scaffali</i>		
<i>Bassa</i>	<i>Altezza occhi</i>	<i>Alta</i>
582	781	524
558	604	499
525	584	521

- a. Create un grafico che mostra le unità vendute per le tre posizioni delle torte negli scaffali e le vendite medie. Il grafico suggerisce qualche effetto significativo sulle vendite?

M. Middleton, Analisi statistica con Excel, Milano, Apogeo, 2004-2006

b. Effettuate un'analisi della varianza e interpretatene il significato statistico. La vostra interpretazione è coerente con l'impressione visiva fornita dal grafico.

2. (Adattato da Cryer, pag. 572) Una salumeria utilizza uno schema a blocchi casuali per valutare l'efficienza di tre nuove macchine di confezionamento che sono state noleggiate da tre diversi fornitori. I blocchi sono costituiti da tre impiegati del reparto di confezionamento. Ogni impiegato è addestrato su una macchina; poi, gli impiegati vengono assegnati casualmente alle macchine (trattamenti) e viene loro chiesto di tenerle in funzione per un'ora. I dati della tabella sono i numeri delle confezioni difettose prodotte in un'ora. Notate che ogni impiegato usa le tre macchine. Effettuate un'analisi della varianza e valutatene il significato statistico.

<i>Operatore</i>	<i>Macchina</i>		
	<i>A</i>	<i>B</i>	<i>C</i>
1	9	8	7
2	10	12	3
3	11	10	9

3. Facendo riferimento all'Esercizio 1 del Capitolo 11, verificate l'ipotesi di uguaglianza delle varianze di due popolazioni, utilizzando un livello di significatività del 5%.
4. (Adattato da Mendenhall, pag. 464) I regolamenti federali richiedono un trattamento che ritardi l'azione delle fiamme per alcuni materiali, come quelli utilizzati nei pigiama da bambino. In due diversi laboratori viene valutata l'efficacia di questo trattamento applicato a tre tipi di materiali. Ogni laboratorio ha effettuato un test su tre campioni per ogni tipo di materiale. I risultati ottenuti indicano la lunghezza della parte di materiale carbonizzato per ogni campione e sono riportati nella tabella. Oltre a determinare se l'effetto del trattamento varia da materiale a materiale, vorremmo anche valutare se i risultati dei laboratori sono coerenti. Effettuate un'analisi della varianza per i seguenti dati, mirata a valutare le differenze significative dovute ai materiali, ai laboratori e alla loro interazione.

<i>Laboratorio</i>	<i>Materiale</i>		
	<i>A</i>	<i>B</i>	<i>C</i>
1	4,1	3,1	3,5
	3,9	2,8	3,2
	4,3	3,3	3,6
2	2,7	1,9	2,7
	3,1	2,2	2,3
	2,6	2,3	2,5