

# CONTRASTI E CONFRONTI MULTIPLI

Andrea Onofri

Dipartimento di Scienze Agrarie ed Ambientali

Università degli Studi di Perugia

Versione on-line: <http://www.unipg.it/~onofri/RTutorial/index.html>

## Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Contrasti ortogonali</b>	<b>3</b>
2.1	I contrasti con R . . . . .	5
<b>3</b>	<b>Test di confronto multiplo</b>	<b>6</b>
<b>4</b>	<b>Referenze bibliografiche</b>	<b>10</b>

## 1 Introduzione

Quando eseguiamo un'analisi di regressione e riusciamo ad individuare l'esistenza di una relazione di dipendenza della  $Y$  sulla  $X$ , non è più necessario (anzi, è sbagliato!) stabilire se vi sia una differenza significativa tra la risposta osservata per ogni livello della variabile  $X$ . Infatti, dato che le risposte giacciono su una retta (o su un'altra funzione della  $X$ ) ad ogni  $X$  corrisponde un diverso  $Y = f(X)$ . Si potrà poi discutere quanto sia importante (grande) questa differenza da un punto di vista biologico, ma essa esiste e vale per ogni livello di  $X$ , comunque scelto.

Quando invece la  $X$  è una variabile qualitativa non esiste necessariamente una relazione o un ordine tra i livelli del trattamento sperimentale. In questo caso, se l'ANOVA è significativa vuol dire che l'effetto del trattamento è maggiore dell'errore sperimentale, ma questo non implica una graduatoria di merito tra i gruppi. Chi è superiore a chi?

Ammettiamo di aver effettuato una prova con quattro trattamenti sperimentali (tesi di concimazione), con i risultati riportati nella tabella 1:

Il primo passaggio dell'analisi è, in genere, l'esecuzione dell'ANOVA, con lo scopo di misurare l'errore sperimentale. Anche se la significatività del test di  $F$  non è strettamente necessaria per l'esecuzione dei contrasti o dei confronti multipli, essa viene in genere considerata prodromica, per

Tabella 1: Esempio di una prova sperimentale di concimazione

Tesi	Produzione
Non concimato	12
Non concimato	15
Non concimato	13
<b>MEDIA</b>	<b>13.33</b>
Concime minerale	20
Concime minerale	21
Concime minerale	23
<b>MEDIA</b>	<b>21.33</b>
Concime minerale a lento rilascio	19
Concime minerale a lento rilascio	18
Concime minerale a lento rilascio	16
<b>MEDIA</b>	<b>17.67</b>
Concime organico	22
Concime organico	19
Concime organico	20
<b>MEDIA</b>	<b>20.33</b>

aumentare la protezione, soprattutto nel caso di procedure come la MDS, che, per definizione, sono abbastanza liberali. L'ANOVA può essere eseguita facilmente con R (si veda l'esempio sottostante e si noti l'uso della funzione `tapply()` per calcolare le medie).

```
> Tesi<-c(1,1,1,2,2,2,3,3,3,4,4,4)
> Tesi<-as.factor(Tesi)
> Prod<-c(12,15,13,20,21,23,19,18,16,22,19,20)
> tapply(Prod,Tesi,mean)
      1      2      3      4
13.33333 21.33333 17.66667 20.33333
> anova(lm(Prod~Tesi))
```

#### Analysis of Variance Table

Response: Prod

Df	Sum Sq	Mean Sq	F value	Pr(>F)
Tesi	3	115.000	38.333	16.429 0.0008821 ***
Residuals	8	18.667	2.333	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

In questo caso, l'analisi della varianza ed il relativo test di F ci dicono che esiste una differenza significativa tra le medie, ma rimane il problema di classificare le soluzioni concimanti in ordine di efficacia. Questo problema può essere risolto in vari modi, come vedremo in seguito.

## 2 Contrasti ortogonali

Senza entrare troppo nel dettaglio numerico (per questo esistono numerosi testi di riferimento) diciamo che si definisce CONTRASTO una combinazione lineare delle medie dei trattamenti, in modo che i coefficienti sommati diano 0. Ad esempio, con i dati nella tabella precedente, una combinazione lineare del tipo:

$$13.33 - (21.33 + 17.67 + 20.33)$$

non è un contrasto perchè è un confronto tra la media del gruppo A e la somma delle medie dei gruppi B, C e D. In effetti, la somma dei coefficienti della combinazione è:

$$1 - (1 + 1 + 1) = -3$$

La seguente combinazione lineare:

$$13.33 - \left( \frac{21.33}{3} + \frac{17.67}{3} + \frac{20.33}{3} \right)$$

è invece un contrasto, in quanto i coefficienti:

$$1 - (1/3 + 1/3 + 1/3) = 0$$

Si può osservare che in questo caso ho confrontato la media del gruppo A, con la media delle medie dei gruppi B, C e D.

Un contrasto ha un grado di libertà e l'effetto, tal quale (E) e centrato (C) è pari a:

$$E = \begin{pmatrix} 13.33 \\ 13.33 \\ 13.33 \\ 13.33 \\ 19.78 \\ 19.78 \\ 19.78 \\ 19.78 \\ 19.78 \\ 19.78 \\ 19.78 \\ 19.78 \\ 19.78 \end{pmatrix} \quad C = \begin{pmatrix} -4.837 \\ -4.837 \\ -4.837 \\ -4.837 \\ 1.613 \\ 1.613 \\ 1.613 \\ 1.613 \\ 1.613 \\ 1.613 \\ 1.613 \\ 1.613 \\ 1.613 \end{pmatrix}$$

la devianza del contrasto è pari al quadrato della norma del vettore  $C$ , cioè 93.61. La varianza corrispondente può essere confrontata con la varianza dell'errore tramite test  $F$ , per verificare la significatività del contrasto.

I contrasti possono essere ortogonali se la somma dei prodotti dei coefficienti è pari a zero. Ad esempio, i due contrasti:

$$13.33 - \left( \frac{21.33}{3} + \frac{17.67}{3} + \frac{20.33}{3} \right)$$

e

$$21.33 - \left( \frac{17.67}{2} + \frac{20.33}{2} \right)$$

sono ortogonali, perchè la somma dei prodotti dei coefficienti è:

$$1 \times 0 + \frac{1}{3} \times 1 - \frac{1}{3} \times \frac{1}{2} - \frac{1}{3} \times \frac{1}{2} = 0$$

Più contrasti si dicono mutualmente ortogonali se tutte le coppie di contrasti sono ortogonali. In un disegno sperimentali con  $p$  tesi esistono solo  $p-1$  contrasti ortogonali, con l'importante caratteristica che le loro devianze individuali (una per ogni contrasto) sommate ricostruiscono esattamente la devianza del trattamento.

L'uso dei contrasti ortogonali è il modo più potente ed elegante per confrontare tra di loro diversi trattamenti sperimentali, anche se richiede un'apposita pianificazione dell'esperimento.

Nel caso in esempio, si potrebbero ipotizzare i seguenti contrasti:

1. non concimato vs concimato
2. concime organico vs. minerali

## 3. minerale tradizionale vs lento rilascio.

La matrice dei coefficienti del contrasto, scritta per comodità con i contrasti sulle colonne e codificata in modo da evitare i decimali, é:

$$C = \begin{pmatrix} 3 & 0 & 0 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \\ -1 & 2 & 0 \end{pmatrix}$$

## 2.1 I contrasti con R

L'esecuzione dei contrasti in R è abbastanza semplice: dopo aver definito la matrice dei contrasti, essa viene assegnata al vettore della variabile esplicativa, tramite la funzione `contrasts()`. Dopo ciò si esegue l'analisi della varianza usando la funzione `aov`, invece che `lm` (stessa sintassi) e si utilizza la funzione `summary.aov` con l'argomento `split` per suddividere la devianza tra i diversi contrasti. In questa procedura si deve far attenzione al fatto che la matrice dei contrasti (che riporta nelle colonne i coefficienti di ogni contrasto ed ha quindi tante righe quante sono le tesi e tante colonne quanti sono i contrasti) viene associata al vettore delle tesi e in modo che ad ogni riga viene a corrispondere una tesi. **Nel fare questa corrispondenza R riorganizza le tesi in ordine alfabetico**, quindi alla n-sima riga viene fatto corrispondere l' n-esimo trattamento in ordine alfabetico, senza considerare la posizione che quel trattamento occupa nel vettore delle tesi. Ciò può dare grossi errori se le tesi non sono ordinate alfabeticamente.

Per quanto riguarda l'astrusa sintassi del comando `split`, converrà ricopiarla così come è, considerando che con essa vengono elencati i contrasti con il loro numero d'ordine (in questo caso da 1 a 3) definendone nel contempo il nome (opzionale).

```
> Tesi<-c(1,1,1,2,2,2,3,3,3,4,4,4)
> Tesi<-as.factor(Tesi)
> Prod<-c(12,15,13,20,21,23,19,18,16,22,19,20)

#Definizione della matrice dei contrasti (uno per ogni colonna)
> con<-matrix(c(3,-1,-1,-1,0,-1,-1,2,0,1,-1,0),4,3)
> con
      [,1] [,2] [,3]
[1,]    3    0    0
[2,]   -1   -1    1
[3,]   -1   -1   -1
[4,]   -1    2    0
>
```

```

#Si associa la matrice dei contrasti alla colonna delle tesi, per
#mezzo della funzione contrasts. In pratica ogni tesi è associata
#ad una riga della matrice dei contrasti
# ATTENZIONE!!! Le tesi sono riorganizzate in ordine alfabetico!!!
> con->contrasts(Tesi)

# Si esegue l'ANOVA con la funzione aov
> mod<-aov(Prod~Tesi)

> summary.aov(mod,split=list(Tesi=list("non concimato vs concimato "=1,
                                         "minerale vs organico "=2,
                                         "normale vs lento rilascio "=3)))

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Tesi	3	115.000	38.333	16.4286	0.0008821 ***
Tesi: non concimato vs concimato	1	93.444	93.444	40.0476	0.0002258 ***
Tesi: minerale vs organico	1	1.389	1.389	0.5952	0.4625785
Tesi: normale vs lento rilascio	1	20.167	20.167	8.6429	0.0187133 *
Residuals	8	18.667	2.333		

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

Il test mostra che la concimazione ha, in media, un effetto significativo, che i concimi organici non differiscono dai minerali (in media) e che il concime a lento rilascio è inferiore a quello normale.

### 3 Test di confronto multiplo

Non sempre le prove sono pianificate in modo da poter eseguire controtesti ortogonali e il ricercatore può essere interessato a confrontare tutte le possibili coppie di trattamenti. In questo caso sono molto comode le procedure di confronto multiplo, anche se esse sono spesso abusate dagli utilizzatori.

La gran parte delle procedure sono basate sul calcolo di una differenza critica minima, da utilizzare come base per il confronto tra due medie. Non è questa la sede opportuna per una disamina del problema, che è stato abbondantemente trattato da svariate rassegne bibliografiche. Basti solo menzionare il fatto che le diverse procedure differiscono per il livello di potenza e protezione e quindi per il rischio di errore di I e II specie.

Le procedure proposte sono molte, tra cui:

1. Minima differenza significativa (MDS);
2. Test di Duncan
3. Test di Newman-Keuls;
4. Honest Significant Difference (HSD) di Tukey
5. Test di confronto multiplo di Tukey;
6. Test di Scheffe;
7. Test di Bonferroni;
8. Test di Dunnet per il confronto con un testimone;
9. LSD di Bayes
10. Analisi cluster di Scott-Knott.

Se consideriamo le medie in tabella 1 (identificandole per brevità con  $A=21.33$ ,  $B=20.33$ ,  $C=17.67$  e  $D=13.33$ ) abbiamo sei possibili confronti.

Questi confronti vengono eseguiti in base ad una differenza critica minima, che cambia a seconda della procedura adottata. Nel caso della MDS il valore critico è piuttosto basso, in quanto viene scelto in modo da garantire una probabilità d'errore del 5% per ogni confronto (tasso d'errore per confronto:  $\alpha$ ). E' chiaro però che se i confronti sono molti (indicativamente più di 10) la probabilità di sbagliarne almeno uno è molto più elevata del 5% prefissato (tasso d'errore per esperimento:  $\alpha$ ).

Per far fronte a questo problema ed assicurare una certa costanza tra il tasso d'errore per confronto e il tasso d'errore per esperimento, le diverse metodiche alternative alla MDS seguono due strade. La prima ipotesi possibile è quella di abbassare il livello di protezione per confronto, così da raggiungere un certo valore di  $\alpha$  prefissato (ad esempio il test di Bonferroni). La seconda ipotesi è quella di abbassare  $\alpha$  solo per quei confronti che hanno delle medie distanti in graduatoria, laddove è più alta la possibilità di commettere un errore di I specie (dichiarare diverse due medie che non lo sono). In questo modo la differenza critica viene ad essere non più unica, ma dipendente dalla distanza tra le 2 medie in graduatoria (ad esempio il test di Duncan e di Newman-Keuls).

Più in dettaglio, la procedura di Newman Keuls e quella di Duncan sono molto simili nel principio, ma quest'ultima è meno conservativa ed adotta differenze critiche più basse; entrambe le procedure sono largamente adottate in quelle prove nelle quali è implicita una graduatoria di merito tra i soggetti (prove di confronto varietale, ad esempio). In entrambi casi, comunque, quando si confrontano due medie contigue in graduatoria si adotta una differenza critica pari alla MDS.

La HSD di Tukey adotta una differenza critica pari a quella più elevata calcolabile con la procedura di Newman-Keuls ed il test, di conseguenza, è molto più protetto che quest'ultimo. Per evitare un'eccessiva probabilità di errore di seconda specie (dichiarare uguali due medie che in realtà sono diverse), Tukey ha proposto il suo test di confronto multiplo (Multiple Range Test), che risulta dalla media tra la HSD e la differenza critica corrispondente di Newman-Keuls.

Tra tutte le procedure finora descritte, quella di Scheffe è la più conservativa e adotta una differenza critica unica, ma fissata in modo estremamente prudenziale.

La procedura di Scott-Knott è molto diversa nel principio, in quanto si tratta di una analisi dei cluster di tipo divisivo, con la quale le medie vengono inizialmente considerate in un unico raggruppamento, da suddividere in due sottogruppi in modo che la distanza tra essi sia la massima possibile. Nel caso in esempio, dopo aver ordinato le medie in senso crescente o decrescente, si valuterebbero le distanze tra gruppi per tutte le possibili divisioni (A vs BCD, AB vs CD, ABC vs D) scegliendo quella che presenta il valore di distanza più elevato. Lo stesso procedimento viene utilizzato per tutti i sottogruppi, interrompendo la divisione sulla base di un'opportuno test d'ipotesi.

Il test di Scott-Knott molto comodo in quanto i gruppi nei quali vengono suddivise le medie non si sovrappongono, anche se il risultato del confronto tra due medie dipende anche dal livello delle medie degli altri trattamenti inseriti in prova (tab. 2).

Tabella 2: Ampiezza delle differenze osservate e delle differenze critiche nelle procedure di confronto multiplo più usualmente impiegate (in grassetto i confronti significativi)

Confr.	Diff. oss.	Differenze critiche						
		LSD	Newman Keuls	Duncan MRT	Tukey's HSD	Tukey's MRT	Scheffe	Bonferroni
A vs B	1.00	2.874	2.874	2.874	3.991	3.432	4.353	4.3358
A vs C	3.66	<b>2.874</b>	<b>3.561</b>	<b>2.995</b>	3.991	3.776	4.353	4.3358
A vs D	8.00	<b>2.874</b>	<b>3.991</b>	<b>3.062</b>	<b>3.991</b>	<b>3.991</b>	<b>4.353</b>	<b>4.3358</b>
B vs C	2.66	2.874	2.874	2.874	3.991	3.432	4.353	4.3358
B vs D	7.00	<b>2.874</b>	<b>3.561</b>	<b>2.995</b>	<b>3.991</b>	<b>3.776</b>	<b>4.353</b>	<b>4.3358</b>
C vs D	4.34	<b>2.874</b>	<b>2.874</b>	<b>2.874</b>	<b>3.991</b>	<b>3.432</b>	4.353	4.3358

Per l'esecuzione di ognuna di queste procedure in R possibile ricorrere alla funzione `mulcomp()`, nel pacchetto `aomisc`.

L'anzidetta funzione accetta come argomenti il vettore delle medie, il numero di repliche (comune per tutte le medie), i gradi di libertà della varianza d'errore, la varianza d'errore dall'ANOVA e il livello di protezione (per default: 5%), e il tipo di procedura (LSD, NewmanKeuls, Duncan, Tu-



keyHSD, TukeyMRT, Scheffe, Bonferroni, rispettivamente per la MDS, il test di Newman-Keuls, il test di Duncan, la HSD di Tukey, il test di confronto multiplo di Tukey, il test di Scheffe e il test di Bonferroni). L'output viene gestito attraverso le lettere di significanza: medie seguite da almeno una lettera in comune non sono significativamente diverse per il livello  $\alpha$  prefissato.

```
library(aomisc)
medie <- c(13.33,17.67,20.33, 21.33)

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="TukeyMRT")
Multiple Comparison testing
  medie test
M4 21.33    a
M3 20.33    a
M2 17.67    a
M1 13.33    b

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="LSD")
Multiple Comparison testing
  medie test
M4 21.33    a
M3 20.33   ab
M2 17.67    b
M1 13.33    c

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="NewmanKeuls")
Multiple Comparison testing
  medie test
M4 21.33    a
M3 20.33   ab
M2 17.67    b
M1 13.33    c

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="Duncan")
Multiple Comparison testing
  medie test
M4 21.33    a
M3 20.33   ab
M2 17.67    b
M1 13.33    c

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="TukeyHSD")
Multiple Comparison testing
```

```

      medie test
M4 21.33      a
M3 20.33      a
M2 17.67      a
M1 13.33      b

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="TukeyMRT")
Multiple Comparison testing
      medie test
M4 21.33      a
M3 20.33      a
M2 17.67      a
M1 13.33      b

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="Scheffe")
Multiple Comparison testing
      medie test
M4 21.33      a
M3 20.33      a
M2 17.67     ab
M1 13.33      b

> mulcomp(medie, 3, 8, 2.33, conf.level=0.05, type="Bonferroni")
Multiple Comparison testing
      medie test
M4 21.33      a
M3 20.33      a
M2 17.67      a
M1 13.33      b

>

```

## 4 Referenze bibliografiche

- Bretz, F., T. Hothorn, and P. Westfall. 2002, On Multiple Comparison Procedures in R. *R News* 2: 14-17.
- Calinsky, T. and L. C. A. Corsten. 1985, Clustering means in ANOVA by simultaneous testing. *Biometrics*, 41, 39-48.
- Cargnelutti, A. F., L. Storck, L. A. Dal'Col, L. Pisaroglo de Carvalho, and P. Machado dos Santos. 2003, A precisao experimental relacionada ao uso de bordaduras nas extremidades das fileiras em ensaios de milho. *Ciencia Rural* 33: 607-614.
- Carmer, S. G. and M. R. Swanson. 1971, Detection of difference between

- mens: a Monte Carlo study of five pairwise multiple comparison procedures. *Agronomy Journal*, 63, 940-945.
- Carmer, S. G. 1976, Optimal significance levels for application of the least significant difference in crop performance trials. *Crop Science*, 16, 95-99.
- Chew, V. 1976, Comparing treatment means: a compendium. *Hortscience*, 11(4), 348-357.
- Cousens, R. 1988, Misinterpretation of results in weed research through inappropriate use of statistics. *Weed Research*, 28, 281-289.
- Edwards, A. W. F. and L. L. Cavalli-Sforza. 1965, A method for cluster analysis. *Biometrics*, 21, 362-375.
- Gates, C. E. and J. D. Bilbro. 1978, Illustration of a cluster analysis method for mean separation. *Agronomy journal*, 70, 462-465.
- O'Neill, R. and G. B. Wetherill. 1971, The present state of multiple comparison methods. *Journ. Roy. Stat. Soc.* , 2, 218-249.
- Petersen, R. G. 1977, Use and misuse of multiple comparison procedures. *Agronomy Journal*, 69, 205-208.
- Scott, A. J. and M. Knott. 1974, A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30, 507-512.
- Willavize, S. A., S. G. Carmer, and W. M. Walker. 1980, Evaluation of cluster analysis for comparing treatment means. *Agronomy journal*, 72,317-320.