

Anova a due vie e fattoriale

L'analisi della varianza a due vie e l'analisi della varianza fattoriale, dal punto di vista delle procedure di analisi e di lettura dei risultati, sono praticamente uguali. Non così per quanto riguarda invece la definizione dei modelli, cosa di cui qui non ci occupiamo specificatamente.

Nell'esempio che segue, utilizzeremo la funzione `aov()` per analizzare le relazioni fra due, e quattro variabili categoriali e una variabile numerica di risposta.

Dati e pacchetti

Tab. 1: Dati: npk

yield	rendimento
block	blocco
N	uso di azoto (nitrogeno)
P	uso di fosfato
k	uso di potassio

```
library(tidyverse)
```

```
npk %>%  
  ggplot(aes(block, yield)) +  
  geom_boxplot()
```

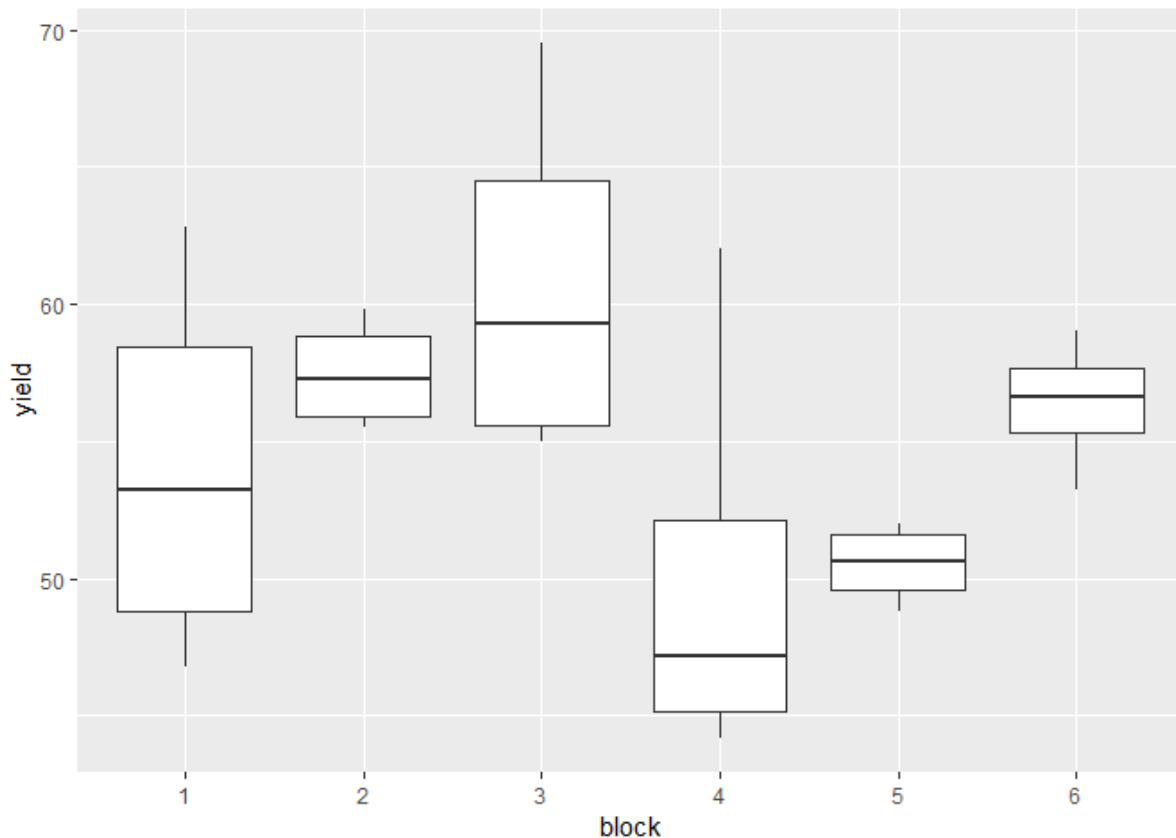


Fig. 1: Boxplot delle due variabili principali

Per rappresentare i dati, possiamo in questo caso utilizzare anche l'[interaction plot](#).

Anova a due vie

$Y \sim A * B$

Vedi: [Formula](#)

L'Anova a due vie equivale ad un'[analisi di regressione lineare multipla](#).

```
fit1 <- aov(yield ~ block * N, data = npk)
```

La tabella dell'Anova

```
summary(fit1)
```

Possiamo ottenere questa stessa tabella con il comando:

```
fit <- lm(yield ~ block * N, data = npk)  
summary.aov(fit)
```

Utilizziamo la versione *tidy* della tabella, per leggere l'output (vedi [pacchetto "broom"](#)):

```
library(broom)
(aov.df <- tidy(fit1)) # con la parentesi, salvo e visualizzo
l'oggetto
```

```
## # A tibble: 4 x 6
##   term          df sumsq meansq statistic p.value
##   <chr>      <dbl> <dbl>  <dbl>    <dbl>   <dbl>
## 1 block          5 343.    68.7     3.36    0.0397
## 2 N              1 189.    189.     9.26    0.0102
## 3 block:N        5  98.5    19.7     0.964    0.477
## 4 Residuals     12 245.    20.4     NA        NA
```

La tabella va letta così:

- la devianza totale (*Total SumSq*) è pari a `sum(aov.df$sumsq)`, ovvero 876,365;
- la devianza non spiegata (residua) è pari a 245,27;
- la devianza spiegata è dunque pari a 631,09, ovvero 876,36 \$- 245,27; di questa:
 - 343,29 è la devianza spiegata dalla prima variabile (*block*);
 - 189,28 è la devianza spiegata dalla seconda variabile (*N*);
 - 98,52 è la devianza spiegata dalla loro interazione (*block:N*);

I contrasti

Quando il modello include più di una variabile indipendente, sono impostati i [I contrasti](#) per tutte le variabili/gruppi:

```
coef(fit1)
```

```
## (Intercept)      block2      block3      block4      block5
block6
##      48.15        7.60       10.75       -3.30        2.00
6.45
##           N1  block2:N1  block3:N1  block4:N1  block5:N1
block6:N1
##      11.75       -8.35       -8.00       -1.20      -11.00
-8.25
```

La tabella dell'Anova con i contrasti ci consente di analizzare il peso di ciascuna modalità:

```
summary(fit1,
        split = list(block = list(1,2,3,4,5)))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## block          5  343.3   68.66   3.359 0.03967 *
## block: C1      1   31.8   31.83   1.557 0.23588
## block: C2      1  205.8  205.76  10.067 0.00803 **
## block: C3      1   36.9   36.93   1.807 0.20378
## block: C4      1   58.0   57.97   2.836 0.11797
## block: C5      1   10.8   10.81   0.529 0.48100
## N              1  189.3  189.28   9.261 0.01021 *
## block:N        5   98.5   19.70   0.964 0.47690
## block:N: C1    1    5.9    5.90   0.288 0.60101
## block:N: C2    1    6.7    6.67   0.326 0.57836
## block:N: C3    1   20.4   20.41   0.999 0.33738
## block:N: C4    1   31.5   31.51   1.542 0.23809
## block:N: C5    1   34.0   34.03   1.665 0.22124
## Residuals     12  245.3   20.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova fattoriale

$Y \sim A + B * C * D$

In questo caso, le variabili indipendenti sono più di due, e le combinazioni fra di loro possono essere svariate. Di base, però, l'interpretazione della tabella dell'Anova e dei contrasti resta la stessa.

Consideriamo il caso con quattro variabili, di cui solo tre interagiscono fra di loro¹⁾:

```
fit2 <- aov(yield ~ block + N * P * K, npk)
```

Nella tabella dell'Anova, troviamo la devianza spiegata dalle tre variabili indipendentemente l'una dall'altra, e quella spiegata dalle interazioni fra i diversi prodotti:

```
summary(fit2)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## block         5  343.3   68.66   4.447 0.01594 *
## N             1  189.3  189.28  12.259 0.00437 **
## P             1    8.4    8.40    0.544 0.47490
## K             1   95.2   95.20    6.166 0.02880 *
## N:P           1   21.3   21.28    1.378 0.26317
## N:K           1   33.1   33.13    2.146 0.16865
## P:K           1    0.5    0.48    0.031 0.86275
## Residuals    12  185.3   15.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Manca l'effetto dell'interazione N:P:K. Se stampiamo direttamente il risultato, notiamo però il messaggio “1 out of 13 effects not estimable”, che ci avvisa appunto che questo effetto non è stato calcolato.

```
fit2
```

```
## Call:
##   aov(formula = yield ~ block + N * P * K, data = npk)
##
## Terms:
##              block          N          P          K          N:P
N:K
## Sum of Squares  343.2950 189.2817   8.4017  95.2017  21.2817
33.1350
## Deg. of Freedom      5          1          1          1          1
1
##              P:K Residuals
## Sum of Squares    0.4817 185.2867
## Deg. of Freedom      1         12
##
## Residual standard error: 3.929447
## 1 out of 13 effects not estimable
## Estimated effects may be unbalanced
```

Script di esempio

E' possibile scaricare ed eseguire lo script dell'esempio:

[Es-anova-fatt.R](#)

```
# pacchetto
library(tidyverse)

# boxplot
npk %>%
  ggplot(aes(block, yield)) +
  geom_boxplot()

# ANOVA A DUE VIE
fit1 <- aov(yield ~ block * N, data = npk)

summary(fit1)

# oppure
```

```
library(broom)
(aov.df <- tidy(fit1))  # con la parentesi, salvo e visualizzo
                        l'oggetto

# coefficienti contrasti
coef(fit1)

# tabella anova con contrasti
summary(fit1,
          split = list(block = list(1,2,3,4,5)))

# ANOVA FATTORIALE
fit2 <- aov(yield ~ block + N * P * K, npk)

summary(fit2)

fit2
```

Analisi multivariata, Modelli lineari

¹⁾

in Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth edition.
Springer

From:

<https://www.agnesevardanega.eu/wiki/> - Ricerca Sociale con R

Permanent link:

https://www.agnesevardanega.eu/wiki/r/modelli/anova_two_way_factorial?rev=1754920218

Last update: 11/08/2025 13:50

