STATISTICA

Verifica di ipotesi - IV

ANalysis **O**f **Va**riance

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\overline{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla **concentrazione media di mercurio** (in ppm) in 4 **specie** di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\overline{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	[illegible]	1.06	0.63
4	[illegible]	0.50	0.36

C'è abbastanza evidenza per concludere che le concentrazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

Variabile qualitativa (di classificazione): ***FATTORE & LIVELLI***

Variabile quantitativa continua

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che **le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse**?

Cohen & Cohen, p.348

CONFRONTO DI MEDIE

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\overline{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

$(X_{1k}, \dots, X_{n_k k}),\ k = 1, \dots, G$ campioni casuali, indipendenti tra loro

$$X_{ik} \sim N(\mu_k, \sigma^2)$$

$H_0 :\ \mu_1 = \cdots = \mu_G$

H_1 : alcune medie differiscono tra loro

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

$(X_{1k}, \dots, X_{n_k k}),\ k = 1, \dots, G$ campioni casuali, indipendenti tra loro

$X_{ik} \sim N(\mu_k, \sigma^2)$

$H_0 : \mu_1 = \cdots = \mu_G$

H_1 : alcune medie differiscono tra loro

$$\bar{\bar{X}} = \frac{1}{n} \sum_{k=1}^{G} n_k \bar{X}_k$$

stimatore della media comune

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

$(X_{1k}, \dots, X_{n_k k}),\ k = 1, \dots, G$ campioni casuali, indipendenti tra loro

$$X_{ik} \sim \boldsymbol{N}(\mu_k, \boldsymbol{\sigma}^2)$$

$H_0 : \mu_1 = \cdots = \mu_G$ $\quad$ H_1 : alcune medie differiscono tra loro

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

$\left(X_{1k},\dots,X_{n_k k}\right),\ k=1,\dots,G$ campioni casuali, indipendenti tra loro

$X_{ik}\sim \boldsymbol{N}\left(\mu_k,\boldsymbol{\sigma^2}\right)$

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}(X_{ik}-\bar{X}_k)^2=\sum_{k=1}^{G}(n_k-1){S_k}^2$$

stimatore "universale" di σ^2

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

$\left(X_{1k}, \dots, X_{n_k k}\right),\ k = 1, \dots, G$ campioni casuali, indipendent

$$\bar{\bar{X}} = \frac{1}{n} \sum_{k=1}^{G} n_k \bar{X}_k$$

$$X_{ik} \sim \boldsymbol{N}\left(\mu_k, \boldsymbol{\sigma^2}\right)$$

$$H_0 \ :\ \mu_1 = \cdots = \mu_G$$

$$\sum_{k=1}^{G} \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2 = \sum_{k=1}^{G} (n_k - 1) {S_k}^2$$

stimatore "universale" di σ^2

$$\sum_{k=1}^{G} \sum_{i=1}^{n_k} \left(X_{ik} - \bar{\bar{X}}\right)^2$$

stimatore di σ^2 sotto H_0

ANOVA a un fattore (*one-way*)

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{\bar{X}}\right)^2=\sum_{k=1}^{G}\sum_{i=1}^{n_k}(X_{ik}-\bar{X}_k)^2+\sum_{k=1}^{G}n_k\left(\bar{X}_k-\bar{\bar{X}}\right)^2$$

$\left(X_{1k},\dots,X_{n_k k}\right),\ k=1,\dots,G$ campioni casuali, indipendent

$$\bar{\bar{X}}=\frac{1}{n}\sum_{k=1}^{G}n_k\bar{X}_k$$

$X_{ik}\sim \boldsymbol{N}\left(\mu_k,\boldsymbol{\sigma^2}\right)$

$H_0:\ \mu_1=\dots=\mu_G$

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}(X_{ik}-\bar{X}_k)^2=\sum_{k=1}^{G}(n_k-1){S_k}^2$$

stimatore "universale" di σ^2

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{\bar{X}}\right)^2$$

stimatore di σ^2 sotto H_0

ANOVA a un fattore (*one-way*)

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{\bar{X}}\right)^2 = \boxed{\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{X}_k\right)^2} + \sum_{k=1}^{G} n_k\left(\bar{X}_k-\bar{\bar{X}}\right)^2$$

$\left(X_{1k},\ldots,X_{n_k k}\right),\ k=1,\ldots,G$ campioni casuali, indipendenti

$X_{ik}\sim \boldsymbol{N}\left(\mu_k,\boldsymbol{\sigma}^2\right)$

$$\bar{\bar{X}} = \frac{1}{n}\sum_{k=1}^{G} n_k \bar{X}_k$$

$H_0 : \mu_1 = \cdots = \mu_G$

$$\boxed{\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{X}_k\right)^2 = \sum_{k=1}^{G}\left(n_k-1\right){S_k}^2}$$

stimatore “universale” di σ^2

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{\bar{X}}\right)^2$$

stimatore di σ^2 sotto H_0

ANOVA a un fattore (*one-way*)

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{\bar{X}}\right)^2=\sum_{k=1}^{G}\sum_{i=1}^{n_k}(X_{ik}-\bar{X}_k)^2+\boxed{\sum_{k=1}^{G}n_k\left(\bar{X}_k-\bar{\bar{X}}\right)^2}$$

$(X_{1k},\dots,X_{n_k k}),\ k=1,\dots,G$ campioni casuali, indipendent

$$\bar{\bar{X}}=\frac{1}{n}\sum_{k=1}^{G}n_k\bar{X}_k$$

$X_{ik}\sim \boldsymbol{N}(\mu_k,\boldsymbol{\sigma^2})$

$H_0 : \mu_1=\dots=\mu_G$

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}(X_{ik}-\bar{X}_k)^2=\sum_{k=1}^{G}(n_k-1){S_k}^2$$

stimatore "universale" di σ^2

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik}-\bar{\bar{X}}\right)^2$$

stimatore di σ^2 sotto H_0

Déjà-vu...

Analisi della varianza

K gruppi di dati, anche di diverse numerosità: $n_1, \dots, n_K$

l'analisi della varianza riguarda le **medie**!

$\bar{x}_i$ e σ^2_i media e varianza distorta(!) nel gruppo *i*-mo

media totale: $\bar{x} = \dfrac{n_1 \bar{x}_1 + \cdots + n_K \bar{x}_K}{n_1 + \cdots + n_K}$

Variabilità delle medie nei **gruppi** rispetto alla **media totale**

varianza totale: $\sigma^2 = \sigma_W{}^2 + \sigma_B{}^2$

Within

Between

media pesata *delle varianze* nei gruppi

normale formula della varianza, ottenuta *come se* tutti i dati in ciascun gruppo coincidessero con la media nel gruppo: *varianza delle medie*

$$\frac{\sigma_B{}^2}{\sigma_W{}^2 + \sigma_B{}^2}$$

(per K=2, è un punto interno, cioè *within*)

$$\sigma_W{}^2 = \frac{1}{n} \sum_{i=1,\dots,K} n_i \sigma^2_i \qquad \sigma_B{}^2 = \frac{1}{n} \sum_{i=1,\dots,K} n_i (\bar{x}_i - \bar{x})^2$$

è piccola se le $\bar{x}_i$ **non** sono molto diverse tra di loro

ANOVA a un fattore (*one-way*)

stimatore di σ^2 sotto H_0: $\mu_1 = \cdots = \mu_G$

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik} - \bar{\bar{X}}\right)^2 = \sum_{k=1}^{G}(n_k-1){S_k}^2 + \sum_{k=1}^{G} n_k\left(\bar{X}_k - \bar{\bar{X}}\right)^2$$

var. ***within***
media delle
varianze

var. ***between***
varianza
delle medie

$\left(X_{1k}, \dots, X_{n_k k}\right),\ k = 1, \dots, G$ campioni casuali, indipendenti tra loro

$$X_{ik} \sim \boldsymbol{N}\left(\mu_k, \boldsymbol{\sigma^2}\right)$$

ANOVA a un fattore (*one-way*)

stimatore di σ^2 sotto H_0: $\mu_1 = \cdots = \mu_G$

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik} - \bar{\bar{X}}\right)^2 = \sum_{k=1}^{G}(n_k - 1){S_k}^2 + \sum_{k=1}^{G} n_k\left(\bar{X}_k - \bar{\bar{X}}\right)^2$$

var. ***within*** — media delle varianze

var. ***between*** — varianza delle medie

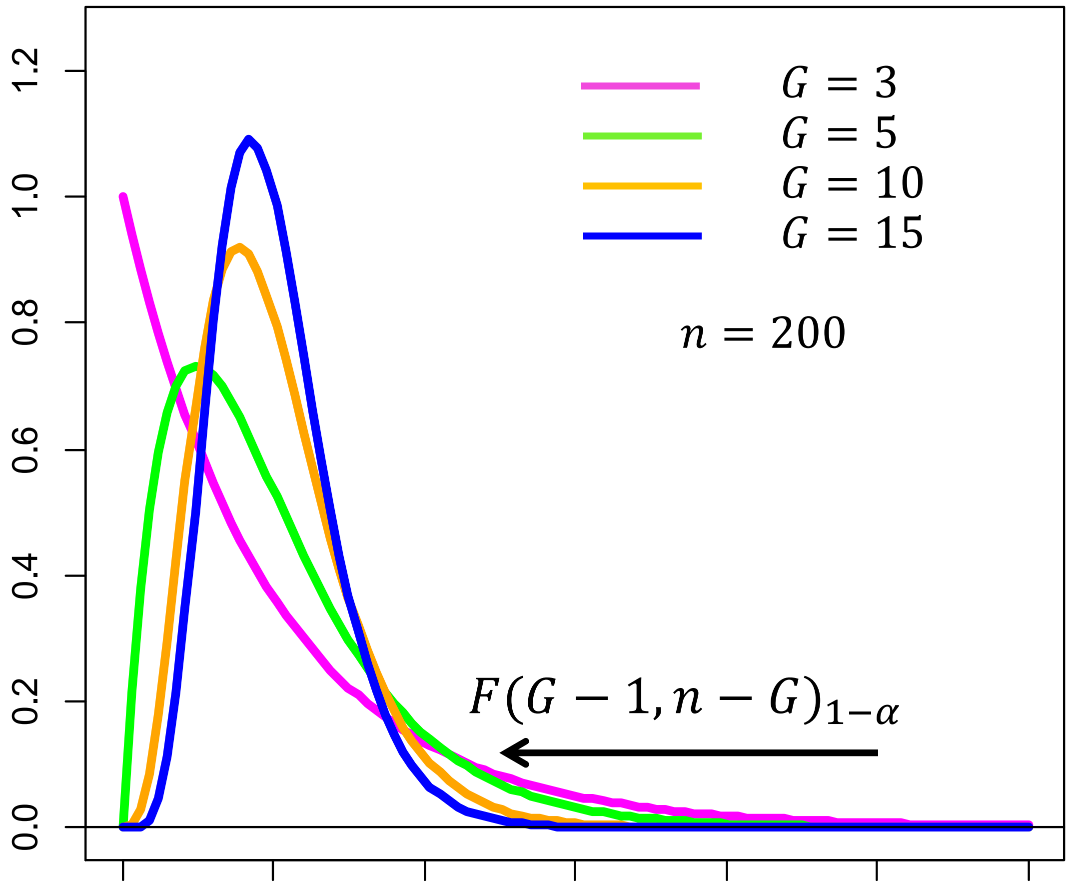
$\left(X_{1k}, \dots, X_{n_k k}\right),\ k = 1, \dots, G$ campioni casuali, indipendenti tra loro

$$X_{ik} \sim \boldsymbol{N}\left(\mu_k, \boldsymbol{\sigma^2}\right)$$

rifiuto H_0 al livello $1 - \alpha$ se:

$$\frac{\sum_{k=1}^{G} n_k(\bar{x}_k - \bar{\bar{x}})^2/(G-1)}{\sum_{k=1}^{G}(n_k - 1){s_k}^2/(n-G)} > F(G-1, n-G)_{1-\alpha}$$

ANOVA a un fattore (*one-way*)



Distribuzione di Fisher

rifiuto H_0 al livello $1-\alpha$ se:

$$\frac{\sum_{k=1}^{G} n_k(\bar{x}_k - \bar{\bar{x}})^2/(G-1)}{\sum_{k=1}^{G}(n_k-1){s_k}^2\,/(n-G)} > F(G-1, n-G)_{1-\alpha}$$

ANOVA a un fattore (*one-way*)

$(X_{11}, \dots, X_{n_1 1}), \dots, (X_{1G}, \dots, X_{n_G G})$ campioni casuali, indipendenti tra loro

$X_{ik} \sim N(\mu_k, \sigma^2) \qquad n = n_1 + \cdots + n_G$

$$\sum_{k=1}^{G}\sum_{i=1}^{n_k}\left(X_{ik} - \bar{\bar{X}}\right)^2 = \sum_{k=1}^{G}\sum_{i=1}^{n_k}(X_{ik} - \bar{X}_k)^2 + \sum_{k=1}^{G} n_k\left(\bar{X}_k - \bar{\bar{X}}\right)^2$$

Fonte	df	SS	MSS=$\frac{SS}{df}$	F-ratio
Between	$G-1$	$SSB = \sum_{k=1}^{G} n_k(\bar{x}_k - \bar{\bar{x}})^2$	$s^2{}_B = \frac{SSB}{G-1}$	$\frac{s^2{}_B}{s^2{}_W}$
Within	$\sum_{k=1}^{G}(n_k - 1) = n - G$	$SSW = \sum_{k=1}^{G}\sum_{i=1}^{n_k}(x_{ik} - \bar{x}_k)^2$	$s^2{}_W = \frac{SSW}{n-G}$	(*residual* sum of squares)
Mean	1	$n\bar{\bar{X}}^2$		
Total	n	$\sum_{k=1}^{K}\sum_{i=1}^{n_k} x_{ik}{}^2$		

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

Facciamo un saltino in R

Script4a.R

$$\frac{\sum_{k=1}^{G} n_k(\bar{x}_k - \bar{\bar{x}})^2/(G-1)}{\sum_{k=1}^{G}(n_k - 1){s_k}^2/(n-G)}$$

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

evidenza fortissima!

QUALI SONO LE MEDIE
TRA LORO DIFFERENTI?

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

evidenza fortissima!

QUALI SONO LE MEDIE TRA LORO DIFFERENTI?

CONFRONTI MULTIPLI

ANOVA a un fattore (*one-way*)

Errore di prima specie nei confronti multipli: prob. di trovare **almeno una differenza significativa** dato che tutte le medie sono uguali.

QUALI SONO LE MEDIE TRA LORO DIFFERENTI?

CONFRONTI MULTIPLI

ANOVA a un fattore (*one-way*)

Errore di prima specie nei confronti multipli: prob. di trovare **almeno una differenza significativa** dato che tutte le medie sono uguali.

H_0

$$\begin{aligned} \mu_1 &= \mu_2 \\ \mu_1 &= \mu_3 \\ \mu_1 &= \mu_4 \\ \mu_2 &= \mu_3 \\ \mu_2 &= \mu_4 \\ \mu_3 &= \mu_4 \end{aligned}$$

QUALI SONO LE MEDIE
TRA LORO DIFFERENTI?

CONFRONTI
MULTIPLI

ANOVA a un fattore (*one-way*)

Errore di prima specie nei confronti multipli: prob. di trovare **almeno una differenza significativa** dato che tutte le medie sono uguali.

=1-prob. di non trovare alcuna differenza significativa

$1-\alpha$ = prob. di non rifiutare $H_0: \mu_i = \mu_j$ quando è vera

$$= 1 - (1-\alpha)^{G(G-1)/2}$$

$$\alpha = per\ comparison\ error$$

$$\mu_1 = \mu_2$$
$$\mu_1 = \mu_3$$
$$\mu_1 = \mu_4$$
$$\mu_2 = \mu_3$$
$$\mu_2 = \mu_4$$
$$\mu_3 = \mu_4$$

H_0

QUALI SONO LE MEDIE TRA LORO DIFFERENTI?

CONFRONTI MULTIPLI

ANOVA a un fattore (*one-way*)

Errore di prima specie nei confronti multipli: prob. di trovare **almeno una differenza significativa** dato che tutte le medie sono uguali.

=1-prob. di non trovare alcuna differenza significativa

$1-\alpha$ = prob. di non rifiutare $H_0: \mu_i = \mu_j$ quando è vera

$$= 1 - (1-\alpha)^{G(G-1)/2}$$

$$\alpha = per\ comparison\ error$$

$$1 - 0.95^6 = 0.265!$$

$$G = 4$$
$$\alpha = 0.05$$

QUALI SONO LE MEDIE TRA LORO DIFFERENTI?

CONFRONTI MULTIPLI

ANOVA a un fattore (*one-way*)

Errore di prima specie nei confronti multipli: prob. di trovare **almeno una differenza significativa** dato che tutte le medie sono uguali.

=1-prob. di non trovare alcuna differenza significativa

$1-\alpha$ = prob. di non rifiutare $H_0 : \mu_i = \mu_j$ quando è vera

$$= 1 - (1-\alpha)^{G(G-1)/2}$$

$$\alpha = per\ comparison\ error$$

$$1 - 0.95^6 = 0.265!$$

$$G = 4$$
$$\alpha = 0.05$$

Bonferroni : $\dfrac{\alpha}{G(G-1)/2}$ da applicare a tutti i t –test per coppie di medie

ANOVA a un fattore (*one-way*)

Errore di prima specie nei confronti multipli: prob. di trovare **almeno una differenza significativa** dato che tutte le medie sono uguali.

=1-prob. di non trovare alcuna differenza significativa

$1-\alpha$ = prob. di non rifiutare $H_0: \mu_i = \mu_j$ quando è vera

$$= 1-(1-\alpha)^{G(G-1)/2}$$

$$\alpha = per\ comparison\ error$$

$$1-0.95^6 = 0.265!$$

$$G = 4$$
$$\alpha = 0.05$$

Bonferroni : $\dfrac{\alpha}{G(G-1)/2}$ da applicare a tutti i t –test per coppie di medie

Tukey (o Tukey-Kramer)

ANOVA a un fattore (*one-way*)

Errore di prima specie nei confronti multipli: prob. di trovare **almeno una differenza significativa** dato che tutte le medie sono uguali.

=1-prob. di non trovare alcuna differenza significativa

$1-\alpha$ = prob. di non rifiutare $H_0 : \mu_i = \mu_j$ quando è vera

$= 1 - (1 - \alpha)^{G(G-}$

$$s^2 = \sum_{k=1}^{G}(n_k - 1){s_k}^2$$

Bonferroni : $\dfrac{\alpha}{G(G-1)/2}$ da applicare a tutti i t –test per coppie di medie

Tukey (o Tukey-Kramer)

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\overline{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

evidenza fortissima!

QUALI SONO LE MEDIE
TRA LORO DIFFERENTI?

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

Commenti:

- Le varianze sono tutte uguali?
- senza dati disgregati non posso verificare la normalità, ma se li ho devo farlo!
- Come incidono le diverse numerosità dei gruppi?

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\overline{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

Commenti:

- Le varianze sono tutte uguali?
  - Bartlett's test per campioni gaussiani (bastano s_i)
  - Hartley's test per campioni gaussiani (bastano s_i)
  - Levene's Method (richiede i dati disgregati)

ANOVA a un fattore (*one-way*)

Esempio: I dati sulla concentrazione media di mercurio (in ppm) in 4 specie di squali catturati al largo della Florida sono riassunti in tabella:

Specie	n_i	$\bar{x}_i$	s_i
1	53	0.77	0.32
2	21	0.77	0.71
3	81	1.06	0.63
4	95	0.50	0.36

C'è abbastanza evidenza per concludere che le concentazioni medie di mercurio nelle diverse specie di squalo sono diverse?

Cohen & Cohen, p.348

Commenti:

- Le varianze sono tutte uguali?
- senza dati disgregati non posso verificare la normalità, ma se li ho devo farlo!
- Come incidono le diverse numerosità dei gruppi?

  - Tukey (o Tukey-Kramer) "approssimato";
  - Robustezza rispetto agli scostamenti dalla normalità

ANOVA a due fattori (*two-way*)

Studio multi-centrico dell'efficacia di un trattamento.

ANOVA a due fattori (*two-way*)

Studio multi-centrico dell'efficacia di un **trattamento**.

FATTORE 1: livelli dati, p.es, dalle dosi

ANOVA a due fattori (*two-way*)

Studio **multi-centrico** dell’efficacia di un trattamento.

FATTORE 2: livelli dati dai diversi centri

FATTORE 1: livelli dati, p.es, dalle dosi

ANOVA a due fattori (*two-way*)

	Fattore C				
Fattore R	X_{11}	$\cdots$	X_{1j}	$\cdots$	X_{1c}
	X_{21}	$\cdots$	X_{2j}	$\cdots$	X_{2c}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{i1}	$\cdots$	X_{ij}	$\cdots$	X_{ic}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{r1}	$\cdots$	X_{rj}	$\cdots$	X_{rc}

(un'osservazione per cella)

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipendenti, $i = 1, \ldots, r$; $j = 1, \ldots, c$

media generale

deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

ANOVA a due fattori (*two-way*)

	Fattore C				
Fattore R	X_{11}	$\cdots$	X_{1j}	$\cdots$	X_{1c}
	X_{21}	$\cdots$	X_{2j}	$\cdots$	X_{2c}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{i1}	$\cdots$	X_{ij}	$\cdots$	X_{ic}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{r1}	$\cdots$	X_{rj}	$\cdots$	X_{rc}

(un'osservazione per cella)

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipendenti, $i = 1, \dots, r$; $j = 1, \dots, c$

media generale

deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

Nel caso *one-way*:
$\mu_i = \mu + (\mu_i - \mu)$

ANOVA a due fattori (*two-way*)

	Fattore C				
Fattore R	X_{11}	$\cdots$	X_{1j}	$\cdots$	X_{1c}
	X_{21}	$\cdots$	X_{2j}	$\cdots$	X_{2c}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{i1}	$\cdots$	X_{ij}	$\cdots$	X_{ic}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{r1}	$\cdots$	X_{rj}	$\cdots$	X_{rc}

(un’osservazione per cella)

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipendenti, $i = 1, \dots, r$; $j = 1, \dots, c$

media generale

deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = 0$$

(condizione necessaria per l’identificabilità dei parametri, che sono $r + c + 1$ rispetto agli $r + c$ valori medi)

ANOVA a due fattori (*two-way*)

	Fattore C				
Fattore R	X_{11}	$\cdots$	X_{1j}	$\cdots$	X_{1c}
	X_{21}	$\cdots$	X_{2j}	$\cdots$	X_{2c}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{i1}	$\cdots$	X_{ij}	$\cdots$	X_{ic}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{r1}	$\cdots$	X_{rj}	$\cdots$	X_{rc}

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipendenti, $i = 1, \dots, r$; $j = 1, \dots, c$

media generale

deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = 0$$

$H_{0,\alpha} : \alpha_i = 0 \ \forall\, i$

$H_{0,\beta} : \beta_j = 0 \ \forall\, j$

ANOVA a due fattori (*two-way*)

$$\bar{X} = \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c} X_{ij}$$

	Fattore C				
Fattore R	X_{11}	$\cdots$	X_{1j}	$\cdots$	X_{1c}
	X_{21}	$\cdots$	X_{2j}	$\cdots$	X_{2c}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{i1}	$\cdots$	X_{ij}	$\cdots$	X_{ic}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{r1}	$\cdots$	X_{rj}	$\cdots$	X_{rc}

$$\bar{X}_{i\cdot} = \frac{1}{c}\sum_{j=1}^{c} X_{ij}$$

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipende $j = 1, \dots, c$

$$\bar{X}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r} X_{ij}$$

$$H_{0,\alpha} : \alpha_i = 0 \ \forall\, i$$

$$H_{0,\beta} : \beta_j = 0 \ \forall\, j$$

ANOVA a due fattori (*two-way*)

$$\hat{\mu} = \bar{X} = \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c} X_{ij}$$

	Fattore C				
Fattore R	X_{11}	$\cdots$	X_{1j}	$\cdots$	X_{1c}
	X_{21}	$\cdots$	X_{2j}	$\cdots$	X_{2c}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{i1}	$\cdots$	X_{ij}	$\cdots$	X_{ic}
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	X_{r1}	$\cdots$	X_{rj}	$\cdots$	X_{rc}

$$\bar{X}_{i\cdot} = \frac{1}{c}\sum_{j=1}^{c} X_{ij}$$

$$\hat{\alpha}_i = \bar{X}_{i\cdot} - \bar{X}$$

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipende ... $j = 1, \dots, c$

$$\bar{X}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r} X_{ij}$$

$$\hat{\beta}_j = \bar{X}_{\cdot j} - \bar{X}$$

$$H_{0,\alpha} : \alpha_i = 0 \;\; \forall\, i$$

$$H_{0,\beta} : \beta_j = 0 \;\; \forall\, j$$

ANOVA a due fattori (*two-way*)

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipendenti, $i = 1, \dots, r;\ j = 1, \dots, c$

Fonte	**df**	**SS**	**MSS=**$\frac{SS}{df}$	***F*-ratio**
Between Riga	$r-1$	$SS_R = c\sum_{i=1}^{r}(\bar{X}_{i\cdot} - \bar{\bar{X}})^2$	$s^2{}_R = \frac{SS_R}{r-1}$	$\frac{s^2{}_R}{s^2{}_{RC}}$
Between Colonna	$c-1$	$SS_C = r\sum_{j=1}^{c}(\bar{X}_{\cdot j} - \bar{\bar{X}})^2$	$s^2{}_C = \frac{SS_C}{c-1}$	$\frac{s^2{}_C}{s^2{}_{RC}}$
Within	$\tau = (r-1)(c-1)$	$SS_{RC} = \sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{\bar{X}})^2$	$s^2{}_{RC} = \frac{SS_C}{\tau}$	
Media	1	$rc\bar{\bar{X}}^2$		
Totale	rc	$\sum_{i=1}^{r}\sum_{j=1}^{c}X_{ij}{}^2$		

ANOVA a due fattori (*two-way*)

$X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ indipendenti, $i = 1, \dots, r;\ j = 1, \dots, c$

Fonte	df	SS	MSS=$\frac{SS}{df}$	*F*-ratio
Between Riga	$r-1$	$SS_R = c\sum_{i=1}^{r}(\bar{X}_{i\cdot} - \bar{\bar{X}})^2$	$s^2{}_R = \frac{SS_R}{r-1}$	$\frac{s^2{}_R}{s^2{}_{RC}}$
Between Colonna	$c-1$	$SS_C = r\sum_{j=1}^{c}(\bar{X}_{\cdot j} - \bar{\bar{X}})^2$	$s^2{}_C = \frac{SS_C}{c-1}$	$\frac{s^2{}_C}{s^2{}_{RC}}$
Within	$\tau = (r-1)(c-1)$	$SS_{RC} = \sum^{r}\sum^{c}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{\bar{X}})^2$	$s^2{}_{RC} = \frac{SS_C}{\tau}$	
Media	1			
Totale	rc	$\sum^{r}\sum^{c} X_{ij}{}^2$		

sotto $H_{0,\alpha}:\ \alpha_i = 0\ \forall i$ $\quad \frac{s^2{}_R}{s^2{}_{RC}} \sim F(r-1, \tau)$

sotto $H_{0,\beta}:\ \beta_j = 0\ \forall j$ $\quad \frac{s^2{}_C}{s^2{}_{RC}} \sim F(c-1, \tau)$

Esempio

In ciascuno di quattro appezzamenti di terreno vengono piantate 5 varietà di avena, ottenendo i seguenti dati di produttività:

	Terreno			
Avena	**1**	**2**	**3**	**4**
1	296	357	340	348
2	402	390	420	335
3	345	342	358	308
4	360	322	336	270
5	324	339	357	308

I dati sono compatibili con l'ip. che la produttività non dipenda

a) dal terreno?

b) dalla varietà?

Esempio

In ciascuno di quattro appezzamenti di terreno vengono piantate 5 varietà di avena, ottenendo i seguenti dati di produttività:

	Terreno			
Avena	**1**	**2**	**3**	**4**
1	296	357	340	348
2	402	390	420	335
3	345	342	358	308
4	360	322	336	270
5	324	339	357	308

I
l'i
d
a
b

Avena	Terreno	Produzione
1	1	296
1	2	357
1	3	340
1	4	348
2	1	402
…	…	…

Facciamo un saltino in R
script4a.R

Esempio

In ciascuno di quattro appezzamenti di terreno vengono piantate 5 varietà di avena, ottenendo i seguenti dati di produttività:

	Terreno			
Avena	**1**	**2**	**3**	**4**
1	296	357	340	348
2	402	390	420	335
3	345	342	358	308
4	360	322	336	270
5	324	339	357	308

I
l'i
d
a
b

Avena	Terreno	Produzione
1	1	296
1	2	357
1	3	340
1	4	348
2	1	402
…	…	…

```
> summary(aov(model))
            Df Sum Sq Mean Sq F value Pr(>F)
avena        4  10234  2558.6   4.470 0.0193 *
terr         3   6380  2126.6   3.715 0.0424 *
Residuals   12   6869   572.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

ANOVA a due fattori (*two-way*)

Confronto di metodi di insegnamento

Metodo A : insegnante 

Metodo B : insegnante 

Metodo C : insegnante 

ANOVA a due fattori (*two-way)*

Confronto di metodi di insegnamento

Metodo A : insegnante 

Metodo B : insegnante 

Metodo C : insegnante 

Stiamo confrontando

i metodi di

insegnamento

o gli insegnanti?

ANOVA a due fattori (*two-way)*

Confronto di metodi di insegnamento

	Prof. 1	Prof. 2	Prof. 3
Metodo A			
Metodo B			
Metodo C			

ANOVA a due fattori (*two-way)*

Confronto di metodi di insegnamento

	Prof. 1	Prof. 2	Prof. 3
Metodo A			
Metodo B			
Metodo C			

C'è un'effetto dovuto alla combinazione di un metodo con un dato insegnante?

ANOVA a due fattori (*two-way*)

	Fattore C				
Fattore R	$\boldsymbol{X}_{11}$	$\cdots$	$\boldsymbol{X}_{1j}$	$\cdots$	$\boldsymbol{X}_{1c}$
	$\boldsymbol{X}_{21}$	$\cdots$	$\boldsymbol{X}_{2j}$	$\cdots$	$\boldsymbol{X}_{2c}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	$\boldsymbol{X}_{i1}$	$\cdots$	$\boldsymbol{X}_{ij}$	$\cdots$	$\boldsymbol{X}_{ic}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	$\boldsymbol{X}_{r1}$	$\cdots$	$\boldsymbol{X}_{rj}$	$\cdots$	$\boldsymbol{X}_{rc}$

$X_{ij} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$ indipendenti, $i = 1, \ldots, r;\ j = 1, \ldots, c$

media generale

deviazione da μ per effetto della interazione tra riga i e colonna j

deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = \sum_{i=1}^{r} \gamma_{ij} (\forall\ j) = \sum_{j=1}^{c} \gamma_{ij} (\forall\ i) = 0$$

ANOVA a due fattori (*two-way*)

$\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijm})$

m osservazioni per ogni combinazione dei fattori

		Fattore C				
		$\cdots$	$\boldsymbol{X}_{1j}$	$\cdots$	$\boldsymbol{X}_{1c}$
		$\cdots$	$\boldsymbol{X}_{2j}$	$\cdots$	$\boldsymbol{X}_{2c}$
			$\cdots$	$\cdots$	$\vdots$
Fatto	$\boldsymbol{X}_{i1}$	$\cdots$	$\boldsymbol{X}_{ij}$	$\cdots$	$\boldsymbol{X}_{ic}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	$\boldsymbol{X}_{r1}$	$\cdots$	$\boldsymbol{X}_{rj}$	$\cdots$	$\boldsymbol{X}_{rc}$

$X_{ij} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$ indipendenti, $i = 1, \dots, r$; $j = 1, \dots, c$

media generale

deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

deviazione da μ per effetto della interazione tra riga i e colonna j

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = \sum_{i=1}^{r} \gamma_{ij} (\forall\, j) = \sum_{j=1}^{c} \gamma_{ij} (\forall\, i) = 0$$

ANOVA a due fattori (*two-way*)

$\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijm})$

m osservazioni per ogni combinazione dei fattori

			Fattore C			
			$\cdots$	$\boldsymbol{X}_{1j}$	$\cdots$	$\boldsymbol{X}_{1c}$
			$\cdots$	$\boldsymbol{X}_{2j}$		
				$\cdots$		
Fatto		$\boldsymbol{X}_{i1}$	$\cdots$	$\boldsymbol{X}_{ij}$	$\cdots$	$\boldsymbol{X}_{ic}$
		$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
		$\boldsymbol{X}_{r1}$	$\cdots$	$\boldsymbol{X}_{rj}$	$\cdots$	$\boldsymbol{X}_{rc}$

$$\bar{X}_{ij.} = \frac{1}{m}\sum_{s=1}^{m} X_{ijs}$$

$X_{ij} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$ indipendenti, $i = 1, \dots, r$; $j = 1, \dots, c$

media generale

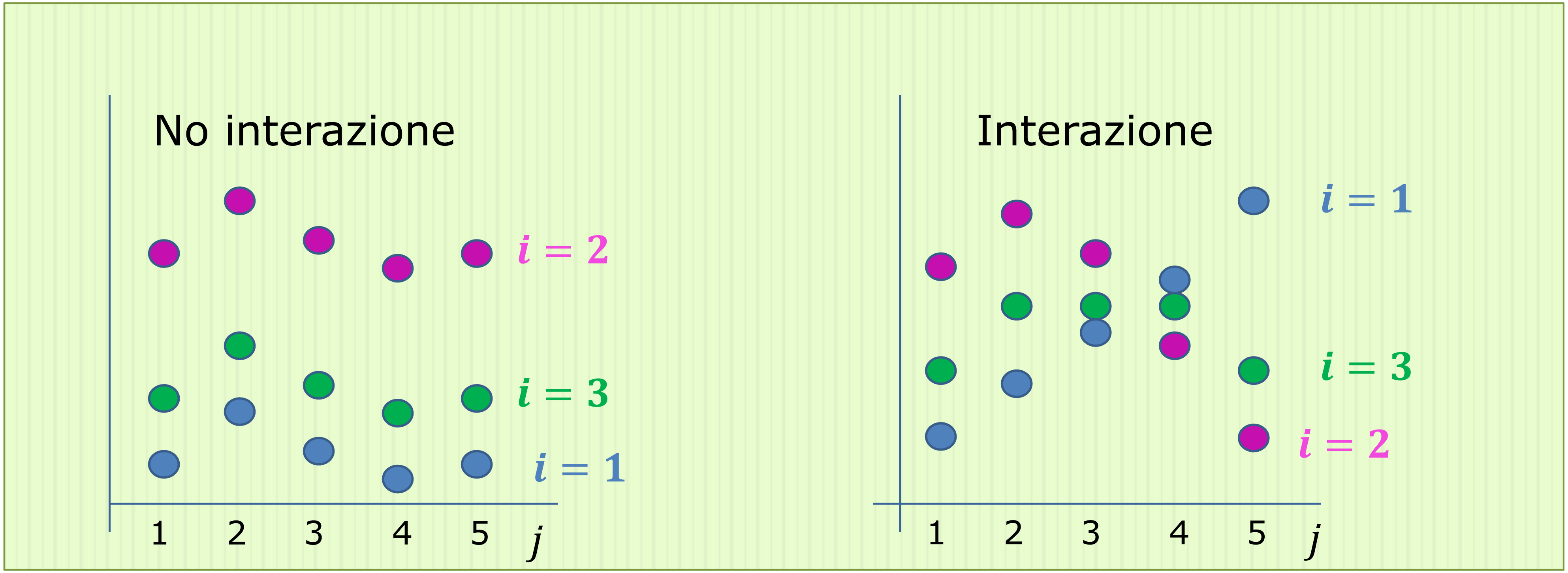
deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

deviazione da μ per effetto della interazione tra riga i e colonna j

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = \sum_{i=1}^{r} \gamma_{ij} (\forall\, j) = \sum_{j=1}^{c} \gamma_{ij} (\forall\, i) = 0$$

ANOVA a due fattori (*two-way*)



$X_{ij} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$ indipendenti, $i = 1, \dots, r;\ j = 1, \dots, c$

media generale

deviazione da μ per effetto del livello i del fattore R

deviazione da μ per effetto del livello j del fattore C

deviazione da μ per effetto della **interazione** tra riga i e colonna j

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = \sum_{i=1}^{r} \gamma_{ij} (\forall\, j) = \sum_{j=1}^{c} \gamma_{ij} (\forall\, i) = 0$$

ANOVA a due fattori (*two-way*)

$$\hat{\mu} = \bar{\bar{X}} = \frac{1}{mrc}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{s=1}^{m} X_{ijs}$$

	Fattore C				
Fattore R	$\boldsymbol{X}_{11}$	$\cdots$	[illegible]	[illegible]	$\boldsymbol{X}_{1c}$
	$\boldsymbol{X}_{21}$	$\cdots$	[illegible]	[illegible]	$\boldsymbol{X}_{2c}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	$\boldsymbol{X}_{i1}$	$\cdots$	$\boldsymbol{X}_{ij}$	$\cdots$	$\boldsymbol{X}_{ic}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	$\boldsymbol{X}_{r1}$	$\cdots$	$\boldsymbol{X}_{rj}$	$\cdots$	$\boldsymbol{X}_{rc}$

$$\bar{X}_{ij\cdot} = \frac{1}{m}\sum_{s=1}^{m} X_{ijs}$$

$$\bar{X}_{i\cdot\cdot} = \frac{1}{mc}\sum_{j=1}^{c}\sum_{s=1}^{m} X_{ijs}$$

$$\hat{\alpha}_i = \bar{X}_{i\cdot\cdot} - \bar{\bar{X}}$$

$X_{ij} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$ indip $= 1, \ldots, c$

$$\bar{X}_{\cdot j\cdot} = \frac{1}{mr}\sum_{i=1}^{r}\sum_{s=1}^{m} X_{ijs}$$

$$\hat{\gamma}_{ij} = \bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{\bar{X}}$$

$$\hat{\beta}_j = \bar{X}_{\cdot j\cdot} - \bar{\bar{X}}$$

ANOVA a due fattori (*two-way*)

$X_{ijs} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$ indipendenti, $i = 1, \dots, r$; $j = 1, \dots, c$; $s = 1, \dots, m$

Fonte	**df**	**SS**	***F*-ratio**
Riga Between	$r - 1$	$SS_R = cm \sum_{i=1}^{r} \left(\bar{X}_{i\cdot\cdot} - \bar{\bar{X}}\right)^2$	$\frac{SS_R/(r-1)}{SS_W/\tau_W}$
Colonna Between	$c - 1$	$SS_C = rm \sum_{j=1}^{c} \left(\bar{X}_{\cdot j\cdot} - \bar{\bar{X}}\right)^2$	$\frac{SS_C/(c-1)}{SS_W/\tau_W}$
Interazione	$\tau_I = (r-1)(c-1)$	$SS_I = m \sum_{i=1}^{r} \sum_{j=1}^{c} \left(\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{\bar{X}}\right)^2$	$\frac{SS_I/\tau_I}{SS_W/\tau_W}$
Within	$\tau_W = (m-1)rc$	$SS_W = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{s=1}^{m} \left(X_{ijs} - \bar{X}_{ij\cdot}\right)^2$	
Media	1	$mrc\bar{\bar{X}}^2$	
Totale	mrc	$\sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{s=1}^{m} {X_{ijs}}^2$	

ANOVA a due fattori (*two-way*)

$X_{ijs} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$ indipendenti, $i = 1, \dots, r;\ j = 1, \dots, c;\ s = 1, \dots, m$

Fonte	**df**	**SS**	***F*-ratio**
Riga Between	sotto $H_{0,\alpha} : \alpha_j = 0 \ \forall j$	$\frac{SS_R/(r-1)}{SSW/\tau_W} \sim F(r-1, \tau_W)$	$\frac{SS_R/(r-1)}{SSS_W/\tau_W}$
Colonna Between	sotto $H_{0,\beta} : \beta_j = 0 \ \forall j$	$\frac{SS_C/(c-1)}{SS_W/\tau_W} \sim F(c-1, \tau_W)$	$\frac{SS_C/(c-1)}{SS_W/\tau_W}$
Interazione	τ_I sotto $H_{0,\gamma} : \gamma_{ij} = 0 \ \forall i, j$	$= m \sum^{r} \sum^{c} \dots + \bar{\bar{X}})^2$ $\frac{SS_I/\tau_I}{SS_W/\tau_W} \sim F(\tau_I, \tau_W)$	$\frac{SS_I/\tau_I}{SS_W/\tau_W}$
Within	$\tau_W = (m-1)rc$	$SS_W = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{s=1}^{m} (X_{ijs} - \bar{X}_{ij\cdot})^2$	
Media	1	$mrc\bar{\bar{X}}^2$	
Totale	mrc	$\sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{s=1}^{m} {X_{ijs}}^2$	

Esercizio

Facciamo un saltino in R

Script4a.R

Confronto di metodi di insegnamento

	Prof. 1	Prof. 2	Prof. 3
Metodo A	95,85,74,74	60,90,80,70	86,77,75,70
Metodo B	90,80,92,82	89,90,91,86	83,70,75,72
Metodo C	70,80,85,85	68,73,78,93	74,86,91,89

ANOVA a due fattori (*two-way*)

Studio **multi-centrico** dell’efficacia di un trattamento.

FATTORE 2: livelli dati dai diversi centri

FATTORE 1: livelli dati, p.es, dalle dosi

ANOVA a due fattori (*two-way*)

Studio **multi-centrico** de[…]efficacia di un trattamento.

RANDOMIZED BLOCK DESIGN: un fattore di primario interesse e una variabile che rappresenta una caratteristica che causa eterogeneità tra le unità sperimentali (fattore di "disturbo")

Randomized block design: esempio.

Cfr rendimento di varie marche di benzina

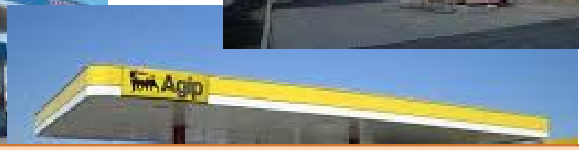
Randomized block design: esempio.

Cfr rendimento di varie marche di benzina

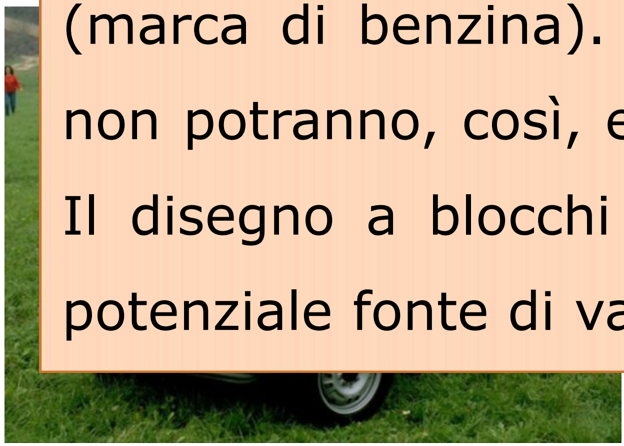
tipo e/o marca di auto influiscono sul rendimento, ma tale variabilità è necessaria!

Randomized block design: esempio.

Cfr rendimento di varie marche di benzina



Ogni marca di benzina viene testata da utilitarie ed auto sportive in ugual proporzione: in ogni *blocco* (tipologia di auto, utilitaria o sportiva nell’esempio) c’è una relativa omogeneità di unità sperimentali (auto), ciascuna delle quali viene assegnata casualmente ad un *trattamento* (marca di benzina). Le eventuali differenze tra le marche di benzina non potranno, così, essere attribuite al tipo di auto usata per la prova. Il disegno a blocchi randomizzato rimuove la tipologia di auto come potenziale fonte di variabilità e come eventuale variabile confondente.



necessaria!