

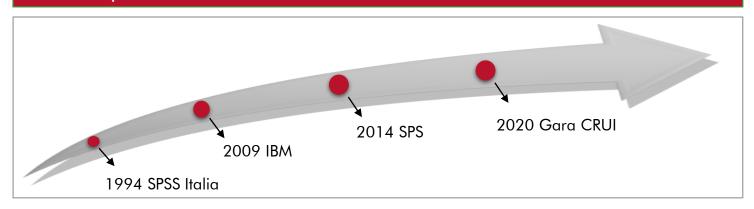
Webinar

l'Analisi della Varianza con IBM SPSS Statistics

Webinar | SPS Srl



Roadmap



Posizionamento

- +25 anni di esperienza sui prodotti SPSS
- IBM BP Software Support Provider con competenza IBM Expert in Data Science and Business Analytics
- Gold Business Partner IBM



Presenza sul me<u>rcato</u>

- Scuole e Università
- Enti di ricerca (CNR, ISS, CREA ecc.)
- Sanità (IRCCS, AO, AOU ecc.)
- Pubblica Amministrazione
- Associazioni, Fondazioni
- Aziende (GdO, Finance, Retail, ecc.)



Webinar | CRUI e SPS S.r.l.



Il contratto stipulato tra CRUI e SPS S.r.l. in data 21 ottobre 2019 si riferisce alla fornitura licenze Campus IBM SPSS Statistics del "Catalogo licenze CRUI Campus PA" e di licenze aggiuntive IBM SPSS Amos e IBM SPSS Statistics per usi amministrativi.

Tutte le informazioni sul contratto sono disponibili al link https://www.spss.it/crui-universita-italiane

Il referente commerciale per l'esecuzione contratto in SPS Srl è: Laura Zerbini (+39 335 1360538, <u>laura.zerbini@spss.it</u>)

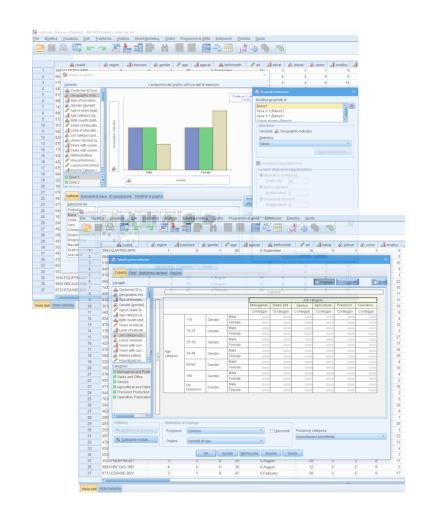




Webinar | IBM SPSS Statistics al lavoro...



Arianna Azzellino docente @SPS





Webinar | Analisi della Varianza



L'Analisi della Varianza (ANOVA)

Breve introduzione al metodo

Le componenti della Varianza

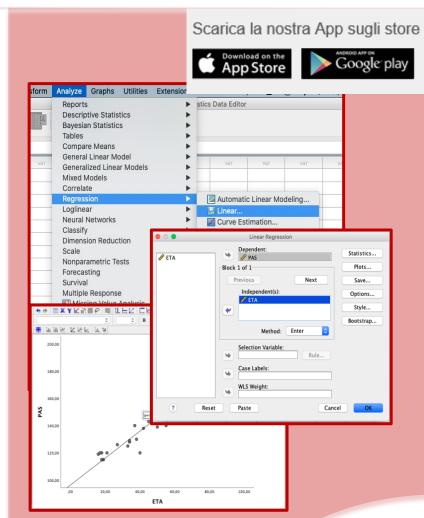
Come interpretare la variabilità, applicata ad esempi pratici

Il campionamento

Come farlo, interpretandone correttamente i dati ottenuti

L'incertezza analitica

Alcuni studi sperimentali con struttura più complessa



Webinar | Analisi della Varianza



Per fare delle analisi statistiche non serve essere statistici...

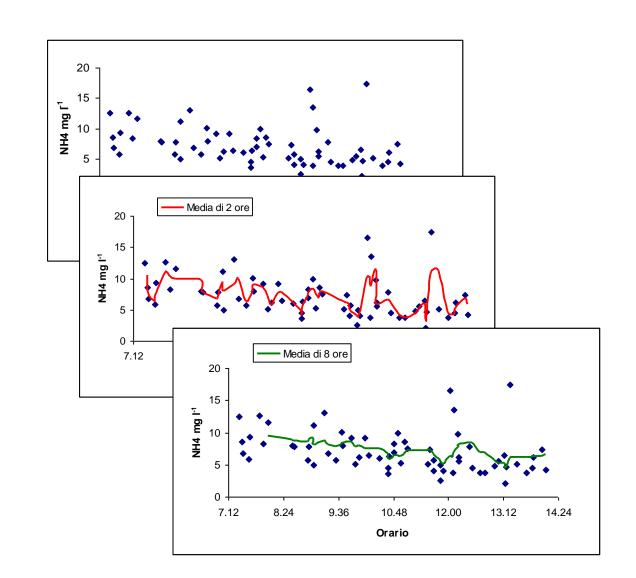
... ma è necessario capire quello che si sta facendo!



Interpretare la variabilità



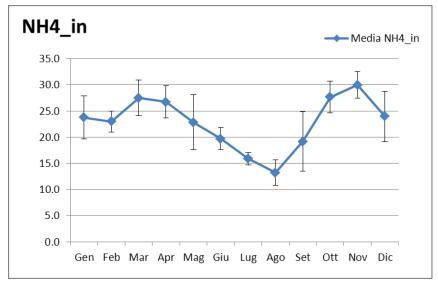
Q	NH4	data giorno set
1000	5.8	13-gen giov
1200	4.9	13-gen giov
1950	3.6	13-gen giov
2100	2.5	13-gen giov
2100	2.1	13-gen giov
1000	12.5	20-gen giov
1200	11.1	20-gen giov
2200	9.2	20-gen giov
2280	7.3	20-gen giov
2380	6.5	20-gen giov
1050	8.6	27-gen giov
1250	7.8	27-gen giov
1700	6.2	27-gen giov
2300	4.1	27-gen giov
2300	3.8	27-gen giov
1600	6.4	10-gen lun
1950	4.1	10-gen lun
1980	4.8	10-gen lun
1500	5.1	17-gen lun
2000	7.5	17-gen lun
2200	4.5	17-gen lun
1800	9.2	24-gen lun
2240	8.5	24-gen lun
2450	7.8	24-gen lun

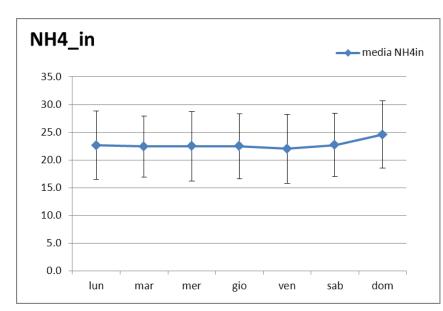


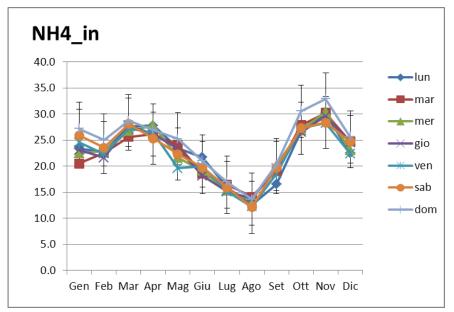


Interpretare la variabilità (2)











La verifica di ipotesi statistiche







Premessa



Argomentazioni logiche



Contraddizione (reduction ab absurdum)



Conclusione

Falsificazione della premessa

STATISTICA



Applicazione di un test



Risultato improbabile (p<0,05)



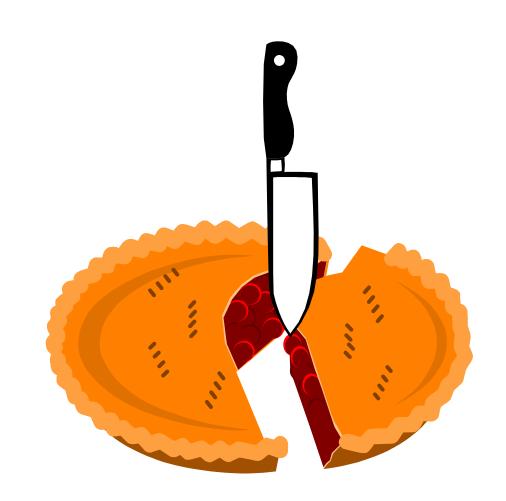
Conclusioni

- 1. L'ipotesi non è rifiutata (si è verificato un risultato altamente improbabile)
- L'ipotesi è rifiutata
 (il risultato osservato è inconsistente con quanto specificato nell'ipotesi)



Webinar | Il Modello di analisi della Varianza







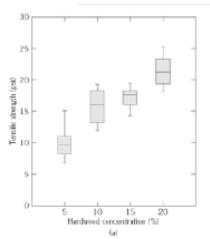
L'Analisi della Varianza (ANOVA)



È utilizzata per verificare se diversi gruppi indipendenti provengono da popolazioni con la stessa media

Esempio: si vuole valutare l'efficacia di diversi tipi di fertilizzazione.

Le colture trattate con diversi tipi di fertilizzazione hanno in media la stessa produzione o ci sono dei trattamenti che sono più efficaci di altri?



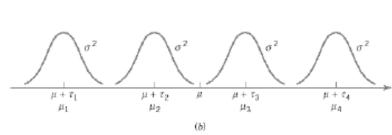


Figure 3-11 (a) Box plots of hardwood concentration data. (b) Display of the model in equation 3-65 for the completely randomized single-factor experiment.

Modello ANOVA ad un fattore



$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Il ricercatore è interessato a valutare se l'effetto α è significativo

$$H_0: \alpha = 0$$

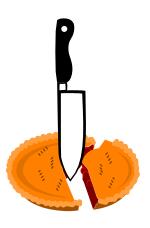
$$\mathbf{H}_0: \boldsymbol{\mu}_{\text{fert1}} = \boldsymbol{\mu}_{\text{fert2}} = \boldsymbol{\mu}_{\text{fert3}} = \boldsymbol{\mu}_{\text{fert4}}$$

Modello ANOVA ad un fattore



Varianza Totale

$$\bar{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$



si scompone la devianza totale i due componenti:

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 n_i + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Dev totale = devianza tra i gruppi + devianza entro i gruppi (devianza spiegata + devianza residua)



SPSS Webinar Esempio sull'ANOVA





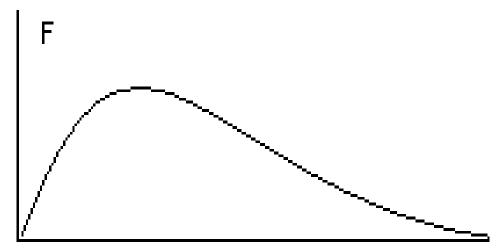
	Α	В	С	D	
1	4	5	7	2	
2	4	5	8	1	
3	5	6	7	2	
4	5	6	9	3	
5	6	7	6	3	
6	3	6	3	4	
7	4	4	2	5	
8	4	5	2	4	
9	3	6	2	4	
10	4	3	3	3	
tra					
media	4.2	5.3	4.9	3.1	media tot
(scarti) ²	0.030625	0.855625	0.275625	1.625625	4.375
Σ(scarti) ²	27.875				gdl
gdl (tra)	4-1				40-1
varianza (tra)	9.292				
entro					varianza tot
	Σ (scarti) ²	Σ (scarti) ²	Σ(scarti) ²	Σ (scarti) ²	2.94
	7.6	12.1	68.9	12.9	Σ (scarti) ²
Σ(scarti) ²	101.5				129.38
gdl (entro)	40-4				
varianza (entro)	2.819				129.38

Esempio sull'ANOVA



Pertanto per saggiare l'ipotesi nulla, è possibile costruire la statistica test:

$$F = \frac{\text{devianza spiegata}}{\text{devianza residua}} \frac{\text{(n° livelli -1)}}{\text{(n° unità - n° livelli)}} = \frac{\text{varianza spiegata}}{\text{varianza residua}}$$



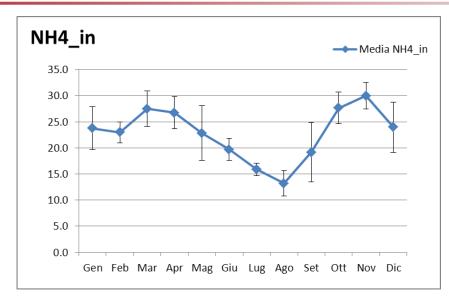
Quanto più il valore empirico di F si allontana da 0 quanto più probabile sarà il rifiuto dell'ipotesi nulla di uguaglianza degli effetti dei due farmaci

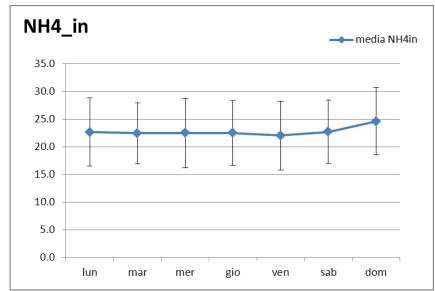


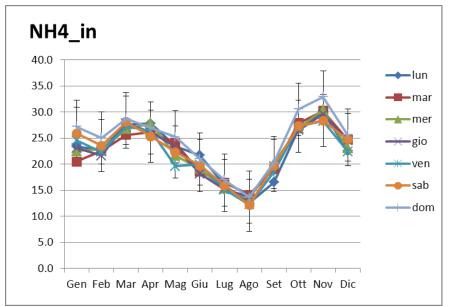


Contaminanti in ingresso impianto di trattamento













Webinar | Confronti multipli (test post-hoc)



Nel caso in cui i livelli del fattore sperimentale siano più di due, è interesse del ricercatore indagare quali tra i gruppi hanno portato al rigetto dell'ipotesi nulla

Tukey

Bonferroni

Scheffè

Questi test mettono in evidenza le coppie di medie statisticamente differenti



Studi sperimentali con struttura più complessa



Il ricercatore è interessato a valutare se nella determinazione dei valori della dipendendente incidono fattori ed effetti multipli

Disegno fattoriale semplice (ANOVA a due vie)

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + \varepsilon_{ijk}$$

- à l'effetto principale del fattore A (es. luogo misura),
- è l'effetto principale del fattore B (metodo campionamento),
- $\alpha\beta$ è l'effetto di interazione dei due fattori

Studi sperimentali con struttura più complessa



- Effetto del Fattore 1 (es. Località misura) ?
- Effetto del Fattore 2 (es. metodo di campionamento) ?
- Effetto combinato dei due fattori?

$$H_0: \alpha = 0$$

$$H_0: \mu_{\text{Località A1}} = \mu_{\text{Località A2}}$$

$$H_0: \beta = 0$$

$$H_0: \mu_{\text{Metodo B1}} = \mu_{\text{Metodo B2}}$$

$$H_0: \alpha\beta = 0$$

$$\mathbf{H}_{0}: \mu_{\text{Loc.A1} \cap \text{MetodoB1}} = \mu_{\text{Loc.A2} \cap \text{MetodoB2}} = \mu_{\text{Loc.A2} \cap \text{MetodoB2}} = \mu_{\text{Loc.A2} \cap \text{MetodoB1}}$$

$$F_{\alpha} = \frac{\text{Varianza Località}}{\text{Varianza Residua}}$$

$$F_{\beta} = \frac{\text{Varianza Metodo Campionamento}}{\text{Varianza Residua}}$$

$$_{\alpha\beta}F = \frac{\text{Varianza Località*Metodo}}{\text{Varianza Residua}}$$



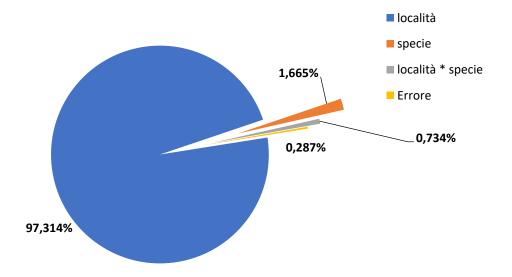


Test di effetti tra soggetti

Variabile dipendente: Cs127sup_b

Origine	Somma dei quadrati di tipo III	gl	Media quadratica	F	Sign.
Modello corretto	8442035691 ^a	5	1688407138	143.596	.000
Intercettazione	1.644E+10	1	1.644E+10	1397.799	.000
località	7965898388	2	3982949194	338.742	.000
spnum	68130441.83	1	68130441.83	5.794	.017
località * spnum	60090230.96	2	30045115.48	2.555	.082
Errore	1552064009	132	11758060.67		
Totale	2.014E+10	138			
Totale corretto	9994099700	137			

a. R-quadrato = .845 (R-quadrato adattato = .839)







Disegni a blocchi randomizzati, disegni gerarchici

Disegni cross-over, modelli a misure ripetute

Analisi della covarianza

Modelli ad effetti misti

Analisi della varianza multivariata

MODELLO LINEARE GENERALIZZATO





Disegno fattoriale (ANOVA a tre vie)

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_n + (\alpha \beta)_{ij} + \varepsilon_{ijnk}$$

- α è l'effetto principale del fattore A (es. località misura)
- B è l'effetto principale del fattore B (specie campionata)
- y è l'effetto principale del fattore C (incertezza metodo analitico)
- $\alpha\beta$ è l'effetto di interazione (località per specie campionata)



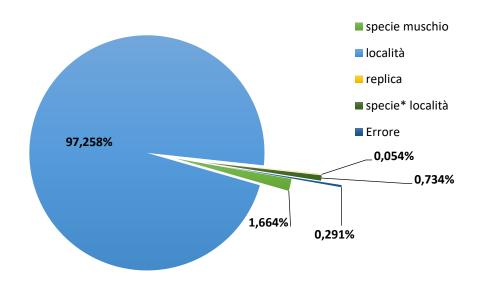


Test di effetti tra soggetti

Variabile dipendente: Cs127sup_b

Origine	Somma dei quadrati di tipo III	gl	Media quadratica	F	Sign.
			<u> </u>		
Modello corretto	8446424392 ^a	7	1206632056	101.353	.000
Intercettazione	1.644E+10	1	1.644E+10	1380.524	.000
spnum	68130441.83	1	68130441.83	5.723	.018
località	7965898388	2	3982949194	334.556	.000
replica	4388700.666	2	2194350.333	.184	.832
spnum * località	60090230.96	2	30045115.48	2.524	.084
Errore	1547675308	130	11905194.68		
Totale	2.014E+10	138			
Totale corretto	9994099700	137			

a. R-quadrato = .845 (R-quadrato adattato = .837)





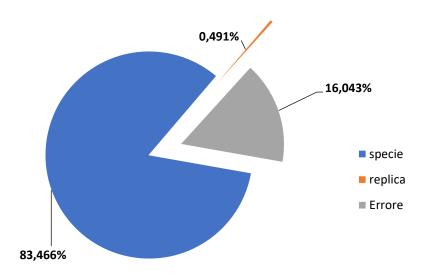


Test di effetti tra soggetti

Variabile dipendente: Cs127sup_b

Origine	Somma dei quadrati di tipo III	gl	Media quadratica	F	Sign.
Modello corretto	377743078 ^a	3	125914359.4	1.755	.159
Intercettazione	1.041E+10	1	1.041E+10	145.095	.000
spnum	373354377.5	1	373354377.5	5.203	.024
replica	4388700.666	2	2194350.333	.031	.970
Errore	9616356622	134	71763855.38		
Totale	2.014E+10	138			
Totale corretto	9994099700	137			

a. R-quadrato = .038 (R-quadrato adattato = .016)





Webinar | Importazione e preparazione dei dati



Ringraziamenti e Conclusioni

LAURA ZERBINI @SPS

