Metodi statistici per l'analisi dei dati

Esperimenti ad un singolo fattore

1

Analisi della Varianza (ANOVA) ad un singolo fattore – Introduzione

- Nell'esempio precedente sono state introdotte le tecniche più adeguate per confrontare **due trattamenti** distinti nella campagna sperimentale.
- I trattamenti possono anche essere visti come due differenti **livelli** di un **fattore** (nell'esempio precedente, il fattore è la concentrazione di additivo presente nella pasta).
- Molti esperimenti implicano però **più di due livelli** di un fattore.
- In questa sezione saranno presentati metodi per la progettazione e l'analisi di esperimenti **ad un singolo fattore** con *a* diversi livelli del fattore (o trattamenti)

2

ANOVA ad un singolo fattore – Esempio introduttivo

Esperimenti ad un singolo fattore

- Un ingegnere tessile intende investigare la resistenza di una nuova fibra sintetica al variare della percentuale di un additivo usato nella miscela.
- A tal riguardo esegue delle prove di resistenza su
  - a=5 diversi livelli di percentuale in peso di additivo:
    - 15%, 20%, 25%. 30% e 35%
  - n=5 diversi modelli
- Le misure totali sono $n \times a = 5 \times 5 = 25$.
- N.B. la successione delle misure è stabilita in modo casuale (*randomizzazione* delle misure)

3

ANOVA ad un singolo fattore – Esempio introduttivo

Esperimenti ad un singolo fattore

- Dati della campagna sperimentale

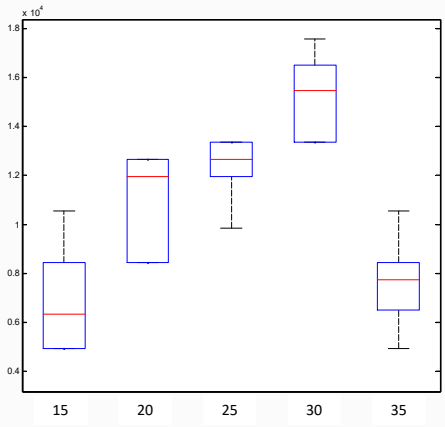
Percentuale di cotone	Osservazioni						
	1	2	3	4	5	Totale	Media
15	4900	4900	10500	7700	6300	34300	6860
20	8436	11900	8436	12600	12600	53972	10794.4
25	9800	12600	12600	13400	13300	61700	12340
30	13400	17600	15000	13300	16200	75500	15100
35	4900	7030	7700	10500	7700	37830	7566
						263302	10532.08

4

ANOVA ad un singolo fattore – Esempio introduttivo

Esperimenti ad un singolo fattore

- L'analisi grafica permette una prima valutazione qualitativa:



5

ANOVA ad un singolo fattore – Procedura

Esperimenti ad un singolo fattore

- Obbiettivo:
- Implementare una procedura **rigorosa** che permetta
  - di stabilire se si osservano trattamenti significativamente diversi o, equivalentemente, se il livello del fattore (la percentuale di additivo) ha un **impatto** sulla qualità del prodotto
  - individuare eventualmente quali sono i trattamenti che differiscono significativamente

6

ANOVA ad un singolo fattore – Nomenclatura

Esperimenti ad un singolo fattore

Percentuale di cotone	Osservazioni				
	1	2	3	4	5
15	4900	4900	10500	7700	6300
20	8436	11900	8436	12600	12600
25	9800	12600	12600	13400	13300
30	13400	17600	15000	13300	16200
35	4900	7030	7700	10500	7700

Ogni singola riga prende il nome di **trattamento**
Ciascun trattamento è costituito da ***n*** osservazioni (nel caso in esame *n* = 5)
L'analisi è svolta su ***a*** differenti trattamenti o livelli (nel caso in esame *a* =5)

La singola osservazione è caratterizzata da due indici:

$$x_{ij}$$

Indice ***i***: si riferisce alla riga i-esima **trattamento**

Indice ***j***: si riferisce alla j-esima **osservazione**

7

ANOVA ad un singolo fattore – Nomenclatura

Esperimenti ad un singolo fattore

- Modelli statistici per i dati sperimentali:
- **Modello delle medie**

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}$$

y_{ij}
j-esima osservazione dal livello i-esimo

μ_i
Media della risposta al livello (trattamento) i-esimo

ε_{ij}
Variabile aleatoria normale associata con la j-esima osservazione

8

ANOVA ad un singolo fattore – Nomenclatura

Esperimenti ad un singolo fattore

- Modelli statistici alternativi per descrivere i dati sperimentali:
- **Modello degli effetti**

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \begin{cases} i = 1,2,...,a \\ j = 1,2,...,n \end{cases}$$

y_{ij}	μ	τ_i	ε_{ij}
j-esima osservazione dal livello i-esimo	Media complessiva	Effetto del livello (trattamento) i-esimo	Variabile aleatoria normale associata con la j-esima osservazione

9

ANOVA ad un singolo fattore – Studio del modello degli effetti

Esperimenti ad un singolo fattore

- Nomenclatura usata nel seguito:

$y_{i\bullet} = \sum_{j=1}^{n} y_{ij}$ → Somma di tutte le osservazioni per il trattamento i-esimo

$y_{\bullet\bullet} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$ → Somma di tutte le osservazioni per tutti trattamenti

$\bar{y}_{i\bullet} = y_{i\bullet}/n$ → Media del trattamento i-esimo

$\bar{y}_{\bullet\bullet} = y_{\bullet\bullet}/N$ → "Grande" media del campione di dati ($N=n\cdot a$)

10

ANOVA ad un singolo fattore – Studio del modello degli effetti

Esperimenti ad un singolo fattore

- Si è interessati a testare l'eguaglianza tra i diversi gruppi.
- Le ipotesi statistiche possono essere scritte:

H_0: $\mu_1=\mu_2= \ldots =\mu_a$

H_1: $\mu_i \neq \mu_m$ per almeno una coppia (i,m)

- o, equivalentemente:

H_0: $\tau_1=\tau_2= \ldots =\tau_a=0$

H_1: $\tau_i \neq 0$ per almeno un valore i

11

ANOVA ad un singolo fattore – Decomposizione della somma totale dei quadrati

Esperimenti ad un singolo fattore

- Si consideri la **somma totale dei quadrati** *SST*:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{\bullet\bullet}\right)^2$$

- È una misura della variabilità complessiva presente nei dati.
- Con qualche passaggio:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{n}\left[\left(y_{ij} - \bar{y}_{i\bullet}\right)+\left(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right)\right]^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{i\bullet}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}\left(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right)^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{i\bullet}\right)\left(y_{i\bullet} - \bar{y}_{\bullet\bullet}\right)$$

$$= 0$$

12

ANOVA ad un singolo fattore – Decomposizione della somma totale dei quadrati

Esperimenti ad un singolo fattore

- In conclusione si ha:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{\bullet\bullet}\right)^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{i\bullet}\right)^2 + n\sum_{i=1}^{a}\left(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right)^2$$

SSE
Somma dei quadrati delle differenze **all'interno** dei trattamenti

$SS_{Treatments}$
Somma dei quadrati delle differenze **tra** i trattamenti

13

ANOVA ad un singolo fattore – Decomposizione della somma totale dei quadrati

Esperimenti ad un singolo fattore

- Interpretazione dei termini – **Somma dei quadrati degli errori:**
- **SSE** ha un numero di gradi di libertà pari a (*N*-*a*)
  - *N* è il numero totale di punti a disposizione
  - *a* è il numero di informazioni usato per calcolare le medie della singola colonna

$$MSE = \frac{SSE}{N-a}$$

Stima della varianza comune **all'interno** dei trattamenti

- Si può **dimostrare** che il valore atteso per MSE coincide con la varianza dell'errore sperimentale:

$$E[MSE] = \sigma^2$$

14

ANOVA ad un singolo fattore – Decomposizione della somma totale dei quadrati

Esperimenti ad un singolo fattore

- Interpretazione dei termini – **Somma dei quadrati dei trattamenti:**
- Analogamente, si può facilmente verificare che il numero di gdl di $\boldsymbol{SS_{Treatments}}$ è pari ad (*a-1*) per cui:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1}$$

⇒ Stima della varianza **tra** i trattamenti

- Anche in questo caso si può dimostrare che

$$E\left[MS_{Trearments}\right] = \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

15

ANOVA ad un singolo fattore – Decomposizione della somma totale dei quadrati

Esperimenti ad un singolo fattore

- Se l'ipotesi nulla **H_0: τ_i=0 per ogni *i* fosse vera**:

$$E\left[MSE\right] = E\left[MS_{Trearments}\right] = \sigma^2$$

- In presenza di almeno un trattamento significativamente diverso da zero ($\tau_i \neq 0$ per almeno un i):

$$E\left[MS_{Trearments}\right] > E\left[MSE\right] = \sigma^2$$

- Intuitivamente, la sorgente di varianza presente **tra** i trattamenti non è della stessa natura della varianza presente **all'interno** dei trattamenti (misura verosimilmente **genuina** dell'errore sperimentale)

16

ANOVA ad un singolo fattore – Decomposizione della somma totale dei quadrati

Esperimenti ad un singolo fattore

- In conclusione la dispersione totale dei dati può essere decomposta in due distinti contributi:

$$SST = SSE + SS_{\text{Treatments}}$$

- Inoltre, in assenza di influenza dei trattamenti, si ha:

$$\frac{SST}{\sigma^2} \sim \chi^2_{N-1} \qquad \frac{SSE}{\sigma^2} \sim \chi^2_{N-a} \qquad \frac{SS_{Treatments}}{\sigma^2} \sim \chi^2_{a-1}$$

- Le VA *SST*, *SSE* e $SS_{\text{Treatments}}$ sono **indipendenti** in quanto soddisfano il teorema di Cochran (vedi lucido successivo)

17

Decomposizione della somma totale dei quadrati – Analisi statistica

Esperimenti ad un singolo fattore

- **Teorema di Cochran**
- Siano $Z_i \sim NID(0,1)$ per i=1,2, …,ν

$$\sum_{i=1}^{\nu} Z_i^2 = Q_1 + Q_2 + \ldots + Q_s$$

- dove s ≤ ν e ciascuna Q_i abbia ν_i g.d.l. (i = 1,2, …, s)
- Allora Q_1, Q_2, …,Q_s sono VA di tipo χ^2 **indipendenti** con $\nu_1,\nu_2,\ldots,\nu_s$ g.d.l. rispettivamente, se e solo se

$$\nu = \nu_1 + \nu_2 + \ldots + \nu_s$$

18

Decomposizione della somma totale dei quadrati – Analisi statistica

Esperimenti ad un singolo fattore

- In conclusione, se l'ipotesi nulla fosse vera, il rapporto delle varianze

$$F_0 = \frac{SS_{Treatments}/(a-1)}{SSE/(N-a)} = \frac{MS_{Treatments}}{MSE}$$

  sarebbe distribuito secondo una F di Fisher a (*a-1,N-a*) gdl
- Valori di F_0»1 sono poco verosimili e portano al rigetto dell'ipotesi nulla di partenza

19

ANOVA ad un singolo fattore – Ricetta 1/2

Esperimenti ad un singolo fattore

1. Scegliere un livello di significatività α del test (in genere α=0.05)
2. Calcolare il **valore critico** $F_{\alpha,a-1,N-a}$ tale che:

$$P(F \le F_{\alpha,a-1,N-a}) = 1-\alpha$$

3. essendo *F* la Fisher a (*a-1,N-a*) gdl

20

ANOVA ad un singolo fattore – Ricetta 2/2

Esperimenti ad un singolo fattore

4. Calcolare il rapporto F_0 delle varianze per il set di dati:

$$F_0 = \frac{SS_{Treatments}/(a-1)}{SSE/(N-a)} = \frac{MS_{Treatments}}{MSE}$$

5. Si confronta il valore F_0 osservato con il valore critico $F_{\alpha,a-1,N-a}$

6. Se

$$F_0 > F_{\alpha,a-1,N-a}$$

Si **rigetta** l'ipotesi nulla ed esiste almeno un trattamento significativamente diverso dagli altri

21

ANOVA ad un singolo fattore – Tabella ANOVA

Esperimenti ad un singolo fattore

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0
Trattamenti	$SS_{Treatments} = n\sum_{i=1}^{a}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$	a-1	$MS_{Treatments}$	$F_0 = \frac{MS_{Treat.}}{MSE}$
Errore	$SSE = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\bullet})^2$	N-a	MSE	
Totale	$SST = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\bullet\bullet})^2$	N-1		

22

ANOVA ad un singolo fattore – Tabella Esempio

- Esperimento – Resistenza della fibra sintetica

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0
Trattamenti	2.325e+08	4	5.81e+07	14.69
Errore	7.909e+07	20	3.95e+06	
Totale	3.115e+08	24		

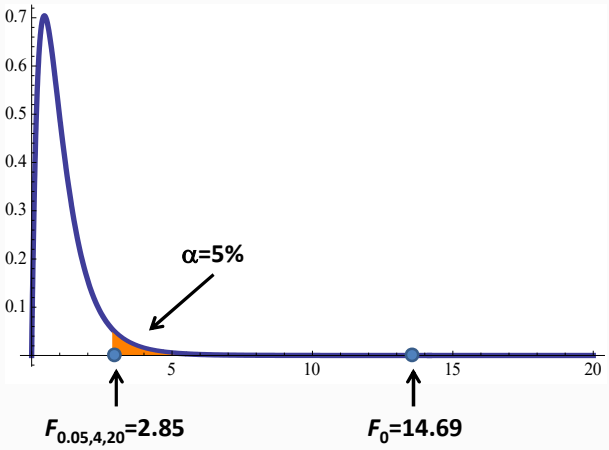
- Dalle tabelle si trova $F_{0.05,4,20}=2.85$
- È possibile anche calcolare il p-value per la statistica test:

P-value=9.41e-06

23

Tabella ANOVA Singolo Fattore – Esempio

- Distribuzione *F* di riferimento per la statistica dell'esempio



24

Tabella ANOVA Singolo Fattore – Calcolo semplificato Somme dei Quadrati

Esperimenti ad un singolo fattore

- In genere si ricorre a software dedicati per il calcolo dei termini presenti nel test ANOVA
- Nel caso si dovesse ricorrere ad un calcolo manuale, è possibile sfruttare delle espressioni più semplici:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{\bullet\bullet}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{N}$$

$$SS_{Treatments} = n\sum_{i=1}^{a}\left(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right)^2 = \frac{1}{n}\sum_{i=1}^{a} y_{i\bullet}^2 - \frac{y_{\bullet\bullet}^2}{N}$$

$$SSE = SST - SS_{Treatments}$$

25

ANOVA ad un singolo fattore – Stima dei parametri del modello

Esperimenti ad un singolo fattore

- Presentiamo ora stimatori per i parametri del modello a effetti a fattore singolo:

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

- Si può dimostrare che:

$$\hat{\mu} = \bar{y}_{\bullet\bullet}$$

$$\hat{\tau}_i = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet} \qquad i = 1,\ldots,a$$

- Inoltre, uno stimatore puntuale di $\mu_i=\mu+\tau_i$ è:

$$\hat{\mu}_i = \bar{y}_{i\bullet}$$

26

ANOVA ad un singolo fattore – Stima degli intervalli di fiducia dei parametri

Esperimenti ad un singolo fattore

- Se si assume che ciascuna misura sia indipendente e normalmente distribuita si ha che, per la singola media:

$$\bar{y}_{i\bullet} \sim N\left(\mu_i, \sigma^2/n\right)$$

- Se σ^2 fosse nota, si potrebbe usare la distribuzione normale per determinare gli intervalli di fiducia.
- Come stima della varianza è possibile comunque usare MSE (misura genuina dell'errore sperimentale)

$$\sigma^2 \rightarrow MSE$$

- e basare il calcolo dell'intervallo di fiducia sulla *t* di student a *N-a* gdl.

27

ANOVA ad un singolo fattore – Stima degli intervalli di fiducia dei parametri

Esperimenti ad un singolo fattore

- Un intervallo di fiducia per la media del trattamento i-esimo è quindi dato da:

$$\bar{y}_{i\bullet} - t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}} \le \mu_i \le \bar{y}_{i\bullet} + t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}}$$

- Analogamente, un intervallo di fiducia per la differenza di due medie μ_i-μ_k di trattamenti è dato da:

$$\left(\bar{y}_{i\bullet} - \bar{y}_{k\bullet}\right) - t_{\alpha/2,N-a}\sqrt{\frac{2MSE}{n}} \le \mu_i - \mu_k \le \left(\bar{y}_{i\bullet} - \bar{y}_{k\bullet}\right) + t_{\alpha/2,N-a}\sqrt{\frac{2MSE}{n}}$$

- Da notare che l'intervallo di fiducia risulta più grande rispetto al caso della singola media

28

Verifica adeguatezza del modello – Analisi dei residui

Esperimenti ad un singolo fattore

- **Definizione**
- Dato il modello

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

- Si definisce **residuo e_{ij}** la distanza tra la generica osservazione e la corrispondente previsione del modello.

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i\bullet}$$

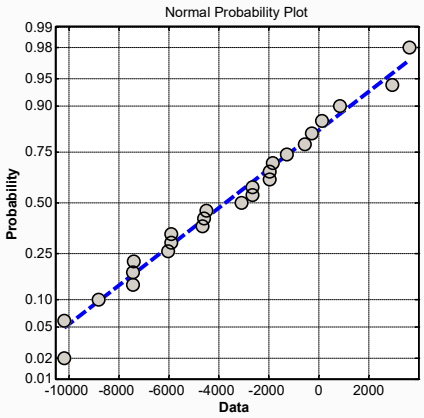
- Se il modello è adeguato, i residui dovrebbero apparire **senza una evidente struttura** (il determinismo è catturato completamente dal modello)

29

Verifica adeguatezza del modello – Analisi dei residui

Esperimenti ad un singolo fattore

- Riportando i residui su una carta probabilistica si può verificare una eventuale deviazione dalla assunzione di normalità



30

Verifica adeguatezza del modello – Analisi dei residui

Esperimenti ad un singolo fattore

- Si possono introdurre i **residui standardizzati**:

$$d_{ij} = \frac{e_{ij}}{\sqrt{MSE}}$$

- Se i residui sono $N(0,\sigma^2)$, i residui standardizzati saranno VA approssimativamente di tipo standard (*MSE* è una stima di σ^2)

$$d_{ij} \sim N(0,1)$$

- e pertanto (nel 95% dei casi circa)

$$-2 < d_{ij} < 2$$

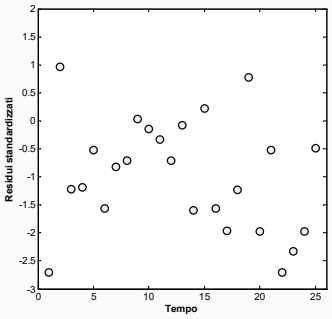
- Residui standardizzati $|d_{ij}| \gg 2$ si ritengono incompatibili con la campagna sperimentale (**outliers**) e sono pertanto da rimuovere

31

Verifica adeguatezza del modello – Analisi dei residui

Esperimenti ad un singolo fattore

- Diagramma dei residui rispetto all'ordine di esecuzione delle prove sperimentali (tempo)



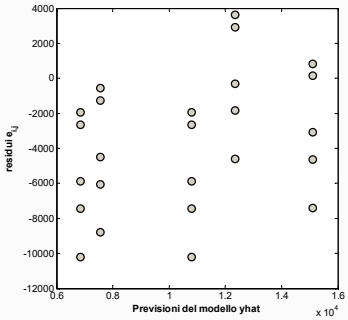
- Eventuali trend negativi/positivi nel grafico dei residui potrebbe suggerire che l'**assunzione di indipendenza** sugli errori è stata violata

32

Verifica adeguatezza del modello – Analisi dei residui

Esperimenti ad un singolo fattore

- Diagramma dei residui rispetto alle previsioni del modello



- Anche in questo caso, non si osserva la presenza di una struttura

33

Confronto tra i diversi trattamenti

Esperimenti ad un singolo fattore

- In seguito al rigetto dell'ipotesi nulla di partenza del test ANOVA, esistono delle procedure per stabilire **quali** siano i trattamenti specifici che differiscono significativamente dagli altri.
- Una possibilità è rappresentata dal confronto tra tutte le coppie possibili delle medie dei trattamenti.
  - N.B. Eseguire tutte le possibili combinazioni di test statistici su due trattamenti non risulta la scelta più adeguata dato che porterebbe ad un'amplificazione notevole dell'errore di tipo I.
- Un test adeguato per il confronto tra le diverse coppie dei trattamenti è il test di **Tukey** (1953).

34

Confronto tra i diversi trattamenti – Test di Tukey

Esperimenti ad un singolo fattore

- Una volta rigettata H_0, si intende eseguire un **test statistico** confrontando tutte le possibili combinazioni di medie dei trattamenti:

$$H_0: \quad \mu_i=\mu_j$$

$$H_1: \quad \mu_i\neq\mu_j$$

  per ogni coppia (*i*,*j*).
- Tukey ha proposto una procedura per questo test delle ipotesi, la cui significatività complessiva è pari proprio ad α (nel caso in cui le dimensioni del campione siano uguali per tutti i trattamenti).

35

Confronto tra i diversi trattamenti – Test di Tukey

Esperimenti ad un singolo fattore

- Si fa riferimento alla distribuzione della seguente **statistica "studentizzata" di intervallo**:

$$q=\frac{\bar{y}_{\max}-\bar{y}_{\min}}{\sqrt{\frac{MSE}{n}}}=\frac{|T|}{\sqrt{\frac{MSE}{n}}}$$

- dove $\bar{y}_{\max}$ e $\bar{y}_{\min}$ sono, rispettivamente, la massima e minima media campionarie, sul gruppo di *a* medie campionarie.
- È possibile ricavare (da tabelle disponibili in letteratura) i valori criciti $q_\alpha(a,f)$ della statistica, dove:
  - f è il numero di gdl associato alla varianza MSE
  - a è il numero di trattamenti presi in considerazione
  - α è il livello di significatività del test

36

Confronto tra i diversi trattamenti – Test di Tukey

Esperimenti ad un singolo fattore

- Il test stabilisce che due medie sono significativamente differenti se il valore assoluto delle loro differenze eccede:

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MSE}{n}}$$

- Nel caso di dimensioni dei campioni non uguali $n_i \neq n_j$:

$$T_\alpha = \frac{q_\alpha(a, f)}{\sqrt{2}}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

37

Confronto tra i diversi trattamenti – Test di Tukey – Esercizio

Esperimenti ad un singolo fattore

- Applichiamo il test di Tukey all'esempio (con un livello di significatività α=0.05):
  - a=5 trattamenti
  - f=20 gdl per l'errore
- Dalle tabelle si trova $q_{0.05}(5,20)$=4.23

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MSE}{n}} = 4.23\sqrt{\frac{8.06}{5}} = 5.37$$

- Per cui, ogni coppia di trattamenti che differisce in valore assoluto per un valore maggiore di 5.37 implica che le corrispondenti medie delle popolazioni sono significativamente differenti.

38

Confronto tra i diversi trattamenti – Test di Tukey – Esercizio

Esperimenti ad un singolo fattore

- Le medie dei cinque trattamenti sono:

$\bar{y}_{1\bullet} = 9.8 \quad \bar{y}_{2\bullet} = 15.4 \quad \bar{y}_{3\bullet} = 17.6 \quad \bar{y}_{4\bullet} = 21.6 \quad \bar{y}_{5\bullet} = 10.8$

- Da cui è possibile valutare quali sono le differenze significative:

$\bar{y}_{1\bullet} - \bar{y}_{2\bullet} = 9.8 - 15.4 = -5.6$ ⇐
$\bar{y}_{1\bullet} - \bar{y}_{3\bullet} = 9.8 - 17.6 = -7.8$ ⇐
$\bar{y}_{1\bullet} - \bar{y}_{4\bullet} = 9.8 - 21.6 = -11.8$ ⇐
$\bar{y}_{1\bullet} - \bar{y}_{5\bullet} = 9.8 - 15.4 = -1.0$
$\bar{y}_{2\bullet} - \bar{y}_{3\bullet} = 15.4 - 17.6 = -2.2$
$\bar{y}_{2\bullet} - \bar{y}_{4\bullet} = 15.4 - 21.6 = -6.2$ ⇐
$\bar{y}_{2\bullet} - \bar{y}_{5\bullet} = 15.4 - 10.8 = 4.6$
$\bar{y}_{3\bullet} - \bar{y}_{4\bullet} = 17.6 - 21.6 = -4.0$
$\bar{y}_{3\bullet} - \bar{y}_{5\bullet} = 17.6 - 10.8 = -6.2$ ⇐
$\bar{y}_{4\bullet} - \bar{y}_{5\bullet} = 21.6 - 10.8 = 10.8$ ⇐

39

Confronto tra i diversi trattamenti

Esperimenti ad un singolo fattore

- Il metodo di Tukey non è l'unico disponibile in letteratura per confrontare coppie di diversi trattamenti:

- Least Significant Difference (LSD), sviluppato da Fisher
- Metodo di Scheffé
- Altri …

40

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- Un ingegnere civile intende confrontare quattro distinti metodi di stima degli scarichi idrici quando sono applicati sullo stesso spartiacque

Metodo di stima	Osservazioni							
	1	2	3	4	5	6	Totale	Media
1	0.34	0.12	1.23	0.7	1.75	0.12	4.26	0.71
2	0.91	2.94	2.14	2.36	2.86	4.55	15.76	2.62666667
3	6.31	8.37	9.75	6.09	9.82	7.24	47.58	7.93
4	17.15	11.82	10.95	17.2	14.35	16.82	88.29	14.715
							155.89	6.49541667

41

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- Tabella ANOVA per l'esempio:

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0	P-value
Metodi	708.35	3	236.116	76.07	4.11e-11
Errore	62.081	20	3.1046		
Totale	770.73	23			

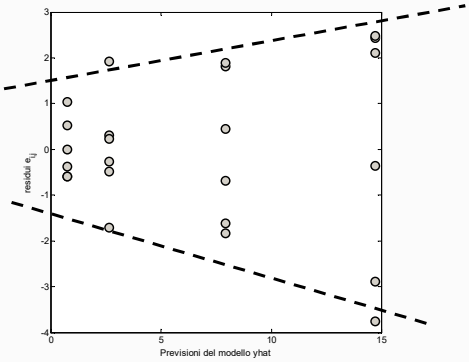
- È evidente che esiste un impatto del metodo: in genere $\mu_i \neq \mu_k$

42

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- Si può inoltre notare che la dispersione dei residui tenda a crescere con *y*



43

Verifica adeguatezza del modello – Analisi dei residui

Esperimenti ad un singolo fattore

- La presenza eventuale di una struttura nei dati può essere dovuta ad una varianza **non costante**.
- La varianza delle osservazioni può, per esempio, crescere con i valori assunti da *y*.
- Varianza non costante può essere indicativa di dati che seguono una distribuzione non-normale, di tipo **asimmetrico**
- In questi frangenti è possibile ricorrere a **trasformazioni non lineari dei dati** per avvicinare la dispersione dei dati ad una popolazione di tipo Gaussiano

44

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- **Trasformazione non lineare dei dati**
- Selezione **empirica** di una espressione **non lineare** per rendere **omogenea** la varianza nei trattamenti
- Si assume che la deviazione standard e la media siano legati da una legge di potenza
  $\sigma_y = \mu^{\alpha}$
- Per semplicità si considerano solo leggi di potenza per la trasformazione:
  $y^* = y^{\lambda}$
- Il che implica che sussiste una relazione del seguente tipo
  $\sigma_{y^*} \propto \mu^{\lambda+\alpha-1}$
- Ponendo $\lambda=1-\alpha$, la varianza dei dati trasformati è costante

45

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- **Trasformazione non lineare dei dati**
- Alcuni esempi di trasformazione:

Relazione tra σ_y e μ	α	$\lambda=1-\alpha$	Trasformazione
$\sigma_y \propto$ costante	0	1	No trasformazione
$\sigma_y \propto \mu^{1/2}$	½	½	Radice quadrata
$\sigma_y \propto \mu$	1	0	Logaritmica
$\sigma_y \propto \mu^{3/2}$	3/2	- ½	Reciproca della radice quadrata
$\sigma_y \propto \mu^2$	2	-1	Reciproca

46

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- Tornando all'esempio
- Si calcola la deviazione standard s_i e media per trattamento
- Se sussiste la seguente relazione per le popolazioni:

$$\sigma_{yi} = \theta \mu_i^{\alpha}$$

- ovvero

$$\log \sigma_{yi} = \log \theta + \alpha \log \mu_i$$

- Si apprezza comunque una dipendenza lineare tra i logaritmi delle deviazioni standard e delle medie per i trattamenti

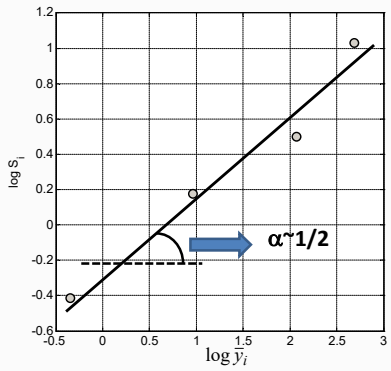
$$\log s_i = \log \theta + \alpha \log \bar{y}_i$$

47

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- Dal diagramma si evince una dipendenza lineare con pendenza ½



- Da cui:

$$\lambda = 1 - \alpha = \frac{1}{2}$$

- Si può applicare la seguente trasformazione non lineare per rendere omogenea la varianza nei dati

$$y^* = y^{\frac{1}{2}} = \sqrt{y}$$

48

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- Ripetendo la procedura con i dati trasformati

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0	P-value
Metodi	32.68	3	10.89	76.99	3.91e-11
Errore	2.688	20	0.1415		
Totale	35.37	23			

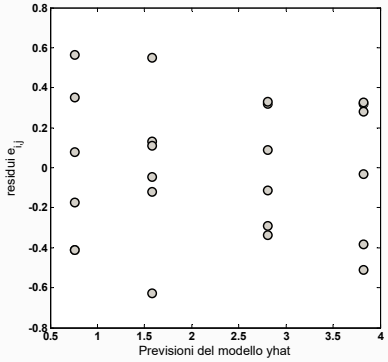
- Dal punto di vista qualitativo, le cose non cambiano in modo significativo

49

Verifica adeguatezza del modello – Analisi dei residui – Esempio (negativo)

Esperimenti ad un singolo fattore

- La dispersione dei residui risulta più omogenea al variare di y



50