Metodi statistici per l’analisi dei dati

Introduzione al Factorial Design

1

Factorial designs – Definizione

Factorial designs

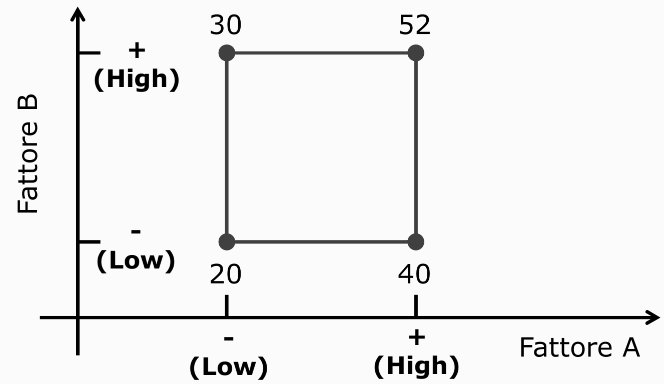
- Per “factorial design” si intende una campagna sperimentale in cui le misure sono eseguite **per tutte le possibili combinazioni** dei livelli dei fattori.
- Esempio:
  - 2 fattori A e B
  - a livelli (trattamenti) del fattore A
  - b livelli (trattamenti) del fattore B
    - Si eseguono a·b combinazioni dei trattamenti
- Quando i fattori sono combinati in questo modo, sono spesso definiti come **incrociati**.

2

Factorial designs – Definizione

Factorial designs

- **Esempio**: Si consideri il factorial design in figura
- Esperimento fattoriale con **due** fattori entrambi a **due** livelli



3

Factorial designs – Definizioni di base

Factorial designs

- Il disegno sperimentale preso in considerazione è il più semplice possibile.
  - i due distinti livelli sono definiti "**low**" e "**high**" e definiti anche con i simboli "**-**" e "**+**", rispettivamente.
- Si definisce **effetto principale del livello** la variazione nella risposta prodotta da una variazione del livello del fattore.
- Nel caso in esame si può calcolare come la differenza tra la risposta media al livello A e la risposta media al livello B.

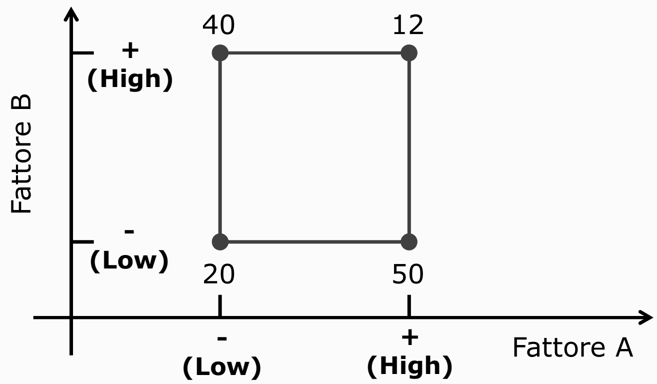
$$A = \frac{52+40}{2} - \frac{30+20}{2} = 21 \qquad B = \frac{52+30}{2} - \frac{20+40}{2} = 11$$

4

Factorial designs – Definizioni preliminari: presenza di interazioni

Factorial designs

- In alcuni casi la determinazione dell'effetto principale non è sufficiente. Si consideri ad esempio il seguente esperimento



5

Factorial designs – Definizioni preliminari: presenza di interazioni

Factorial designs

- È possibile calcolare, ancora gli effetti principali:

$$A = \frac{50+12}{2} - \frac{40+20}{2} = 1 \qquad B = \frac{40+12}{2} - \frac{20+50}{2} = -9$$

- ma, si può notare come:
  - a livello più basso di *B* (B^-), l'effetto di A sia

$$A\left(B^-\right) = 50 - 20 = 30 > 0$$

  - al livello più alto (B^+) si ha invece

$$A\left(B^+\right) = 12 - 40 = -28 < 0$$

- È chiara la presenza di **interazioni** tra i due fattori.

6

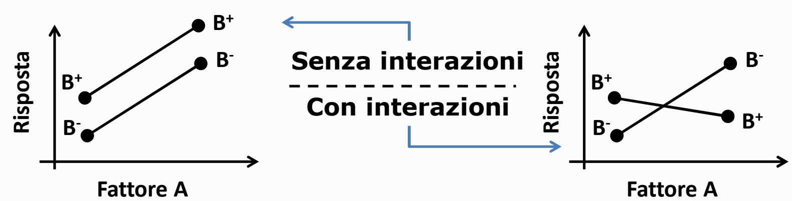
Factorial designs – Definizioni preliminari: presenza di interazioni

Factorial designs

- L'interazione presente può essere quantificata come la differenza **media** dei due effetti di A in corrispondenza di B^+ e B^-

$$AB = \frac{A(B^+) - A(B^-)}{2} = \frac{-28-30}{2} = -29$$

- La presenza di interazioni può essere apprezzata anche per via grafica:



Metodi statistici per l'analisi dei dati
10-14 febbraio 2020
7

7

Factorial designs – Definizioni preliminari: presenza di interazioni

Factorial designs

- Si può essere più dettagliati nel caso in cui i fattori siano **quantitativi** (es: temperatura, pressione, tempo etc.).
- In questo caso si può scrivere un modello di **regressione**:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

- dove:
  - y è la risposta misurata
  - x_1 e x_2 rappresentano i fattori
    - N.B. In genere, si suggerisce di lavorare con variabili **scalate**: $x_i \in [-1,1]$
  - β sono i parametri da determinare
  - ε è l'errore sperimentale

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020
8

8

Factorial designs – Definizioni preliminari: presenza di interazioni

Factorial designs

- Nel caso dell'esempio (estremamente **banale**: **due misure per due livelli**) la stima dei coefficienti è immediata:
  - I coefficienti β_i ($i \neq 0$) sono stimati come la metà dei valori dei corrispondenti effetti principali
  - Il coefficiente di interazione β_{12} corrisponde alla metà dell'effetto di interazione
  - L'intercetta β_0 è la media di tutte le risposte.

9

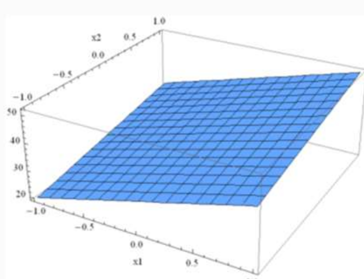
Factorial designs – Definizioni preliminari: presenza di interazioni

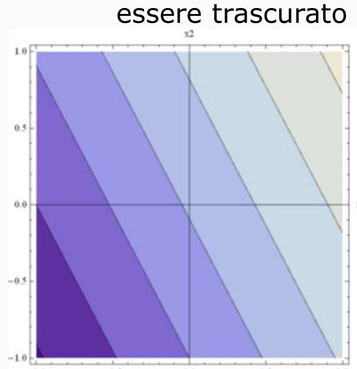
Factorial designs

- Caso 1: Stima modello di regressione

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + \cancel{0.5x_1x_2}$$

Il termine di interazione può essere trascurato





10

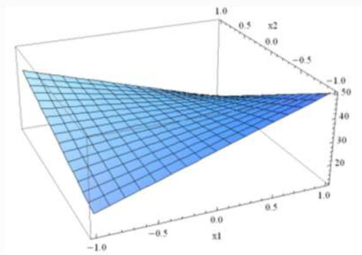
Factorial designs – Definizioni preliminari: presenza di interazioni

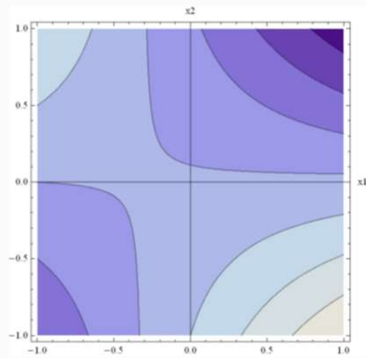
Factorial designs

- Caso 2: Stima modello di regressione

$$\hat{y} = 30.5 + 0.5x_1 - 4.5x_2 - 14.5x_1x_2$$

→ Il termine di interazione **non** può essere trascurato





La presenza di interazioni nei fattori porta ad una **curvatura** nel modello

11

Factorial design a due Fattori –Esempio

Factorial designs

- Un ricercatore sta studiando un nuovo enzima in grado di degradare un dato substrato. In particolare vuole investigare la resa al variare di
  - Temperatura di esercizio (tre diversi livelli: 15°C, 70°C, 125°C)
  - supporto specifico usato (tre diverse formulazioni)
- Per ogni combinazione delle condizioni sperimentali sono replicate 4 misure sperimentali

12

Factorial design a due Fattori –Esempio

Factorial designs

Tipo di supporto	Temperatura					
	15		70		125	
1	130	155	34	40	30	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

- **Domande:**
- Quali sono gli effetti della temperatura e del supporto sulla resa?
- Esiste una scelta ottimale del supporto che garantisca una resa uniforme rispetto alla temperatura?

13

Factorial design a due Fattori – Definizioni

Factorial designs

- Sia

  y_{ijk}

- la generica misura sperimentale in corrispondenza di:
  - livello i-esimo del fattore A con i=1,2, …,a
  - livello j-esimo del fattore B con i=1,2, …,b
  - replica k-esima della combinazione sperimentale con k=1,2,…,n
- Il numero di misure sperimentali totali è quindi $N=n \cdot a \cdot b$

14

Factorial design a due Fattori – Definizioni

Factorial designs

- **Modello statistico a effetti** per il disegno fattoriale a due fattori:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \\ k = 1,2,\ldots,n \end{cases}$$

Effetto della media totale	Effetto dell'i-esimo livello del fattore **A**	Effetto del j-esimo livello del fattore **B**	Effetto **interazione** tra τ_i e β_j	Componente errore casuale

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 15

15

Factorial design a due Fattori – Definizioni

Factorial designs

- Per avere una soluzione univoca, è necessario porre dei vincoli.
- Come già visto nel caso del singolo fattore:

$$\sum_{i=1}^{a} \tau_i = 0 \qquad \sum_{j=1}^{b} \beta_j = 0$$

- Anche le interazioni sono soggette a vincolo

$$\sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$$

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 16

16

Factorial design a due Fattori – Definizioni

Factorial designs

- L'ipotesi nulla da verificare è ancora sulla eguaglianza degli effetti dei trattamenti:
- Rispetto al fattore A:

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

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

- e rispetto al fattore B:

H_0: $\beta_1=\beta_2= \dots =\beta_b =0$

H_1: $\beta_j\neq 0$ per almeno un *j*

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020
17

17

Factorial design a due Fattori – Definizioni

Factorial designs

- Inoltre, si può essere interessati ad una eventuale interazione tra fattori:

H_0: $(\tau\beta)_{ij}=0$ per tutte le coppie (i,j)

H_1: $\tau_i\neq 0$ per almeno un *(i,j)*

- Tali ipotesi possono essere investigate con un **test analisi della varianza a due fattori**

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020
18

18

Analisi della varianza a due fattori – Procedura

Factorial designs

- Definizioni grandezze utili per la statistica 1/2

$$y_{i\bullet\bullet} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$$ → Totale osservazioni per il trattamento i-esimo Fattore A

$$y_{\bullet j\bullet} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}$$ → Totale osservazioni per il trattamento j-esimo Fattore B

$$y_{ij\bullet} = \sum_{k=1}^{n} y_{ijk}$$ → Totale osservazioni nella cella (i,j)

$$y_{\bullet\bullet\bullet} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$$ → Totale di tutte le osservazioni

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

19

19

Analisi della varianza a due fattori – Procedura

Factorial designs

- Definizioni grandezze utili per la statistica 2/2

$$\bar{y}_{i\bullet\bullet} = \frac{y_{i\bullet\bullet}}{bn}$$ → Media osservazioni i-esimo Fattore A

$$\bar{y}_{\bullet j\bullet} = \frac{y_{\bullet j\bullet}}{an}$$ → Media osservazioni j-esimo Fattore B

$$\bar{y}_{ij\bullet} = \frac{y_{ij\bullet}}{n}$$ → Media osservazioni cella (i,j)

$$\bar{y}_{\bullet\bullet\bullet} = \frac{y_{\bullet\bullet\bullet}}{abn}$$ → “Grande” Media per tutte le osservazioni

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

20

20

Analisi della varianza a due fattori – Procedura

Factorial designs

- La somma corretta dei quadrati può essere scritta come:

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

- Espandendo il secondo membro e verificando che i prodotti misti si annullano

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

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

21

21

Analisi della varianza a due fattori – Procedura

Factorial designs

- Interpretazione dei termini

Termine	Effetto	Gradi di libertà
$SSA = bn\sum_{i=1}^{a}\left(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}\right)^2$	A	a-1
$SSB = an\sum_{j=1}^{b}\left(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}\right)^2$	B	b-1
$SSAB = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}\right)^2$	Interazioni AB	(a-1)(b-1)
$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}_{ij\bullet}\right)^2$	Errore	ab(n-1)
$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}_{\bullet\bullet\bullet}\right)^2$	Totale	abn-1

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

22

22

Analisi della varianza a due fattori – Procedura

Factorial designs

- Tabella ANOVA per il caso di due fattori

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0
Trattamento A	SSA	a-1	MSA=SSA/(a-1)	*MSA/MSE*
Trattamento B	SSB	b-1	MSB=SSB/(b-1)	*MSB/MSE*
Interazione	SSAB	(a-1)(b-1)	MSAB=SSAB/(a-1)(b-1)	*MSAB/MSE*
Errore	SSE	ab(n-1)	MSE=SSE/(ab(n-1))	
Totale	SST	abn-1		

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 23

23

ANOVA a 2 fattori– Formule per il calcolo manuale

Factorial designs

- La procedura ANOVA è in genere eseguita con software specifici
- Nel caso si debba procedere al calcolo manuale può essere conveniente ricorrere ad alcune formule semplificative per le somme dei quadrati:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn}$$

$$SSA = \frac{1}{bn}\sum_{i=1}^{a} y_{i\bullet\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn} \qquad SSB = \frac{1}{an}\sum_{j=1}^{b} y_{\bullet j\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn}$$

$$SS_{Subtotal} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn} \qquad SSAB = SS_{Subtotal} - SSA - SSB$$

$$SSE = SST - SSAB - SSA - SSB = SST - SS_{Subtotal}$$

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 24

24

ANOVA a 2 fattori– Esercizio

Factorial designs

- Esempio di applicazione – sintesi enzimi

Tipo di supporto	Temperatura									
	15			70			125			$y_{i..}$
1	130	155	539	34	40	229	30	70	230	998
	74	180		80	75		82	58		
2	150	188	623	136	122	479	25	70	198	1300
	159	126		106	115		58	45		
3	138	110	576	174	120	583	96	104	342	1501
	168	160		150	139		82	60		
$y_{.j.}$			1738			1291			770	y...=3799

- I totali delle singole celle sono i numeri cerchiati

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

25

25

ANOVA a 2 fattori– Esercizio

Factorial designs

- Applicando le formule si perviene a

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn} = (130)^2 + (155)^2 + \ldots (60)^2 - \frac{(3799)^2}{36} = 77{,}646.97$$

$$SSA = \frac{1}{bn}\sum_{i=1}^{a} y_{i\bullet\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn} = \frac{1}{(3)(4)}\left[(998)^2 + (1300)^2 + (1501)^2\right] - \frac{(3799)^2}{36} = 10{,}683.72$$

$$SSB = \frac{1}{an}\sum_{j=1}^{b} y_{\bullet j\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn} = \frac{1}{(3)(4)}\left[(1738)^2 + (1291)^2 + (770)^2\right] - \frac{(3799)^2}{36} = 39{,}118.72$$

$$SSAB = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn} - SSA - SSB = 9613.78$$

$$SSE = SST - SSA - SSB - SSAB = 18{,}230.75$$

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

26

26

ANOVA a 2 fattori– Esercizio

Factorial designs

- Tabella per l'esempio

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0	P-value
Supporto	10,683.72	2	5,341.86	7.91	0.0020
Temperatura	39,118.72	2	19,559.36)	28.97	0.0001
Interazione	9,613.78	4	2,403.44	3.56	0.0186
Errore	18,230.75	27	675.21		
Totale	77,646.97	35			

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

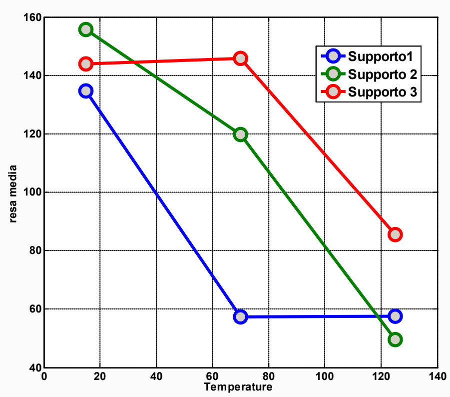
27

27

ANOVA a 2 fattori– Esercizio

Factorial designs

- Rappresentazione delle rese (medie) al variare dei fattori



- La **perdita di parallelismo** tra le rette è un indizio di **presenza di interazione**
- Resa migliore a basse temperature
- Su questo aspetto il supporto 3 risulta il supporto che garantisce rese migliori

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

28

28

ANOVA a 2 fattori– Esercizio Confronti multipli

Factorial designs

- Una volta concluso che esistono differenze significative tra le colonne e le righe, ha senso confrontare le **singole righe/colonne** per individuare le differenze specifiche.
- A tal proposito può essere usato il **test di Tukey**.
- **Da rimarcare che la presenza di interazioni non permette uno studio analogo al caso del singolo fattore.**
- Procedura possibile:
  – Si fissa uno specifico livello del fattore B e si applica il test alle medie del fattore A al dato livello.
  – Esempio: applicazione del test al solo livello 2 (T=70˚C)
  – Stima usata per la varianza: MSE (assunzione che la varianza dell'errore sia uniforme per tutte le combinazioni dei trattamenti)

ANOVA a 2 fattori– Esercizio Confronti multipli

Factorial designs

- Medie dei supporti a 70˚C

$$\bar{y}_{12\bullet} = 57.25 \quad \text{(supporto 1)}$$
$$\bar{y}_{22\bullet} = 119.75 \quad \text{(supporto 2)}$$
$$\bar{y}_{32\bullet} = 145.75 \quad \text{(supporto 3)}$$

- Inoltre, si può calcolare il valore critico:

$$T_{0.05} = q_{0.05}(3,27)\sqrt{\frac{MSE}{n}} = 3.50\sqrt{\frac{675.21}{4}} = 45.47$$

- dove $q_{0.05}(3,27)$=3.50 è disponibile da tabelle.

ANOVA a 2 fattori– Esercizio
Confronti multipli

Factorial designs

- Il confronto a coppie porta ai seguenti risultati:

$$3 \text{ vs. } 1 \quad 145.75 - 57.25 = 88.50 > T_{0.05} = 45.47$$
$$3 \text{ vs. } 2 \quad 145.75 - 119.75 = 26.00 < T_{0.05} = 45.47$$
$$2 \text{ vs. } 1 \quad 119.75 - 57.25 = 62.50 > T_{0.05} = 45.47$$

Si può concludere che il **supporto di tipo 1** differisce in maniera significativa dagli altri due ed ha una performance significativamente inferiore

31

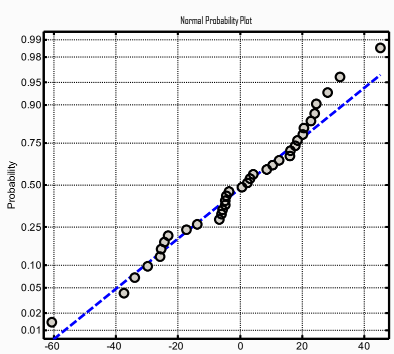
ANOVA a 2 fattori– Esercizio
Analisi dei residui

Factorial designs

- Per definizione:

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} = y_{ijk} - \bar{y}_{ij\bullet} \quad \left(\hat{y}_{ijk} \equiv \bar{y}_{ij\bullet}\right)$$

- **Rappresentazione su carta probabilistica**



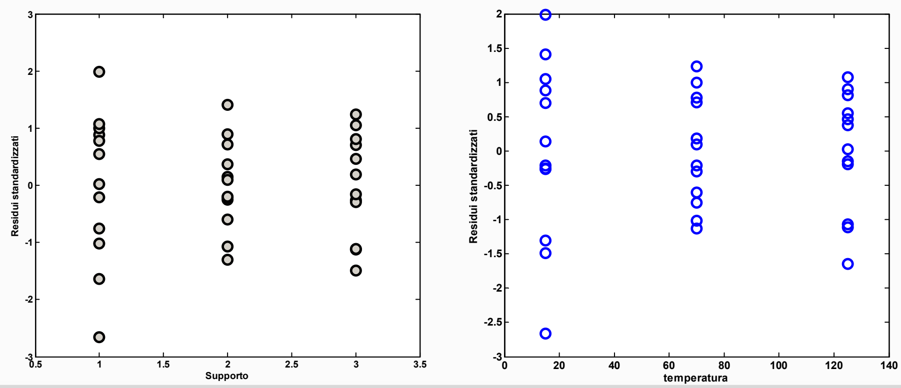
- Non si evidenziano particolari difetti nei dati
- Il residuo e_{ijk}~-60, (cui corrisponde un residuo standardizzato d_{ijk}~-2.34) nonostante sia molto grande in valore assoluto è comunque ancora ammissibile

32

ANOVA a 2 fattori– Esercizio Analisi dei residui

Factorial designs

- La combinazione **supporto 1** e **basse temperatura** presenta una maggiore varianza: tale combinazione *pare* produrre maggiore incertezza nella resa



33

Factorial designs – Stima dei parametri del modello

Factorial designs

- I parametri nel modello a effetti per un disegno sperimentale a due fattori

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

- possono essere stimati per mezzo del **metodo dei minimi quadrati**.

Modello a (**ab+b+a+1**) parametri → Richiede la scrittura di (**ab+b+a+1**) **equazioni normali**

34

Disegni fattoriali – Stima dei parametri del modello

Factorial designs

- Scrittura equazioni normali

1 equazione:

$$\mu: \quad abn\,\hat{\mu} + bn\sum_{i=1}^{a}\hat{\tau}_i + an\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\hat{i}_{\tau\beta}\right)_{ij} = y_{\bullet\bullet\bullet}$$

a equazioni:

$$\tau_i: \quad bn\,\hat{\mu} + bn\hat{\tau}_i + n\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{j=1}^{b}\left(\hat{i}_{\tau\beta}\right)_{ij} = y_{i\bullet\bullet} \quad i = 1,\dots,a$$

b equazioni:

$$\beta_j: \quad an\,\hat{\mu} + n\sum_{i=1}^{a}\hat{\tau}_i + an\hat{\beta}_j + n\sum_{i=1}^{a}\left(\hat{i}_{\tau\beta}\right)_{ij} = y_{\bullet j\bullet}, \qquad j = 1,\dots,b$$

ab equazioni:

$$(\tau\beta)_{ij}: \quad n\,\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_j + n\left(\hat{i}_{\tau\beta}\right)_{ij} = y_{ij\bullet}, \qquad \begin{cases} i = 1,2,\dots,a \\ j = 1,2,\dots,b \end{cases}$$

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

35

35

Factoriali designs – Stima dei parametri del modello

Factorial designs

- Ci sono (*a*+*b*+*1*) dipendenze lineari nel sistema di equazioni.
- Non esiste soluzione univoca a meno che non si fissano dei vincoli. Ad esempio:

$$\sum_{i=1}^{a}\hat{\tau}_i = 0 \qquad \textbf{1} \text{ vincolo}$$

$$\sum_{j=1}^{b}\hat{\beta}_j = 0 \qquad \textbf{1} \text{ vincolo}$$

$$\left.\begin{aligned} &\sum_{i=1}^{a}\left(\hat{i}_{\tau\beta}\right)_{ij} = 0 \quad j = 1,2,\dots,b \\ &\sum_{j=1}^{b}\left(\hat{i}_{\tau\beta}\right)_{ij} = 0 \quad i = 1,2,\dots,a \end{aligned}\right\} \; \boldsymbol{a+b-1} \text{ vincoli}$$

a+b+1 vincoli

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020

36

36

Factoriali designs – Stima dei parametri del modello

Factorial designs

- Applicando i vincoli le equazioni normali ammettono soluzione:

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

$$\hat{\tau}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}, \quad i = 1,2,\ldots,a$$

$$\hat{\beta}_j = \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}, \quad j = 1,2,\ldots,b$$

$$\left(\hat{i}_{\tau\beta}\right)_{ij} = \bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 37

37

Factoriali designs – Stima dei parametri del modello

Factorial designs

- Usando le equazioni precedenti è possibile trovare i **valori predetti dal modello**

$$\begin{aligned}\hat{y}_{ijk} &= \mu + \hat{\tau}_i + \hat{\beta}_j + \left(\hat{i}_{\tau\beta}\right)_{ij} \\ &= \bar{y}_{\bullet\bullet\bullet} + \left(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}\right) + \left(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}\right) + \left(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}\right) \\ &= \bar{y}_{ij\bullet}\end{aligned}$$

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 38

38

Assenza di interazioni in modelli a due fattori

Factorial designs

- In alcuni casi, è possibile descrivere i dati con un modello a due fattori senza interazioni.

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots b \\ k = 1,2,\ldots,n \end{cases}$$

- N.B. Tale modello richiede estrema cautela nella sua analisi, dato che come ampiamente discusso, le interazioni tra fattori possono essere importanti.
- Il modello senza interazioni ha una risoluzione immediata.

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 39

39

Assenza di interazioni in modelli a due fattori

Factorial designs

- Da notare che il vecchio termine di interazione è andato a sommarsi alla somma degli errori

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0	P-value
Supporto	10,683.72	2	5,341.86	5.95	0.0065
Trattamento B	39,118.72	2	19,559.36)	21.78	$2.33 \cdot 10^{-6}$
Errore	27,844.52	31	898.21		
Totale	77,646.97	35			

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 40

40

Assenza di interazioni in modelli a due fattori

Factorial designs

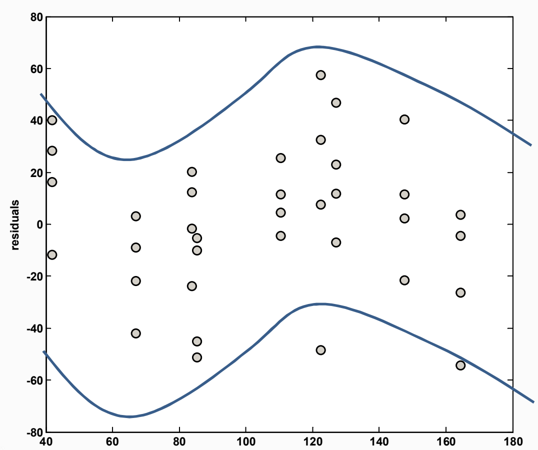
- Come precedentemente discusso, si nota che i due effetti principali sono significativi.
- Anche in questo caso si possono calcolare i residui.
- I valori previsti dal modello in questo caso sono dati dalla espressione:

$$
\begin{aligned}
\hat{y}_{ijk} &= \mu + \hat{\tau}_i + \hat{\beta}_j \\
&= \bar{y}_{\bullet\bullet\bullet} + \left(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}\right) + \left(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}\right) \\
&= \bar{y}_{i\bullet\bullet} + \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}
\end{aligned}
$$

Assenza di interazioni in modelli a due fattori

Factorial designs

- Analisi dei residui



- La presenza di una struttura nei residui è plausibilmente dovuta alla necessità del termine di interazione nel modello.

Modelli a due fattori con una osservazione per cella

Factorial designs

- Caso in cui si effettua una osservazione per combinazione delle condizioni sperimentali.
- La **varianza non è più stimabile** come nel caso generale, dato che non si può ricorrere alla espressione:

$$MSE = \frac{SSE}{ab(n-1)} = \frac{1}{ab(n-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij\bullet})^2$$

- È possibile comunque considerare un modello con la presenza dei soli effetti principali.

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

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 43

43

Modelli a due fattori con una osservazione per cella

Factorial designs

- In questo caso è possibile scrivere la tabella ANOVA

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0
Trattamento A	SSA	a-1	MSA=SSA/(a-1)	*MSA/MSE*
Trattamento B	SSB	b-1	MSB=SSB/(b-1)	*MSB/MSE*
Errore	SSE=SST-SSA-SSB	(a-1)(b-1)	MSE=SSE/((a-1)(b-1))	
Totale	SST	ab-1		

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 44

44

Modelli a due fattori con una osservazione per cella

Factorial designs

- Per stabilire se le interazioni sono presenti o meno, si può ricorrere ad un test sviluppato da Tukey (1949).
- La procedura assume che il termine di interazione sia di una forma particolarmente semplice:

$$(i_{\tau\beta})_{ij} = \gamma \tau_i \beta_j$$

- Si può considerare un test statistico che partiziona la somma dei quadrati dei residui in due termini
  - Una componente che descrive il contributo non additivo (termine di interazione) che viene formulato sulla base del tipo di interazione assunto (1 grado di libertà)
  - Il resto della somma dei residui

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 45

45

Modelli a due fattori con una osservazione per cella

Factorial designs

- Somma dei quadrati dovuta alla «non additività» (1 gdl):

$$SSN = \frac{1}{ab\, SSA\, SSB}\left[\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} y_{i\bullet} y_{\bullet j} - y_{\bullet\bullet}\left(SSA + SSB + \frac{y_{\bullet\bullet}^2}{ab}\right)\right]^2$$

- Somma degli errori ad *(a-1)(b-1)-1* gdl

$$SSE = SS_{\text{Residual}} - SSN \qquad SS_{\text{Residual}} = SST - SSA - SSB$$

- Per testare l'interazione si deve valutare la statistica:

$$F_0 = \frac{SSN/1}{SSE/(a-1)(b-1)-1}$$

Metodi statistici per l'analisi dei dati
10-14 febbraio 2020 46

46

Modelli a due fattori con una osservazione per cella – Esempio

Factorial designs

- Le impurità presenti in un prodotto chimico possono essere dovute a due fattori:
  - temperatura
  - pressione
- I dati provenienti da una campagna sperimentale con una sola replicazione sono riportati in tabella:

Temperatura	**Pressione**					
	25	30	35	40	45	$y_{i\cdot}$
100	5	4	6	3	5	**23**
125	3	1	4	2	3	**13**
150	1	1	3	1	2	**8**
$y_{\cdot j}$	**9**	**6**	**13**	**6**	**10**	**$y_{\cdot\cdot}$=44**

Modelli a due fattori con una osservazione per cella – Esempio

Factorial designs

- Le somme dei quadrati sono

$$SSA = \frac{1}{b}\sum_{i=1}^{a} y_{i\bullet}^2 - \frac{y_{\bullet\bullet}^2}{ab} = 23.33$$

$$SSB = \frac{1}{a}\sum_{j=1}^{b} y_{\bullet j}^2 - \frac{y_{\bullet\bullet}^2}{ab} = 11.60$$

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{ab} = 36.93$$

$$SS_{\text{Residuals}} = SST - SSA - SSB = 2.00$$

Modelli a due fattori con una osservazione per cella – Esempio

Factorial designs

- La somma dei quadrati **SSN** per la **non additività** è calcolata come da equazione:

$$SSN = \frac{1}{ab\,SSA\,SSB}\left[\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}y_{i\bullet}y_{\bullet j} - y_{\bullet\bullet}\left(SSA + SSB + \frac{y_{\bullet\bullet}^2}{ab}\right)\right]^2 = 0.0985$$

- Da cui è possibile ricavare la somma dei quadrati degli errori come:

$$SSE = SS_{\text{Residuals}} - SSN = 2.0 - 0.0985 = 1.9015$$

- Come si può vedere dalla tabella successiva, l'ipotesi di mancanza di interazioni sembra ragionevole.

49

Modelli a due fattori con una osservazione per cella – Esempio

Factorial designs

- I risultati possono essere sintetizzati nella tabella ANOVA

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Varianza	F_0	P-value
(A) Temperatura	SSA=23.33	a-1=2	MSA=SSA/(a-1) =11.67	*42.97*	*0.0001*
B (Pressione)	SSB=11.60	b-1=4	MSB=SSB/(b-1) =2.90	*10.68*	0.0042
«Interazione»	SSN=0.0985	1	MSN=SSN/1 =0.0985	0.36	0.5674
Errore	SSE = SS_{Res}-SSN=1.90	(a-1)(b-1)-1 =7	MSE=SSE/((a-1)(b-1)-1)=0.2716		
Totale	SST	ab-1=14			

50