CAPITOLO 4 - Controllo statistico della qualità dei prodotti e dei processi produttivi
Paragrafo 4.2 - Metodi off line e ANOVA

4

PEARSON

Paragrafo 4.2
Metodi off-line e analisi della varianza (ANOVA)

Argomenti

- Obiettivi del controllo off-line
- La programmazione degli esperimenti: concetti base
- ANOVA a una via
- Analisi post-hoc
- ANOVA a due vie

Miglioramento della qualità coi metodi off-line

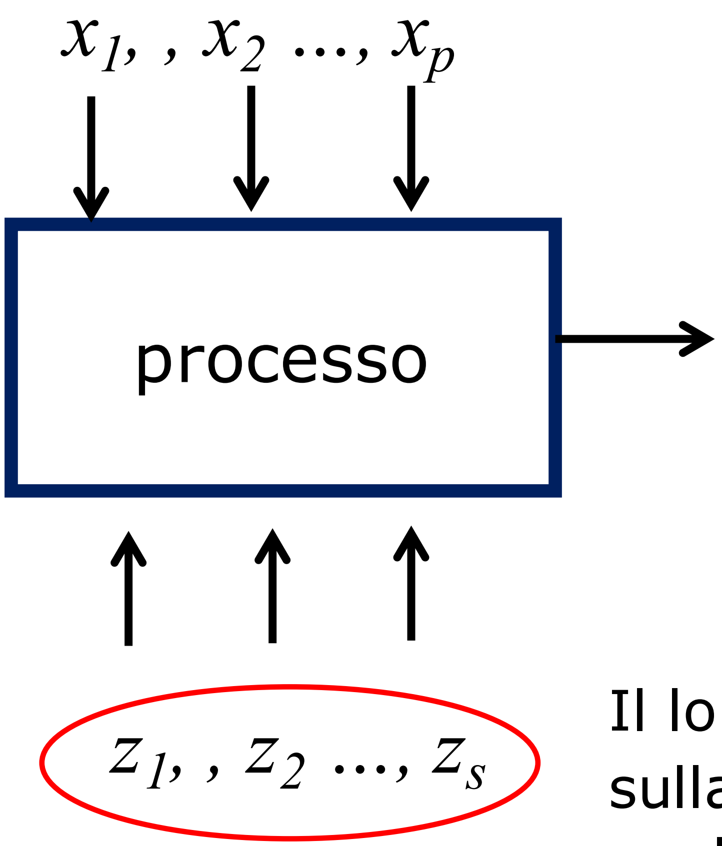
E’ importante trovare in **fase di progettazione** del processo le condizioni operative (tipo di materiali, taratura delle macchine, ecc.) che consentono di ottenere un **processo capace**.

Un modo per procedere è quello di effettuare degli **esperimenti programmati** per valutare **se** e in **quale misura** certi fattori del processo influenzano la caratteristica di qualità.

metodi off-line

Il processo produttivo

input controllabili



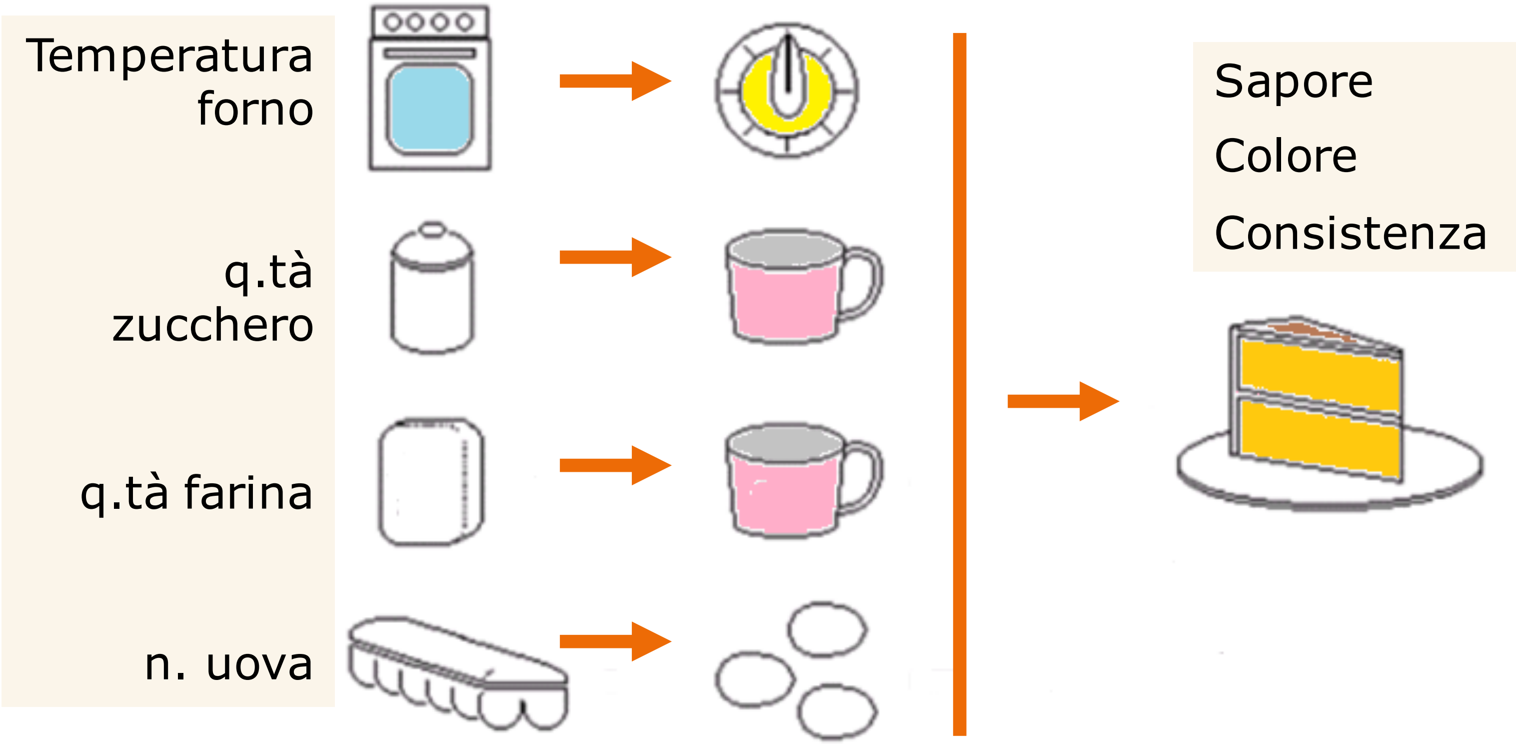
Caratteristica di qualità Y

input non controllabili

Il loro effetto complessivo sulla Y si assume essere quello di un disturbo accidentale

Esempio: processo di produzione di una torta

Fattori controllabili | Livelli | Variabile risposta (Y)



PEARSON

Obiettivo dell'esperimento

L'esperimento è una serie di prove in cui vengono **deliberatamente** fatti variare i **livelli** dei **fattori controllabili** di processo in modo da poter osservare le corrispondenti variazioni sulla **variabile risposta** (la misura di qualità).

L'esperimento dovrà rispondere ai seguenti quesiti:

1. quali fattori influenzano la Y?

2. quali livelli devono avere questi fattori in modo che la Y soddisfi il più possibile le specifiche ?

Esempio: produzione di torte

Temperatura	Nr. uova	N. tazze zucchero	N. tazze farina		Consistenza
180°C	3	3	3	→	Y_1
185°C	3	3	3	→	Y_2

In questo esperimento, si pongono fissi i livelli di **3 fattori** (n. uova, n. tazze di zucchero, n. tazze di farina) e si fa variare la **temperatura** esaminando come varia in corrispondenza la consistenza delle torte prodotte (**variabile risposta**).

Controllare i fattori (es. il livello della temperatura) significa che si posso tenere costanti o farli variare in modo controllato.

Esempio: produzione di torte

Temperatura	Nr. uova	N. tazze zucchero	N. tazze farina		Consistenza
180°C	3	3	3	→	Y_1
185°C	3	3	3	→	Y_2
180°C	3	3	2	→	Y_3
185°C	3	3	2	→	Y_4

In questo esperimento, si pongono fissi i livelli di 2 fattori (nr. uova, n. tazze di farina) e facciamo variare la **temperatura** e il **n. di tazze di zucchero**, esaminando come varia in corrispondenza la consistenza delle torte prodotte.

Concetti base della pianificazione sperimentale (1/3)

Unità sperimentale: unità (oggetto) su cui viene condotto l'esperimento o che viene prodotta nell'esperimento.

Prova dell'esperimento: consiste nel fissare i livelli dei fattori sperimentali, condurre l'esperimento e misurare la variabile risposta su ciascuna unità sperimentale.

Fattore sperimentale: fattore che viene fatto variare nell'esperimento (es. temperatura); è un carattere misurato su **scala nominale** (o trattato come tale).

Livello del fattore: es. temperatura bassa (180° C), temperatura alta (185° C).

Variabile risposta: è la caratteristica di qualità misurata su ogni unità sperimentale (variabile misurata su scala a intervallo).

Concetti base della pianificazione sperimentale (2/3)

Trattamento: qualifica la prova a cui si sottopone l'unità.

Con **1 solo** fattore sperimentale, si hanno tanti trattamenti quanti sono i *livelli* del fattore.
Esempio. Fattore *temperatura*; 2 livelli (2 trattamenti): 180° C, 185° C

Con **2** o **più fattori** sperimentali si hanno tanti trattamenti quanti sono le *combinazioni dei livelli* dei fattori.
Esempio. 2 fattori: *temperatura* (livelli: 180° C, 185° C) e *numero tazze di farina* (livelli: 2 tazze, 3 tazze).
4 trattamenti: (180° C, 2 tazze), (185° C, 2 tazze), (180° C, 3 tazze), (185° C, 3 tazze).

Concetti base della pianificazione sperimentale (3/3)

Replicazione: è la ripetizione dell'esperimento su più unità sperimentali per ogni livello del fattore. Il numero delle prove effettuate per ogni trattamento è detto **numero delle replicazioni**.
Si parla di **dati bilanciati** se si ha lo stesso numero di replicazioni per ogni trattamento.

Casualizzazione: consiste nell'assegnare le unità sperimentali ai diversi trattamenti in modo casuale.

Controllo locale: controllo di fattori intervenienti che non sono però di interesse dell'esperimento (fattori **sub-sperimentali**)

L'esperimento. Formalizzazione. Esempio peso dei palloni

Materiale			
L1 (Fornitore 1)	**L2 (Fornitore 2)**	**L3 (Fornitore 3)**	**L4 (Fornitore 4)**
$y_{11}, y_{12}, \dots, y_{1n}$	$y_{21}, y_{22}, \dots, y_{2n}$	$y_{31}, y_{32}, \dots, y_{3n}$	$y_{41}, y_{42}, \dots, y_{4n}$

Un solo fattore (tipo di materiale) con **4** livelli.

n: numero di replicazioni per ogni trattamento (dati bilanciati)

y_{ij}: peso del pallone j del trattamento i che è la determinazione della variabile casuale:

$$Y_{ij} \sim N(\mu_i;\sigma^2) \qquad j=1,\dots,n$$

PEARSON

L’ipotesi distributiva sulla variabile risposta

$Y_{ij} \sim N(\mu_i; \sigma^2) \qquad j=1,\ldots,n$

i: indica il trattamento, j: indica l’unità sottoposta al trattamento i

L’esperimento deve garantire che:

le variazioni nel peso dei palloni prodotti dallo **stesso trattamento**, sono dovute **solo** a fattori accidentali

Si assume varianza **costante** σ^2
(che non dipende dal trattamento: **omoschedasticità**).

Valutazione dell'effetto del fattore sperimentale mediante l'ANOVA (ANalysis Of VAriance)

$Y_{ij} \sim N(\mu_i; \sigma^2) \qquad i=1,\ldots,K; \quad j=1,\ldots,n$

i: indica il trattamento; j: indica l'unità sottoposta al trattamento i

L'ANOVA consiste nel condurre il seguente test

$$\begin{cases} H_0: \ \mu_1 = \mu_2 = \ldots = \mu_K = \mu \\ H_1: \text{ almeno una media è diversa dalle altre} \end{cases}$$

che equivale al test

$$\begin{cases} H_0: \ \mu_i - \mu = 0 \ \text{ per ogni } i=1,\ldots,K \\ H_1: \ \mu_i - \mu \neq 0 \ \text{ per almeno un } i \end{cases}$$

$$\boxed{\mu = \frac{\mu_1 + \mu_2 + \ldots + \mu_K}{K}}$$

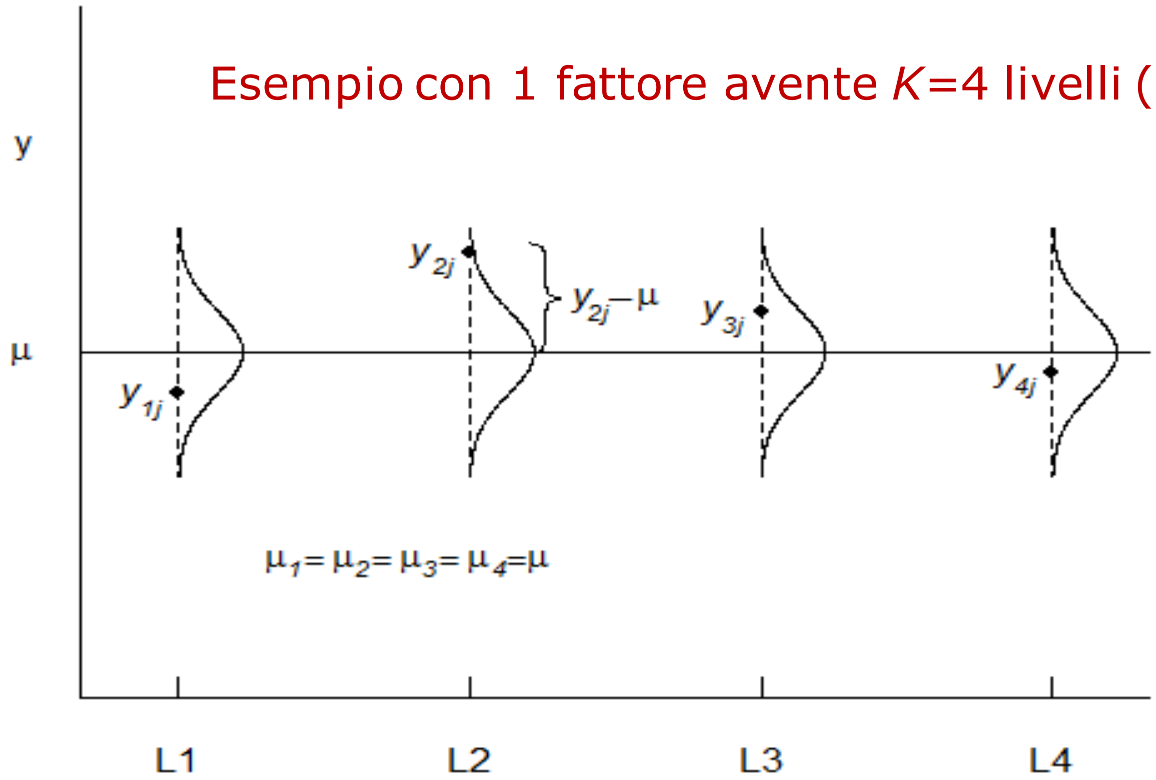
Situazione sotto l'ipotesi nulla $H_0: \mu_1 = \mu_2 = . . . = \mu_K = \mu$

$Y_{ij} \sim N(\mu; \sigma^2) \qquad i=1,\ldots,K; \quad j=1,\ldots,n$

$Y_{ij} = \mu + \varepsilon_{ij} \qquad\qquad \varepsilon_{ij} = Y_{ij} - \mu$

$Y_{ij} - \mu$: è dovuto solo a cause accidentali (errore sperimentale)

Esempio con 1 fattore avente K=4 livelli (trattamenti)



Ipotesi alternativa. Una possibile situazione.

$Y_{ij} \sim N(\mu_i; \sigma^2)$ $\qquad i=1,\ldots,K; \quad j=1,\ldots,n$

$Y_{ij} = \mu_i + \varepsilon_{ij} \quad \rightarrow \quad Y_{ij} - \mu = (\mu_i - \mu) + \varepsilon_{ij}$

Per almeno un i: $(Y_{ij} - \mu)$ è dovuto a cause accidentali (all'errore sperimentale) e **anche** all'effetto del livello i del fattore perché $(\mu_i - \mu) \neq 0$

Esempio con 1 fattore avente 4 livelli (trattamenti)

Riepilogo: la variabilità intorno a μ

H_0 vera

$$E(Y_{ij}-\mu)^2=\sigma^2$$

La variabilità intorno a μ è dovuta solo a cause accidentali

H_0 falsa (per almeno un livello i si ha: $\mu_i \neq \mu$)

$$E(Y_{ij}-\mu)^2=E(Y_{ij}-\mu_i+\mu_i-\mu_i)^2=E(Y_{ij}-\mu_i)^2+(\mu_i-\mu)^2=\sigma^2+(\mu_i-\mu)^2$$

La variabilità intorno a μ è dovuta a cause accidentali e anche all'effetto del livello i del fattore

Esperimento con 1 fattore (4 livelli e dati bilanciati)

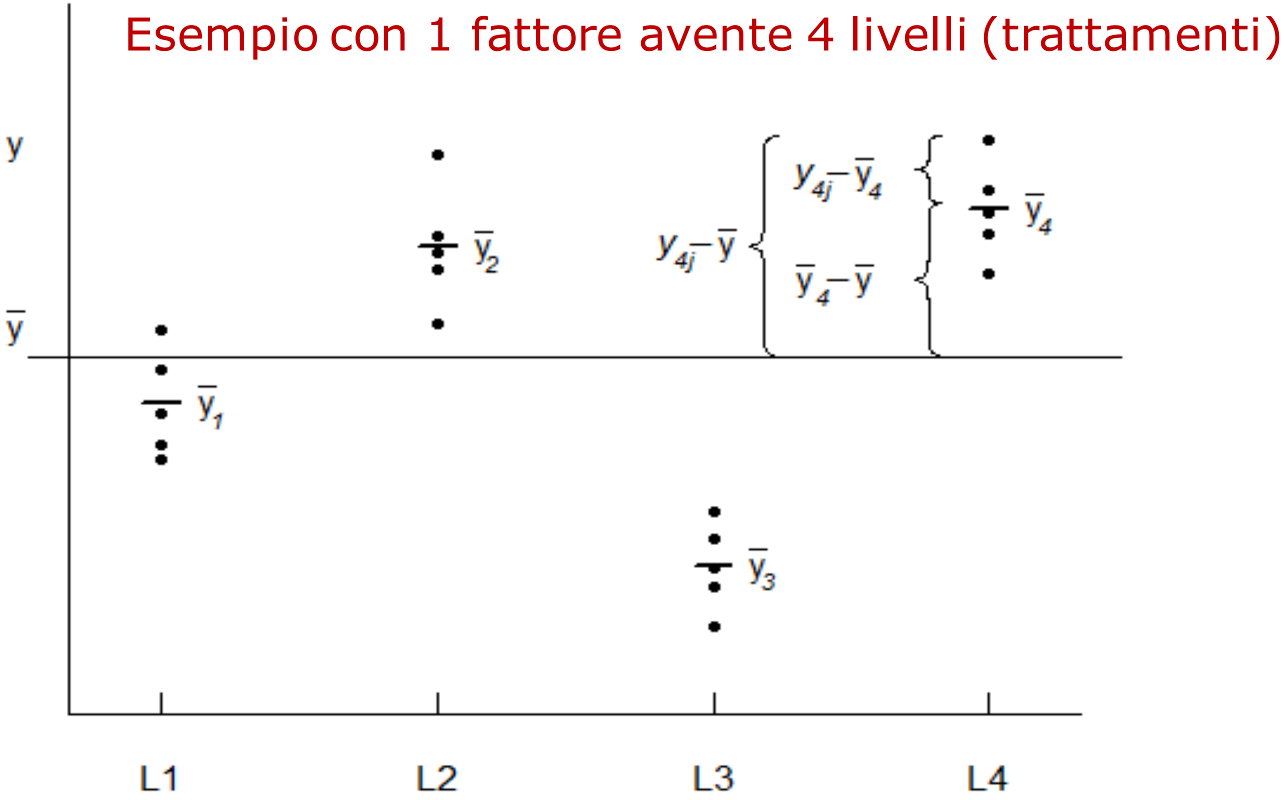
Materiale			
L1 (Fornitore 1)	L2 (Fornitore 2)	L3 (Fornitore 3)	L4 (Fornitore 4)
$y_{11}, y_{12}, \ldots, y_{1n}$	$y_{21}, y_{22}, \ldots, y_{2n}$	$y_{31}, y_{32}, \ldots, y_{3n}$	$y_{41}, y_{42}, \ldots, y_{4n}$

$$\bar{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$$ media per il trattamento i: stima di μ_i

$$\bar{y} = \frac{1}{K}\sum_{i=1}^{K} \bar{y}_i$$ media generale: stima di μ

I risultati dell'esperimento e la scomposizione dello scarto dalla media generale

$$(y_{ij} - \bar{y}) = (\bar{y}_i - \bar{y}) + r_{ij} \qquad r_{ij} = y_{ij} - \bar{y}_i$$



Interpretazione della scomposizione

$$y_{ij} - \bar{y} = \underbrace{(\bar{y}_i - \bar{y})}_{\text{}} + \underbrace{(y_{ij} - \bar{y}_i)}_{\text{}}$$

Contiene componente accidentale e, se H_0 è **falsa**, anche componente sistematica. Indica la variabilità **fra** le medie dei trattamenti *(between)*

Contiene solo la componente accidentale. Indica la variabilità **interna** ai gruppi *(within)* individuati dai trattamenti

Il procedimento del test confronta gli scostamenti *between* (che potrebbero contenere anche gli effetti sistematici) con quelli *within* che sono dovuti **solo** al caso.

La scomposizione della variabilità intorno alla media generale

Per i dati del trattamento i

$$\sum_{j=1}^{n}(y_{ij}-\bar{y})^2=\sum_{j=1}^{n}\left[(\bar{y}_i-\bar{y})+(y_{ij}-\bar{y}_i)\right]^2=n(\bar{y}_i-\bar{y})^2+\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2$$

Per tutti i dati

$$\underbrace{\sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij}-\bar{y})^2}_{sst}=\underbrace{\sum_{i=1}^{K}n(\bar{y}_i-\bar{y})^2}_{ssb}+\underbrace{\sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2}_{ssw}$$

$$sst = ssb + ssw$$

sum of squares **total**

sum of squares **between**

sum of squares **within**

I residui del modello e la stima della varianza σ^2

I residui del modello sono:

$$r_{ij} = y_{ij} - \bar{y}_i$$

La stima (da stimatore **non distorto** anche se **H_0 è falsa**) di σ^2

$$msw = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n} r_{ij}^{\ 2}}{Kn-K} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2}{Kn-K} = \frac{ssw}{Kn-K}$$

msw: *mean squares within*

$Kn-K$: gradi di libertà associati alla devianza within

La devianza between

Se **H_0 è vera**, possiamo ricavare un'altra stima di σ^2 da stimatore **non distorto**:

$$msb = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n} (\bar{y}_i - \bar{y})^2}{K-1} = \frac{\sum_{i=1}^{K} n(\bar{y}_i - \bar{y})^2}{K-1} = \frac{ssb}{K-1}$$

msb: *mean squares between*

$K-1$: gradi di libertà associati alla devianza between (fra gruppi o fra medie)

Il test dell'ANOVA

$$\begin{cases} H_0: \ \mu_1 = \mu_2 = \ldots = \mu_K = \mu \\ H_1: \text{almeno una media è diversa dalle altre} \end{cases}$$

Statistica test:

$$F_{oss} = \frac{msb}{msw} = \frac{ssb/(K-1)}{ssw/(Kn-K)}$$

Si rifiuta H_0 se $F_{oss} > F_\alpha$ dove:

$P(F_{(K-1),(Kn-K)} > F_\alpha) = \alpha$ (probabilità di rifiutare H_0 quando è vera)

$F_{(K-1),(Kn-K)}$ è la v.c. di Fisher con $(K-1),(Kn-K)$ gradi di libertà.

Tavola dell'ANOVA (ANOVA table)

Effetti	Devianza (sum of squares)	Gradi di libertà	Varianza (mean squares)	F di Fisher (F_{oss})	*p-level*
Trattamenti	*ssb*	K-1	*msb*	*msb/msw*	$P(F_{K-1,Kn-K}>F_{oss})$
Errore	*ssw*	$Kn-K$	*msw*		
Totale	*sst*	$Kn-1$			

stima di σ^2

NOTA BENE. Posto F_α tale che $P(F_{(K-1),(Kn-K)} > F_\alpha) = \alpha$,

si rifiuta H_0 se $F_{oss} > F_\alpha$

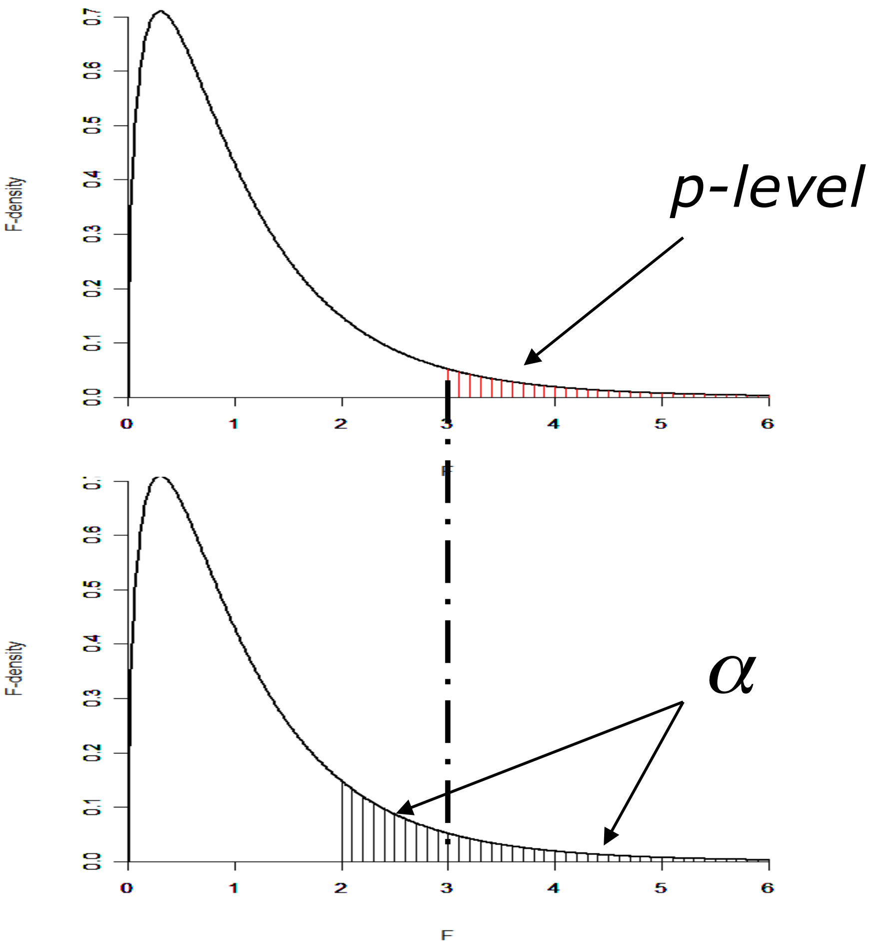
ovvero

si rifiuta H_0 se $p\text{-}level < \alpha$

è equivalente

PEARSON

Criterio dell'F_{oss} e del *p-level*



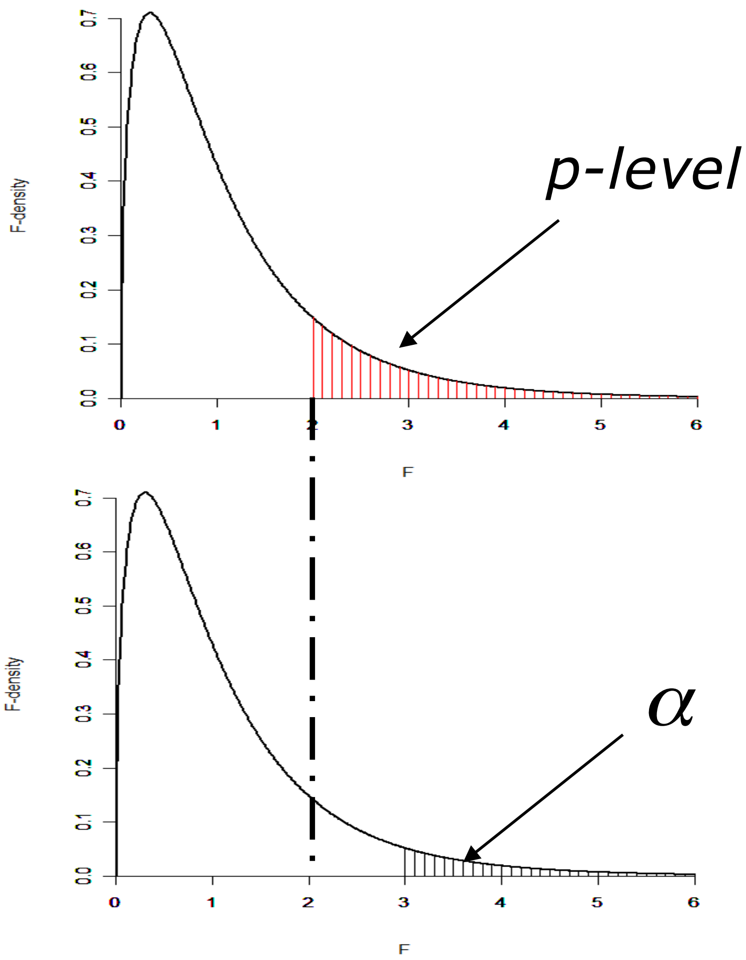
$F_{oss}=3, \quad F_{\alpha}=2$

$F_{oss}>F_{\alpha}$

$p\text{-}level<\alpha$

si rifiuta l'ipotesi nulla

Criterio dell'F_{oss} e del *p-level*



$F_{oss}=2, \quad F_{\alpha}=3$

$F_{oss}<F_{\alpha}$

$p\text{-}level>\alpha$

si accetta l'ipotesi nulla

PEARSON

I gradi di libertà (gdl)

$$sst = \sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij} - \bar{y})^2 \rightarrow \boxed{\begin{array}{l} gdl = Kn - 1 \\ Kn\text{: nr. totale unità} \end{array}}$$

$$ssb = \sum_{i=1}^{K}\sum_{j=1}^{n}(\bar{y}_i - \bar{y})^2 \rightarrow \boxed{gdl = K - 1}$$

$$ssw = \sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2 \rightarrow \boxed{gdl = Kn - K}$$

K: n. trattamenti, n: nr. replicazioni per trattamento

Un po' di numeri per capire: processo di produzione di lastre sottili di gres laminato

Variabile risposta: spessore (in mm.).

Fattore: temperatura di lavorazione

Livelli di fattore: Bassa: 1200° C, Media: 1220° C, Alta:1240° C

Replicazioni per livello: 4

Due esempi

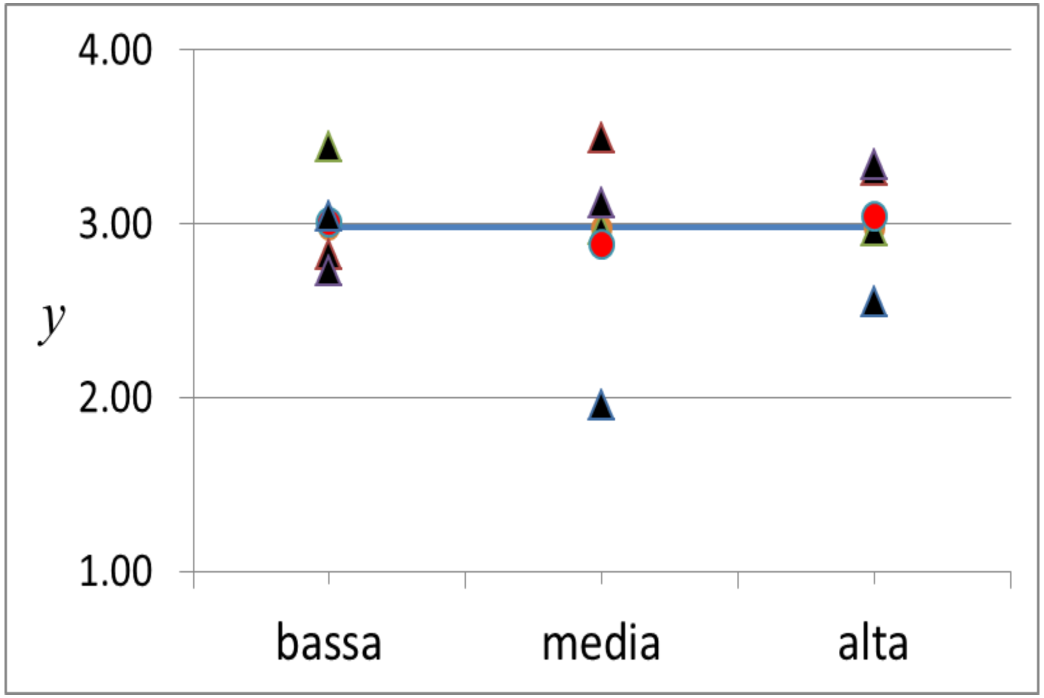
Esempio 1: il fattore non ha effetti sulla variabile risposta

Esempio 2: il fattore ha effetti sulla variabile risposta

Un po' di numeri per capire – Esempio 1

Temperatura		
bassa	media	alta
3.04	1.96	2.55
2.82	3.49	3.31
3.44	2.97	2.96
2.72	3.12	3.34
$\bar{y}_i$ **3.01**	**2.89**	**3.04**

4 replicazioni (prove) per ogni trattamento



Le medie osservate (rosso) non variano molto fra loro e non sono molto diverse dalla media generale (2.98, linea azzurra)

Esempio 1 – ANOVA e regola del *p-value*

α=0.05

Effetti	*SS*	*gdl*	*Varianza*	*F-value*	*p-value*
Trattamenti	0.0529	2	0.0264	0.1190	0.8892
Errore	1.9994	9	0.2222		
Totale	2.0523	11			

p-value$>\alpha$ → si accetta H_0 e si conclude che le medie sono uguali

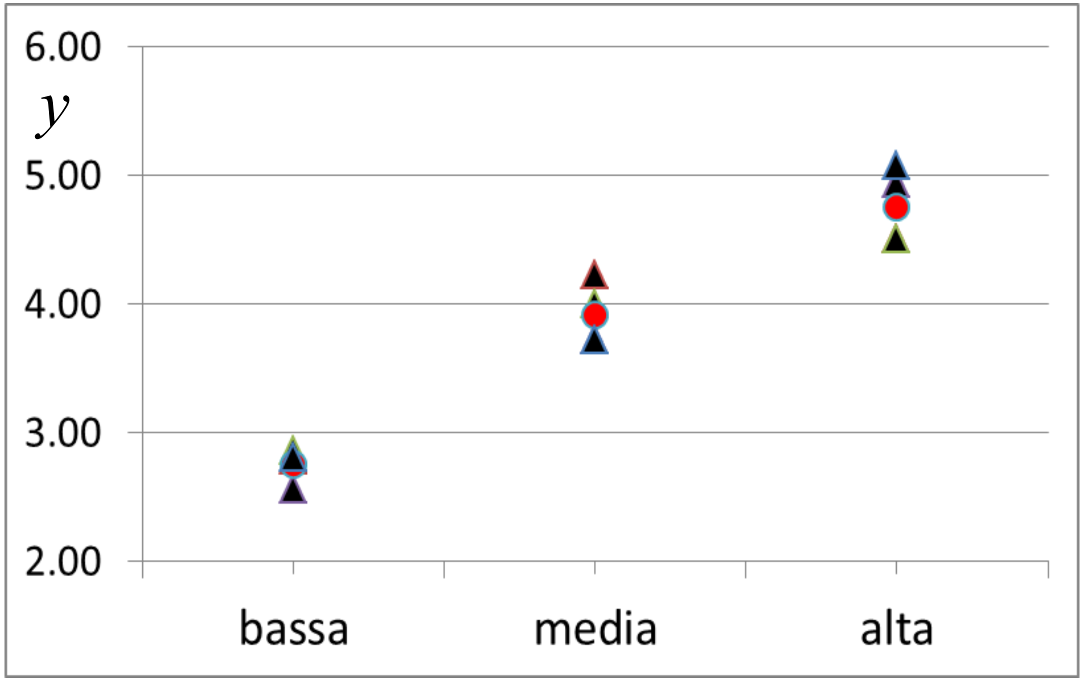
N.B. *F-value*: *F* osservato

Un po' di numeri per capire – Esempio 2

	Temperatura		
	bassa	media	alta
	2.81	3.73	5.08
	2.79	4.22	4.51
	2.86	4.01	4.51
	2.56	3.72	4.94
$\bar{y}_i$	**2.76**	**3.92**	**4.76**

4 replicazioni (prove) per ogni trattamento

Le medie osservate (rosso) appaiono alquanto diverse fra loro.



Esempio 2 – ANOVA e regola del *p-value*

α=0.05

Effetti	*SS*	*gdl*	*Varianza*	*F-value*	*p-value*
Trattamenti (between)	8.1105	2	4.0552	74.8966	2.5E-06
Errore (within)	0.4873	9	0.0541		
Totale	8.5978	11			

p-value $< \alpha$ → si rifiuta H_0 e si conclude che **almeno 2 medie** sono diverse fra loro ovvero che il fattore influenza lo spessore del gres.

ANOVA: tre puntualizzazioni

1) Questa tecnica è basata sulla scomposizione della devianza (e sul calcolo delle varianze between, interna e totale) della variabile risposta.

2) L'ipotesi nulla assume **indipendenza in media** ovvero al variare del dal livello del fattore sperimentale la media della variabile risposta non varia

3) Il fattore sperimentale è trattato come una variabile qualitativa.

Quali medie sono diverse fra loro ?

Quando rifiutiamo H_0 concludiamo che **almeno 2 medie sono diverse** (H_1) e **NON** che *tutte le medie sono diverse fra loro*.
E' necessario condurre l'analisi ***post-hoc*** (confronti multipli).

Analisi post-hoc*: confronto a coppie delle medie mediante* ***t-test*** per verificare l'uguaglianza delle medie a 2 a 2.

Con *K* livelli del fattore, abbiamo *K(K*–1*)*/2 coppie di medie da confrontare e quindi con *K*=3 abbiamo 3 *t-test* da condurre

$$\boxed{\begin{array}{l} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{array}} \qquad \boxed{\begin{array}{l} H_0: \mu_1 = \mu_3 \\ H_1: \mu_1 \neq \mu_3 \end{array}} \qquad \boxed{\begin{array}{l} H_0: \mu_2 = \mu_3 \\ H_1: \mu_2 \neq \mu_3 \end{array}}$$

Fattore: temperatura.
Livello 1: bassa ; Livello 2: media; Livello 3: alta

t-test e analisi post hoc

$$\begin{cases} H_0: \mu_i - \mu_h = 0 & i \neq h, \ i,h = 1,\ldots,K \\ H_1: \mu_i - \mu_h \neq 0 \end{cases}$$

Sotto H_0 si ha: $\bar{Y}_i - \bar{Y}_h \sim N\left(0; 2\frac{\sigma^2}{n}\right)$

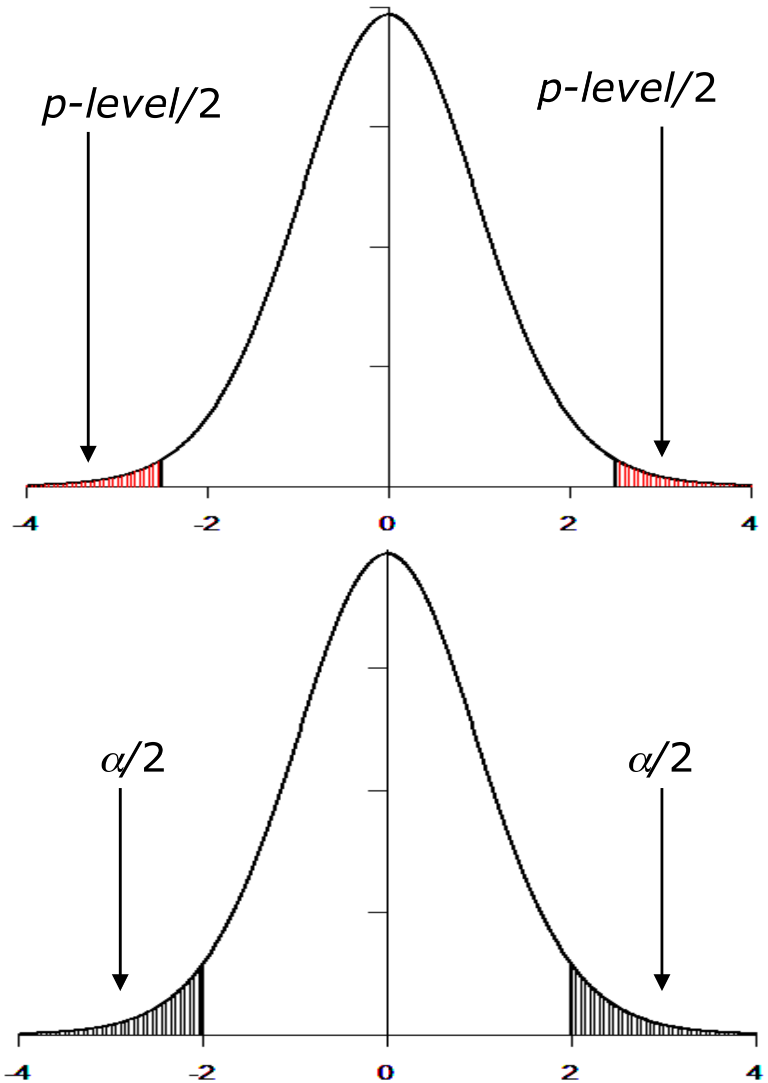
Si usa msw come stima di σ^2

Si usa il test *t*-Student per l'uguaglianza di due medie

$$t_{oss} \sim \frac{\bar{y}_i - \bar{y}_h}{\sqrt{\dfrac{2\,msw}{n}}}$$

sotto H_0 è la determinazone di una v.c t_{nK-K}

Criterio del t_{oss} e del *p-level*



t_{oss}=2.5 oppure t_{oss}=–2.5

$t_{\alpha/2}=2$

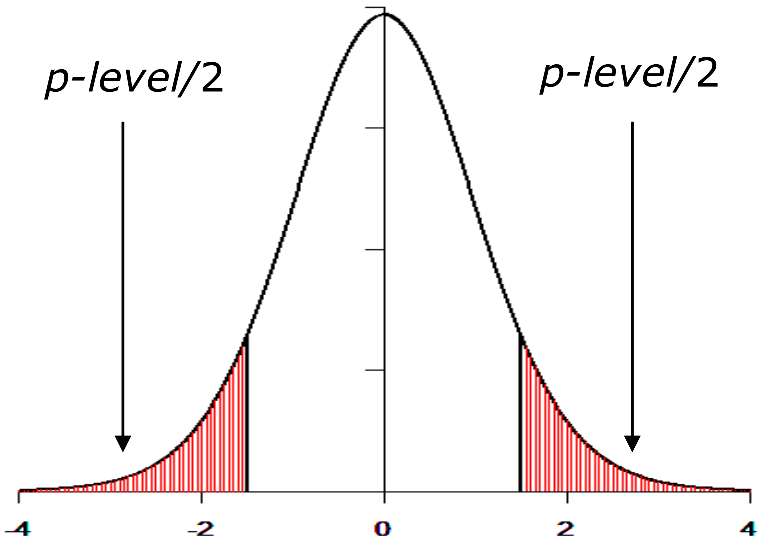
$|t_{oss}|>t_{\alpha/2}$

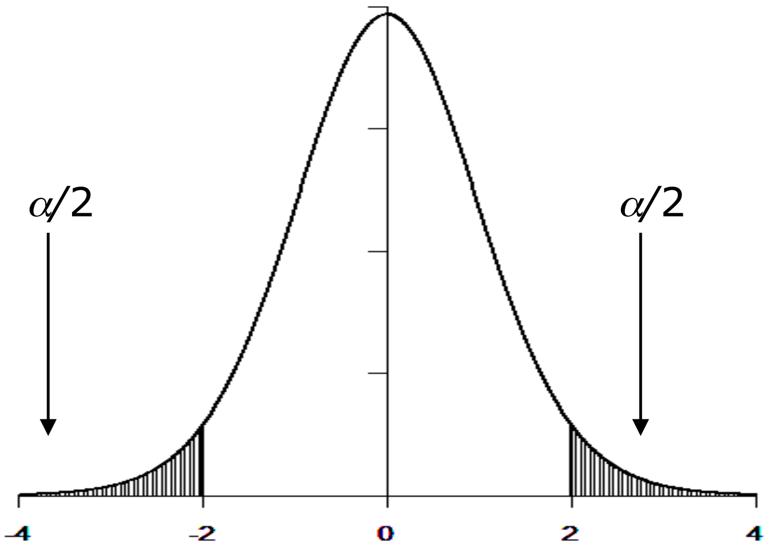
p-level$<\alpha$

si **rifiuta** l'ipotesi nulla

PEARSON

Criterio del t_{oss} e del *p-level*





t_{oss}=1.5 oppure t_{oss}=–1.5

$t_{\alpha/2}=2$

$|t_{oss}|<t_{\alpha/2}$

p-level$>\alpha$

si **accetta** l’ipotesi nulla

Dati dell'esempio 2 – Fattore: temperatura

	bassa	media	alta
$\bar{y}_i$	**2.76**	**3.92**	**4.76**

H_0	t-value	p-value
$\mu_1-\mu_2=0$	-7.0805	5.79E-05
$\mu_1-\mu_3=0$	-12.186	6.75E-07
$\mu_2-\mu_3=0$	-5.1053	6.41E-04

$\alpha=0.05$

$n=4$

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

$$t_{oss} = \frac{2.76-3.92}{\sqrt{0.0541\frac{2}{4}}} = -7.0805$$

I *p-value* sono tutti minori di α e quindi le medie sono tutte significativamente diverse fra loro.

Una precisazione:
t-test: il valore α nell'analisi post hoc

Con K livelli → $c=K(K-1)/2$ confronti simultanei di coppie di medie

Se si pone uguale ad α la probabilità di rifiutare H_0 (errore di I tipo) in **ognuno** dei c test, nel **complesso** dei c confronti, si ha una probabilità di commettere un errore del I tipo pari a:

$$1-(1-\alpha)^c > \alpha$$

Metodo Bonferroni

Posto α il valore globale *(familywise)* per i c test, per ciascuno di questi si usa il livello α/c.

Esempio: con $\alpha=0.05$ e $K=3$ livelli del fattore si ha $c=3$ deve usare per ogni test $\alpha/3=0.017$.

Fasi per condurre un'ANOVA con 1 fattore sperimentale

1. Effettuare il test F

1.1 Se l'ipotesi nulla (uguaglianza di tutte le medie) viene rifiutata→ analisi post-hoc

1.2 Se l'ipotesi nulla non viene rifiutata ci sono 2 possibilità:

i) le medie sono uguali fra loro (H_0 è vera)

ii) le medie sono diverse fra loro ma il test non è stato in grado di dimostrarlo.

1.2 Si accetta H_0: caso i)

Se si sospetta che *le medie siano uguali fra loro* (**caso i**), ci sono due possibilità.

a) Si ritiene che il fattore non abbia effetti sulla variabile risposta e quindi tale fattore non incide sulle prestazioni del processo;

b) Le medie associate ai livelli scelti nell'esperimento sono uguali fra loro ma **si ritiene che il fattore abbia effetti.**

In questo caso si conduce un **nuovo esperimento** scegliendo altri livelli del fattore.

1.2 Si accetta H_0: caso ii)

Se si sospetta che le medie associate ai livelli scelti del fattore sono diverse fra loro ma il test non è stato in grado di dimostrarlo (**caso ii**) ci due possibilità.

a) Il test F ha fornito un valore F_{oss} inferiore a F_{α} perché si è ottenuta una stima troppo elevata della varianza di errore (denominatore di F_{oss}) a causa di una bassa efficienza dello stimatore.

In questo caso si conduce un **nuovo esperimento** con un **numero di replicazioni** maggiore.

b) Il test F ha fornito un valore F_{oss} inferiore a F_{α} perché si è ottenuta una stima troppo elevata della varianza di errore (denominatore di F_{oss}) a causa dell'effetto di **fattori intervenienti** che non sono stati considerati o controllati nell'esperimento.

In questo caso si passerà all'**ANOVA e più vie**.

Competenze acquisite

- la logica del miglioramento di qualità con i metodi off line;
- linguaggio tecnico degli esperimenti;
- analisi della varianza (ANOVA)
- analisi post-hoc

ANALISI DELLA VARIANZA
esempio 2

Scostamenti da un target specificato (in ml) sotto 4 velocità di produzione

	VELOCITÀ DI PRODUZIONE (IN BOTTIGLIE PER MINUTO)			
	210	**240**	**270**	**300**
	−3.5	3.4	−1.4	4.3
	2.0	−2.1	3.2	3.3
	−4.8	0.6	−1.2	2.0
	−2.1	−4.5	2.7	−0.8
	−4.0	−1.6	0.9	2.5
Media	$\overline{X}_1 = -2.48$	$\overline{X}_2 = -0.84$	$\overline{X}_3 = 0.84$	$\overline{X}_4 = 2.26$
Varianza	$S_1^2 = 7.237$	$S_2^2 = 8.903$	$S_3^2 = 4.553$	$S_4^2 = 3.683$
Dev. st.	$S_1 = 2.69$	$S_2 = 2.98$	$S_3 = 2.13$	$S_4 = 1.92$

Nota: uno scostamento negativo dal target denota una bottiglia "troppo vuota", mentre uno scostamento positivo dal target una bottiglia riempita eccessivamente (gli scostamenti sono misurati in millilitri).

ANALISI DELLA VARIANZA
esempio 2

Analisi della varianza per il problema di imbottigliamento industriale

FONTE	GRADI DI LIBERTÀ	SOMMA DEI QUADRATI	MEDIA DEI QUADRATI (VARIANZA)	*F*	*p*-VALUE
Fra i gruppi (velocità)	4 − 1 = 3	63.286	21.095	3.46	0.041
Nei gruppi (velocità)	20 − 4 = 16	97.504	6.094		
Totale	20 − 1 = 19	160.790			

Al livello di significatività α=0,05 rifiuto l'ipotesi di uguaglianza delle medie a favore dell'ipotesi che non tutte le medie sono uguali (p-value = 0,041 < 0,05). Il valore della distribuzione F con 3 e 16 g.l. che lascia a destra una probabilità di 0,05 è F_u=3,24<3,46.