

Testo a cura di [Luca Biasibetti](#), *Mathematical Engineering*

---

## ANALISI DELLA VARIANZA (TEST ANOVA)

---

Nel presente articolo verrà trattato il tema dell'analisi della varianza. In particolare dopo una breve introduzione teorica verrà proposto un esempio concreto e una simulazione in Minitab del test ANOVA relativo ad un esperimento a singolo fattore (One-Way ANOVA).

L'analisi della varianza (ANOVA, Analysis of Variance) è una tecnica statistica di analisi dei dati che consente di verificare ipotesi relative a differenze tra le medie di due o più popolazioni. Permette quindi di studiare due o più gruppi di dati confrontandone sia le variabilità interne sia quelle tra i gruppi stessi.

Prima di effettuare un test ANOVA è necessario verificare che le ipotesi richieste dal modello siano effettivamente verificate dal data-set a disposizione.

### **Ipotesi:**

- i dati di ogni gruppo devono avere la stessa origine (nonché la stessa distribuzione stocastica) e le differenze osservate tra i gruppi devono essere casuali;
- le popolazioni considerate devono avere distribuzione Normale o quasi Normale (questa ipotesi è facilmente verificabile effettuando un Normal Probability Plot);
- le popolazioni confrontate nel test devono avere varianza omogenea.

Relativamente a quest'ultima ipotesi, poiché le varianze delle popolazioni non sono nella gran parte dei casi note a priori, tipicamente si effettuano degli ulteriori test di ipotesi sulle varianze. Per semplicità, molto spesso ci si limita ad utilizzare la seguente regola empirica: se il rapporto tra la deviazione standard campionaria dei dati più alta e quella più bassa è minore di 2, allora il test ANOVA può essere utilizzato.

Esistono diversi tipi di test ANOVA ed è possibile classificarli in base al numero di variabili indipendenti e dipendenti presenti nel modello oggetto di studio:

- modelli che prevedono una sola variabile indipendente: disegni a una via (One-Way ANOVA);
- modelli che prevedono due o più variabili indipendenti: disegni fattoriali;
- modelli che prevedono una sola variabile dipendente: analisi della varianza univariata;
- modelli che prevedono due o più variabili dipendenti: analisi della varianza multivariata (MANOVA - Multivariate Analysis of Variance).

Dopo questa breve premessa passiamo ora alla descrizione formale del modello a una via (One-Way ANOVA) e alla presentazione di un esempio applicativo.

## Descrizione formale del modello

Consideriamo  $k$  campioni indipendenti ottenuti da  $k$  diverse popolazioni con distribuzione Normale, con media  $\mu_k$  e con la stessa varianza (non nota a priori).

*Osservazioni*

Campione 1:  $y_{11}$  .....  $y_{1n}$

.....

Campione  $k$ :  $y_{k1}$  .....  $y_{knk}$

dove:

$y_{ij}$  rappresenta la  $j$ -esima osservazione nel campione estratto dalla popolazione  $i$ ;

$n_i$  rappresenta la numerosità del campione  $i$  (assumiamo  $n_1 + \dots + n_k = N$ ).

Il test ANOVA viene utilizzato per verificare le seguenti ipotesi:

**$H_0$ :  $\mu_1 = \mu_2 = \dots = \mu_k$**

**$H_1$ : almeno una delle  $\mu$  è diversa**

La statistica del test  $F$  è definita come il rapporto fra la misura di variabilità fra i gruppi  $MS_k$  e la misura di variabilità all'interno dei gruppi  $MS_e$ :

$$F = \frac{MS_k}{MS_e}$$

La distribuzione di questa statistica test, quando  $H_0$  è vera, è quindi una distribuzione  $F$  con gradi di libertà  $k-1$  e  $N-k$ .

## Un esempio pratico: One-Way ANOVA

*Descrizione dell'esperimento*

Un fabbricante di carta impiegata nei sacchetti per alimenti, è interessato a migliorare la resistenza a trazione dei suoi prodotti. Assumiamo che la resistenza a trazione sia funzione della concentrazione di legno di latifoglio presente nell'impasto con cui si realizza la carta.

Prendiamo in considerazione quattro livelli di concentrazione tra il 5% e il 20% (intervallo di interesse pratico definito dai tecnici addetti) e prepariamo sei provini per ciascun livello di concentrazione. Tutti i 24 provini vengono quindi sottoposti al test di resistenza di laboratorio in ordine casuale (la casualizzazione è molto importante in quanto bilancia l'effetto di ogni variabile di disturbo che può influenzare la resistenza alla trazione osservata).

Di seguito il data set rilevato.

CONCENTRAZIONE LEGNO (%)	Osservazione 1	Osservazione 2	Osservazione 3	Osservazione 4	Osservazione 5	Osservazione 6
5	7	8	15	11	9	10
10	12	17	13	18	19	15
15	14	18	19	17	16	18
20	19	25	22	23	18	20

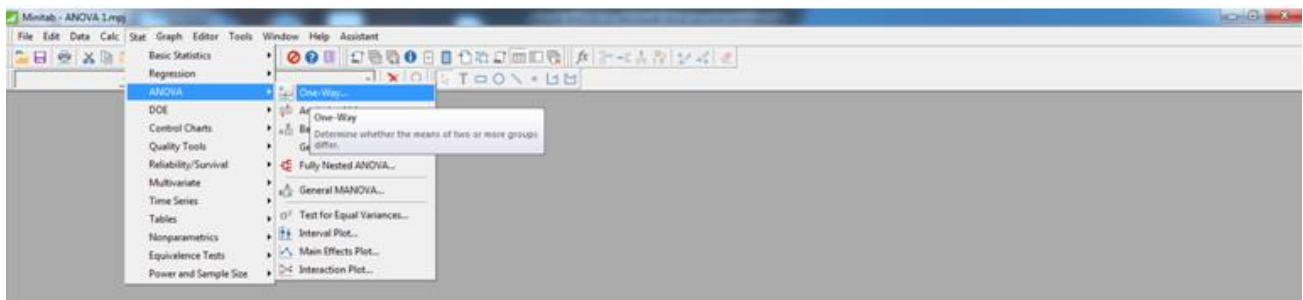
Questo appena esposto è un tipico esempio di esperimento completamente casualizzato a singolo fattore, con quattro livelli del singolo fattore (concentrazioni del legno).

## Analisi con Minitab

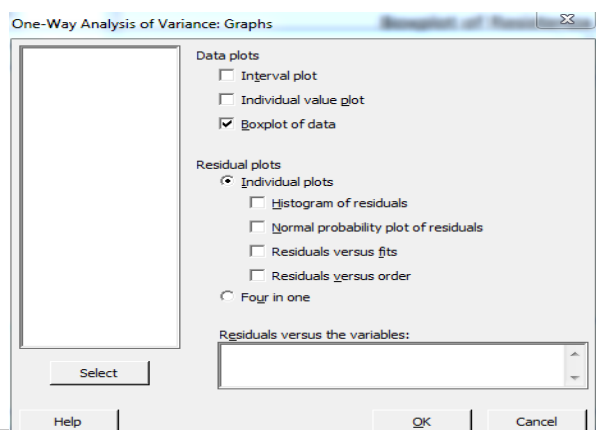
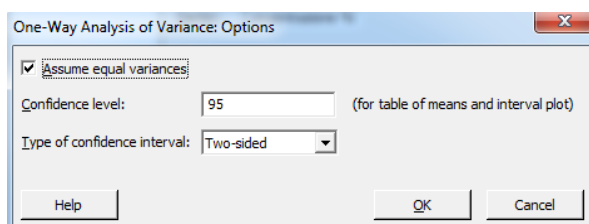
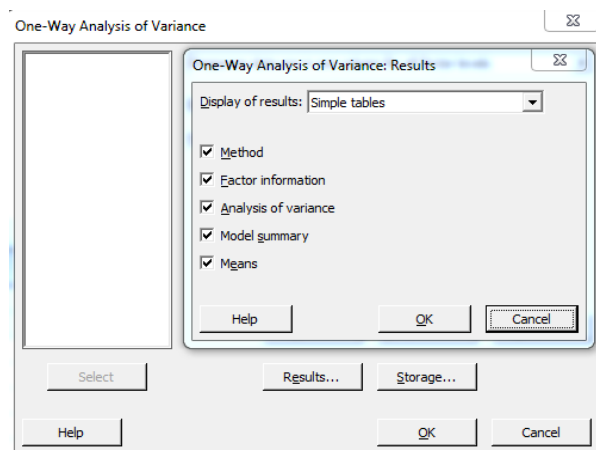
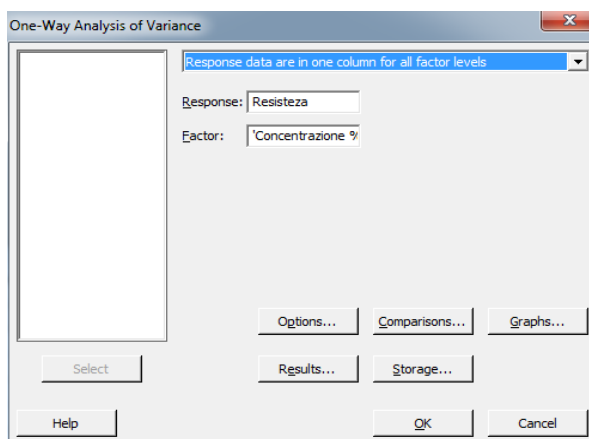
Vediamo ora come i dati ricavati da un esperimento casualizzato a singolo fattore possano essere analizzati statisticamente.

Inseriamo la tabella relativa al data set in Minitab riportando in due colonne separate la concentrazione percentuale del legno e la corrispettiva resistenza a trazione per ogni osservazione.

Selezioniamo il menù: *Stat – ANOVA - One-Way*.



Appare quindi la seguente finestra in cui vengono mostrate le variabili del modello (Response e Factor) e da cui è possibile scegliere e configurare l'output del test grafico e analitico.



Una volta configurato l'output, l'elaborazione del data set di Minitab fornirà la seguente analisi.

## One-way ANOVA: Resistenza versus Concentrazione %

Method

**Null hypothesis** All means are equal  
**Alternative hypothesis** At least one mean is different  
**Significance level**  $\alpha = 0,05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Concentrazione %	4	5; 10; 15; 20

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Concentrazione %	3	382,8	127,597	<b>19,61</b>	<b>0,000</b>
Error	20	130,2	6,508		
Total	23	513,0			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2,55114	74,62%	70,82%	63,46%

Means

Concentrazione %	N	Mean	StDev	95% CI
5	6	10,00	2,83	( 7,83; 12,17)
10	6	15,67	2,80	( 13,49; 17,84)
15	6	17,000	1,789	(14,827; 19,173)
20	6	21,17	2,64	( 18,99; 23,34)

Pooled StDev = 2,55114

## Analisi dei risultati

Come è possibile osservare, poiché il p-Value riportato nell'output è uguale a zero (quindi inferiore a un ipotetico livello di soglia scelto  $\alpha=0,05$ ) affermiamo che l'ipotesi nulla ( $H_0$ ) non è vera e concludiamo che la concentrazione di legno nell'impasto influisce in modo significativo sulla resistenza della carta.

Un ulteriore criterio di rifiuto dell'ipotesi nulla si ha quando la statistica del test F, presenta un valore maggiore del suo valore critico  $f_{\alpha,b,c}$  (punto percentuale superiore della distribuzione F).

In questo caso abbiamo  $f_{0,05,4,6} = 3,10$  (dato tabulato) ed  $F = 19,61$  (output ANOVA). Segue che  $F > f$  e quindi rifiutiamo  $H_0$ .

Notiamo inoltre che Minitab fornisce un'informazione riassuntiva su ciascun livello di concentrazione di legno: media dei dati, deviazione standard e intervallo di confidenza.

E' anche molto interessante analizzare graficamente i dati ricavati dall'esperimento.

Utilizzando ad esempio come output grafico il Boxplot della resistenza a trazione per i quattro livelli di concentrazione di legno di latifoglio, notiamo che la variazione di tale concentrazione influenza la resistenza alla trazione. Più precisamente, più alte concentrazioni di legno di latifoglio portano a osservare valori più elevati di resistenza a trazione.

Inoltre è evidente che la distribuzione della resistenza a trazione a un particolare livello di legno è ragionevolmente simmetrica e la variabilità della stessa non cambia drasticamente al variare della concentrazione di legno.

Il Boxplot mostra quindi sia la variabilità delle osservazioni entro un trattamento sia la variabilità tra i trattamenti effettuati.

