

Analisi della varianza (ANOVA) con R

L'analisi della varianza (ANOVA) è una tecnica utilizzata per confrontare le medie e le varianze di due o più gruppi, e per valutare se tali differenze siano statisticamente significative. L'Anova si utilizza quindi quando la variabile o le variabili indipendenti sono di tipo [categoriale](#), e la variabile dipendente è [cardinale](#).

R prevede diverse funzioni per condurre diversi tipi di ANOVA. Per una introduzione a queste tecniche, si rimanda alla relativa [voce di Wikipedia](#).

Test degli assunti

Ai fini della scelta delle tecniche da utilizzare, è importante ricordare gli **assunti** dell'Anova:

- ***normalità***: la distribuzione della variabile deve avere un andamento normale nei gruppi (è possibile effettuare il **[test di Shapiro-Wik](r:test_statistici:shapiro-test)**);
- ***omoschedasticità***: le varianze dei gruppi devono essere uguali (è possibile effettuare il **[test F delle varianze](r:test_statistici:var-test)**);
- ***sfericità*** (solo per l'Anova **entro casi**): le covarianze dei gruppi sono omogenee.

Vedi anche, su Wikipedia, [Analisi della varianza](#).

Inoltre, quando si procede ad una **Anova fattoriale**, vale anche il requisito:

- I gruppi devono avere la stessa numerosità (circa).

Per soddisfare questo requisito, è sufficiente selezionare in maniera casuale un sotto-campione, oppure ricorrere alla [Ponderazione](#) dei casi.

Anova non parametrica

Quando uno o entrambi di questi requisiti non sono rispettati, si devono utilizzare [tecniche non parametriche](#) di analisi della varianza [(vedi anche: http://www.creative-wisdom.com/teaching/WBI/parametric_test.shtml)].

Ci sono pacchetti dedicati alle diverse tecniche di analisi della varianza non parametrica. Vedi in particolare [il pacchetto WRS2](#)

L'Anova in R

- **//One-way Anova//**, o **[[Anova_one_way|Anova a una via]]**: una variabile

indipendente (se le modalità sono due, si usa anche il t-test);

- **//Two-way Anova//**, o **[[r:analisi_multivariata:anova_two_way_factorial|Anova a due vie]]**: due variabili indipendenti, con o senza analisi degli effetti di interazione;
- **//Factorial Anova//**, o **[[r:analisi_multivariata:anova_two_way_factorial|Anova fattoriale]]**: quando si vogliono studiare gli effetti di interazione fra due o più variabili indipendenti;
- **Anova multivariata (MANOVA)**: quando le variabili dipendenti sono più di una.

Distinguiamo inoltre fra:

- *Anova tra casi*: quando i casi sono indipendenti fra di loro;
- *Anova entro casi* o *Repeated Measures*: quando i casi non sono indipendenti fra di loro (ad esempio quando si esegue un test prima-dopo, e dunque i gruppi sono composti dagli stessi casi).

In generale, è possibile utilizzare la funzione `lm()` o la funzione `aov()`, che consente anche l'analisi a due o più vie.

Interpretazione della tabella Anova

È possibile ottenere una tabella dell'analisi della varianza (o della devianza) con la funzione `anova()`, applicata ad un modello di analisi lineare (`lm` e `glm`, ad esempio):

```
anova(lm.res)
```

Poniamo ad esempio che "y" sia una variabile numerica, e "A" una variabile categoriale. Otterremmo una tabella così composta:

Analysis of Variance Table

Response: mydata\$y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
mydata\$A	1	0.23	0.2335	0.1144	0.7353
Residuals	627	1279.93	2.0414		

Dove:

- `mydata$A` = variabile esplicativa
- `Df` = gradi di libertà
- `Sum Sq` = devianza (`mydata$A` = entro gruppi, o spiegata, o sistematica; `Residuals` = residua, o non spiegata, o stocastica)
- `Mean Sq` = varianza (`Sum Sq / Df`)
- `F value` e `Pr(>F)` = test F: varianza spiegata / varianza residua, e significatività statistica

Ai fini dell'interpretazione, si deve ricordare che l'ipotesi nulla è che le varianze siano uguali

fra di loro, e che dunque la variabile indipendente *non* produca effetti su quella dipendente. La probabilità che sia vera l'ipotesi nulla è indicata dal valore Pr (altrove, p). Nel caso in esempio, la relazione non è significativa ed anzi le due variabili sono quasi perfettamente indipendenti, in quanto $Pr = 0.7353$: c'è il 73,5% di probabilità che sia vera l'ipotesi nulla.

Costruzione del modello e formula

Vedi:

- [formula](#)
- [Anova a una via](#)
- [Anova a due vie e fattoriale](#)
- [Contrasti](#)

Tests Post-hoc (parametrici)

- [test di Tukey](#)
- [pairwise t-test](#)
- test di Scheffé sui [contrasti](#), vedi pacchetto DescTools, [funzione ScheffeTest\(\)](#)
- su Statistica con R, si trovano le funzioni per eseguire [test di Bonferroni e di Scheffé](#)
- pacchetto DescTools, [funzione PostHocTest\(\)](#), consente di eseguire diversi test post-hoc

Con RCommander

Con RCommander è possibile svolgere da menu:

- i tests per i controlli degli assunti;
- l'Anova a due vie, con i tests post-hoc (confronto di medie a coppie);
- l'Anova a più vie , senza tests post-hoc;

Non è invece possibile (da menu):

- utilizzare dati ponderati;
- specificare il modello fattoriale.

Per approfondire

Le procedure per eseguire l'Analisi della varianza in R sono disponibili su Wikipedia all'indirizzo: [W Analisi della varianza](#)

Sul blog [Statistica con R](#), sono inoltre disponibili i seguenti posts di approfondimento (in [Anova](#)):

- [Anova a due vie](#)
- [Confronto fra più gruppi, metodo parametrico](#)
- [Confronto fra più gruppi, metodo non parametrico](#)
- [Quando le assunzioni dell'Anova sono violate:tests e post-hoc tests](#), per i metodi non parametrici.
- [Post-hoc tests \(1\)](#) e [Post-hoc tests\(2\)](#)

[Analisi bivariata](#), [Analisi multivariata](#), [Test statistici](#), [Regressione lineare](#)

From:

<https://www.agnesevardanega.eu/wiki/> - **Ricerca Sociale con R**

Permanent link:

<https://www.agnesevardanega.eu/wiki/r/modelli/anova?rev=1743612470>

Last update: **02/04/2025 16:47**

