**Appunti lab MRP 17 dicembre 2013**

L’analisi della varianza

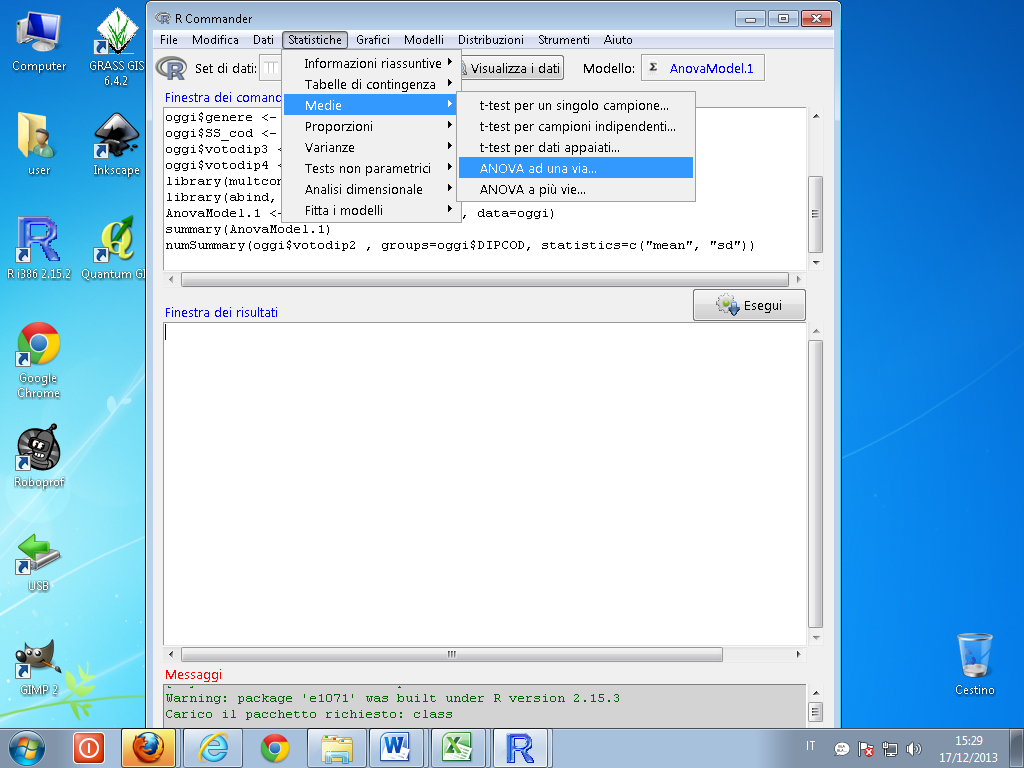
Si applica per verificare se ci sono differenze nei valori medi di una variabile dipendente tra gruppi diversi quando ci sono più di due gruppi. La situazione è molto simile a quella del t-test per campioni indipendenti, con la differenza che quest’ultimo si applica quando i gruppi sono solo due.

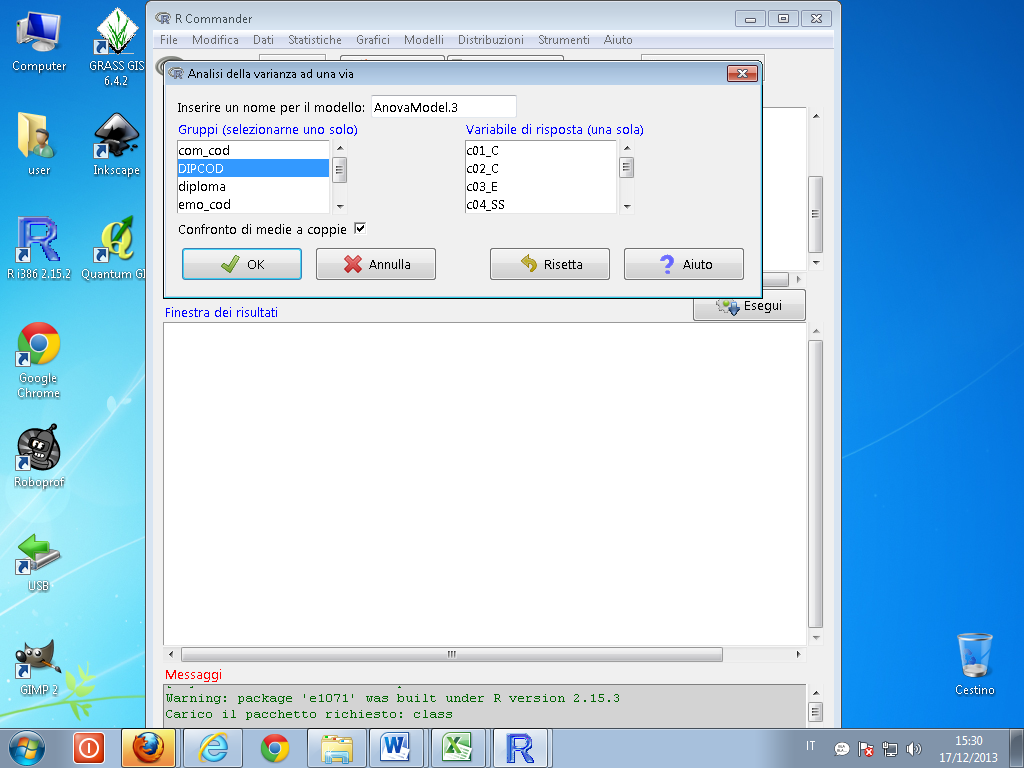
ESEMPIO

Voglio verificare se esiste una relazione tra Tipo di diploma e Voto di diploma

Poiché la variabile “TIPO DI DIPLOMA” (DIPCOD) comprende più di due livelli, applico l’ANOVA a una via.

L’ipotesi nulla è che le medie non siano diverse, l’ipotesi alternativa è che almeno una media sia diversa da almeno un’altra.





Risultati:

> AnovaModel.3 <- aov(votodip2 ~ DIPCOD, data=oggi)

> summary(AnovaModel.3)

Df Sum Sq Mean Sq **F value** **Pr(>F)**

DIPCOD 5 0.1607 0.03213 **3.131** **0.0103** \*

Residuals 142 1.4571 0.01026

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

2 observations deleted due to missingness

**RIFIUTO L’IPOTESI NULLA perché il valore p è < 0,05**

**Significa che almeno una delle medie considerate è deversa da almeno una delle altre.**

**In questo caso il valore della statistica calcolata è F=3,131**

**Per vedere quale, considero i confronti a coppie effettuati dopo l’ANOVA**

> numSummary(oggi$votodip2 , groups=oggi$DIPCOD, statistics=c("mean", "sd"))

mean sd data:n data:NA

ALTRO 0.7527803 0.10836054 15 0

COM 0.7122222 0.06591240 9 0

IT 0.7808709 0.09694975 37 0

LIC 0.7152778 0.08351171 36 1

SOC 0.7328846 0.10607708 26 1

SPP 0.8000000 0.12816006 25 0

> .Pairs <- glht(AnovaModel.3, linfct = mcp(DIPCOD = "Tukey"))

> summary(.Pairs) # **pairwise tests**

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = votodip2 ~ DIPCOD, data = oggi)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

COM - ALTRO == 0 -0.040558 0.042711 -0.950 0.9299

IT - ALTRO == 0 0.028091 0.031007 0.906 0.9421

LIC - ALTRO == 0 -0.037503 0.031131 -1.205 0.8281

SOC - ALTRO == 0 -0.019896 0.032845 -0.606 0.9900

SPP - ALTRO == 0 0.047220 0.033084 1.427 0.7022

IT - COM == 0 0.068649 0.037650 1.823 0.4442

LIC - COM == 0 0.003056 0.037752 0.081 1.0000

SOC - COM == 0 0.020662 0.039177 0.527 0.9947

SPP - COM == 0 0.087778 0.039378 2.229 0.2234

LIC - IT == 0 -0.065593 0.023714 -2.766 0.0658 .

SOC - IT == 0 -0.047986 0.025923 -1.851 0.4267

SPP - IT == 0 0.019129 0.026226 0.729 0.9770

SOC - LIC == 0 0.017607 0.026071 0.675 0.9836

**SPP - LIC == 0 0.084722 0.026372 3.213 0.0188 \***

SPP - SOC == 0 0.067115 0.028375 2.365 0.1694

---

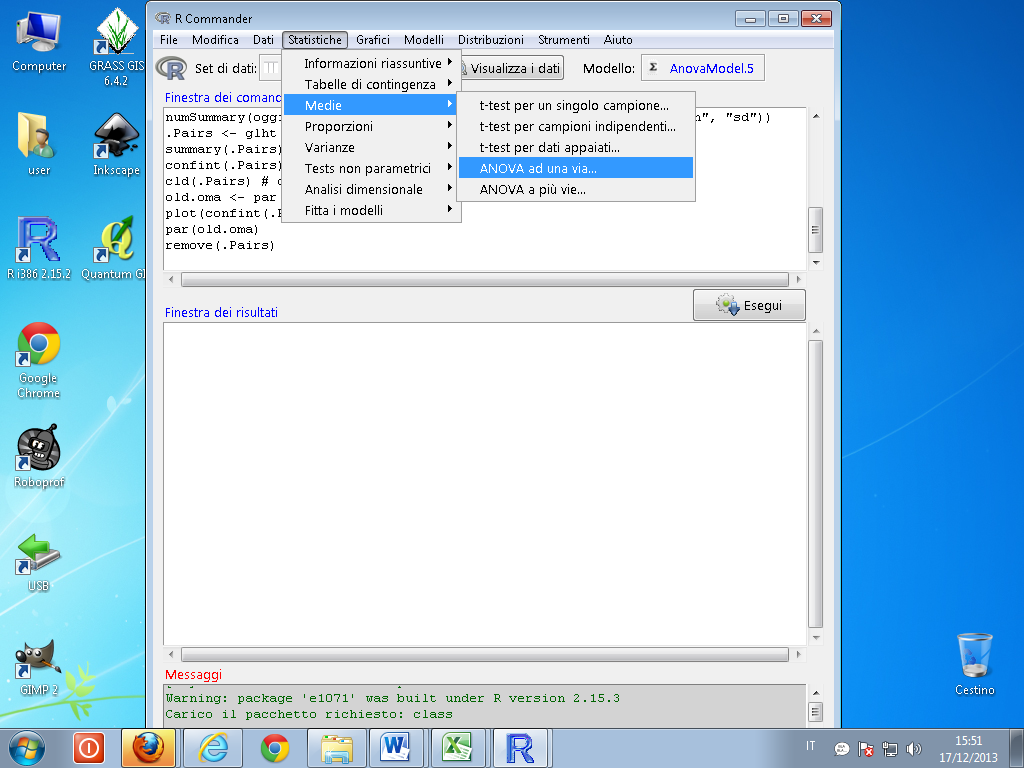
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

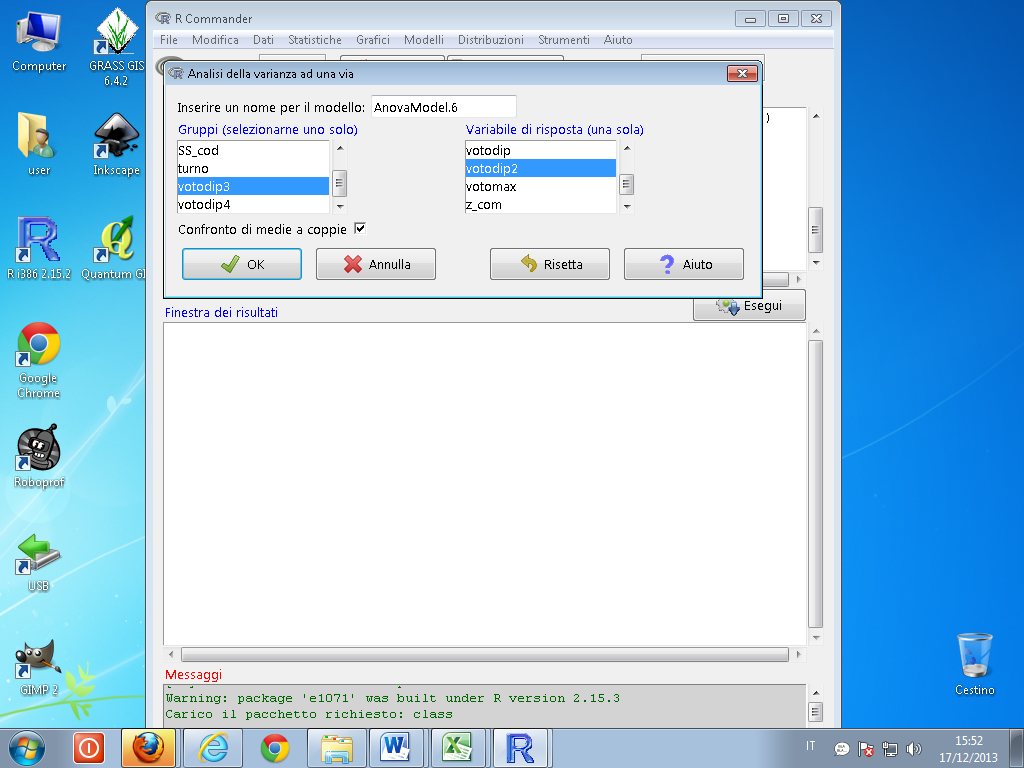
(Adjusted p values reported -- single-step method)

**L’unica differenza significativa è quella del voto medio alla maturità di SPP confrontato con il voto medio alla maturità di LIC**

ESERCIZIO

Voglio verificare se i partecipanti alla ricerca hanno un voto di diploma diverso a seconda del gruppo di voto di diploma a cui sono stati assegnati a posteriori (Votodip 3)





> AnovaModel.6 <- aov(votodip2 ~ votodip3, data=oggi)

> summary(AnovaModel.6)

Df Sum Sq Mean Sq **F value** **Pr(>F)**

votodip3 4 1.4925 0.3731 **425.7 <2e-16 \*\*\***

Residuals 143 0.1253 0.0009

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

2 observations deleted due to missingness

**RIFIUTO L’IPOTESI NULLA perché il valore p è inferiore a 0,05 (Ovvio, dato che “votodip 3” è stata costruita sulla base di “votodip 2”.**

**In questo caso il valore della statistica calcolata è F=425,7**

> numSummary(oggi$votodip2 , groups=oggi$votodip3, statistics=c("mean", "sd"))

mean sd data:n

6 0.6397636 0.03505818 47

7 0.7407933 0.02681412 52

8 0.8361765 0.02735393 34

9 0.9378788 0.02873094 10

10 1.0000000 0.00000000 5

> .Pairs <- glht(AnovaModel.6, linfct = mcp(votodip3 = "Tukey"))

> **summary(.Pairs) # pairwise tests**

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = votodip2 ~ votodip3, data = oggi)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

7 - 6 == 0 0.101030 0.005958 16.956 < 1e-04 \*\*\*

8 - 6 == 0 0.196413 0.006665 29.469 < 1e-04 \*\*\*

9 - 6 == 0 0.298115 0.010310 28.916 < 1e-04 \*\*\*

10 - 6 == 0 0.360236 0.013926 25.868 < 1e-04 \*\*\*

8 - 7 == 0 0.095383 0.006529 14.609 < 1e-04 \*\*\*

9 - 7 == 0 0.197086 0.010222 19.280 < 1e-04 \*\*\*

10 - 7 == 0 0.259207 0.013861 18.700 < 1e-04 \*\*\*

9 - 8 == 0 0.101702 0.010650 9.550 < 1e-04 \*\*\*

10 - 8 == 0 0.163824 0.014180 11.553 < 1e-04 \*\*\*

10 - 9 == 0 0.062121 0.016215 3.831 0.00156 \*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Adjusted p values reported -- single-step method)

**Tutti i successivi confronti a coppie mostrano che ogni media è significativamente diversa da ogni altra.**