

Se ipotesi bidirezionale

H0:

Alfa = 0,05 Zc = 1,96

Alfa = 0,01 Zc = 2,58

Alfa = 0,001 Zc = 3,29

**Es 8.21**

Se bidirezionale

H0: media = 3,5

HA: media ≠ 3,5

Se monodirezionale (RISPOSTA CORRETTA)

H0: media ≤ 3,5 (affermazione della casa farmaceutica)

HA: media > 3,5

ERRORE DI PRIMO TIPO: rifiuto H0 quando è vera

Nel caso specifico, non permetto alla casa farmaceutica di fare pubblicità anche se avevano ragione

ERRORE DI SECONDO TIPO: accetto H0 quando è falsa

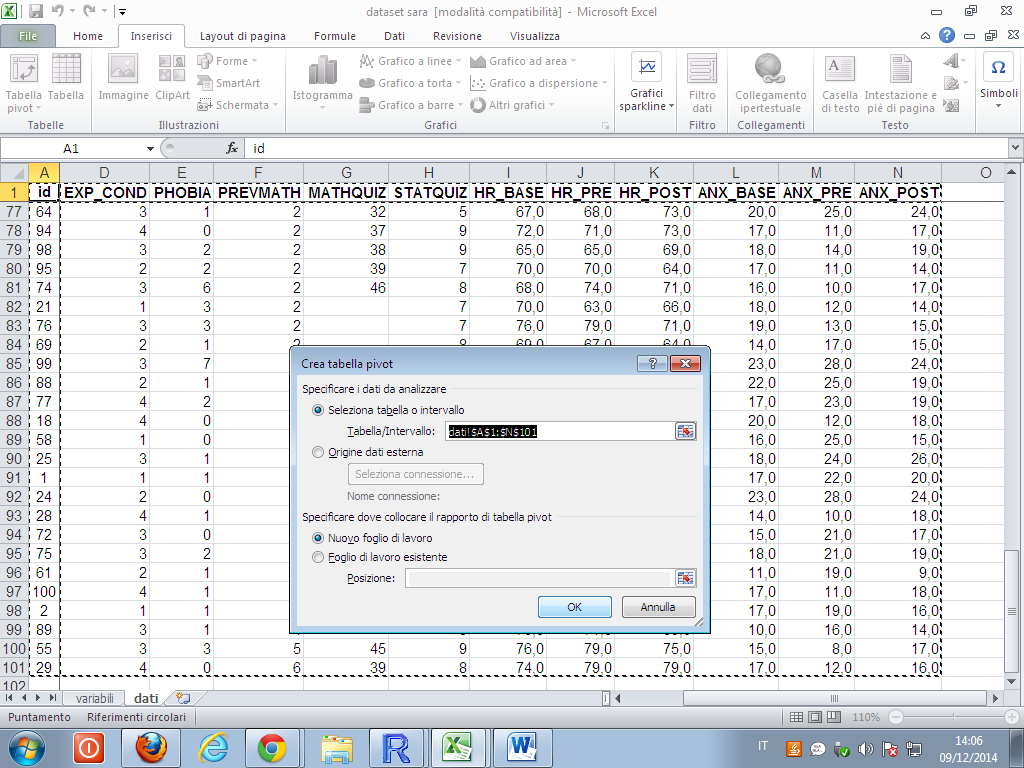
Permetto alla casa farmaceutica di fare pubblicità anche se avevano torto

**Appunti 4 dicembre (mancano)**

**Appunti 9 dicembre**

Come fare per

Tabella Pivot in Excel per statistiche descrittive e grafici



Carico il dataset Sara con il programma R

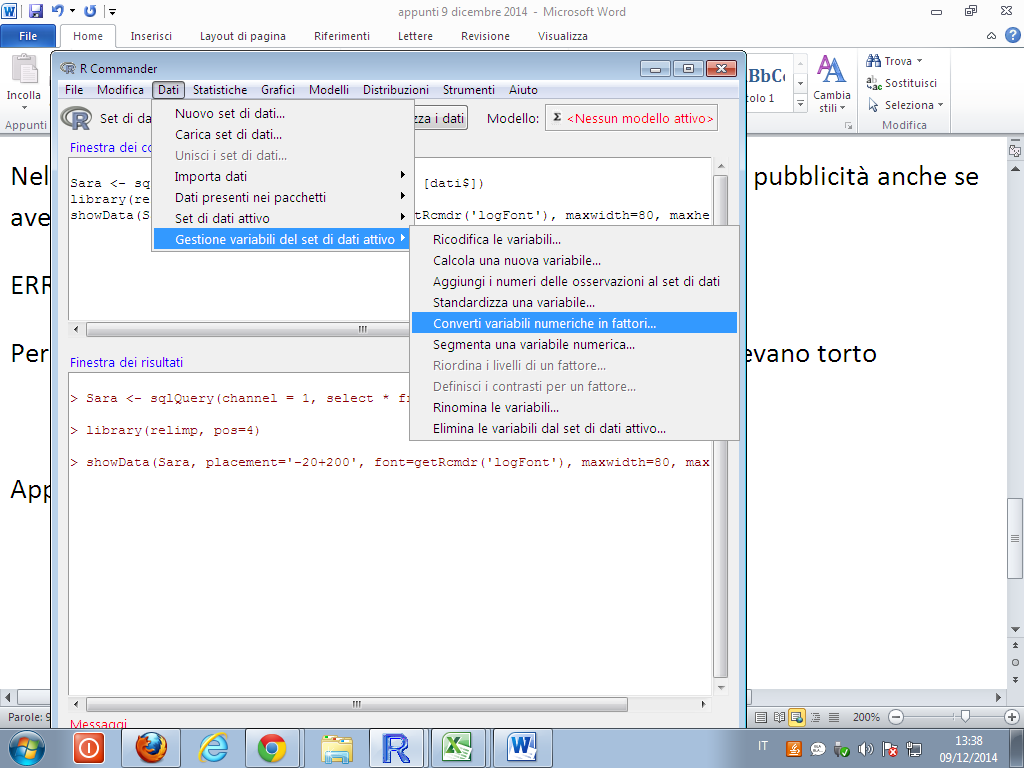
Come prima operazione converto le variabili numeriche in fattori,

e in particolare:

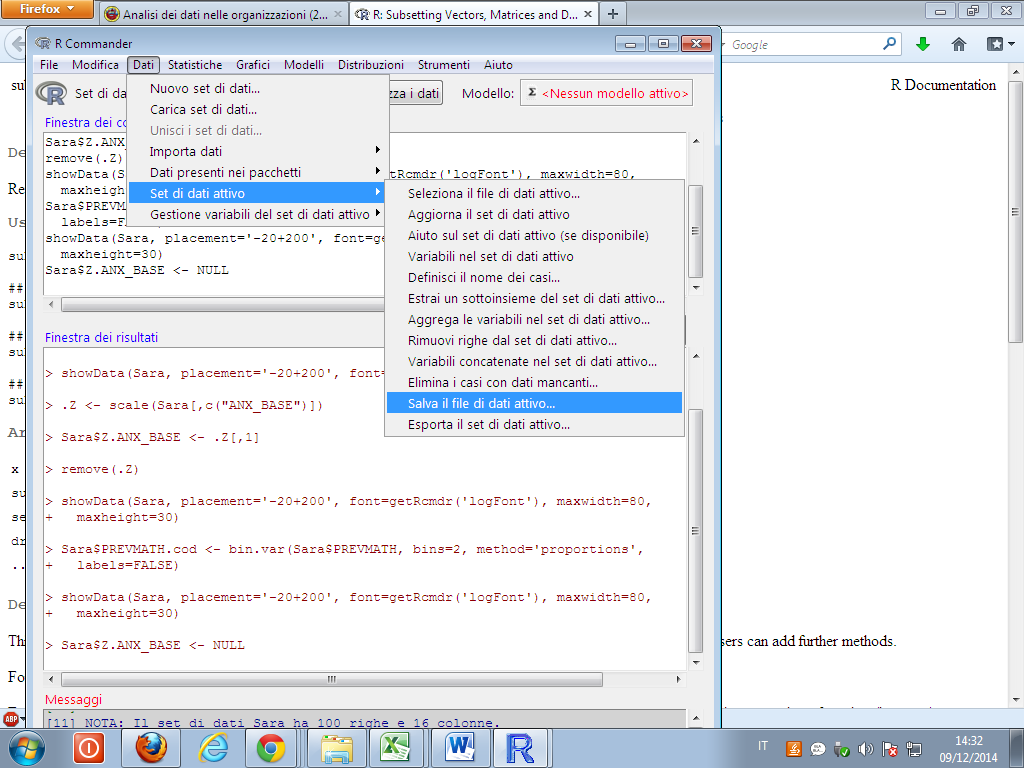
GENDER: 1=M; 2=F

UGMAJOR: 1=PSI; 2=MED; 3=BIO; 4=SOC; 5=ECO

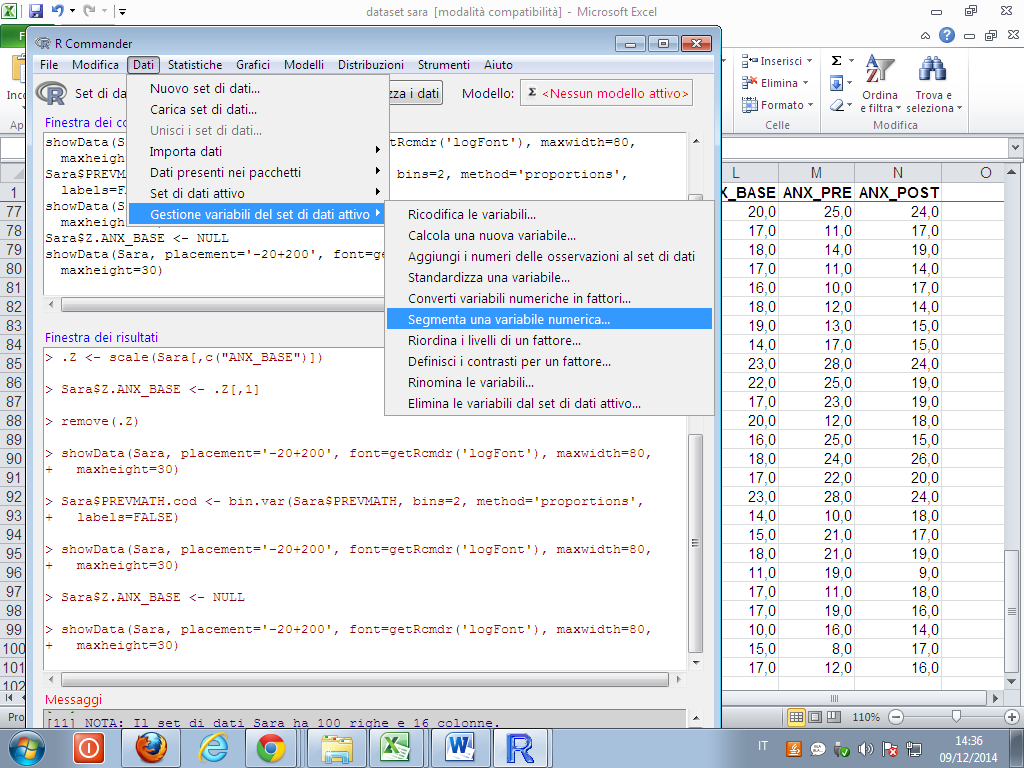
EXP\_COND: lasciare i numeri come nomi dei livelli (1, 2, 3, 4)



Come salvare il dataset in R



Creare classi da una variabile numerica



data: HR\_BASE by GENDER

F = 0.7823, num df = 56, denom df = 42, p-value = 0.3877

alternative hypothesis: true ratio of variances is not equal to 1

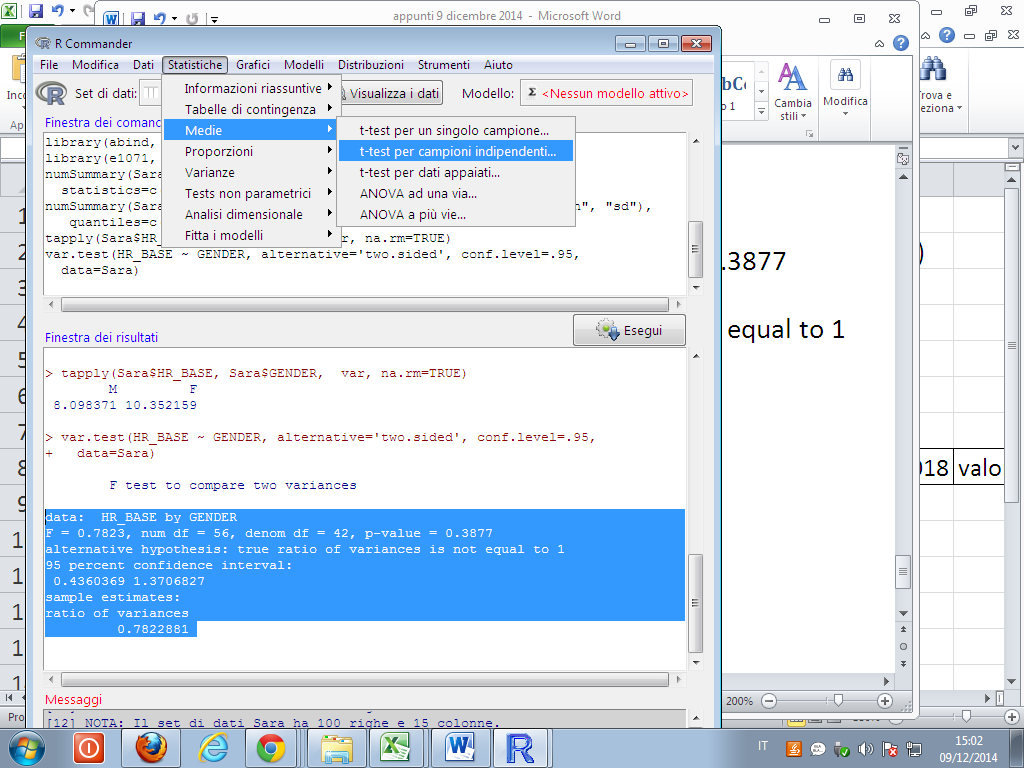
95 percent confidence interval:

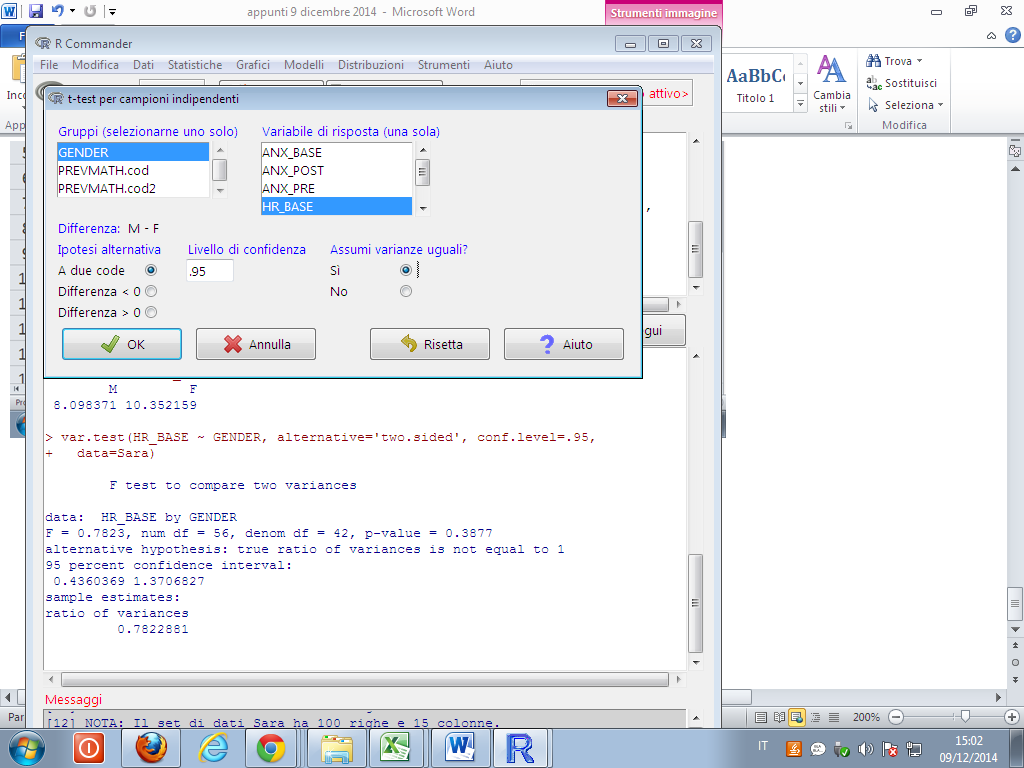
0.4360369 1.3706827

sample estimates:

ratio of variances

0.7822881





**DOMANDA: C’è differenza tra maschi e femmine quanto a frequenza cardiaca misurata in condizioni base?**

1. quale è il livello di scala delle variabili di cui sto testando l’esistenza di una relazione

Genere: variabile qualitativa

Frequenza cardiaca: variabile quantitativa

*In questo caso sto confrontando due valori medi (media della frequenza cardiaca), uno relativo al gruppo dei maschi, uno relativo al gruppo delle femmine. Quindi mi trovo nella situazione di applicare un* ***t-test per campioni indipendenti*** *(fatta salva l’assunzione di normalità delle distribuzioni delle due popolazioni)*

1. per applicare un t-test per campioni indipendenti devo scegliere tra la situazione “varianze uguali” e “varianze diverse”, per cui è necessario fare un **test di verifica di ipotesi sulle varianze** (test F)
2. In base all’esito del test sulle varianze (accetto H0 : le due varianze sono uguali; oppure rifiuto H0 : le due varianze non sono uguali) cambia il modo di calcolare l’errore standard e i g.d.l.
3. Un’altra decisione da prendere è se si tratta di una ipotesi monodirezionale o bidirezionale. Nel primo caso ipotizzo esplicitamente una direzione della differenza (il più delle volte per motivi teorici); nel secondo caso ipotizzo solo la differenza tra le due medie. Nel caso del presente esempio si tratta dunque di una **ipotesi bidirezionale**.

> **t.test(HR\_BASE~GENDER, alternative='two.sided', conf.level=.95,**

**+ var.equal=TRUE, data=Sara**)

Two Sample t-test

data: HR\_BASE by GENDER

t = 3.8651, df = 98, p-value = 0.0001997 (cfr. p.108)

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

1.143659 3.557279

sample estimates:

mean in group M mean in group F

73.28070 70.93023

> t.test(STATQUIZ~PREVMATH.cod, alternative='less', conf.level=.95,

+ var.equal=TRUE, data=Sara)

Two Sample t-test

data: STATQUIZ by PREVMATH.cod

t = -3.3782, df = 98, p-value = 0.0005239

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -0.5953797

sample estimates:

mean in group 1 mean in group 2

6.485294 7.656250

1. Poiché il valore p calcolato in base ai dati campionari è inferiore al p critico (cioè alfa, 0,05), rifiuto l’ipotesi nulla e quindi le medie sono diverse.

**DOMANDA: Mi chiedo se l’ansia pre, dato che siamo in una situazione pre-esame, e quindi potenzialmente con ansia elevata, sia maggiore dell’ansia post, ovvero del livello di ansia misurato dopo che la prova è stata sostenuta.**

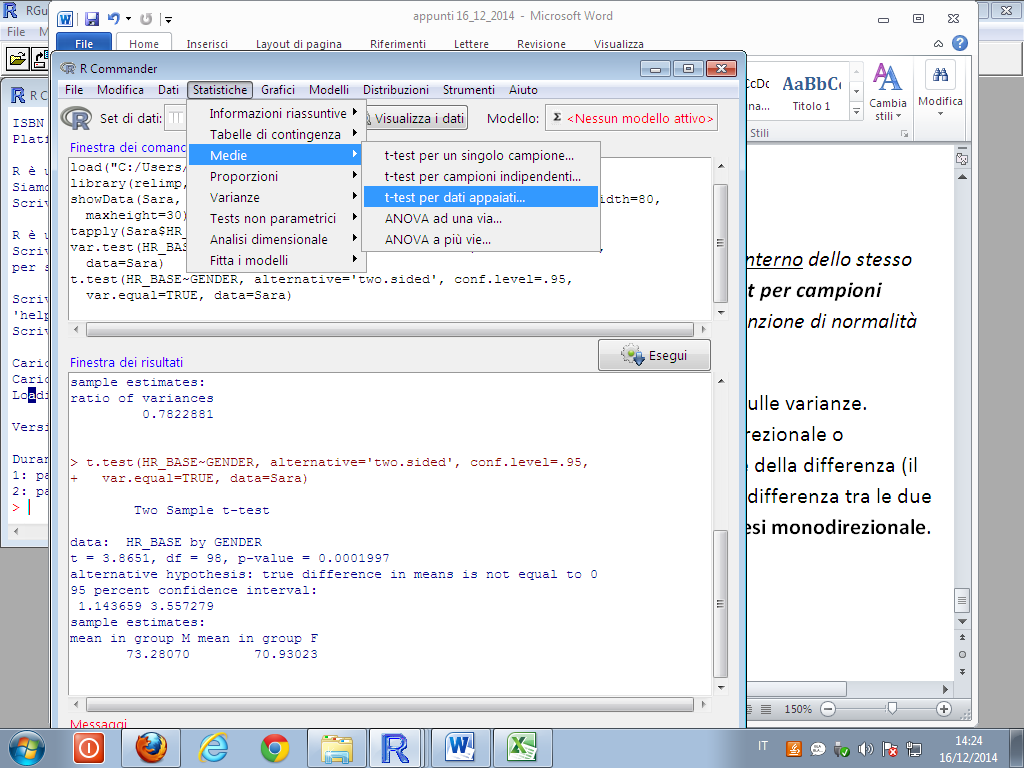
1. quale è il livello di scala delle variabili di cui sto testando l’esistenza di una relazione?

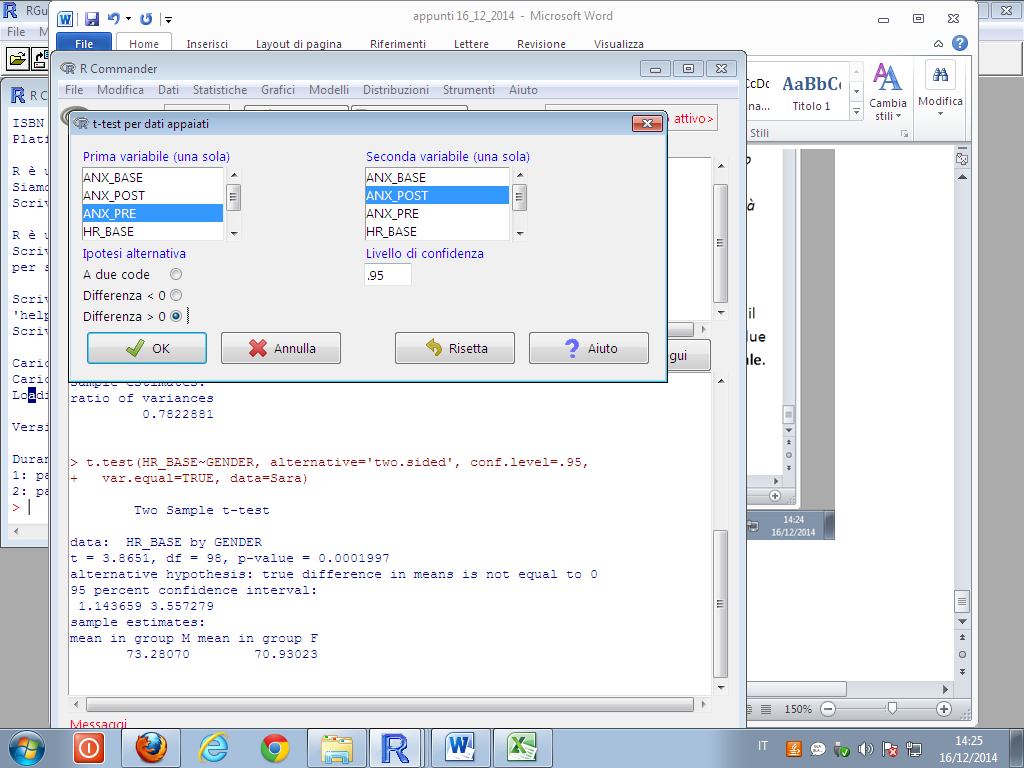
Ansia pre: variabile quantitativa

Ansia post: variabile quantitativa

*In questo caso sto confrontando due valori medi sempre all’interno dello stesso gruppo. Quindi mi trovo nella situazione di applicare un* ***t-test per campioni appaiati (dipendenti, paired, accoppiati)*** *(fatta salva l’assunzione di normalità delle distribuzioni delle due popolazioni)*

1. per applicare un t-test per campioni appaiati non devo fare il test sulle varianze.
2. Un’altra decisione da prendere è se si tratta di una ipotesi monodirezionale o bidirezionale. Nel primo caso ipotizzo esplicitamente una direzione della differenza (il più delle volte per motivi teorici); nel secondo caso ipotizzo solo la differenza tra le due medie. Nel caso del presente esempio si tratta dunque di una **ipotesi monodirezionale**.





> t.test(Sara$ANX\_PRE, Sara$ANX\_POST, alternative='greater', conf.level=.95,

+ paired=TRUE)

Paired t-test

data: Sara$ANX\_PRE and Sara$ANX\_POST

t = 0.4018, df = 99**, p-value = 0.3443**

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

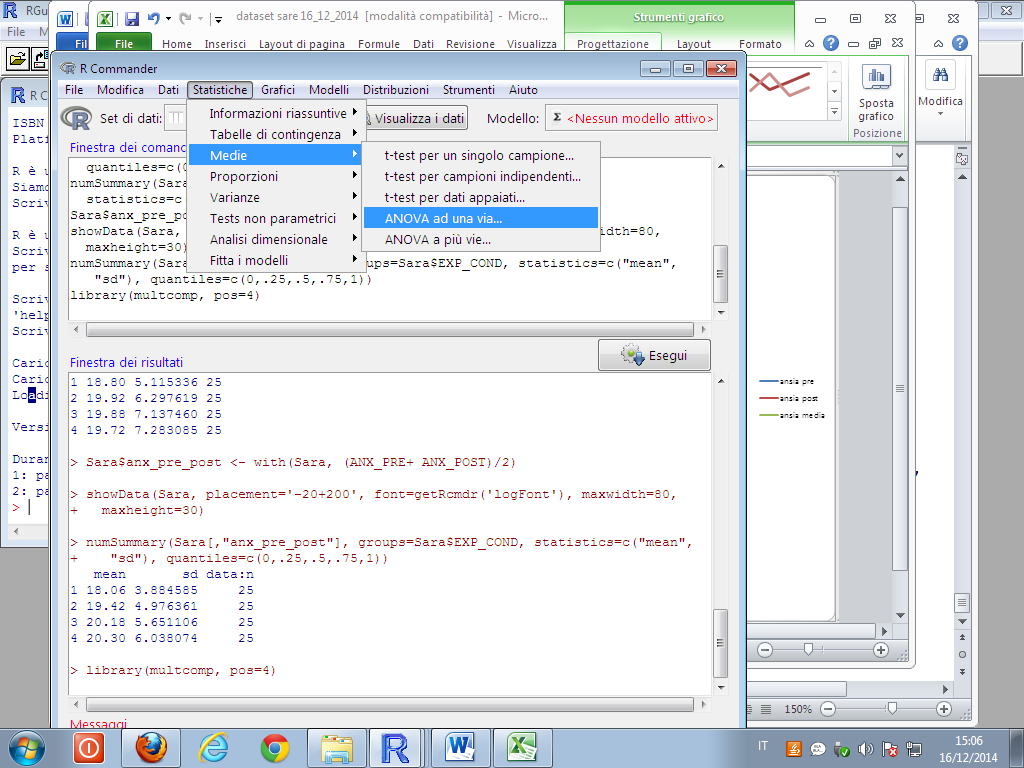
-0.563816 Inf

sample estimates:

mean of the differences

0.18

1. Poiché il valore p calcolato in base ai dati campionari è superiore al p critico (cioè alfa, 0,05), accetto l’ipotesi nulla e il punteggio medio di ansia prima dell’esame non è maggiore del punteggio medio di ansia dopo l’esame.



DOMANDA: Mi chiedo se c’è un effetto della condizione sperimentale sull’ansia post, ovvero se almeno una delle coppie possibili di valori di ansia post esame ha valori diversi.

> AnovaModel.2 <- aov(ANX\_POST ~ EXP\_COND, data=Sara)

> summary(AnovaModel.2)

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

EXP\_COND 3 197.8 65.95 3.082 0.031 \*

Residuals 96 2054.2 21.40

---

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

> numSummary(Sara$ANX\_POST , groups=Sara$EXP\_COND, statistics=c("mean", "sd"))

mean sd data:n

1 17.32 3.613862 25

2 18.92 4.545327 25

3 20.48 4.646863 25

4 20.88 5.502424 25

> .Pairs <- glht(AnovaModel.2, linfct = mcp(EXP\_COND = "Tukey"))

> summary(.Pairs) # pairwise tests

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = ANX\_POST ~ EXP\_COND, data = Sara)

Linear Hypotheses:

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

2 - 1 == 0 1.600 1.308 1.223 0.6139

3 - 1 == 0 3.160 1.308 2.415 0.0808 .

4 - 1 == 0 3.560 1.308 2.721 0.0382 \*

3 - 2 == 0 1.560 1.308 1.192 0.6332

4 - 2 == 0 1.960 1.308 1.498 0.4427

4 - 3 == 0 0.400 1.308 0.306 0.9900

---

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

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

> confint(.Pairs) # confidence intervals

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = ANX\_POST ~ EXP\_COND, data = Sara)

Quantile = 2.615

95% family-wise confidence level

Linear Hypotheses:

Estimate lwr upr

2 - 1 == 0 1.6000 -1.8213 5.0213

3 - 1 == 0 3.1600 -0.2613 6.5813

4 - 1 == 0 3.5600 0.1387 6.9813

3 - 2 == 0 1.5600 -1.8613 4.9813

4 - 2 == 0 1.9600 -1.4613 5.3813

4 - 3 == 0 0.4000 -3.0213 3.8213

> cld(.Pairs) # compact letter display

1 2 3 4

“a" "ab" "ab" "b"

> old.oma <- par(oma=c(0,5,0,0))

> plot(confint(.Pairs))

> par(old.oma)

> remove(.Pairs)