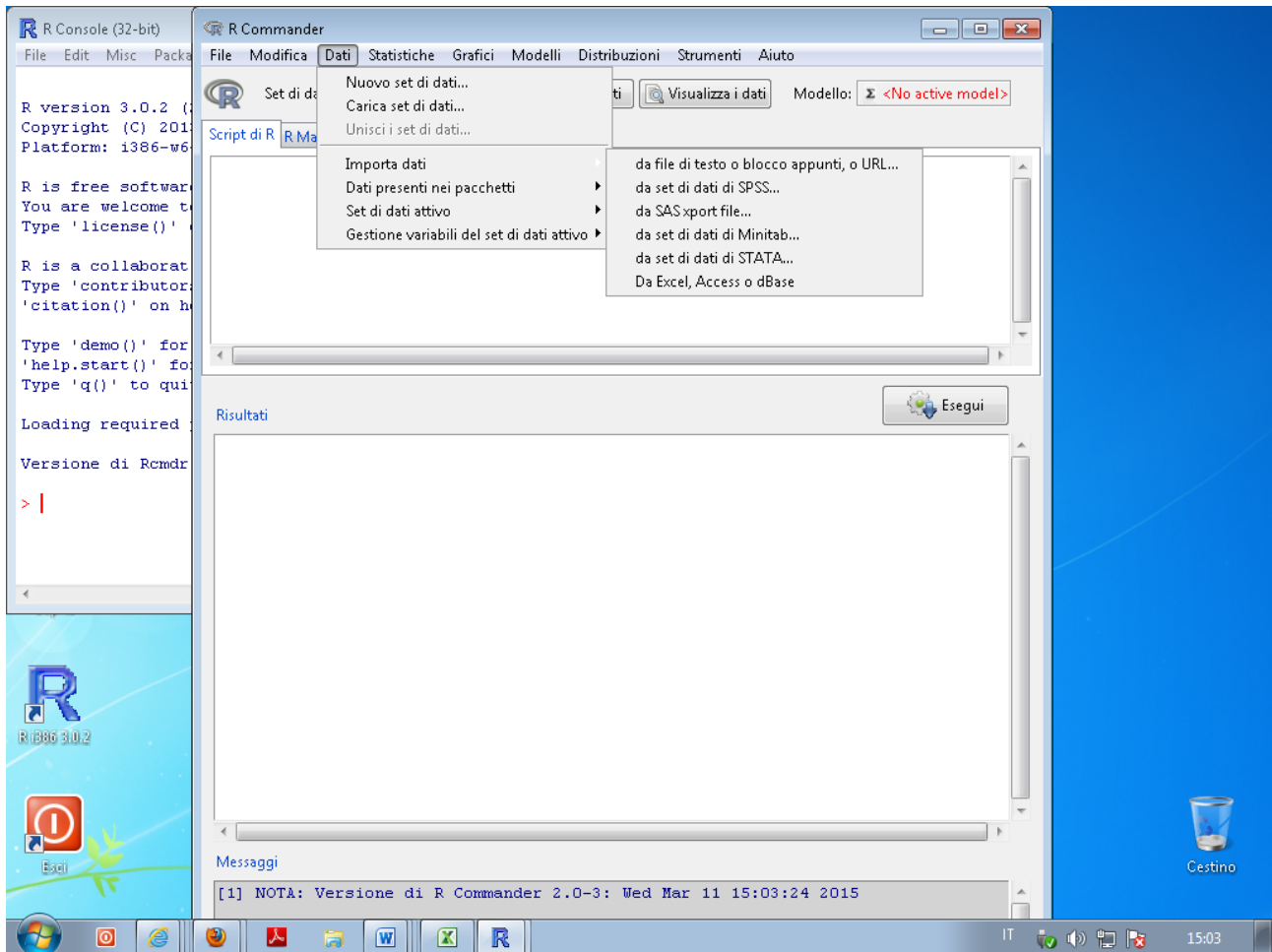
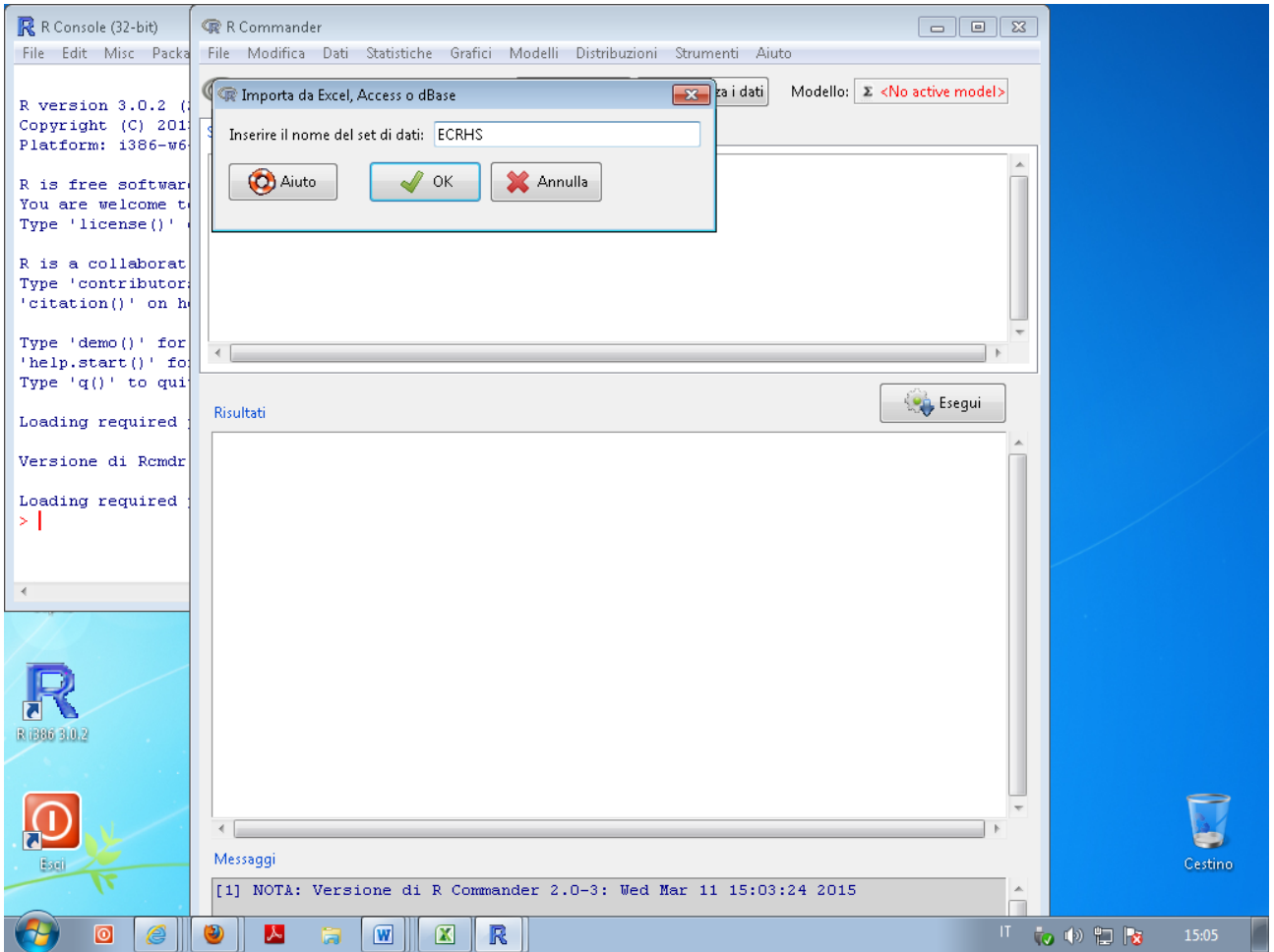


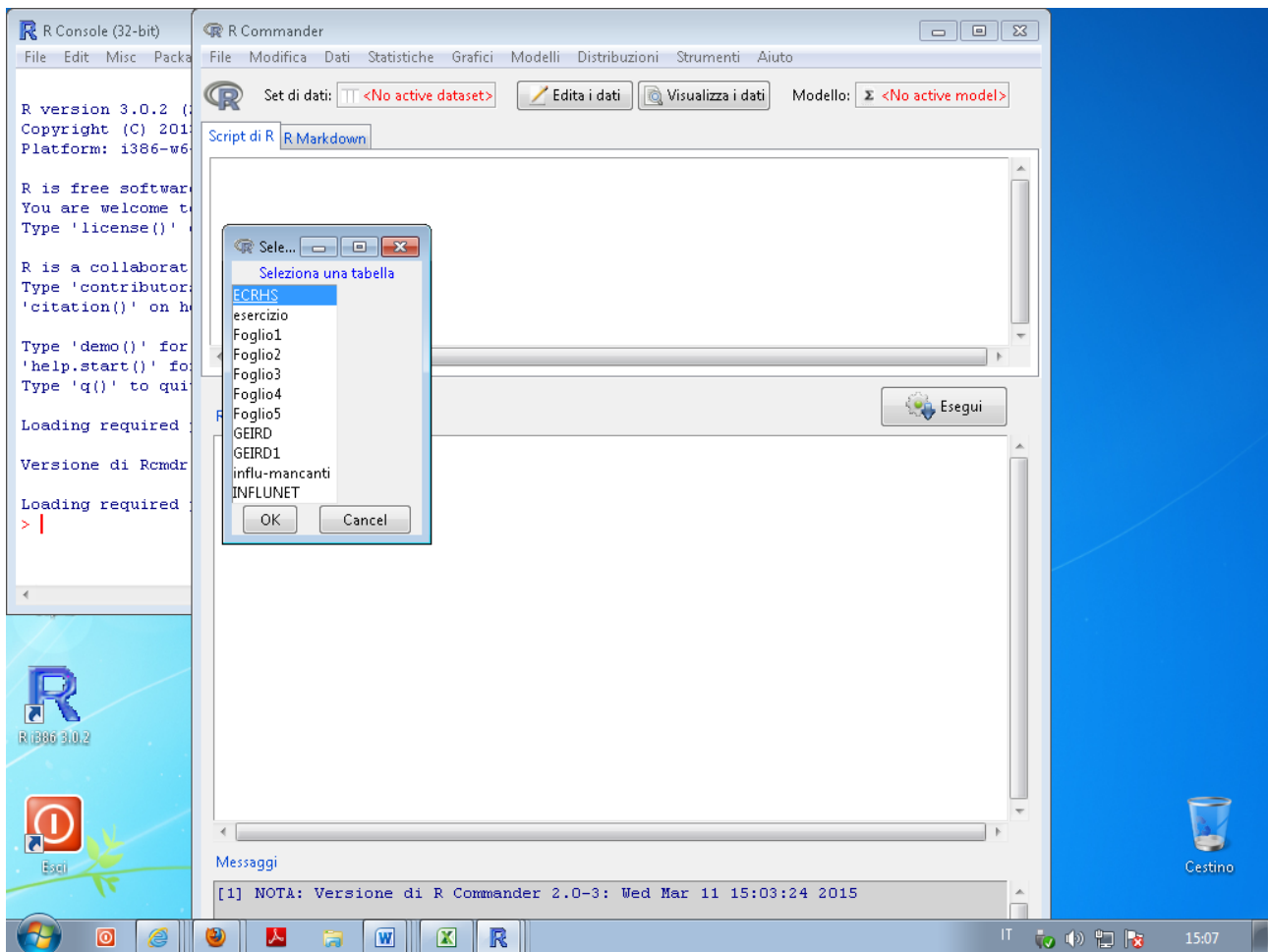
Per caricare un set di dati:

Dati/Importa dati/Da Excel, Access o dBase



Nomino il dataset che sto caricando in R con un nome per me significativo, in questo esempio ECRHS





Cliccando su Visualizza i dati posso verificare se il foglio di lavoro che ho selezionato era quello con i dati che mi interessano

R Commander interface showing R code execution and data output.

```
ECRHS <- sqlQuery(channel = 1, select * from [ECRHS$])
names(ECRHS) <- make.names(names(ECRHS))
library(relimp, pos=4)
showData(ECRHS, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
maxheight=30)
```

Results:

```
> ECRHS <- sqlQuery(
> names(ECRHS) <- ma
> library(relimp, po
> showData(ECRHS, pl
+ maxheight=30)
```

Messages:

```
These have been chan
ever asthma
```

id	sex	age	height	weight	fev1	fvc1	wheezing	thightness	dispnoea	ever.asthma
1	1	43.72	184	81	489	562	1	1	1	1
2	1	42.10	NA	NA	NA	NA	2	1	1	1
3	1	44.85	NA	NA	NA	NA	1	1	1	1
4	1	44.03	174	68	363	446	2	1	1	1
5	1	42.07	182	84	431	499	1	1	1	1
6	1	42.95	177	58	380	474	1	1	1	1
7	1	44.25	164	63	358	436	1	1	1	1
8	1	43.23	168	73	422	477	1	1	1	1
9	1	45.13	161	72	392	450	1	1	1	1
10	1	44.18	178	78	420	464	1	1	1	1
11	1	41.38	177	77	478	600	1	1	1	1
12	1	44.03	179	83	477	593	2	1	1	2
13	1	44.34	182	96	349	475	1	1	1	1
14	1	44.59	177	86	454	526	1	1	1	1
15	1	43.39	185	97	NA	NA	1	1	1	1
16	1	44.65	177	92	377	522	2	1	1	1
17	1	44.62	170	70	375	482	1	1	1	1
18	1	43.50	169	77	460	498	1	1	1	1
19	1	44.16	163	66	324	419	1	1	1	1
20	1	42.02	160	86	367	430	1	1	1	1
21	1	42.79	166	79	436	527	2	2	1	1
22	1	44.33	169	72	354	422	1	1	1	1
23	1	41.31	172	80	316	403	1	1	1	1
24	1	43.15	160	60	289	325	1	1	1	1
25	1	43.85	171	70	NA	NA	2	2	1	1
26	1	41.64	NA	NA	NA	NA	1	1	1	1
27	1	41.15	NA	NA	NA	NA	1	1	1	1
28	1	45.50	182	85	468	589	1	1	1	1
29	1	43.39	174	82	415	475	1	1	1	1
30	1	41.26	177	70	414	512	1	1	1	1

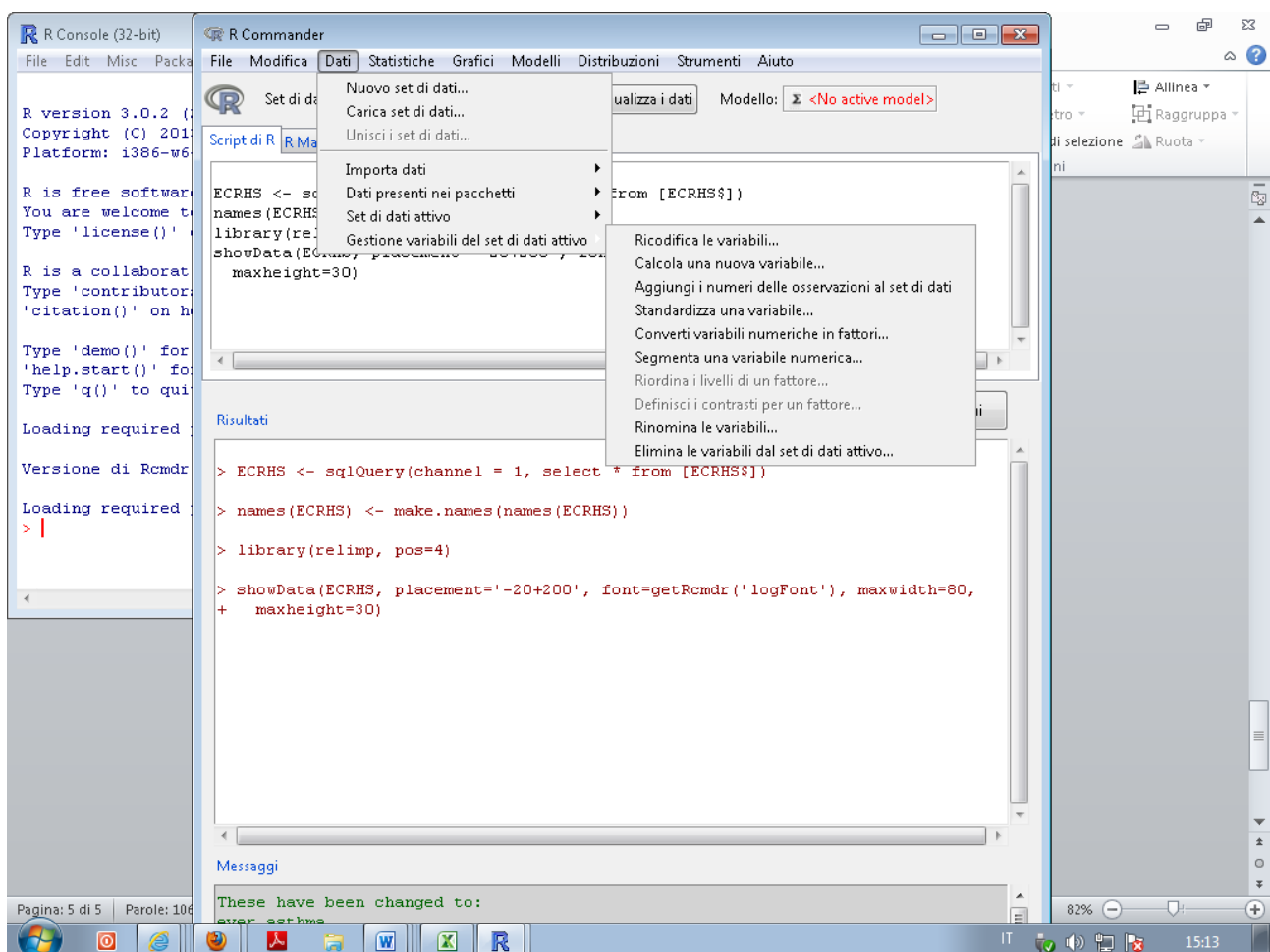
TEST DI VERIFICA DI IPOTESI SULLA DIFFERENZA TRA DUE MEDIE

CASO DI DUE CAMPIONI INDIPENDENTI

La prima cosa da fare per poter applicare il test con il programma R e il pacchetto Rcmdr (R commander) è quella di convertire le variabili numeriche eventualmente presenti in “fattori”, ovvero variabili misurate su scala nominale.

Per fare questo seguire il percorso:

Dati/Gestione variabili del set di dati attivo/converti variabili numeriche in fattori::



Decido di assegnare nomi ai livelli e di sovrascrivere la variabile precedente

The image shows the R Commander interface. A dialog box titled "Converti variabili numeriche in fattori" is open. It has two sections: "Variabili (una o più)" and "Livelli del fattore".

Variabili (una o più): A list of variables: height, id, sex, thightness, weight, wheezing. The variable "sex" is selected.

Livelli del fattore: Two radio buttons: "Assegna nomi ai livelli" (selected) and "Utilizza i numeri".

Nuovo nome della variabile o prefisso per variabili multiple: A text field containing "<same as variables>".

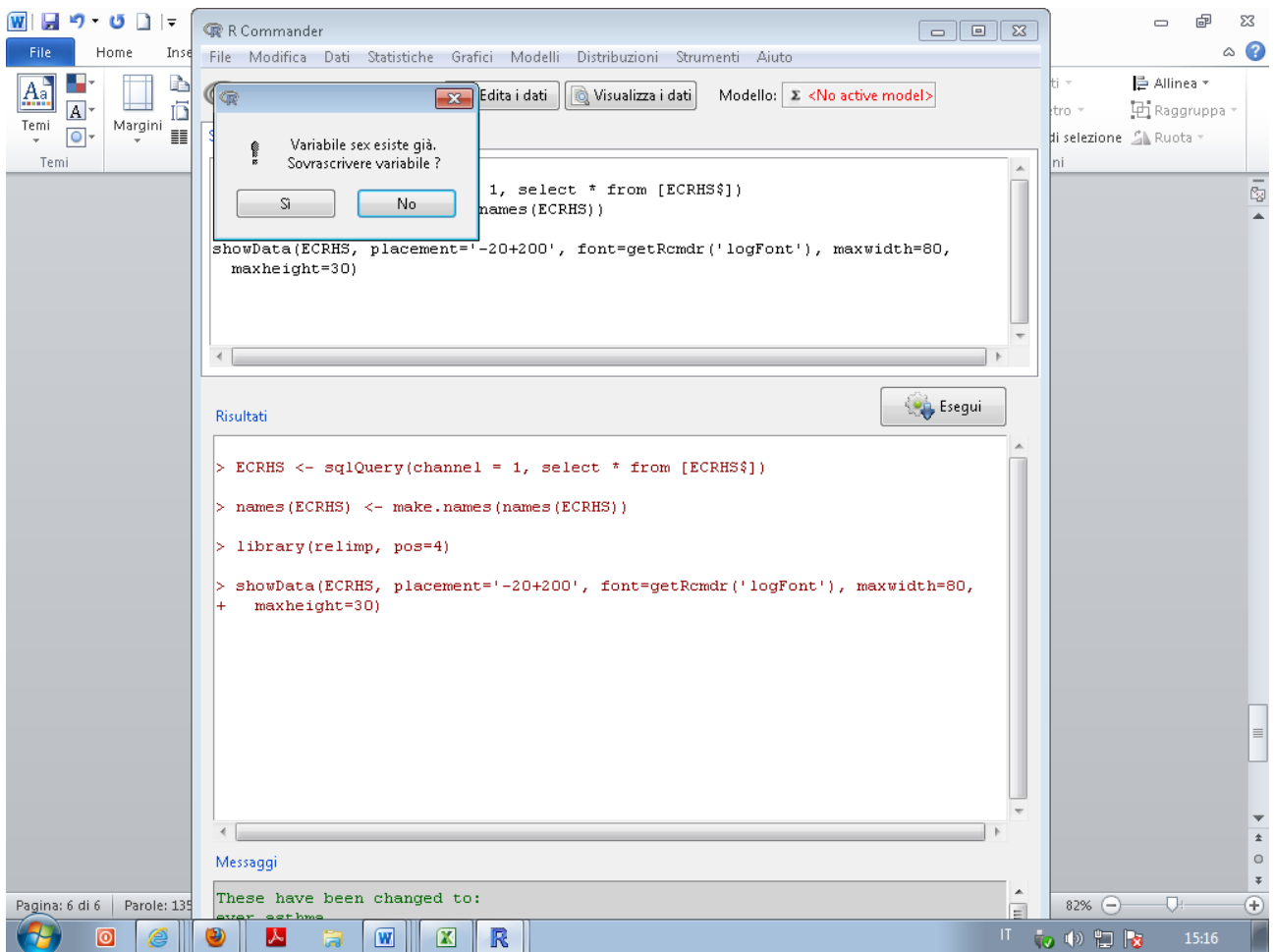
Buttons: "Aiuto", "OK", "Annulla".

Risultati: A console window showing the following R code:

```
> ECRHS <- sqlQuery(channel = 1, select * from [ECRHS$])
> names(ECRHS) <- make.names(names(ECRHS))
> library(relimp, pos=4)
> showData(ECRHS, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
+ maxheight=30)
```

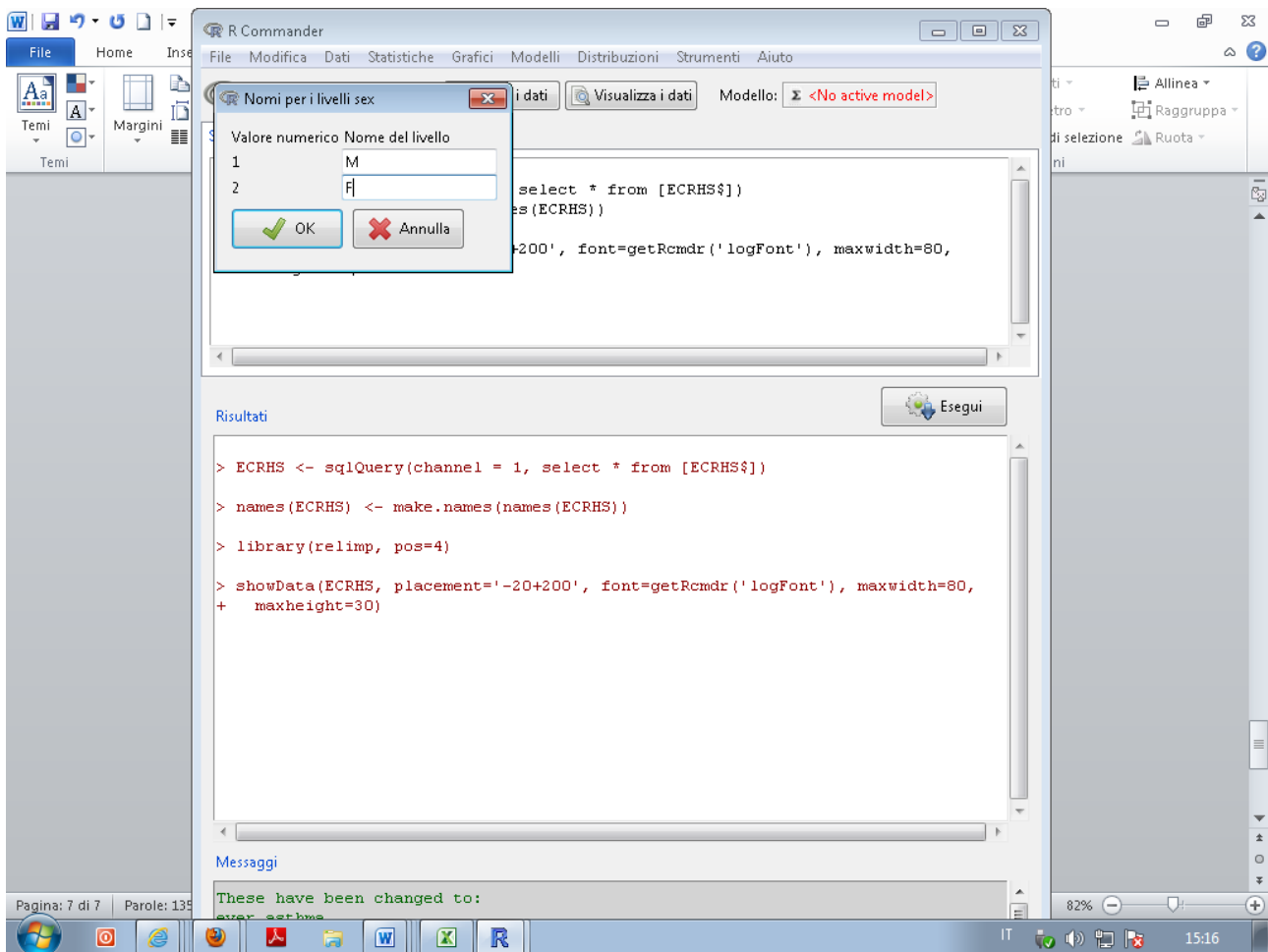
Message: A message box at the bottom says "These have been changed to: ever_asthma".

The taskbar at the bottom shows the Windows Start button, taskbar icons for Internet Explorer, R Commander, and other applications. The system tray shows the date and time as 15:15.



Rispondo "Sì"

Decido che 1 corrisponde a M e 2 corrisponde a F



Domanda:

Sono interessato a vedere se c'è differenza tra maschi e femmine quanto al peso.

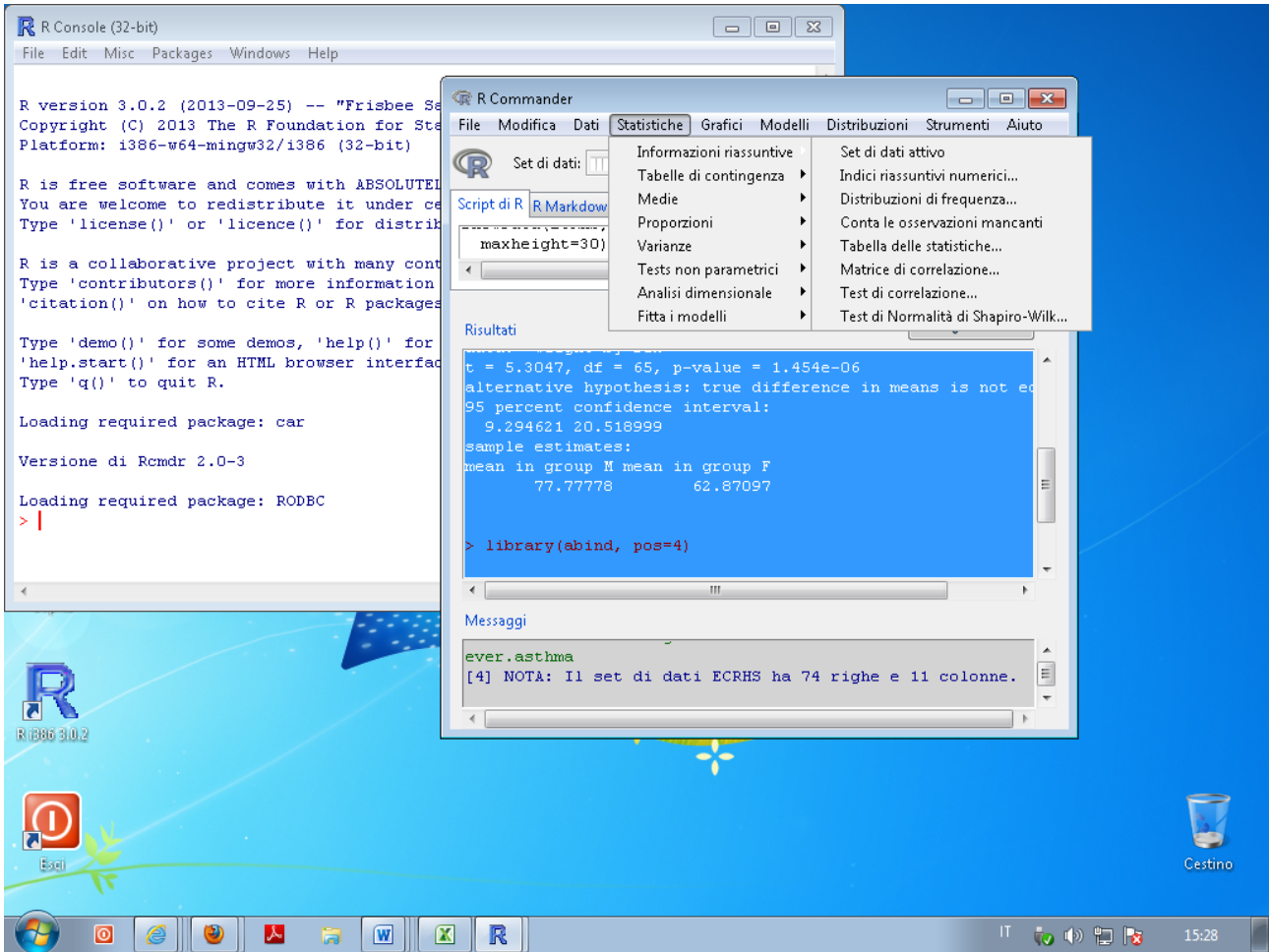
Dal momento che la variabile considerata (ovvero “weight”, il peso) è una variabile quantitativa, io potrò calcolare il valore medio.

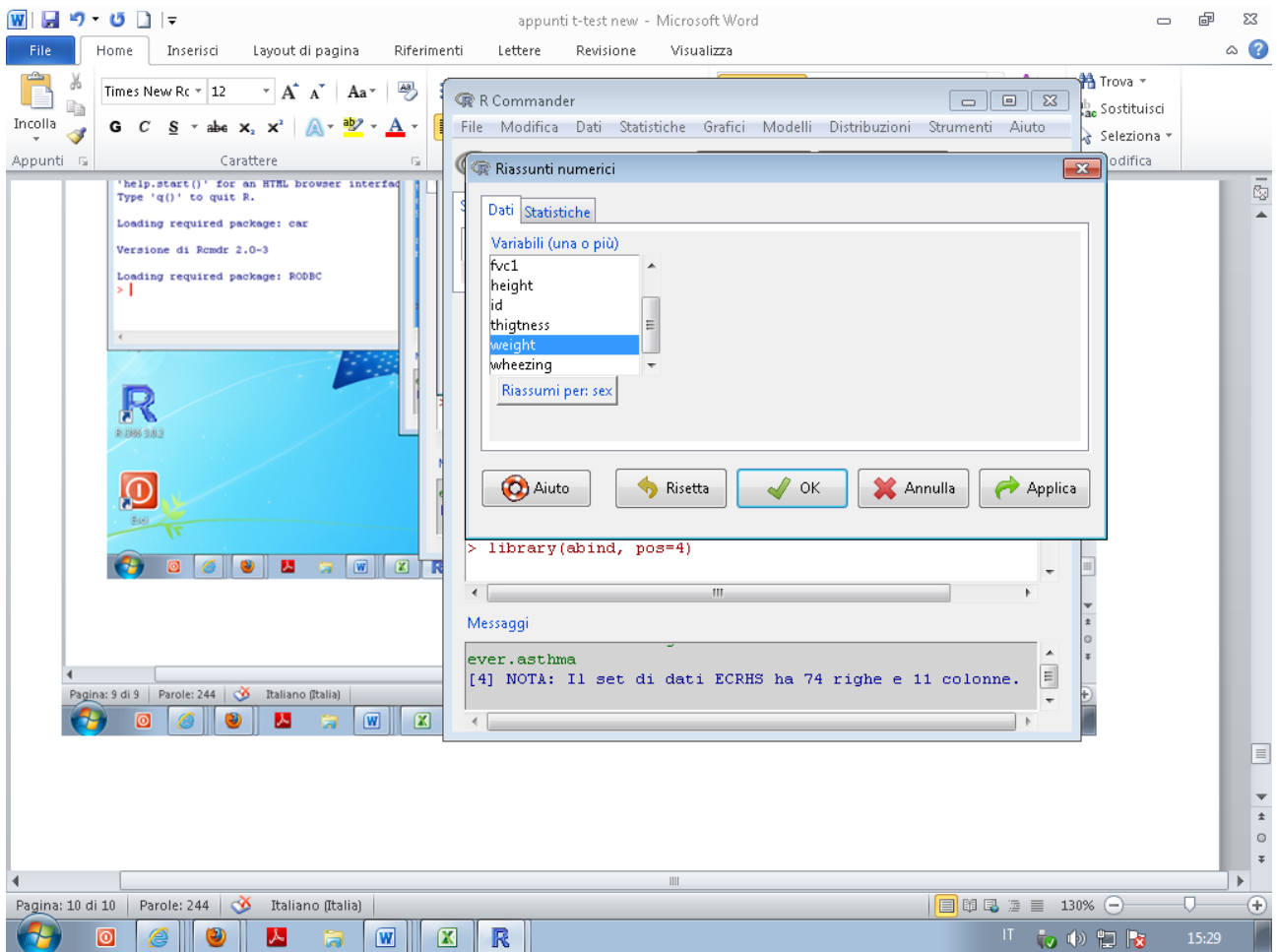
Dal momento che voglio verificare la differenza tra maschi e femmine, dovrò avere a disposizione due valori medi, uno per i maschi e uno per le femmine (la variabile “sex” è la variabile di raggruppamento, una variabile misurata su scala nominale).

Posso dunque calcolare media e deviazione standard della variabile “weight”.

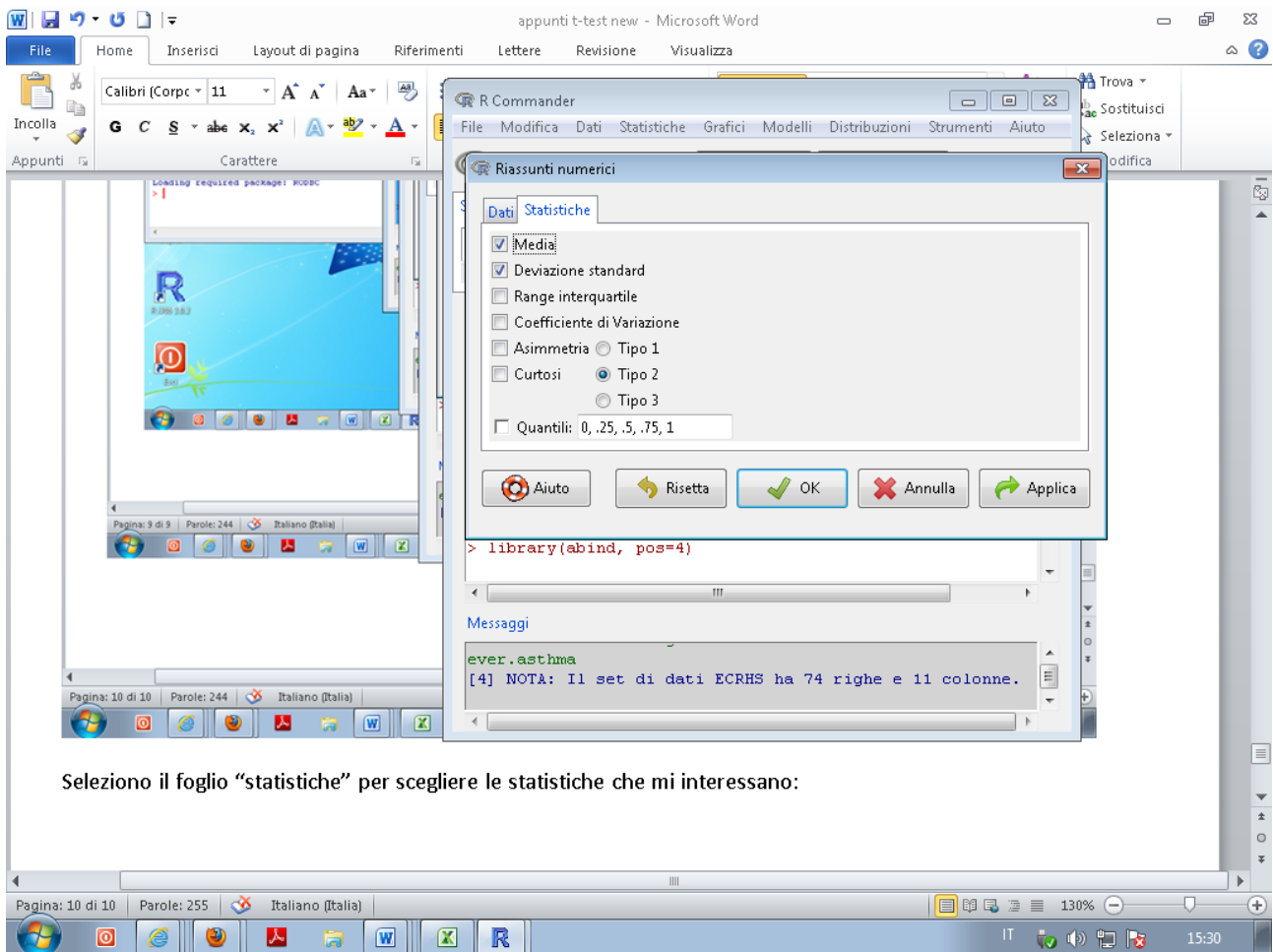
Per farlo uso i seguenti comandi:

Statistiche/informazioni riassuntive/indici riassuntivi numerici





Seleziono il foglio "statistiche" per scegliere le statistiche che mi interessano (in questo caso media e deviazione standard):



Questo sarà, in questo caso illustrato, l'esito finale:

```
> numSummary(ECRHS[, "weight"], groups=ECRHS$sex,
+ statistics=c("mean", "sd"), quantiles=c(0,.25,.5,.75,
+ 1))
```

	Mean	sd	data:n	data:NA
M	77.77778	9.931511	36	5
F	62.87097	13.035188	31	2

Individuo il sistema di ipotesi, ipotesi nulla e ipotesi alternativa:

- IPOTESI NULLA (H_0):** QUANTO ALLA VARIABILE PESO, LA MEDIA DEI MASCHI E' UGUALE ALLA MEDIA DELLE FEMMINE
- IPOTESI ALTERNATIVA (H_1):** QUANTO ALLA VARIABILE PESO, LA MEDIA DEI MASCHI E' DIVERSA DALLA MEDIA DELLE FEMMINE

In questo caso si tratta di una ipotesi alternativa bidirezionale, perché non ho una ipotesi specifica sulla direzione della differenza.

Poiché le medie sono due, posso applicare il t-test per campioni indipendenti (se fossero più di due medie non potrei, ma dovrei applicare l'ANOVA)

I t-test per campioni indipendenti prevede un algoritmo di calcolo differente a seconda che le varianze siano uguali o diverse. Per questo eseguo un test di verifica di ipotesi sulle varianze.

ipotesi nulla: le due varianze sono uguali

ipotesi alternativa: le due varianze sono diverse³

Per fare il test sulle varianze seguo la seguente procedura:
Statistiche/varianze/test F per due varianze:

The screenshot shows the R Commander interface. The 'Statistiche' menu is open, with 'Test F per due varianze...' selected. The results window displays the following data:

	mean	sd	data:n	data:NA
M	77.77778	9.931511	36	5
F	62.87097	13.035188	31	2

The code executed in the console is:

```
> numSummary(ECRHS[, "weight"], groups=ECRHS$sex,  
+ statistics=c("mean", "sd"), quantiles=c(0,.25,.5,.75,  
+ 1))
```

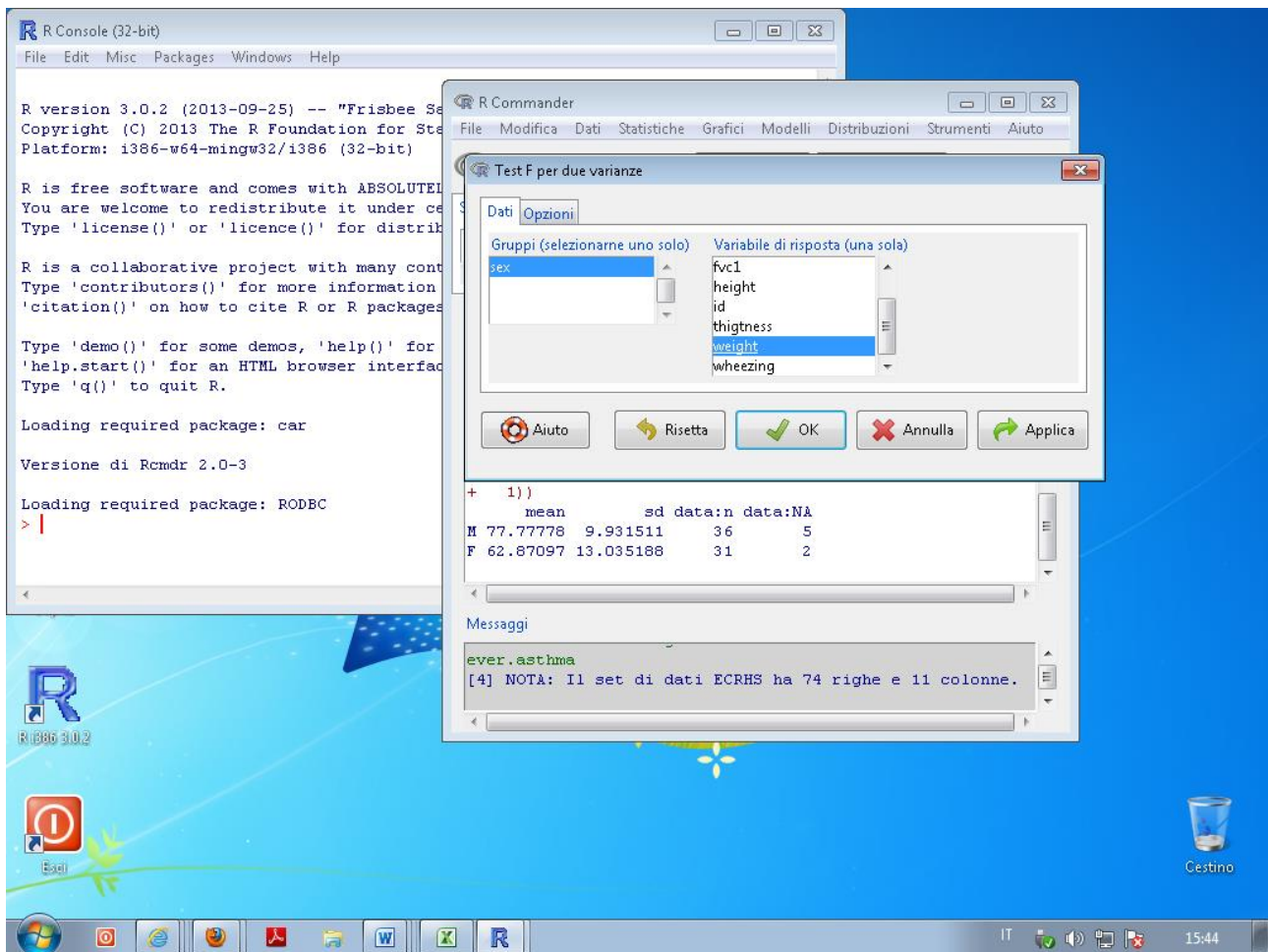
The results window also shows a message: [4] NOTA: Il set di dati ECRHS ha 74 righe e 11 colonne.

I t-test per c
varianze sia

ipotesi nulla
ipotesi alter

Per fare il te
Statistiche/

te a seconda che
testi sulle varian



L'esito nel seguente caso sarà:

```
> var.test(weight ~ sex, alternative='two.sided',
+ conf.level=.95, data=ECRHS)
```

F test to compare two variances

data: weight by sex

F = 0.5805, num df = 35, denom df = 30, **p-value = 0.1225**
 alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2849485 1.1601836

sample estimates:

ratio of variances

0.5804918

Poiché il valore p è superiore a 0,05 non rifiuto (accetto) l'ipotesi nulla, quindi significa che le due varianze possono essere considerate uguali.

A questo punto posso procedere con l'esecuzione del t-test per campioni indipendenti, caso delle varianze uguali.

Statistiche/medie/t-test per campioni indipendenti:

The screenshot shows the R Commander interface. The main window displays a document with text including "ratio of varian", "0.58049", "Poiché il valor", "varianze poss", "A questo punt", "varianze ugua", and "Statistiche/me". The "t-test per campioni indipendenti" dialog box is open, showing the following configuration:

- Dati:** Gruppi (selezionarne uno solo) - `sex`
- Opzioni:** Variabile di risposta (una sola) - `weight`

The console output shows:

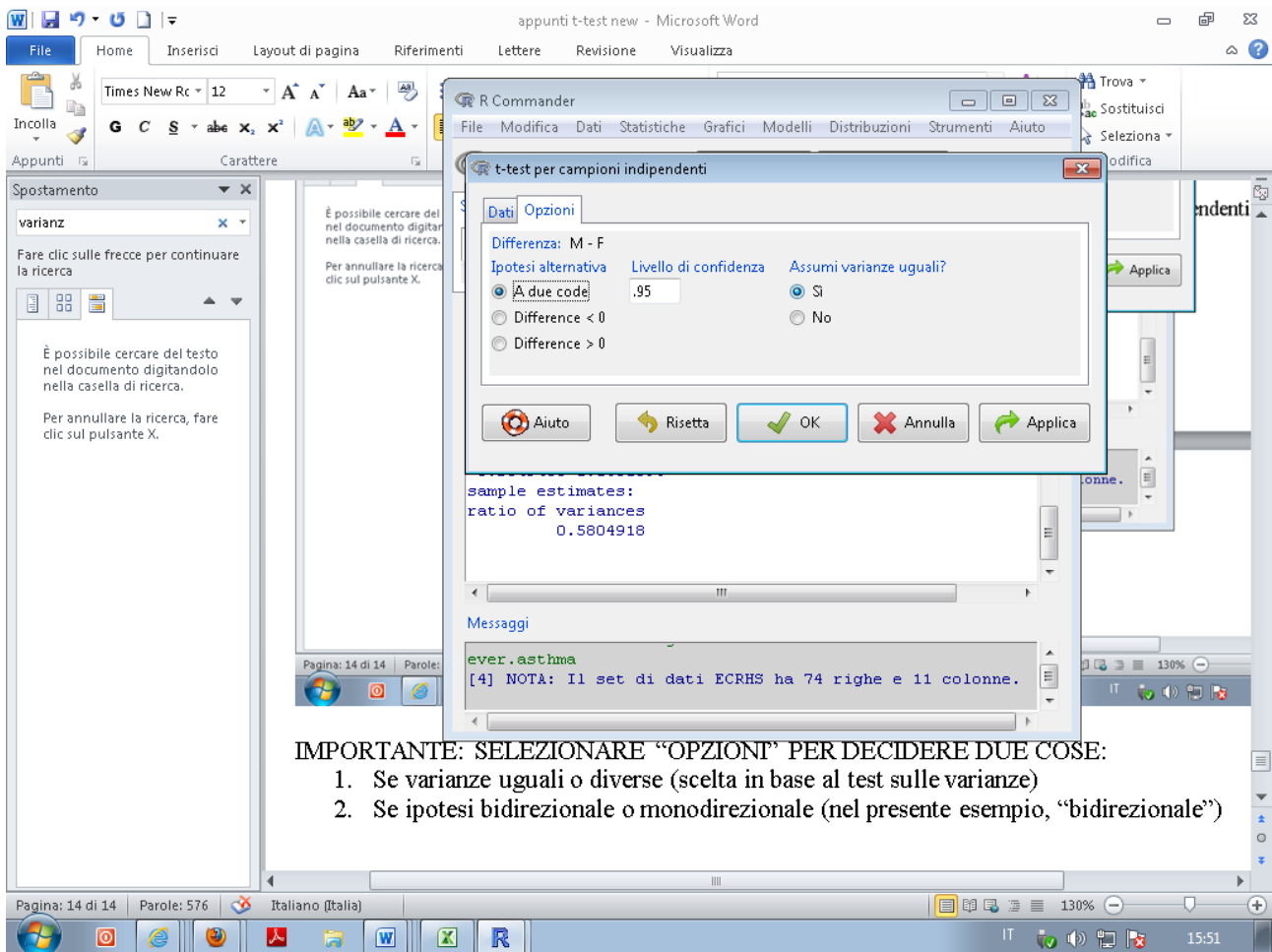
```
sample estimates:
ratio of variances
      0.5804918
```

and a message:

```
[4] NOTA: Il set di dati ECRHS ha 74 righe e 11 colonne.
```

IMPORTANTE: SELEZIONARE “OPZIONI” PER DECIDERE DUE COSE:

1. Se varianze uguali o diverse (scelta in base al test sulle varianze)
2. Se ipotesi bidirezionale o monodirezionale (nel presente esempio, “bidirezionale”)



IMPORTANTE: SELEZIONARE "OPZIONI" PER DECIDERE DUE COSE:
 1. Se varianze uguali o diverse (scelta in base al test sulle varianze)
 2. Se ipotesi bidirezionale o monodirezionale (nel presente esempio, "bidirezionale")

Ecco l'esito:

```
> t.test(weight~sex, alternative='two.sided',
+ conf.level=.95, var.equal=TRUE, data=ECRHS)
```

Two Sample t-test

data: weight by sex

t = 5.3047, df = 65, **p-value = 1.454e-06 (=0,000001454) (valore p calcolato)**

(t = 5.3047 : t calcolato)

df = 65 : gradi di libertà cioè nM + nF -2)

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

95 percent confidence interval:

9.294621 20.518999

sample estimates:

mean in group M mean in group F

77.77778 62.87097

Confronto il valore p calcolato con il valore p critico di 0,05 e rifiuto l'ipotesi nulla se, come in questo caso, il valore calcolato è inferiore a 0,05

Questo significa che il peso medio dei maschi è DIVERSO dal peso medio delle femmine.

T test per campioni appaiati

In questo caso mi trovo nella situazione di confrontare due misure, tra loro comparabili, per le quali dispongo dei dati per tutti i soggetti.

Per esempio potrebbe essere un valore su cui vorrei che incidesse una certa terapia/formazione/intervento, in modo tale da poterlo modificare.

Disponerei dunque del dato prima dell'intervento, e dello stesso dato dopo l'intervento.

Si tratta di variabili di tipo quantitativo, e nel dataset spesso sono indicate come due colonne.

L'esempio che segue è fatto su due variabili del dataset ECRHS, specificamente fev1 e fvc1.

```
> t.test(ECRHS$fev1, ECRHS$fvc1, alternative='less',  
+ conf.level=.95, paired=TRUE)
```

Paired t-test

data: ECRHS\$fev1 and ECRHS\$fvc1

t = -16.635, df = 62, p-value < 2.2e-16 (0,00000000000000022)

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -65.90082

sample estimates:

mean of the differences

-73.25397