Confidence intervals

Marc Brysbaert

Aims

- To show you how you can add confidence intervals to your means, so that you have an idea of the preciseness of your estimates.
- This is straightforward for between-subjects designs; a bit trickier (harder) for repeated measures.

The Central limit theorem

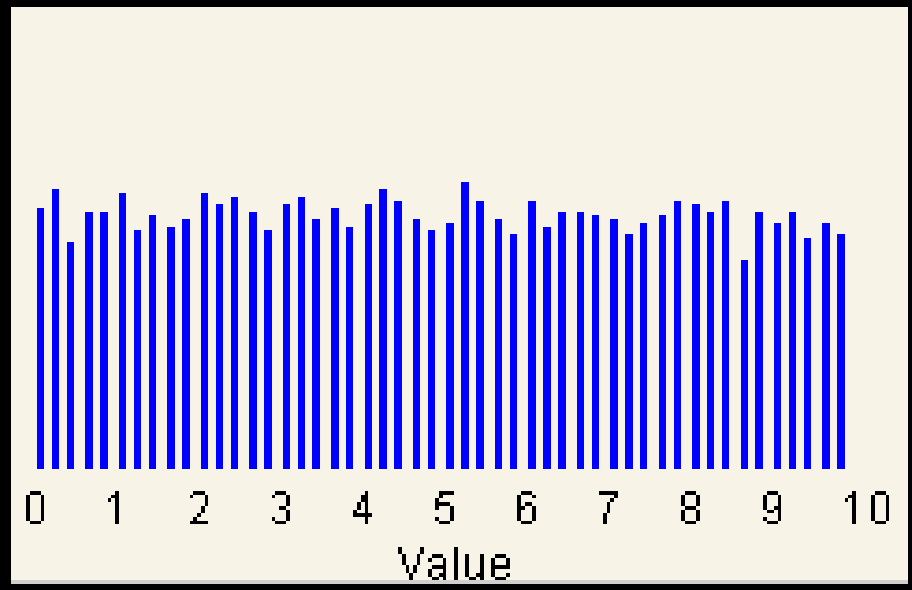
- If we take many samples from a population, the means of the samples are likely to differ from one sample to the next.
- The smaller the sample size n, the larger the variability of the sample means.

The Central limit theorem cont.

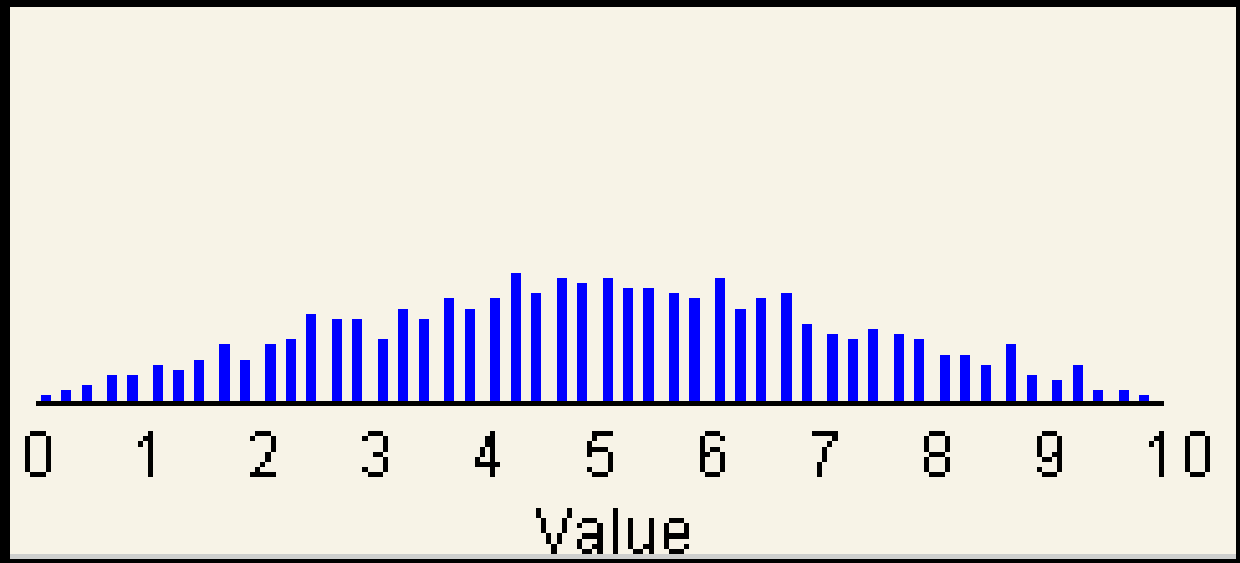
- The standard deviation of the sample mean (the standard error) is related to the standard deviation in the population according to the following equation:

$$SE = \frac{SD}{\sqrt{n}}$$

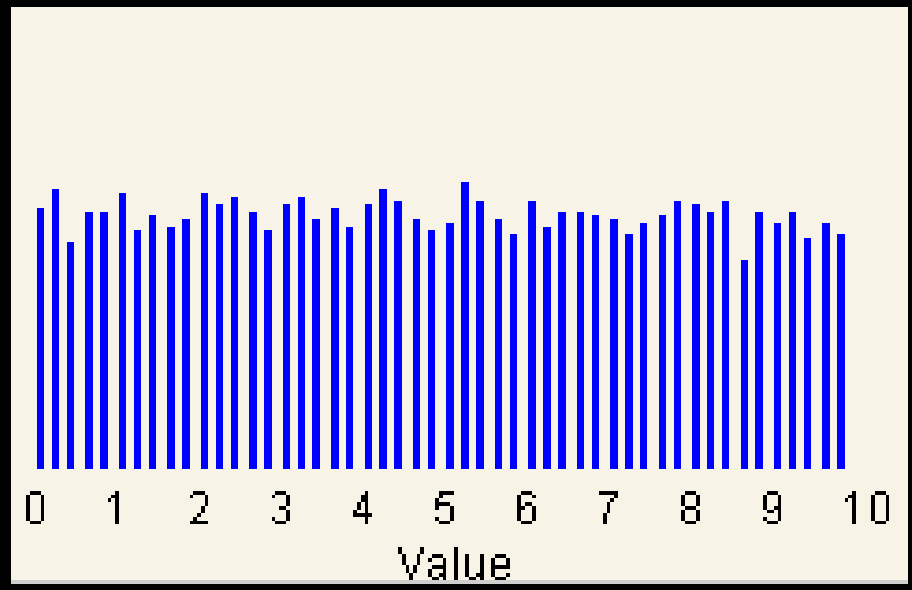
Population



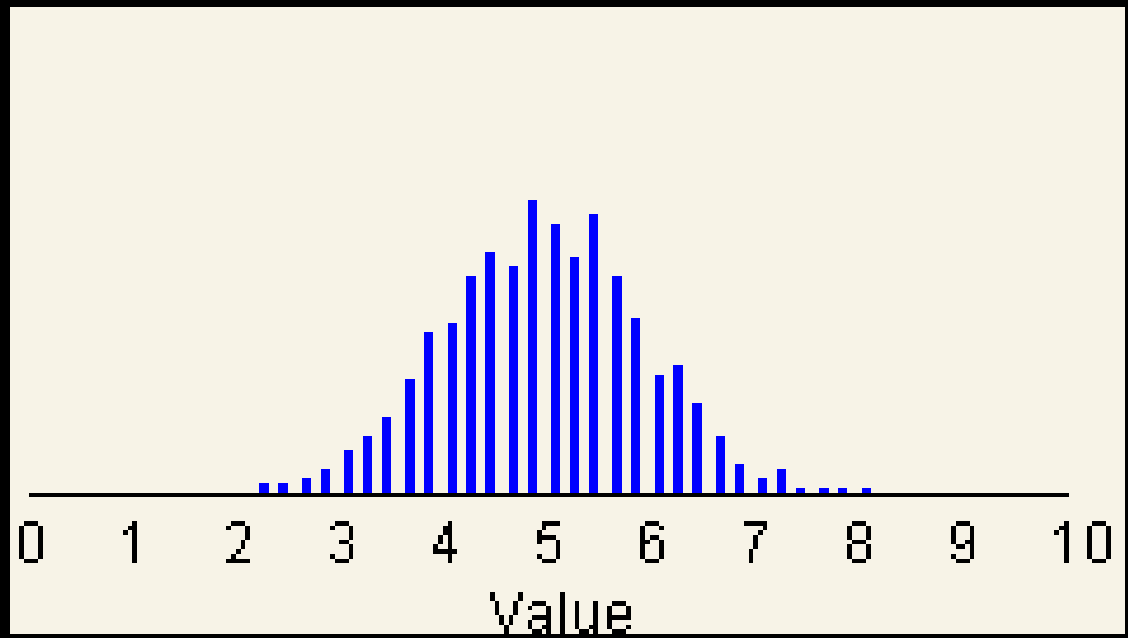
Distribution of sample means
n = 2



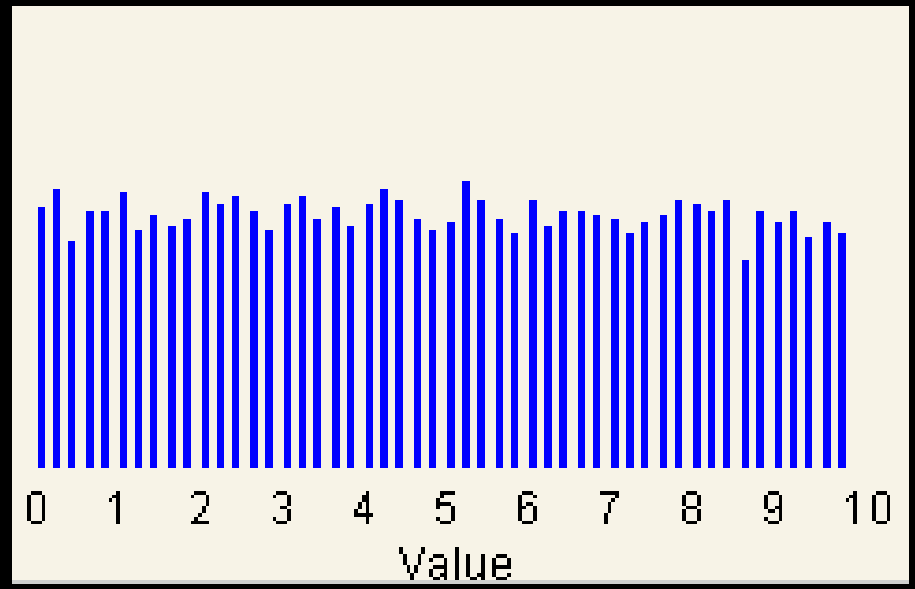
Population



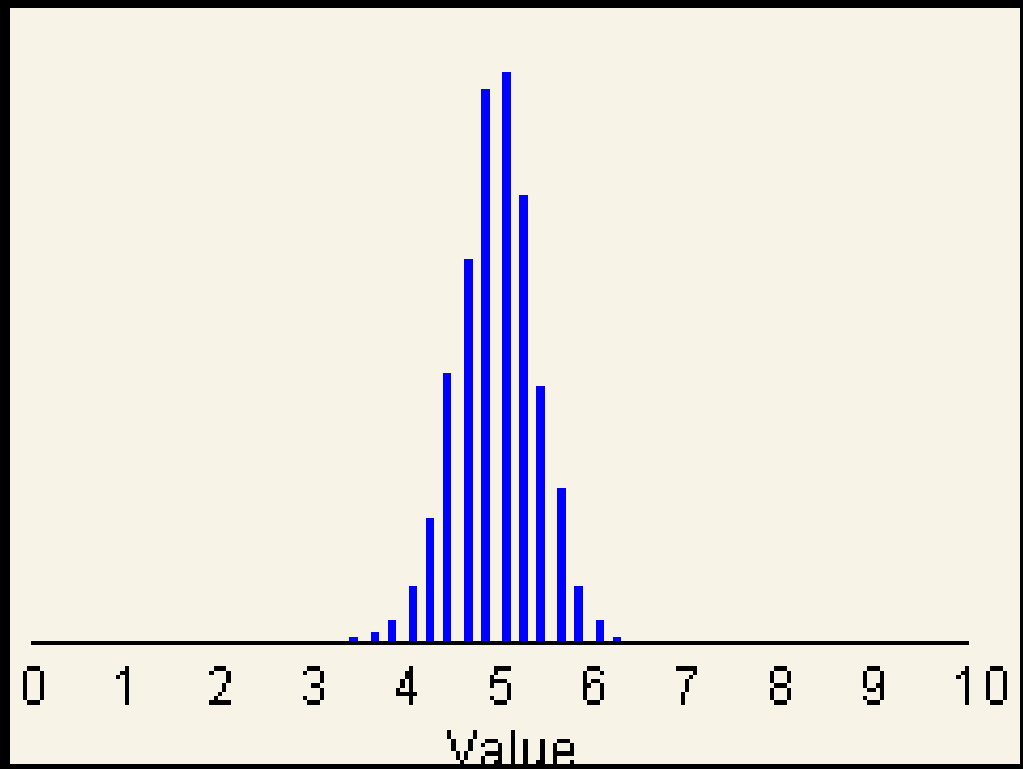
Distribution of sample means
n = 10



Population



Distribution of sample means
n = 50

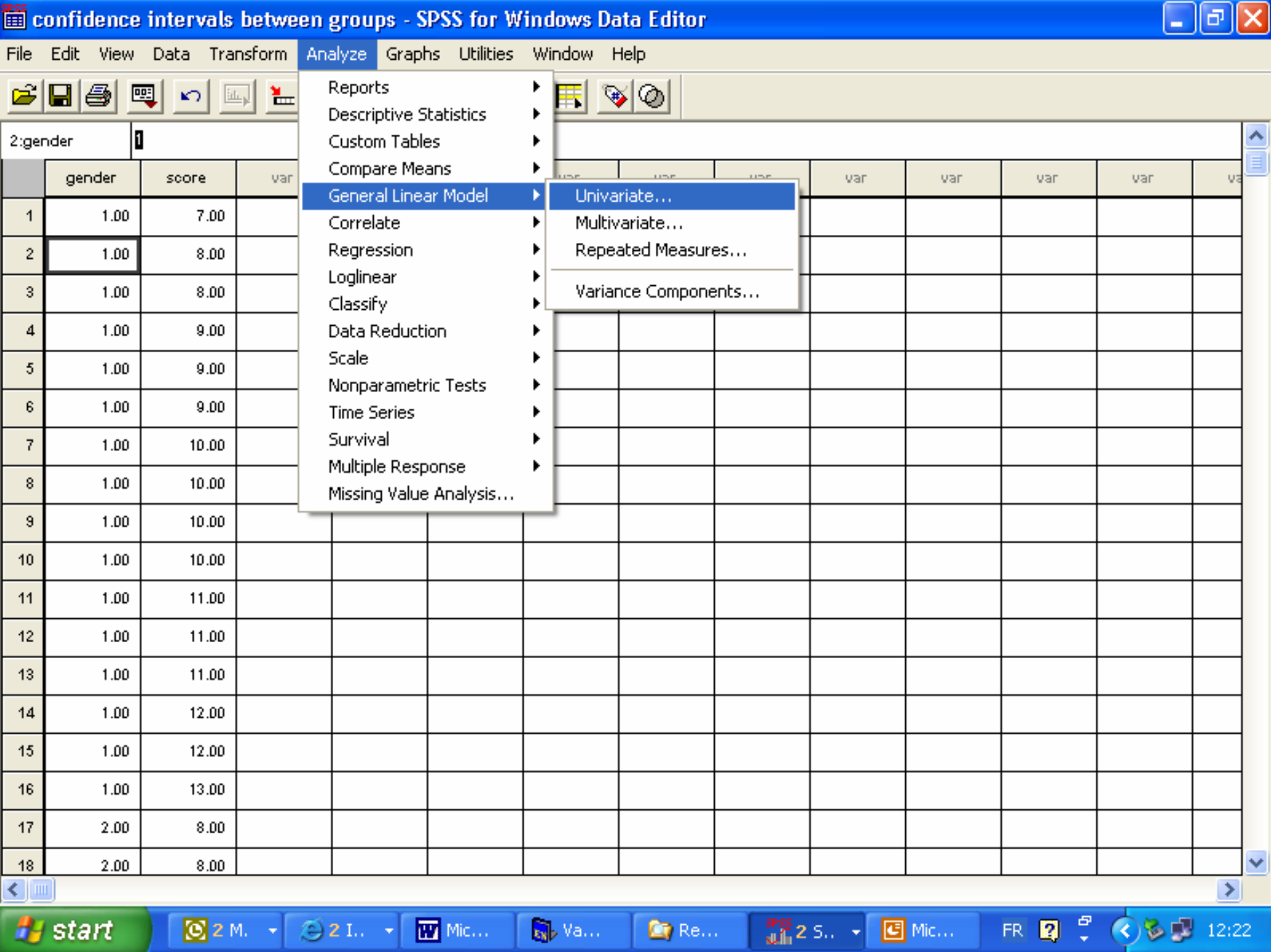


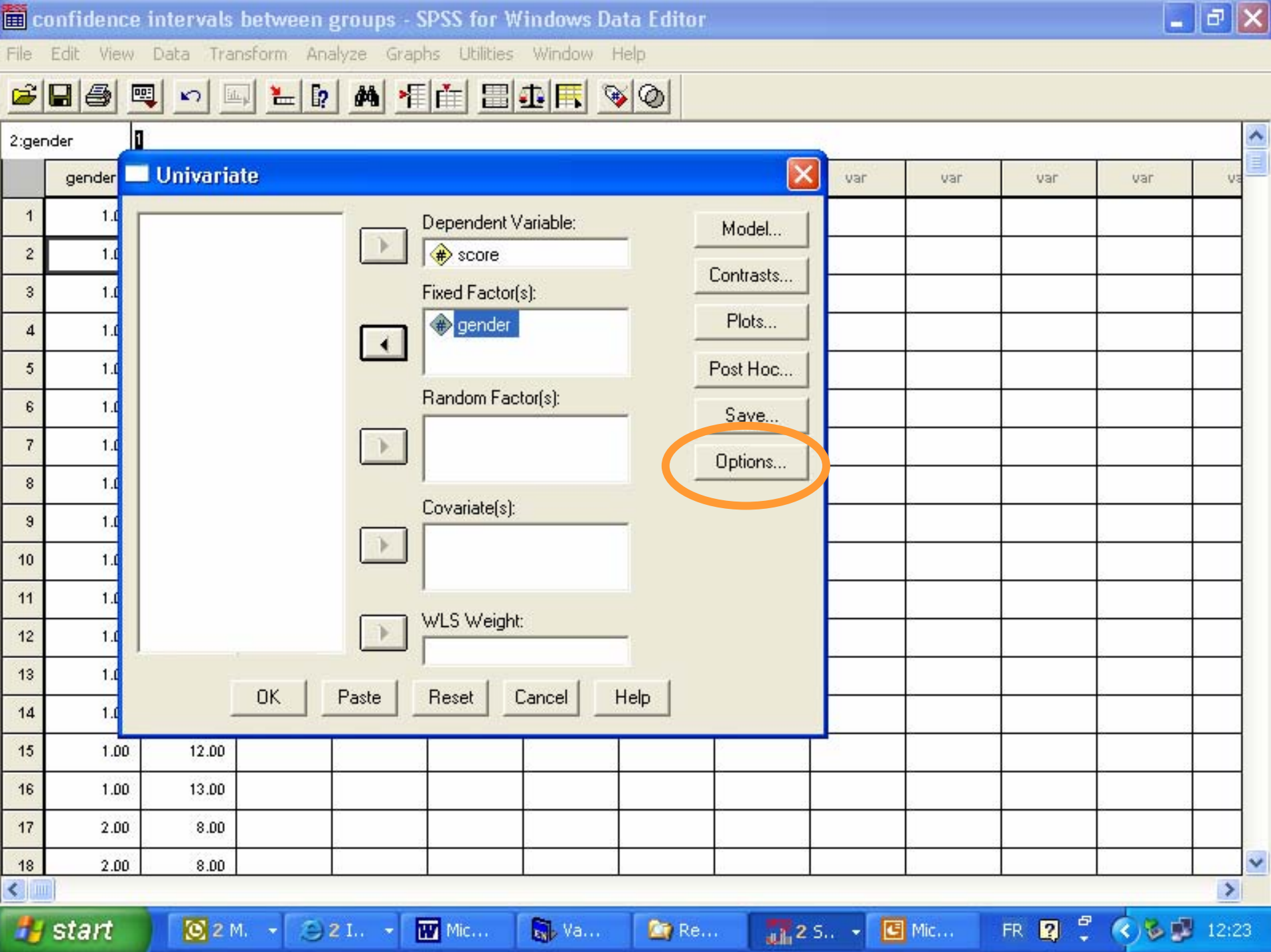
Confidence interval

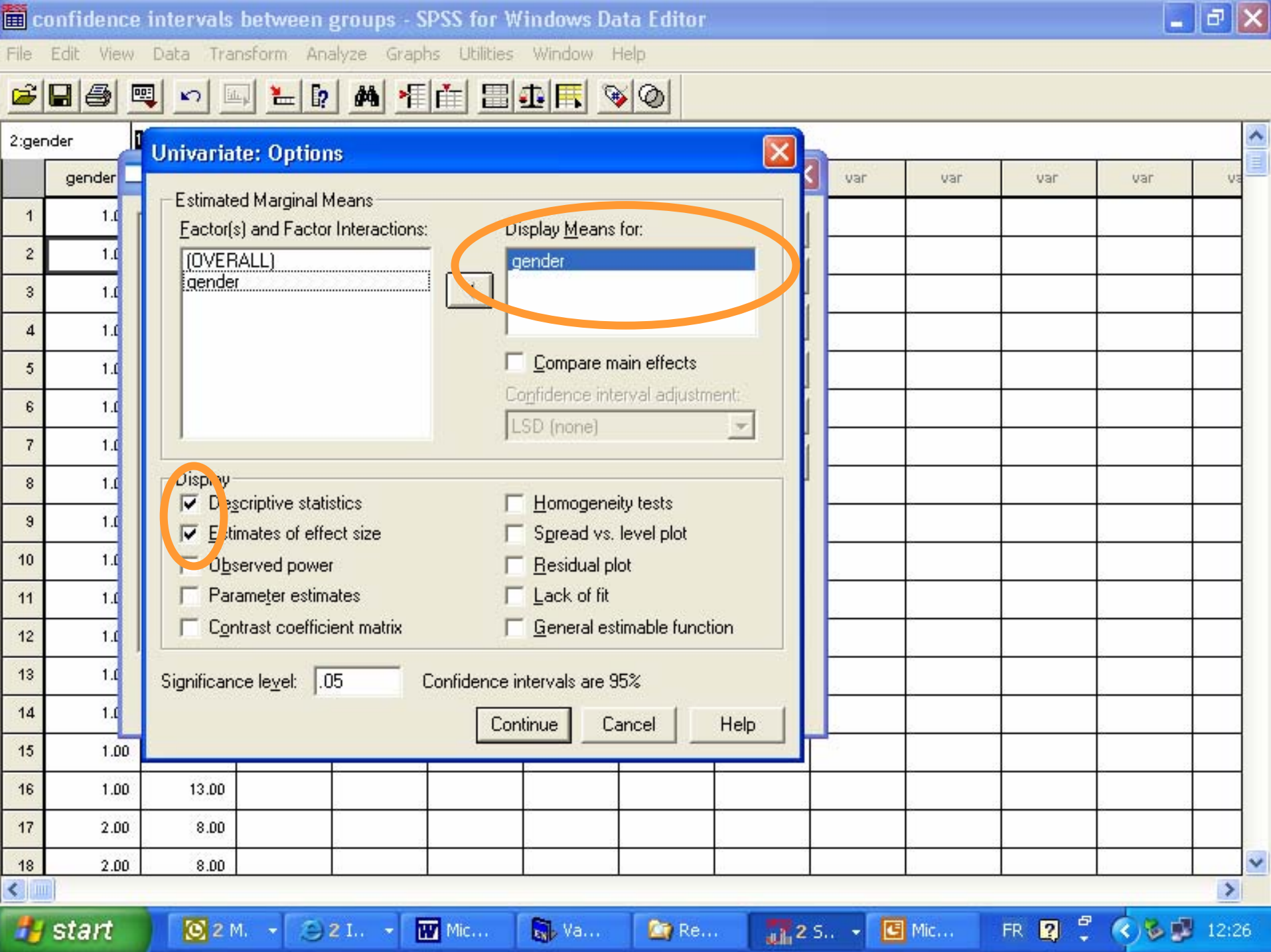
- A confidence interval (CI) is a range of plausible values of the population mean μ. Values outside the CI are relatively implausible.
- The larger the sample size, the smaller the CI.
- We will work with the 95% CI ($\alpha = .05$)

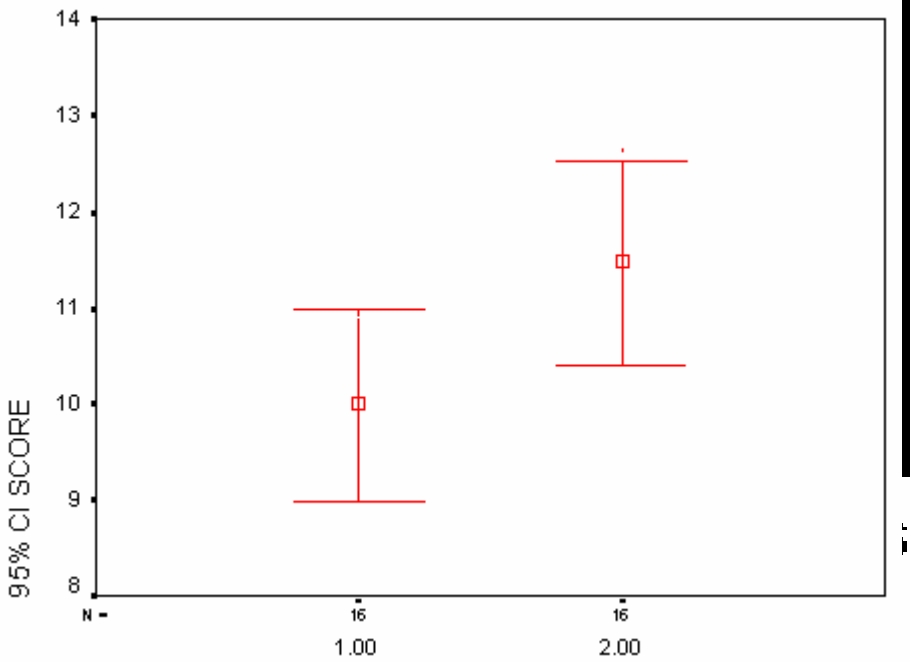
Between-subjects designs

- Two conditions, formed by two different groups of participants (men and women)
- n1 = n2 = 16
- Sample 1 = 7, 8, 8, 9, 9, 9, 10, 10, 10, 10, 11, 11, 11, 12, 12, 13
- Sample 2 = 8, 8, 9, 9, 10, 10, 11, 11, 12, 12, 13, 13, 14, 14, 15, 15


confidence intervals between groups - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Univariate...
Multivariate...
Repeated Measures...
Variance Components...
2:gender
gender score var
1 1.00 7.00
2 1.00 8.00
3 1.00 8.00
4 1.00 9.00
5 1.00 9.00
6 1.00 9.00
7 1.00 10.00
8 1.00 10.00
9 1.00 10.00
10 1.00 10.00
11 1.00 11.00
12 1.00 11.00
13 1.00 11.00
14 1.00 12.00
15 1.00 12.00
16 1.00 13.00
17 2.00 8.00
18 2.00 8.00
start
12:22


confidence intervals between groups - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
2:gender
gender
var
Univariate
Dependent Variable:
score
Fixed Factor(s):
gender
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
OK
Paste
Reset
Cancel
Help
1 1.0
2 1.0
3 1.0
4 1.0
5 1.0
6 1.0
7 1.0
8 1.0
9 1.0
10 1.0
11 1.0
12 1.0
13 1.0
14 1.0
15 1.00 12.00
16 1.00 13.00
17 2.00 8.00
18 2.00 8.00
start
2 M.
2 I..
Mic...
Va...
Re...
2 S..
Mic...
FR
12:23


confidence intervals between groups - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
2:gender
gender
var
Univariate: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
gender
Display Means for:
gender
Compare main effects
Confidence interval adjustment:
LSD (none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
Contrast coefficient matrix
Homogeneity tests
Spread vs. level plot
Residual plot
Lack of fit
General estimable function
Significance level: .05
Confidence intervals are 95%
Continue
Cancel
Help
1.00
2.00
13.00
8.00
start
12:26



$CI = 11.038 - 10$

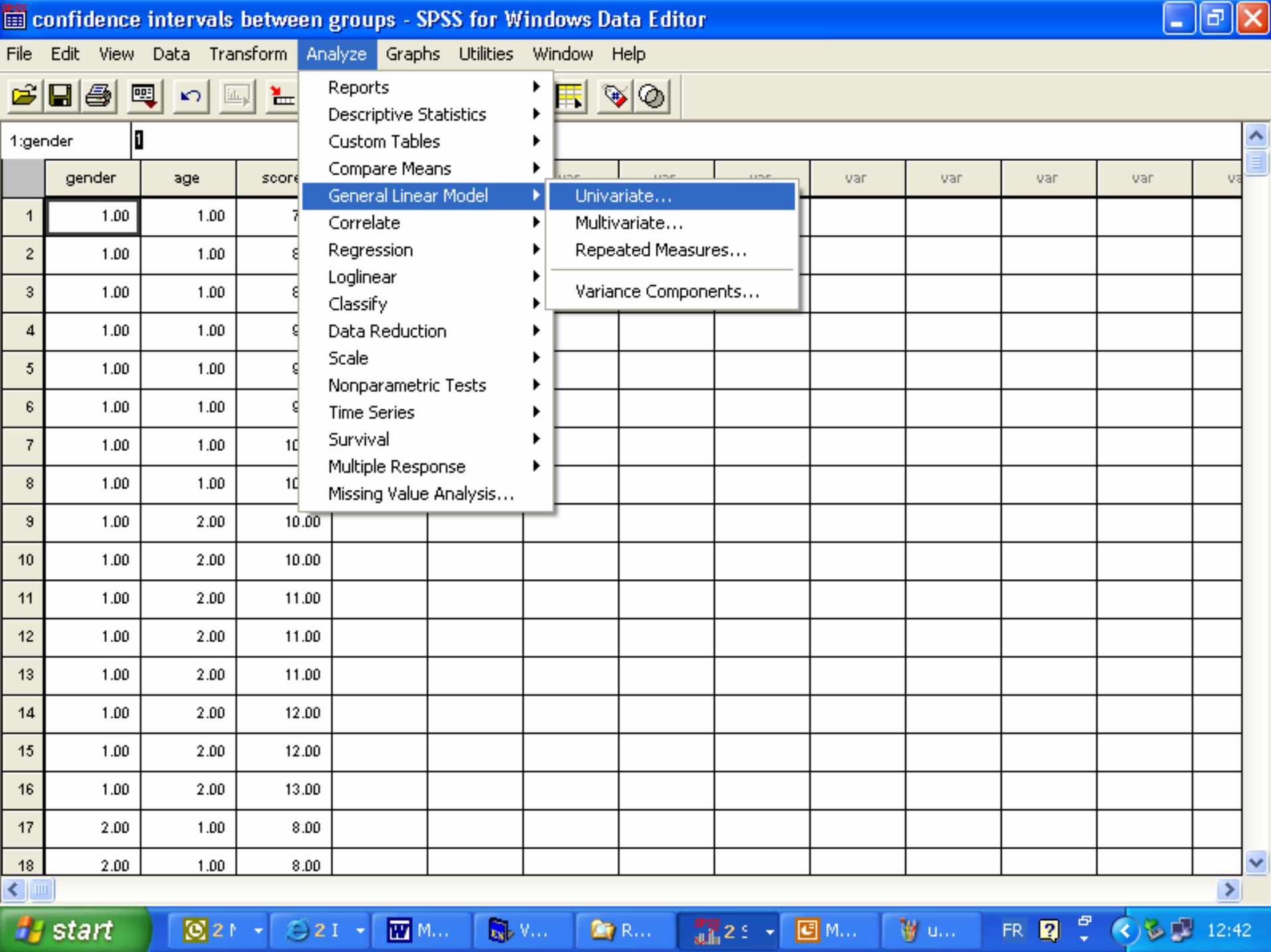
$= 1.038$

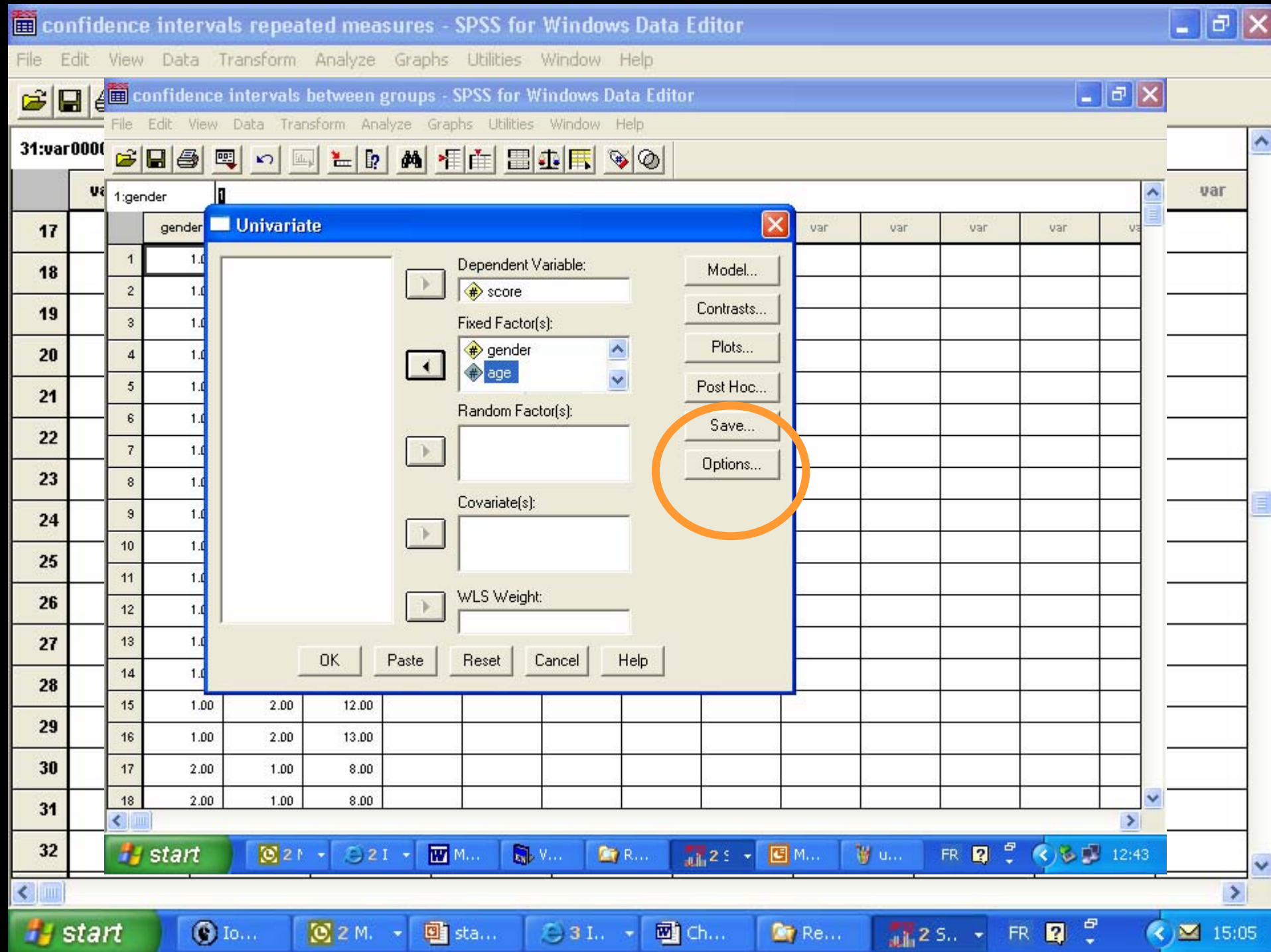
Dependent Variable: SCORE

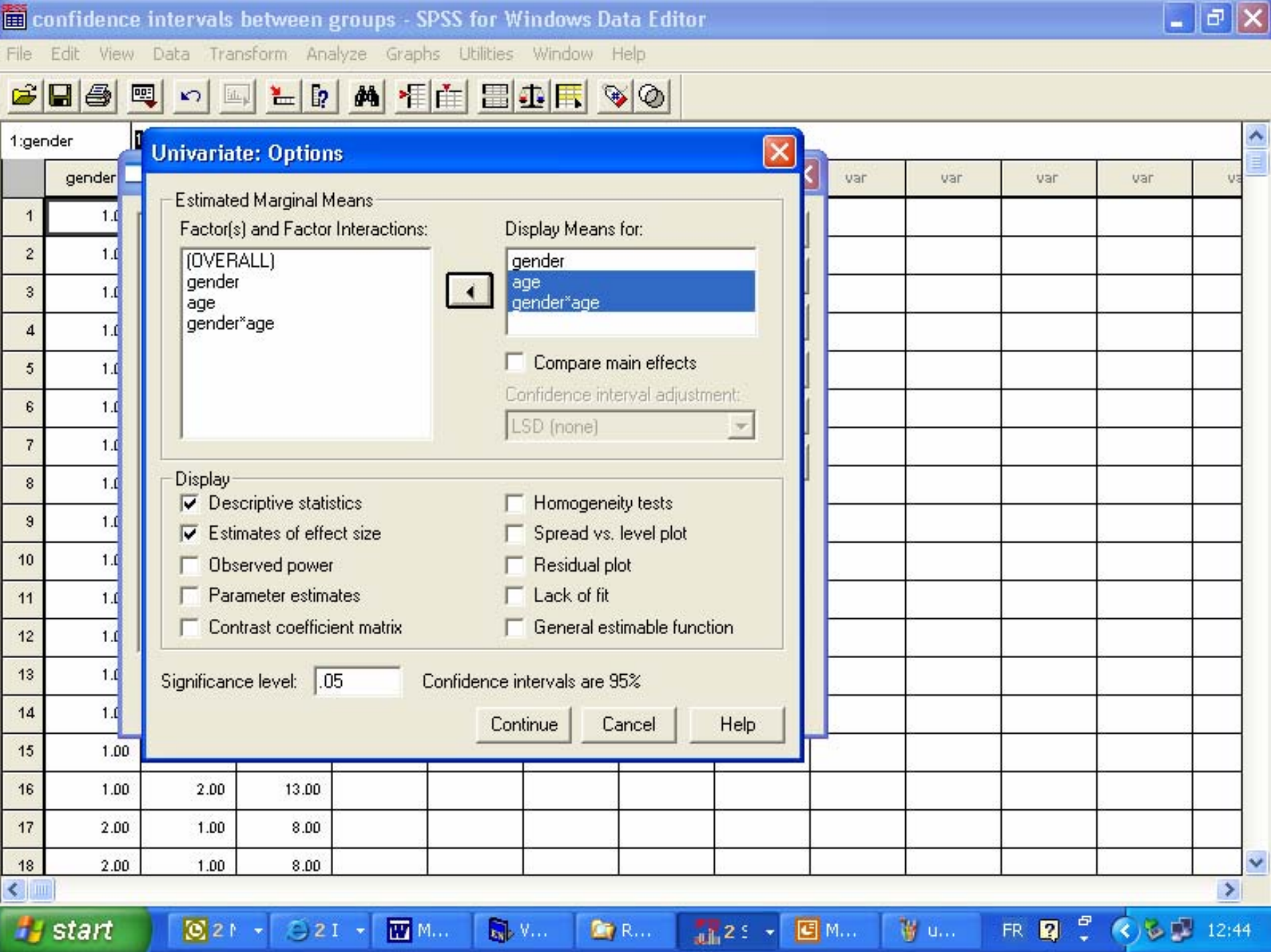
GENDER	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	10.000	.508	8.962	11.038
2.00	11.500	.508	10.462	12.538

2 x 2 between-participants design

- Same approach; you just add a new column with the second variable (e.g., age: young / old)


confidence intervals between groups - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Univariate...
Multivariate...
Repeated Measures...
Variance Components...
1:gender
gender age score
1.00 1.00 7
1.00 1.00 8
1.00 1.00 8
1.00 1.00 9
1.00 1.00 9
1.00 1.00 9
1.00 1.00 10
1.00 1.00 10
1.00 2.00 10.00
1.00 2.00 10.00
1.00 2.00 11.00
1.00 2.00 11.00
1.00 2.00 11.00
1.00 2.00 12.00
1.00 2.00 12.00
1.00 2.00 13.00
2.00 1.00 8.00
2.00 1.00 8.00
var
start
FR
12:42


confidence intervals repeated measures - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
confidence intervals between groups - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
31:var0000
1:gender
gender
Univariate
Dependent Variable:
score
Fixed Factor(s):
gender
age
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
OK
Paste
Reset
Cancel
Help
start
12:43
15:05

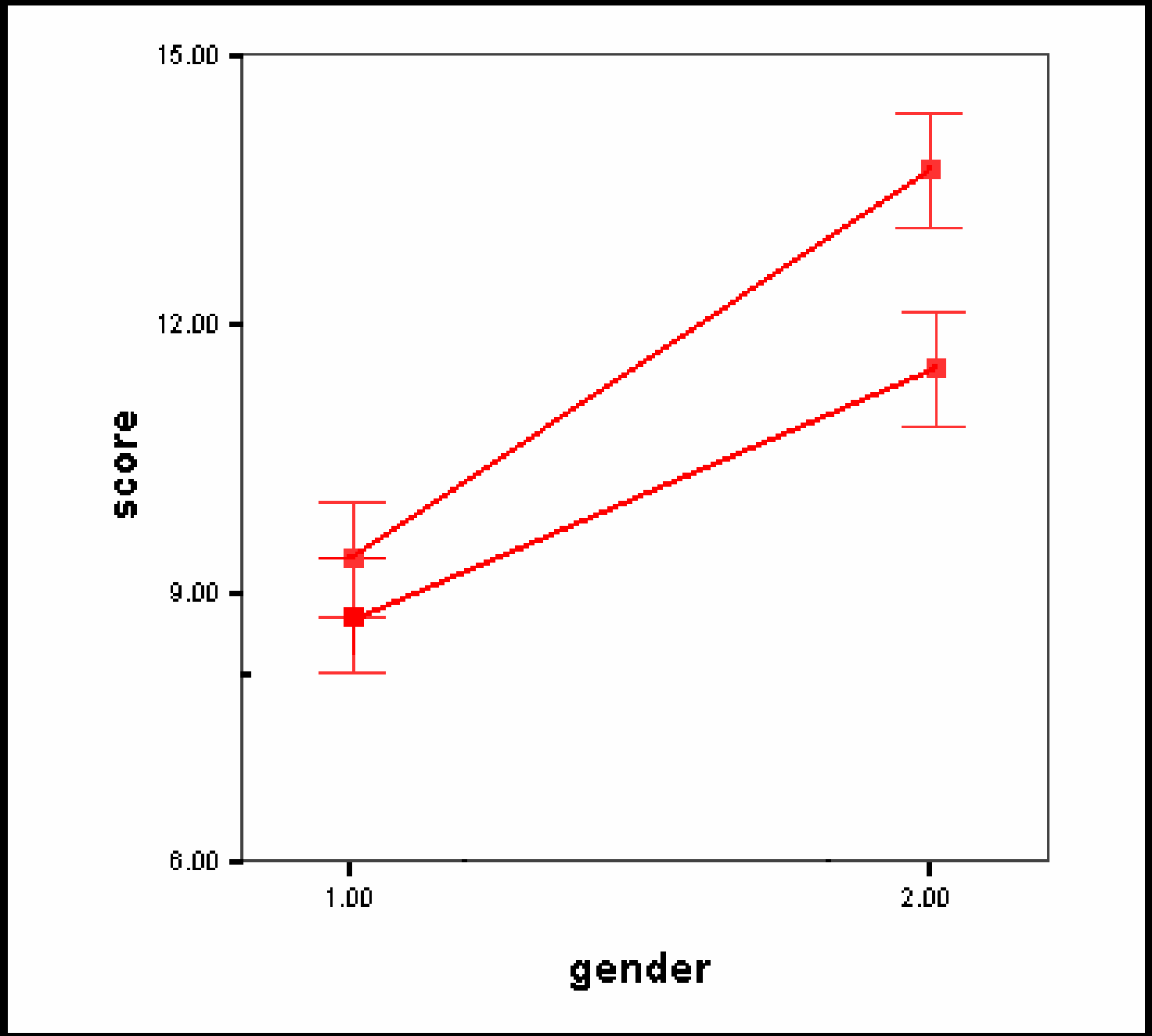

confidence intervals between groups - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:gender
gender
var
Univariate: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
gender
age
gender*age
Display Means for:
gender
age
gender*age
Compare main effects
Confidence interval adjustment:
LSD (none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
Contrast coefficient matrix
Homogeneity tests
Spread vs. level plot
Residual plot
Lack of fit
General estimable function
Significance level: .05
Confidence intervals are 95%
Continue
Cancel
Help
1.00 2.00 13.00
2.00 1.00 8.00
2.00 1.00 8.00
start
FR
12:44

3. GENDER * AGE

Dependent Variable: SCORE

GENDER	AGE	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1.00	1.00	8.750	.395	7.940	9.560
	2.00	11.250	.395	10.440	12.060
2.00	1.00	9.500	.395	8.690	10.310
	2.00	13.500	.395	12.690	14.310

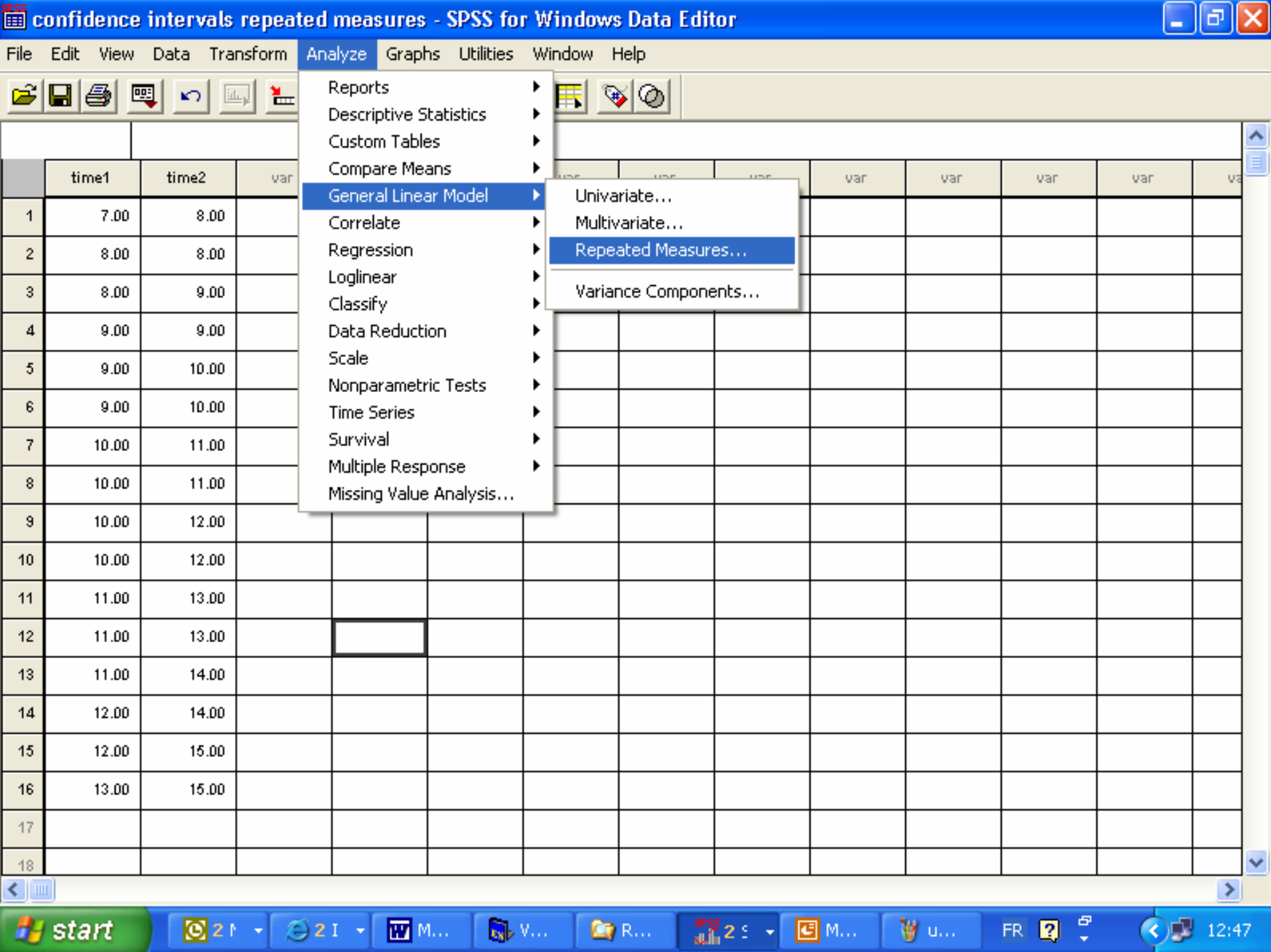
Mean scores as a function of gender and age, together with the 95% Confidence Intervals based on the pooled error term.

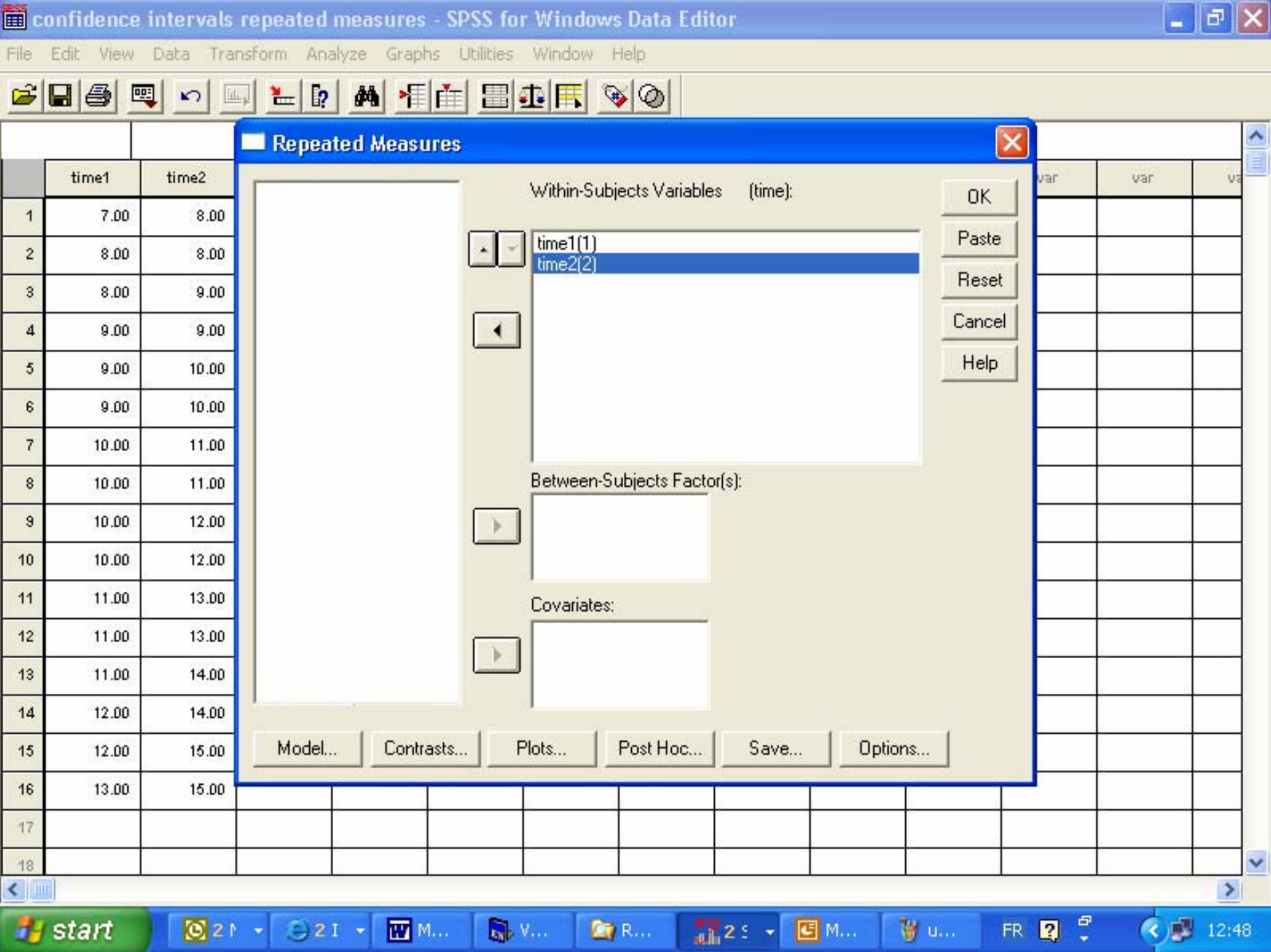

15.00
12.00
9.00
6.00
score
1.00
2.00
gender

Repeated measures

- Much more difficult!
- SPSS gives you the wrong confidence intervals (for between subjects designs)
- So, you will have to do it by hand.
- Based on Masson & Loftus (2003)

  http://web.uvic.ca/psyc/masson/ML.pdf
- Same data as before but now 1 repeated measure variable (time1 and time2; 16 participants)


confidence intervals repeated measures - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Univariate...
Multivariate...
Repeated Measures...
Variance Components...
time1 time2 var
1 7.00 8.00
2 8.00 8.00
3 8.00 9.00
4 9.00 9.00
5 9.00 10.00
6 9.00 10.00
7 10.00 11.00
8 10.00 11.00
9 10.00 12.00
10 10.00 12.00
11 11.00 13.00
12 11.00 13.00
13 11.00 14.00
14 12.00 14.00
15 12.00 15.00
16 13.00 15.00
17
18
start
FR
12:47


confidence intervals repeated measures - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Repeated Measures
Within-Subjects Variables (time):
time1(1)
time2(2)
Between-Subjects Factor(s):
Covariates:
OK
Paste
Reset
Cancel
Help
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
time1
time2
1 7.00 8.00
2 8.00 8.00
3 8.00 9.00
4 9.00 9.00
5 9.00 10.00
6 9.00 10.00
7 10.00 11.00
8 10.00 11.00
9 10.00 12.00
10 10.00 12.00
11 11.00 13.00
12 11.00 13.00
13 11.00 14.00
14 12.00 14.00
15 12.00 15.00
16 13.00 15.00
17
18
start
FR
12:48

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
TIME	Sphericity Assumed	18.000	1	18.000	45.000	.000	.750
	Greenhouse-Geisser	18.000	1.000	18.000	45.000	.000	.750
	Huynh-Feldt	18.000	1.000	18.000	45.000	.000	.750
	Lower-bound	18.000	1.000	18.000	45.000	.000	.750
Error(TIME)	Sphericity Assumed	6.000	15	.400			
	Greenhouse-Geisser	6.000	15.000	.400			
	Huynh-Feldt	6.000	15.000	.400			
	Lower-bound	6.000	15.000	.400			

$$CI = M \pm \sqrt{\frac{MS_{error}}{n_{each\_condition}}} \times t(df_{error})$$

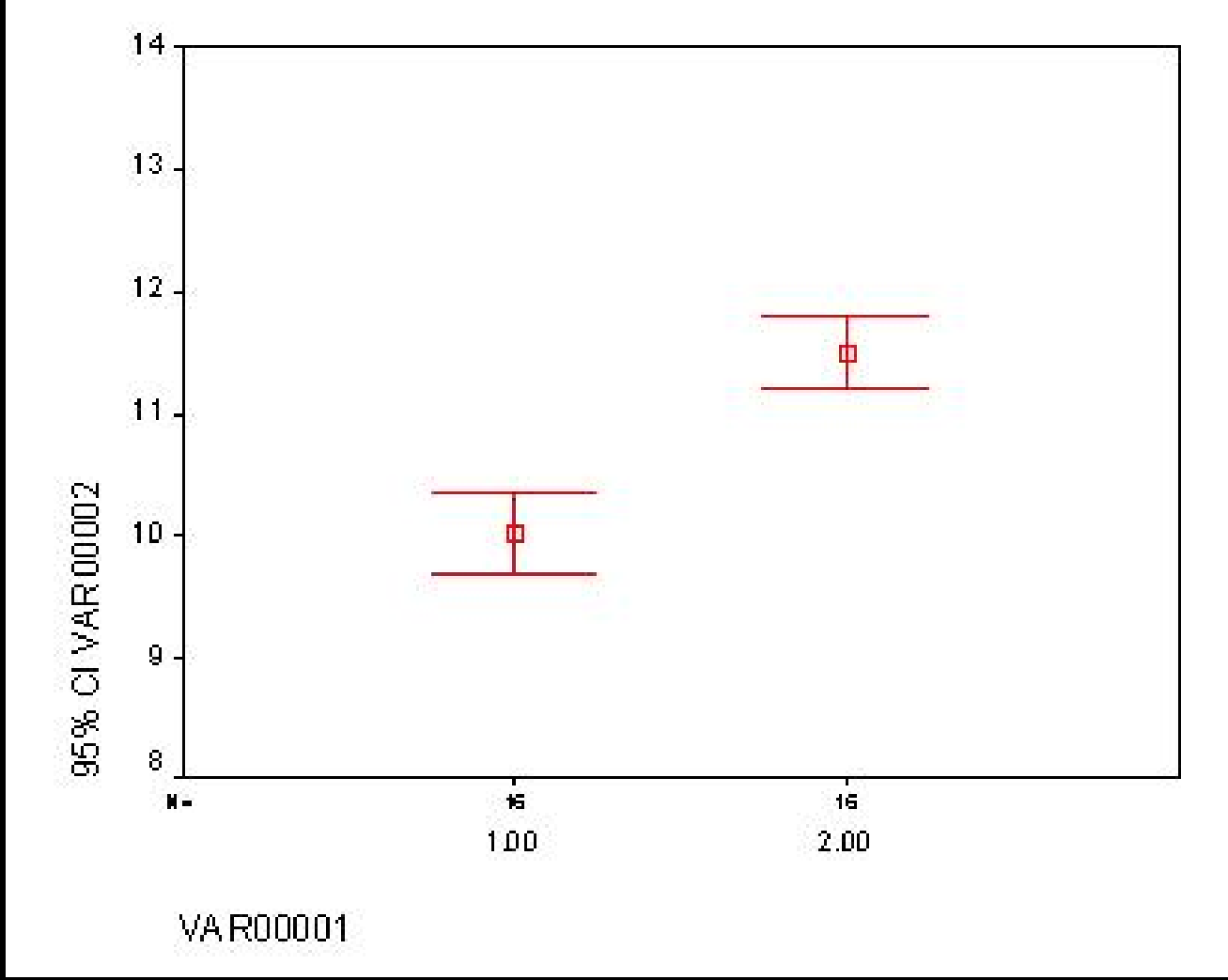
16

$$CI = M \pm \sqrt{\frac{MS_{error}}{n_{each\_condition}}} \times t(df_{error})$$

$$CI = M \pm \sqrt{\frac{.400}{16}} \times t(15)$$

t(15) = critical t-value for p = .05, two-tailed = 2.131 (Stangor, p 398)

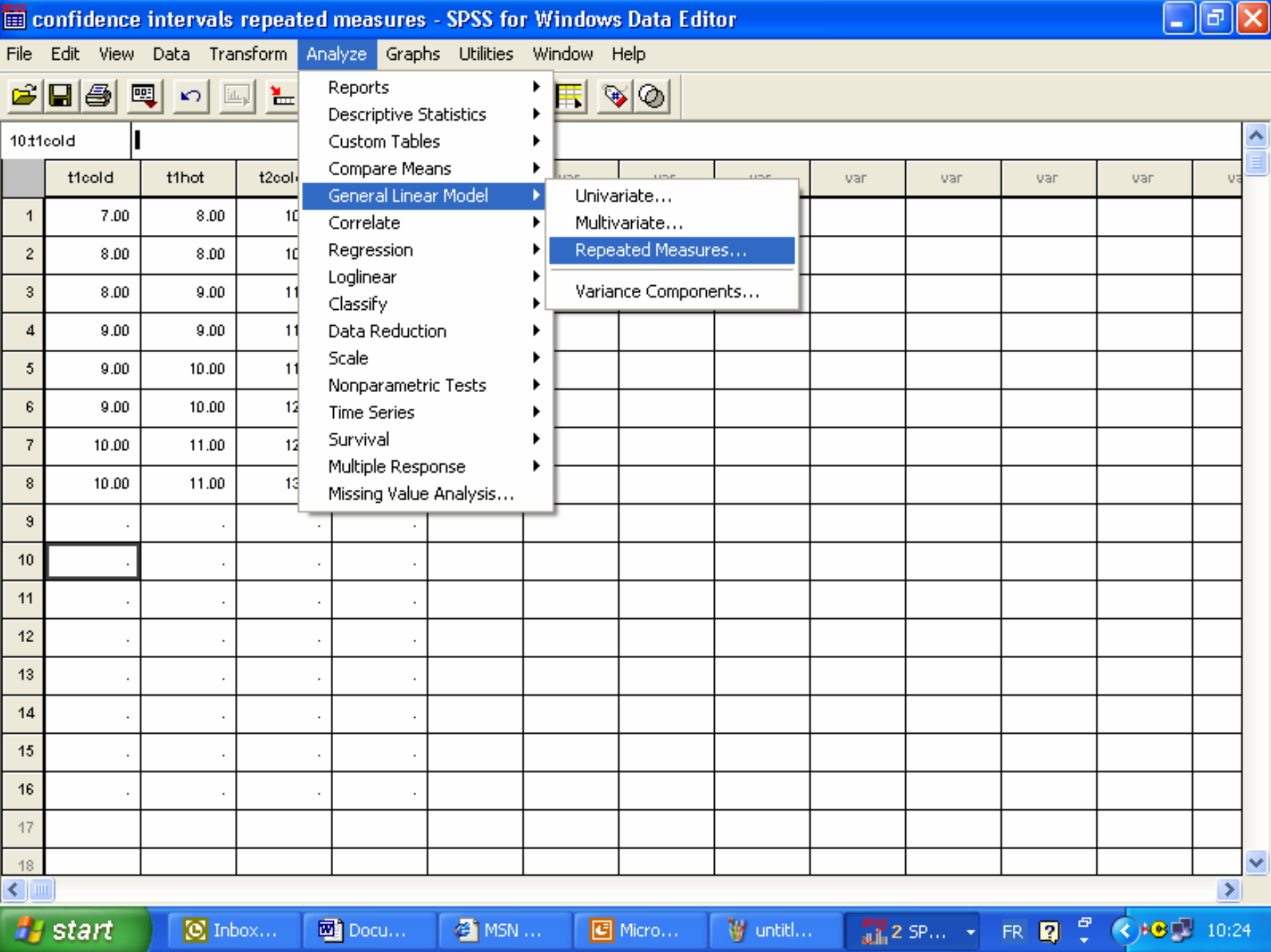
$$CI = M \pm \sqrt{\frac{.400}{16}} \times 2.131 = M \pm .337$$

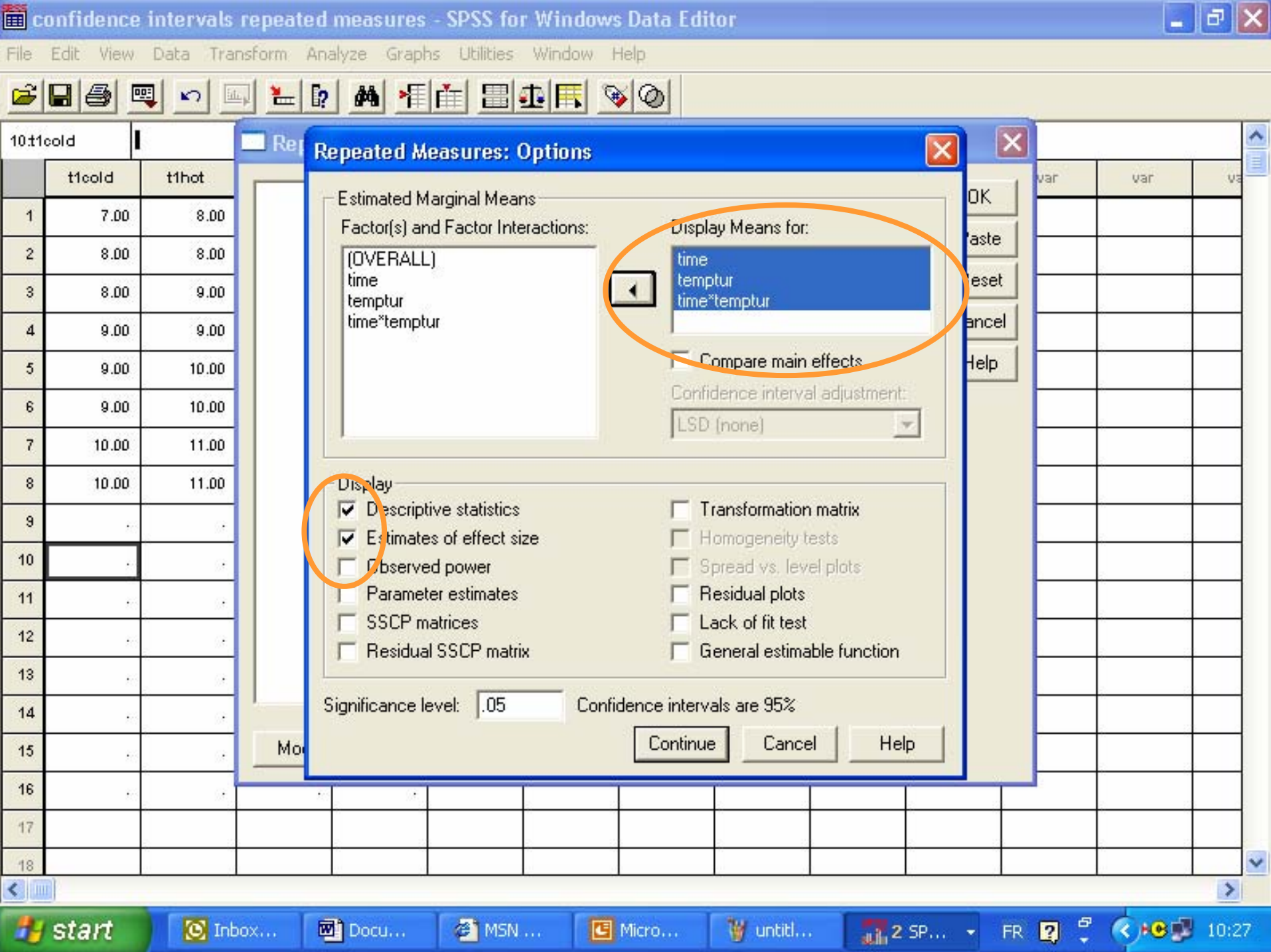


Mean scores as a function of time, together with the 95% Confidence Intervals based on the error term of the repeated measures design.

2x2 repeated measures design

- 8 participants are tested under 4 conditions:
  - Time 1 (morning): cold room
  - Time 1 (morning): hot room
  - Time 2 (evening): cold room
  - Time 2 (evening): hot room


confidence intervals repeated measures - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Univariate...
Multivariate...
Repeated Measures...
Variance Components...
10:t1cold
t1cold t1hot t2col
7.00 8.00
8.00 8.00
8.00 9.00
9.00 9.00
9.00 10.00
9.00 10.00
10.00 11.00
10.00 11.00
var
start
Inbox...
Docu...
MSN ...
Micro...
untitl...
2 SP...
FR
10:24


confidence intervals repeated measures - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
10:t1cold
t1cold t1hot
1 7.00 8.00
2 8.00 8.00
3 8.00 9.00
4 9.00 9.00
5 9.00 10.00
6 9.00 10.00
7 10.00 11.00
8 10.00 11.00
Repeated Measures: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
time
temptur
time*temptur
Display Means for:
time
temptur
time*temptur
Compare main effects
Confidence interval adjustment:
LSD (none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
SSCP matrices
Residual SSCP matrix
Transformation matrix
Homogeneity tests
Spread vs. level plots
Residual plots
Lack of fit test
General estimable function
Significance level: .05
Confidence intervals are 95%
Continue
Cancel
Help
OK
Paste
Reset
Cancel
Help
start
Inbox...
Docu...
MSN ...
Micro...
untitl...
2 SP...
FR
10:27

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
TIME	Sphericity Assumed	84.500	1	84.500	1183.000	.000	.994
	Greenhouse-Geisser	84.500	1.000	84.500	1183.000	.000	.994
	Huynh-Feldt	84.500	1.000	84.500	1183.000	.000	.994
	Lower-bound	84.500	1.000	84.500	1183.000	.000	.994
Error(TIME)	Sphericity Assumed	.500	7	7.143E-02			
	Greenhouse-Geisser	.500	7.000	7.143E-02			
	Huynh-Feldt	.500	7.000	7.143E-02			
	Lower-bound	.500	7.000	7.143E-02			
TEMPTUR	Sphericity Assumed	18.000	1	18.000	126.000	.000	.947
	Greenhouse-Geisser	18.000	1.000	18.000	126.000	.000	.947
	Huynh-Feldt	18.000	1.000	18.000	126.000	.000	.947
	Lower-bound	18.000	1.000	18.000	126.000	.000	.947
Error(TEMPTUR)	Sphericity Assumed	1.000	7	.143			
	Greenhouse-Geisser	1.000	7.000	.143			
	Huynh-Feldt	1.000	7.000	.143			
	Lower-bound	1.000	7.000	.143			
TIME * TEMPTUR	Sphericity Assumed	4.500	1	4.500	63.000	.000	.900
	Greenhouse-Geisser	4.500	1.000	4.500	63.000	.000	.900
	Huynh-Feldt	4.500	1.000	4.500	63.000	.000	.900
	Lower-bound	4.500	1.000	4.500	63.000	.000	.900
Error(TIME*TEMPTUR)	Sphericity Assumed	.500	7	7.143E-02			
	Greenhouse-Geisser	.500	7.000	7.143E-02			
	Huynh-Feldt	.500	7.000	7.143E-02			
	Lower-bound	.500	7.000	7.143E-02			

$$CI = M \pm \sqrt{\frac{MS_{error}}{n_{each\_condition}}} \times t(df_{error})$$

MSerror and dferror must be estimated on the basis of the three error terms in the ANOVA

$$MS_{error} = \frac{SS_{error1} + SS_{error2} + SS_{error3}}{df_1 + df_2 + df_3}$$

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
TIME	Sphericity Assumed	84.500	1	84.500	1183.000	.000	.994
	Greenhouse-Geisser	84.500	1.000	84.500	1183.000	.000	.994
	Huynh-Feldt	84.500	1.000	84.500	1183.000	.000	.994
	Lower-bound	84.500	1.000	84.500	1183.000	.000	.994
Error(TIME)	Sphericity Assumed	.500	7	7.143E-02			
	Greenhouse-Geisser	.500	7.000	7.143E-02			
	Huynh-Feldt	.500	7.000	7.143E-02			
	Lower-bound	.500	7.000	7.143E-02			
TEMPTUR	Sphericity Assumed	18.000	1	18.000	126.000	.000	.947
	Greenhouse-Geisser	18.000	1.000	18.000	126.000	.000	.947
	Huynh-Feldt	18.000	1.000	18.000	126.000	.000	.947
	Lower-bound	18.000	1.000	18.000	126.000	.000	.947
Error(TEMPTUR)	Sphericity Assumed	1.000	7	.143			
	Greenhouse-Geisser	1.000	7.000	.143			
	Huynh-Feldt	1.000	7.000	.143			
	Lower-bound	1.000	7.000	.143			
TIME * TEMPTUR	Sphericity Assumed	4.500	1	4.500	63.000	.000	.900
	Greenhouse-Geisser	4.500	1.000	4.500	63.000	.000	.900
	Huynh-Feldt	4.500	1.000	4.500	63.000	.000	.900
	Lower-bound	4.500	1.000	4.500	63.000	.000	.900
Error(TIME*TEMPTUR)	Sphericity Assumed	.500	7	7.143E-02			
	Greenhouse-Geisser	.500	7.000	7.143E-02			
	Huynh-Feldt	.500	7.000	7.143E-02			
	Lower-bound	.500	7.000	7.143E-02			

$$MS_{error} = \frac{SS_{error1} + SS_{error2} + SS_{error3}}{df_1 + df_2 + df_3}$$

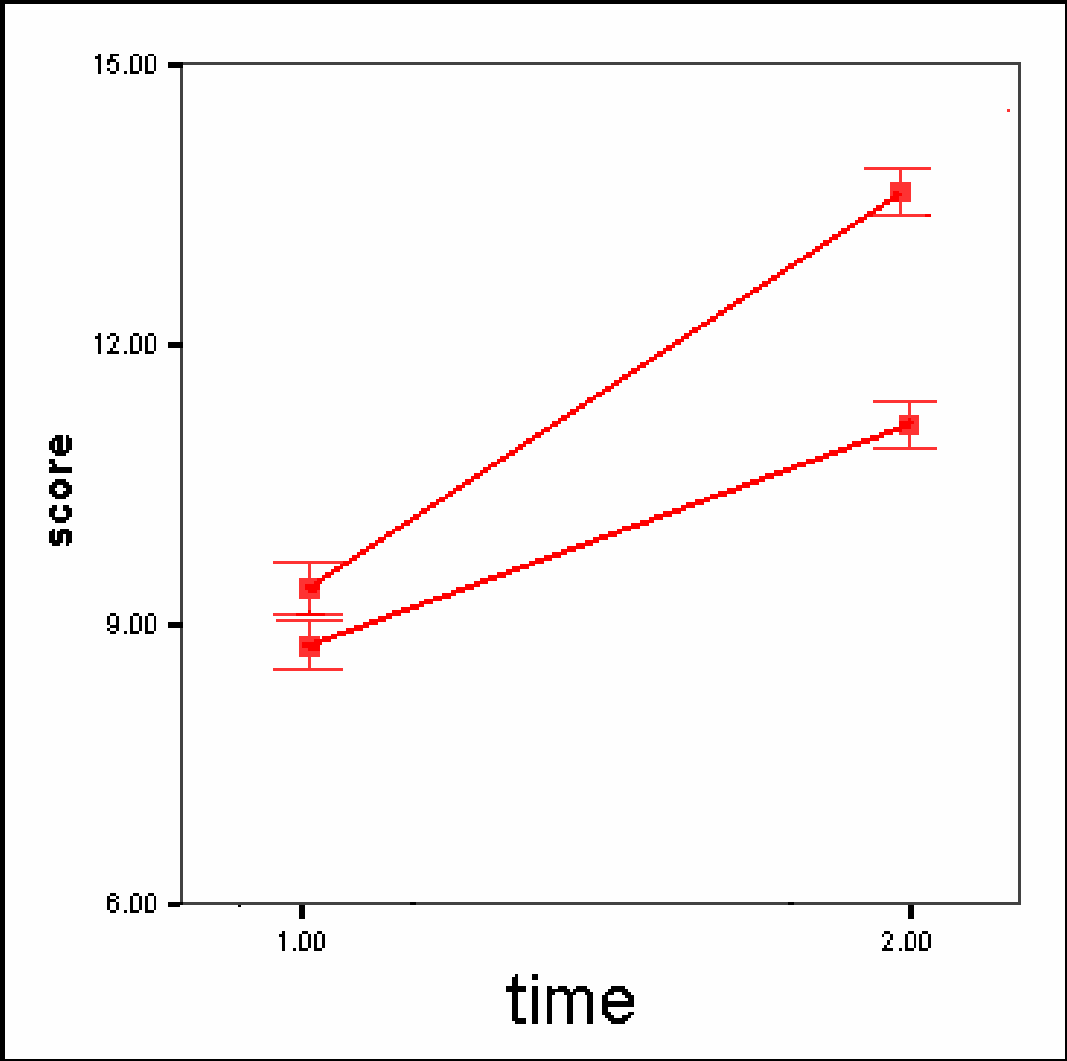
$$MS_{error} = \frac{1 + .5 + .5}{7 + 7 + 7} = .095$$

$$CI = M \pm \sqrt{\frac{MS_{error}}{n_{each\_condition}}} \times t(df_{error})$$

$$CI = M \pm \sqrt{\frac{MS_{error}}{n_{each\_condition}}} \times t(df_{error})$$

$$CI = M \pm \sqrt{\frac{.095}{8}} \times t(21)$$

$$CI = M \pm \sqrt{\frac{.095}{8}} \times 2.080 = .227$$



Mean scores as a function of time and surrounding, together with the 95% Confidence Intervals based on the pooled error terms of the repeated measures design.

2x2 mixed design

- Between-participants variable: gender
- Repeated measure: time
- 16 participants (8 of each gender)
- TRICKY!!! : you get different confidence intervals for the between-participants variable on the one hand and the within- and interaction effects on the other hand

2x2 mixed design

- Between-participants (gender): you can use the confidence intervals given by SPSS
- Within and interaction:
  - You work with the error term given in the SPSS table and do the same as for the repeated measures design

1. GENDER

Measure: MEASURE_1

GENDER	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	9.125	.387	8.296	9.954
2.00	12.375	.387	11.546	13.204

$$CI = 9.954 - 9.125$$

$$= .829$$

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
TIME	Sphericity Assumed	18.000	1	18.000	168.000	.000	.923
	Greenhouse-Geisser	18.000	1.000	18.000	168.000	.000	.923
	Huynh-Feldt	18.000	1.000	18.000	168.000	.000	.923
	Lower-bound	18.000	1.000	18.000	168.000	.000	.923
TIME * GENDER	Sphericity Assumed	4.500	1	4.500	42.000	.000	.750
	Greenhouse-Geisser	4.500	1.000	4.500	42.000	.000	.750
	Huynh-Feldt	4.500	1.000	4.500	42.000	.000	.750
	Lower-bound	4.500	1.000	4.500	42.000	.000	.750
Error(TIME)	Sphericity Assumed	1.500	14	.107			
	Greenhouse-Geisser	1.500	14.000	.107			
	Huynh-Feldt	1.500	14.000	.107			
	Lower-bound	1.500	14.000	.107			

$$CI = M \pm \sqrt{\frac{MS_{error}}{n_{each\_condition}}} \times t(df_{error})$$

$$= \sqrt{\frac{.107}{8}} \times t(14) = .116 \times 2.145 = .248$$

Conclusion
Confidence intervals mixed design

- Between-variable: CI = .829 (to be used when you talk about the main effect of the gender)
- Repeated-measure: CI = .248 (to be used when you talk about the main effect of time or the interaction between time and gender)