Analysis of Variance: repeated measures

Tests for comparing three or more groups or conditions:

(a) Nonparametric tests:
Independent measures: Kruskal-Wallis.

Repeated measures: Friedman's.

(b) Parametric tests:
One-way independent-measures Analysis of Variance (ANOVA).

One-way repeated-measures ANOVA.

Logic behind ANOVA:

ANOVA compares the amount of systematic variation (from our experimental manipulations) to the amount of random variation (from the participants themselves) to produce an *F*-ratio:

$$F = \frac{\text{systematic variation}}{\text{random variation ("error")}}$$

$$F = \frac{\text{systematic variation}}{\text{random variation ("error")}}$$

Large value of *F*: a lot of the overall variation in scores is due to the experimental manipulation, rather than to random variation between participants.

Small value of *F*: the variation in scores produced by the experimental manipulation is small, compared to random variation between participants.

ANOVA is based on the *variance* of the scores. The variance is the standard deviation squared:

$$\boxed{\text{variance}} = \frac{\sum (X - \overline{X})^2}{N}$$

In practice, we use only the top line of the variance formula (the *"Sum of Squares"*, or "*SS*"):

$$\text{sum of squares} = \sum (X - \overline{X})^2$$

We divide this by the appropriate "*degrees of freedom*" (usually the number of groups or participants minus 1).

One-way Repeated-Measures ANOVA:

Use this where you have:
(a) *one* independent variable (with 2 or more levels);

(b) *one* dependent variable;

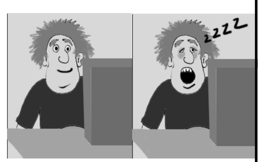
(c) each participant participates in *every* condition in the experiment (repeated measures).

A one-way repeated-measures ANOVA is equivalent to a repeated-measures *t*-test, except that you have more than two conditions in the study.

Effects of sleep-deprivation on vigilance in air-traffic controllers:

No deprivation vs. 12 hours' deprivation:

One Independent Variable, 2 levels – use repeated-measures *t*-test.



No deprivation vs. 12 hours vs. 24 hours:

One Independent Variable, 3 levels (differing quantitatively) – use one-way repeated-measures ANOVA.

Effects of sleep deprivation on vigilance:
Independent Variable: length of sleep deprivation (0, 12 hours and 24 hours). Dependent Variable: 1 hour vigilance test (number of planes missed).
Each participant does all 3 conditions, in a random order.

Participant	0 hours	12 hours	24 hours
1	3	12	13
2	5	15	14
3	6	16	16
4	4	11	12
5	7	12	11
6	3	13	14
7	4	17	16
8	5	11	12
9	6	10	11
10	3	13	14

0 hours:
Mean = 4.6
standard deviation = 1.43.

12 hours:
Mean = 13.0
standard deviation = 2.31.

24 hours:
Mean = 13.3
standard deviation = 1.83.

"Partitioning the variance" in a one-way repeated-measures ANOVA:

Total SS

Between Subjects SS

(usually uninteresting: if it's large, it just shows that subjects differ from each other overall)

Within Subjects SS

SS Experimental (*systematic* within-subjects variation that reflects our experimental manipulation)

SS Error (*unsystematic* within-subjects variation that's not due to our experimental manipulation)

Another look at the table: *Effects of sleep deprivation on vigilance*

between subjects variability

within subjects variability

Participant	0 hours	12 hours	24 hours
1	3	12	13
2	5	15	14
3	6	16	16
4	4	11	12
5	7	12	11
6	3	13	14
7	4	17	16
8	5	11	12
9	6	10	11
10	3	13	14

0 hours:
Mean = 4.6
standard deviation = 1.43.

12 hours:
Mean = 13.0
standard deviation = 2.31.

24 hours:
Mean = 13.3
standard deviation = 1.83.

The ANOVA summary table:

Source:	SS	df	MS	F
Between subjects	48.97	9	5.44	
Within subjects	534.53	20		
Experimental	487.00	2	243.90	92.36
Error	47.53	18	2.64	
Total	584.30	29		

Total SS: reflects the total amount of variation amongst all the scores.

Between subjects SS: a measure of the amount of unsystematic variation between the subjects.

Within subjects SS:

> ***Experimental SS:*** a measure of the amount of systematic variation within the subjects. (This is due to our experimental manipulation).
>
> ***Error SS:*** a measure of the amount of unsystematic variation within each participant's set of scores.

Total SS = Between subjects SS + Within subjects SS

Assessing the significance of the F-ratio (by hand):
The bigger the F-ratio, the less likely it is to have arisen merely by chance.

Use the between-subjects and within-subjects degrees of freedom to find the critical value of F.

Your F is significant if it is *equal to or larger* than the critical value in the table.

Here, look up the critical *F*-value for 2 and 18 degrees of freedom

Columns correspond to EXPERIMENTAL degrees of freedom

Rows correspond to ERROR degrees of freedom

Here, go *along* 2 and *down* 18: critical *F* is at the intersection

Our obtained *F*, 92.36, is *bigger* than 3.55; it is therefore significant at $p<.05$. (Actually it's bigger than the critical value for a *p* of 0.0001)

	1	2	3
1	161.448	199.5	215.707
2	18.513	19	19.164
3	10.128	9.552	9.277
4	7.709	6.944	6.591
5	6.608	5.786	5.409
6	5.987	5.143	4.757
7	5.591	4.737	4.347
8	5.318	4.459	4.066
9	5.117	4.256	3.863
10	4.965	4.103	3.708
11	4.844	3.982	3.587
12	4.747	3.885	3.49
13	4.667	3.806	3.411
14	4.6	3.739	3.344
15	4.543	3.682	3.287
16	4.494	3.634	3.239
17	4.451	3.592	3.197
18	4.414	3.555	3.16
19	4.381	3.522	3.127
20	4.351	3.493	3.098

Interpreting the Results:

A significant F-ratio merely tells us that there is a statistically-significant difference between our experimental conditions; it does not say *where* the difference comes from.

In our example, it tells us that sleep deprivation affects vigilance performance.

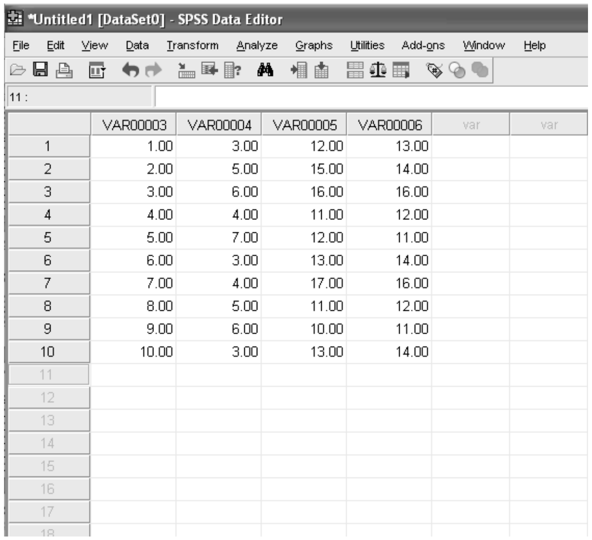
To pinpoint the source of the difference:

(a) *planned comparisons* - comparisons between groups which you decide to make *in advance* of collecting the data.

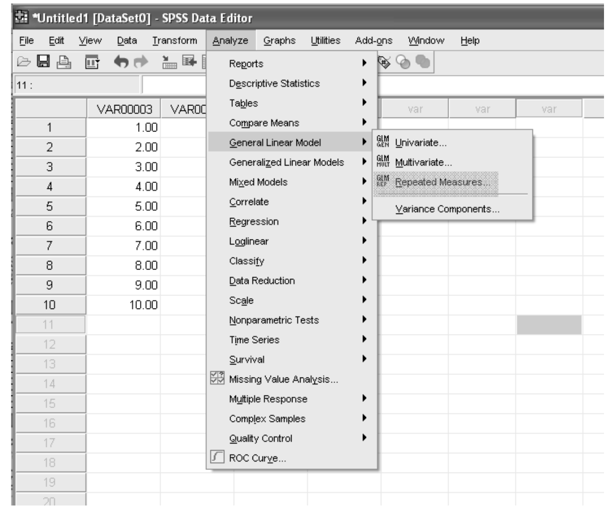
(b) *post hoc tests* - comparisons between groups which you decide to make *after* collecting the data:
Many different types - e.g. Newman-Keuls, Scheffé, Bonferroni.

Using SPSS for a one-way repeated-measures ANOVA on effects of fatigue on vigilance

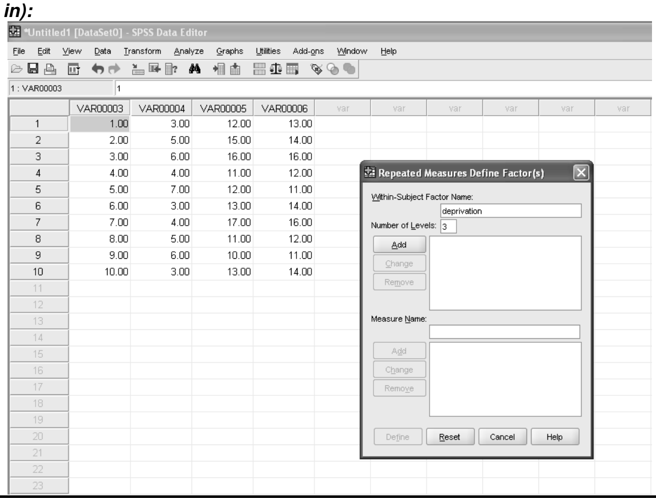
Data entry



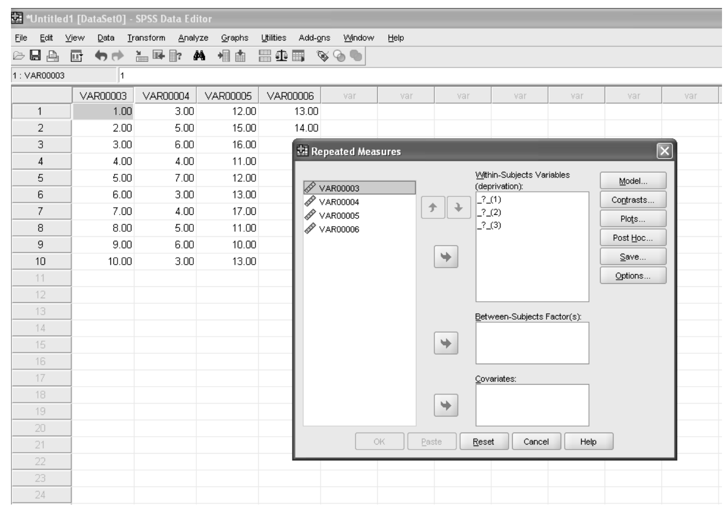
Go to: Analyze > General Linear Model > Repeated Measures…

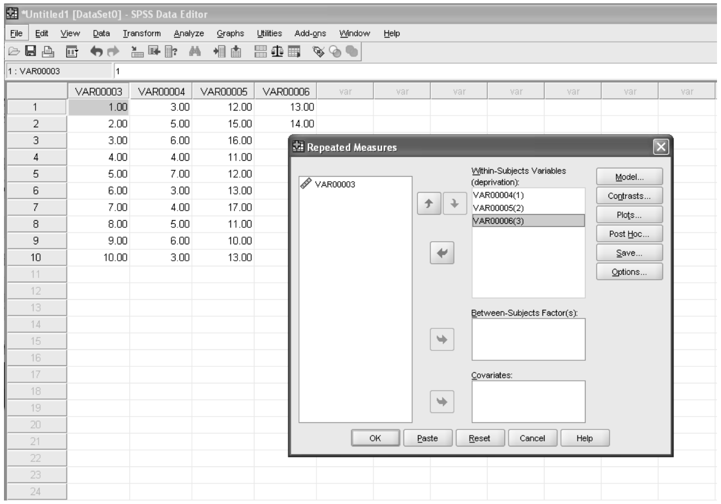


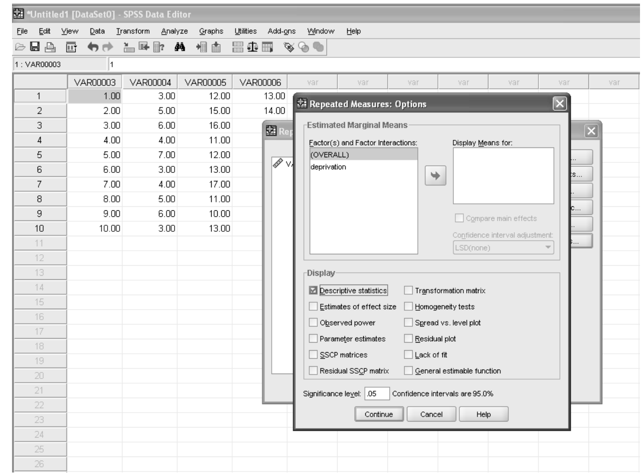
Tell SPSS about your within-subjects Independent Variable (i.e. number of levels; and which columns the levels of the independent variable are in):



Move VAR 4, VAR 5 and VAR 6 into the 'Within-Subjects Variables' box by pressing the top arrow; then press 'options…' button







Then click continue and OK

The SPSS output (ignore everything except what's shown here!):

Descriptive Statistics

	Mean	Std. Deviation	N
No sleep deprivation	4.6000	1.42984	10
12 hours' deprivation	13.0000	2.30940	10
24 hours' deprivation	13.3000	1.82878	10

Similar to Levene's test - if significant, shows inhomogeneity of variance.

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
deprivation	.306	9.475	2	.009	.590	.627	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: deprivation

SPSS ANOVA results:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
deprivation	Sphericity Assumed	487.800	2	243.900	92.360	.000
	Greenhouse-Geisser	487.800	1.181	413.186	92.360	.000
	Huynh-Feldt	487.800	1.254	388.985	92.360	.000
	Lower-bound	487.800	1.000	487.800	92.360	.000
Error(deprivation)	Sphericity Assumed	47.533	18	2.641		
	Greenhouse-Geisser	47.533	10.625	4.474		
	Huynh-Feldt	47.533	11.286	4.212		
	Lower-bound	47.533	9.000	5.281		

Use Sphericity Assumed F-ratio if Mauchly's test was NOT significant. Significant effect of sleep deprivation ($F_{2, 18} = 92.36$, $p<.0001$)

OR, (if Mauchly's test was significant) use Greenhouse-Geisser ($F_{1.18, 10.63} = 92.36$, $p<.0001$).

This is not too interesting; this just tells us that the subjects are significantly different from each other.

Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	3182.700	1	3182.700	584.975	.000
Error	48.967	9	5.441		