

[Page One](#)
[Campus Computing News](#)
[Changes in Authentication Services Greet the New Academic Year](#)
[They're Outta Here!](#)
[What the General Access Lab Staff Did During Your Summer Vacation](#)
[Computer Buying Tips for Students](#)
[Free Software Licenses: A General Overview](#)
[Today's Cartoon](#)
RSS Matters
[The Network Connection](#)
[Link of the Month](#)
[WWW@UNT.EDU](#)
[Short Courses](#)
[IRC News](#)
[Staff Activities](#)
[Subscribe to Benchmarks Online](#)

Research and Statistical Support University of North Texas

RSS Matters

Link to the last RSS article here: [R Commander: A Simple Windows Interface for R on the Windows Platform](#) - Ed.

One-way Repeated Measures and Corresponding Multiple Comparisons using SPSS and R

By [Mike Clark](#), Research and Statistical Support Services Consultant

Hello, my name is Mike Clark and I'm the newest addition to the RSS group here on campus. Having settled in I decided it'd probably be worthwhile to begin contributing to the benchmarks articles and hopefully help others out with problems I come across or feel are probably worthy of discussion.

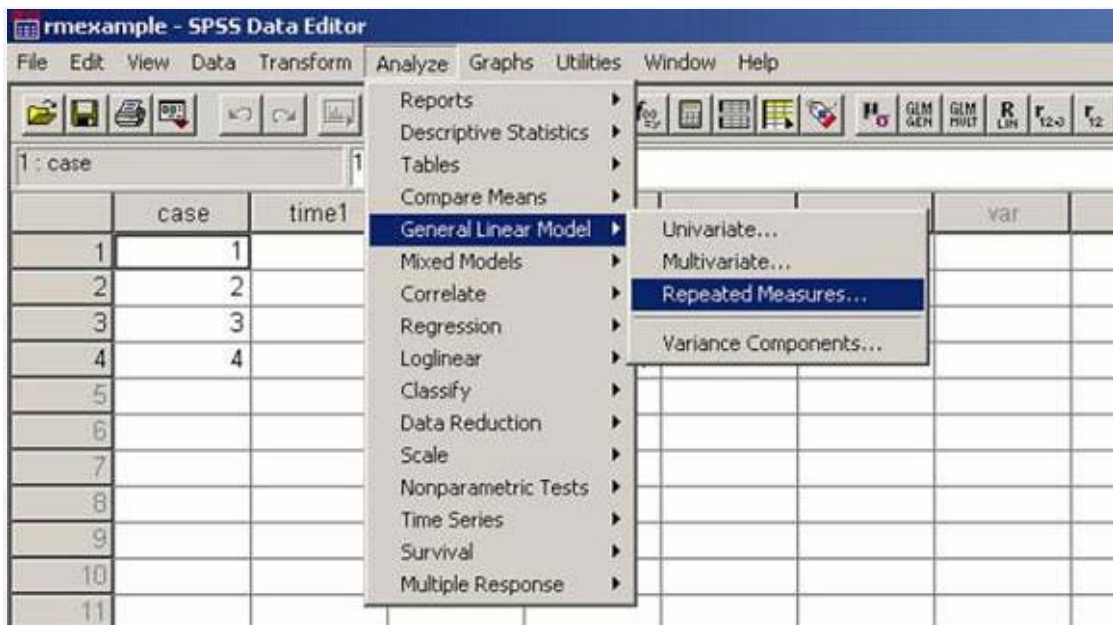
One of the first problems I tackled in my new position involved performing a one-way repeated measures in SPSS. Though not necessarily a technically difficult procedure compared to many, it can be a trying experience attempting to implement software in order to give good estimates of multiple comparisons. In order to begin we'll need to cover the basics so we know what we're doing and trying to accomplish.

In the one-way RM design, each person or subject under study is scored on multiple occasions of some measure. In other words they have several scores on some dependent variable. For example, perhaps they are given some test on different occasions over time, or are tested on multiple items related to some construct (e.g. a depression inventory). To contrast this with a typical Analysis of Variance (ANOVA), all participants undergo these multiple treatments, rather than be relegated to belonging to one treatment or another. The table below shows how one set of data might look:

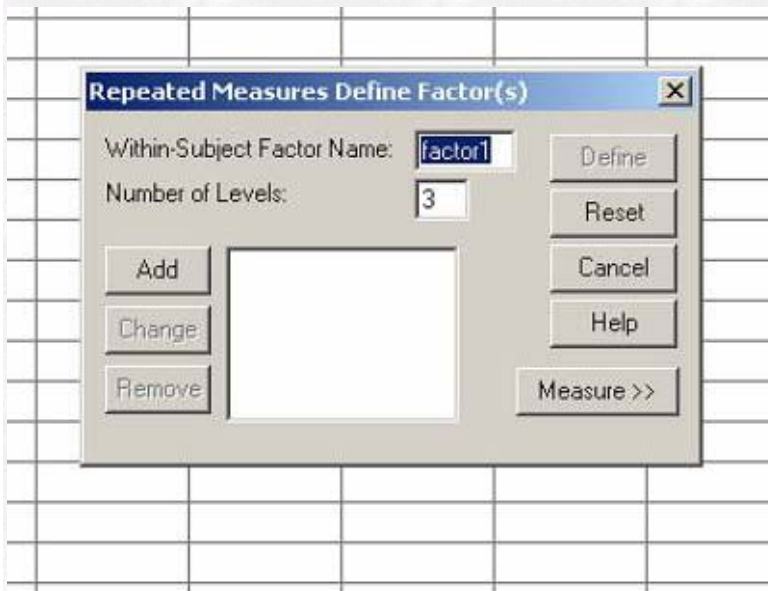
Case	Time1	Time2	Time3
1	5	6	7
2	4	4	6
3	3	5	7
4	2	4	7

First of all we'll want to know if there is an overall effect of some change over time. That is what our basic F test will tell us in this situation. In SPSS our mouse clicking will go something like this:

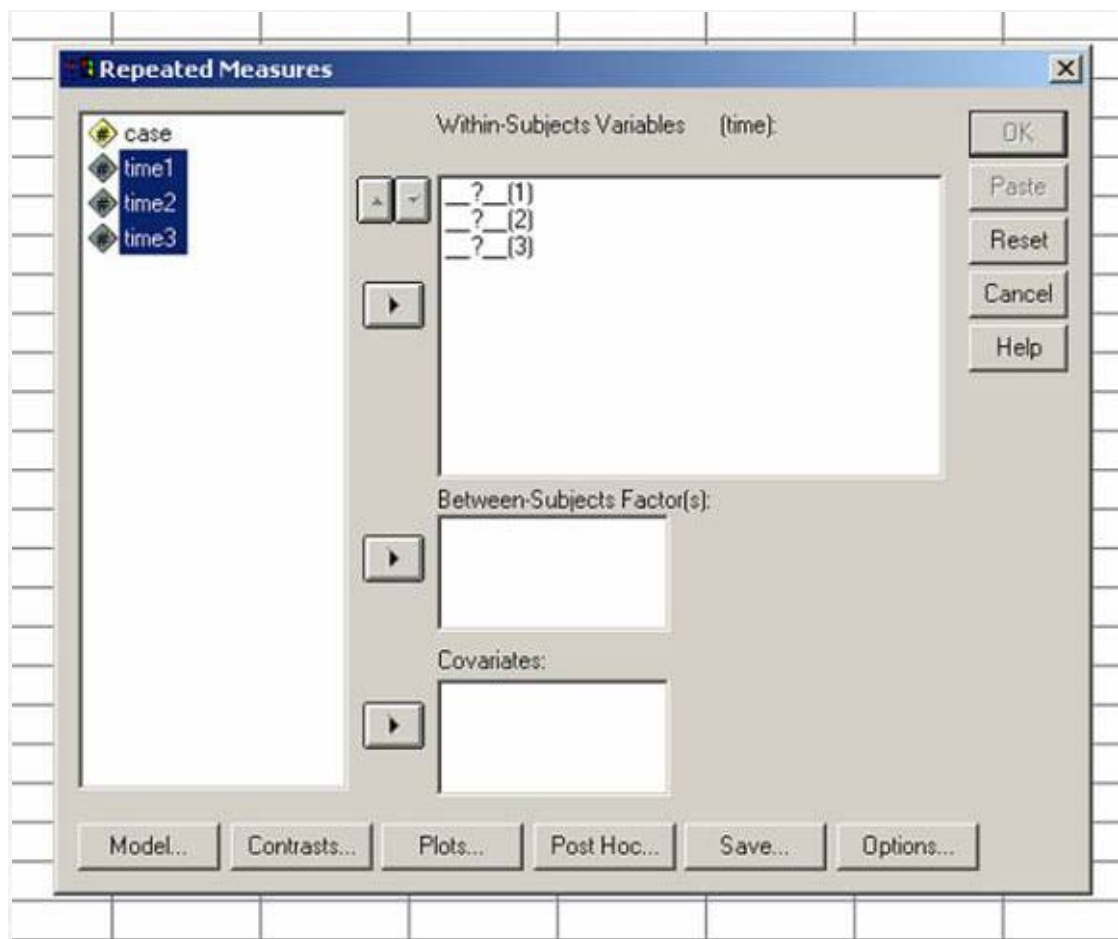
Analyze/General Linear Model/Repeated Measures



Next we'll have to define the factor of study. From the table above we have 3 levels of treatment (i.e. time 1,2,3), and we'll call the factor TIME (instead of factor1). Click "Add" after giving the name and number of levels:



Now we click "Define" and we're all set. Now on the box we'll need to highlight our 3 variables of interest (time1-3), and move them over to the "Within Subjects Variables" area by clicking on the arrow between where the variables are on the left and where they're going on the right.



Now at this point we could just click ok and go to our output, but we might have another question on our mind. Perhaps we want to know about specific differences between, for example, time1 and time2 or time2 and time3. In order for this to work out we'll need some sort of multiple comparison feature.

If you had some particular relationship in mind you want to test due to theoretical reasons (e.g. a linear trend over time) one could test that by doing contrast analyses available in the contrast option before clicking ok. One that might be useful in the above scenario is the "Repeated" contrast that would compare time1 to time2, time2 to time3 and so on.

Here's the gist comparing the different contrasts available which are deviation, simple, difference, Helmert, repeated, and polynomial.

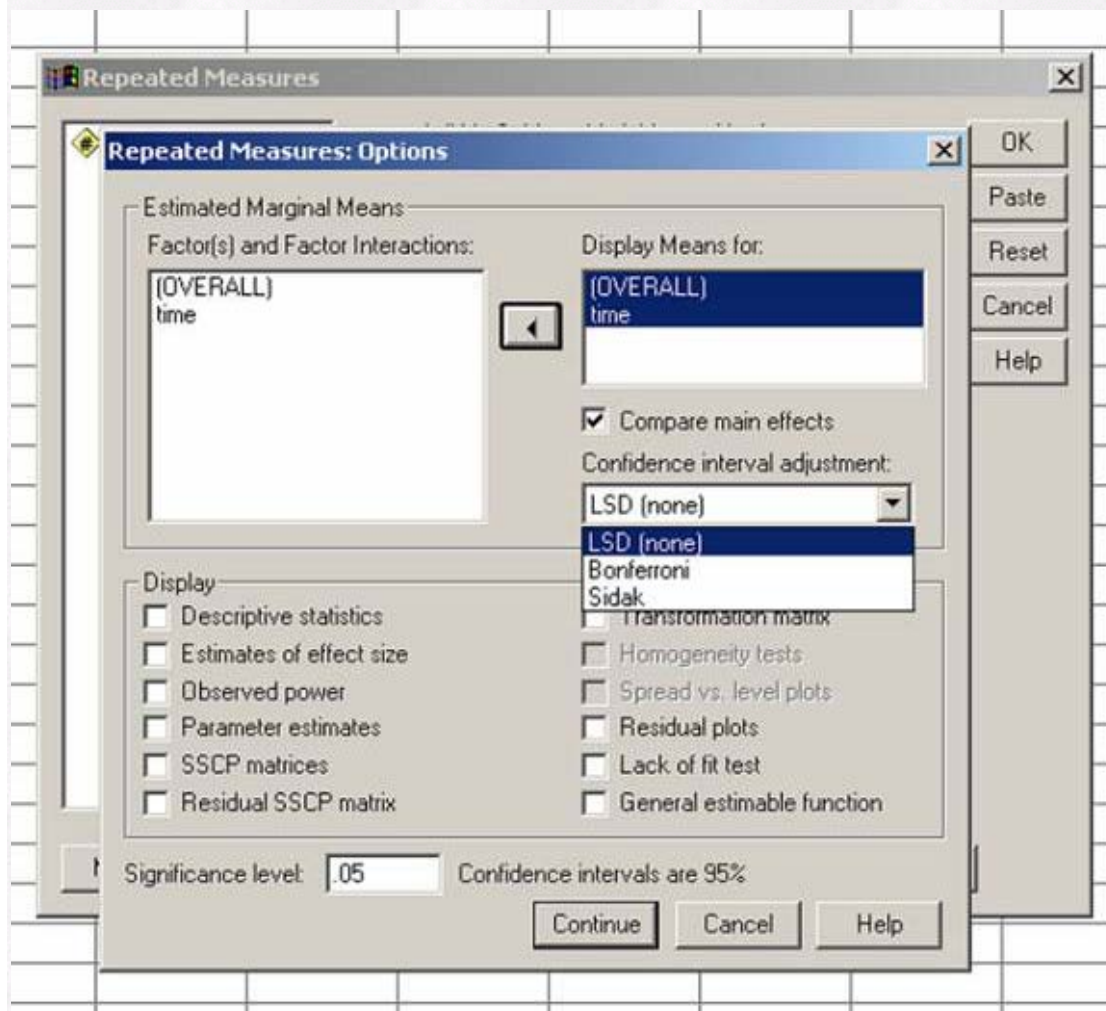
Deviation	Compares the mean of one level to the mean of all levels (grand mean)
Simple	Compares each mean to some reference mean (either the first or last category e.g. a control group)
Difference	Compares level 1 to 2, level 3 with the mean of the previous two etc.
Helmert	Compares level 1 with all later, level 2 with the mean of all later, level 3 etc.
Repeated	Compares level 1 to level 2, level 2 to level 3, 3 to 4 and so on
Polynomial	Tests for trends (e.g. linear) across levels

If on the other hand you aren't dealing with a time based model (e.g. you're dealing with particular items from some measurement scale) or just don't have any preconceived notions of what to expect you'll have to perform some sort of post hoc analysis. One's first inclination might be to click Post Hocs and look for the good ol' Tukey test. However you'll be left in the lurch in this case if you do so.

So now how are you going to do conduct a post hoc analysis? Technically you could flip your data so

that items are in the rows with their corresponding score, run a regular one-way ANOVA, and do Tukey's as part of your analysis. However along with there not being a non-tedious way to go about this you would still be in trouble because the appropriate error term would not be used in the analysis. In fact, in treating the data as a between groups design you'd have a larger error term which would make it more difficult to find a significant difference. In short, the result would be a more conservative estimate but what you gain in terms of type I error you lose in type II error.

What you'll want to do then to maintain (maximum!) power is to perform your comparisons within the repeated measures design. To begin with, at your Repeated Measures dialog box click on options. Along with descriptives, effect size, etc. that you'll want to click you'll notice that there is a white space for displaying the factor means and overall mean for the items/levels/times you're looking at. Go ahead and move the options on the left into that space and observe that the option to "Compare main effects" underneath it becomes available. Click that option. Now you have three options to choose from for these comparisons, only one of which you'd probably use. LSD or Least Significant Difference (no not that other stuff) just goes about the t-tests comparing one level to the other with no correction involved at all. Of course your type I error rate goes through the roof and you end up saying lots of things are significant when they really probably are not (to find alpha rate for multiple t-tests figure $1-(1-\alpha)^c$ where c is the number of tests you perform).



So what we're concerning ourselves with here is controlling our overall error rate by changing the error rate used for each comparison. The Bonferroni attempts to correct this by dividing your alpha by the number of comparisons. In our example this would be 3 (i.e. 1 vs. 2, 2 vs. 3, 3 vs.1) so the alpha for each comparison would be $.05/3$ or $.0166$. Although this seems ok, in many cases we have many more than just 3 levels of a repeated measure we are looking at. For example, if we had 8 levels of a factor we'd need 28 comparisons, which by using the Bonferroni method we'd test each at the $.00179$ alpha level. Yikes! So if anything is "significant" great but good luck finding it. This is

perhaps the primary criticism of this method is that it is too conservative in many cases, leading to type II error or essentially throwing something that might be *practically* significant out the window. It also doesn't make much sense to base your finding of significance in one test exclusively on how many other tests you do.

The only option left is the Sidak.. It's correction is $\alpha_c = 1 - (1 - \alpha_e)^{1/j}$. In our current situation that would test comparisons at $1 - (1 - .05)^{1/3}$ or .0169. Not really different than the Bonferroni correction. In the 8 level scenario discussed previously we'd test at the .00183. Again not much different but if we want the best shot at finding a significant difference (maximum power!) we'd want to use the Sidak option in SPSS.

So is that it then? As far as SPSS is concerned yes, that's how we'd do our post hocs for a one-way repeated measures design. However there would be an even better way to do so using the statistical program R which unlike SPSS and other statistical packages is completely free. Using R one can do the Bonferroni, Sidak and other more (statistically) powerful corrections all at the same time. In our 3 level case let's say we did our comparisons and got a .02 p-value for each. Before we jump 'significantly' up and down we know we're going to have to deal with a correction, and as we already know that the .02 level doesn't cut it for Bonferonni or Sidak's corrected p-value needed.

Now if I punched these three p-values into an R program here's what I'd come up with.

	rawp	Bonferroni	Holm	Hochberg	SidakSS	SidakSD	BH	BY
[1,]	0.02	0.06	0.06	0.02	0.058808	0.058808	0.02	0.03666667
[2,]	0.02	0.06	0.06	0.02	0.058808	0.058808	0.02	0.03666667
[3,]	0.02	0.06	0.06	0.02	0.058808	0.058808	0.02	0.03666667

What this shows is my raw p-values and what they'd really be after correcting for overall alpha (in other words we're just thinking in general rather than per comparison). Again with Bonferroni and Sidak (both of them) we still miss the magical .05 level. But take a look at the BY at the end there. That refers to a correction (BY stands for Benjamini and Yekutieli) that takes into account the dependent nature of our variables i.e. that the same people are tested at the levels being compared. And what do you know? Significance! Even after the correction the "true" p-value would be .037, less than .05!

Now how do we do this? Well the good doctor [Richard Herrington](#) has already gone into glorious detail about it [here](#). Not only that but you can actually run the program for there on the webpage, you don't even have to have R on your machine! Essentially you do the noncorrected (LSD) version in SPSS and then just put your p-values into the program. Nothing too difficult.

That's about it for doing post hocs in a one way repeated measures design. Hope you were able to get something out of it. We'll see you next time.

Web and other references

On the problems of Bonferroni's correction:

Perneger T.V. (1998) BMJ, 316:1236-1238 (<http://bmj.com/cgi/content/full/316/7139/1236>)

An alternative to Bonferroni:

Benjamini, Y., Hochberg, Y. (1995). J.R. Stat. Soc. B, Vol 57, pp. 289-300.

Yekutieli, D. & Benjamini Y. (1999). J Stat Plann Inference, 82(1-2), pp. 171-90.

Speaking of Dr. Herrington, he says you should all check this out:

http://linuxtoday.com/news_story.php3?ltsn=2003-08-20-006-26-OS-SV-DV