Feb. 3 Homework: Dependent Sample Project

For this assignment, the key aim is to ***simulate data*** with characteristics you deem to be realistic for comparing treatments (paradigm 2a, P & H article) in an *experiment using dependent samples*. It is assumed that you will use R, and analyze your data using **`granova.ds`** (as well as whatever other functions you see as needed or appropriate).

Think first of two treatments you would 'really like to compare'. Describe the context you have in mind in sufficient detail that it communicates all features of relevance. This description should provide answers to these questions: Why do you have an interest in comparing these two treatments? What, exactly, characterizes these treatments, including how long the experiment would 'run', what 'subjects' would experience in each one, whether others have made this comparison before (but perhaps outside of a true experiment, or not using dependent samples (and give references if you can)) and why results of this comparison are likely to be meaningful or useful. You might also think about things you judge most likely to go wrong, to interfere with either the execution or interpretation of results, and describe steps you could take to reduce the effects of such problems.

Because you will be thinking in terms of dependent samples, a key step is to identify or describe a way of ranking all units (there being 2*n of them) based on a variable that you would in practice measure at the outset of the study. In my A1c example I used pre-experimental A1c's; in yours, you might use children's test scores, but in any case the central idea is to choose a variable that you would in practice have reason to believe would co-relate 'highly' with ultimate post-treatment scores (where I used 'co-relate' because the association could be curvilinear or linear – the method does not care about linearity). Of course you should also describe the outcome measure(s) you would aim to use; the central ideas here would be to ensure that the outcome is sensitive to whatever influence the treatments are likely to 'cause', is likely to be 'psychometrically sound', and likely to be accepted as sound by the audience(s) you most care to see your ultimate report.

The key function to use for simulating data is mvrnorm (from MASS package) since this function provides an easy way to create dependency between two columns of scores. The command **`str`** shows **`function (n = 1, mu, Sigma, tol = 1e-06, empirical = FALSE)`** which shows you that you need merely set the arguments n, mu and Sigma to run **`mvrnorm`**. Choose **`n`** to be in the range of say, 10 to 50, and look at results before a final choice. **`mu`**, a vector with two values, should also be set after some trials (**`mu = c(mu.1, mu.2)`** also, where the **`mu.j`** are two population treatment means. These means should be chosen with respect to you judge to be 'realistic' differences between population means for the comparison you have in mind, this difference also being based largely on the variances of the columns of scores. Think in terms of an effect size (standardized mean difference) being in a range such as .4 to .9 say, and choose the mean difference accordingly. The easiest choice for **`Sigma`** is to define it as a correlation matrix, where I'll use the notation **`Sig = matrix(c(1, r, r, 1), ncol=2)`** in which case you can set the degree of dependency with your choice of the scalar **`r`**. See the history file for the last class, posted on our wiki. (You may try **`empirical = TRUE`**, but the default should work fine.) It might be useful to try **`r`** = 0, .5, .7, .9 or similar values in this range. Pay attention to details too, as in rounding simulated scores effectively (round function), titling your plots and tables, and most importantly, describing how results of your simulated experiment should be interpreted. Say enough to be comprehensive, but be succinct. Think in terms of what my various **`granova.ds`** figures or plots looked like, based on real data, and read the interpretations you see in the P & H article. In the end you should have a 5-8 page report, well organized, and readily interpreted.