Assignment for Feb10, Modern Graphics

The first part of this assignment can begin from dependent sample data you have simulated (perhaps redoing your first assignment), or a new version of that based on use of the function **dep.dat** below. Alternatively, you may use real data or a textbook source for you data. Since you have already run and interpreted **granova.ds**, you should be able to do this relatively quickly again, and proceed to write at least 2-4 sentences about what the numerical and graphical results tell you. (Note that I decided to ask for more attention to the ds case, analyzed two ways, and then to ask you to consider .2w analyses – see the Optional Exercise at the end here. I made some changes, accordingly to the BloodLead pdf.)

Next, use the function **twoCols2.3** to put the ds data in a format that can be read by **granova.2w** to get an alternative analysis and a moving graphic for your data. I strongly suggest that you run your **.2w** analysis two ways, once w/ the fit argument set at its default value, and then with **fit='quadratic'** – for which a screen snapshot is provided in the BloodLead pdf. (Given the two functions I've provided, the computer work should take relatively little time (not more than an hour I would think). Interpret each **.2w** as well as you can, paying particular attention to the ideas of 'main effects' and 'interaction"; try especially to make clear how the ds and 2w analyses differ AND complement one another; also how they are similar in what they say about the research question that (one might presume, especially for the simulated version) that drove the initial data collection.

In the end you should have, say, a 6-8 page report that you have written after reading the Guidelines4Homework. Most importantly, I'd like to have at least two class presentations (say 10 – 15 minutes) based on what you have done. *Be sure to include your R code.*

After completing the preceding, if you have time, consider going on to do the *Optional Exercise,* below.

```
dep.dat =function(n=1,r,mu1,mu2,es){

#routine generates an object the key part of which is xx, dependent
sample data

#You need to enter all six arguments (scalars) to run it

mu=c(mu1,mu2)

dif=diff(mu)

spool=dif/es

   sigma=matrix(c(1,r,r,1),ncol=2)

sigma=(1/sqrt(2))*spool^2 * sigma

sigma=round(sigma,2)

xx=mvrnorm(n,mu=c(mu1,mu2),Sigma=sigma)

xx=round(xx[,2:1],2)

list(es=es,mus=mu,sigma=sigma,xx=xx)  }
```

Example (cf. An A1c example similar to that in Elemental Graphics paper)

```
xx.15.7.5 = dep.dat(15, 0.7, 8, 9, .5) #where I name the object #based
on inputs to facilitate comparison across different choices for args.
```

```
# ---------------------------------------

twoCols2.3 = function (xx) {

# xx is assumed to be a matrix w/ two columns of (quantitative) data

    xv = c(xx[, 1], xx[, 2])  #same as xv = as.vector(xx)

    ncx = nrow(xx)

    xx3 = data.frame(xv, rep(1:ncx, 2), rep(1:2, ea = ncx))

    dimnames(xx3)[2]=list(c("Response","FactorA","FactorB"))

    xx3  }
```

Optional Exercise*: Find at least one example of real data for a two-way anova. (Online books such as that of Lowry perhaps, or textbooks, R packages whatever; the ISwR package has at least one data set,* **coking**, *that is already setup (but read it as* **coking**[,3:1]) for .2w.*) Either data for a randomized blocks design or a factorial. It will help if n's for cells are all the same, and n is not greater than 10 or so per cell; dataset should have no more than around say 80 points or data values. Set the data up in a form that corresponds to what you see on the right side of p. 2 in the BloodLead pdf and produce analyses that include numerical outputs, and – if you can – a snapshot or two from the dynamic graphic that* ***.2w*** *produces. If you'd write a report for your analysis, and the data are real, describe the scientific questions that your analysis aims to answer. I will provide feedback. Finally interpret your results in two ways: 1. Concentrating on the statistical aspects, re: matters of inference, and so forth, 2. Concentrating on what the results with respect to the initial scientific questions.*

Software Note: (Best to read this w/ the .2w help open.)

The first column of the input data for a .2w run should always contain the vector of observations for all groups, in this case w/ equal n's in all groups (unequal n's complicate things considerably). Columns 2 and 3 contain the subscripts the indicate rows (i, say) and columns (j, say). Beyond the (right side 3 column version of the blood lead data, you might look at data(warpbreaks) and data(rat), both of which have help files in R, and both of which can be used w/out changes in .2w runs. If you want to you can simulate balanced data using say, xx.datfrm = data.frame(rnorm(N, mu, sigma), rep(1:I), rep(1:J, each=n)) where N = I*J (product of the upper limits of i and j subscripts) if n = 1; etc.

This version sets n = 2 per cell (try to generalize it); e.g. xx.df = data.frame(Resp=rnorm(24,30,2),factA=rep(1:3,8),factB=rep(1:4,ea=6))

You can then run > granova.2w(Resp ~ factA + factB,data=xx.df)

The preceding simulates data w/ NULL population effects. Now, change the Response variable to make effects non-null, say as follows. Let Resp =

Resp = rnorm(24,30,2)+ rowSums(cbind(rep(1:3,8),rep(1:4,ea=6)))

then xx.dfC = data.frame(Resp, factA=rep(1:3,8),factB=rep(1:4,ea=6))) Again, you may set fit = 'linear' (default) or 'quad') in a run such as

granova.2w(Resp ~ factA + factB,fit='quad',data=xx.dfC) #Try these!

It might be helpful to round responses first; e.g. Resp=round(Resp,2)