STAT 8200 — Design of Experiments for Research Workers
Lab 3 – Due Monday, Sept. 16, 2013

Example:

An experiment was conducted in which subjects were randomly assigned to one of four levels of vitamin D supplementation: 2000, 1000, 400, or 0 International Units (IUs) per day. After 12 weeks of daily supplementation, blood plasma samples were taken from each subject and, among other variables, the concentration of 1,25-dihydroxyvitamin D (1,25D) was measured. The experiment was large, involving many subjects, but for the purposes of this example, we will assume there were 8 subjects per treatment. Hypothetical data from this experiment appear below.

Dose			
A	B	C	D
74.61	72.56	37.91	33.70
53.53	54.21	65.04	44.71
⋮	⋮	⋮	⋮
73.41	76.59	56.47	45.64

The file "vitD.sas" contains a SAS program for analyzing these data with a one-way ANOVA model. In this lab we concentrate on the use of orthogonal polynomials to analyze the effect of dose on the mean response. Copy vitD.sas from the course website to your flash drive and open the program in SAS.

Much of the code in vitD.sas has been commented out. What remains just reads the data in to a dataset called vitD and then calls PROC GLM to fit a one-way analysis of variance model that allows for different means for each level of dose. Note that in the dataset the variable DOSE is coded with non-numeric levels A, B, C and D. In the "Example" part of this lab, we will pretend that the numeric levels of dose corresponding to A, B, C and D are equally spaced, even though they are not. In the "Exercise" portion of the lab, we will see how to modify our analysis to deal appropriately with non-equally spaced levels of this factor.

There is a line in the code that reads

```
ods pdf file="mypath\vitD.pdf";
```

Edit this line and change "mypath" to a folder on your flash drive. Then run the SAS program and examine its output.

- On p.2 of the SAS output we see that the test of equal means across the four doses is rejected at $\alpha = 0.05$ ($F_3, 28 = 3.24$, $p = 0.0370$).

  Given that the mean response depends on dose, we may be interested in whether the mean increases linearly in dose. If the doses were equally spaced (and for now we are pretending

that they are), the linear, quadratic, and cubic orthogonal polynomial contrasts for this problem are as follows:

$$\psi_{lin} = -3\mu_1 - \mu_2 + \mu_3 + 3\mu_4$$

$$\psi_{quad} = \mu_1 - \mu_2 - \mu_3 + \mu_4$$

$$\psi_{cub} = -\mu_1 + 3\mu_2 - 3\mu_3 + \mu_4$$

The hypothesis of no linear trend in the mean response with dose is $H_0 : \psi_{lin} = 0$. If we were to reject this hypothesis then it still may be possible that there is some nonlinearity to how the mean response depends on dose, so we would also want to test the hypothesis of "no nonlinearity" or $H_0 : \psi_{quad} = \psi_{cub} = 0$. These hypotheses are tested using the CONTRAST statement in PROC GLM.

- We reject $H_0 : \psi_{lin} = 0$ ($F_{1,28} = 6.25$, $p = 0.0186$) and conclude that there is a linear trend in the mean response with dose.

- We fail to reject $H_0 : \psi_{quad} = \psi_{cub} = 0$ ($F_{2,28} = 1.74$, $p = .1946$) and conclude that there is non-significant evidence of nonlinearity in the mean trend with dose.

- Taken together, these two hypothesis tests lead us to conclude that the pattern of dependence of the mean response on dose is adequately described as linear.

  Notice that our call to PROC GLM also asks for estimates of each treatment mean and the ODS GRAPHICS produce side-by-side boxplots of the data at each level of dose and a plot of the estimated mean for each dose.

STAT 8200 — Lab 3

Name:_____

Exercise:

In reality the doses used in the vitamin D experiment were not equally spaced. Orthogonal polynomial contrasts can be used for non-equally spaced quantitative factors too, but the contrast coefficients are not tabulated in standard references because they depend upon the particular spacing in any given problem. In such cases we must rely on computer routines to generate the coefficients.

Let's now re-analyze the vitamin D data using the correct spacing for dose when testing linearity/non-linearity of the dose effect.

- First, uncomment the code starting on line 70 of vitD.sas and ending on line 125 (in the SAS editor, the line number on which the cursor currently rests is shown in the extreme right lower corner of the SAS window). Then highlight the code that calls PROC IML (lines 72–79) and run it. PROC IML is SAS' "Interactive Matrix Language". It can be used for all sorts of computations with matrices and vectors. Here, we use it to generate the orthogonal polynomial coefficients when the doses are 2000, 1000, 400 and 0. The function orpol() does the actual work of computing the coefficients. Then they are printed out as the columns of the matrix P. For instance, the coefficients for the linear contrast are $(.763, .0996, -.299, -.564)$.

- The next chunk of code (lines 81–93) calls PROC GLM and repeats the analyses that were done previously in the "Example" portion of this lab. However, before running this code, you must specify the orthogonal polynomial coefficients correctly now that we are using the true, non-equally spaced levels of dose. On lines 87–90, replace c1, c2, c3, c4 with the appropriate contrast coefficients. If done this for you already on line 86, for the linear contrast. Then run lines 81–93 and examine the output.

1. What is the test statistic and p-value for testing whether the dose effect has a linear trend? What is the test statistic and p-value for the hypothesis that the trend is nonlinear? Putting these results together, what do you conclude about whether the effect of dose on the mean response can be accurately described as linear?

Notice that the horizontal axis in the graphs produced by PROC GLM treat the four doses as equally spaced even though they are not. This gives a misleading impression about the linearity/nonlinearity of the dose effect. Even if the values of dose had been quantitative and unequally spaced (e.g., 2000, 1000, 400 and 0), the plots would still misleadingly show the values of dose to be equally spaced. This is because dose appears on the CLASS statement and is therefore treated as qualitative, even though it is not. If we want a true picture of the quantitative effects of dose, we should redraw the plot of estimated means using the actual values of dose. We will do this next.

- The ODS statement on line 92 asks for the least-squares means (the estimates of the treatment means) to be output to a dataset that I have called mnout. The call to PROC PRINT on lines 104–107 prints this dataset out. We can plot the estimated means from this dataset, but we will need to add the numeric values of dose to the dataset first. To do this, we create a new dataset called mnout2, copy mnout into mnout2, and then create a new numeric version of dose, which I've called ndose, that contains the actual dose values.

- Finally, PROC GPLOT is used to plot the estimated means versus the numeric values of ndose. Run this code and look at the plot.

2. Does the plot of the mean response versus ndose seem to show a nonlinear pattern, or a linear one? That is, is the picture consistent with your results from question 1? Briefly explain.

Please hand in pp.3–4, including your answers. Remember to write your name at the top. You may keep pages 1–2 for your notes.