

STAT 8200 — Design of Experiments for Research Workers
Lab 8 – Due: Friday, Oct. 25

This lab is intended to show you another example of a multi-way layout and also illustrate a phenomenon that sometimes occurs in such designs: a one-cell interaction. This example is from our text (§9.2.2).

- Retrieve the SAS program onecell1.sas from the course website, copy it to your USB drive, and run it in SAS. Be sure to redirect the SAS output to an appropriate pdf file by changing “mypath” to a valid path on your USB drive.

The data from this experiment, which involves 4 factors, each with 2 levels (1=Low, 2=High), are given below.

Table 9.2: Data from a replicated four-factor experiment.
 All factors have two levels, labeled low and high.

A	B	C	D			
			Low		High	
low	low	low	26.1	27.5	23.5	21.1
low	low	high	22.8	23.8	30.6	32.5
low	high	low	22.0	20.2	28.1	29.9
low	high	high	30.0	29.3	38.3	38.5
high	low	low	11.4	11.0	20.4	22.0
high	low	high	22.3	20.2	28.7	28.8
high	high	low	18.9	16.4	26.6	26.5
high	high	high	29.6	29.8	34.5	34.9

First, we will pretend this is a 3-factor experiment (a $2 \times 2 \times 2$) by ignoring the data collected at the High level of factor D. We do this just have a 3-way layout to consider first. Eventually we will analyze the full dataset.

- In the first call to PROC GLM in onecell1.sas, we fit a 3-way anova model to the data when D=High. The model is

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{ijk} + (\alpha\beta\gamma)_{ijk} + e_{ijkl}$$

(cf. p.157 of the lecture notes).

- Note that here the right hand side of the model is written “ $A|B|C$ ”. The vertical bar notation is a shortcut that generates all the main effects and interactions among the factors. That is, “ $A|B|C$ ” is short-hand for “ $A B C A*B A*C B*C A*B*C$ ”.

- From the SAS output, p.2, we see that the three-way A*B*C interaction is significant ($F_{1,8} = 23.02$, $p = .0014$).

What does this mean?

- It means that the nature of the two-way relationship between any two factors depends upon the third. This can be seen by estimating the treatment means and then examining the B*C profile plot at each level of A¹.
 - These plots appear on pp.4–5, and show that there is a strong qualitative interaction between B& C when A=Low, but B & C do not seem to interact much when A=High.
- Under these circumstances we could go further with the analysis along the lines of our paint on pavement example. However, this is really a four-way layout, so now we'll analyze the full dataset.

Now that we have seen what a 3-way interaction looks like, let's fit the four-way anova model. This is done in the second call to PROC GLM in onecell1.sas.

- Note that now the 4-way A*B*C*D interaction is significant ($F_{1,16} = 13.39$, $p = .0021$). This means the the nature of the 3-way interaction (between A, B & C, say) depends upon the fourth factor (D). Such a result is really difficult to make sense of.
- To visualize the 4-way interaction it can be helpful to collapse two factors down to one. E.g, instead of thinking of A and B as two separate factors with 2 and 2 levels respectively, think of the combinations of A and B as four levels of a single factor "AB". Do the same for a factor "CD". Then draw an "AB" by "CD" profile plot to try and see what is going on in our 4 way interaction.
- Such a profile plot appears on p.12 of the output. From this plot it is apparent that there is just one treatment that is primarily responsible for making the profiles non-parallel. That treatment corresponds to AB=1 and CD=1, which happens when factors A, B, C, and D are all at the Low level.

The profile plot suggests that all of the significant interactions in the anova table (p.7) may be due to one treatment. Perhaps, apart from that one treatment, the four factors have additive (non-interacting) effects on the mean response. If so, such a situation would be much simpler to understand and describe than one in which we have to interpret several significant 2-, 3-, and 4-way interactions.

¹ Alternatively, we could look at the A*B plot for each level of C, or the A*C plot for each level of B. With any of these approaches, the three-way interaction should be apparent.

To see whether these data exhibit a one-cell interaction as we have hypothesized by examining the profile plot, we can check to see whether a model fits that allows no interaction among the four factors except that it permits a unique effect in the A=B=C=D=Low condition. That is, we can fit a model of the form:

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta\gamma\delta)_{ijkl} + e_{ijklm}, \quad (*)$$

where $(\alpha\beta\gamma\delta)_{ijkl} = 0$ unless $i = j = k = l = 1$.

Notice that this model is a special case of the full 4-way anova model

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{il} + (\beta\gamma)_{jk} + (\beta\delta)_{jl} + (\gamma\delta)_{kl} \\ + (\alpha\beta\gamma)_{ijk} + (\alpha\beta\delta)_{ijl} + (\beta\gamma\delta)_{jkl} + (\alpha\beta\gamma\delta)_{ijkl} + e_{ijklm}.$$

Therefore, we can test whether all interactions are 0 other than the one-cell interaction by fitting these two models and testing them via our F test of nested models (the principle of conditional error, see p.59 of the lecture notes).

- To see how to fit the one-cell interaction model (*), note that it can be written equivalently in a different parameterization as

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \theta x_{ijklm} + e_{ijklm},$$

where x_{ijklm} is an indicator or dummy variable equal to 1 if $i = j = k = l = 1$, and 0 otherwise. To fit the model, we just include the indicator x in our model as a covariate in addition to the main effects for A, B, C and D.

- This is done in the final call to PROC GLM in onecell1.sas. The corresponding ANOVA table is given on p.14 of the SAS output. Our F test of nested models gives

$$F = \frac{(SS_E(\text{partial}) - SS_E(\text{full})) / (df_E(\text{partial}) - df_E(\text{full}))}{MS_E(\text{full})} \\ = \frac{(33.21 - 16.47) / (26 - 16)}{1.029} = 1.63$$

This page intentionally blank.

STAT 8200 — Lab 8

Name: _____

Exercise:

1. Compute the p value for the F test that compares the one-cell interaction model to the full 4-way anova model.
2. What do you conclude? Does the one-cell interaction model fit significantly worse than the full 4-way model, or is it an adequate description of the effects of factors A,B,C, and D on the mean response?

In the last call to PROC GLM in onecell1.sas, I used the OUTPUT statement to create a dataset call “pred” in which the predicted or fitted values are saved as the variable “preds”. There is one predicted value for each observation in the input dataset, but of course the predicted values for the two replicates in each treatment are the same; these predicted values are just the estimated treatment means according to the model. These predicted values are printed out on p.15 of the output and then again, after deleting duplicate observations (leaving just one estimated mean per treatment) on p.16. It is interesting to plot these estimated treatment means in the same way as we did for those from our full 4-way model. Such a plot is given on p.17 of the output.

3. Take a look at this profile plot (p.17) and compare it to that of the 4-way model (on p.12). Does it look like the much simpler one-cell model fits similarly to the full 4-way model?

Please hand in p.5, including your answers. Remember to write your name at the top. You may keep pages 1–4 for your notes.