Unit 11:
A Mixed Three-Factor ANOVA Model

11.1. The Data and the Main Effects

The data below were collected as part of a research program aimed at developing bacteriological tests for milk. Each test assembly has a Tube and a Bottle. Three types of Tubes and two types of Bottles are under test. These are fixed effects. Twelve randomly chosen milk Samples are taken (random effect). Each sample is split into 60 subsamples, ten of which are run with each combination of bottle type and tube type. The data recorded below are Counts, the number of runs out of 10 in each "cell" that grew bacteria. Thus the table, based on these counts, records one "observation" for each combination of Bottle, Tube, and Sample. (The number of bottle-and-tube assemblies used in this experiment is 2 x 3 x 12 x 10 = 720. This is also the total number of bacteria cultures examined.)

Counts of Assemblies Growing Bacteria Cultures Out of 10 Runs
For Each of 72 Bottle/Tube/Sample Combinations

		Milk Sample											
Bottle	**Tube**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
I	**A**	1	3	3	2	2	1	5	1	0	3	0	0
	B	1	4	2	4	1	1	5	1	1	4	0	1
	C	1	2	4	1	3	2	5	1	2	5	4	2
II	**A**	1	2	3	1	2	0	3	0	2	1	0	0
	B	3	1	3	0	4	2	5	2	2	1	2	3
	C	2	3	6	0	6	1	5	0	2	3	1	1

Data reported in Brownlee: *Statistical Theory and Methodology in Science and Engineering, 2nd ed.,* page 516, Wiley, 1960, New York.

Problems:

11.1.1. One row at a time, cut the data from the table above and paste it into a Minitab worksheet: you will need columns labeled `Count`, `Bottle`, `Tube`, and `Sample`.

11.1.2. Using the Minitab worksheet of the previous problem, make a table that contains the data. Compare with the data table printed in this section; proofread.

11.2. Model, Analysis and Conclusions

This is a mixed, three-way ANOVA with Bottles and Tubes fixed, Samples random, and one observation per cell. The complete model in this case includes three two-factor interactions. It is not possible to include the three-way interaction in the model because we have collapsed the 10 Bernoulli observations in each cell into a single Binomial observation. Thus, for purposes of the ANOVA procedure, we have only one observation per cell. Computationally, three-way interaction is indistinguishable from random error. There are two ways to specify the correct model here:

- Complete list: `Bottle Tube Sample Bottle*Tube Bottle*Sample Tube*Sample`
- Vertical bar: `Bottle | Tube | Sample - Bottle*Tube*Sample`. The vertical bars indicate all possible interactions; the minus sign removes the three-way interaction.

The output from Minitab's Balanced ANOVA procedure, with the appropriate model and specifications, is as follows:

```
Factor  Type    Levels  Values
Bottle  fixed        2  1, 2
Tube    fixed        3  1, 2, 3
Sample  random      12   1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12

Analysis of Variance for Count

Source         DF       SS     MS     F      P
Bottle          1    0.347  0.347  0.14  0.715
Tube            2   14.528  7.264  7.01  0.004
Sample         11   93.486  8.499  3.54  0.036 x
Bottle*Tube     2    1.694  0.847  0.77  0.476
Bottle*Sample  11   27.153  2.468  2.23  0.052
Tube*Sample    22   22.806  1.037  0.94  0.559
Error          22   24.306  1.105
Total          71  184.319

x Not an exact F-test.

S = 1.05109   R-Sq = 86.81%   R-Sq(adj) = 57.44%

                    Variance  Error  Expected Mean Square for Each
   Source          component   term  Term (using unrestricted model)
1  Bottle                         5  (7) + 3 (5) + Q[1,4]
2  Tube                           6  (7) + 2 (6) + Q[2,4]
3  Sample            1.01641      *  (7) + 2 (6) + 3 (5) + 6 (3)
4  Bottle*Tube                    7  (7) + Q[4]
5  Bottle*Sample     0.45455      7  (7) + 3 (5)
6  Tube*Sample      -0.03409      7  (7) + 2 (6)
7  Error             1.10480         (7)

* Synthesized Test.

Error Terms for Synthesized Tests

                                  Synthesis of
Source    Error DF  Error MS  Error MS
3 Sample      8.75     2.400  (5) + (6) - (7)
```

The conclusions are as follows:

- There are significant differences among the types of Tubes,
- The Sample random effect contributes a significant amount of variability to the observed Counts,
- There is no significant difference between the two types of Bottles, and
- None of the two-way interactions is significant.

Problems:

11.2.1. Write the model for this experiment, including ranges of subscripts and distributional assumptions. For uniformity, use β and *i* for bottles, τ and *j* for tubes, and *S* and *k* for samples. If you were to try including the three-way interaction in the model, what would its subscripts have to be? What other term in the model has these same subscripts?

11.2.2. Perform the Minitab procedure that produces output identical to that shown in this section. Declare Sample as random, choose the unrestricted model, make the EMS table, and store the residuals. What happens in Minitab if you try to include the disallowed three-way interaction term in the Minitab model?

11.2.3. Translate Minitab's notation in the EMS table into the symbols you used in your model in Problem 11.2.1. Verify that each F-ratio in the ANOVA table uses the denominator suggested by the EMS table. Show how the Error MS for testing the Sample factor was found, and how the corresponding F-ratio was computed. (The method of finding the Error DF is not discussed in this course.)

11.2.4. Make a normal probability plot of the residuals and give the P-value of the Anderson-Darling test. Also check for equality of variances among levels of bottle and tube.

11.2.5. Because none of the interactions in the model is significant (5% level), the issue that disorderly interaction might complicate the interpretation of the Bottle effect does not arise. Nevertheless, make the complete set of interaction (profile) plots. Comment as appropriate.

11.2.6. Remove the non-significant interaction terms from the model and run the resulting ANOVA. (Or maybe there is one of them you'll want to keep.) Is the interpretation the same as for the full model?

11.2.7 (Optional) The normal probability plot of Problem 11.4.2 shows that the data are not far from being normal. Even so, for binomial data differences in population means imply differences in population variances and so some statisticians would say a variance stabilizing transformation is appropriate. Variances for binomial counts can be stabilized by transforming the data as follows: first divide each count by 10 to get a proportion, then take the square root of each proportion, and finally take the arcsine of each result. (For short, this transformation is often called the "arcsine" transformation.) In Minitab you can make this transformation by following the menu path **`CALC ► Calculator`** and selecting the appropriate operations or (maybe more easily) with the command `let c11 = asin(sqrt(c1/10))`. Analyze the transformed data. Is the interpretation the same as for the untransformed data? Does the normal probability plot of residuals look substantially different than the one for the original count data?

11.2.8. Why would it be incorrect to treat these data as a two-way ANOVA with factors Bottle and Tube, with a Bottle*Tube interaction, and with 12 replications per cell. (Try this incorrect analysis in Minitab and see what you get.) Why can't you take the Samples in this experiment as replications in a two-factor model? What fundamental model assumption is violated if you do?

11.2.9. How would your model change in each of the following hypothetical circumstances? (Consider them separately, one at a time.) Write the model and obtain the EMS table from Minitab.

(a) You are running a dairy, you depend on 12 suppliers for almost all of the milk you process, and one of the 12 milk samples in this experiment comes from each supplier.

(b) Three rolls of tubing are selected at random and labeled A, B, C. For each assembly a tube of appropriate length is cut from one of these three rolls. The labels in the data table match the labels on the rolls of tubing, but do not represent different *types* of tubing.

Minitab Notes for Statistics 3503/6305 by Bruce E. Trumbo,
Department of Statistics, CSU East Bay, Hayward CA, 94542,
Email: bruce.trumbo@csueastbay.edu.
Comments and corrections welcome.

 Preparation of the first version of these notes was partially supported by NSF grant USE-9150433.