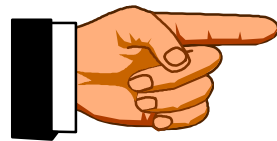


Statistical Techniques II

EXST7015

Experimental Design Examples



Example 2

- **Example 2: Snedecor & Cochran, 1980 (pg 248)**
- **Examine the calcium content of Turnip leaves. There are 4 plants, 3 leaves from each plant and 2 samples of 100 mg from each leaf.**
- **This experiment was probably done not so much to test leaves as to estimate the relative variability among the 3 levels of sampling.**

Example 2 (*continued*)

- We want to test plants with the error term leaf(plant) and not the error term sample(leaf plant).
- This is a CRD with a nested error term (2 levels).
 - ▶ $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$
- A fully nested design is one place where the TYPE I SS can be used in Designed experiments. TYPE III are used for everything else.

Example 2 (*continued*)

- **There are two ways to do this.**
 - ▶ **The TEST statement - where you specify the H=effect to be tested and the E= effect to use as error term. You can also specify a Type SS (HTYPE and ETYPE).**
 - ▶ **A test statement was used for this problem.**
 - ▶ **The RANDOM statement, with the test option, will determine the error term to be used and correct for unbalanced data. The test option was not requested for this example.**

Example 2 (*continued*)

- Note that SAS always tests by default with the residual error term. This is not correct in this example.
- The correct test was requested with the test statement, and Type I SS used.
 - ▶ **TEST H=PLANT E=LEAF(PLANT) /
HTYPE=1 ETYPE=1;**

Example 2 (*continued*)

- **The Random statement in PROC GLM would specify which of the components in the model are random.**
- **Note that SAS will provide you with the EMS with numeric coefficients for the RANDOM statement (with or without the test option).**

Example 2 (*continued*)

- **PROC MIXED is a newer alternative to solving this type of problem. This procedure works differently from the usual least squares procedures.**
 - ▶ **It is not a least squares solution, it is an iterative solution (maximum likelihood).**
 - ▶ **It estimates the random variance components (σ^2) instead of the EMS.**
 - ▶ **Then it tests any fixed effect components to the model.**

Example 2 (*continued*)

- **PROC MIXED works a little differently from PROC GLM. In PROC MIXED the fixed components ONLY go in the model and the random effects go in the RANDOM statement. There is not a "test" option in PROC MIXED.**

Example 2 (*continued*)

- **In the Turnip calcium example there are no fixed components, so only random components are estimated.**
- **Confidence limits for the random variance components were requested.**

Example 3 (*continued*)

- **Example 3: Snedecor & Cochran, 1980 (pg 293)**
- **This is an example of a CRD with different numbers of observations at different levels. Wheat yields were available for 6 districts in the midwest. There were UP TO 10 farms per district and UP TO 3 fields per farm.**
- **The model for this design is**
 - ▶ $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$

Example 3 (*continued*)

- **In this example the TEST statement was requested, but it actually gives the wrong answers because the unbalanced design has no clear error term estimated (see EMS coefficients).**
- **The TEST option on the RANDOM statement will make a simple algebraic adjust for unbalanced designs. Occasionally negative F tests can result.**

Example 3 (*continued*)

- **The TEST statement produced a test of DISTRICTS with a P-value of 0.3056.**
- **The TEST option on the RANDOM statement causes the tests to be calculated (with the appropriate error terms).**
- **This test also adjusts the tests to account for the unequal coefficients on the EMS. The P-value was 0.4601.**

Example 3 (*continued*)

- **The MIXED model again estimates the variance components with confidence intervals. Note that the confidence intervals do not include zero (a negative value should not be possible with variance components), but it is very wide and overlaps with the other variance components.**
- **Note that algebraic calculations of the Variance components from GLM give different results.**

Example 4

- **Example 4: Snedecor & Cochran, 1980 (pg 256)**
- **The experiment tests the failure of soybean seeds to germinate after treatment with one of 4 fungicides and a control.**
- **Randomized block design without replication within blocks. Five blocks and four treatment levels in each blocks.**

Example 4 (*continued*)

- The model for this design is
 - ▶ $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$
- The error term is the cell to cell variation estimated by the "interaction" term. Since there is no error for testing the interaction we ASSUME that this variation represents only random variation and that there is no real interaction.

Example 4 (*continued*)

- **The GLM tests for this model are correct because the lone error term is used for both of the sources in the model.**
- **Note that the output from the test option on the RANDOM statement provides the same results.**
- **The coefficients on the BLOCK and TREATMENT EMS are both 5. The 5 blocks are the treatment reps and the 5 treatments are the block reps.**

Example 4 (*continued*)

- **The PROC MIXED was fitted with the treatments as random effects. The output gives the same test as the GLM. This is often true for simple designs.**
- **A histogram is the only graphic output needed to express the difference between the main effects.**

Example 5

- **Example 5: Snedecor & Cochran, 1980 (pg 267)**
 - ▶ **The experiment tests efficacy of fumigants on wire worms (nematodes). There are two fumigants (C and S) and a control (0).**
 - ▶ **There are 5 blocks.**
 - ▶ **This is a Randomized block design with replication within blocks. Five blocks with three treatment levels in each blocks and 4 replicates for each treatment in each block.**

Example 5 (*continued*)

- **The model for this design is**
 - ▶ $Y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$
- **If the replicates are replicated experimental error we can use it to test the interaction. If they are sampling units within plots the test makes less sense.**

Example 5 (*continued*)

- In this case the test was done with the TEST statement instead of the RANDOM statement test option. The TEST statement should be adequate since the appropriate error term can be specified (see EMS) and the design is balanced.
- The MIXED model analysis of this simple problem will also give the same result.
- A histogram shows the differences between the main effects.

Example 5 (*continued*)

- **The test of the Fumigant*Block interaction is done with the residual error. This test is provided by default in SAS, and is correct (see EMS).**
- **This test indicates that there is a significant difference. If the replicates are samples within plots this is expected, sampling error is expected to be smaller than experimental error.**

Example 5 (*continued*)

- **It was noted that there appeared to be a pattern of the means and variances, such that the variance increased as the mean increased. This suggests nonhomogeneous variance, and was addressed by a log transformation.**
- **NOTE: HOV tests are only available for CRD (One-way ANOVA).**

Example 5 (*continued*)

- The model run on logarithms (Y_{ij+1}) was the same. Note that the variance no longer appears nonhomogeneous.
- The results changed little. The P-value for the arithmetic results was 0.0258 and for the logarithms was 0.0112.
- The PROC MIXED gave the same result.
- Although the results did not change much, the second application better meets the assumptions.

Example 6 (*continued*)

- **Example 6: Snedecor & Cochran, 1980 (pg 271)**
- **The experiment tests for differences in yield of millet for different row spacings. There are 5 spacings at 2, 4, 6 8 and 10 inches.**
- **This is a Latin Square Design because the investigators decided that the rows and columns were sufficiently heterogeneous to justify blocking.**

Example 6 (*continued*)

- **Since there are 5 treatments we need 5 rows and 5 columns for a Latin Square Design.**
- **Since there are now replicated samples within the 25 cells we have only the "interaction" as an error term.**
- **The 12 d.f. for the remaining interaction could be obtained by interacting any two components (rows, columns or treatments) if we were to put them in the model.**

Example 6 (*continued*)

- **However, since it is the only error term we leave it off the model and let SAS pick it up as the residual error term.**
- **This error is the appropriate error for all three main effect components of the model (see EMS).**
- **No additional test statements are needed.**
- **The PROC MIXED gives exactly the same results for this simple Latin Square Design.**

Example 7 (*continued*)

- **A series of Latin Squares is a number of separate Latin Squares with the same treatments in each square. The separate squares are usually blocks, but they could be another treatments. We will discuss experiments with several treatments soon.**
- **Since we do not expect ROWS and COLUMNS to necessarily be the same in each square, we nest these effects.**

Example 7 (*continued*)

- **We do expect the treatments to be the same in each square, so the treatment and the squares can be factored out as a main effects.**
- **We can also get an interaction between the squares and the treatments.**
- **The remnants of the remaining interaction are pooled into a single error term.**

Example 7 (*continued*)

- **One last unfortunate fact is apparent from the EMS. For higher order designs with more terms we often see that there is no one correct error term. Note that there is not proper error term for the squares.**
- **This is not too important if these are blocks.**
- **The TEST option on the RANDOM statement will calculate an error term for this test.**

Example 7 (*continued*)

- **PROC MIXED has no difficulty with this problem. If SQUARES were fixed, it would get a proper error term.**
- **Another type of problem is apparent from the PROC MIXED estimates of the variance components. Some variance components cannot be estimated and are set to zero. Examine the comparable components in the PROC GLM. Could you estimate these algebraically?**
- **End Design discussion!**