Estimation and Hypothesis Testing

SAS ESTIMATION and
CONSTRAST
- *Supplemental Material* -

What are Tests of Hypotheses?

- Let's start with a situation in which we have a field where there are 2 cultivars and 3 planting dates arranged in a randomized complete block design (with 4 blocks). Before we begin the experiment, we have some specific questions of interest:
  - Is cultivar 1 different from cultivar 2 (for example, less disease and/or higher yield)?
  - Is cultivar 1 different from cultivar 2, but conditional on the planting date?
- In SAS, such questions (hypotheses) may be examined using **CONTRAST** and **ESTIMATE** statements.
- The principle idea is that contrasts enable a researcher to assess linear combinations of treatment means. But, in order to see how a contrast (or estimate) works, we must first be able to translate the "verbal" hypothesis into a mathematical hypothesis based on the fitted model!

Mathematical Notation

- Let's start with a treatment mean averaged over the two factors: μ_{ij} (for example: *i = cultivar; j = planting date*)
- A linear combination of the treatment mean(s) is defined as: $\sum_{i,j} c_{ij}\mu_{ij}$
- c_{ij} represents the coefficients of a contrast and obeys the following restrictions::

$$\sum_i c_{ij} = \sum_j c_{ij} = \sum_{i,j} c_{ij} = 0$$

For Example

- Let's define two factors, A and B (each at two levels).

- If we have c_{11} = 1 and c_{12} = -1 and $c_{11}+c_{12}$ = 0, the contrast defines the simple effect $B|A_1$.

- Or, if we have c_{11} = c_{22} = 1 and c_{12} = c_{21} = -1, this defines an interaction.

- The main effect of A is defined by the coefficients c_{11} = c_{22} = 1 and c_{21} = c_{22} = -1 (also can be defined as c_{11} = c_{12} = ½ and c_{21} = c_{22} = -½)

Hypotheses and SAS

- In SAS, there are numerous ways we can test hypotheses of interest.
- One is using a **LSMeans** statement, but with large experiments, this can become very tedious (although we have seen tools to help simplify the organizing of multiple comparisons – see mult.sas).
- Using **CONTRAST** and **ESTIMATE** statements
  - This will be illustrated by first examining hypotheses for a factorial design and then expanding the concept for repeated measures studies.
- Furthermore, concepts presented here will be only in the PROC MIXED procedure. With PROC GLIMMIX, new tools are available for hypothesis testing (see Schabenberger 2007, SAS Global Forum 2007).

CONTRAST

- Useful in SAS for obtaining custom hypothesis tests
  - Greater than one hypothesis may be tested at the same time
- **General statement form** (see SAS help for further information)**:**

```
CONTRAST 'label' <fixed-effect values…>
  <|random-effect values…>, .. </options>;
```

- The procedure in PROC MIXED is similar to that of PROC GLM
  - Today we will not discuss much regarding the incorporation of random effects with CONTRAST (or ESTIMATE) that are important for understanding Broad, Intermediate, and Narrow sense inference. For further information, we recommend you to consult Littell et al. 2006 or Schabenberger and Pierce 2002 for further details (see references).

ESTIMATE

- Similar to the CONTRAST statement
- Key difference: **only** <u>one-row *L*-matrices</u> are allowed
- **General statement form:**

ESTIMATE 'label' <fixed-effect values…> <|random-effect values...>, …</options>;

- In SAS MIXED, the degrees of freedom is selected to match those displayed in the "Tests of Fixed Effects"

Illustration – Factorial Example

- To begin, let's say we have a factorial with two levels of factor A and two levels of factor (plus the A × B interaction term) as well as that there were four blocks (considered as random effects).
- Response may be yield or disease (logit-transformed).
- Model form (effects model):
  - $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + b_k + e_{ijk}$
    - α_i = effect of treatment factor A
    - β_j = effect of treatment factor B
    - $(\alpha\beta)_{ij}$ = interaction of trt A and B
    - b_k = random effect for block, $b_k \sim N(0, \sigma_b^2)$
    - e_{ijk} = overall random error term, $e_{ijk} \sim N(0, \sigma_e^2)$

PROC MIXED

Our general form for coding with SAS may look like:

```
proc mixed data = factorial;
class block a b;
model response = a|b / ddfm=KR;
random block;
run;
```

Let's incorporate some questions of interest….

1) What is the effect of treatment B given treatment A_1?

2) What is the effect of treatment A?

3) Is there is there an interaction between treatments A and B?

Q1: Simple Effect

- Here, the question is what is the effect of treatment B given treatment A_1?

- H_0: $\mu_{11} - \mu_{12}$

- We can write out the factor effects form for each μ as:
  - $\mu_{11} \rightarrow [\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}]$
  - $\mu_{12} \rightarrow [\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}]$

  and,

  $\mu_{11} - \mu_{12} = [\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}] - [\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}]$
  $= \beta_1 - \beta_2 + (\alpha\beta)_{11} - (\alpha\beta)_{12}$

  In SAS, this would be coded (using ESTIMATE) as:

  ```
  estimate 'simple effect of b given a_1' b 1 -1 a*b 1 -1 0 0;
  (contrast 'simple effect of b given a_1' b 1 -1 a*b 1 -1 0 0;)
  ```

Q2: Main Effect of A

- Here, the question is what is the simple effect of treatment A?

- Now, we work with marginal means: $\mu_{1\bullet}$ and $\mu_{2\bullet}$

- In factor effect form, for $\mu_{1\bullet}$: $\mu_{1\bullet} = \mu + \alpha_1 + \bar{\beta}_\bullet + \left(\overline{\alpha\beta}\right)_{1\bullet}$

  - $\bar{\beta}_\bullet = \left(\frac{1}{b}\right)\sum_j \beta_j$ The mean of the B effects

  - $\left(\overline{\alpha\beta}\right)_{1\bullet} = \left(\frac{1}{b}\right)\sum_j (\alpha\beta)_{1j}$ The average of the A × B effects within level A_1

- To compare A_1 and A_2, we get:

$$\mu_{1\bullet} - \mu_{2\bullet} = \left[\mu + \alpha_1 + \bar{\beta}_\bullet + \left(\overline{\alpha\beta}\right)_{1\bullet}\right] - \left[\mu + \alpha_2 + \bar{\beta}_\bullet + \left(\overline{\alpha\beta}\right)_{2\bullet}\right] = \alpha_1 - \alpha_2 + \left(\overline{\alpha\beta}\right)_{1\bullet} - \left(\overline{\alpha\beta}\right)_{2\bullet}$$

SAS Coding – Main Effects

- There are different ways we can obtain the same answer based on the question, "What is the main effect of A?":

```
estimate 'main effect of a' a 1 -1 a*b 0.5 0.5 -0.5
   -0.5;

estimate 'main effect of a' a 2 -2 a*b 1 1 -1 -
   1/divisor=2;

estimate 'main effect of a' a 1 -1;
```

ESTIMATE, by default, computes coefficients for higher-order interactions (i.e., A×B) that contain effects already mentioned. This is why the last estimate statement will give you the same solution as the previous two.

Q3: Interactions

- Interactions are defined by the difference between simple effects and can be written as: $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$
- In factor effect form, we get:

$[\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}] - [\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}]$
$- [\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}] + [\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}] =$
$(\alpha\beta)_{11} - (\alpha\beta)_{12} - (\alpha\beta)_{21} - (\alpha\beta)_{22}$

```
estimate 'a x b interaction' a*b 1 -1 1 -1;
contrast 'a x b interaction' a*b 1 -1 1 -1;
```

General Comments

- ESTIMATE provides the estimate effect of interest and a *t*-statistic for the test

- CONTRAST provides a *F*-statistic test of the hypothesis, but does not provide an estimate of the effect

Hypotheses for Repeated Measurements

- Typically, the following questions are of interest for repeated measurement studies:
  - To estimate and test simple effects, where simple effects are either among different treatments holding time points constant (or vice versa).

  - To SLICE apart and test effects of treatment at a given time point, or to test a time point given the treatment – the former portion is more common as the latter can be examined using regression methods (see Littell et al. 2006).

  - To perform simple effect tests (i.e., contrasts) defined based on specific treatment $\times$ time interactions.

Illustration Using repeatEx2m

Recall the mixed model code as:

```
proc mixed data=b covtest;
class block cultivar days;
model logit = cultivar|days / ddfm=kr;
random block;
repeated / subject=block*cultivar
  type=arh(1);
run;
```

We will also see how this is accomplished for regression coefficients when we perform a covariance analysis.

Different covariance structures will have different estimates of error, but if data are balanced, the estimated values will be the same.

Let's Start by Slicing "Time"

- To begin, we might want to examine if there are differences between treatments (cultivars) at the different assessment times (days). We will use the results to help formulate different testable hypotheses.
- To slice an interaction, we can use LSMeans and the SLICE option:

```
lsmeans cultivar*days/slice=days;
```

SAS Output and Discussion

In all situations, it appears that on all assessment days, there are significant cultivar differences. Notice that the Num DF = 3. This means that the cultivar comparisons were from a 3 degree-of-freedom test. Thus, we can construct three 1 df tests using CONTRAST or ESTIMATE.   For example, cultivar 1 versus cultivar 2; cultivar 1 versus cultivar 3;…

```
                  Tests of Effect Slices

                           Num     Den
Effect           days       DF      DF     F Value    Pr > F

cultivar*days      3         3     11.5      10.22     0.0015
cultivar*days      7         3     11.3      69.84     <.0001
cultivar*days     10         3     12.3      59.68     <.0001
cultivar*days     13         3     12.6     208.25     <.0001
cultivar*days     18         3     12.9      55.44     <.0001
cultivar*days     21         3     12.8     159.33     <.0001
cultivar*days     26         3     12.2     943.36     <.0001
```

Breaking Down the "SLICE"

- Given the results in the previous slide, let's formulate both the CONTRAST and ESTIMATE statements to compare:
  - Cultivar 1 versus 2, average over 7 days
  - Cultivar 1 versus Cultivar 2 at days = 3 and 26

Our initial test examines if the two cultivars are different when averaged across days. From this we can then break down specific days for comparison.

```
estimate 'c1 vs c2 averaged over 7 days' cultivar 1 –1 0 0;
contrast 'c1 vs c2 at day = 3' cultivar 1 –1 0 0
   cultivar*days 1 0 0 0 0 0 0 –1 0 0 0 0 0 0;
estimate 'c1 vs c2 at day = 3' cultivar 1 –1 0 0
   cultivar*days 1 0 0 0 0 0 0 –1 0 0 0 0 0 0;
contrast 'c1 vs c2 at day = 26' cultivar 1 –1 0 0
   cultivar*days 1 0 0 0 0 0 0 –1 0 0 0 0 0 0;
estimate 'c1 vs c2 at day = 26' cultivar 1 –1 0 0
   cultivar*days 0 0 0 0 0 0 1 0 0 0 0 0 0 –1;
```

SAS reads coefficients from the order of the effects in the CLASS statement. So, this means the levels of days are nested within the levels of cultivar.

Covariance Analysis
(Mix of Class and Continuous Variables)

- In the notes, you were presented a discussion of repeated measurements and covariance analysis (Example: repeatEx3). In much the same way that we have seen already, we can perform contrasts (estimates). Now, we are comparing estimated intercepts slopes between different treatments.

```
estimate 'interc1' int 1 treat 1 0;
estimate 'slope1' t 1 treat*t 1 0;
estimate 'int1-int2' treat 1 -1;
estimate 'slope1-slope2' treat*t 1 -1;
```

Covariance Analysis
(Mix of Class and Continuous Variables)

Recall from the notes on covariance analysis, there is an overall intercept and overall slope, which are then adjusted for each treatment effect. This is then why the ESTIMATE code needs to include the "int" and "t" terms when estimating individual components. However, where we compare two intercepts or slopes, then the overall term cancels out, which is why we no longer need to incorporate that into the statement.

```
estimate 'interc1' int 1 treat 1 0;
estimate 'slope1' t 1 treat*t 1 0;
estimate 'int1-int2' treat 1 -1;
estimate 'slope1-slope2' treat*t 1 -1;
```

1. An estimate of the intercept for treatment 1.
2. An estimate of the slope for treatment 1.
3. A comparison of intercepts between treatments 1 and 2.
4. A comparison of the slopes between treatments 1 and 2.

Effect of Covariance Structure

- Different covariance structures produce different standard errors of estimates as discussed in the main notes
  - Keep in mind that the best structure may not provide the smallest standard errors
- Also, when you have balanced data, the estimates of effects or differences will be the same regardless of covariance structure