Chapter 22: Analysis of Covariance (ANCOVA)

• A combination of a regression model and ANOVA model – still in the class of the "General Linear Model".

• We have two variables affecting the response – one is a factor and one is a continuous variable.

ANCOVA model

• In a situation with a response and one primary factor of interest, we may have a blocking variable that is continuous (not a factor).

• This continuous blocking variable is called a concomitant variable, or a covariate.

• Accounting for this variable in our model will reduce experimental error variance, just like in block designs.

Principles:
(1) The covariate should have a true effect on the response (otherwise we could use a simpler ANOVA model).
(2) The covariate should be observed before the study (or at least should not be influenced by the treatments).
Otherwise, the effect of the treatments on the response may be impossible to separate from the effect of the covariates.
Example: Response = Improvement from Training Program
Factor = Instructional Method
Covariate = Amount of Time Spent Studying

Problem?

Better covariate choice?

• We can check the assumption of no treatment-covariate association using a symbolic scatter plot:
• Look for a similar distribution of points along the *X* space across the different treatments.

Single Factor ANCOVA Model

Equation:

<u>ANCOVA or blocks?</u>

<u>Example</u>: Study analyzing blood pressure reduction (*Y*) in patients. Factor is the type of drug (3 different drugs).
• However, weight of patient (continuous) will also affect BP reduction. We need to account for weight in the model.

• One possibility: Break weight into categories (levels) and make weight a <u>blocking factor</u>.

<u>Problems with this</u>:
(1) May not be enough people in some weight categories.
(2) We may not know weight is affecting BP reduction until experiment is ongoing.
(3) Weight is inherently continuous, it may be hard to naturally categorize it.

<u>On the other hand</u>:
• ANCOVA assumes the relationship between *Y* and *X* is specified (usually <u>linear</u>) – a randomized block design makes no such restrictive assumption about functional form.

<u>Note</u>: Could assume a nonlinear relationship between *Y* and *X* (Example: include

• Could include more than one covariate (Example:

• Note that our basic ANCOVA model assumes the slope of the regression line relating *Y* and *X* is the same (= γ) for <u>all levels</u> of the factor.

• Fitting the model is done with the "regression approach" using indicators for the treatments.

Testing for Treatment Effects

We test

• We may also test $H_0: \gamma = 0$ to determine whether the covariate has a significant effect.
• This is not of primary interest, but it could indicate whether a one-factor ANOVA model is appropriate.

Example (Cracker data):

Response *Y* = sales of crackers (in cases)

Treatment Factor =

Covariate =

Experimental Units = 15 stores

• Is the covariate associated with the treatment factor?

• Test for Significant Promotion Type Effect:

• Bonferroni simultaneous 95% CIs for these differences can be found:

• Testing whether previous sales has a significant effect on mean sales:

• Evidence that ANCOVA model may be _________ useful than the one-factor ANOVA model here.

• Can examine residual plots to check model assumptions.

• The *i*-th "adjusted estimated treatment mean" is the estimate of the mean response for the *i*-th treatment, when the covariate is at its sample mean value.
• PROC GLM provides these estimates, plus standard errors:

Testing for Equal Slopes

• We can extend the model to allow the lines relating *Y* and *X* to have unequal slopes across the treatment levels.

• If the slopes are <u>not</u> parallel, the effect of *X* on E(*Y*) is <u>not</u> the same at <u>each level</u> of the factor
→

• Include

Previous example

Model:

• Equal-slopes would be fine if

• To test whether unequal-slopes model is needed, we test

• SAS does this if we include

<u>Cracker example</u>:

• If the equal-slopes model does not fit, then it is not appropriate to make direct comparisons across treatments.
• Section 22.4 discusses ANCOVA with more than one factor.

Chapter 26: Nested Designs

• In the factorial designs studied previously, every level of one factor appears with each level of the other factor – that is, the factors are crossed.

• In other designs, the factors may be nested.

Example (Athletics Spending Study):

Response: Athletics expenditure by college sports supporters
Factor A: Conference (e.g., SEC, ACC, Big Ten, …)
Factor B: College (e.g.,

• The first college in Conference 1 is not the same as the first college in Conference 2. The levels of "college" are nested within the levels of "conference".

Example (Training School):

Response: Class Learning Scores
Factor A: School (Atlanta, Chicago, San Francisco)
Factor B: Instructor (2 at each school)

• Each instructor taught 2 classes; classes had been randomly assigned to instructors within each school.

Data:

Note: Instructor 1 is Atlanta has nothing to do with instructor 1 in Chicago.

- **Instructors are nested within schools.**

- **There are, in fact, ______ distinct instructors in this study.**

Two-Factor Nested Design – Notation

- **This notation assumes factor B is nested within factor A.**

- **Denote the levels of A by 1, …, *a*.**
- **Denote the levels of B (within each level of A) by 1, …, *b*.**

μ_{ij} =

$\mu_{i\cdot}$ =

α_i =

Note

- **We do not examine the main effect of, say, level *j* of B, because level *j* at one level of A is not the same as level *j* at another level of A.**

- **Instead, we examine**

Example (Training school):

$\beta_{2(1)}$ is the difference between

Nested Design Model

- **Data could be unbalanced, but computations are messier.**

- **Interaction between A and B is not estimated in a nested design.**

- **If factor(s) have random levels, the usual changes to the model are made (see p. 1103-1104 for more discussion):**

ANOVA for Nested Designs

Least Squares Estimates of Parameters:

• The ANOVA table partitions SSTO into SSA + SSB(A) + SSE:

• Formulas for sums of squares are on page 1094.

• If B has random levels, then the correct F* for the test about factor A is

Model Assumptions may be checked by:
(1) Residuals plotted vs. Fitted values
(2) Residuals plotted at each level of factor A
(3) Normal Q-Q plot of Residuals

<u>SAS example</u> (Training Data):

• The ANOVA table is found in PROC GLM with SCHOOL and INSTRUCTOR(SCHOOL) in the MODEL statement.

Testing for school effect:

Testing for instructor effect within schools:

• **We may investigate this further by breaking SSB(A) into SSB(A_1), SSB(A_2), SSB(A_3) and testing instructor effects separately within each school, using**

Example:

Diagnostic Plots:

Further Analysis:

If desired, we may:

• Compare $\mu_{i\cdot}$ and $\mu_{i'\cdot}$ ($i \neq i'$) pairs using multiple comparison procedures.
• Compare μ_{ij} and $\mu_{ij'}$ ($j \neq j'$) pairs, separately for each level of A, using multiple comparison procedures (Bonferroni easiest).

Previous example:

• Sec. 26.6 discusses Unbalanced Nested Two-Factor Designs, for which the "regression approach" and "Full vs. Reduced" F-tests are appropriate.
• Sec. 26.9 discusses three-factor "partially nested" designs, in which some pairs of factors are crossed and some are nested.