Ch 22 – Analysis of Covariance (ANCOVA)

ANCOVA is ANOVA adjusted for one or more covariates. A *covariate* is a regressor one anticipates or knows is related to the response variable, so one anticipates the model having a smaller error variance by virtue of inclusion of the covariate. This is the anticipated advantage. As disadvantages, inclusion of one or more covariates in the model generally complicates the data analysis, as orthogonality may be lost and such methods as Tukey's may no longer be known to be valid. While one hopes the covariate is related to the response variable in a useful way, the covariate should not be affected by the treatments.

Example: Section 22.3, p926. See data plot in **ch22eg1.txt**.

Single Factor Models: Covariate uncentered or centered.

$$Y_{ij} = \mu + \tau_i + \gamma x_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \text{ and independent}$$

$$Y_{ij} = \mu + \tau_i + \gamma (x_{ij} - \bar{x}_{..}) + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \text{ and independent}$$

What do these look like graphically?

The latter is better computationally.

LOF test: Either model includes a rather strong assumption—namely, that there is a simple linear regression relation between the covariate and mean response, and the slope is the same for each treatment. One could evaluate this based on the data plot, residual plots, or lack-of-fit (LOF) tests. For example, it is common to do a LOF test for unequal slopes. With unequal slopes, the first model above becomes:

$$Y_{ij} = \mu + \tau_i + \gamma_i x_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \text{ and independent}$$

Example: ch22eg1.txt

Residual plots: One can evaluate LOF due related to the covariate from a plot of the raw data (e.g. y*x=A) or by plotting residuals versus the covariate (e.g. e*x=A). If the model seems to fit the data okay, one can proceed to check the equal variance assumption by plotting residuals against fitted values, and if that's also okay then one can examine a normal probability plot of the residuals.

ANCOVA: See **ch22eg1.txt**, and test for treatment effects adjusted for the covariate.

Multiple Comparisons: Scheffé's and Bonferroni's methods are applicable. It remains an open problem to show that Tukey's method is exact or conservative in this setting, since inclusion of the covariate in the model causes the treatment effect estimates to be correlated. Also, the treatment effect estimates are adjusted for the covariate, so in PROC GLM one must use the LSMEANS statement—not the MEANS statement—to get least squares estimates of treatment contrasts.

Example: ch22eg1.txt

Extensions: The above discussion pertains to study of a single factor in a CRD with a single covariate. It is a direct extension to include multiple covariates in a CRD, or to use ANCOVA for a factorial experiment (though adjusting for the covariate would complicate the analysis as would unequal sample sizes), or to use ANCOVA for a RCBD, though in testing or comparing treatments one would adjust for covariates and perhaps also for blocks.

Blocking and use of covariates as alternatives: These are two ways to use ancillary variables (e.g. blocking factor, covariate) to reduce the error variance in the model. If the covariate is known in advance of data collection, one could use the covariate as the basis for blocking, in which case one has a choice between blocking and use of ANCOVA. The ANCOVA model probably adds fewer parameters to the model, preserving more error degrees of freedom. However, blocking imposes fewer assumptions so is more likely to have a clean analysis. Hence, blocking seems preferable given a choice.

Example: ch22eg2.txt — reanalyze the ANCOVA data as a RCBD. This is somewhat artificial, since one would be able to block more effectively based on the covariate prospectively.

Homework: 22.9, 22.10ac*def

*For 22.10c, test the treatment effects, but you need not fit the corresponding full and reduced models as regression models.