1 Introduction

In analysis of covariance we wish to test equality of group means, after first adjusting for the effect of a continous covariate. In our first (educational) example, we see that the raw y = post-test values look quite similar between groups. However, y is linearly related to x = the pre-test score, so that $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. In the example the group means $\overline{x}_1$ and $\overline{x}_2$ were not equal, and this inequality of the x values obscured the true group difference for y. Even if the pre-test means were equal for the two groups, variability among the x values can also obscure the true group difference for y by creating large within-group variation. Thus the x variable is a nuisance variable, and we wish to adjust for it as we test hypotheses about the y variable. This approach is quite similar to using a blocking factor in a randomized block design to remove nuisance variation.

2 Two assumptions

To perform a standard ANCOVA, we will make two assumptions. The first of these assumptions we can check with a statistical test discussed below. The first assumption is that of parallelism. We assume that the slope β_1 is equal for each group. This assumption allows us to make a global conclusion about the treatment that is applicable for all values of x. If this assumption does not hold we then have to qualify any statements about the treatment, because it can change as x changes. A test for the parallelism assumption is discussed below. The second assumption is that the covariate x is unaffected by the treatment. As an example if we are comparing the weight of two species of mice at 1 year of age, we may consider using their weights at 6 months as a covariate. The problem is that if weight differs at 1 year by species, it will also likely differ at 6 months as well. If we adjust for weight at 6 months, we will probably remove the true difference due to species. This assumption must be considered each time ANCOVA is used.

3 ANCOVA as a general linear model, assuming parallelism

In the most common ANCOVA application, we have a linear relationship between x and y, so that $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ in each group. If the slope (β_1) is equal in the groups then the regression lines are parallel, and the test for equality of group means is equivalent to equality of the intercepts (β_0) among groups. This test can be conveniently arranged using dummy coding. Suppose, for the education example above, that we call the pre-test score x_1, and we define a dummy variable x_2 :

$$x_2 = \begin{Bmatrix} 0 \text{ if in group 1} \\ 1 \text{ if in group 2} \end{Bmatrix}.$$

Then we have the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i.$$

The test of a difference in group means is then $H_0 : \beta_2 = 0$ (a partial F or partial t test). In general, if we have k groups, then we have $k - 1$ dummy variables, and the test of the null hypothesis of equality of group means is performed using a multiple partial F test. We can then perform the ANCOVA using a regression program (such as Proc REG in SAS). Most statistical packages also have programs for ANOVA that will perform ANCOVA as well (Proc GLM in SAS performs ANCOVA so that the user does not need to define dummy variables).

4 Testing the parallelism assumption

Complications arise if the regression lines for the separate groups are not all parallel. When parallelism fails, then the comparison of y values depends on the x values, as the y comparison will change with x. We can test this assumption by introducing a new crossproduct term:

$$x_3 = x_1 x_2, \text{ and the new model } y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i.$$

Now the parallelism assumption can be tested by the null hypothesis $H_0 : \beta_3 = 0$. As in the tests above, if we have more than two groups we simply define additional crossproduct terms and perform a partial multiple F test.

5 Comparison of means

If we have more than 2 groups, if the parallelism assumption is satisfied and we have rejected the global null hypothesis of equalilty of means, we may then want to do additional tests such as pairwise tests of means. Multiple comparison tests should be based on the adjusted group means. For example, with the 2 group educational data above, we had:

$$\mu_{1i} = \beta_0 + \beta_1 x_{1i} \text{ and } \mu_{2i} = (\beta_0 + \beta_2) + \beta_1 x_{1i}.$$

The adjusted group means are then : $\overline{y}_{1(adj)} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{x}_1$ and $\overline{y}_{2(adj)} = (\widehat{\beta}_0 + \widehat{\beta}_2) + \widehat{\beta}_1 \overline{x}_1$,

where $\overline{x}_1$ is the mean of x_{1i} for the entire (combined) data set.

6 Final remarks

1) We have discussed ANCOVA for the completely randomized design, but it can be used with other ANOVA designs as well.

2) We have required that the relationship between y and x be linear, but ANCOVA can be used with nonlinear patterns as well. In these cases it remains important to have parallelism.

3) It is preferred that the ranges of the x values overlap between different treatment groups. If they do not overlap at all, then the ANCOVA test is extrapolating beyond observed x values, which may not be justified.

4) Comparing ANCOVA to blocking: If y and x are linearly related then ANCOVA is more efficient than using a randomized block design, if not then a randomized block design may be better.

5) ANCOVA versus differencing: Differencing (analyzing $y - x$) is a special case of ANCOVA in which $\beta_1 = 1$.

6) For heterogeneous slopes, an article by Hendrix *et al.* (*Biometrics*, 1982, pp 641-650) discusses some approaches to data analysis.