ANCOVA

ANCOVA is used to increase power in a one-way or two-way ANOVA by adding a second or third variable as a covariate.

It is also used to control for initial differences in pretest scores in quasi-experimental designs. This removes "covariate bias" or "selection bias". Selection bias weakens internal validity.

Key points:

- ANCOVA's usefulness depends on a good correlation between the covariate and the DV.
- ANCOVA has the same 3 assumptions as ANCOVA, plus 3 more:
    1. independence (each subject is in only one cell)
    2. normality of population from which sample is chosen
    3. homogeneity of variance across all cells
    4. a linear relationship between covariate and dependent variable
    5. homogeneity of regression coefficients: the correlations are the same
    6. the treatment IV has no effect on the covariate. (they're independent)

General Linear Model

- For a 1-way ANOVA (one independent variable), the general linear model is
  Xij = mu + group effect (alpha j or MSb) + error (eij or MSw)
- For a 2-way ANOVA on the same results (two independent variables), the general linear model becomes
  Xijk = mu + group effect #1 (alpha j) + group effect #2 (beta k) + interaction term + smaller error (eijk)
- For ANCOVA with one IV and a covariate, the general linear model is:
  Yij = mu + group effect + beta(Xij - X.) + eij
- Here X is a predictor, not a factor.
- Beta coefficient is the slope of the line of best fit: b = rxy(Sy/Sx)
- The error term thus is broken into the amount of variance by the covariate X and the residual, smaller error (eij) that doesn't depend on the covariate.

The F ratio

- For ANCOVA we use the adjusted F ratio:
  F = MSb'/MSe'
- MSe' = MSe * (1 - r**2)
- Note that this lowers the residual error by that (1 - r**2) factor, which is why ANCOVA increases power. THIS IS VERY IMPORTANT.

- Unless the r correlations are high, ANCOVA won't do very much.

The scattergrams for quasi-experiments

- In a true experiment, when you graph the DV scores on Y and the IV scores on X, all the scattergrams line up above one another. There is no separation among the pretest scores - no right-left shift.
- This right-left shift comes in when you run a quasi-experiment. Look at the scattergrams.
  - If the slope is the same, the homogeneity of regression assumption is OK.
  - If they are skinny enough and in the same direction, the linearity assumption is satisfied.
  - Then, ANOVA splits the grand mean on X, extends both regression lines to the vertical line through that grand mean, and reflects them leftwards to the Y axis.
- The result is that the grand mean becomes the adjusted X'. The Y intercepts of those horizontal lines become the new adjusted Y's. The separation between the adjusted Y's will be less than between the unadjusted Y's.
- This means the MSb' goes down somewhat, but the MSe' - the error term - goes down even more. This is how the power gets increased.
- When you write up an ANCOVA you always report the adjusted means.
- If you had little or no correlation, then ANCOVA would gain you nothing above a one-way ANOVA.
- Note that just because the correlation between the covariate and the DV is very high, and the error goes down, that does not make the quasi-experiment the same as a true experiment. It just lowers the error.

Covariate scattergrams

- A good choice of covariate has two scattergrams which look the same but are displaced with a shift along the X axis.
- The size and direction of the rs should be the same. That gives power.
- As a result, there will be a clear difference in the adjusted means.
- If there is no shift along the X axis, you have a true experiment, but you can still use ANCOVA to decrease error and increase power.

Choice of a good covariate

- A covariate is a source of variation that is not controlled for in the design of the experiment, but which does affect the dependent variable.
- It is used when you have intact groups, as in quasi-experiments. For example, section I may have higher reading ability than section II.
- In ANCOVA, the dependent variable is adjusted statistically to remove the effects of the portion of uncontrolled variation represented by the covariate. Basically, the covariate is used to:
  1. reduce error variance

2. take any preexisting mean group difference on the covariate into account
3. take into account the relationship between the covariate and the dependent variable, and
4. yield a more precise and less biased estimate of the group effects.

- The covariate should be independent of the independent variable, and it shouldn't correlate highly with any other covariates.
- Adding a covariate complicates the design. It also means you'll probably need more subjects so you won't get empty cells in the new design.

Lorraine Sherry
lsherry@carbon.cudenver.edu
Updated April 1, 1997