

Matching and ANCOV with Confounded Variables[®]

Suppose we are interested in the relationship between some categorical variable and some continuous variable. We have available data on an extraneous variable that we can use for matching subjects or as a covariate in an ANCOV. If we were manipulating the categorical variable, we could match subjects on the covariate, and then, within each block, randomly assign one subject to each treatment group. If our covariate is well correlated with the dependent variable, but not correlated with the independent variable, the randomized blocks design or ANCOV removes from what would otherwise be error variance the variance due to the covariate, thus increasing power. If we measure the covariate prior to administering our experimental treatment and then randomly assign subjects to treatment groups (within each block for a randomized blocks design), then any apparent correlation between covariate and independent variable is due to sampling error, and statistically removing the effect of the covariate removes only error variance.

If, however, we cannot randomly assign subjects to levels of the independent variable or if our covariate is measured after administering the treatments, then removing the effect of the covariate may also result in removing the effect of the treatment. In other words, when the independent variable and the extraneous variable are correlated (confounded), you cannot remove from the dependent variable variance due to the extraneous variable without also removing variance due to the independent variable.

Nonexperimental Research: Contrived Data

Now consider the case of nonexperimental research. Suppose that we have a nonmanipulated dichotomous variable and continuous data on a covariate and a comparison variable. Imagine that comparison variable is score on a reading aptitude test, the covariate is number of literature courses taken, and the grouping variable is sex/gender. Run the program **Confound.sas** from my [SAS programs page](#). Look at the data, which are included within the program. The first three columns of scores are sex/gender (1 is female), number of courses, and aptitude. We match participants on number of courses (before looking at their aptitude scores), obtaining 10 pairs of participants perfectly matched on the covariate. The 4th column of scores indicates matched pair number. Participants with a missing value code (a dot) in this column could not be matched, so they are excluded from the matched pairs analysis. Note that this excludes from the analysis the female participants with very high covariate scores (and, given a positive correlation with the criterion variable, with high aptitude as well) and the male participants with very low covariate (and criterion) scores. The last three columns of data are scores on the criterion variable for matched participants (female, male) followed by the difference score.

Matched Pairs

Look at the output from Proc Corr. Number of courses is indeed well correlated with aptitude, and the women scored higher than the men on both courses and aptitude (the negative sign of the point biserial correlation coefficients indicating that the gender 2 scores are lower than the gender 1 scores).

The Ttest output shows us again that women score significantly higher than men on both courses and aptitude, and gives us the means etc. Note that the analyses so far are based on all 34 cases.

The Proc Means output shows us that with the matched pairs data, men have reading aptitude ($M = 42.5$) significantly greater than that of women ($M = 37.5$). Now, can we make sense out of this?

Ignoring the covariate, women had a significantly higher mean than did men, but if we “control” the covariate by matching (excluding high scores from one group and low scores from the other group), we not only remove Group 1’s superiority, but we get Group 2 having the significantly higher mean. In other words, if the two groups did not differ on the covariate, Group 2 would have the higher mean -- but the two groups do differ on the covariate, so asking if the groups would differ on reading aptitude if they did not differ on number of literature courses may be a rather strange question to ask.

ANCOV (analysis of covariance)

Now, let us do a quick ANCOV using all 34 participants. I used the GLM (General Linear Model) procedure, which is especially convenient for testing linear models which have a mixture of categorical and continuous predictor variables, the latter generally being referred to as covariates. PROC GLM is first used to test an interaction term. Look back at the data step to see how I defined the interaction term -- the product of gender and courses. Gender is identified as a CLASSification (categorical) variable, aptitude as the comparison variable. SS1 indicates that I want sequential sums of squares (each effect in the model being adjusted to exclude overlap with effects to its left but not effects to its right, which is computationally more efficient than other types of sums of squares). The *F* reported for the interaction component tests the null hypothesis that the slope for predicting aptitude from courses is the same in women as in men. This must be so if we are to do a standard ANCOV. The *F* is clearly nonsignificant, so we go on to do the ANCOV with the interaction term dropped from the model. In this analysis, the effect of the covariate is first removed. The LSMEANS are our estimates of what the group means on aptitude would be if the groups did not differ on number of literature courses taken. Since the women had high covariate scores and the men had low covariate scores, the adjusted mean on the comparison variable was lowered in the women and raised in the men. The *F* reported for gender in this analysis tests the null that the two **adjusted means** (given under LSMEANS) are equal in the population. After taking out the “effect” of number of literature courses, men have a mean reading aptitude that is significantly higher than that of women. Once again, statistically controlling the covariate with these confounded data has resulted not only in removing Group 1’s superiority but in producing a significant difference in the opposite direction. Later we shall discuss such results in terms of the “reversal paradox” and “net suppression.”

Please beware the use of matching or ANCOV in circumstances like this. I have contrived these data to make a point, exaggerating the degree of confounding likely with real data, but we shall see this problem with real data too. For our contrived data, women have significantly higher reading aptitude unless we statistically remove the “effect” of taking more literature courses. Does this mean that men really have higher reading aptitude that is just masked by their not taking many literature courses? I doubt it. People generally take more courses in areas where their aptitude is high rather than low, so statistically removing the gender difference in number of literature courses taken also removes (or reduces or even reverses) the (real, unadjusted) sex/gender difference in aptitude. Suppose Group 1 was men, Group 2 women, the covariate a measure of amount eaten daily, and the criterion body weight. Men are significantly heavier than women, but if we statistically hold constant amount eaten, women have higher weights than do men. If women ate as much as men, they would weigh more than men. So what, not eating that much is part of being a woman, women eat significantly less than men do!

Despite numerous warnings from statisticians about such use of matching and ANCOV, psychologists persist in doing it. You be a critical reader and be aware of the severe limitations of such research when you encounter it.

Nonexperimental Research: Actual Data

Lest I have overstated the case against ANCOV and matching with covariates confounded with the categorical variable, let me state that I believe such analyses can be informative when interpreted

with caution and understanding. Multiple regression (which is really what we are doing here) generally involves obtaining partialled (adjusted) statistics (reflecting the contribution of each predictor variable partialled for some or all of the other predictor variables). Such analyses are especially useful with nonexperimental data, where causal attribution is slippery at best. Consider the data collected by statistics student Dechanile Johnson, and used in PSYC 6430 as her Personal Data Set. The data are in the file **Weights.sas** on my SAS programs page. Run the program. The variables are gender, height, and weight. The program does an ANCOV to compare the genders on weight, using height as the covariate. Proc Corr shows us that height is well correlated with weight, and that men are significantly taller and heavier than women (point biserial correlations). Proc Ttest gives us the means by gender along with associated statistics. Proc GLM shows us that the slope for predicting weight from height does not differ significantly between men and women, and that men still weigh significantly more than women after adjusting for height. The means show us that the men averaged $163.76 - 123.36 = 40.4$ lb. heavier than the women and $70.571 - 64.893 = 5.678$ inches taller. These are quite large differences, 2.5 standard deviations in the case of weight, 2.34 in the case of height. LSMEANS shows us that the adjusted means differ by less, by only 35.2 lb (160.8 - 125.6).

Removing the effect of height did not make the weight difference nonsignificant (if it did, would we conclude that men don't really weigh more than women?), but it did reduce the difference from 40.4 to 35.2. In other words, some part of the sex difference in weight is due to men being taller, but even if we statistically hold height constant, men are significantly heavier. Why? Well, men have stockier builds and perhaps more dense tissue (more muscle, less fat, not to mention denser crania ☺).

Including Covariate in Model May Reduce Effect to Nonsignificance

Colom, R., Escorial, S., & Rebollo, I., (2004, Sex differences on the Progressive Matrices are influenced by sex differences on spatial ability, *Personality and Individual Differences*, 37, 1289-1293. [doi:10.1016/j.paid.2003.12.014](https://doi.org/10.1016/j.paid.2003.12.014)) noted that men score higher on women on the Progressive Matrices Test (PM), which is thought to measure general intelligence (g). It is, however, generally acknowledged that g does not differ by any meaningful amount between men and women. The authors administered the Advanced Progressive Matrices Test (APM) to a sample of undergraduate university students. As expected the men scored significantly higher than did the women. They also administered the Spatial Rotation Test (SRT) from the Primary Mental Abilities Battery. An ANCOV comparing the sexes on APM indicated that the sexes did not differ significantly when SRT was entered as a covariate. These results were interpreted as indicating that the APM, as a measure of g, is biased against women because of its visuospatial format.

[Return to Wuensch's Stats Lessons Page](#)

Copyright 2014, Karl L. Wuensch - All rights reserved.