Change Score Analysis versus ANCOVA in Pretest/Posttest Designs: The Assignment Mechanism Matters

Justin C. Mary, Dale E. Berger, Giovanni Sosa, & Christopher Pentoney
Claremont Graduate University

Gain Scores VS. ANCOVA

Two common methods for comparing change across paired observations for two groups are the analysis of change or gain scores (Posttest – Pretest) and Analysis of Covariance (ANCOVA).

For example, suppose we wish to assess gender differences in the effects of an intervention on test performance. Participants were tested both before and after engaging in the intervention.

Should we use ANCOVA to covary out pretest scores, or should we conduct a *t*-test on change scores?

For over 40 years gain score analysis (GSA) has been criticized for possible bias, low reliability, and low power. These criticisms have been successfully challenged as being too broad.

Gain scores are often biased when assignment to treatment groups is based on pretest scores. However, ANCOVA is often biased when intact groups are studied. Graphs of these two testing situations are shown here to demonstrate the bias inherent in both methods.

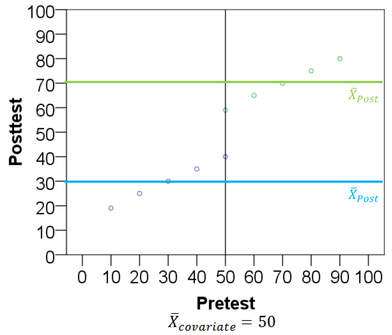
Mary, J.C., Berger, D.E., Sosa, G.W., & Pentoney, C. (2012, April). Change score analysis versus ANCOVA in pretest/posttest designs: The assignment mechanism matters. Poster presented at the meeting of the Western Psychological Association, Burlingame, CA.

ANCOVA is Often Biased when Intact Groups are Studied

ANCOVA adjusts the means for men and women to be equal on the covariate (50 is the overall mean on the covariate). Imagine sliding the points along the regression line and lining them up at the covariate mean.

Women: Pre = 70, Post = 70, Difference = 0
Men: Pre = 30, Post = 30, Difference = 0

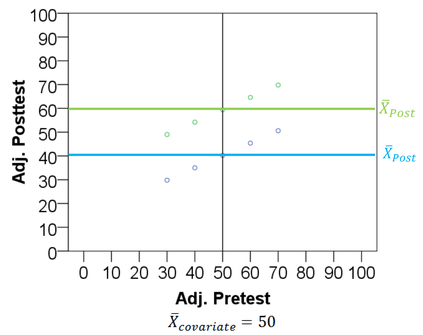
GSA Unbiased. Difference scores both zero



Although neither group improved from pretest to posttest, ANCOVA shows the test of the group effect on the posttest is significant after covarying out pretest scores.

Women: Adj Pre = 50, Adj Post = 60
Men: Adj Pre = 50, Adj Post = 40

ANCOVA Adjusted Posttest means are not equal.



ANCOVA Interpretation: After equating men and women on pretest score, the effect of the intervention was larger for women than for men.

Gain Score interpretation: There was no gain for either men or women.
Post – Pre = 0 for both men and women.

This discrepancy in findings between gain score analysis and ANCOVA is known as Lord's Paradox[3] Using the ANCOVA approach, a researcher may falsely conclude that there are gender differences in gains, when in fact there were none. This is called a "regression artifact" or pseudo effect[3] because ANCOVA under-corrects for pretest effects.

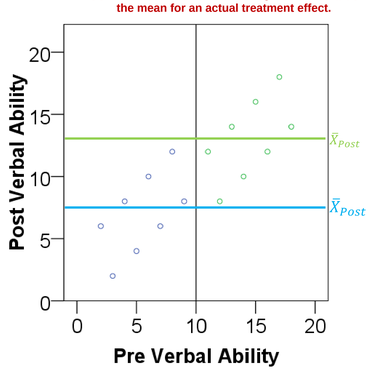
Lord's paradox can be attributed to measurement error[4] and trait instability[2]; this paradox can be eliminated only when the slope of the within-group regression lines = 1.[2]

Gain Scores are Biased when Assignment is Based on Pretest Scores

There is no overall treatment effect (overall pre and posttest means are both 10).

Control: Pre = 14.5, Post = 13.0, Difference= -1.5
Intervention: Pre = 5.5, Post = 7.0, Difference = 2.5

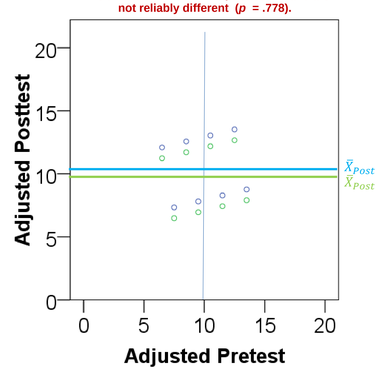
GSA is biased. GSA mistakes regression towards the mean for an actual treatment effect.



ANCOVA appropriately adjusts for regression towards the mean

Control: Adj Pre = 10.0, Adj Post = 9.6, Difference = -.4
Intervention: Adj Pre = 10.0, Adj Post = 10.4, Difference = +.4

ANCOVA is unbiased. The posttest means are not reliably different (p = .778).



ANCOVA Interpretation: When the posttest scores are adjusted to the predicted values if the two groups had the same average pretest score, there is no reliable difference between the control group and the intervention group on posttest performance

Gain Score interpretation: Participants in the treatment group had larger gain scores than the control group.

In this example, assignment was based on pretest scores and regression towards the mean artificially increased scores in the Intervention condition and decreased scores in the Control condition.

ANCOVA assumes that any pretest differences are due to sampling error and corrects for regression towards the mean by adjusting scores for the group with the highest mean downward and adjusting scores from the group with the lower mean upwards.

GSA does not adjust scores to account for regression towards; therefore, it is biased.

Scenario 1: ANCOVA is Biased

ANCOVA assumes that any group differences on the covariate are due to sampling error[1]; posttest means are adjusted to account for regression towards the mean. The test of significance in ANCOVA is performed on these adjusted means[2].

When groups are assigned non-randomly, the assumption that groups have the same population mean on the covariate may be invalid and any adjustments are suspect.

Gain scores analysis does not assume that pretest scores are equivalent across groups. Gain score analysis treats any differences between groups as a real effect.

Thus, when pretest differences are real, gain scores are unbiased and ANCOVA is biased. These findings have been confirmed by several simulation studies[4,5] .

Scenario 2: GSA is Biased

When participants are assigned to groups based upon the pretest, regression towards the mean is likely to occur and ANCOVA will be unbiased, while GSA will be biased[1]

For example, assume students (N = 16) were randomly assigned to an intervention group or a control group based upon pretest scores. That is, any student scoring above the mean score of 10 was placed in the control group and everyone below the mean into the intervention group.

Regression towards the mean suggests that students with extremely low scores are likely to have higher posttest scores and students with extremely high scores on the pretest are likely to have lower scores at posttest. Consequently, **even in the absence of a treatment effect**, students in the treatment group will likely have increased performance and students in the control group will likely have worse performance on the posttest. ANCOVA corrects for regression to the mean, while GSA does not.