Does Group-Mean Centering Always Inflate Type I Error Rates in Multiple Regression Analyses?

Robert E. Larzelere and Hua Lin
Oklahoma State University

Huitema (2011) proposed quasi-ANCOVA to increase statistical power in randomized designs by controlling for covariates that were measured after the start of treatment. Quasi-ANCOVA accomplishes this with group-mean centering, thereby removing any effect of treatment on the covariate (which would otherwise bias the analysis toward non-significance). Lin (2018; Lin & Larzelere, 2020) extended that strategy to dual-centered ANCOVA by centering the posttest scores as well as the pretest scores on the pretest group means. That strategy produces robust treatment estimates from ANCOVA and difference-score analyses of the dual-centered data. Those robust results replicate the treatment effect from the original difference-score analysis prior to group-mean centering. Following Huitema (2011), we assumed that the use of ANCOVA to analyze dual-centered data increased the statistical power for predicting difference-score estimates of treatment effects. However, Lin's simulations indicate that the standard deviation of treatment effects across 1000 replications corresponds to the standard deviation of treatment effects for the original difference-score analysis rather than the SD of treatment effects under the original standard ANCOVA (Table 1).

Since submitting this proposed presentation for this conference, B. Wade Brorsen (Brorsen et al., 2023) showed us that these results occur because group-mean centering is an example of a *generated regressor* (Pagan, 1984). A generated regressor is a predictor in one regression equation that is generated from another regression equation. The correct standard error for treatment effects in models with generated regressors, such as quasi-ANCOVA and dual-centered ANCOVA, must be estimated with two-stage least squares (or maximum likelihood in some cases), not with ordinary least squares regression.

Generated regressors can produce incorrect standard errors in randomized pre-post designs as well as nonequivalent comparison-group designs based on Lin's simulations. That is, OLS regression can produce incorrect standard errors in regression equations with generated regressors even in designs lacking systematic differences between the pretest group means. Any pretest group differences in randomized pre-post designs occur only due to random variations around the same grand mean at the population level (expected mean values).

How then does standard ANCOVA and dual-centered ANCOVA differ from each other in the randomized pre-post design? Standard ANCOVA correctly predicts the shrinkage in the difference between the pretest group means from the pretest to the posttest, as a function of the difference between the pretest group means and the within-group regression coefficient predicting posttest scores from pretest scores (Figure 1). Note that larger discrepancies from the expectation of equal pretest group means $\bar{\delta}_0$ result in larger regression toward the grand mean on the posttest $\tilde{\delta}_1$. But mean values on the random variations in pretest group means (i.e., equal pretest group means, $\bar{d}_0 = 0$) do not regress toward the grand mean, because they are already at the grand mean (of positive or negative signed differences between the group means).

Paper for the Modern Modeling Methods conference, Storrs, CT, June 28, 2023. We are grateful for help from David Kenny, Joshua Habiger, Isaac Washburn, & Tianyu Liu and for funding from NICHD grant 5 R03 HD107307

In contrast to standard ANCOVA, group-mean-centered ANCOVA always forces the pretest group means to be equal to each other. It is as though group-mean centering makes each centered pretest group mean predict the posttest score as though it were the grand mean of random variations around an expectation of equal group means (second panel, Figure 1). Under the null hypothesis then, dual-centered ANCOVA always predicts that the two original pretest group means will maintain their difference from each other on the posttest, an expectation that is identical to the expected posttest scores of the original differences-score analysis under its null hypothesis (i.e., parallel slopes from pretest to posttest).

This explanation implies that when group-mean centering is incorporated into a regression-based analysis and is analyzed as an OLS regression analysis, the standard error of estimate can be under-estimated, because OLS regression models treat it like an ordinary ANCOVA-like prediction of a residualized score, when the correct standard error should be the same as the corresponding standard error in a two-stage least squares analysis.

Pagan (1984, Model 4, pp. 232-233) considers the case of multilevel modeling (nested data) and concludes that OLS regression produces the correct standard error of estimate for residuals from the generated regressor, but not for the generated regressor itself.

Do multilevel modeling programs base their statistical tests on the correct standard error at higher levels (e.g., Level 2)? Hua Lin has started doing some simulation analyses in R to help to answer this question. She is now trying to check her simulations, including verifying that she can get the same result in corresponding linear growth models.

In any case, we need to be sure that the standard errors are correct in general linear models that incorporate generated regressors, such as group-mean centering, in their models.

## References


Brorsen, B. W., Larzelere, R. E., & Lin, H. (2023). *Critique of enhanced power claimed for Quasi-ANCOVA and Dual-Centered ANCOVA*. Oklahoma State University.

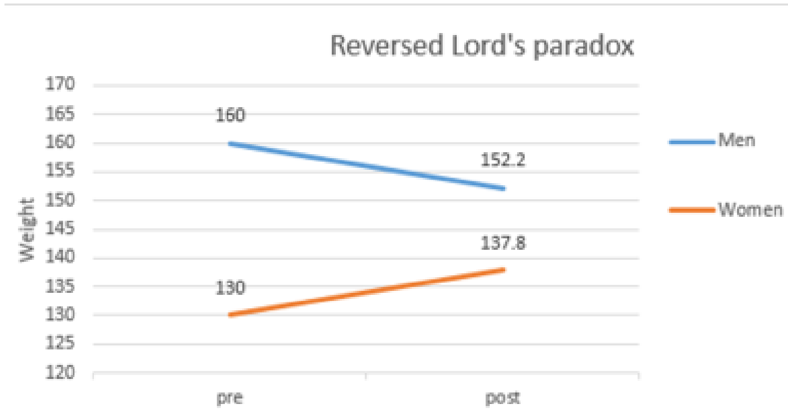
Huitema, B. E. (2011). *The analysis of covariance and alternatives* (2nd ed.). Wiley. https://doi.org/10.1002/9781118067475

Lin, H. (2018). *Revealing and resolving contradictory ways to reduce selection bias to enhance the validity of causal inferences from non-randomized longitudinal data* [Dissertation, Oklahoma State University]. Stillwater, OK.

Lin, H., & Larzelere, R. E. (2020). Dual-centered ANCOVA: Resolving contradictory results from Lord's paradox with implications for reducing bias in longitudinal analyses. *Journal of Adolescence*, *85*, 135-147. https://doi.org/10.1016/j.adolescence.2020.11.001

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, *25*(1), 221-247.

## Null Hypothesis for ANCOVA




## Null Hypothesis for Difference-Score Analysis

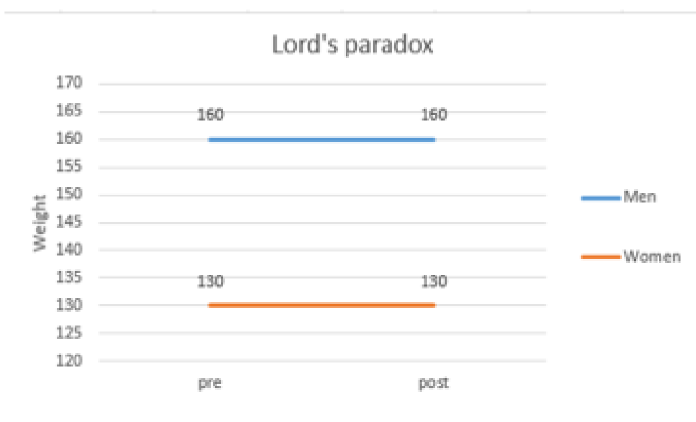



Figure 1. Distinct null hypotheses for ANCOVA and for Difference-Score Analysis (Illustrated for simulating the null hypotheses of the two change-score analyses of Lord's Paradox. Randomized pre-post designs would have a similar pattern relative to each other for their respective null hypotheses for an extreme random variation around the expected grand means of 145 for both groups at the Pretest.)

Paper for the Modern Modeling Methods conference, Storrs, CT, June 28, 2023. We are grateful for help from David Kenny, Joshua Habiger, Isaac Washburn, & Tianyu Liu and for funding from NICHD grant 5 R03 HD107307

Table 1. Simulated results of five analyses of randomized pre-post data generated under the null hypothesis.

| id | mod_name | model | rho | b_y0 | s_b_y0 | b_Tx | s_b_Tx | sd_b_Tx | TSS | SSREG_y0 | SSREG_Tx | SSE | MSE | v_error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ANOVA | y1 ~ Tx | 0.00 | - | - | -0.009 | 0.9482 | 0.9760 | 224651.4 | - | 237.92 | 224413.5 | 224.86 | 225 |
| 2 | ANCOVA | y1~y0+Tx | 0.00 | 0.00 | 0.032 | -0.011 | 0.9487 | 0.9767 | 224651.4 | 220.2 | 237.98 | 224193.3 | 224.87 | 225 |
| 3 | Quasi-ANCOVA | y1 ~ y0_c + Tx | 0.00 | 0.00 | 0.032 | -0.009 | 0.9482 | 0.9760 | 224651.4 | 220.2 | 237.92 | 224193.3 | 224.87 | 225 |
| 4 | dif_in_dif | y10 ~ Tx | 0.00 | - | - | -0.023 | 1.3403 | 1.3209 | 448827.1 | - | 435.91 | 448391.2 | 449.29 | 450 |
| 9 | DC_ANCOVA | y1_c ~ y0_c + Tx | 0.00 | 0.00 | 0.032 | -0.023 | 0.9482 | 1.3209 | 224849.4 | 220.2 | 435.91 | 224193.3 | 224.87 | 225 |
| 1 | ANOVA | y1 ~ Tx | 0.20 | - | - | -0.008 | 0.9482 | 0.9820 | 224655.2 | - | 240.87 | 224414.4 | 224.86 | 225 |
| 2 | ANCOVA | y1~y0+Tx | 0.20 | 0.20 | 0.031 | -0.012 | 0.9295 | 0.9506 | 224655.2 | 9207.9 | 228.49 | 215218.9 | 215.87 | 216 |
| 3 | Quasi-ANCOVA | y1 ~ y0_c + Tx | 0.20 | 0.20 | 0.031 | -0.008 | 0.9290 | 0.9820 | 224655.2 | 9195.5 | 240.87 | 215218.9 | 215.87 | 216 |
| 4 | dif_in_dif | y10 ~ Tx | 0.20 | - | - | -0.021 | 1.1988 | 1.1815 | 359061.7 | - | 348.73 | 358713.0 | 359.43 | 360 |
| 9 | DC_ANCOVA | y1_c ~ y0_c + Tx | 0.20 | 0.20 | 0.031 | -0.021 | 0.9290 | 1.1815 | 224763.1 | 9195.5 | 348.73 | 215218.9 | 215.87 | 216 |
| 1 | ANOVA | y1 ~ Tx | 0.40 | - | - | -0.006 | 0.9482 | 0.9876 | 224650.5 | - | 243.62 | 224406.9 | 224.86 | 225 |
| 2 | ANCOVA | y1~y0+Tx | 0.40 | 0.40 | 0.029 | -0.012 | 0.8694 | 0.8827 | 224650.5 | 36147.2 | 200.13 | 188303.2 | 188.87 | 189 |
| 3 | Quasi-ANCOVA | y1 ~ y0_c + Tx | 0.40 | 0.40 | 0.029 | -0.006 | 0.8690 | 0.9876 | 224650.5 | 36103.7 | 243.62 | 188303.2 | 188.87 | 189 |
| 4 | dif_in_dif | y10 ~ Tx | 0.40 | - | - | -0.018 | 1.0382 | 1.0232 | 269296.3 | - | 261.55 | 269034.7 | 269.57 | 270 |
| 9 | DC_ANCOVA | y1_c ~ y0_c + Tx | 0.40 | 0.40 | 0.029 | -0.018 | 0.8690 | 1.0232 | 224668.4 | 36103.7 | 261.55 | 188303.2 | 188.87 | 189 |
| 1 | ANOVA | y1 ~ Tx | 0.60 | - | - | -0.004 | 0.9481 | 0.9927 | 224635.2 | - | 246.12 | 224389.0 | 224.84 | 225 |
| 2 | ANCOVA | y1~y0+Tx | 0.60 | 0.60 | 0.025 | -0.012 | 0.7589 | 0.7644 | 224635.2 | 81030.6 | 152.86 | 143451.7 | 143.88 | 144 |

Paper for the Modern Modeling Methods conference, Storrs, CT, June 28, 2023. We are grateful for help from David Kenny, Joshua Habiger, Isaac Washburn, & Tianyu Liu and for funding from NICHD grant 5 R03 HD107307

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Quasi-ANCOVA | y1 ~ y0_c + Tx | 0.60 | 0.60 | 0.025 | -0.004 | 0.7585 | 0.9927 | 224635.2 | 80937.4 | 246.12 | 143451.7 | 143.88 | 144 |
| 4 | dif_in_dif | y10 ~ Tx | 0.60 | - | - | -0.015 | 0.8477 | 0.8354 | 179530.9 | - | 174.37 | 179356.5 | 179.72 | 180 |
| 9 | DC_ANCOVA | y1_c ~ y0_c + Tx | 0.60 | 0.60 | 0.025 | -0.015 | 0.7585 | 0.8354 | 224563.4 | 80937.4 | 174.37 | 143451.7 | 143.88 | 144 |
| 1 | ANOVA | y1 ~ Tx | 0.80 | - | - | -0.002 | 0.9480 | 0.9970 | 224603.3 | - | 248.23 | 224355.0 | 224.80 | 225 |
| 2 | ANCOVA | y1~y0+Tx | 0.80 | 0.80 | 0.019 | -0.009 | 0.5691 | 0.5681 | 224603.3 | 143841.3 | 86.57 | 80675.4 | 80.92 | 81 |
| 3 | Quasi-ANCOVA | y1 ~ y0_c + Tx | 0.80 | 0.80 | 0.019 | -0.002 | 0.5688 | 0.9970 | 224603.3 | 143679.6 | 248.23 | 80675.4 | 80.92 | 81 |
| 4 | dif_in_dif | y10 ~ Tx | 0.80 | - | - | -0.010 | 0.5994 | 0.5907 | 89765.4 | - | 87.18 | 89678.2 | 89.86 | 90 |
| 9 | DC_ANCOVA | y1_c ~ y0_c + Tx | 0.80 | 0.80 | 0.019 | -0.010 | 0.5688 | 0.5907 | 224442.2 | 143679.6 | 87.18 | 80675.4 | 80.92 | 81 |

Notes
y0: pretest
y1: posttest
y0_c: pretest centered on pretest group mean
y1_c: posttest centered on pretest group mean
Tx: treatment

rho: within-group correlation between pre and post
b_y0: estimated coefficient for the covariate (pretest)
s_b_y0: standard error of the estimated coefficient for the covariate (OLS)
b_Tx: estimated treatment effect
s_b_Tx: standard error of the estimate of the treatment effect (OLS)
sd_b_Tx: SD of the 1000 estimated treatment effects
TSS: total sum of squares
SSREG_y0: sum of squares for the pretest
SSREG_Tx: sum of squares for Tx
SSE: sum of squares for error
MSE: mean squared error from the simulation
v_error: variance of error based on calculation

| | |
|---|---|
| mean of pretest for Tx | 130 |
| mean of pretest for Cntrl | 130 |
| mean of posttest for Tx | 130 |
| mean of posttest for Cntrl | 130 |
| SD for the pre- and post-test scores within a group at one time | 15 |

dif_in_dif: differences in differences
DC_ANCOVA: Dual-centered ANCOVA