

Baseline Mean Centering for Analysis of Covariance (ANCOVA) Method of Randomized Controlled Trial Data Analysis

Jennifer Scodes, New York State Psychiatric Institute, New York, NY

ABSTRACT

Many analytical approaches exist to compute treatment effects and within-group changes from baseline for data analysis of randomized controlled trials with multiple follow-up visits. One of these approaches is the analysis of covariance (ANCOVA) method, in which baseline values are included as a covariate instead of as an outcome. Using the ANCOVA method, the treatment effects can be easily computed from model estimates; however, within-group changes from baseline cannot be directly computed in SAS procedures without centering the outcome measures by the overall baseline mean. This paper will present a macro that can be used to analyze data from two-arm randomized controlled trials using the ANCOVA method to compute and present both treatment effects and within-group changes from baseline using baseline mean centering. This paper is intended for all levels of SAS® users that analyze clinical trial data.

INTRODUCTION

When analyzing data from a standard two-arm randomized controlled trial (RCT), there are always two main questions that investigators ask: 1) what are the treatment effects, and 2) what are the within group changes from baseline. The first question asks whether or not one treatment group is doing better than the other, known as the between-group difference, and the second question asks by how much has each treatment group changed on average since study entry, known as the within-group difference.

There are many analytical methods that are commonly used to assess the between-group and within-group differences, but there is much debate about which is the most appropriate method. A few of the more common ways that investigators compute these estimates include: longitudinal analysis of covariance (ANCOVA), where follow-up values are outcomes and baseline value is a covariate; simple analysis of change, where change scores (follow-up values - baseline values) are outcomes and the baseline value is a covariate; or repeated measure analysis, where follow-up and baseline values are outcomes (1). Each of these methods have been widely used to analyze study aims, but current literature is in favor of models that adjust for baseline values because they have more power to find the treatment effect significant (2, 3). Although the treatment effect can be easily computed from model estimates from each of these methods, the problem lies in computing the within-group change when baseline is being included as a covariate in the model.

In order to compute within-group effects, the baseline and follow-up measures need to be preprocessed by centering with the baseline mean. By centering both the outcome and baseline measures with the baseline mean value, the model-estimated adjusted means (i.e., LSMeans) of the treatment effect (or treatment by time interaction, if more than one follow-up assessment is modelled) can then be directly interpreted as the mean difference from baseline after controlling for baseline values. This allows the within-group effects to be easily computed even in models that include baseline values as a covariate.

This paper will present how the ANCOVA method with baseline mean centering can be used to compute the between-group and within-group effects of a standard two-arm RCT with multiple follow-up visits. Example data is from a randomized trial comparing extended-release naltrexone to buprenorphine for individuals with opioid use disorders. This paper will also present a macro to compute and report both the between-group and within-group effects along with the appropriate tests of significance.

UNDERSTANDING ANALYSIS OF COVARIANCE (ANCOVA) MODELS

In general, an ANCOVA is an analysis in which measures that are associated with the outcome variable are included in the model as covariates in order to explain more of the variability in the outcomes. Doing this decreases the variability of the error terms which in turn gives more power to find treatment effects

significant. In RCTs, the baseline outcome measure is typically highly associated with follow-up outcome measures, and inclusion of the baseline outcome measure as a covariate therefore gives a more precise treatment effect.

The typical repeated-measures ANCOVA regression model with two treatment groups, denoted by *trt* {0,1}, a baseline visit, and two follow-up visits, denoted by *time* {0, 1}, is as follows:

$$(1) \quad Y_{ij} = \beta_0 + b_{0i} + \beta_1 * Y_{i0} + \beta_2 * trt_i + \beta_3 * time_j + \beta_4 * trt_i * time_j$$

In model (1), the subscript *i* is the index for subject, *j* is the index for follow-up time point. The variable *Y* is the outcome value, β_0 is the overall model intercept, b_{0i} is the random intercept for subject *i* since we will use longitudinal data, β_1 is the slope of the association between the outcome at baseline and follow-up, β_2 is the treatment effect at the referent time point, β_3 is the difference in outcome between follow-up visits for the referent treatment group, and β_4 is the difference in treatment effect between the follow-up visits. With this model, the between group treatment effects at each follow up time point are computed as: β_2 for the first follow-up visit and $\beta_2 + \beta_4$ for the second follow-up visit. The within-group effects are not easily computed without the following adjustment to the model:

$$(2) \quad Y_{ij} - \bar{Y}_0 = \beta_0 + b_{0i} + \beta_1 * (Y_{i0} - \bar{Y}_0) + \beta_2 * trt_i + \beta_3 * time_j + \beta_4 * trt_i * time_j$$

Model (2) is similar to (1) except that it contains an additional parameter of \bar{Y}_0 which is the grand mean of the baseline outcome measures across all subjects. The two models are identical in that they produce the same error variance estimates (4). However, in order to compute the overall mean change from baseline to follow-up visit by treatment group (i.e., when $(Y_{0i} - \bar{Y}_0) = 0$), we have:

TREATMENT EFFECTS (BETWEEN-GROUP EFFECTS)

- Treatment effect at follow-up visit 1 = β_2
- Treatment effect at follow-up visit 2 = $\beta_2 + \beta_4$

WITHIN-GROUP EFFECTS

- Change from baseline to follow-up visit 1 for referent group = β_0
- Change from baseline to follow-up visit 2 for referent group = $\beta_0 + \beta_3$
- Change from baseline to follow-up visit 1 for non-referent group = $\beta_0 + \beta_2$
- Change from baseline to follow-up visit 2 for non-referent group = $\beta_0 + \beta_2 + \beta_3 + \beta_4$

All of these contrast can be easily computed using the LSMeans statement in PROC GLIMMIX.

MACRO FOR ANCOVA WITH BASELINE MEAN CENTERING

The macro below performs four main tasks:

1. Computes Baseline Mean: It computes the mean of the outcome measure at baseline across groups and saves these values as macro variables. Then it centers the outcome and baseline measure by this grand mean value.
2. Runs ANCOVA Model: It runs the ANCOVA model using PROC GLIMMIX to allow for fixed and random effects, if necessary, and uses the LSMEANS statements to output all of the necessary contrasts for treatment effects, and within-group changes from baseline.
3. Organizes Results: It uses data steps to organize the treatment effects and within group differences that are produced from the LSMEANS statement.
4. Presents Results: It outputs the model results including the model estimated means and mean differences using PROC REPORT.

To use the macro, the data needs to be in long format, include a column with the baseline value of the outcome (to be used as a predictor), and the treatment variable should be dichotomized as 0 or 1.

The code for the macro is presented in the appendix.

EXAMPLE USING CLINICAL TRIAL NETWORK DATA

The following example uses data from a two-arm randomized trial of extended release naltrexone versus buprenorphine for individuals with opioid use disorder. There were 570 individuals enrolled in the study that were randomly assigned to receive either a once a month injection of naltrexone or a once daily dose of buprenorphine. The individuals were then followed for 24 weeks. During study entry and follow-up visits, study sites collected data such as demographic information, substance use, treatment history, and physical and mental health. Specific information about the study and primary aims are presented elsewhere (5).

One specific measure that was collected included the Stroop Color and Word test, which is a test that measures cognitive processing abilities. Each individual was administered this task at study entry and follow-up. Using these data, we assessed whether there was a treatment effect on cognitive processing based on the Stroop task, and by how much did cognition improve during trial follow-up within each treatment group. The results are presented in Table 1 and model estimated adjusted means are presented in Figure 1.

Time	Within Group Differences								Treatment Effect (Bup vs Nal)	
	Buprenorphine Group				Naltrexone Group					
	Adj Mean	SE	Difference from BL	p-value	Adj Mean	SE	Difference from BL	p-value	Mean Difference	p-value
0	41.62				41.62					
4	44.62	0.56	3.00	<.001	45.67	0.65	4.05	<.001	-1.05	0.222
8	47.86	0.60	6.24	<.001	47.31	0.73	5.69	<.001	0.55	0.563
16	48.88	0.68	7.26	<.001	48.95	0.80	7.33	<.001	-0.08	0.942
24	48.35	0.61	6.74	<.001	48.09	0.67	6.47	<.001	0.27	0.768

Table 1. Model Results for Stroop Color and Word test

There was no significant treatment difference between Stroop Color and Word test scores between the buprenorphine and naltrexone group at any of the follow-up visits. However, within both the buprenorphine and naltrexone groups there was a significant increase in scores from baseline to each follow-up time point (all p-values <.001).

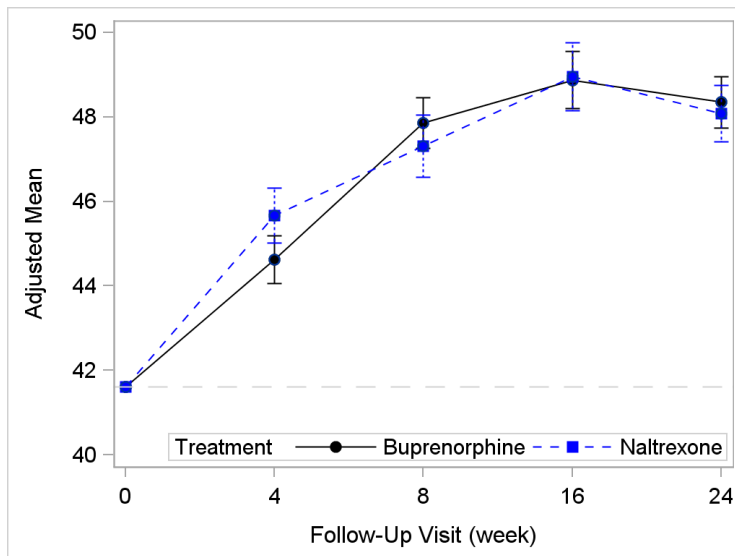


Figure 2. Model Estimated Means with Standard Error Bars for Stroop Color and Word Test

CONCLUSION

Even though current literature shows that adjusting for baseline is the most powerful approach to detect treatment effects in RCTs, some researchers still do not apply this method to analyze RCT data. This might be the case because within-group changes are more difficult to compute in SAS and other statistical software without applying baseline mean centering. However, hopefully with more widespread knowledge of how baseline mean centering solves this computational problem, adjusting for baseline outcome measures will become a more common practice in analyzing randomized trial data.

REFERENCES

1. Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. 2018. "Different ways to estimate treatment effects in randomised controlled trials." *Contemporary Clinical Trials Communications*, 10: 80–85.
2. Van Breukelen G. 2006. "ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies." *Journal of Clinical Epidemiology*, 59: 920-925.
3. Nunes EV, Pavlicova M, Hu MC, Campbell A, Miele G, Hien D, and Klein DF. 2011. "Baseline matters: The importance of covariation for baseline severity in the analysis of clinical trials." *The American Journal of Drug and Alcohol Abuse*, 37(5): 446–452.
4. Mulligan NW and Wiesen C. 2003. "Using the analysis of covariance to increase the power of priming experiments." *Canadian Journal of Experimental Psychology*, 57: 152–166.
5. Lee J, Nunes EV, Novo P, et. al. 2018. "Comparative effectiveness of extended-release naltrexone versus buprenorphine-naloxone for opioid relapse prevention (X:BOT): a multicentre, open-label, randomized controlled trial." *Lancet*, 391: 309–18.

ACKNOWLEDGMENTS

Thank you to my colleagues at NYSPI who reviewed this paper and provided helpful insight and suggestions.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jennifer Scodes
Jms2374@cumc.columbia.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX

The following code is to run a centered ANOVA regression model for a two-armed clinical trial. The macro variables are:

- *Data*: the name of the data file
- *ID*: variable name for Subject ID
- *Y*: the outcome variable
- *Bly*: the baseline value of the outcome variable
- *BLtime*: numerical value of the baseline time-point
- *Time*: the time variable
- *Trx*: the treatment variable

```
%macro CenteredANCOVA(Data, ID, Y, BLY, BLtime, Time, Trx);

/*Step 1 - Compute Baseline Mean*/

*Find the mean across treatment groups for all participants of the outcome
variable at baseline;
proc means data=&data maxdec=3 n mean stddev;
  where &time =&BLtime;
  var &y ;
  output out=summary mean=mean stddev=stddev;
run;
data _null_;
  set summary;
  call symput("mean", Mean); run;

*Center the outcome and baseline measures by the baseline grand mean;
data long_analysis;
  set &data;
  c_&y = &y - &mean;
  c_&BLY = &BLY - &mean;
run;

/*Step 2 - Run ANCOVA model*/

*Run the model controlling for baseline, and output the LSmeans and
differences of LSmeans;
proc glimmix data=long_analysis plots=RESIDUALPANEL order=data;
  where &time ne &BLtime;
  class &id &trx &time /ref=last;
  model c_&y = c_&BLY &trx|&time / dist=normal solution;
  random int / subject=&id ;
  lsmeans &trx*&time / at c_&BLY=0 plot = mean(sliceby = treatment join) cl
diff OM;
  ods output lsmeans=lsmeans0 (where=(&trx=0))
lsmeans=lsmeans1 (where=(&trx=1)) diffs=diffs (where=(&time=_&time));
run;
```

```

/*Step 3 - Organize results*/

*Treatment effects- between group differences;
data BGD;
  set diffs(rename=(estimate=est_BGD df=df_BGD tvalue=t_BGD probt=p_BGD));
  keep &time est_BGD p_BGD;
run;

*Within group differences: Treatment=0 group;
data WGD0;
  set lsmeans0(rename=(estimate=est_WGD0 stderr=se_WGD0 df=df_WGD0
tvalue=t_WGD0 probt=p_WGD0));
  where treatment = 0;
  mean_WGD0 = est_WGD0 + &mean;
  keep visit est_WGD0 se_WGD0 df_WGD0 t_WGD0 p_WGD0 mean_WGD0;
run;

*Within group differences: Treatment=1 group;
data WGD1;
  set lsmeans1(rename=(estimate=est_WGD1 stderr=se_WGD1 df=df_WGD1
tvalue=t_WGD1 probt=p_WGD1));
  where treatment = 1;
  mean_WGD1 = est_WGD1 + &mean;
  keep visit est_WGD1 se_WGD1 df_WGD1 t_WGD1 p_WGD1 mean_WGD1;
run;

*Add row for baseline mean value;
data BL;
  visit=&BLtime;
  mean_WGD0=&mean;
  mean_WGD1=&mean;
run;

*Merge results to report into one data file;
data report_table;
  merge BL WGD0 WGD1 BGD;
  by &time;
run;

/*Step 4 - Output Results */

*Output a table that has the within group and between group means,
differences, and p-values;
proc report data=report_table nowd;
  column &time ("Within Group Differences" ("Trx=0 Group" mean_WGD0 se_WGD0
est_WGD0 p_WGD0)
  ("Trx=1 Group" mean_WGD1 se_WGD1 est_WGD1 p_WGD1))
  ("Treatment Effect" est_BGD p_BGD);
  define &time /group order=internal;
  label &time="Time" mean_WGD0="Adjusted/Mean" se_WGD0="SE"
est_WGD0="Difference/from BL" p_WGD0="p-value" mean_WGD1="Adjusted/Mean"
se_WGD1="SE" est_WGD1="Difference/from BL" p_WGD1="p-value"
est_BGD="Mean/Difference" p_BGD="p-value";
  format p_WGD0 p_WGD1 p_BGD PVALUE6.3 mean_WGD0 se_WGD0 est_WGD0 mean_WGD1
se_WGD1 est_WGD1 est_BGD 8.2;
run;
%mend;

```