

Nonparametric Test for a Semiparametric Mixed ANCOVA Model for a Nested Design

By

Maricar Moreno, MSc

Joseph Ryan Lansangan, PhD

Erniel Barrios, PhD

Presented by

Maricar Moreno

UP School of Statistics

Motivation

- In most multifactor experiments, factorial designs are often used to properly measure the response variable. When factors are crossed, all possible combination of the independent variables are of concern which also allows examination of interaction among factors. However, when some factors cannot be crossed to another factor, and each of these factors appear at only one level of another factor, then factors are said to be nested. (Schielzeth and Nakagawa, 2013).
- With the presence of covariates in a design, there is no existing model yet that properly measures the treatment effects of a nested experimental design, which also specifies the relationship of the response variable and covariate component in a nonparametric form.

Objective

- To estimate nonparametrically the covariate effect to capture all the possible heterogeneity among the experimental units and to estimate the nesting effect, then set these aside to properly test for the significance of the treatment effects.

Semiparametric Analysis of Covariance Model

- A semiparametric ANCOVA model is a model which specifies the relationship between the **response** variable and the **covariate** component in a **nonparametric** form.
- Alao (2016) postulated a semiparametric mixed ANCOVA model with nonparametric part corresponding to the fixed covariate and the parametric part corresponding to the random effects. The model is given by:

$$Y_{ijk} = f(X_{ijk}) + \tau_j + \delta_k + (\tau\delta)_{jk} + \epsilon_{ijk}, \begin{cases} i = 1, 2, \dots, n \text{ observations} \\ j = 1, 2, \dots, p \text{ treatments} \\ k = 1, 2, \dots, q \text{ treatments} \end{cases}$$

Postulated Semiparametric Mixed ANCOVA model for Nested Design

- The parametric part corresponds to the treatment effects and nested effect while the nonparametric part corresponds to the fixed covariate. In this postulated model, two factors cross with each other, and one factor is nested in one of them.

The model is defined as:

$$Y_{(jkl)i} = f(X_{(jkl)i}) + \tau_j + \delta_k + (\tau\delta)_{jk} + \gamma_{(k)l} + \epsilon_{(jkl)i}$$

$$\begin{cases} i = 1, 2, \dots, n \text{ observations} \\ j = 1, 2, \dots, p \text{ treatments} \\ k = 1, 2, \dots, q \text{ treatments} \\ l = 1, 2, \dots, m \text{ treatments} \end{cases}$$

Estimation Procedure

- The postulated model will be estimated through a hybrid backfitting algorithm by Alao and Barrios (2016), called B-ARMS (Bootstrapped Analysis of Covariance via REML with splines).

Algorithm of ARM

- **Step 1** Fit the nesting part of the model using REML method by ignoring $f(X_{(ijk)l})$ and the parametric part $\tau_j + \delta_k + (\tau\delta)_{jk}$ so that the model would be $Y_{(ijk)l} = \gamma_{k(l)} + \epsilon_{(jkl)i}$. This will contain estimates of the components which we are of interest.
- **Step 2:** Compute the partial residuals $p_{(ijk)l} = Y_{(ijk)l} - \hat{Y}_{(ijk)l}$. These partial residuals now contain information on $f(X_{(ijk)l})$ and $\tau_j + \delta_k + (\tau\delta)_{jk}$ which will then be used to estimate $f(X_{(ijk)l})$ and $\tau_j + \delta_k + (\tau\delta)_{jk}$ consecutively.

Algorithm of ARMS

- **Step 3:** Using smoothing spline, estimate $f(X_{(ijk)l})$ nonparametrically.
- **Step 4:** Compute the new partial residuals $p_{(ijk)l}^* = p_{(ijk)l} - \hat{f}(X_{(ijk)l})$. These partial residuals contain information on $\tau_j + \delta_k + (\tau\delta)_{jk}$ and thus, will be used to estimate the main effects.
- **Step 5:** Estimate the parametric part containing the main effects using REML. The random model would be $p_{(ijk)l}^* = \tau_j + \delta_k + (\tau\delta)_{jk} + \epsilon_{(jkl)i}$.
- **Step 6:** Compute for the new partial residuals $e_{(ijk)l} = Y_{(ijk)l} - \hat{f}(X_{(ijk)l}) - \hat{Y}_{k(l)}$. These partial residuals contain information on the nesting effect and then be used to estimate the components in Step 1.
- **Step 7:** Repeat steps 1 to 6 until convergence, i.e. the change in the new estimates from the previous estimates do not change more than the tolerance level (0.001)

Algorithm of B-ARMS

- **Step 1:** Generate the initial estimates and residuals $\emptyset_{(ijk)l} = Y_{(ijk)l} - \hat{Y}_{(ijk)l}$ by fitting the model (2) using ARMS. 200 bootstrap samples of residuals in Step 2 below are obtained using these residuals, while the initial estimates are used to compute for the new values of dependent variables in Step 3 below.
- **Step 2:** Generate new set of residuals by obtaining samples of $\emptyset_{(ijk)l}$ from Step 1 with replacement from $\{1, 2, \dots, n\}$. These residuals will be used to compute the new values of Y in Step 3.
- **Step 3:** Compute for the new values of the dependent variable,
 - $Y_{(ijk)l}^* = \hat{f}(X_{(ijk)l}) + \hat{\tau}_j + \hat{\delta}_k + (\widehat{\tau\delta})_{jk} + \hat{Y}_{k(l)} + \epsilon_{(ijk)l}$, where $\hat{f}(X_{(ijk)l})$, $\hat{\tau}_j$, $\hat{\delta}_k$, $(\widehat{\tau\delta})_{jk}$ and $\hat{Y}_{k(l)}$ are estimates from Step 1 and new residuals $\emptyset_{(ijk)l}^*$ from Step 2.
- **Step 4:** Fit the model from the pseudo data in Step 3 using ARMS. This contains the estimates of the components from the dependent variable $Y_{(ijk)l}^*$.

Power and Size of the Test

- In order to test the significance of the nesting effect given the postulated model, an empirical distribution of the test statistic will be obtained to test if the l^{th} nested factor has no effect on the response by obtaining an empirical distribution of the test statistic by repeating Steps 2 to 4 of B-ARMS algorithm R times, and then constructing a 95% bootstrap confidence interval using the empirical distribution. If the interval contains 0, we reject the null hypothesis.
- To evaluate the test, power and size of the test will be computed. The power of the test will be computed by constructing the confidence interval 200 times and the proportion of confidence interval that does not contain 0, or the rejected null hypothesis is the power of the test. For size of the test, data will be generated without the treatment effects and then confidence interval will be constructed using B-ARMS. The proportion of rejected null hypothesis is the size of the test.

Results

- The proposed test is correctly sized for random and mixed effects model. Size is also comparable to the baseline tests except for fixed effects model where the test is found to be incorrectly sized.
- The proposed test is found to be more powerful than the baseline test across the different percentage differences in treatment levels compared to the baseline test procedure.
- In the presence of contamination of the covariate, dominating covariate effect, misspecification error and non-normal errors, the proposed test procedure performs better than the baseline test.
- Notably, the proposed test is more powerful when the replicate size is small ($r=2$), than when the replicate size is large ($n=5$).