University of Twente

Bachelorthesis

Multivariate models for pretest posttest data and a comparison to univariate models

Sven Kleine Bardenhorst
s1543377

January 30, 2017

supervised by
Prof. Dr. Ir. Jean-Paul Fox
Drs. Sebie Oosterloo

Contents

Abstract

The pretest-posttest control group design is a popular design and frequently discussed in the literature. In this study multivariate models are investigated for pretest posttest data and a comparison is made with univariate methods, the change score method and the regressor variable method. Three simulation studies were conducted to investigate differences. The first study provides a basic comparison, while the second study focuses on the investigation of heterogeneous treatment effects. In the third simulation study, the analysis of two dependent variables are investigated. The results showed no significant differences between univariate and multivariate methods in the studies one and two. Nevertheless, the third study showed that the multivariate method provides higher power and a better Type-1 error rate compared to univariate methods when investigating two dependent variables. The multivariate method can provide better results in analyzing multivariate pretest posttest data, when compared to the univariate methods.

1 Introduction

1.1 Pretest-Posttest Control Group Design

In a wide variety of scientific fields like psychology, education and bio-medical sciences, the so-called "pretest-postest control group design" is frequently used to investigate the effects of a treatment given to participants. In this design, participants are (randomly) assigned to either the treatment group or the control group. The number of groups is not necessarily limited to two, an assignment to three or more groups is also possible.

The variable of interest is measured at two points in time. The first measurement is conducted before the treatment is given (pre) and the second afterwards (post). One advantage of the pretest-posttest design, compared to a simple posttest only design, is the possibility to take pretest differences into account, when analyzing the resulting data. Specifically when the treatment group is compared to a control group, this design allows the researcher to control for previous prevalent group differences, when investigating between-subject effects. Furthermore, pretest-posttest control group designs are able to establish causality between two investigated variables, when the subjects are randomly assigned to either the treatment group or the control group. As mentioned by Allison (1990), this design allows to rule out the possibility that variable Y causes variable X given the situation that the researcher is interested in the hypothesis that variable X causes variable Y. Furthermore, it also reduces the chance of spurious effects of confounding variables that may influence an effect of either the dependent variable X or the independent variable Y. Effects that occur due to multiple testing, like maturation or test effects, can also be assumed to be the same across groups (Campbell & Stanley, 1975).

An important condition of the prestest-posttest control group design is that subjects are randomly assigned to either the treatment group or the control group. However, a random assignment is not always possible but depends on the type of research one is conducting. Sometimes it is inevitable to make use of preexisting groups and quasi-experimental designs. For example, in clinical research, it can be ethical and morally mandatory to assign people to the treatment group due to their specific illness and their need for treatment. Another example is the evaluation of educational programs. To evaluate the effectiveness of a program, one group of students will be part of the program, while the other group will be the control group. It is quite obvious that randomly assigning students to one of these groups is impractical. It would be necessary to split existing classes. In practice, existing classes are used for the sake of convenience.

The problem with the non-randomized assignment is the possibility of pre-existing group differences. One of the classes could score initially higher on the target variable. Thus, the comparison would be biased when not controlling for these pre-existing differences.

1.2 Univariate methods

Due to the popularity of this research design, different methods have been developed in order to evaluate the data while accounting for non-randomization. When the variable of interest is continuous, the most commonly used statistical approach is the change score method for which the treatment effects are analyzed as a function of the difference between the pretest and the posttest (Brognan & Kutner, 1980). Another commonly used method includes the pretest measurement as a covariate in the analysis of the posttest measurement. This approach is referred to as the regressor variable method. Henceforth, the terms change score method and regressor variable method will be used to refer to the respective approaches.
The regressor variable, as well as the change score method, are both univariate methods, which are able to investigate the effects of one or more independent variables on one particular outcome variable. When the researcher is interested in the treatment effect on more than one dependent variable, a univariate method is applied to every dependent variable separatly. However, the use of multiple univariate analyses will implicitly result in an inflated Type-1 error rate and an increased rate of false positives (Wang et al., 2015). To account for this inflated error rate, also referred to as familywise error rate, a Bonferroni correction can be applied. A more in-depth view of the Bonferroni inequality and its issues will be discussed later.

Another known issue of the unviariate methods is Lord‘s Paradox. This paradox refers to the issue that there are specific cases in which the different approaches lead to different results (Lord, 1967). Although this paper will not cover this issue in particular, it is important to mention it, since it fosters the need for new methods. Despite the fact that the change score method as well as the regressor variable method are widely used, there are further issues and limitations, which have been frequently discussed in the literature.

1.2.1 Regressor variable method

As previously mentioned, the regressor variable method treats the pretest measurement as a covariate in the analysis. It can be expressed in a regression equation as

$$Y_{ij2} = \beta_0 + \delta T_{ij} + \beta_1 Y_{ij1} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

where Y_{ij2} represents the posttest score of person i in group j, Y_{ij1} the pretest score of person i in group j, T_i the treatment indicator of person i in group j (0 = control group, 1 = treatment group), and δ the treatment effect. The ε_{ij} is the error that is present and not controlled for, which is assumed to be normally distributed with mean 0 and variance σ^2.

1.2.2 Change score method

The change score method investigates the treatment effect as a comparison of changes from the baseline between treatment and control group. The baseline refers to the pre-measurement value of each person in each group. Thus, the relative change between the pre-measurement and the post-measurement is the subject of interest. In terms of regression equations the change score method can be given by

$$(Y_{ij2} - Y_{ij1}) = \beta_0 + \delta T_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

where Y_{ij1} represents the pretest score of person i in group j and Y_{ij2} the posttest score of person i in group j. Hence, $(Y_{ij2} - Y_{ij1})$ is the measurement of change from baseline for person i in group j. Again, T_{ij} is the group assingment of person i in group j and ε_{ij} is the error that is present but not specifically controlled for.

As with the regressor variable approach, researchers need to deal with some issues when using this approach. First, the change score method is known to be less reliable than their component variables, especially when the pretest and posttest scores are highly correlated (Kessler, 1977). Furthermore, regression towards the mean will influence the outcome of change score analyses (Allison, 1990). Regression towards the mean refers to the phenomenon that subjects with a relatively high score on the pretest will tend to score lower on the posttest. Also, subjects that scored relatively low on the pretest will therefore tend to score higher on the posttest. For the parameter in our model this means that $Y_{ij2} - Y_{ij1}$ is expected to be negatively correlated with Y_{ij1} and that any variable related to Y_{ij1} will indirectly influence the change score.

1.3 Multivariate method

A multivariate analysis is expected to provide advantages when compared to the univariate methods. In terms of regression equation the multivariate model is given by

$$\begin{gathered} Y_{ij1} = \beta_{01} + \varepsilon_{ij1} \\ Y_{ij2} = \beta_{02} + \delta T_{ij} + \varepsilon_{ij2} \\ \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{pmatrix} \sim MVN\left(0, \Sigma\right), \Sigma = \begin{pmatrix} \sigma^2, \rho \\ \rho, \sigma^2 \end{pmatrix} \end{gathered}$$

The pre- and post-measurement are dependent variables with the correlated error terms ε_{ij1} and ε_{ij2} that are assumed to be multivariate normally distributed with mean 0 and variance Σ.

The multivariate method is able to model a correlation of the error terms in the equations for the pretest and the posttest scores. Furthermore, when investigating more than one variable of interest, a multivariate approach is expected to prevent inflated Type-1 error rates, which would be the result of using univariate analyses for each dependent variable.

2 Research question

A multivariate approach is described to analyze the data of pretest-postest control group designs and compared to the univariate methods in terms of the estimated treatment effect, power to detect the treatment effect, and the Type-1 error rates. The multivariate approach is expected to perform better given the advantages, when comparing it to univariate methods. A simulation study was used to investigate specific differences between the univariate and the multivariate methods.

3 Methods

To be able to systematically compare the univariate methods to the multivariate method, data was simulated under three different predefined conditions. The statistical programming language R and the packages "CAR" and "MASS" were used to simulate and analyze the data. The three simulation studies differ on specific parameters regarding the simulation of the data, while other parameters maintained constant across all simulation studies. The regressor variable method, the change score method and the multivariate method were fitted to the data. For each model, the estimated treatment effect, its standard deviation, and the mean squared error were stored in order to compare them across models. This process was iterated 1,000 times. For each replication, a sample size of $n = 1,000$ was used.

In each of the three simulation studies, the intercept as well as the treatment effect was set to a fixed value. Furthermore, a gender effect was simulated with males scoring higher and they were also over-represented in the treatment group. Additionally, this gender effect differed from pre- to posttest to include effects of regression towards the mean.

In the first simulation study, a comparison was made without introducing additional effects. Hence, in this simulation study, attention was focused on the respective pre-measurement and post-measurement scores, the estimated treatment and the estimated gender effect. In the second simulation study, the data included a heterogeneous treatment effect to provide a more elaborate design. Therefore, the design contained three groups instead of two groups. That is one control group and two treatment groups with different treatment effects. In the last simulation study, there were two different outcome variables of interest. This was different from the situation considered in simulation study two, where for the outcome variable two different treatment effects were simulated. Note further, in the third study the heterogeneous treatment effect from study two was not present and a homogeneous treatment effect was simulated. For the third study, the p-values of each test, and for each model, were stored to examine the power and Type-1 error rate. To investigate the Type-1 error rate, the percentage of significant tests across replications

were calculated when there was no treatment effect simulated. The power was investigated by computing the percentage of significant tests when a treatment effect was simulated.

For every simulation study, the data was analyzed by the three different methods. At first, the regressor variable method was fitted to the data with the pre-measurement score treated as a covariate in the model. Then, the change score method was applied to investigate the treatment effect in terms of a change from pre- to post-test measurement. At last, the multivariate method was applied to the data, treating both the pre-test as well as the post-test score as dependent variables. Note, that in the multivariate approach within-person effects were investigated, since it concerns a change in score over time and not across persons, which is the more common situation.

4 Simulation study

The data generation was iterated 1,000 times with $n = 1,000$ for every condition. The data was generated according to the specific needs of every study. In the first study, data was simulated with a treatment effect of $\delta = .2$. Additionally, a time-specific gender effect was simulated with males scoring 1.9 on the prestest and 1.1 on the posttest, while females scored .3 on the pretest and .6 on the posttest. The covariance between the pre-test score and the post-test score was 0.5. This gender effect was also present in the simulation of study two and three.

To introduce a heterogenous treatment effect an outcome variable was simulated with two different treatment effects of $\delta_1 = .2$ and $\delta_2 = .4$. The covariance between pre-test and post-test remained unchanged to the first study.

In the last study, two outcome variables were simulated. Therefore, the treatment effect was $\delta_1 = .2$ for the first outcome variable and $\delta_2 = .2$ for the second outcome variable. The covariance between the two outcome variables as well as the covariance between pre-test score and post-test score for both outcome variables was 0.5.

After the simulation process, the three different methods were applied to analyze the data generated under each condition.

4.1 Study 1

At first, the regressor variable method with the regression equation was fitted to the data using the model

$$Y_{ij2} = \beta_0 + \delta T_{ij} + \beta_1 Y_{ij1} + \beta_2 X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

with T_{ij} being the indicator of treatment of person i in group j, Y_{ij1} the pretest score of person i in group j and X_{ij1} the indicator of the gender effect of person i in group j. Thus, the dependent variable Y_{i2} is the posttest score of person i in group j.

Then, the change score method with the regression equation

$$Y_{ij2} - Y_{ij1} = \beta_0 + \delta T_{ij} + \beta_1 X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

was fitted to the data, with the pretest score Y_{ij1} being on the left-hand side of the equation. Therefore, the variable of interest was, like previously mentioned, the change from pretest measurement to posttest measurement.

As a third model, the multivariate regression model is given by

$$Y_{ij1} = \beta_{01} + \beta_{11} X_{ij1} + \varepsilon_{ij1}$$
$$Y_{ij2} = \beta_{02} + \beta_{12} X_{ij2} + \delta T_{ij} + \varepsilon_{ij2}$$

$$\begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{pmatrix} \sim MVN\left(0, \Sigma\right), \Sigma = \begin{bmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{bmatrix}$$

was fitted to the data with correlated error terms ε_{ij1} and ε_{ij2} that are assumed to be multivariate normally distributed with mean 0 and covariance Σ.

4.2 Study 2

In the second simulation study, the focus was on a heterogeneous treatment effect that was present in the simulated data.

Again, the three models were fitted to the data, starting with the regressor variable method

$$Y_{ij2} = \beta_0 + \delta_1 T_{ij1} + \delta_2 T_{ij2} + \beta_1 X_{ij} + \beta_2 Y_{ij1} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

with T_{ij1} indicating the assignment of subject i in group j to the treatment group with a treatment effect of $\delta_1 = .2$ and T_{ij2} indicating the assignment to the treatment group with treatment effect $\delta_2 = .4$. Each person is assigned to one of the treatment groups or to the control group.

The change score method:

$$Y_{ij2} - Y_{ij1} = \beta_0 + \delta_1 T_{ij1} + \delta_2 T_{ij2} + \beta_1 X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

was fitted to the data analyzing the change score $Y_{ij2} - Y_{ij1}$ instead of treating the pretest as a covariate.

At last, the multivariate regression model

$$Y_{ij1} = \beta_{01} + \beta_{21} X_{ij} + \varepsilon_{ij1}$$
$$Y_{ij2} = \beta_{02} + \delta_{12} T_{ij1} + \delta_{22} T_{ij2} + \beta_{22} X_{ij} + \varepsilon_{ij2}$$

$$\begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{pmatrix} \sim MVN\left(0, \Sigma\right), \Sigma = \begin{bmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{bmatrix}$$

was applied with correlated error terms.

4.3 Study 3

Like in the previous simulation studies, the three methods were applied to the data. Due to the two dependent variables, for the univariate methods two analyses were needed in order to investigate both treatment effects separately.

To investigate the expected differences in power and Type-1 error rate, this simulation study consisted of two conditions. In the first condition, the parameter indicating the treatment effect were fixed to $b_1 = .2$ and $c_1 = .2$. In the second condition, the parameters were set to $b_1 = 0$ and $c_1 = 0$ to simulate data without treatment effect.

At first, the regressor variable method was fitted to the data, with one equation for each dependent variable,

$$\begin{aligned} Y_{ij3} &= \beta_{03} + \delta_1 T_{ij} + \beta_{13} Y_{ij1} + \beta_{23} X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma_1^2) \\ Y_{ij4} &= \beta_{04} + \delta_2 T_{ij} + \beta_{14} Y_{ij2} + \beta_{24} X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma_2^2) \end{aligned}$$

where Y_{ij3} represented the posttest score of of outcome 1 and Y_{ij1} the pretest score of outcome 1, whereas Y_{i4} and Y_{i2} represented the pretest and posttest score of outcome 2.

Thereafter, the change score method was fitted to the data, to analyze each dependent variable,

$$\begin{aligned} (Y_{ij3} - Y_{ij1}) &= \beta_{03} + \delta_1 T_{ij} + \beta_{13} X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma_1^2) \\ (Y_{ij4} - Y_{ij2}) &= \beta_{04} + \delta_2 T_{ij} + \beta_{14} X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma_2^2) \end{aligned}$$

Thus, $Y_{ij3} - Y_{ij1}$ represents the change score for the dependent variable 1, while $Y_{ij4} - Y_{ij2}$ represents the change score for the second dependent variable.

Finally, the multivariate method was applied to the data.

$$\begin{aligned} Y_{ij1} &= \beta_{01} + \beta_{11} X_{ij} + \varepsilon_{ij1} \\ Y_{ij2} &= \beta_{02} + \beta_{12} X_{ij} + \varepsilon_{ij2} \\ Y_{ij3} &= \beta_{03} + \delta_3 T_{ij} + \beta_{13} X_{ij} + \varepsilon_{ij3} \\ Y_{ij4} &= \beta_{04} + \delta_4 T_{ij} + \beta_{14} X_{ij} + \varepsilon_{ij4} \end{aligned}$$

$$\begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \\ \varepsilon_{ij3} \\ \varepsilon_{ij4} \end{pmatrix} \sim MVN\left(0, \Sigma\right)$$

Again, the error terms are assumed to be multivarite normally distributed with mean 0 and covariance $\Sigma = \begin{bmatrix} \sigma^2 & \rho & \rho & \rho \\ \rho & \sigma^2 & \rho & \rho \\ \rho & \rho & \sigma^2 & \rho \\ \rho & \rho & \rho & \sigma^2 \end{bmatrix}$.

5 Results

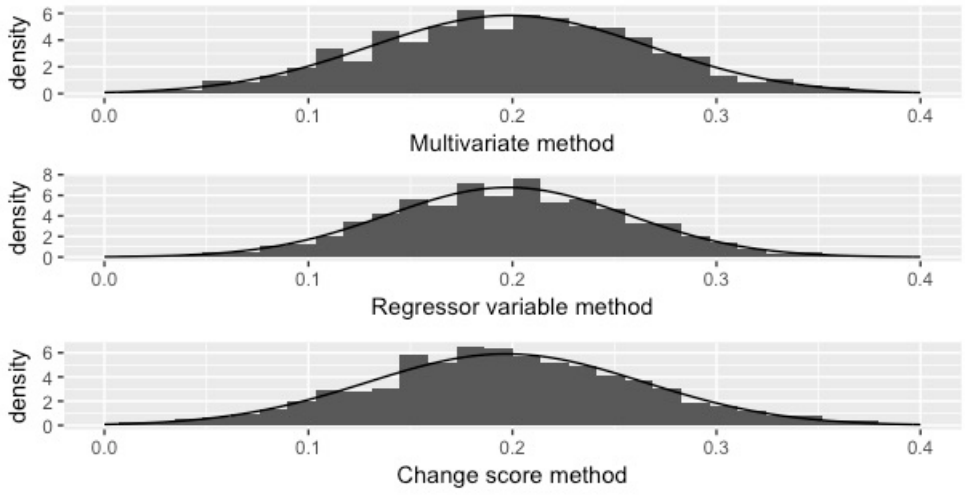

Figure 1: Estimation of treatment effect - Simulation study 1

In the first simulation study with a treatment effect of $\delta = .2$, the multivariate approach provided an average estimation of the treatment effect across all 1000 replications of $\overline{\delta} = .199$ ($SD = .068$). The univariate approaches provided comparable results. With the regressor variable method a treatment effect of $\overline{\delta} = .198$ ($SD = .059$) was estimated. The change score provided an average coefficient of $\overline{\delta} = .199$ ($SD = .067$). Represented in Figure 1, there are no significant differences in the estimation of the treatment effects between the three methods.

In the second simulation study, the focus was on the heterogeneous treatment effect with $\delta_1 = .2$ and $\delta_2 = .4$. In the multivariate method, the treatment effects were accurately estimated with an average treatment effect of $\overline{\delta}_1 = .199$ ($SD = .082$) and $\overline{\delta}_2 = .399$ ($SD = .080$). Again, the univariate approaches provided similarly precise estimates of the treatment effects. The regressor variable method estimated the treatment effects on average with $\overline{\delta}_1 = .198$ ($SD = .071$) and $\overline{\delta}_2 = .398$ ($SD = .071$), while the change score approach provided estimates of $\overline{\delta}_1 = .196$ ($SD = .081$) and $\overline{\delta}_2 = .398$ ($SD = .081$).

The third simulation study focused on the estimation of the treatment effect on two outcomes. The treatment effect was the same across both variables $\delta_1 = \delta_2 = .2$ in the first condition and $\delta_1 = \delta_2 = 0$ in the second condition.

In the first condition, for the multivariate method average treatment effects of $\overline{\delta}_1 = .202$ ($SD = .067$) and $\overline{\delta}_2 = .2$ ($SD = .066$) were estimated, while the regressor variable method provided treatment effects of $\overline{\delta}_1 = .202$ ($SD = .059$) and $\overline{\delta}_2 = .201$ ($SD = .056$). Furthermore, for the change score method a treatment effect of $\overline{\delta}_1 = .201$ ($SD = .069$) and $\overline{\delta}_2 = .201$ $SD = .066$) was estimated. In the second condition, for each method it was concluded that there was no treatment effect present in the generated data.

When looking at the power and Type-1 error rate, the univariate methods were applied

to each outcome with an applied Bonferroni correction. Therefore, a significance level of $\alpha = 0.05$ was used for the multivariate method and a significance level of $\alpha = 0.025$ was used for the univariate methods. The regressor variable approach yielded a power of $1 - \beta_1 = .88$ for the first treatment effect and $1 - \beta_2 = .88$ for the second treatment effect, while the change score method provided a power of $1 - \beta_1 = .78$ and $1 - \beta_2 = .76$ for the treatment effects. The multivariate approach provided a power of $1 - \beta = .99$ in the analysis of both treatment effects.

Regarding the Type-1 error rate, the univariate approaches provided error rates of $\alpha_1 = .024$ and $\alpha_2 = .027$ for the regressor variable method, and $\alpha_1 = .026$ and $\alpha_2 = .025$ for the change score method. The multivariate approach led to an Type-1 error rate of $\alpha = .046$. This result showed accurate Type-1 error rates for all methods. However, with a lower sample size of $n = 400$, differences were obtained. In this condition, the regressor variable method yielded Bonferroni corrected Type-1 error rates of $\alpha_1 = .026$ and $\alpha_2 = .036$, while the change score method provided Type-1 error rates of $\alpha_1 = .021$ and $\alpha_2 = .029$. However, the multivariate approach provided an error rate of $\alpha = .049$, which shows that the multivariate approach provided a better Type-1 error rate, especially for smaller sample sizes.

6 Discussion

The results of the simulation showed that there are no significant differences between the multivariate method and the univariate methods for the conditions of the first simulation study. All methods were able to give a correct estimation of the treatment effect. When it comes to the investigation of heterogeneous treatment effects, the multivariate method provided no additional value compared to the univariate methods. For both treatment effects the univariate as well as the multivariate approaches gave accurate estimates. When the research design asks for the investigation of multiple outcome variables, the multivariate method provided the initially expected advantages in comparison with the univariate methods. The power to detect significant differences between control group and treatment group was higher in the multivariate approach. Furthermore, also the Type-1 error rate was better in the multivariate analysis. Especially, when the sample size decreased, the advantages of the multivariate method significantly increased. As mentioned before, the univariate methods depend on procedures such as Bonferroni corrections in order to correct for an inflated Type-1 error rate. Bonferroni corrections are known to be quite conservative and tend to underestimate the signifcance of group differences. Although the Bonferroni correction is the most widely used and known method to correct for familywise error rates, there are several developments aimed to provide more appropriate results, for instance Holm's sequential rejective multiple test procedure (Holm, 1979) or the improved

Bonferroni procedure (Simes, 1986). Nevertheless, also these procedures were criticized. For example, the improved Bonferroni procedure by Simes creates a new overall hypothesis that consists of several individual hypotheses, so that it is not clear how inferences about individual hypotheses can be done (Hommel, 1988). Furthermore, as Cohen (1994) concluded, any correction for multiple testing will result in a lower power to detect significant differences between treatment and control group. Therefore, it is intuitively the superior option to avert multiple testing and to avoid the occurrence of these issues.

Furthermore, it is important to note that in this study, only simulation studies were conducted in which the variable of interest differed in one dimension only. All changes that were introduced to the variable only affected the mean value of the variable. For future research, it would be interesting to investigate differences between univariate and multivariate methods under conditions with more elaborate designs, such as changes to the variance of the dependent variable or inclusion of unknown random error that is not specifically modeled in the regression model. Another case of interest would be the investigation of multilevel data structures. The advantages of this approach are the possibility to account for nested data and to model error terms on multiple levels of the data.

References

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20*, 93-114.

Brognan, D. R., & Kutner, M. H. (1980, nov). Comparative analyses of pretest-posttest research design. *The American Statistican*, *34*(4).

Campbell, D., & Stanley, J. (1975). Experimental and quasi-experimental designs for research.

Cohen, J. (1994). The earth is round (p <.05). *American Psychologist*, *49*(12), 997-1003.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Board of the Foundation of the Scandinavian Journal of Statistics*, *6*(2), 65-70.

Hommel, G. (1988, jun). A stagewise rejective multiple test procedure based on a modified bonferroni test. *Biometrika*, *75*(2), 383-386.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*(5), 304-305.

Simes, R. J. (1986, dec). An improved bonferroni procedure for multiple tests of significance. *Biometrika*, *73*(3), 751-754.

Wang, D., Li, Y., Wang, X., Liu, X., Fu, B., Lin, Y., ... Offen, W. (2015). Overview of multiple testing methodology and recent development in clinical trials. *Contemporary Clinical Trials*, *45*, 13-20.

A Appendix

A.1 R code of the first simulation study

```
## construct design matrix
tijd <- ordered(rep(1:2,1))
idata2 <- data.frame(tijd)
idata2 # defines the within-subject factors

# Define sample size n
n <- 1000

# Set number of iterations
reps <- 1000

# Fix randomization to assure reproducibility of results
set.seed(10)

# Set intercept parameters with b1 representing the intercept
# for the treatment effect
b0 <- 0
b1 <- .2

# Generate matrix for treatment indicator X
# (0 = control group, 1 = treatment)
X <- matrix(0,ncol=1,nrow=n)
X[1:(n/2)] <- 0
X[((n/2)+1):n] <- 1

# Random binominal distribution with higher propability
# for males to appear in treatment group
Tr0 = n/2
Tr1 = n/2
G <- c(rbinom(Tr0,1,0.3),rbinom(Tr1,1,0.7))

# Set gender effect for the pretest measurement
Z1 <- G*1.9
Z1 <- ifelse(Z1==0,0.3,Z1)

# Set gender effect for the posttest measurement
Z2 <- G*1.1
Z2 <- ifelse(Z2==0,0.6,Z2)

# Generate mean structure to be used in data simulation
mean <- matrix(0,ncol=2,nrow=n)
mean[,1] <- b0 + Z1
mean[,2] <- b0 + b1*X + Z2

# Generate covariance matrix to be used in data simulation
Sigma <- matrix(diag(2)+.5,ncol=2,nrow=2) - diag(2)*.5

# Prepare data file to store generated data of each replication
data <- matrix(0,ncol=2,nrow=n)
data <- data.frame(data,X)

# Prepare outcome storage
outcome = matrix(0,ncol=9,nrow=reps)
datalist <- list(reps)
```

```
# Loop to iterate data generation, data analaysis
# and storage of results

for(i in 1:reps){

  # Simulate data
  for(ii in 1:n){
    data[ii,1:2] <- mvrnorm(1,mu=mean[ii,],Sigma=Sigma)
  }

  data <- data.frame(data)

  # Estimate models
  # Regressor variable method
  mod.rv <- lm(X2~X+X1+G, data=data)

  # Change score method
  mod.cs <- lm(X2-X1~1+X+G, data=data)

  # Multivariate method
  mod.f <- lm(cbind(X1,X2)~1+X+G, data=data)

  # Store data
  datalist[[i]] <- data
  outcome[i,1] <- mod.f$coefficients[5]
  outcome[i,2] <- summary(mod.f)$"Response X2"$coefficients[2,2]
  outcome[i,3] <- mean(mod.f$residuals^2)
  outcome[i,4] <- mod.rv$coefficients[2]
  outcome[i,5] <- summary(mod.rv)$coefficients[3,2]
  outcome[i,6] <- mean(mod.rv$residuals^2)
  outcome[i,7] <- mod.cs$coefficients[2]
  outcome[i,8] <- summary(mod.cs)$coefficients[2,2]
  outcome[i,9] <- mean(mod.cs$residuals^2)

}

# Set column names of the outcome matrix
colnames(outcome) <- c("MV.COEF","MV.SD","MV.MSE",
                       "RV.COEF","RV.SD","RV.MSE",
                       "CS.COEF", "CS.SD", "CS.MSE")

# Compute means, variance and standard deviation
parametermean <- apply(outcome, 2, mean)
parametervar <- apply(outcome, 2, var)
paramtersd <- sqrt(parametervar)

# Print outcome
outcome
parametermean
parametervar
paramtersd
```

A.2 R code of second simulation study

```
# Construct design matrix
tijd <- ordered(rep(1:2,1))
idata2 <- data.frame(tijd)
idata2 # defines the within-subject factors

# Define sample size n
n <- 1000

# Set number of iterations
reps <- 1000

# Fix randomization to assure reproducibility of results
set.seed(10)

# Set intercept parameters with b1 representing the treatment
# effect 1 and b2 the treatment effect 2
b0 <- 0
b1 <- .2
b2 <- .4

# Generate matrix for treatment indicator X
# (0 = control group, 1 = treatment 1, 2 = treatment 2)
X <- matrix(0,ncol=1,nrow=n)
X[1:(n/2)] <- 0
X[((n/2)+1):((n/4)*3)] <- 1
X[(((n/4)*3)+1):n] <- 2

# Dummy variable for treatment assignment
# (X00 = treatment 1, X01 = treatment 2)
X00 <- as.numeric(X == 1)
X01 <- as.numeric(X == 2)

# Random binominal distribution with higher propability for males
# in treatment group
G <- c(rbinom(Tr0,1,0.3),rbinom(Tr1,1,0.7))
Tr0 = n/2
Tr1 = n/2

# Set gender effect for the pretest measurement
Z1 <- G*1.9
Z1 <- ifelse(Z1==0,0.3,Z1)

# Set gender effect for the posttest measurement
Z2 <- G*1.1
Z2 <- ifelse(Z2==0,0.6,Z2)

# Generate mean structure to be used in data simulation
mean <- matrix(0,ncol=2,nrow=n)
mean[,1] <- b0 + Z1
mean[,2] <- b0 + b1*X00 + b2*X01 +Z2

# Generate covariance matrix to be used in data simulation
Sigma <- matrix(diag(2)+.5,ncol=2,nrow=2) - diag(2)*.5

# Prepare data file to store generated data of each replication
data <- matrix(0,ncol=2,nrow=n)
data <- data.frame(data,X)
```

```
# Prepare outcome storage
outcome = matrix(0,ncol=15,nrow=reps)
datalist <- list(reps)

# Loop to iterate data generation, data analaysis
# and storage of results

for(i in 1:reps){

  # Simulate data
  for(ii in 1:n){
    data[ii,1:2] <- mvrnorm(1,mu=mean[ii,],Sigma=Sigma)
  }

  data <- data.frame(data)

  # Estimate models
  # Regressor variable method
  mod.rv <- lm(X2~X00+X01+G+X1, data=data)

  # Change score method
  mod.cs <- lm(X2-X1~1+X00+X01+G, data=data)

  # Multivariate method
  mod.f <- lm(cbind(X1,X2)~1+X00+X01+G, data=data)

  # Store data
  datalist[[i]] <- data

  # Results of multivariate method
  outcome[i,1] <- mod.f$coefficients[6] #X00
  outcome[i,2] <- summary(mod.f)$"Response X2"$coefficients[2,2]
  outcome[i,3] <- mod.f$coefficients[7] #X01
  outcome[i,4] <- summary(mod.f)$"Response X2"$coefficients[3,2]
  outcome[i,5] <- mean(mod.f$residuals[,2]^2)

  # Results of regressor variable method
  outcome[i,6] <- mod.rv$coefficients[2] #X00
  outcome[i,7] <- summary(mod.rv)$coefficients[2,2]
  outcome[i,8] <- mod.rv$coefficients[3] #X01
  outcome[i,9] <- summary(mod.rv)$coefficients[3,2]
  outcome[i,10] <- mean(mod.rv$residuals^2)

  # Results ofchange score method
  outcome[i,11] <- mod.cs$coefficients[2] #X00
  outcome[i,12] <- summary(mod.cs)$coefficients[2,2]
  outcome[i,13] <- mod.cs$coefficients[3] #X01
  outcome[i,14] <- summary(mod.cs)$coefficients[3,2]
  outcome[i,15] <- mean(mod.cs$residuals^2)

}

# Set column names of the outcome matrix
colnames(outcome) <- c("MV.COEFX00","MV.SDX00","MV.COEFX01","MV.SDX01",
                       "MV.MSE","RV.COEFX00","RV.SDX00","RV.COEFX01",
                       "RV.SDX01", "RV.MSE","CS.COEFX00", "CS.SDX00",
                        "CS.COEFX01", "CS.SDX01", "CS.MSE")
```

```
# Compute means, variance and standard deviations
parametermean <- apply(outcome, 2, mean)
parametervar <- apply(outcome, 2, var)
paramtersd <- sqrt(parametervar)


# Print outcome
outcome
parametermean
parametervar
paramtersd
```

A.3 R code of the third simulation study

```
# Import functions for test statistics for multivariate analysis
Pillai <- car:::Pillai
Wilks <- car:::Wilks
HL <- car:::HL
Roy <- car:::Roy
car:::summary.Anova.mlm()

# Create function to be able to retrieve
# p-values for multivariate method

################################################

Anovamlm <- function (object, test.statistic,
 univariate = TRUE, multivariate = TRUE,...)
{
  GG <- function(SSPE, P) {
    p <- nrow(SSPE)
    if (p < 2)
      return(NA)
    lambda <- eigen(SSPE %*% solve(t(P) %*% P))$values
    lambda <- lambda[lambda > 0]
    ((sum(lambda)/p)^2)/(sum(lambda^2)/p)
  }
  HF <- function(gg, error.df, p) {
    ((error.df + 1) * p * gg - 2)/(p * (error.df - p * gg))
  }
  mauchly <- function(SSD, P, df) {
    if (nrow(SSD) < 2)
      return(c(NA, NA))
    Tr <- function(X) sum(diag(X))
    p <- nrow(P)
    I <- diag(p)
    Psi <- t(P) %*% I %*% P
    B <- SSD
    pp <- nrow(SSD)
    U <- solve(Psi, B)
    n <- df
    logW <- log(det(U)) - pp * log(Tr(U/pp))
    rho <- 1 - (2 * pp^2 + pp + 2)/(6 * pp * n)
    w2 <- (pp + 2) * (pp - 1) * (pp - 2) * (2 * pp^3 + 6 *
          pp^2 + 3 * p + 2)/(288 * (n * pp * rho)^2)
    z <- -n * rho * logW
    f <- pp * (pp + 1)/2 - 1
    Pr1 <- pchisq(z, f, lower.tail = FALSE)
    Pr2 <- pchisq(z, f + 4, lower.tail = FALSE)
    pval <- Pr1 + w2 * (Pr2 - Pr1)
    c(statistic = c(W = exp(logW)), p.value = pval)
  }
  if (missing(test.statistic))
    test.statistic <- c("Pillai", "Wilks", "Hotelling-Lawley",
                        "Roy")
  test.statistic <- match.arg(test.statistic,
  c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"), several.ok = TRUE)
  nterms <- length(object$terms)
  summary.object <- list(type = object$type, repeated = object$repeated,
       multivariate.tests = NULL, univariate.tests = NULL,
        pval.adjustments = NULL, sphericity.tests = NULL)
```

```
  if (multivariate) {
    summary.object$multivariate.tests <- vector(nterms, mode = "list")
    names(summary.object$multivariate.tests) <- object$terms
    summary.object$SSPE <- object$SSPE
    for (term in 1:nterms) {
      hyp <- list(SSPH = object$SSP[[term]], SSPE = if (object$repeated)
       object$SSPE[[term]] else object$SSPE,
       P = if (object$repeated) object$P[[term]] else NULL,
       test = test.statistic, df = object$df[term],
       df.residual = object$error.df, title = object$terms[term])
      class(hyp) <- "linearHypothesis.mlm"
      summary.object$multivariate.tests[[term]] <- hyp
    }
  }
  if (object$repeated && univariate) {
    singular <- object$singular
    error.df <- object$error.df
    table <- matrix(0, nterms, 6)
    table2 <- matrix(0, nterms, 4)
    table3 <- matrix(0, nterms, 2)
    rownames(table3) <- rownames(table2) <- rownames(table) <-
                          object$terms
    colnames(table) <- c("SS", "num Df", "Error SS", "den Df",
                         "F", "Pr(>F)")
    colnames(table2) <- c("GG eps", "Pr(>F[GG])", "HF eps",
                          "Pr(>F[HF])")
    colnames(table3) <- c("Test statistic", "p-value")
    if (singular)
      warning("Singular error SSP matrix:\nnon-sphericity test
      and corrections not available")
    for (term in 1:nterms) {
      SSP <- object$SSP[[term]]
      SSPE <- object$SSPE[[term]]
      P <- object$P[[term]]
      p <- ncol(P)
      PtPinv <- solve(t(P) %*% P)
      gg <- if (!singular)
        GG(SSPE, P)
      else NA
      table[term, "SS"] <- sum(diag(SSP %*% PtPinv))
      table[term, "Error SS"] <- sum(diag(SSPE %*% PtPinv))
      table[term, "num Df"] <- object$df[term] * p
      table[term, "den Df"] <- error.df * p
      table[term, "F"] <- (table[term, "SS"]/table[term,
          "num Df"])/(table[term, "Error SS"]/table[term, "den Df"])
      table[term, "Pr(>F)"] <- pf(table[term, "F"], table[term,
         "num Df"], table[term, "den Df"], lower.tail = FALSE)
      table2[term, "GG eps"] <- gg
      table2[term, "HF eps"] <- if (!singular)
        HF(gg, error.df, p)
      else NA
      table3[term, ] <- if (!singular)
        mauchly(SSPE, P, object$error.df)
      else NA
    }
    table3 <- na.omit(table3)
    if (nrow(table3) > 0) {
      table2[, "Pr(>F[GG])"] <- pf(table[, "F"], table2[,
        "GG eps"] * table[, "num Df"], table2[, "GG eps"] *
```

```
        table[, "den Df"], lower.tail = FALSE)
      table2[, "Pr(>F[HF])"] <- pf(table[, "F"], pmin(1,
      table2[, "HF eps"]) * table[, "num Df"], pmin(1,
      table2[, "HF eps"]) * table[, "den Df"], lower.tail = FALSE)
      table2 <- na.omit(table2)
      if (any(table2[, "HF eps"] > 1))
        warning("HF eps > 1 treated as 1")
    }
    class(table3) <- class(table) <- "anova"
    summary.object$univariate.tests <- table
    summary.object$pval.adjustments <- table2
    summary.object$sphericity.tests <- table3
  }
  class(summary.object) <- "summary.Anova.mlm"
  summary.object
  return(table)
}

################################################

## Construct design matrix
type <- factor(rep(c("var1", "var2"), 2), levels=(c("var1", "var2")))
tijd <- ordered(c(1,1,2,2))
idata2 <- data.frame(type, tijd)
idata2 # defines the within-subject factors

# define sample size n
n <- 1000

# Set number of iterations
reps <- 1000

# Fix randomization
set.seed(16)

# Set intercept parameters with b1 representing the treatment effect
# of outcome variable 1  and c1 the treatment effect for
# outcome variable 2
# To invetigate the Type-1 error rate b1 and c1
# were set to .0 in the second condition
b0 <- .0
b1 <- .2
c0 <- .0
c1 <- .2

# Generate matrix for treatment indicator X
# (0 = control group, 1 = treatment)
X <- matrix(0,ncol=1,nrow=n)
X[1:(n/2)] <- 0
X[((n/2)+1):n] <- 1


# Random binominal distribution with higher propability
# for males in treatment group
Tr0 = n/2
Tr1 = n/2
G <- c(rbinom(Tr0,1,0.3),rbinom(Tr1,1,0.7))

# Set gender effect for the pretest measurement
```

```
Z1 <- G*1.9
Z1 <- ifelse(Z1==0,0.3,Z1)

# Set gender effect for the posttest measurement
Z2 <- G*1.1
Z2 <- ifelse(Z2==0,0.6,Z2)


# Generate mean structure to be used in data simulation.
# The first to columns represent the pretest scores,
# and the last two columns represent the posttest scores
mean <- matrix(0,ncol=4,nrow=n)
mean[,1] <- b0 + Z1
mean[,2] <- c0 + Z1
mean[,3] <- b0 + b1*X + Z2
mean[,4] <- c0 + c1*X + Z2

# Generate covariance matrix to be used in data simulation
Sigma <- matrix(diag(4)+.5,ncol=4,nrow=4) - diag(4)*.5

# Prepare data file to store generated data of each replication
data <- matrix(0,ncol=4,nrow=n)
data <- data.frame(data,X)

# Prepare outcome storage
outcome.mv = matrix(0,ncol=6,nrow=reps)
outcome.rv = matrix(0,ncol=6,nrow=reps)
outcome.cs = matrix(0,ncol=6,nrow=reps)
outcome.pval = matrix(0,ncol=5, nrow=reps)
datalist <- list(reps)

# Loop to iterate data generation, data analaysis
# and storage of results

for(i in 1:reps){
  #simulate data

  for(ii in 1:n){
    data[ii,1:4] <- mvrnorm(1,mu=mean[ii,],Sigma=Sigma)
  }

  data <- data.frame(data)

  # Estimate models
  # Regressor variable method
  mod.rv1 <- lm(X3~X+X1+G, data=data)
  mod.rv2 <- lm(X4~X+X2+G, data=data)

  # Change score method
  mod.cs1 <- lm(X3-X1~1+X+G, data=data)
  mod.cs2 <- lm(X4-X2~1+X+G, data=data)

  # Multivariate analysis
  mod.f <- lm(cbind(X1,X2,X3,X4)~1+X+G, data=data)
  mod.f.lht <- Anova(mod.f, type=c("III"),
                     idata=idata2,idesign=~type*tijd)

  # Store data
  datalist[[i]] <- data
```

```
  # Store results of themultivariate method
  # Outcome1
  # Mean
  outcome.mv[i,1] <- summary(mod.f)$"Response X3"$coefficients[2,1]
  # SD
  outcome.mv[i,2] <- summary(mod.f)$"Response X3"$coefficients[2,2]
  # MSE
  outcome.mv[i,3] <- mean(mod.f$residuals[,3]^2)

  # Outcome2
  # Mean
  outcome.mv[i,4] <- summary(mod.f)$"Response X4"$coefficients[2,1]
  #SD
  outcome.mv[i,5] <- summary(mod.f)$"Response X4"$coefficients[2,2]
  #MSE
  outcome.mv[i,6] <- mean(mod.f$residuals[,4]^2)

  # Store results of the regressor variable method
  # Outcome1
  outcome.rv[i,1] <- summary(mod.rv1)$coefficients[2,1]   #Mean
  outcome.rv[i,2] <- summary(mod.rv1)$coefficients[2,2]   #SD
  outcome.rv[i,3] <- mean(mod.rv1$residuals^2)            #MSE

  # Outcome2
  outcome.rv[i,4] <- summary(mod.rv2)$coefficients[2,1]   #Mean
  outcome.rv[i,5] <- summary(mod.rv2)$coefficients[2,2]   #SD
  outcome.rv[i,6] <- mean(mod.rv2$residuals^2)            #MSE

  # Store results of the change score method
  # Outcome1
  outcome.cs[i,1] <- summary(mod.cs1)$coefficients[2,1]   #Mean
  outcome.cs[i,2] <- summary(mod.cs1)$coefficients[2,2]   #SD
  outcome.cs[i,3] <- mean(mod.cs1$residuals^2)            #MSE

  # Outcome2
  outcome.cs[i,4] <- summary(mod.cs2)$coefficients[2,1]   #Mean
  outcome.cs[i,5] <- summary(mod.cs2)$coefficients[2,2]   #SD
  outcome.cs[i,6] <- mean(mod.cs2$residuals^2)            #MSE

  # Store p-values of every analysis

  # Regressor variable method
  # Outcome 1
  outcome.pval[i,1] <- summary(mod.rv1)$coefficients[2,4]
  # Outcome 2
  outcome.pval[i,2] <- summary(mod.rv2)$coefficients[2,4]
  # Change score method
  # Outcome 1
  outcome.pval[i,3] <- summary(mod.cs1)$coefficients[2,4]
     Outcome 2
  outcome.pval[i,4] <- summary(mod.cs2)$coefficients[2,4]
  #Multivariate Method
  outcome.pval[i,5] <- Anovamlm(mod.f.lht)["X:tijd","Pr(>F)"]

}

# Set column names of the outcome matrix
colnames(outcome.mv) <- c("V1.COEF","V1.SD", "V1.MSE",
```

```
                              "V2.COEF","V2.SD", "V2.MSE")
colnames(outcome.rv) <- c("V1.COEF","V1.SD","V1.MSE",
                              "V2.COEF","V2.SD","V2.MSE")
colnames(outcome.cs) <- c("V1.COEF","V1.SD","V1.MSE",
                              "V2.COEF","V2.SD","V2.MSE")

# Compute means, variance and standard deviations

parametermean.mv <- apply(outcome.mv, 2, mean)
parametervar.mv <- apply(outcome.mv, 2, var)
parametermean.rv <- apply(outcome.rv, 2, mean)
parametervar.rv <- apply(outcome.rv, 2, var)
parametermean.cs <- apply(outcome.cs, 2, mean)
parametervar.cs <- apply(outcome.cs, 2, var)
parametersd.mv <- sqrt(parametervar.mv)
parametersd.rv <- sqrt(parametervar.rv)
parametersd.cs <- sqrt(parametervar.cs)


#compute power or alpha-error
#regressor variable method
(table(outcome.pval[,1]<(0.05/2)))/reps    #treatment effect 1
(table(outcome.pval[,2]<(0.05/2)))/reps    #treatment effect 2
#change score method
(table(outcome.pval[,3]<(0.05/2)))/reps    #treatment effect 1
(table(outcome.pval[,4]<(0.05/2)))/reps    #treatment effect 2
#Multivairate method
(table(outcome.pval[,5]<(0.05)))/reps

#Print outcome
parametermean.mv
parametersd.mv
parametermean.rv
parametersd.rv
parametermean.cs
parametersd.cs
```