

Statistical Approaches for Analyzing a Continuous Outcome in Experimental Studies

Métodos Estadísticos para Analizar un Resultado Continuo en Estudios Experimentales

Antonio Sanhueza^{*,**}; Tamara Otzen^{**,**}; Carlos Manterola^{****,*****} & Nelson Araneda^{*****}

SANHUEZA, A.; OTZEN, T.; MANTEROLA, C. & ARANEDA, N. Statistical approaches for analyzing a continuous Outcome in experimental studies. *Int. J. Morphol.*, 32(1):339-350, 2014.

SUMMARY: In experimental studies the outcome variable is measured at initial time, usually called "baseline", and then in several times called "follow-up" measurement(s). The study question of interest in an experimental study is whether there is a significant difference effect between treatment and comparison group, after intervention. In addition, one wants to estimate the difference effect between groups. This paper studies some of the strategies, including a simulation process, that one can be used for analyzing data coming from an experimental study as above, and considers using or not using the baseline measurements. Three parametric and two non-parametric strategies are evaluated considering only one follow-up measurement. The baseline measurement is incorporated in context in these strategies.

KEY WORDS: Experimental Studies; Follow-Up Studies; Biostatistics"Statistics; Nonparametric.

INTRODUCTION

In experimental studies (ES), the term "baseline" is used for the measurements of a participant before the start of intervention. These measurements are the basis for characterizing and describing the population in the study. In addition, the investigators compare the distributions of baseline characteristics in the treatment group with the comparison group. In ES, if randomization worked, one expects that there will be no meaningful differences in these characteristics between the groups. However, if there are big differences, randomization can be called into question (Friedman *et al.*, 1985; Piantadosi *et al.*, 1997).

This paper only considers those ES which randomly assign patients to one of two groups, treatment group or comparison group, and where the outcome of interest is a continuous variable. Under this design, the outcome variable is measured at two times; one time before the intervention (baseline data) and the other time after the intervention (follow-up data).

In these ES the study question of interest is whether there is a significant difference effect between treatment and comparison groups, after intervention. Also, one wants to estimate the difference effect between groups.

In both type of study design, there are different ways to analyze the question of interest. For instance, one can use Nonparametric Analysis or Parametric Analysis. In addition, one of the considerations to deal with is whether to use or not use the baseline data in the analysis.

This manuscript refers to some of the strategies that one can use for analyzing data coming from an ES, and considers using or not using the baseline measurements. Section 2 and 3 show the parametric and nonparametric methods used in the analysis of the data from an ES. In Section 4, we describe some guideline to use these procedures.

* Department of Mathematics and Statistics, Universidad de La Frontera, Temuco, Chile.

** Department of Psychology, Universidad Autónoma de Chile, Temuco, Chile.

*** Ph.D. Program in Medical Sciences, Universidad de La Frontera, Temuco, Chile.

**** Department of Surgery and Traumatology, Universidad de La Frontera, Temuco, Chile.

***** Centro de Investigación en Ciencias Biomédicas, Universidad Autónoma de Chile, Temuco, Chile.

***** Department of Education, Universidad de La Frontera, Temuco, Chile.

MATERIAL AND METHOD

Parametric methods. In an experimental study, patients are randomly assigned to the comparison group or treatment group, before the intervention is applied. It is assumed that there are continuous measures on each participant at two times, before intervention and after intervention, for an outcome variable of interest. Suppose that measurements of the outcome variable on patients at the baseline (before intervention) for control group and treatment group are $Y_{ctrl, b, i}$, $i = 1, \dots, n_{ctrl}$ and $Y_{treat, b, i}$, $i = 1, \dots, n_{treat}$, respectively. Similarly, let the corresponding measurements at the follow-up (after intervention) for control group and treatment group be $Y_{ctrl, f, i}$, $i = 1, \dots, n_{ctrl}$ and $Y_{treat, f, i}$, $i = 1, \dots, n_{treat}$, respectively. Suppose that the variance-covariance matrix in the two groups is identical and equal to:

$$\Sigma_{b,f} = \begin{pmatrix} \sigma^2 & \sigma^2 \rho \\ \sigma^2 \rho & \sigma^2 \end{pmatrix}$$

where $\sigma^2 = \text{Var}(Y_{ctrl, b, i}) = \text{Var}(Y_{treat, b, i})$ and $\rho = \text{Corr}(Y_{ctrl, b, i}; Y_{treat, b, i})$. This assumption means that the variance of the continuous outcome at baseline and at follow-up is the same, σ^2 , and the correlation between both outcomes is ρ .

The interest of this experiment is the assessment of the difference effect between treatment and control group. It is possible to analyze the difference effect between these groups by using the following strategies:

Simple model. This strategy uses measures on each participant at follow-up only (no baseline data). It can be represented by $Y_f = \alpha_0 + \alpha_1 T + e$, where Y_f is the outcome at follow-up, α_0 is the intercept term, α_1 is the effect of difference between treatment group and control group, T is an indicator variable of group, and e is the error term that follows a normal distribution with mean 0 and variance σ^2 . In this model, the unbiased estimate of α_1 is given by:

$$\hat{\alpha}_1 = \bar{Y}_{treat} - \bar{Y}_{ctrl}$$

where \bar{Y}_{treat} and \bar{Y}_{ctrl} are the average values of the outcome at follow-up in the treatment group and control group, respectively. The meaning of this estimator is just the difference of the two means without controlling for the baseline measurements. It is assumed that the randomization produces balance of the outcome between the groups. The variance of the estimate of α_1 is:

$$\text{Var}(\hat{\alpha}_1) = \sigma^2 \left(\frac{1}{n_{treat}} + \frac{1}{n_{ctrl}} \right)$$

and if the sample sizes of the treatment and control groups are equal ($n_{treat} = n_{ctrl} = n$) we obtained the known expression:

$$\text{Var}(\hat{\alpha}_1) = \frac{2}{n} \sigma^2$$

Most standard analysis of ES state that randomization takes care of baseline differences and thus this Simple model approach is appropriate.

Difference Score model. This method employs differences of the measures of follow-up and baseline on each participant. The following model is used: $Y_f - Y_b = \beta_0 + \beta_1 T + e$, where $(Y_f - Y_b)$ is the difference between the outcome variable at follow-up and baseline, β_0 is the intercept term, β_1 is the effect of difference of treatment and control group, and T and e are as before. The estimated difference of treatment effect is given by:

$$\hat{\beta}_1 = \bar{Z}_{treat} - \bar{Z}_{ctrl}$$

where, $Z_{treat} = Y_{treat, f} - Y_{treat, b}$, $Z_{ctrl} = Y_{ctrl, f} - Y_{ctrl, b}$ and $Y_{ctrl, f}$ and $Y_{treat, f}$ are the sample means of the outcome variables at follow-up for control group and treatment group, and $Y_{ctrl, b}$ and $Y_{treat, b}$ are the sample means at baseline for control group and treatment group, respectively. This estimator considers the potential imbalance in the outcomes due to the randomization process. Note that, β_1 compare the differences between the means Z_{treat} and Z_{ctrl} adjusted by the averages at baseline.

The variance of the estimate of β_1 is given by:

$$\text{Var}(\hat{\beta}_1) = 2\sigma^2(1 - \rho) \left(\frac{1}{n_{treat}} + \frac{1}{n_{ctrl}} \right)$$

and if the sample size are equals, then:

$$\text{Var}(\hat{\beta}_1) = \frac{4}{n} \sigma^2(1 - \rho)$$

where ρ is the within-subject correlation between baseline and follow-up measurements. In general, this correlation is most likely to be positive, reducing the variance of β_1 .

This approach is often more efficient than using only

a single follow-up measurement because the standard error of the effect of difference of treatment and control is usually reduced as the result of using two measurements from each participant.

Analysis of Covariance (ANCOVA) model. This method employs the baseline measurement as a covariate. One fits the following regression model: $Y_f = \delta_0 + \delta_1 T + \delta_2 Y_b + e$, where δ_0 is the intercept term, δ_1 is the effect of the difference between the treatments, δ_2 is the effect of the baseline measurement, and T and e are as before. This model depends on several assumptions, including normality of error terms, equality of error variances for different treatments, equality of slopes for the different treatment regression lines, and linearity of regression. The unbiased estimate of δ_1 is given by:

$$\hat{\delta}_1 = \bar{Y}_{treat,f} - \bar{Y}_{ctrl,f} - \hat{\delta}_2 (\bar{Y}_{treat,b} - \bar{Y}_{ctrl,b}) \quad \text{with}$$

$$\hat{\delta}_2 = \frac{(n_{ctrl} - 1)S_{Y_{b,ctrl}}^2 \hat{\delta}_{1,ctrl} + (n_{treat} - 1)S_{Y_{b,treat}}^2 \hat{\delta}_{1,treat}}{(n_{ctrl} - 1)S_{Y_{b,ctrl}}^2 + (n_{treat} - 1)S_{Y_{b,treat}}^2}$$

where $\hat{\delta}_{1,ctrl}$ and $\hat{\delta}_{1,treat}$ are the estimate slopes on separate line fits for control group and treatment group, $S_{Y_{b,ctrl}}^2$ and $S_{Y_{b,treat}}^2$ are the sample variances of the outcome variable at baseline for control group and treatment group, $\bar{Y}_{ctrl,f}$ and $\bar{Y}_{treat,f}$ are the sample means of the outcome variables at follow-up for control group and treatment group, and $\bar{Y}_{ctrl,b}$ and $\bar{Y}_{treat,b}$ are the sample means at baseline for control group and treatment group, respectively. The estimator $\hat{\delta}_1$ is a generalization of the previous estimators shown in 1.1 and 1.2 (Kleinbaum *et al.*, 1998).

Considering that here we have two groups (control and treatment) the two adjusted regression lines by fitting the ANCOVA model are:

Comparison group (T=0):

$$\hat{Y}_f = \hat{\delta}_0 + \hat{\delta}_2 Y_b$$

Treatment group (T=1):

$$\hat{Y}_f = (\hat{\delta}_0 + \hat{\delta}_1) + \hat{\delta}_2 Y_b$$

An alternative way to calculate the unbiased estimate of δ_1 is :

$$\hat{\delta}_1 = \bar{Y}_{treat,f}(adj\ ustd) - \bar{Y}_{ctrl,f}(adj\ ustd)$$

where:

$$\bar{Y}_{treat,f}(adj\ ustd) = (\hat{\delta}_0 + \hat{\delta}_1) + \hat{\delta}_2 \bar{Y}, \quad \bar{Y}_{ctrl,f}(adj\ ustd) = \hat{\delta}_0 + \hat{\delta}_2 \bar{Y} \quad \text{with}$$

$$\bar{Y} = \frac{n_{ctrl} \bar{Y}_{ctrl,b} + n_{treat} \bar{Y}_{treat,b}}{n_{ctrl} + n_{treat}}$$

So, the variance of the estimate of δ_1 is given by:

$$Var(\hat{\delta}_1) = \sigma^2(1 - \rho^2) \left(\frac{1}{n_{treat}} + \frac{1}{n_{ctrl}} \right)$$

and if $n_{treat} = n_{ctrl} = n$ then

$$Var(\hat{\delta}_1) = \frac{2}{n} \sigma^2(1 - \rho^2)$$

Thus, the covariance analysis reduces the variance of the treatment effect estimate and thereby is a more powerful statistical test (provides narrower confidence intervals) for examining the difference between groups (Koch *et al.*, 1982).

Non-parametric methods. Based on randomization in the study design, the analysis can be nonparametric. For instance, it is possible to use the Wilcoxon-Mann-Whitney test to analyze the strategies of the Simple model and Difference Score model above, and the Rank Analysis of Covariance to analyze the ANCOVA model (Wilcoxon *et al.*, 1945).

The Wilcoxon-Mann-Whitney test can test the null hypothesis that the distribution of an ordinal scale response variable is the same in two independent groups. This statistical test is sensitive to the alternative hypothesis that there is a location difference between the two groups. Also, this statistical test can be used when the t-test is appropriate (Wilcoxon *et al.*).

The Wilcoxon-Mann-Whitney test converges to the Mantel-Haenszel mean score statistic for the special case of one stratum when rank scores are used, if the sample size is large. Thus, another way to analyze the data under the Simple model and Difference Score model is using the Mantel-Haenszel mean score statistic (Wilcoxon *et al.*).

As mentioned before, the ANCOVA model depends on several assumptions which one must prove before fitting the model. In situations in which these assumptions are not satisfied, it can be used the Rank Analysis of Covariance (Quade *et al.*, 1982). This technique can be combined with

the extended Mantel-Haenszel statistics to establish nonparametric comparisons between treatment groups, after adjusting for the effect of the covariate (Koch *et al.*, 1982, 1990).

The advantages of nonparametric methods include higher statistical power under certain conditions, exact p-values for the test when sample size is small, and no assumptions of any kind of distribution. However, the disadvantage of this method is the lack of estimates of the magnitude of treatment effects.

Relative Efficient. The relative efficiency (Reffic) of the three parametric methods above can be calculated. The relative efficiency of the Difference Score model relative to the Simple model, the relative efficiency of ANCOVA relative to the Simple model, and the relative efficiency of the Difference Score relative to ANCOVA, is going to be calculated. The Reffic is defined in terms of the ratio of the variance of the effect based on each method.

Let Y_{bi} indicate the baseline measurement and Y_{fi} the follow-up measurement, then the difference score is: $d_i = Y_{fi} - Y_{bi}$. Assume normality for Y_{bi} and Y_{fi} , with $Var(Y_{bi}) = Var(Y_{fi}) = \sigma^2$ and correlation between Y_{bi} and Y_{fi} given by ρ . The variance of d is given by $Var(d_i) = 2(1 - \rho)\sigma^2$, with the restriction of equal variance for Y_{bi} and Y_{fi} . So, the efficiency of the Difference score relative to the Simple model is equal to:

$$\begin{aligned} Reffic(DifferenceScore, Simple model) &= \frac{Var(DifferenceScoreModel)}{Var(SimpleModel)} \\ &= 2(1 - \rho) \frac{Var(Y_{fi})}{Var(Y_{bi})} \\ &= 2(1 - \rho) \end{aligned}$$

A correlation between Y_{bi} and Y_{fi} greater than 0.5 would make the Difference Score model more efficient than the Simple model. Also, a correlation less than 0.5 insures that the Simple model is more efficient than the Difference Score model.

Under the ANCOVA model, the variance of Y_{fi} given Y_{bi} is:

$$Var\left(\frac{Y_{fi}}{Y_{bi}}\right) = (1 - \rho^2)Var(Y_{fi})$$

Then the efficiency of ANCOVA relative to the Simple model is given by:

$$\begin{aligned} Reffic(ANCOVA, Simple model) &= Var(ANCOVA) / Var(Simple model) \\ &= (1 - \rho^2) \end{aligned}$$

$$Reffic(ANCOVA, Simple model) = Var(ANCOVA) / Var(Simple model) = (1 - \rho^2)$$

This value is always less than or equal to 1. Therefore, ANCOVA is never less efficient than the Simple model. The efficiency of the Difference Score relative to ANCOVA is equal to:

$$\begin{aligned} Reffic(Difference Score, ANCOVA) &= Var(Difference Score) / Var(ANCOVA) \\ &= 2(1 - \rho) / (1 - \rho^2) \\ &= 2 / (1 + \rho) \end{aligned}$$

which is always greater than or equal to 1. So, ANCOVA never has less efficiency than the Difference score model.

Applications

Example: A Experimental Study to Compare Two Treatments of Cholesterol

This example is based on data from an experimental community-based trial to compare the efficacy of a school-based treatment with a placebo group for reducing cholesterol levels in children (Harrell *et al.*, 1996). By randomization, 617 children were assigned to the control group and 546 children were assigned to the treatment group. The primary outcome variable was level of cholesterol for the 1163 children, measured at two times, before intervention and after intervention.

The analysis plan for this experimental study identified 5 covariables at baseline as relevant candidates for adjustment, which were Height (cm), Weight (kg), VO_2 Max (aerobic capacity, ml/kg/min), Skinfold Sum (mm), and Systolic BP (mmHg).

Table I describes the characteristics of the children at baseline. From this table, one can see the imbalance in the distribution of baseline Cholesterol values for the two groups. The average cholesterol level was 164.9 mg/dl and 168.2 mg/dl in the control group and the treatment group, respectively. The distributions of the covariables Height, Weight, VO_2 Max, Skinfold Sum, and Systolic BP do not vary much between the treatment group and control group at baseline (Table I). The statistical analysis for this ES considers the nonparametric and parametric methods mentioned in section 1.

3.1. Parametric Analysis

The three models mentioned in section 1.1 are applied to the data in this section. PROC GLM in SAS is used to

Table I. Baseline Characteristics of the 1163 Children (mean±SD).

Variables	Control group (n = 617)	Treatment group (n = 546)
Cholesterol (mg/dl)	164.9±27.81	168.2±30.81
Height (cm)	136.6±7.21	135.6±7.11
Weight (Kg)	34.9±9.30	33.9±8.34
VO ₂ Max (ml/kg/min)	41.5±10.11	42.5±9.59
Skinfold Sum (mm)	26.2±14.58	25.5±13.68
Systolic BP (mmHg)	104.0±10.06	103.5± 9.94

calculate the corresponding statistics for each model. Also, an estimate of the difference effect between groups will be provided.

Table II shows the results of the analysis for the three models mentioned. For the Simple Model, the effect of the difference between groups is not significant (p-value=0.1168). The estimated effect difference in cholesterol at follow-up of the two groups is 2.73. This analysis is equivalent to using a t-test statistic for two independent samples.

Table II. Parametric Analysis Using Linear Models.

Model Specification	Model terms	Parameter estimate	Standard error	p-values
Simple model	α_0	161.25	1.268	< 0.0001
	α_1	2.73	1.741	0.1168
Difference score model	β_0	-6.70	0.866	< 0.0001
	β_1	6.05	1.188	< 0.0001
	δ_0	31.07	3.328	< 0.0001
ANCOVA model	δ_1	5.30	1.125	< 0.0001
	δ_2	0.77	0.019	< 0.0001

The second model, Difference Score, shows that the difference between the groups is significant (p-value < 0.0001). The estimate of the difference in cholesterol of the groups is 6.05. The adjusted linear model for the difference score is given by:

$$\hat{Y}_f - Y_b = -6.70 + 6.05T$$

The ANCOVA model shows that Cholesterol at follow-up depends strongly on the Cholesterol at baseline (p-value < 0.0001) and the difference between groups, after accounting for the baseline Cholesterol level, is significant (p-value < 0.0001). The two adjusted regression lines are: Comparison group:

$$\hat{Y}_f = 31.07 + 0.77Y_b$$

and Treatment group:

$$\hat{Y}_f = 36.37 + 0.77Y_b$$

, respectively and the estimate of difference between treatments after accounting for the baseline Cholesterol level is 5.30.

From the Difference Score model and the ANCOVA model, one can note that there is no difference in the conclusions for this analysis, because the p-values associated with the difference in the groups are almost the same. However, the value of the statistic associated with the difference in groups for the Difference Score model is greater than the value associated with the ANCOVA model, but the standard error of the estimate of difference between the groups for the Difference Score model is less than the standard error of the estimate of the difference for the ANCOVA model (Table II).

The analysis also considered adjusting simultaneously for the covariables Height, Weight, VO₂ Max, Skinfold Sum, and Systolic BP at baseline. Table IV and Table V show the results of this analysis. All of the models are ANCOVA models; however, keep the same names as given above (Simple model, Difference Score model, ANCOVA model) in order to facilitate comparisons among the models using additional covariables and those not using the covariables. The Simple model shows no difference between the groups (p-value = 0.1031). The difference score model shows a significant difference between the groups (p-value = 0.0001). The ANCOVA model shows that the effect of the difference between groups is significant (p-value=0.0001).

The conclusions related to the difference between groups obtained from those models with the 5 additional covariables (Height, Weight, VO₂ Max, Skinfold Sum, and Systolic BP) at baseline are the same as those without the additional covariables (see Tables IV and V). Also, the estimates of the difference effects between groups are almost the same for those models with the 5 additional covariables and those without the additional covariables. In other words, it is not necessary to adjust for the 5 covariables in the three models used, because the distribution of these 5 covariables at baseline is balanced between the two groups.

Table III. Wilcoxon-Mann-Whitney Test Using the Mantel-Haenszel Score Chi-Square Statistic.

Model	Mantel-Haenszel Statistic	df	p-value
Simple model	2.40	1	0.121
Difference Score model	26.36	1	0.001

Table IV. Rank Analysis of Covariance combined with Mantel-Haenszel Statistic.

Specification of the covariables	Mantel-Haenszel	df	p-value
Cholesterol at baseline (Y_b)	15.79	1	0.001
Y_b , Height, Weight, VO_2 Max, Skinfold Sum, Systolic BP	18.49	1	0.001

3.2 Non-parametric Analysis. The Wilcoxon-Mann-Whitney test was used to analyze the Simple model (no baseline data) and the Difference Score model. This test converges to the Mantel-Haenszel mean score statistic for the special case of one stratum when rank scores are used. In this example, the sample size is large enough, so that Mantel-Haenszel is appropriate for providing confirmatory inferences for treatment group being better than control group. PROC FREQ in SAS was employed to calculate the Mantel-Haenszel statistic.

For the Simple model, Table III shows the Mantel-Haenszel mean score statistic, which indicates that there is no a significant difference between treatment group and control group (chi-square = 2.401, 1 df, p-value = 0.121). Also, this table shows that for the Difference Score model, the Mantel-Haenszel mean score statistic is equal to 26.359

with 1 df, corresponding to a p-value of 0.001. Therefore, the cholesterol level change differs between the treatment group and the control group.

The ANCOVA analysis was carried out using Rank Analysis of Covariance combined with the extended Mantel-Haenszel statistics. This methodology can be implemented in the SAS System, by using PROC RANK, PROC REG, and PROC FREQ.

The ANCOVA analysis results with cholesterol at baseline as the only covariable is given in Table IV. The Mantel-Haenszel statistic is equal to 15.79, with 1 df, and a p-value of 0.001; which indicates a clearly significant difference between treatment group and control group after accounting for the cholesterol level at baseline.

In the ANCOVA analysis with cholesterol, Height, Weight, VO_2 Max, Skinfold Sum, and Systolic BP at baseline as the covariables; the Mantel-Haenszel statistic was equal to 18.49, 1 df, with corresponding p-value of 0.001. Thus, there is a significant difference between treatment group and control group after adjusting for the additional baseline covariables. The Mantel-Haenszel statistic from this model is just slightly larger than for the ANCOVA model with

Table V. Analysis Parametric Using Additional Covariables at Baseline.

Model Specification	Parameter Standard			
	Covariable	Estimate	Error	p-values
Simple model	Program	2.79	1.713	0.1031
	Height	-0.65	0.21	0.0026
	Weight	0.16	0.296	0.5896
	Uptake	-0.16	0.120	0.1829
	Skinfold Sum	0.30	0.145	0.042
	SBP	0.05	0.093	0.578
Difference Score model	Program	6.36	1.182	0.0001
	Height	-0.17	0.148	0.2532
	Weight	0.12	0.204	0.5606
	Uptake	0.17	0.083	0.0359
	Skinfold Sum	-0.001	0.100	0.9932
	SBP	-0.19	0.064	0.0024
ANCOVA model	Program	5.56	1.124	0.0001
	Cholesterol at Bl.	0.77	0.020	0.0001
	Height	-0.28	0.141	0.0495
	Weight	0.13	0.194	0.5087
	Uptake	0.10	0.079	0.2105
	Skinfld Sum	0.07	0.096	0.4892
	SBP	-0.14	0.061	0.0225

cholesterol as the only covariable. One of the reasons for this is the balance in the baseline distribution of these additional covariables for the two groups (see Table I).

The results from nonparametric methods agree with the results obtained using parametric methods.

Simulations

Simulation Process. This section considers the estimation of the difference effect between groups for the three models mentioned in section 1 using a simulation process. The data utilized for fitting the three models (Simple model, Difference Score model, and ANCOVA model) is generated in the following manner:

A) The outcome variable at baseline, Y_{bi} , is generated as following:

$$Y_{bi} = 165 + \{1 - \rho\} \sigma^2 \left\{ Z_{1i} + \left(\frac{\rho}{1 - \rho} \right)^{1/2} Z_{3i} \right\}, \quad i = 1, 2, \dots, n$$

Where Z_{1i} and Z_{3i} are independent N (0, 1).

The random variables Z_{1i} and Z_{3i} were created using the function RANNOR in SAS.

B) By using randomization, half of the n data generated at baseline is assigned to the control group and the other half is assigned to the treatment group. PROC PLAN in SAS was used to assign at random.

C) The outcome variable at follow-up, Y_{fi} , is generated in the following manner:

i) If the participant belong to the control group, then

$$Y_{fi} = 165 + \{1 - \rho\} \sigma^2 \left\{ Z_{2i} + \left(\frac{\rho}{1 - \rho} \right)^{1/2} Z_{3i} \right\}, \quad i = 1, 2, \dots, n/2$$

Where Z_{2i} and Z_{3i} are independent N (0, 1).

The random variable Z_{2i} was created using the function RANNOR in SAS, and Z_{3i} has values as before.

ii) If the participant belong to the treatment group, then

$$Y_{fi} = \beta + 165 + \{1 - \rho\} \sigma^2 \left\{ Z_{2i} + \left(\frac{\rho}{1 - \rho} \right)^{1/2} Z_{3i} \right\}, \quad i = n/2 + 1, \dots, n$$

where Z_{2i} and Z_{3i} are as before; and β represents the difference effect between the two groups.

Note that the outcome variable at baseline and follow-up as generated above have the same variance σ^2 , and the correlation between these two variables is equal to ρ .

The estimation of the difference effect between groups can be affected by different conditions, such as:

a) The standard deviation of the outcome variables at baseline and follow-up, which is assumed to be the same. Two different values will be used: $\sigma = 10$ and 30

The correlation, ρ , between the outcome variables at baseline and follow-up. The simulation will use the following values: $\rho = 0.4$ and 0.7

The difference effect between groups, β . Three different values will be used: $\beta = 2, 5,$ and 10

The sample size, n . Two different sample sizes will be used: $n = 100$ and 1000

Thus, combining all the possible values for σ, ρ, β and n , there are 24 different scenarios that one can analyze for comparing the three models of interest. For each of these 24 scenarios, three hundred simulations will be performed.

RESULTS

This section presents the results of the simulation, which are tabulated for the 24 scenarios mentioned before. Appendix contains these tables that present the estimation of the difference effect between groups, the statistical tests, and the p-values for the three models fitted.

In general, the results for the three models show that the estimates of the difference effect between groups are unbiased. Also, when the correlation between the outcome variable at baseline and follow-up is small ($\rho = 0.4$), the estimated variance of the estimated difference effect between groups under the Simple model is less than that under the Difference Score model. However, when this correlation is high ($\rho = 0.7$) the estimated variance of the estimated difference effect between groups under the Simple model is greater than that under the Difference Score model. The results also show that the estimated variance of the estimated difference of group effects under ANCOVA model is the smallest.

For a small value of the correlation between outcome variable at baseline and follow-up, $\rho = 0.4$, on average the p-value associated with the difference effect between groups

under the Simple model is less than that under the Difference Score model. However, on average the p-value associated with the difference of group effects under ANCOVA model is the lowest, but this value does not vary much with respect to the p-value under the Simple model. Thus, when ρ is small one can fit the Simple model or ANCOVA model and get the same conclusion with respect to the difference in groups. Also, when the value of ρ is high, on average the p-values associated with the difference effect between groups under Difference Score model are similar to those under ANCOVA model, so one can fit any of these two models. Therefore, if ρ is small one can say that the p-value obtained under the Difference Score model is the most conservative. But, if ρ is large, the p-value under the Simple model is the most conservative.

One can see that the p-values (for the three models) associated with the difference of groups effect when the Standard Deviation is small, SD=10, are less than when the Standard Deviation is high, SD=30.

For a small sample size, $n = 100$, and a small value of β , $\beta = 2$, there are no significant p-values (< 0.05), i.e. no significant difference effect between groups. Now, for a small sample size and a medium value of β , $\beta = 5$, there are significant p-values under the Difference Score model and ANCOVA model, only when the Standard Deviation is small and the correlation between the outcome variable at baseline and follow-up is large. When the sample size is small and β is large, there are significant p-values for the three models only when the Standard Deviation is small.

For a large sample size, $n = 1000$, and a small value of β there are significant p-values under the three models only when the Standard Deviation is small. Also, for a large sample size and a medium value of β , there are significant p-values under the three models when the Standard Deviation is small; but also under the Difference Score model and ANCOVA model when the Standard Deviation and ρ are large. If the sample size and β are large, then all of the p-values are significant.

APPENDIX

This appendix contains the tables that present the estimation of the difference effect between groups, the statistical tests, and the p-values for the three parametric models fitted using simulation.

Table Scenario 1 and 2: $n = 100$, $r = 0.40$, and $\beta = 2$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	2.13	2.106	1.98	6.063
BETA-DIFFERENCE	2.15	2.163	1.86	7.214
BETA-ANCOVA	2.13	1.885	1.88	5.797
TEST-SIMPLE	1.06	1.057	0.33	1.021
TEST-DIFFERENCE	0.98	0.992	0.28	1.111
TEST-ANCOVA	1.15	1.023	0.34	1.063
P-VALUES-SIMPLE	0.35848	0.30005	0.47036	0.28277
P-VALUES-DIFFERENCE	0.37286	0.30355	0.46891	0.30416
P-VALUES-ANCOVA	0.33605	0.29843	0.46766	0.29641

Table Scenario 3 and 4: $n = 100$, $\rho = 0.70$, and $\beta = 2$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	1.89	2.028	1.94	5.996
BETA-DIFFERENCE	2.00	1.573	1.84	4.499
BETA-ANCOVA	1.96	1.433	1.86	4.169
TEST-SIMPLE	0.95	1.015	0.32	1.005
TEST-DIFFERENCE	1.29	1.024	0.41	0.992
TEST-ANCOVA	1.37	1.006	0.44	0.991
P-VALUES-SIMPLE	0.36872	0.28276	0.47383	0.29576
P-VALUES-DIFFERENCE	0.30409	0.29458	0.47793	0.28864
P-VALUES-ANCOVA	0.28777	0.29518	0.46265	0.28229

Table Scenario 5 and 6: $n = 100$, $\rho = 0.40$, and $\beta = 5$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	4.82	2.019	4.84	5.826
BETA-DIFFERENCE	4.95	2.186	5.12	6.089
BETA-ANCOVA	4.86	1.865	4.92	5.214
TEST-SIMPLE	2.43	1.029	0.81	0.981
TEST-DIFFERENCE	2.27	1.011	0.78	0.929
TEST-ANCOVA	2.66	1.033	0.89	0.947
P-VALUES-SIMPLE	0.07872	0.15343	0.41614	0.31024
P-VALUES-DIFFERENCE	0.09707	0.16577	0.43133	0.31093

Table Scenario 7 and 8: $n = 100$, $\rho = 0.70$, and $\beta = 5$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	5.14	1.838	5.22	6.605
BETA-DIFFERENCE	5.03	1.609	5.22	4.840
BETA-ANCOVA	5.07	1.440	5.29	4.583
TEST-SIMPLE	2.59	0.954	0.88	1.127
TEST-DIFFERENCE	3.27	1.077	1.13	1.050
TEST-ANCOVA	3.56	1.055	1.24	1.073
P-VALUES-SIMPLE	0.05799	0.11358	0.36145	0.29418
P-VALUES-DIFFERENCE	0.02131	0.05847	0.33487	0.29490
P-VALUES-ANCOVA	0.01134	0.03742	0.30168	0.28182

Table Scenario 9 and 10: $n = 100$, $\rho = 0.40$, and $\beta = 10$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	10.04	2.012	10.33	6.121
BETA-DIFFERENCE	10.07	2.265	10.58	6.578
BETA-ANCOVA	10.03	1.899	10.44	5.648
TEST-SIMPLE	5.07	1.133	1.75	1.039
TEST_DIFFERENCE	4.61	1.062	1.62	1.018
TEST-ANCOVA	5.50	1.161	1.91	1.041
P-VALUES-SIMPLE	0.00067	0.00499	0.20891	0.26155
P-VALUES-DIFFERENCE	0.00088	0.00324	0.22627	0.26307
P-VALUES-ANCOVA	0.00015	0.00153	0.17873	0.24662

Table Scenario 11 and 12: $n = 100$, $\rho = 0.70$, and $\beta = 10$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	9.86	1.966	10.12	6.017
BETA-DIFFERENCE	9.87	1.635	10.10	4.712
BETA-ANCOVA	9.85	1.488	10.15	4.267
TEST-SIMPLE	4.98	1.036	1.70	1.033
TEST-DIFFERENCE	6.39	1.181	2.17	1.018
TEST-ANCOVA	6.91	1.179	2.36	1.014
P-VALUES-SIMPLE	0.00069	0.00416	0.19848	0.23299
P-VALUES_DIFFERENCE	7.64E-6	0.00004	0.12490	0.20897
P-VALUES-ANCOVA	2.93E-6	0.00003	0.09195	0.16248

Table Scenario 13 and 14: $n = 1000$, $\rho = 0.40$, and $\beta = 2$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	1.97	0.666	2.04	1.971
BETA-DIFFERENCE	1.96	0.752	1.91	2.060
BETA-ANCOVA	1.97	0.627	1.98	1.751
TEST-SIMPLE	3.11	1.055	1.08	1.039
TEST-DIFFERENCE	2.83	1.091	0.92	0.990
TEST-ANCOVA	3.39	1.087	1.14	1.007
P-VALUES-SIMPLE	0.03570	0.09729	0.35508	0.30024
P-VALUES-DIFFERENCE	0.05372	0.13366	0.37632	0.29190
P-VALUES-ANCOVA	0.02632	0.09137	0.34435	0.30417

Table Scenario 15 and 16: $n = 1000$, $\rho = 0.70$, and $\beta = 2$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	1.99	0.627	2.13	1.878
BETA-DIFFERENCE	1.99	0.459	1.96	1.506
BETA-ANCOVA	1.99	0.425	2.01	1.402
TEST-SIMPLE	3.14	0.985	1.12	0.989
TEST-DIFFERENCE	4.08	0.943	1.34	1.029
TEST-ANCOVA	4.41	0.945	1.49	1.037
P-VALUES-SIMPLE	0.02572	0.06895	0.35095	0.30204
P-VALUES-DIFFERENCE	0.00260	0.00817	0.28877	0.28628
P-VALUES-ANCOVA	0.00100	0.00407	0.26459	0.28391

Table Scenario 17 and 18: $n = 1000$, $\rho = 0.40$, and $\beta = 5$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	5.01	0.609	5.04	1.973
BETA-DIFFERENCE	5.02	0.682	4.97	2.263
BETA-ANCOVA	5.01	0.564	5.01	1.843
TEST-SIMPLE	7.92	0.979	2.65	1.036
TEST-DIFFERENCE	7.25	1.003	2.39	1.085
TEST-ANCOVA	8.65	0.999	2.87	1.052
P-VALUES-SIMPLE	8.5E-10	7.33E-9	0.06329	0.14258
P-VALUES-DIFFERENCE	3.93E-8	3.87E-7	0.10024	0.19833
P-VALUES-ANCOVA	5.5E-11	7.6E-10	0.05272	0.14443

Table Scenario 19 and 20: $n = 1000$, $\rho = 0.70$, and $\beta = 5$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	4.96	0.605	5.06	1.832
BETA-DIFFERENCE	4.97	0.443	5.01	1.463
BETA-ANCOVA	4.96	0.409	5.02	1.322
TEST-SIMPLE	7.83	0.977	2.67	0.966
TEST-DIFFERENCE	10.13	0.933	3.41	1.006
TEST-ANCOVA	10.97	0.950	3.71	0.983
P-VALUES-SIMPLE	2.43E-9	2.39E-9	0.05610	0.12928
P-VALUES-DIFFERENCE	5.4E-16	7.9E-15	0.01535	0.05037
P-VALUES-ANCOVA	1.4E-18	0.00000	0.00715	0.02379

Table Scenario 21 and 22: $n = 1000$, $\rho = 0.40$, and $\beta = 10$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	10.02	0.637	9.80	1.823
BETA-DIFFERENCE	10.04	0.702	9.94	2.131
BETA-ANCOVA	10.03	0.582	9.85	1.727
TEST-SIMPLE	15.88	1.067	5.17	0.978
TEST-DIFFERENCE	14.52	1.047	4.77	1.018
TEST-ANCOVA	17.34	1.061	5.66	1.001
P-VALUES-SIMPLE	0.00000	0.00000	0.00019	0.00144
P-VALUES-DIFFERENCE	0.00000	0.00000	0.00086	0.00784
P-VALUES-ANCOVA	0.00000	0.00000	0.00007	0.00103

Table Scenario 23 and 24: $n = 1000$, $\rho = 0.70$, and $\beta = 10$.

	SD = 10		SD = 30	
	Mean	STD Error	Mean	STD Error
BETA-SIMPLE	9.96	0.655	10.08	1.858
BETA-DIFFERENCE	9.95	0.471	9.99	1.508
BETA-ANCOVA	9.95	0.441	10.01	1.408
TEST-SIMPLE	15.77	1.085	5.32	0.982
TEST-DIFFERENCE	20.34	1.081	6.80	1.031
TEST-ANCOVA	22.05	1.102	7.39	1.048
P-VALUES-SIMPLE	0.00000	0.00000	0.00014	0.00127
P-VALUES-DIFFERENCE	0.00000	0.00000	7.67E-7	7.11E-6
P-VALUES-ANCOVA	0.00000	0.00000	8.61E-9	5.23E-8

DISCUSSION

Under the three parametric models one can get an unbiased estimator for the difference effect between groups. When the correlation of the outcome variable at baseline and follow-up is less than 0.5, the variance of the estimator of the difference effect under the Simple model is less than that under the Difference score model. But if the correlation is bigger than 0.5, the variance of the estimator under the Difference Score model is less than that under the Simple model. However, in general the variance of the estimator of the difference effect under the ANCOVA model is the smallest.

The Simple model should be used under the belief that randomization produces balance in the outcome variable at baseline.

The Difference Score model should be used when randomization produces an imbalance in the outcome variable for treatment and control group, especially in smaller studies or also when a "change" variable is the outcome of interest.

The ANCOVA model is recommended when randomization produces an imbalance in the baseline value

of the outcome variable for treatment and control group, especially in smaller studies.

From the simulation approach, if the correlation of the outcome variable at baseline and follow-up is small, one can fit either the Simple model or the ANCOVA model. Also, if this correlation is large one can fit either the Difference Score model or the ANCOVA model.

The overall recommendation of this paper is the complementary use of nonparametric methods and parametric methods for analyzing the data coming from an experimental study.

SANHUEZA, A.; OTZEN, T.; MANTEROLA, C. & ARANEDA, N. Métodos estadísticos para analizar un resultado continuo en estudios experimentales. *Int. J. Morphol.*, 32(1):339-350, 2014.

RESUMEN: En estudios experimentales, la variable resultado se mide en el momento inicial y luego en diversas ocasiones. De este modo, se habla de mediciones de "línea de base" y seguimiento respectivamente. Lo interesante de esta materia es poder determinar si una vez aplicada una intervención, existen di-

ferencias significativas entre el grupo al que se asignó un tratamiento de prueba y el grupo de comparación. En este manuscrito se exponen algunas de las estrategias utilizadas para tal propósito; las que incluyen un proceso de simulación mediante datos obtenidos a partir de un estudio experimental. Tres estrategias paramétricas y dos no paramétricas se evalúan teniendo en cuenta sólo una medida de seguimiento. La medida de referencia se incorpora en el contexto de estas estrategias.

PALABRAS CLAVE: Estudio experimental; Estudios de seguimiento; Bioestadística; Estadísticas no-paramétricas.

REFERENCES

- Friedman, L.; Furberg, C. & DeMets, D. *Fundamentals of Clinical Trials*. 2nd ed. St Louis, Mosby-Year Book, 1985.
- Harrell, J. S.; McMurray, R. G.; Bangdiwala, S. I.; Frauman, A. C.; Gansky, S. A. & Bradley, C. B. Effects of a school-based intervention to reduce cardiovascular disease risk factors in elementary-school children: The Cardiovascular Health in Children (CHIC) Study. *J. Pediatr.*, 128(6):797-805, 1996.
- Kleinbaum, D.; Kupper, L.; Muller, K. & Nizam, A. *Applied Regression Analysis and Multivariable Methods*. 3rd ed. Pacific Grove, Brooks/Cole Publishing Company, 1998.
- Koch, G. G.; Amara, I. A.; Davis, G. W. & Gillings, D. B. A review of some statistical methods for covariance analysis of categorical data. *Biometrics*, 38(3):563-95, 1982.
- Koch, G.; Carr, G.; Amara, I.; Stokes, M. & Uryniak, T. *Categorical data analysis in Statistical Methodology in the Pharmaceutical Sciences*. New York, Marcel Dekker Inc., 1990. pp.391-475.
- Piantadosi, S. *Clinical Trials: a methodologic Perspective*. 2nd ed. New York, John Wiley & Sons Inc., 1997.
- Quade, D. Nonparametric analysis of covariance by matching. *Biometrics*, 38(3):597-611, 1982.
- Wilcoxon, F. Individual comparison by ranking methods. *Biometrics*, 1(6):80-3, 1945.

Correspondence to:
Dr. Antonio Sanhueza
Department of Mathematics and Statistics
Universidad de La Frontera
Temuco
CHILE

Email: antonio.sanhueza@ufrontera.cl

Received: 29-06-2013

Accepted: 19-12-2013