Articles

New Possibilities of Analysis of Experimental Data in Pedagogical Research

Vladimír Polášek, Lubomír Sedláček

Abstract

The aim of this paper is to present new possibilities of the statistical analysis of pretest-posttest data types measured in a pedagogical experiment. These possibilities will allow us to eliminate the effect of so-called disruptive independent variables that could adversely affect a dependent variable and thus, the research results. By using a model example, we present the design of a suitable experimental plan. Further, we illustrate advanced statistical techniques – Analyses of Variance and of Covariance using a real example.

Key words: Pedagogical experiment, pretest, posttest, analysis of variance for difference scores, analysis of covariance.

Introduction

Although pedagogical experiment is one of the basic methods of educational research and there are such kinds of pedagogical problems that cannot be credibly answered without making an experiment (e. g. verification of new forms and methods of teaching, implementation of major changes in education reform, etc.), contemporary Czech educators tend to neglect this method (Průcha, 2000, p. 197). Jan Průcha (2000, p. 198)

reports the relatively high complexity of this research procedure as the main reason. Its application must meet a number of requirements; the satisfaction of which is linked with considerable problems.

Pedagogical experiment is used to research the learning process. It studies the behavior of entities – i.e. individuals participating in the experiment with the introduction of a controlled change. The experimenter intentionally changes the conditions by manipulations with variables and examines the effects of this experimental intervention and its results. So, experiment is a research method for the verification of scientific hypotheses about causal relationships between phenomena (Průcha, 2003).

Scheme of the research method and its characteristics

As the most fundamental problem in the implementation especially of a field experiment conducted in the natural environment of the studied subjects (i.e. in vivo); we considered the design of such an experimental plan that would have the maximum ability to eliminate the effect of so-called disruptive variables which could adversely affect the dependent variable, and thus significantly skew the results of the research. Part of that plan is the appropriate use of advanced statistical methods and statistical programs for the analysis of experimental data. The following text, which provides two different evaluation examples of a pedagogical experiment, could serve as instructions to deal with these difficulties.

The basic procedures are demonstrated on the model example. We have formulated the following research problem: *Do the results of computer-assisted teaching differ from the traditional frontal teaching of geometry with the dominant teacher interpretation?* Thus, we will compare the traditional teaching method typical for the Czech educational environment and so-far less used method: computer-assisted teaching implemented using some of the programs in the field of Dynamic Geometry, e.g. GeoGebra, Cabri Geometry II Plus, etc. We formulate the null hypothesis for this research problem as follows: H_0: *Increase of knowledge and skills of the pupils acquired during computer-assisted teaching of "Congruent Projection" is equal to the increase in pupils' knowledge and skills gained in the frontal teaching of "Congruent Projection".*

For the best research method in this case, we consider pedagogical experiment which is realized by the technique of parallel groups (Chráska, 2007, p. 29). This type of experiment is based on the fact that a researcher is working with two groups. The control group is not exposed to the intervention whose effects are researched. The experimental group is exposed to the studied type of intervention. The other conditions are as equal as possible; it ensures that the incurred differences can be attributed only to the debit of examined interventions (Hendl, 2004, p. 60). In our case, we handle in the experimental group with the independent variable, which is a teaching method, and

we examine its influence on the dependent variable which is an increase in students' knowledge and skills. The teaching method is the only element that differentiates two groups from each other.

As already mentioned above, the main problem with this type of experiment is to design its plan in such a way that the elimination of the effect of the other disruptive independent variables (covariates) on the dependent variable is maximized. If we cannot separate or measure this effect on the dependent variable and adjust the results when evaluating with an appropriate statistical method, we often approach randomization which is the process of a random placing of subjects to the groups being compared. In our case, we propose to use a combination of these two possible procedures and to structure the experimental plan so that the way of experiment arrangement and the used evaluation method would allow us to take the effect of covariates into account and allow us to estimate the "net" effect of the intervention.

Therefore, we propose to divide the arrangement of the whole research into the following phases.

Randomization

We will **randomly** select two classes at the relevant school. One class will be determined as "experimental", while the other one as "control". In no case do we try to divide the students into two equal groups using an assessment test – according to their current level of geometry knowledge and skills – before the start of the experiment. This method of rearranged or paired groups leads to fragmentation of previously organized classes, which is very inconvenient and, from an administrative viewpoint, practically impossible. For this reason, it is better to proceed to the use of modern statistical procedures (Analysis of Covariance, Linear Mixed Model, Analysis of Variance, see below), which allow us to perform an equally reliable experiment without regrouping the pupils into new groups (Chráska, 2007, p. 31).

We further recommend that each class be taught by different, **randomly** selected Mathematics teachers. The teaching both groups by one teacher could affect the experiment due to greater teacher interest in one of the teaching methods.

Pretest

We will measure the current level of knowledge and skills in the field acquired so far classically through the curriculum of geometry in each student from the both groups by using didactic input test.

Application of Intervention

This phase of the experiment is crucial. It is characterized by performing the intervention in the experimental group by manipulating the independent variable, which is here the way geometry is taught. All other elements of the teaching process remain the same in both groups. In the experimental group, geometry is taught in the mathematics classroom, equipped with a computer and a data projector and using programs from the dynamic geometry field; while in the control group, geometry is taught using traditional frontal teaching methods with the dominant interpretation of the teacher, standard writing and drawing on the board.

Figure 1
Diagram of disruptive variables impact elimination
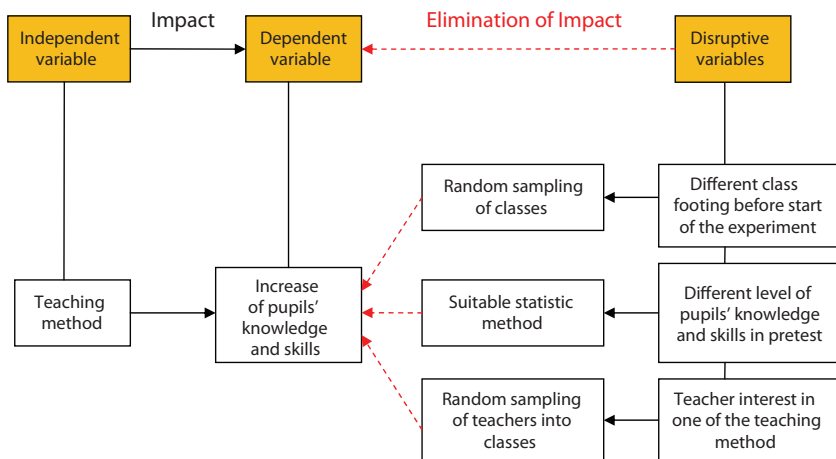
Comparison

The content of this penultimate stage of the teaching experiment is the statistical analysis of the data obtained in the didactic tests and their comparison in both groups.

Conclusion

In the final phase of the experiment, we will formulate the consequences of the intervention under study in the experimental group, i. e. whether significant differences in target geometry knowledge and skills in each group were measured, i. e. the acceptance or rejection of the null hypothesis.

Selection of a Statistical Method for Data Analysis

When selecting statistical methods, it was necessary to realize that we are working with uneven, unpaired groups; which means that the initial level of geometry knowledge and skills of students before the start of the experiment may be different in different groups. However, this can significantly affect the credibility of the whole research. Therefore, when analyzing the measured pretest – posttest data, the pretest values are considered to be one of the disruptive variables (covariates) and we must correct this undesirable effect.

There are two basic possibilities. At first sight, **analysis of variance for the difference scores** seems to be the easier procedure. We analyze the differences of pretest and posttest data by testing the differences of averages for the experimental and control groups. We determine change (i. e. the output – input difference) with each student and – using variance analysis, we compare the rates of changes in these groups. Another possibility is **analysis of covariance**, which combines regression analysis and analysis of variance. In this method, we make an adjustment to the output evaluation against the input evaluation with each student and compare only these adjusted values. By adjusting, we will obtain data that is not dependent on the students' initial knowledge and skills level. From this point-of-view, we will attain two similar groups without any comparison of them in advance. The principle of this method consists in the fact that we estimate regression of the final measured values to the initial values in every particular group. This yields estimates of the regression coefficients for both the experimental and control groups. Then, we will discover whether it is possible to consider the slopes of the regression lines as estimates of the same constant so that we test the hypothesis validity by the relevant statistics, which assumes that the real estimated slopes are identical. If we do not refute this hypothesis, we will assume that the two slopes are identical, thus the output test in both groups reacts to changes in the input test as sensitive. The finding is a necessary condition for the model application of covariance analysis. We assume therefore, that there is no real difference in the regression in both groups, the regression coefficients in the groups are equal, and the regression lines in both groups are parallel. It remains to decide whether we can identify them, or whether they are necessarily displaced relative to one another. Analysis of covariance

precisely consists in making this decision. In the event of a significant displacement, we will have a reason to reject the null hypothesis, which argues that the knowledge level of students in the experimental and control groups is the same, and we can lean toward the alternative hypothesis. We can therefore conclude that the output level of the students' knowledge is different in both the groups. Then it only remains to show which group of the students achieved significantly better results.

We consider the analysis of covariance as the optimal approach to the assessing the comparative experiments. However, it is necessary to meet the basic assumptions of this method. Only when they are fulfilled, is it possible to interpret the results. The basic assumptions are:

- classes were selected randomly to the groups,
- regression coefficients in the groups are equal,
- regression is linear,
- variance in the groups is homogeneous (homoscedasticity),
- normal classification in both the groups.

We can note that for both methods, the analysis of difference scores focused on the change between the pretest and the posttest, is simpler and does not require any assumptions of the regression conformity of the posttest to pretest data in the individual groups, which is a condition for the more complicated analysis of the covariance. However, the interpretation of differential scores is less valid if there are differences between the groups in the pretests. On the other hand, for equivalent groups, we gain no advantage when using covariance analysis. Therefore, it is recommended to make both the ways of analysis and compare their results. If the results come out the same, we have reason to consider them robust (Hendl, 2004, p. 395).

Figure 2
Diagram of a teaching experiment
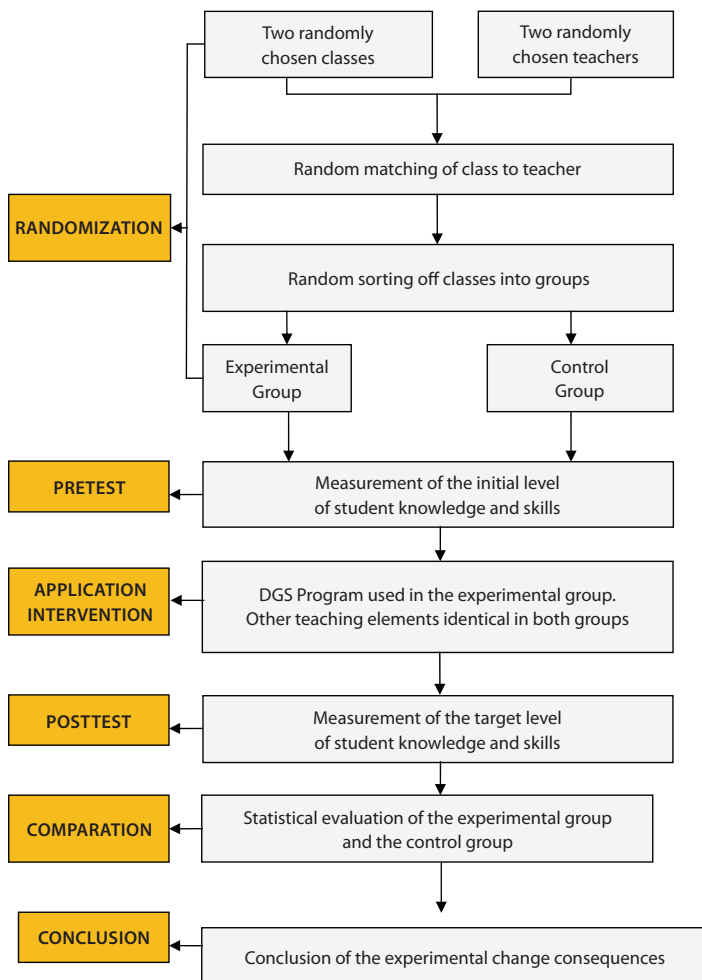
Statistical Evaluation of a Teaching Experiment

Now, we will demonstrate the basic procedures of both the methods with specific values. As mentioned above, we have data of two groups of the students (Experimental, Control), each student was evaluated by two tests (Input, Output). The goal is to compare these two groups.

Analysis of Variance for Difference Scores

We check for the difference between values of the output and the input tests with each student. These differences can be characterized by the following descriptive statistics:

Group	Average	Standard Deviation
Experimental	1.40	4.97
Control	−1.48	4.84

At first sight, it is apparent that the variability is comparable in both the groups (the classical F-test gives a value $F = 1.05$, p-value of $p = 85.1\%$), so it can be used as the Student's t-test. The test statistics $T = 2.96$ gives a p-value of $p = 0.4\%$, so – at the commonly used 5% level, we have **demonstrated a difference between the groups**. In the experimental group, a greater difference between the values of the output and the input test is demonstrated. A common estimate of the standard deviation of the two selections gives $S = 4.90$.

Analysis of Covariance

It is useful to complete the analysis of the covariance model by a graphical representation. The observations of the experimental group are expressed by triangles in Fig. 3; the observations of the control group by crosses. Each observation is illustrated by a point whose x-coordinate is given by the value of the input test and y-coordinate by the value of the output test.

Firstly, let's suppose that we interpolate the data in each group by the individual line. We get estimates of dependence in the form:
Output = 2.11 + 0.94* *Input* for the experimental group
Output = 1.68 + 0.77* *Input* for the control group

The slopes of the regression lines (0.94 and 0.77) are similar, so it is important to discover whether we can consider them to be two estimates of the same constants. The relevant

test statistics $T = -0.83$ lead to a p-value of $p = 41.0\%$, so the hypothesis according to which the real estimate slopes are compared is identical, we cannot refute this.

Furthermore, we will assume that both the slopes are identical, and thus the output tests in both the groups react equally sensitively to changes in the input tests. We assume therefore, that both the lines are parallel, and it remains to decide whether we can identify them with each other, or whether they are necessarily shifted relative to one another.

The parallel lines drawn on the diagram of Fig. 3 have the following equations:
Output = 3.42 + 0.83* *Input*   (Experimental)
Output = 0.77 + 0.83* *Input*   (Control)

The test of the hypothesis that the displacement of the lines in the vertical direction by $3.42 - 0.77 = 2.65$ is random leads to test statistics $T = 2.72$ with a p-value of $p = 0.8\%$. We have therefore, at the 5% level, proven that **both the compared groups differ from each other**. The resulting estimate of the standard deviation is equal to $S = 4.87$, so this is only slightly smaller than with the analysis of variance for differential scores.

Figure 3
Graphical representation of the results dependence of the output test on the input test, assuming that the slopes are identical
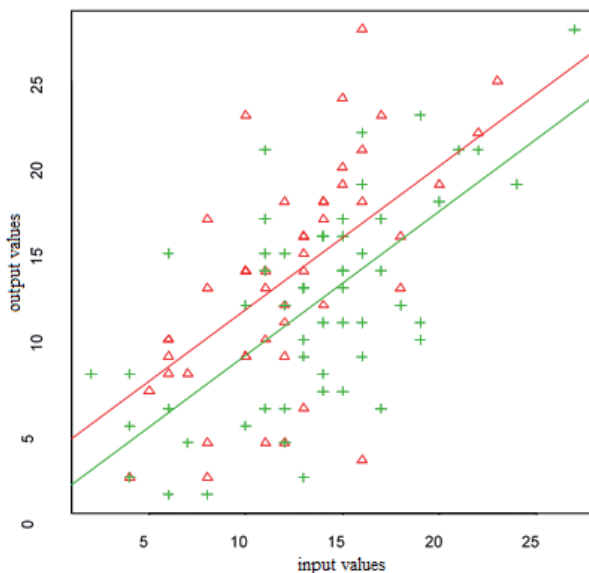
We can then verify for sure whether we can assume the fulfillment of the common assumptions for the application of a linear model. Shapir's Test of the distribution normality of the random component of the model leads to a p-value of $p = 71.4\%$, so it does not indicate a problem. Similarly, homoscedasticity (i. e. constant dispersion) tests can be $p = 88.9\%$ (versus the alternative of the variance dependence on the value of regression function), $p = 68.8\%$ (versus the alternative of different variances in these groups), $p = 99.6\%$ (according to the alternative of the variance dependence on the input test values). So, no problem appears even in the potential non-constant variance. The application of analysis of covariance of the model is therefore adequate.

Conclusion

The test results comparison by the variance analysis for differential scores can be considered a special case of the analysis of the covariance model, where we do not estimate a common slope of the regression lines but we choose it firmly straight number one. The hypothesis test that the real slope of the parallels equal one, leads to statistics $T = 1.62$ with a p-value of $p = 10.9\%$. We did not prove therefore, that it would be necessary to use the analysis of the covariance model. The ninety-five percentage reliability interval for the common slope is (0.63, 1.04), so it covers the value 1. We can consider the number one as a possible common slope.

Both the proposed methods are therefore applicable. Both also give results of the same quality. We have already demonstrated the difference between the experimental and control groups. In the analysis of difference scores, the released p-value of 0.4% was with 100 degrees of freedom, whereas in the analysis of covariance model the released $p = 0.8\%$ was with 99 degrees of freedom.[1]

Now, it remains to decide whether statistically significant differences between the achieved levels of knowledge and skills in the students can be attributed to the teaching methods used in the experimental and control groups. This can be only stated when the research plan was strictly respected and the influence of possible disruptive variables was eliminated. If the decision was to be complicated by the fact that each group consisted of two or more randomly selected classes, we would have to take account of this fact. The results of the tests within the given class might be more similar to each other than the results of pupils from different classes. The values from one class could therefore be dependent among themselves. For this reason, it would not be possible to use the classical analysis of the covariance model. It would be necessary to proceed to the so-called Mixed Linear Model, which combines fixed effects (i. e. constants characterizing the level of output knowledge and skills, a coefficient determining to what

---

[1]   All the calculations were performed using the freeware program R, which is freely available on the Internet (http://cran.r-project.org/).

extent the result of the entrance test affects the result of the output test) and random effects (i.e. expressing the fluctuation of the resulting test between classes) (Sedláček, 2009, p. 97).

Acknowledgement

Literature

Bertrand, Y. (1998). *Soudobé teorie vzdělávání.* Praha: Portál.
Gavora, P. (2000). *Úvod do pedagogického výzkumu*. Brno: Paido.
Chráska, M. (2007). *Metody pedagogického výzkumu*. Olomouc: Grada Publishing.
Hendl, J. (2004). *Přehled statistických metod zpracování dat.* Praha: Portál.
Průcha, J. (2002). *Moderní pedagogika.* 2nd ed. Praha: Portál.
Průcha, J. (2000). *Přehled pedagogiky.* 2nd ed. Praha: Portál.
Průcha, J.; Walterová, E., & Mareš, J. (2003). *Pedagogický slovník. 4th revised ed.* Praha: Portál.
Sedláček, L. (2009). *A Study of the Influence of Using Dynamic Geometric Systems in Mathematical Education on the Level of Knowledge and Skills of Students*. Acta Didactica Universitatis Comenianae – Mathematics, 9, pp. 81–108.
Zvára, K. & Štěpán, J. (2001). *Pravděpodobnost a matematická statistika*. 2nd ed. Praha: MATFYZPRESS.

Contacts:
Mgr. Vladimír Polášek, Ph.D., Mgr. Lubomír Sedláček, Ph.D.
E-mail: vpolasek@fai.utb.cz, lsedlacek@fai.utb.cz
Dept. of Mathematics, Faculty of Applied Informatics
Tomas Bata University in Zlín
Nad Stráněmi 4511
760 05, Zlín, Czech Republic