ELEMENTS OF VARIANCE ANALYSIS, EVALUATION OF DIFFICULTIES BY QUESTIONNAIRE

Osmar D. Vera[1] and Carmen Díaz Batanero[2]
[1]University of Quilmes, Argentina; [2]University of Huelva, Spain
overa@unq.edu.ar

In this paper we present a study where we assess the difficulties and errors in understanding some elements of Variance Analysis (ANOVA) by a sample of 224 Psychology students, after finishing a Data Analysis course. We analyze the selection of an ANOVA model, the understanding of assumptions, and the associated linear model, the computation of ANOVA in different models, the interpretation of results. These results provide information in an area where no much previous research is available.

INTRODUCTION AND BACKGROUND

Statistical inference plays a prominent role in various human sciences, including Psychology, with its research based on data collected in population samples, which needs to extend its conclusions. However, the use and interpretation of statistics for publications in Psychology is not always appropriate, as shown in several reviews (Borges, San Luis Sanchez, & Cañadas, 2001; Batanero & Diaz, 2006). Also occur in college students, as shown by many studies (Vallecillos 1994), most of which focus on understanding the concept of level of significance, and essential interpretation to determining if the presence of some factor is significant for Variance Analysis. But while, thanks to statistical software, performing the calculations associated with variance analysis is very easy today, teaching concepts and inferential reasoning is much more complex, which explains the many difficulties currently apparent in how inference is applied. In this paper we evaluate how certain basic concepts of variance analysis are employed by a group of Spanish psychology students who, after completing a course on inference, were given a short questionnaire designed to assess how well they understood the following: what determines the selection of a particular ANOVA model; how assumptions regarding the appropriate associated linear model are understood; how ANOVA is calculated in different models, and how results are interpreted. Below we describe the background for our investigation and the sample and method used, discuss the results, and conclude with some reflections on the implications of our research for teaching.

Are almost nonexistent research related to learning or teaching variance analysis, its various models, choice of model assumptions and verification thereof or the interpretation and understanding of how a table of variance analysis or experiment´s design is obtained by which we think that our work provide new results. Rubin and Rosebery (1990) Planned and found a teaching experiment aimed at studying the difficulties of interpretation of some of the basic ideas of experimental design. One is the difference between dependent and independent variable. Another important factor in the variance analysis is the concept of interaction. The presence of interaction is what stands out in ANOVA models on several factors, as it allows to evaluate the effect varies as a factor on the response at different levels of another. Rosnow and Rosenthal (1991, p. 574) indicated that the interaction is the result most universally misunderstood in Psychology. Green (2007) also emphasized that students have difficulty understanding the concept of interaction in variance analysis. In sum, the number of elements involved makes research on ANOVA complex, and few studies have been done that directly monitor what students actually learn or test whether learning objectives have been achieved at the end of the course. In our research project, we, in collaboration with the art teacher, monitored students' progress during a two-year data analysis course and included relevant questions as part of their final evaluation. Below we describe the methodology employed and discuss the results.

RESEARCH METHODOLOGY AND MATERIAL

The sample consisted of 224 students in the second year of the Bachelor of Psychology at the University of Huelva, attending a course on Data Analysis II, in the second year of study where they had studied issues related to statistical inference, including ANOVA emphasized at the elementary level. The questionnaire used (Figure 1) consisted of eight multiple-choice items (three

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

choices per item). To develop the questionnaire rigorously, it began with a semantic definition of the construct "understanding of elementary objects ANOVA", defining the content to be assessed.

Item 1. To improve the motor skills of elementary school children, a teacher believes that new physical activities help them. The teacher divides his working group randomly into three equal parts. Each group uses a different type of exercise because she wants to know what kind of exercises will give you better results. Statistical techniques that follow what the teacher should be applied to check whether the methods applied are different?

a. Hypothesis testing *t* on independent means.
b. Hypothesis testing *t* on related samples.
c. **One Way ANOVA completely randomized .**

Item 2. A researcher uses an variance analysis of two factors, fixed effect and fully randomized when:

a. The study has an independent variable, with two randomly selected levels.
b. The study has two independent variables, each with two or more levels.
c.The study has two dependent variables.

Item 3. The assumptions for the application of two-way ANOVA, fixed effects and completely random are:

a. Independence of observations, normal distribution, and additivity.
b. Independence of observations, equal variances and additivity.
c. Independence of observations, normal distribution and equal variances.

Item 4. Two-way ANOVA factors, fixed effects, decomposes the total variability in the following components :

a. Total variability = V. between groups + V. error.
b. Total variability = V. between groups + V. between subjects + V. error.
c. Total variability = V. factor A + V. factor B + V. interaction + V. error.

Item 5. If a one-way ANOVA, and repeated measures find that empirical or observed F takes a value of 8.16 meaning that:

a. CM between groups / CM error = 8,16.
b. CM between groups / CM error = 8,16.
c. CM between groups / CM inter subject = 8,16.

You want to study the effect of certain motivational variables on performance in achievement tasks. Two variables are manipulated: "Type of motivational training" (A1: instrumental, A2 attributional and A3: control) and "classroom climate" (B1: cooperative, B2 competitive and B3 Individual). 45 subjects were selected and divided into groups for each experimental condition. Here are incomplete ANOVA table.

Table ANOVA

Source of variation	SC	Gl	CM	F
Factor A	70			
Factor B			20	
Interaction AB				3,91
Error	46		1,278	
Total	176	44		

From the information in Table ANOVA answers the questions of items 6, 7 and 8.

Item 6. The value of the square sum for factor B (see Table 1) is :

a. 15,65
b. 35
c. 40

Item 7. The value of the mean square for factor A (see Table 1) is :

a. 5
b. 35
c. 15,65

Item 8. One of the conclusions of the study is (alfa = 0,05) (Table 1 using full)

a. There is effect of factor A ("training") on performance in achievement tasks.
b. No effect of factor B ("classroom climate") on performance in achievement tasks.
c. No interaction of factors.

Figure 1. Questionnaire

The first item assesses students' understanding about the basic elements of elementary statistical model variance of analysis, and differences between this model and the one used for a t-test. The correct answer is (c). Both the option (a) and (b) would choose students who confuse the ANOVA model with the two-sample t test. The second item assesses whether the student understands when a researcher will apply a two-way ANOVA with fixed effects, as well as the difference between dependent and independent variables and the role these play in the model is chosen. The correct answer is (c). A student chooses the option (a) not know the assumptions of ANOVA two-factor model, confusing it with the paired samples t test; while those who choose the option (b) confuse the two-factor ANOVA model with a completely randomized factor with fixed effects. Item 3 assesses the understanding that the group has on the assumptions that must be met to apply the observations given a model ANOVA. The correct answer is (c). Students who choose the option (a) forget that to apply this model it is necessary that the variances taken for each level must be statistically equal, while those who choose the option (c) forget the normality assumption that must be met for this model. Item 4 we assess the understanding of the student group to associate a statistical model, according to the study that you want to perform data. The correct answer is (c).

Students who choose the option (a) does not consider the variability for each factor separately, and does not consider interaction; while those who choose the option (b) confuse the decomposition of the variability of a two-factor model with which corresponds to a one-factor model and repeated measures. Item 5 assesses the group's understanding of some of the calculations to interpret an ANOVA table in the design of a factor with repeated measures, results of which will depend on the decision making. The correct answer is (a). In the option (b) the variance between groups to the variance between subjects is confused, which gives us an idea of misunderstanding of the calculation method with the left-sided test. Those who choose the option (c) think that the variance analysis for repeated measures must observe the interaction between the subject-group variables and do not consider this to be an additive model. Both in item 6 and 7 we interpret student responses to questions asked on the calculation of the values that form a two-factor ANOVA table asking the various components involved in this case (item 6) a sum of squares. For item 7 asking about the value of the mean square for a factor. The correct answer is (c) item 6, and (b) item 7. Item 6: the option a) gives the observed value of F for that factor, as if the sum of the square, confusing that with an element of calculation which requires more complex, since this is the ratio of two mean squares. For the option (b) the value of the square sum for factor B is confused with the mean square for the other factor. Item 7: a student who chooses the option (a) would be confusing with the average square of the interaction, while those who choose the option (c) confused with the observed F for factor B. Finally, in item 8 we evaluate how well students perform when making a decision on the effect of the factors involved in responses. The correct answer is (a), since the p-value calculated for this factor is very close to 0. A student who chose options (b) or (c) errs in not considering interaction as another relevant factor. The interaction for 5% and the factor B is highly significant.

RESULTS AND DISCUSSION

Students completed the questionnaire as part of course evaluation. To prevent random responses, they were warned they would be penalized for incorrect answers. Table 2 shows the percentages of correct responses per item, with the appropriate option figuring in boldface.

Table 2. Percentages Item responses (n = 224)

Resp	Ítem 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
a	23,7	4,9	11,2	5,8	**57,1**	3,6	3,1	**33,0**
b	7,6	9,8	4,0	12,5	7,1	4,0	**72,3**	6,3
c	**50,9**	**63,8**	**43,3**	**69,2**	21,9	**72,3**	3,1	12,5
Unansw.	17,9	21,4	41,5	12,5	13,8	20,1	21,4	48,2

In item 1, 50.9% of students chose the correct answer. 23.7% of the sample chose option (a); although able to differentiate in a problem with independent and related samples, this group didn't choose the appropriate procedure, a difficulty noted by Rubin and Rosebery (1990). Another 7.6% incorrectly selected contrast t, thus further confusing independent and related samples. 17.9% seemed to ignore the differences among assumptions. Item 2, which also refers to the choice of models, was answered correctly by more than half of the students (63.8% Table 2). 4.9%, (option (a)) unknown model assumptions of two-factor variance analysis, confusing it with the paired samples t test. 9.8% (option (b)) confused the two-way ANOVA model with a one-factor, completely randomized one with fixed effects. No respond to the question (21.4%). In item 3 43.3% (option (c)) answered correctly. 11.2% chose option (a), forgetting the principle of homoscedasticity for this model. Another 4% chose option (b), forgetting the assumption of normality. Did not response (41.5%). In item 4, based on the association with a statistical model, 69.2% responded correctly. 5.8% mistook it for a single-factor ANOVA model. 12.5% of the sample confused a variability decomposition model with a repeated measures factor. While more than half of the evaluated sample was able to associate an appropriate statistical model with two-factor variance analysis, 18.3% (option (a)+(b)) of the sample did not associate the interaction with full-model variance analysis with two fixed factors. In this regard, several authors have pointed to the interaction between interpreted factors as the worst result among psychology students (Rosnow and Rosenthal, 1991; Green, 2007; Pardo et al, 2007). Also 12.5% did not respond. Item 5, which assesses how well the calculations required to complete an ANOVA table are understood, 57.1% responded correctly.7.1% chose option (b), confusing variance among groups with variance among subjects, a consequence of having mistaken what calculation method to use. 21.9% chose option (c), thus forgetting that a repeated-measures ANOVA model is only additive. As we found no other research on this question with which to compare our students' responses, we appear to be breaking new ground in this regard. Item 6, where we analyze how well interpreting the values in an ANOVA table is understood by 72.3% of students. 3.6% chose option (a), giving the observed value of F for that factor. Another 4% confused the value of the square sum for factor B with the mean square for the other factor. Item 7, which asks for the value of the mean square, has 72.3% correct responses. 3.1% (option a) confused the mean square of the interaction with factor A. We also found a low percentage, 3.1%, confused the factor A mean square with factor B F observed (option c). Did not respond (20.4%). In item 8 how the values obtained in the full-table ANOVA are interpreted in decision-making is evaluated. We found a low percentage of correct responses, 33%, and a high percentage of non-responses (48.2%).

CONCLUSION

In our investigation more than acceptable results are obtained, even though our sample consists exclusively of psychology students who are not academically accustomed to manipulating mathematical models. In our opinion, these satisfactory results are due to the fact that, prior to the evaluation, our students had taken two courses in statistics. Despite the high degree of non-response, items 6 and 7 showed the facility with which students resolved questions directly related to calculation. Not all errors were eradicated during the two courses taken by students. We found that some errors persisted after completing the two-year cycle, which, as noted above. We found no research evaluating the understanding by students of assumptions in ANOVA tests. We believe it is important to emphasize the meaning of these concepts, as faulty comprehension makes researchers unable to interpret the results of their work or to critically evaluate the results of research published in journals in their field. This confusion can cause students and researchers to make wrong decisions. We agree with Vallecillos (1994) and Diaz, Batanero and Wilhelmi (2008) that the value p and the level of significance are concepts that should be sufficiently clear and well-connected because correct decision-making ultimately depends on them. We need to review how statistical inference is taught, as stated in Vera, Diaz and Batanero (2011). It would be important to begin introducing these objects informally in middle school.

REFERENCES

Batanero, C., & Díaz, C. (2006). Methodological and didactical controversies around statistical inference. *Actes du 36iémes Journées de la Societé Française de Statistique* [CD-ROM]. Paris: Societé Française de Statistique.

Borges, A., San Luis, C., Sánchez, J. A., & Cañadas, I. (2001). El juicio contra la hipótesis nula: muchos testigos y una sentencia virtuosa. *Psicothema, 13*(1), 174-178.

Díaz, C., Batanero, C., & Wilhelmi, M. R. (2008). Errores frecuentes en el análisis de datos en educación y Psicología. *Publicaciones, 35*, 109-133.

Green, K. E. (2007). Assessing understanding of the concept of iteraction in analysis of variance. Trabajo presentado en la *IASE/ISI-Satellite Conference on Assessing Student Learning in Statistics*. Guimarães, Portugal: International Statistical Institute.

Rosnow, R. L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at *only* the interaction (unless all main effects are zero). *Psychological Bulletin, 110*, 574-576.

Rubin, A., & Rosebery, A. S. (1990). Teachers' misunderstandings in statistical reasoning; evidence from a field test of innovative materials. In A. Hawkins (Ed.), *Training teachers to teach Statistics* (pp. 72-89). Voorburg, The Netherlands: ISI.

Vallecillos, A. (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis doctoral. Universidad de Granada.

Vera, O. D., Díaz, C., & Batanero, C. (2011). Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología. *Unión, 27*. www.fisem.org/web/union