A Visualization Tool for One- and Two-Way Analysis of Variance

Rachel Sturm-Beiss  
Kingsborough Community College  
(City University of New York)

Journal of Statistics Education Volume 13, Number 1 (2005),  
ww2.amstat.org/publications/jse/v13n1/sturm-beiss.html

Copyright © 2005 by Rachel Sturm-Beiss, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: ANOVA; Java applet

Abstract

Analysis of variance (ANOVA), a technique included in many introductory statistics courses, analyzes the relationship between a quantitative dependent variable and one or more independent qualitative variables. The nature of the relationship is expressed in a model with unknown parameters. Many textbooks emphasize the mechanics of the technique while the model and its parameters remain abstractions or theoretical entities. We introduce a Java Applet that allows a student to profitably explore the features and factors of one and two-way ANOVA tables together with representational models and model parameters.

1. Introduction

ANOVA is a technique often applied to compare the means of several “treatment groups” (Hogg and Craig, 1995 and Neter, Wasserman and Kutner 1990). The study of cholesterol levels of treatment groups of patients receiving various cholesterol medications is a fair example. In cases where data of only two groups are available one may perform the intuitive t-test to determine if the means of the two groups are equal. If there are more than two treatment groups, then ANOVA may be applied. The dependence of cholesterol level on medication can be expressed in terms of a linear statistical model. The quantitative dependent variable, Y, represents cholesterol level and the qualitative independent variable has levels corresponding to the cholesterol medications that determine treatment groups. If other qualitative (factor) variables are introduced then combinations of levels determine treatment groups.

The ANOVA model is the simplest linear statistical model with qualitative independent variables. ANOVA, together with simple linear regression, form a foundation for the study of general linear
models. However, exercises relating ANOVA model parameters and calculated quantities are not easy to formulate. As a result, typical textbook exercises emphasize calculations, leaving the model as an abstract entity. Therefore, it is of pedagogical value to have ancillary materials that help students visualize model parameters and their relationship to sample observations. We present a tool (in the form of a java applet) that emphasizes the ANOVA probabilistic model by placing model parameters alongside observations, and giving the student the ability to manipulate values and to observe resulting effects, thus removing some of the abstraction. Taur (1999) introduced an excellent example of such a teaching tool for nonlinear regression, called “Visual Fit” and Anderson-Cook and Dorai-Raj (2003) reviewed such java applets that demonstrate the power of a test.

The ANOVA one-way approach here recommended starts with treatment group means $m_i$ for $i=1,2,3$ and variance $\sigma^2$ (the two-way model consists of two factor variables with three and two factor levels each). Random normal $N(m_i, \sigma^2)$ samples are generated within each treatment group and estimated sample means, variances and other quantities are calculated and displayed. We display data, actual parameter values, and estimated parameter values in a scatter plot augmented with graphs, ANOVA table, explanations, and exercises. The student can change parameter values (through standard window’s interface such as “drag-and-drop”, text boxes, list boxes etc.) and observe the effect on calculated quantities and model significance. Guided exercises and explanations help the student in this process. This visual teaching tool differs from traditional exercises in which sample observations (but not parameter values) are available. In particular, the student is able to generate many random samples for the same set of parameters and to get a feel for statistical significance as a phenomenon that emerges over a large number of random samples.

2. The ANOVA Visualization Tool

The ANOVA visualization tool demonstrates both one and two-way (fixed factor levels) ANOVA models. We limit our discussion to two-way models in order to avoid repetition. The two-way tool generates sample data consisting of $n$ observations per treatment group, where the size of $n$ is chosen by the student. There are six groups corresponding to various combinations of the three levels of factor variable $A$ and the two levels of factor variable $B$. Actual treatment group means $\mu_{ij}$, $i = 1, 2, 3$ and $j = 1, 2$ are either generated randomly within the fixed range: [-212,212], or are set by the student. The parameter $\mu$ is the overall mean of the $\mu_{ij}$, $i = 1, 2, 3$ and $j = 1, 2$ (note: the parameter $\mu$ is equal to $E(Y)$, the expected value of $Y$, only if each treatment group is equally represented within the $Y$ population). Our two-way ANOVA model is of the form:

$$E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

where

$Y_{ijk}$ = the $k^{th}$ observation in the $ij^{th}$ group

$\mu + \alpha_i + \beta_j + \alpha\beta_{ij} = \mu_{ij}$ = the mean of the treatment group corresponding to the $i^{th}$ level of $A$ and the $j^{th}$ level of $B$

$\alpha_i$, $i = 1, 2, 3$ are factor $A$ main effects
\( \beta_j, j = 1, 2 \) are factor B main effects

\( \alpha \beta_{ij}, i = 1, 2, 3 \) and \( j = 1, 2 \) are the AB interaction effects

We assume that factor variables A and B may interact. If there is no interaction, then the interaction effects are all 0, and the model is additive: \( \mu_y = \alpha_i + \beta_j \). The following figures and the comments below them illustrate some of the features of the ANOVA tool.

Figure 1. Two-way ANOVA

1. A parameter treatment group mean for the population corresponding to the \( i = 3^{rd} \) level of A and the \( j = 2^{nd} \) level of B, i.e. \( \mu_{32} \) (Drag and Drop point)
2. The value of \( \mu_{32} \)
3. The sample treatment group mean \( \bar{Y}_{32} \)
4. Observation \( Y_{12k} \) (the number of observations is set by the user)
5. A main effect of factor variable B, i.e. \( \beta_2 \) (Drag and Drop point)
6. A main effect of factor variable A, i.e. \( \alpha_2 \) (Drag and Drop point)
7. The parameter treatment group mean \( \mu_{31} \) (Drag and Drop point)
8. The overall mean \( \mu \)
9. The sum of squares bar graph
10. The text area for exercises and explanations
11. The ANOVA table
12. User controls:
   - Choose one or two-way model.
   - Set the value of sigma.
   - View explanations of various aspects of ANOVA.
   - Choose another sample.
   - Choose a guided exercise.
   - Choose the sample size.
   - Request an explanation of the two plots.
   - Display data values in the text area.

Figure 2. One-way ANOVA

1. A parameter treatment group mean for the population corresponding to the $i = 3^{rd}$ level of $A$, i.e. $\mu_3$ \textit{(Drag and Drop point)}

2. The value of $\mu_3$

3. The sample treatment group mean $\overline{Y}_3$

4. The overall mean $\mu$

5. A main effect of factor variable $A$, i.e. $\alpha_1$ \textit{(Drag and Drop point)}

6. Observation $Y_{1k}$ (the number of observations is set by the user)

7. The sum of squares bar graph
8. The text area for exercises and explanations
9. The ANOVA table
10. User controls:
   - Choose one or two-way model.
   - Set the value of sigma.
   - View explanations of various aspects of ANOVA.
   - Choose another sample.
   - Choose a guided exercise.
   - Choose the sample size.
   - Request an explanation of the two plots.
   - Display data values in the text area.

---

3. Suggested Exercises

The “Exercises” choice box in the ANOVA applet (see Figure 1 or Figure 2) has eighteen exercises (nine for one-way and nine for two-way) designed to guide the student’s exploration of ANOVA parameters and calculated quantities. Here is a sample of some of the exercises.

3.1 Compare population means to sample means - One Way Model:

Adjust the treatment group means (blue squares) so that they are aligned horizontally. Notice that the sample group means (gray rectangles) are not aligned horizontally even though they were drawn from populations with equal means. Press the “New Sample” button repeatedly to draw new samples from the same populations. Notice that the population means stay the same but the sample means change and that some samples have means that are not close approximations of the population means. Choose a larger sample size. Repeat the exercise. You will find that on the average the sample means are closer approximations when the sample size is large.

3.2 Demonstrate a type I error - One Way Model:

A type I error occurs when a true null hypothesis is rejected. The ANOVA null hypothesis states that treatment group means are equal. Align the treatment group means horizontally (so that they are all equal). As expected, the ANOVA table F-test will probably not be significant, indicating that there is not enough evidence to conclude that the treatment group means differ. Choose a sample size of 4. Press the “New Sample” button repeatedly until you get a significant F-test. Now, treatment group means are equal, however, the significant F-test incorrectly rejects the hypothesis stating that the means are equal. This is an example of a type I error: a true null hypothesis is rejected.

3.3 Estimate the probability of a type II error - One Way Model:

A type II error occurs when a false null hypothesis is not rejected. For one-way ANOVA the error occurs when the means are unequal however, the hypothesis of equal means is not rejected. Adjust the means (blue squares) so that they are close in value but not equal. Choose a sample size of 4. Estimate the probability of a type-two error by pressing the “New Sample” button repeatedly and calculate the percentage of insignificant F-tests. This percentage is the estimate of a type II error for the given treatment group means. Now move the means further apart and repeat the exercise. Now choose a larger sample size and repeat the exercise.
3.4 Show that when factor variables do not interact the two-way model is additive -
Two Way Model:

Align the line graph corresponding to the first level of B (connecting the dark purple means) and the line
graph corresponding to the second level of B (connecting the light purple means) so that they are
parallel. This is the case when there is no interaction between factor variables A and B. Now notice that
\[ \mu_q = \mu + \alpha_i + \beta_j, \quad i = 1, 2, 3 \text{ and } j = 1, 2. \]
All the values are displayed on the plot. The values \( \mu_q \), \( i = 1, 2, 3 \) and \( j = 1, 2 \) are the light gray numbers next to each treatment mean square. Thus, we see that
when there is no interaction we have an “additive model”.

4. Conclusion

We demonstrated this applet at a professional development seminar (CyberProf) given recently at City
University of New York (CUNY). The participants had varied science backgrounds. They expressed
interest in using the applet for their statistic students. Those who had some exposure to ANOVA were
pleased to have their knowledge of the technique enhanced by the demonstration.

Applets such as the one described here are tools that could easily be incorporated into lectures and
assignments as accessibility through the internet is simple and not costly. We believe that the ANOVA
and other related java applet visual teaching tools can help students of statistics and students from
related disciplines gain a better understanding of ANOVA and other statistical techniques.

The ANOVA Visualization Tool can be viewed by clicking on Anova Applet or by going to the author's
web site at www.kingsborough.edu/academicDepartments/math/faculty/rsturm/anova/Anova0126.html

References

Anderson-Cook, C. M. and Dorai-Raj, S (2003), “Making the Concepts of Power and Sample Size
Relevant and Accessible to Students in Introductory Statistics Courses using Applets,” Journal of
Statistics Education [Online] 11 (3) (ww2.amstat.org/publications/jse/v11n3/anderson-cook.html)

Hogg, R., and Craig, A. (1995), Introduction to Mathematical Statistics (5th ed.), New York:
Macmillan.

Neter, J., Wasserman, W., and Kutner, M.H. (1990), Applied Linear Statistical Models, Chicago:
Richard D. Irwin, Inc.

Taur, Y., and McCulloch, C. (1999), “A Teaching Tool for Nonlinear Regression: Visual Fit,” Journal
of Statistics Education [Online], 7 (2) (ww2.amstat.org/publications/jse/secure/v7n2/taur.cfm)

Rachel Sturm-Beiss
Department of Mathematics and Computer Science
Kingsborough Community College
City University of New York
2001 Oriental Boulevard

6 of 7
