# AnoGen: A Program for Generating ANOVA Data Sets

# Version 1.3

Jeff Miller
Department of Psychology
University of Otago
Dunedin, New Zealand

August, 1998

# 1   Introduction

This program was designed for use in teaching the statistical procedure known as *Analysis of Variance* (ANOVA). In brief, it generates appropriate data sets for use as examples or practice problems, and it computes the correct ANOVA for each data set. It handles between-subjects, within-subjects, and mixed designs, and can go up to six factors, with some restrictions on the numbers of levels, subjects, and model terms.

The program can be run in either of two modes: one designed for use by students; the other, by teachers.

The student mode is simpler: The student simply specifies the experimental design, and AnoGen generates an appropriate random data set. The student can then view the data set and answers, and save them to a file. Thus, students can fairly simply get as much computational practice as they want.

The teacher mode is more complicated: The teacher not only specifies the experimental design, but also controls the the cell means and error variance to obtain whatever $F$ values are desired for the example. Considerable familiarity with ANOVA is needed to use this mode.

# 2   Step-by-Step Instructions: Student Mode

1. Start the program as is appropriate on your computer system (e.g., by typing AnoGen at a DOS prompt).

2. Once AnoGen is running, type S to enter the student mode.

3. Specify the design:

   (a) To set the number of within-subjects factors, type W, and then type the number you want, followed by enter.

   (b) Similarly, type B to set the number of between-subjects factors,

   (c) Similarly, type S to set the number of subjects per group. A "group" is defined by one combination of levels of the between-subjects factors. For example, if you have between-subjects factors of Male/Female and Young/Old, then there are four groups corresonding to the four combinations.

   Note that you can set these numbers in any order, and you can change each one as often as you like. After you have the settings you want, type ctrl-Q to move on to the next step.

4. Now specify the number of levels of each factor. Type the letter corresponding to the factor you want to change (A, B, . . . ), and then enter the number of levels you want. Again, after you have the settings as you want them, type ctrl-Q to move on to the next step.

5. Type P to display the problem (i.e., the data set). Ideally, you would now do the computations by hand, for practice. (The information given in the problem display is intended to be self-explanatory, but some explanation is given in Section 3.)

Table 1: An example of a problem display. This design has two between-subjects factors (A and B) with two levels each, and three subjects per group.

| | |
|---|---|
| | Group A1B1: |
| | Sub 1:     95 |
| | Sub 2:     78 |
| | Sub 3:     97 |
| | |
| | Group A2B1: |
| | Sub 1:     -19 |
| | Sub 2:     -37 |
| | Sub 3:     -10 |
| | |
| | Group A1B2: |
| | Sub 1:     55 |
| | Sub 2:     64 |
| | Sub 3:     73 |
| | |
| | Group A2B2: |
| | Sub 1:     58 |
| | Sub 2:     63 |
| | Sub 3:     71 |

6. Type S to display the solution (i.e., cell means, ANOVA table, etc). This is where you check your solution and make sure you've done it correctly. The solution contains the various parts that I use in teaching ANOVA using the general linear model. (More explanation of the information given in the solution display is given in Section 4.)

7. If you want, type F to save the problem and solution to a file. (The main reason to for doing this is to get a printed version of the problem and solution.) Enter the name of the file to which you want the information saved.

8. Type ctrl-Q to quit when you are done with this problem. AnoGen will then ask if you want to start over again: Type Y if you want to do another problem, or N to quit.

## 3   Explanation of Problem Display

Table 1 shows an example of a problem display. There is one line per subject, and the different groups correspond to the different levels of the between-subjects factor(s). For this example, the problem display fits on a single screen; with larger designs (i.e., more groups or more subjects per group), the problem display may be split across several screens.

Table 2 shows an example of a problem display for a more complex experimental design. Note that the different conditions tested within-subjects are listed across the line, and the different subjects and groups organized as in the between-subjects design.

Table 2: An example of a problem display. This design has a within-subjects factor (A) with two levels, two between-subjects factors (B and C) with two levels each, and three subjects per group.

```
                 Group B1C1:
                         A1    A2
                 Sub 1:   77    53
                 Sub 2:   84    56
                 Sub 3:  103    41

                 Group B2C1:
                         A1    A2
                 Sub 1:   77    65
                 Sub 2:   54    64
                 Sub 3:   73    69

                 Group B1C2:
                         A1    A2
                 Sub 1:  103    75
                 Sub 2:  100    78
                 Sub 3:   97    57

                 Group B2C2:
                         A1    A2
                 Sub 1:   72    10
                 Sub 2:   74    18
                 Sub 3:   58     2
```

# 4 Explanation of Solution Display

The solution display has several components, as described below. Some of these components may be omitted if they do not fit well with the way your instructor teaches the material.

## 4.1 Design

This shows a list of factors with the number of levels per factor. Also shown is the number of subjects per group.

## 4.2 Cell Means

The cell means are given in a table of this form (these are the means for the problem in Table 2):

| Cell: | Mean |
|-------|------|
| u | 65 |
| A1 | 81 |
| A2 | 49 |
| B1 | 77 |
| B2 | 53 |
| A1 B1 | 94 |
| A1 B2 | 68 |
| A2 B1 | 60 |
| A2 B2 | 38 |
| C1 | 68 |
| C2 | 62 |
| A1 C1 | 78 |
| … | |

The first line (u) shows the overall mean across all conditions. The next two lines (A1 and A2) show the means of all scores at levels 1 and 2 of factor A, respectively. The next two lines (B1 and B2) show the means of all scores at levels 1 and 2 of factor B, respectively. The next line (A1 B1) shows the mean of all scores at level 1 of factor A and level 1 of factor B, and then the next three lines show means for the other combinations of levels on these two factors. And so on.

## 4.3 Model

The model section shows the form of the general linear model appropriate for this design. The main effect and interaction terms are denoted by capital letters (A, B, AB, etc), S is for subjects, and the subscripts are denoted by lower-case letters (i, j, k, etc).

## 4.4 Estimation Equations

The estimation equations section shows the equation used to estimate each term in the linear model. The period subscript is used to denote averaging across levels of the factor corresponding to that subscript.

## 4.5 Decomposition Matrix

The decomposition matrix shows the breakdown of all data values (numbers of the left sides of the equals signs) into the estimated values corresponding to each term in the linear model. The order of the numbers on the line is the same as the order of the terms in the model.

## 4.6 ANOVA Table

The ANOVA table is in a relatively standard format. The $F$ value is marked with one asterisk if it is significant at the level of $p < .05$ and two asterisks if significant at $p < .01$. The error term used to compute each $F$ is shown at the far right side of the table.