

# Data Analysis Exercises for Chapter 8: *Applied Regression Analysis, Generalized Linear Models, and Related Methods, Third Edition* (Sage, 2016)

John Fox

Last modified: 2015-01-29

**Exercise D8.1** Each of the following data sets contains at least one quantitative variable that can be treated as a response variable and at least one categorical variable that can be treated as an explanatory variable (i.e., a factor):

<i>Data Set</i>	<i>Suggested Response Variable</i>
<code>Angell.txt</code>	moral integration
<code>Duncan.txt</code>	prestige
<code>Leinhardt.txt</code>	infant mortality
<code>Ornstein.txt</code>	number of interlocks
<code>Robey.txt</code>	total fertility rate
<code>UnitedNations.txt</code>	total fertility rate, or expectation of life for males or females

Alternatively, select an appropriate data set of interest to you.

- (a) Find the mean and standard deviation for the response in each level of the factor. Draw parallel boxplots of the response by the factor. Comment on the results.
- (b) Confirm that identical sums of squares are produced by the following three computational methods:
  - (i) Set  $\alpha_m = 0$  in the analysis-of-variance model; use the 0/1 coding scheme of Chapter 7 to produce dummy regressors; and fit the model by least-squares regression.
  - (ii) Set  $\alpha_m = -\sum_{i=1}^{m-1} \alpha_i$ ; use the 1, 0, -1 coding scheme of the current chapter to produce deviation regressors; and fit the model to the data by least-squares regression.
  - (iii) Use the formulas in Table 8.1 to calculate the sums of squares directly.
- (c) Having obtained the regression and residual sums of squares, summarize this information in an ANOVA table, testing the null hypothesis of no differences in the population mean response across the levels of the factor. What can you conclude from this test?

**Exercise D8.2** The following “balanced” (i.e., equal-cell-frequencies) data are from an experiment reported by Fox and Guyer (1978). Twenty four-person groups of subjects played 30 trials of a “prisoner’s dilemma” game. On every trial of the experiment, each subject selected either a competitive or a cooperative choice. The value reported below for each group is the number of cooperative choices (of the 120 choices) made by subjects in that group. The data are also in `Guyer.txt`. Ten of the groups recorded their choices anonymously, while the remaining 10 groups made public choices (i.e., subjects’ choices were made known to other group members);

half of the groups were composed of males and half of females. The experimenters expected to observe a higher level of cooperation in the public-choice condition, but did not make predictions about sex effects or sex-by-condition interaction.

<i>Condition</i>	<i>Sex</i>	
	Male	Female
Public Choice	49	54
	64	61
	37	79
	52	64
	68	29
Anonymous	27	40
	58	39
	52	44
	41	34
	30	44

- (a) Make a scatterplot of cooperation versus sex, using different colors or symbols to plot the points for the public-choice and anonymous conditions. Calculate the mean and standard deviation for cooperation in each cell, and graph the cell means. Comment on the descriptive results of the experiment: How do sex and condition appear to affect cooperation?
- (b) Using the usual sigma constraints, form deviation regressors for condition, sex, and the condition  $\times$  sex interaction. Show that the correlations among these regressors are all zero (by virtue of the fact that there are equal numbers of observations in all of the cells of the design).
- (c) Confirm that the sums of squares for the ANOVA can be obtained in the following four ways:
  - (i) As  $SS(\alpha|\beta)$ ,  $SS(\beta|\alpha)$ ,  $SS(\gamma|\alpha, \beta)$ .
  - (ii) As  $SS(\alpha|\beta, \gamma)$ ,  $SS(\beta|\alpha, \gamma)$ ,  $SS(\gamma|\alpha, \beta)$ .
  - (iii) As  $SS(\alpha)$ ,  $SS(\beta)$ ,  $SS(\gamma)$ . Note: This third approach is valid only for balanced data.
  - (iv) Using the special formulas given in Section 8.2.5.
- (d) Form the ANOVA table for the experiment. What conclusions can you draw from the ANOVA about the effects of condition and sex on cooperation?

**Exercise D8.3** The data in the table below (and in `Adler.txt`) are from a social-psychological experiment, reported by Adler (1973), on “experimenter effects” in social research — that is, how researchers’ expectations can influence the data that they collect. Adler recruited research assistants, who showed photographs of individuals’ faces to respondents; the respondents were asked by the research assistants to rate the apparent “successfulness” of the individuals in the photographs. In fact, Adler chose photographs that were average in their appearance of success, and the true subjects in the study were the research assistants.

Adler manipulated two factors, named *expectation* and *instruction* in the data set:

*expectation*. Some assistants were told to expect high ratings, while others were told to expect low ratings.

*instruction*. In addition, the assistants were given different instructions about how to collect data: Some were instructed to try to collect “good” data; others were instructed to try to collect “scientific” data; and still others were given no special instruction of this type.

Adler randomly assigned 18 respondents to each of the six resulting experimental conditions — combinations of the two categories of the factor *expectation* (*high* or *low*) and the three categories of the factor *instruction* (*good*, *scientific*, or *none*). I deleted some of these observations at random to produce the unbalanced dataset in the following table:

<i>Instruction</i>	<i>Expectation</i>					
	High			Low		
Good	25	0	-16	-25	-20	-2
	5	11	-6	-23	-24	12
	42	-2	-13	-28	-24	-8
	14	4		-22	-22	-17
	19	6		-22	-23	-30
	-22	-6		-10	-19	
Scientific	-19	5	-13	6	-22	-5
	-24	-1	-1	-5	7	
	-4	-9	-3	14	14	
	-24	-5	-11	-11	15	
	0	-6	-6	14	-6	
	-4	4	-4	-5	9	
None	-26	-21	-10	-12	-4	20
	-1	-19	-37	-4	-10	9
	22	-12	0	13	-3	-8
	3	9	-10	-27	-11	8
	-26	-9		-7	2	-6
	4	-27		-20	-9	6

- Treating *instruction* as the row factor and *expectation* as the column factor, calculate the mean and standard deviation of scores in each cell. Count the number of observations in each cell. Graph the cell means and comment on the apparent results of the experiment.
- Construct a two-way analysis-of-variance table for Adler's data, testing hypotheses about the main effects and interactions of *instruction* and *expectation*. What conclusions would you draw from this ANOVA?

**Exercise D8.4** The following table shows the results of an experiment on interpersonal attraction reported by Riordan, Quigley-Fernandez, and Tedeschi (1982). Subjects in the study interacted with an experimenters' confederate whose attitudes were manipulated to be either similar or dissimilar to those of the subjects. In the course of the study, it was arranged for some subjects to request the confederate's help in completing a task, while other subjects did not ask for help. Finally, the confederate provided help to some subjects but not to others. These three factors combine to produce eight experimental conditions. Nine subjects were assigned to each condition. (Actually, two conditions contained eight subjects, but I will disregard this slight complication.) At the beginning and again at the end of the study, subjects rated their attraction to the confederate on a two-item scale, with possible composite scores ranging from 2 through 14. The table reports means and standard deviations for changes in attraction over the course of the study.

Attitude Similarity	Help Requested	Help Provided	Attraction Change	
			Mean	Standard Deviation
Similar	Yes	Yes	0.22	1.2
		No	-4.62	1.3
	No	Yes	-0.12	1.2
		No	-2.56	1.4
Dissimilar	Yes	Yes	1.44	1.9
		No	-2.22	2.3
	No	Yes	2.10	2.1
		No	-1.98	1.3

Riordan et al. (1982: 364) make the following predictions with regard to these changes:

*Changes in attraction should be associated with an interaction of all three factors. Subjects should show an increment in attraction when help is provided by a dissimilar other and the increment should be greater when the help was not requested than when it was requested. A decrement in attraction should occur when help is not given by a similar other and the decrement should be greater when help was requested. No change should occur when a dissimilar other does not provide help and it was not requested, and very little change in a negative direction when the dissimilar other does not provide help and it was requested. No change is also predicted when a similar other provides help and it was requested, and very little change in a positive direction when it was not requested.*

- (a) Graph the cell means from the table. Do the results of the study appear to confirm the authors' predictions?
- (b) Using the results for three-way balanced ANOVA derived in Exercise 8.7, compute an analysis-of-variance table for Riordan et al.'s data. What conclusions would you draw?

**Exercise D8.5** Pecknold et al. (1982) describe an experiment in which psychiatric patients suffering from anxiety were assigned randomly to one of three groups: (i) a treatment group that received a standard anti-anxiety drug (diazepam); (ii) a treatment group that received a new anti-anxiety drug (fenobam); and (iii) a control group that received a placebo. At several points in the study, subjects were administered standard tests for anxiety, including the Hamilton Rating Scale; high scores on the scale are indicative of severe anxiety. After three weeks of treatment, the experiment produced the following results:

	Group		
	Diazepam	Fenobam	Placebo
Mean Hamilton Score	16.00	15.11	23.54
Standard Deviation	8.33	7.14	12.30
Number of Subjects	9	9	11

Use the method derived in Exercise 8.10 to test the null hypotheses that (i) the average anxiety score of patients receiving the drugs is no different from that of patients receiving the placebo; and (ii) the average anxiety score of patients receiving the new drug is the same as that of patients receiving diazepam. Note that the residual mean square can be calculated according to the formula

$$S_E^2 = \frac{\sum_{j=1}^m S_j^2(n_j - 1)}{n - m}$$

where  $S_j^2$  is the variance within group  $j$ . \*Why does this formula work?