

Homework Assignment 8: Fair's Affairs

36-402, Advanced Data Analysis

Due at the start of class, 29 March 2011

In 1969, the magazine *Psychology Today* did a survey of its readers that included questions about (among other things) how often the respondents had had extra-marital sex in the previous twelve months. In 1978 the economist Ray C. Fair used this data to develop a “theory of extramarital affairs”¹, with the idea that people optimize a trade-off between working, spending time with their spouse, and spending time with a “paramour”. The model and data have become very well known (there are at least a hundred later papers and books which reference it), and is available as **Affairs** in the package **AER** on CRAN.

The variable **affairs** records the answer to “How often did you engage in extramarital sexual intercourse during the past year”, with values of “once a month”, or more frequently, all coded as 12. Other variables are sex, age, how many years the respondent had been married², whether they had children, how religious they were (on a scale of 1–5), their level of education, how much prestige their occupation had (on a scale of 1–7), and how happy they were with their marriage (on a scale of 1–5).

1. (30 points) *Two specifications*
 - (a) (15 points) Using logistic regression, fit a model for the *number* of times respondents said they had extramarital sex during the previous year. Describe, in words, the predictions of the model. Which variables are significant predictors?
 - (b) (15 points) Repeat (1a), but use logistic regression to fit a model for *whether* respondents said they had extramarital sex at all during the previous year.
2. (10 points) Are the same variables significant in both models in problem 1? Do they have the same signs in both models? Should the models match in this way? Explain.
3. (20 points) *Comparing predictions*

¹*Journal of Political Economy* **86** (1978): 45–61; a reprint is available from Prof. Fair's website, <http://fairmodel.econ.yale.edu/rayfair/pdf/1978A200.PDF>. This paper also used a similar survey of readers of *Redbook* in 1974, not part of this data set.

²Prof. Fair removed respondents who had never married, or had married more than once.

- (a) (5 points) For each person in the data set, calculate the predicted probability, under both models, that they did *not* have an affair.
 - (b) (10 points) Plot these against each other. Describe the plot in words.
 - (c) (5 points) Do the models agree with each other in their predictions? Should they? Explain.
4. (20 points) *Calibration*
- (a) (2 points) Consider all the people for whom the predicted probability of an affair, according to the model from problem (1a), is less than 10%. What fraction of them report having affairs?
 - (b) (3 points) Repeat this calculation for predicted probabilities between 10% and 20%, 20% and 30%, etc. Plot the actual frequencies against the predicted probabilities.
 - (c) (10 points) Make similar plots for the other model. (You can combine the plots, if the result is clear.)
 - (d) (5 points) For which model do the predictions seem to match the data best? Explain with reference to your plots.
5. (10 points) Download Fair's paper and read Table I (p. 53). Does it make sense to use a linear response for all of the variables (as in problem 1 above), or should some variables be treated as categorical? Explain.
6. (10 points) *Evaluation*
- (a) (5 points) Do either of these models seem to provide an adequate description of the data? (Explain.) If not, what else could one try?
 - (b) (5 points) Is it reasonable to use this data to develop theories about contemporary behavior? Explain.