

STA 2201S Assignment 1.

due Friday, February 10 at the beginning of class

When answering questions requiring numerical work, the results are to be reported in a narrative summary, in your own words. Tables and Figures may be included, but must be formatted along with the text. DO NOT include in this summary printouts of computer code with the relevant selections highlighted. All computer code used to obtain the results summarized in the response should be provided as an appendix. In this appendix you may highlight the relevant results.

1. Exercise 9.1.1, Davison (p.425)
2. Exercise 9.2.6, Davison (p. 438)
3. Exercise 9.4.3, Davison (p.462)
4. Problem 9.6.7, Davison (p. 466)
5. Exercises 10.3.1 and 10.2.2, Davison (p. 486 and 479)
6. The data file for the “incentives to publish” paper by Franzoni et al.¹ is posted on the course web page. You can read this into R by using the command `read.table("http://www.utstat.utoronto.ca/reid/sta2201s/incentives.data");` the file `incentives.csv`, as well as Franzoni et al.’s original spreadsheet, are also posted, for those of you who are using SAS or SPSS or ... The original article and supplementary material are also on the web page.

- (a) One of the authors’ analyses is a linear regression using the logarithm of the acceptance rate as the response variable, defined as:

$$accrate_{it} = \log \left(\frac{accepted_{it} + 1}{submitted_{it}} \right),$$

where i indexes countries ($i = 1, \dots, 30$), t indexes time (either year, or year–1, depending on the variable). Write the mathematical model implied by their Specifications I, II, and III as described in the supporting online material, carefully defining and describing the variables and parameters you use.

- (b) Fit these three models to the data, give a table of estimated parameters from your models in (a), and give an analysis of variance table that includes separate sums of squares for the the relevant explanatory variables.
- (c) Provide two or three plots that serve to check the assumptions of the models, and summarize your conclusions about the use of the linear regression model in non-technical language.

¹*Science* August 2011, p. 702