

### Exercise Sheet 4 — Linear Model

**Problem 1. Discrimination in Salaries** The data set `salaries.csv` contains data on the salaries of 52 tenure-track faculty in a small college (Weisberg, 1985). The question is to assess whether women professors are paid less than their male counterparts in comparable positions. The variables are

`sex` Gender

`rank` Rank coded as 1 for assistant professor, 2 for associate professor and 3 for full professor

`years` Number of years at current rank

`degree` Highest degree. 1 for doctorate, 0 for masters

`years_since_degree` Number of years since highest degree was earned

`salary` Academic year salary, in dollars

- (a) Fit a linear model with `salary` as response and gender as explanatory variables. Is there evidence of gender discrimination?
- (b) Display a “residuals versus fitted values” plot. Interpret.
- (c) For answering the question we need to control for differences in qualifications and experience. We could adjust for either rank, years at rank and its interaction, or for highest degree, years since the degree was earned and its interaction. Fit these two models and comment. Do we need the interaction terms? If there is also discrimination in promotion, which model is likely to be misleading?
- (d) Perform regression diagnostics on the two models of question c).
- (e) A transformation of the response  $g(y)$  is sometimes a good remedy for fixing problems with the residuals. Regress the log of salary on the variable gender. Describe the gender effect on *salary* (not log-salary.)

**Problem 2.** The file `df.csv` contains a simulated data set with variables

`y` response variable (continuous)

`x` an explanatory variable (continuous)

`treatment` an explanatory variable (ordinal with levels 0, 1, 2)

Fit a linear model including `x` and `treatment` as explanatory variables. Plot the residuals vs fitted values as colour the points according to `treatment`. What does the graphic suggests?