Homework Problems on Linear Models Using Matrices

Soc. 761

Fall 2014

The problems in this assignment use the data frame `Prestige` (with data on the rated prestige and other characteristics of about 100 Canadian occupations) in the `car` package. To access the data, first load the package with the `library(car)` command. Once you load the package, entering `help(Prestige)` or `?Prestige` at the R command prompt will give you detailed information about the data set.

The `Prestige` data set contains some missing data (in the variable `type`). To remove observations with missing data, you can, for example, enter the command `Prestige.2 <- na.omit(Prestige)`.

1. Using matrix operations, compute the linear least squares regression of `prestige` on the quantitative explanatory variables `income`, `education`, and `women`.

2. Find the estimated error variance, the estimated covariance matrix of the coefficients, and the standard errors of the coefficients for the model fit in question 1. Compute t-tests for the null hypotheses that each of the coefficients of `income`, `education`, and `women` are 0.

3. Verify that the results that you obtained in questions 1 and 2 are correct by fitting the same regression model with the `lm` function:

```
mod <- lm(prestige ~ income + education + women, data=Prestige.2)
summary(mod)
```

4. (optional) Treating blue-collar occupations as the baseline category, construct dummy regressors for the professional and white-collar categories of the `type` factor [e.g., `prof <- as.numeric(Prestige$type == "prof")` and `wc <- as.numeric(Prestige$type == "wc")`]. Then, using matrix operations, regress `prestige` on `income`, `education`, `women`, and the dummy regressors for `type`, finding the estimated coefficients, error variance, coefficient covariance matrix, and coefficient standard errors for the model. (Do not include interactions in the model.) As before, use `lm` to verify that the results that you obtained are correct; to add `type` to the model, you can either respecify the model from scratch or use the `update()` function: `mod.2 <- update(mod, . ~. + type)`.

5. (optional) Using matrix operations, test the linear hypothesis that the coefficients of the dummy regressors for `type` in the model fit in question 4 are both 0. Verify that the F-test that you obtain is correct by comparing your result with the output from the `Anova()` function [i.e., `Anova(mod.2)`] — the two F-tests should be the same.