

Assignment 01

Matrix Algebra and Regression

This goal of this assignment is to have you use matrix algebra and computation to compute values from a standard regression output. In the spirit of this, use matrix algebra rather than the `lm()` function to compute your answers. (Note: You can use the `lm()` function to check your responses, so you should know whether many of your responses are correct.)

When we ask you to *show your work*, you can either show your work like you would in a high-school mathematics class—although we ask that you typeset all mathematics using an equation editor—or you can provide the relevant part of your script file. Please submit your responses to each of the questions below in a printed document. Also, please adhere to the following guidelines for further formatting your assignment:

- All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Equation Editor, Markdown, or L^AT_EX.
- All syntax included should be typeset in a monospaced font, appropriately commented and follow the Data Camp Style Guide (<http://docs.datacamp.com/teach/style-guide.html>).

This assignment is worth 20 points. Each question is worth 1 point unless otherwise noted.

Use the following data to answer the questions below.

```
##    wage age sex
##  12.00  32  M
##   8.00  33  F
##  16.26  32  M
##  13.65  33  M
##   8.50  26  M
```

Main-Effects Model

In this section, all questions will refer to the regression model that includes an intercept, and the main-effects of `age` and `sex` to predict `wage`.

1. Report the design matrix (i.e., the X-matrix) for the model.
2. What are the dimensions of the design matrix?
3. Compute and report the **b** vector (i.e., the vector of the regression coefficients). Show any relevant work.
4. Compute and report the standard errors for each of the regression coefficients in the model. Show any relevant work.

5. Using the values from Questions 3 and 4, compute and report the t -statistic for each of the regression coefficients. Show any relevant work. You may need to refresh your memory about what how a t -value is computed. One place to start may be your introductory statistics textbook, or any of a number of websites online.
6. Use the `pt()` function to compute the p -value (two-sided) for each of the regression coefficients. Show any relevant work. Again, you may need to refresh your memory about what a p -value is, and how they are computed.
7. Write a short narrative, meant for publication, that reports the regression results. This narrative should inform readers of the model that was fitted, as well as provide information about each of the regression coefficients, standard errors, t -statistics, and p -values. In addition it should include interpretations for any of the effects that are statistically reliable. **(3pts)**

Interaction Model

In this section, all questions will refer to the regression model that includes an intercept, the main-effects of `age` and `sex`, and the interaction effect between `age` and `sex` to predict `wage`.

8. Write out the design matrix for the model.
9. What are the dimensions of the design matrix?
10. Try to compute the `b` vector. You get an error message saying: `Error in solve.default(t(X) %*% X) : system is computationally singular`. Explain, using the language of matrix algebra, what this error message means.
11. Fit this model using the `lm()` function. Report the regression summary output for the coefficients. Note that the same problem exists when using the `lm()` function. However, the model does produce output for the main-effects. Which model was fitted to produce these values?

Quadratic Model

In this section, all questions will refer to the regression model that includes an intercept, as well as, the linear and quadratic effect of `age` to predict `wage`.

12. Write out the design matrix for the model.
13. What are the dimensions of the design matrix?
14. Compute and report the variance–covariance matrix for the regression coefficients. Show any relevant work. **(2pts)**
15. Compute and report the correlation matrix for the regression coefficients. Show any relevant work. (Hint: Compute each correlation individually. Also, since the variance–covariance matrix is larger than 2x2, be careful that you are using the appropriate elements in each computation.) **(2pts)**
16. What does the correlation matrix for the coefficients suggest about independence of the coefficients? Explain.