INFX574 lab 2

Your name:

Deadline: Wed, Apr 11, 11:59pm

Introduction

Please submit the completed lab by end of the day (April 4th, 2018). You should submit a) your code (the notebook or whatever you are using) and b) the lab in a final output form (html or pdf).

Note: you may want to do some of it on paper instead of computer. You are welcome to do it but please include the result as an image into your final file.

Working together is fun and useful but you have to submit your own work. Discussing the solutions and problems with your classmates is all right but do not copy-paste their solution! Please list all your collaborators below:

1.
2. ...

1 LPM for Titanic Survival

Your task is to predict survival in the Titanic's sinking in 1912 using LPM (= linear probability model, i.e. linear regression) and python. Let's ignore the issues with training/testing and overfitting here. Just use the full data.

1. Load the data and ensure you know the meaning and coding scheme of the main variables.
2. Estimate a LPM model in the form
$$\text{surved}_i = \beta_0 + \beta_1 \cdot \text{pclass}_i + \beta_2 \text{sex}_i + \epsilon_i$$
3. Use the model to predict the survival
$$\widehat{\text{survived}}^* = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{pclass}_i + \hat{\beta}_2 \cdot \text{sex}_i$$
4. Analyze your predicted survival values. What are the maximum and minimum of these? Is any of these equal to 0 or 1?
5. Now re-compute the survival in the form
$$\widehat{\text{survived}}_i = \mathbb{1}(\widehat{\text{survived}}_i > 0.5)$$
6. Compare the actual and predicted survival rates. What is the accuracy?

Next, let's play with interaction.

7. Estimate a similar LPM, but now allow the passenger class effect to differ between men and women.
8. Interpret the coefficients
9. Compute the accuracy of your predictions as above. How much did improve by allowing for different gender effects?

2 Extra Task (not graded)

When arranging people into the lifeboats, the captain gave the order "women and children first". Hence we expect both children and women to have noticeably higher survival rate than men. Can you find the age boundary of "children" that the ship officers were using when deciding who has a prority to get into lifeboats?