
COURSE: MATH 50 DARTMOUTH COLLEGE (MWF 11:15 AM-12:20 PM), FALL 2015

INSTRUCTOR: NISHANT MALIK

IN CLASS EXAM : 1

DATE AND TIME: 10/07/2015; 11:15 AM - 12:20 PM (65 MIN)

GRADES: ACCOUNTS FOR 15% OF THE TOTAL GRADES

Directions:

1. All the problems below are to be solved using IPython Notebook (Jupyter).
 2. All the problems should be solved in one single IPython Notebook file.
 3. The output file should be named using the convention exam1<your full name> and it should be submitted as html or pdf, use <https://dropitto.me/m50f15> to upload it.
 5. Use of emails or any other form of internet communication during the exam is NOT allowed.
-

1. A criminologist studying the relationship between level of education and crime rate in medium sized U.S. counties collected the following data for a random sample of 84 counties; column 2 is the percentage of individuals in the county having at least a high-school diploma, and column 1 is the crime rate (crimes reported per 100,000 residents).

Data source: <https://netfiles.umn.edu/users/nacht001/www/nachtsheim/Kutner/Chapter%20%201%20Data%20Sets/CH01PR28.txt>

*This data can also be accessed using the function `read_tb_data(ch,pr)` with `ch=1` and `pr=28`.

(a) Test whether or not there is a linear association between crime rate and percentage of high school graduates, using a t test with $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

(b) Estimate β_1 , with a 99 percent confidence interval. Interpret your interval estimate.

[Points: 10]

2. Consider a normal error regression model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$. Where ϵ_i are independent and distributed as $N(0,1)$, $\beta_0 = 0.2$, $\beta_1 = 0.95$, $i = 1, \dots, 5$ and in these 5 trials X_i takes the values $\{5, 10, 15, 20, 25\}$. You are given that for every trial i there are k different observations of Y_i .

(a) Generate a figure in ipython notebook (Jupyter) with following two subplots:

(i) A scatter plot for $k = 10$ between X and Y .

(ii) A box plot for every trial i when $k = 25$.

(b) Let say a maximum likelihood estimate for the above regression line is given by $\hat{Y}_i = 0.25 + 0.9X_i$, for $k = 10$. Obtain the numerical value of the maximum likelihood estimate of the variance for this case.

[Points: 5]