Final Exam: Math 50 Dartmouth College (MWF 11:15 AM-12:20 PM), Fall 2015
Instructor: Nishant Malik
Date and Time: 11/20/2015; 8:00 AM - 11:00 AM (3 hrs)
Grades: Accounts for 35% of the total grades

Directions:
1. All the problems below are to be solved using one single IPython Notebook (Jupyter).
2. Create a new folder finalexam<your full name> and do all your final exam related work inside this new folder. If you have a functions file then please copy this file into this folder. You should name your IPython notebook (Jupyter) file finalexam<your full name>.ipynb. After you finish the exam convert your finalexam<your full name>.ipynb into html or pdf, and zip your folder finalexam<your full name> with all the files inside it. Upload this zipped folder to https://dropitto.me/m50f15.
3. Please do not use internet for any other purpose except for downloading the data.

---

1. Airfreight breakage. A substance used in biological and medical research is shipped by air- freight to users in cartons of 1,000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route (X) and the number of ampules found to be broken upon arrival (Y). Assume first-order linear regression model is appropriate.

Data source:
`https://netfiles.umn.edu/users/nacht001/www/nachtsheim/Kutner/Chapter%20%201%20Data%20Sets/CH01PR21.txt`
* *This data can also be accessed using the function* read_tb_data(ch,pr) *with ch=1 and pr=21.*
* *Column 1 is Y and column 2 is X*

(a) Will b_0 and b_1 tend to err in the same direction or in opposite directions here?

(b) Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 99 percent family confidence coefficient. Interpret your confidence intervals.

(c) It is desired to obtain interval estimates of the mean number of broken ampules when there are 0, 1, and 2 transfers for a shipment, using a 95 percent family confidence coefficient. Obtain the desired confidence intervals, using the Working-Hotelling procedure.

(d) The next three shipments will make 0, 1, and 2 transfers, respectively. Obtain prediction intervals for the number of broken ampules for each of these three shipments, using the Scheffé procedure and a 95 percent family confidence coefficient.

(e) Would the Bonferroni procedure have been more efficient in developing the prediction intervals in part (d)? Explain.

[Points: 10]

2. Bid preparation. A building construction consultant studied the relationship between cost of bid preparation (Y_1,) and amount of bid (Y_2) for the consulting firm's clients. In a sample of 103 bids prepared by clients, $r_{12} = .87$. Assume that bivariate normal model applies.

(a) Test whether or not $\rho_{12} = 0$; control the risk of Type I error at .10. State the alternatives, decision rule, and conclusion. What would be the implication if $\rho_{12} = 0$?

(b) Obtain a 90 percent confidence interval for ρ_{12} Interpret this interval estimate.

(c) Convert the confidence interval in part (b) to a 90 percent confidence interval for ρ_{12}^2

[Points: 5]

---

3. Typographical errors. Shown below are the number of galleys for a manuscript (X) and the total dollar cost of correcting typographical errors (Y) in a random sample of recent orders handled by a firm specializing in technical manuscripts. Since Y involves variable costs only, an analyst wished to determine whether regression-through-the-origin model is appropriate for studying the relation between the two variables.

i	1	2	3	4	5	6	7	8	9	10	11	12
X_i	7	12	10	10	14	25	30	25	18	10	4	6
Y_i	128	213	191	178	250	446	540	457	324	177	75	107

(a) Fit regression-through-the-origin model and state the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function through the origin appear to provide a good fit here? Comment.

(b) In estimating costs of handling prospective orders, management has used a standard of \$17.50 per galley for the cost of correcting typographical errors. Test whether or not this standard should be revised; use $\alpha = .02$. State the alternatives, decision rule, and conclusion.

(c) Obtain a prediction interval for the correction cost on a forthcoming job involving 10 galleys. Use a confidence coefficient of 98 percent.

(d) Obtain the residuals e_i. Do they sum to zero? Plot the residuals against the fitted values $\hat{Y}_i$. What conclusions can be drawn from your plot?

(e) Conduct a formal test for lack of fit of linear regression through the origin; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

[Points: 10]

4. Sales growth. A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where X is the year (coded) and Y is sales in thousands of units:

i	1	2	3	4	5	6	7	8	9	10
X_i	0	1	2	3	4	5	6	7	8	9
Y_i	98	135	162	178	221	232	283	300	374	395

(a) Prepare a scatter plot of the data. Use *lowess* method to determine whether a linear relation is adequate here?

(b) Use the Box-Cox procedure to find an appropriate power transformation of Y. Evaluate SSE for $\lambda = .3, .4, .5, .6, .7$. What transformation of Y is suggested?

(c) Use the transformation $Y' = \sqrt{Y}$ and obtain the estimated linear regression function for the transformed data. Plot the estimated regression line and the transformed data. Does the regression line appear to be a good fit to the transformed data?

(d) Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plots show?

(e) Express the estimated regression function in the original units.

[Points: 10]