Statistics 512: Problem Set 4
Due Thursday, March 8, 2018 11:59 PM

Important Note – *Every graph or plot you create should have your name printed as a subtitle. Consequently, any graph with no name will result in a* **20% points loss**. *Also, please attach your code at the end; any homework with no code provided will result in a* **50% points loss** *on the entire assignment.*

1. Consider the following SAS output giving 5 confidence intervals for the mean of Y. If you wanted to guarantee that **joint** coverage of the five confidence intervals was at least 95%, what confidence level would you use when forming each interval, using the Bonferroni correction? Compute this adjusted confidence interval for the mean of Y when $X = 4$. (Note that some observations have been omitted from the output.)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	16183	16183	805.62	<.0001
Error	16	321.39597	20.08725		
Corrected Total	17	16504			

Root MSE	4.48188	R-Square	0.9805
Dependent Mean	64.00000	Adj R-Sq	0.9793
Coeff Var	7.00294		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.32215	2.56435	-0.91	0.3786
x	1	14.73826	0.51926	28.38	<.0001

Output Statistics

Obs	x	Dep Var y	Predicted Value	Std Error Mean Predict	95% CL Mean		Residual
3	5	78.0000	71.3691	1.0878	69.0630	73.6752	6.6309
4	1	10.0000	12.4161	2.1021	7.9598	16.8724	-2.4161
6	4	62.0000	56.6309	1.0878	54.3248	58.9370	5.3691
8	3	39.0000	41.8926	1.3125	39.1103	44.6750	-2.8926
10	2	33.0000	27.1544	1.6737	23.6064	30.7024	5.8456

2. Based on the following small data set, construct the design matrix, $\mathbf{X}$, its transpose $\mathbf{X}'$, and the matrices $\mathbf{X}'\mathbf{X}$, $(\mathbf{X}'\mathbf{X})^{-1}$, $\mathbf{X}'\mathbf{Y}$, and $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. (Chapter 5 in the book discusses finding the inverse of a matrix.)

X	Y
0	1
1	4
2	7
3	9
4	10
5	8

For the following 5 problems, consider the data given in the file `CH06PR18.DAT`, which describes a data set ($n = 24$) used to evaluate the relation between intermediate and senior level annual salaries of bachelor's and master's level mathematicians (Y, in thousand dollars) and an index of work quality (X_1), number of years of experience (X_2), and an index of publication success (X_3).

3. Run the multiple linear regression with quality, experience, and publications as the explanatory variables and salary as the response variable. Summarize the regression results by giving the fitted regression equation, the value of R^2, and the results of the significance test for the null hypothesis that the three regression coefficients for the explanatory variables are all zero (give null and alternative hypotheses, test statistic with degrees of freedom, p-value, and brief conclusion in words).

4. Give 85% confidence intervals (do not use a Bonferroni correction) for regression coefficients of quality, experience, and publications based on the multiple regression. Describe the results of the hypothesis tests for the individual regression coefficients (give null and alternative hypotheses, test statistic with degrees of freedom, p-value, and a brief conclusion in words). What is the relationship between these results and the confidence intervals?

5. Plot the residuals versus the *predicted* salary and *each* of the explanatory variables (i.e., 4 residual plots). Are there any unusual patterns?

6. Examine the assumption of normality for the residuals using a qqplot and histogram. State your conclusions.

7. Predict the salary for a mathematician with quality index equal to 5.2, 15 years of experience, and publication index equal to 6.5 . Provide a 85% prediction interval with your prediction.