Department of Mathematics and Statistics
The University of Toledo

Master's Comprehensive Examination
Applied Statistics

December 13, 2014

Instructions:

Do all four problems;

Show all of your computations;

Prove all of your assertions or quote appropriate theorems;

This is three-hour open book examination.

1. (20 points) Do twins have the same IQ? To study this issue, a random sample of 7 pairs of twins was taken and their full scale IQ were recorded. Below the results are summarized in a table:

Older twin	96	89	102	104	129	98	91
Younger twin	89	87	103	96	125	101	96

Find the exact p-value and use it to make your decision.

(a) (10 points) Use Wilcoxon signed rank test at level $\alpha = 0.10$ to decide whether there is a significant difference between twins IQ. Find the exact p-value and use it to make your decision.

(b) (10 points) Repeat part (a) by using normal approximation. Do you get the same conclusion from (a)?

2. (30 points) Use the data set csdata.dat for next three problems. The data have id-a numerical identifier for each student; dependent variable: GPA-dependent variable, the grade point average after three semesters; six explanatory variables. HSM, HSS, HSE, SATM, SATV and GENDER - coded as 1 for men and 2 for women.

(a). (5 points) Give the equation of the fitted regression line using all six explanatory variables.

(b). (5 points) Use the C_p criterion to select the best subset of variables for this problem. Summarize the results and explain your choice of the best model.

In the following problems, use the model which only have HSM and HSE as explanatory variables to predict the response GPA.

(c). (5 points) Obtain the variance inflation factors. Are there indications that serious multicollinearity problems exist here? Explain.

(d). (5 points) Obtain the studentized deleted residuals for observation 8. Use the Bonferroni outlier test procedure with $\alpha = 0.10$ to identify whether it is an outlying Y oberservation.

(e). (5 points) Use the diagonal elements of the hat matrix to identify whether observation 8 is an outlying X observation. State the decision rule and conclusion.

(f). (5 points) Obtain DFFITS, DFBETAS and Cook's distance values for observation 8 to assess its influence. What do you conclude? (Hint: $F_{3,221}(0.5) = 0.7911$)

Appendix for Problem 2

SAS Output

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: GPA

Number of Observations Read 224
Number of Observations Used 224

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	28.68514	4.78086	9.72	<.0001
Error	217	106.77765	0.49206		
Corrected Total	223	135.46279			

Root MSE	0.70147	R-Square	0.2118
Dependent Mean	2.63522	Adj R-Sq	0.1900
Coeff Var	26.61907		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.27864	0.43369	0.64	0.5212
HSM	1	0.14423	0.03979	3.62	0.0004
HSS	1	0.03827	0.03874	0.99	0.3244
HSE	1	0.05103	0.04228	1.21	0.2287
SATM	1	0.00100	0.00071725	1.40	0.1633
SATV	1	-0.00041086	0.00059323	-0.69	0.4893
GENDER	1	0.03237	0.11148	0.29	0.7718

The REG Procedure
Model: MODEL1
Dependent Variable: GPA

Number of Observations Read 224
Number of Observations Used 224

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	27.30349	13.65175	27.89	<.0001
Error	221	108.15930	0.48941		
Corrected Total	223	135.46279			

Root MSE	0.69958	R-Square	0.2016
Dependent Mean	2.63522	Adj R-Sq	0.1943
Coeff Var	26.54718		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	0.62423	0.29172	2.14	0.0335	.	0
HSM	1	0.18265	0.03196	5.72	<.0001	0.80029	1.24954
HSE	1	0.06067	0.03473	1.75	0.0820	0.80029	1.24954

Model: MODEL1
Dependent Variable: GPA

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	3.3200	3.0575	0.0792	0.2625	0.695	0.378	| | |	0.001
2	2.2600	2.0235	0.1100	0.2365	0.691	0.342	| | |	0.001
3	2.3500	2.5708	0.0477	-0.2208	0.698	-0.316	| | |	0.000
4	2.0800	2.6928	0.0695	-0.6128	0.696	-0.880	| * | |	0.003
5	3.3800	2.5708	0.0477	0.8092	0.698	1.159	| |** |	0.002
6	3.2900	2.9361	0.0723	0.3539	0.696	0.509	| |* |	0.001
7	3.2100	2.5102	0.0582	0.6998	0.697	1.004	| |** |	0.002
8	2.0000	1.5362	0.1592	0.4638	0.681	0.681	| |* |	0.008

Dependent Variable: GPA

Output Statistics

Obs	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS Intercept	DFBETAS HSM	DFBETAS HSE
1	0.3770	0.0128	1.0249	0.0429	-0.0299	0.0130	0.0229
2	0.3416	0.0247	1.0378	0.0544	0.0509	-0.0129	-0.0367
3	-0.3157	0.0046	1.0170	-0.0216	-0.0064	0.0040	-0.0006
4	-0.8799	0.0099	1.0131	-0.0879	-0.0230	-0.0489	0.0603
5	1.1603	0.0046	1.0000	0.0792	0.0237	-0.0147	0.0022
6	0.5077	0.0107	1.0211	0.0528	-0.0120	0.0402	-0.0199
7	1.0039	0.0069	1.0069	0.0838	0.0483	0.0097	-0.0481
8	0.6800	0.0518	1.0623	0.1589	0.1295	-0.1373	0.0033

3. Let $X_1, X_2, \ldots, X_{15}$ be a random sample from the exponential distribution with pdf $f(x) = \theta e^{-\theta x}$ for $x > 0$ and $\theta > 0$. To test $H_0 : \theta = 1/5$ versus $H_a : \theta < 1/5$, you will use $X_{(1)} = \min\{X_1, X_2, \ldots, X_{15}\}$ as a test statistic. If $X_{(1)} \geq 1$, you will reject the null hypothesis.

(a) Find the distribution of $X_{(1)}$.

(b) Compute the probability of a Type I error.

(c) Find the power of the test if, in fact, $\theta = 1/25$.

(d) Find the power function of the test.

4. Suppose the height and weight of 30 girls in a city are measured. The mean and standard deviation of the heights are 46 and 7 inches, respectively, and the mean and standard deviation of the weights are 94 and 15 pounds, respectively. Suppose the sample correlation coefficient between height and weight is 0.75.

(a) Find the equation of the least-squares regression line of weight against height.

(b) Find the predicted weight of a girl who is 5 feet (60 inches) tall.

(c) What percentage of the variation in weight is explained by the fitted least-squares regression line in part (a)?

(d) Create the analysis of variance table.

(e) What is the value of the regression standard error s? The statistic s^2 is an unbiased estimate of some parameter—what is the parameter?

(f) Test the null hypothesis that there is no linear relationship between the height and weight against the two-sided alternative. Give the results of the significance test.