22S:30/105, Statistical Methods and Computing
Spring 2008, Instructor: Cowles
Final Exam
PRACTICE PROBLEMS for 2009 FINAL EXAM

Name: _____ Course no. (30 or 105) _____

Secret number for posting grade: _____

1. The Core Plus Mathematics Project (CPMP) is an innovative approach to teaching mathematics that engages students in group investigations and mathematical modeling. Math educators wished to determine whether students in CPMP developed better algebra skills than students taught with traditional methods. After field tests in 36 high schools over a 3-year period, researchers gave an applied algebra test to a sample of 320 CPMP students and to a sample of 272 students from a traditional math program. The individual test scores were numeric scores ranging from 0 to 100.

   (a) What is/are the population(s) of interest?

   (b) What is the variable of interest?

   (c) What is/are the parameter(s) of interest?

   (d) Which hypothesis testing procedure would be most likely to apply in this case? (Circle one)
      i. single sample z test
      ii. two independent sample z test
      iii. single sample t test
      iv. two independent sample t test
      v. paired t test
      vi. chi square test
      vii. ANOVA

   (e) Briefly justify your choice from the preceding question.

   (f) If the data did not justify the assumptions of the test that you chose above, which of the following nonparametric tests would be most appropriate? (Circle one)

      i. Wilcoxon signed rank test
      ii. Wilcoxon rank sum test
      iii. sign test
      iv. Kruskal Wallis test

   (g) Briefly justify your choice from the preceding question.

2. Identify the data type of each of the variables described below. Choose from binary, nominal, ordinal, discrete quantitative, and continuous quantitative.

   (a) (1) ratings of instructors' teaching (can take on values "unsatisfactory," "satisfactory," "excellent")

   (b) (1) whether a student has received a failing grade in any course taken at the UI (yes or no)

   (c) (1) the number of failing grades a student has received in any courses taken at the UI (can take on values 0, 1, 2, ...)

   (d) (1) the department within the College of Business in which a student is majoring (can take on values "Accounting," "Economics," "Finance," "Management and Organization," "Management Information Systems," "Marketing")

3. (3) The National Collegiate Athletic Association (NCAA) requires Division I athletes to score at least 820 on the combined mathematics and verbal parts of the SAT exam to compete in their first college year. In 2002, the scores of the entire 1.3 million students taking the SATs were approximately Normal with mean 1020 and standard deviation 207. What percent of all students who took the SATs in 2002 had scores less than or equal to 820? (Numeric answer; show your work.)

4. A political scientist was interested in what factors determine the amount of public expenditures by state and local governments. He assembled a dataset containing per capita state and local public expenditures and associated state demographic and economic characteristics in 1960. There is one observation for each state in the U.S. Two of the variables are:

   (a) EX: Per capita state and local public expenditures in dollars

(b) ECAB: Economic ability index, in which income, retail sales, and the value of output (manufactures, mineral, and agricultural) per capita are equally weighted.

Refer to the attached SAS code and output in answering the following questions:

(a) (1) Based on the boxplot and stemplot of the `EX` variable, what is the approximate numeric value of the first quartile of these data?

(b) (1) Which phrase best describes the shape of the distribution of values of `EX`? (Circle one)

i. perfectly symmetric
ii. slightly right skewed
iii. slightly left skewed
iv. extremely right skewed
v. extremely left skewed
vi. more than one mode

(c) (1) In the scatterplot, which variable, `EX` or `ECAB`, is treated as the response variable?

(d) (1) In the residual plot at the end of the output, what is plotted on the y-axis?

‘

(e) (1) The null hypothesis is that there is no linear relationship between `EX` and `ECAB`. Write this null hypothesis as a statement about a population parameter. Use conventional symbols.

(f) (1) The political scientist would be interested in a linear relationship between economic ability and expenditures regardless of whether the relationship was positive or negative. Write his alternative hypothesis as a statement about a population parameter. Use conventional symbols.

(g) (1) Give a point estimate and a 95% confidence interval for the population slope (numeric answers taken from the SAS output).

(h) (2) Does your answer to the preceding question provide evidence in favor of the alternative hypothesis? (yes/no) Explain briefly. (If you could not answer the previous question, pretend that the point estimate is 1 and the confidence interval is (0,2) and answer this question accordingly.)

(i) (1) What are the numeric values of the test statistic and the p-value for the two-sided test of no linear relationship between educational level and crime rate (numeric answers from SAS output)?

(j) (2) Based on your answer to the preceding question, would you reject the null hypothesis at significance level $\alpha = .05$? (yes/no) Briefly explain. (If you could not answer the previous question, pretend that the test statistic is 2.5 and the p-value is 0.001, and answer accordingly.)

(k) (1) Use the estimated regression equation to estimate the public expenditure in an individual state with economic ability = 87. Show your calculations.

(l) (1) Give the interval in which you are 95% confident that the public expenditure of an individual state with economic ability equal to 87 would lie. Note that observation number 3 in the dataset has `ECAB` = 87. (Copy the endpoints from the SAS output).

(m) (1) What is the estimated value of the standard deviation of points around the regression line (regression standard error)? (numeric answer)

(n) (1) R^2 given in the regression output is .40. Does this indicate a strong or weak relationship between the `EX` and `ECAB`? Explain briefly.

5. An occupational safety expert wishes to study the level of knowledge of safe equipment use amongst employees of a tool and dye factory. She randomly selects 15 factory employees gives each one a written test on equipment safety practices. The range of possible scores on the written test is 0 to 50. She is specifically interested in making inference about the population mean μ of scores on this test in the population of all employees of such factories. She wants to test the hypotheses:

$$H_0: \ \mu = 35$$
$$H_A: \ \mu < 35$$

She is strongly convinced that the population standard deviation is $\sigma = 7$.

(a) (2) The safety expert wants to carry out her test at significance level $\alpha = .05$. What is the critical value of the sample mean $\bar{x}$ such that she should reject H_0 for all values of $\bar{x}$ less than or equal to the critical value? (numeric answer; show your work)

(b) (2) If the safety expert uses the critical value you found in the previous question, what is the power of her test against the alternative that $\mu = 30$? (numeric answer; show your work)

6. Before an intensive TV advertising campaign, the producers of Nike athletic shoes find that 29 out of a random sample of 200 upper-income adults are aware of their new leisure shoe line. A second random sample of 300 such adults is taken after the campaign. Now 96 of the 300 sampled can identify the new line.

(a) Give a 99% confidence interval for the increase in the proportion of upper-income adults showing brand awareness. (Numeric answer; show your work.)

(b) What quantity are we 99% confident lies in your interval? (Circle one).

i. $\bar{x}_{after} - \bar{x}_{before}$
ii. $\mu_{after} - \mu_{before}$
iii. $\hat{p}_{after} - \hat{p}_{before}$
iv. $p_{after} - p_{before}$

7. Researchers interviewed 10 young married couples, wife and husband separately. One question asks how important the attractiveness of their spouse is to them on a scale of 1 (not important) to 10 (extremely important). Here are the responses:

	Couple									
	1	2	3	4	5	6	7	8	9	10
Husband	7	7	7	3	9	5	10	6	6	7
Wife	4	2	5	2	2	2	4	7	1	5

We wish to use these data to test the null hypothesis that there is no difference between husbands and wives in how important attractiveness of their spouse is to them, versus the alternative hypothesis that attractiveness of their spouse is more important to husbands than to wives.

A nonparametric test is a good idea in this case because the data are ordinal rather than truly quantitative.

(a) Is this a two-independent sample problem or a paired-sample problem?

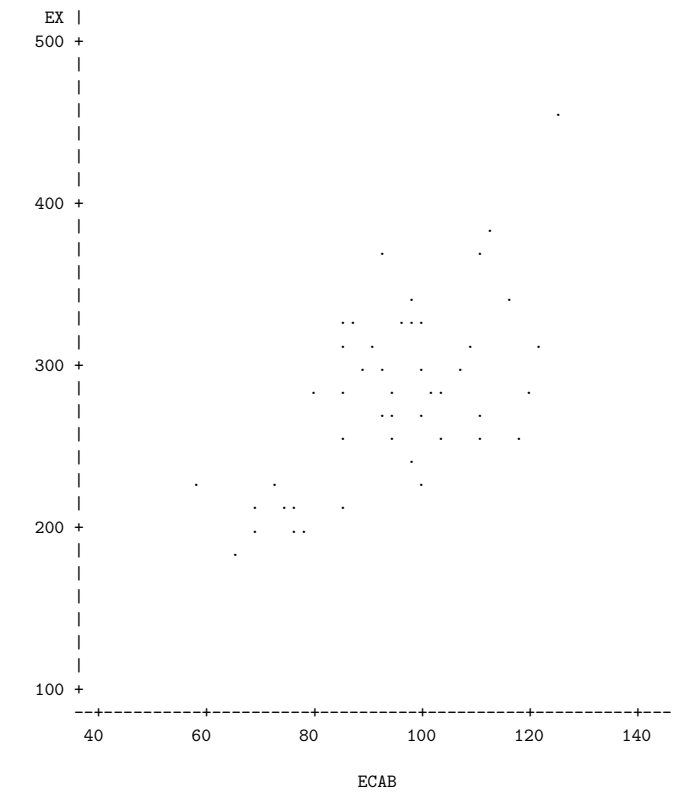
(b) To which variable should the testing procedure be applied?

i. husbands' responses
ii. wives' responses
iii. differences between husbands' and wives' responses
iv. none of the above

(c) Refer to the last page of SAS output to find the p-value for the hypothesis test. Give the numeric value.

(d) At the .05 significance level, should you reject the null hypothesis? (yes/no) Explain what your conclusion means in terms of husbands, wives, and the importance of spouse attractiveness.

Plot of EX*ECAB. Symbol used is '.'.



NOTE: 1 obs hidden.

The REG Procedure
Model: MODEL1
Dependent Variable: EX

Number of Observations Read	47
Number of Observations Used	47

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	57407	57407	29.82	<.0001
Error	45	86629	1925.08573		
Corrected Total	46	144036			

Root MSE	43.87580	R-Square	0.3986
Dependent Mean	283.78723	Adj R-Sq	0.3852
Coeff Var	15.46081		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	70.96941	39.49381	1.80	0.0791
ECAB	1	2.25321	0.41261	5.46	<.0001

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	-8.57521	150.51403
ECAB	1	1.42216	3.08425

The REG Procedure
Model: MODEL1
Dependent Variable: EX

Output Statistics

Obs	ECAB	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean	
1	85.5	256.0000	263.6186	7.3892	248.7361	278.5012
2	94.3	275.0000	283.4469	6.4002	270.5561	296.3376
3	87	327.0000	266.9984	7.1001	252.6981	281.2988
4	107.5	297.0000	313.1892	8.3635	296.3442	330.0342
5	94.9	256.0000	284.7988	6.4026	271.9032	297.6943
6	121.6	312.0000	344.9594	12.9013	318.9748	370.9441
7	111.5	374.0000	322.2020	9.5103	303.0474	341.3567
8	117.9	257.0000	336.6225	11.6005	313.2579	359.9872
9	103.1	257.0000	303.2751	7.3277	288.5164	318.0338
10	116.1	336.0000	332.5668	10.9887	310.4344	354.6991
11	93.4	269.0000	281.4190	6.4146	268.4993	294.3387
12	77.2	213.0000	244.9170	9.5721	225.6378	264.1963
13	108.4	308.0000	315.2171	8.6073	297.8811	332.5530
14	111.8	273.0000	322.8780	9.6022	303.5382	342.2178
15	110.8	256.0000	320.6248	9.2987	301.8963	339.3532
16	120.9	287.0000	343.3822	12.6514	317.9010	368.8633
17	104.3	290.0000	305.9789	7.5811	290.7097	321.2481
18	85.1	217.0000	262.7173	7.4730	247.6659	277.7688
19	76.8	198.0000	244.0157	9.6955	224.4880	263.5434
20	75.1	217.0000	240.1853	10.2329	219.5752	260.7953
21	78.7	195.0000	248.2968	9.1213	229.9256	266.6680
22	65.2	183.0000	217.8785	13.6612	190.3634	245.3937
23	73	222.0000	235.4535	10.9224	213.4546	257.4524
24	80.9	283.0000	253.2539	8.4984	236.1373	270.3705
25	69.4	217.0000	227.3420	12.1573	202.8559	251.8281
26	57.4	231.0000	200.3035	16.5733	166.9231	233.6839
27	95.7	329.0000	286.6013	6.4207	273.6695	299.5332
28	100.2	294.0000	296.7408	6.8254	282.9937	310.4878
29	99.1	232.0000	294.2623	6.6812	280.8056	307.7189
30	93.4	369.0000	281.4190	6.4146	268.4993	294.3387
31	88.2	302.0000	269.7023	6.9001	255.8047	283.5999
32	99.1	269.0000	294.2623	6.6812	280.8056	307.7189
33	102.2	291.0000	301.2472	7.1542	286.8380	315.6564
34	86	323.0000	264.7452	7.2883	250.0659	279.4245
35	68.6	198.0000	225.5394	12.4392	200.4856	250.5933
36	84.9	282.0000	262.2667	7.5160	247.1287	277.4047

37	98.8	246.0000	293.5863	6.6467	280.1991	306.9735
38	86.2	309.0000	265.1959	7.2491	250.5954	279.7964
39	90.2	309.0000	274.2087	6.6360	260.8432	287.5742
40	97.6	334.0000	290.8824	6.5305	277.7293	304.0355
41	93.9	284.0000	282.5456	6.4040	269.6473	295.4439
42	125.8	454.0000	354.4229	14.4317	325.3560	383.4898
43	98	344.0000	291.7837	6.5653	278.5605	305.0070
44	92.5	307.0000	279.3911	6.4504	266.3994	292.3828
45	100.4	333.0000	297.1914	6.8545	283.3857	310.9971
46	98	343.0000	291.7837	6.5653	278.5605	305.0070
47	112.6	380.0000	324.6805	9.8507	304.8402	344.5209

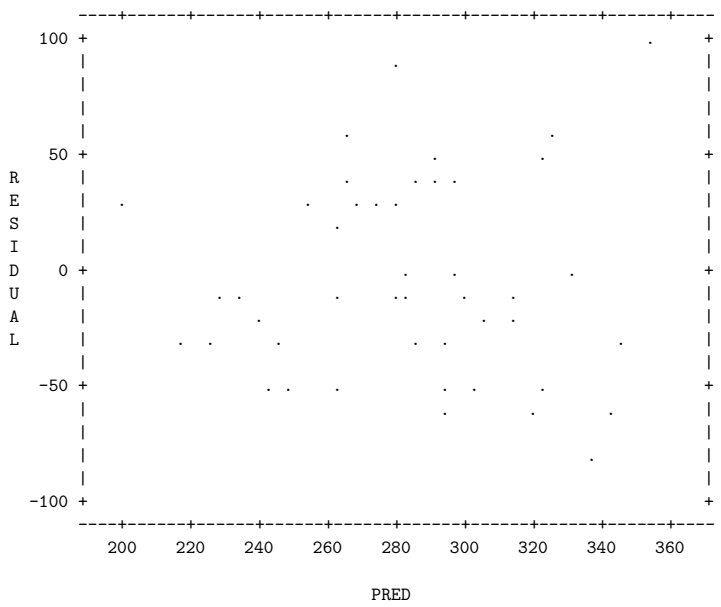
The REG Procedure
Model: MODEL1
Dependent Variable: EX

Output Statistics

Obs	ECAB	95% CL Predict		Residual
1	85.5	174.0038	353.2335	-7.6186
2	94.3	194.1412	372.7525	-8.4469
3	87	177.4785	356.5184	60.0016
4	107.5	223.2276	403.1507	-16.1892
5	94.9	195.4924	374.1051	-28.7988
6	121.6	252.8479	437.0709	-32.9594
7	111.5	231.7795	412.6245	51.7980
8	117.9	245.2156	428.0295	-79.6225
9	103.1	213.6807	392.8694	-46.2751
10	116.1	241.4670	423.6665	3.4332
11	93.4	192.1091	370.7288	-12.4190
12	77.2	154.4680	335.3660	-31.9170
13	108.4	225.1623	405.2719	-7.2171
14	111.8	232.4161	413.3399	-49.8780
15	110.8	230.2916	410.9579	-64.6248
16	120.9	251.4114	435.3529	-56.3822
17	104.3	216.2991	395.6588	-15.9789
18	85.1	173.0743	352.3604	-45.7173
19	76.8	153.5135	334.5180	-46.0157
20	75.1	149.4433	330.9272	-23.1853
21	78.7	158.0370	338.5566	-53.2968
22	65.2	125.3236	310.4334	-34.8785
23	73	144.3861	326.5210	-13.4535
24	80.9	163.2411	343.2667	29.7461
25	69.4	135.6420	319.0420	-10.3420
26	57.4	105.8388	294.7682	30.6965
27	95.7	197.2898	375.9129	42.3987
28	100.2	207.3075	386.1740	-2.7408
29	99.1	204.8732	383.6513	-62.2623
30	93.4	192.1091	370.7288	87.5810
31	88.2	180.2458	359.1588	32.2977
32	99.1	204.8732	383.6513	-25.2623
33	102.2	211.7098	390.7846	-10.2472
34	86	175.1639	354.3265	58.2548
35	68.6	133.6862	317.3927	-27.5394
36	84.9	172.6091	351.9243	19.7333
37	98.8	204.2076	382.9649	-47.5863

38	86.2	175.6275	354.7643	43.8041
39	90.2	184.8333	363.5841	34.7913
40	97.6	201.5385	380.2263	43.1176
41	93.9	193.2388	371.8523	1.4544
42	125.8	261.3949	447.4509	99.5771
43	98	202.4295	381.1380	52.2163
44	92.5	190.0708	368.7114	27.6089
45	100.4	207.7491	386.6337	35.8086
46	98	202.4295	381.1380	51.2163
47	112.6	234.1103	415.2508	55.3195

The REG Procedure
Model: MODEL1



The UNIVARIATE Procedure
Variable: EX

Stem	Leaf	#	Boxplot
44	4	1	0
42			
40			
38	0	1	
36	94	2	
34	34	2	
32	379346	6	+-----+
30	278992	6	| |
28	23470147	8	*--+--*
26	9935	4	| |
24	666677	6	+-----+
22	212	3	
20	3777	4	
18	3588	4	

Multiply Stem.Leaf by 10**+1

SAS output for question regarding importance of attractiveness of spouses.

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	4.256342	Pr > |t|	0.0021
Sign	M	4	Pr >= |M|	0.0215
Signed Rank	S	26	Pr >= |S|	0.0059