PRACTICE PROBLEMS for FINAL 2006
22S:30/105, Statistical Methods and Computing
Spring 2005, Instructor: Cowles
Final Exam

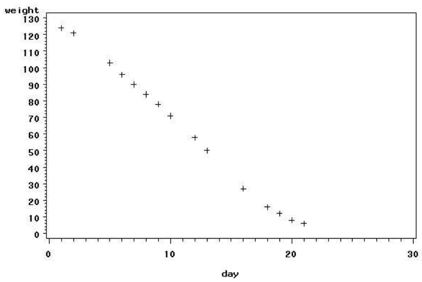
Name: ____ Course no. (30 or 105) _____

Secret number for posting grade: _______

Note: The number of points for each question is indicated in parentheses.

1 General questions

1. Mr. Rex Boggs from Australia weighed the bar of soap in his shower stall each morning before showering. The weight went down as the soap was used. On some days he forgot to weigh the soap. The scatterplot below shows the weight (in grams) versus the day of measurement (numbered 1 through 21).



   (a) (1) Based on the scatterplot, the sample correlation is (circle one)
      i. close to 1
      ii. positive but not close to 1
      iii. close to 0
      iv. negative but not close to -1
      v. close to -1

   (b) (1) Briefly explain your answer.

2. For a biology project, you measure the thorax length (in millimeters) and the weight (in milligrams) of 12 bees of the same species. What units of measurement do each of the following have:

   (a) (1) the sample mean weight

   (b) (1) the first quartile of weight

   (c) (1) the standard deviation of weight

   (d) (1) the correlation between thorax length and weight

3. Digoxin is a drug often prescribed for patients with heart disease. It is taken in pill form, and patients are instructed to drink a full glass of water when they take their digoxin.

   Researchers (Parker et al., *Pharmacotherapy*, 2003) were interested in whether the concentration of digoxin in the bloodstream would be higher if people drank grapefruit juice instead of water when they took their digoxin.

   Seven volunteers participated in the study. Subjects took digoxin with water for 2 weeks, no digoxin for 2 weeks, and digoxin with grapefruit juice for 2 weeks. The response variable — peak concentration of digoxin in the blood plasma (Cmax) — was measured on each patient during the water period and again during the grapefruit juice period. Cmax is a continuous quantitative variable.

   We wish to determine whether their data give evidence at the .05 significance level that Cmax is *higher* when digoxin is taken with grapefruit juice than when it is taken with water.

   (a) (1) Which type of problem is this? (Circle one)
      i. single sample
      ii. paired sample
      iii. two independent sample
      iv. none of the above

   (b) (2) Of the statistical tests that we have studied, the one most likely to be useful for addressing this problem is a paired t-test. Which of the following assumptions need to be met for the paired t-test to give reliable results in this problem? (Circle as many as apply.)
      i. The distribution of Cmax must be approximately normal in the population of all people who take digoxin with water and in the population of all people who take digoxin with grapefruit juice.
      ii. The population distribution of differences between Cmax when digoxin is taken with water and Cmax when digoxin is taken with grapefruit juice must be approximately normal.

iii. The population standard deviations of Cmax must be approximately equal in the population of all people who take digoxin with water and in the population of all people who take digoxin with grapefruit juice.

iv. $n\ p$ and $n\ (1-p)$ must both be greater than or equal to 5.

v. none of the above.

4. Psychiatrists wish to determine the effects of different types of lighting (full spectrum light, regular fluorescent light, and regular incandescent light) and supplementation with Omega 3 fatty acids on depression. Forty eight people who have been diagnosed with mild depression are recruited into the study. They are randomly assigned to six groups, with 8 people in each group. Each subject is given a light bulb to install in the place where he/she spends the most time during each day. Each subject is also given a bottle of pills, of which they are to take one each day. The groups receive the following:

   - full spectrum light; Omega 3 fatty acid supplements
   - regular fluorescent light; Omega 3 fatty acid supplements
   - regular incandescent light; Omega 3 fatty acid supplements
   - full spectrum light; placebo
   - regular fluorescent light; placebo
   - regular incandescent light; placebo

   The subjects are not told which type of lightbulb they have been given and whether their bottle of pills is real supplements or placebos.

   The subjects are given a a written depression inventory test at the beginning of the study and again after a month on the light/supplements regimen. The researchers are interested in whether the changes in depression scores are different in the different groups.

   (a) (1) Is this an experiment or an observational study?

   (b) (1)What are the experimental units?

   (c) (1) What are the factors?

   (d) (1) What are the levels?

   (e) (1) What are the treatments?

   (f) (1) What is the response variable?

5. A historian examining British colonial records for the Gold Coast in Africa suspects that the death rate was higher among African miners than among European miners. In the year 1936 there were

   223 deaths among 33,809 African miners

   7 deaths among 1541 European miners

   in the Gold Coast. (Data courtesy of Raymond Dumett, Purdue University).

   Consider this year as a sample from the prewar era in Africa. We wish to determine whether the data provides good evidence that the proportion of African miners who died during a year was higher than the proportion of European miners who died.

   (a) (2) State the null and alternative hypotheses, using conventional symbols.

   (b) (3) Calculate a test statistic. Show your work and give a numeric result.

   (c) (1) Give a p-value as exact as the tables in the text allow. (numeric result)

   (d) (2) State your conclusion in terms of this application. (If you could not get the p-value in the preceding question, pretend that it was .008 and answer this question accordingly.)

   (e) (3) Give a 95% confidence interval for the difference between the proporation of African miners who would die in a year and the proportion of European miners who would die in a year. (numeric answer)

6. A car salesman would like to estimate the proportion of all UI faculty who have not purchased a car in the last 5 years. He will select a simple random sample of UI faculty and will ask each person in the sample whether he or she has purchased a car in the last 5 years. The salesman wants to calculate a 90% confidence interval with margin of error no greater than 0.03.

   (a) (2) How large a simple random sample of UI faculty will he need if he is pretty sure that the true population proportion is close to .15?

   (b) (2) How large a simple random sample of UI faculty will he need if he has no preliminary idea about the population proportion?

   (c) (2) The total number of faculty at the UI is about 1200. If the car salesman does obtain a sample of the size you calculated in the second part of this problem, should he use normal approximations to calculate his confidence interval? (yes/no) Why or why not?

7. For each of the following variables, state which data type it is (binary, nominal, ordinal, quantitative continuous, quantitative discreet).

   (a) (0.5) hair color (evaluated on a sample of human beings)

   (b) (0.5) boiling temperature of water (evaluated at a number of different elevations in the mountaints)

8. Every spring, Nenana, Alaska, hosts a contest in which participants try to guess the exact minute that a wooden stand placed on the frozen Tanana River will fall through the breaking ice. The contest started in 1917 as entertainment for railroad engineers. It has grown into an event in which hundreds of thousands of entrants enter their guesses on the Internet and compete for prizes of more than $300,000. Because so much money depends on the time of ice breakup, it has been recorded to the nearest minutie with great accuracy ever since 1917.An article in *Science* ("Climate Change in Nontraditional Datasets," Oct. 2001, p. 811) used the data to investigate global

warming by asking the question whether ice breakup had tended to occur earlier over time.

The dataset available to us contains two variables:

- `year`
- `julian` – the number of days from midnight on Jan 1 until the time of ice breakup

Refer to the SAS output provided to answer the following questions.

   (a) (1) The null hypothesis is that there is no linear relationship between year and time of ice breakup. Write this null hypothesis as a statement about a population parameter. Use conventional symbols.

   (b) (1) The alternative hypothesis is that time of ice breakup decreases linearly over time. Write this alternative hypothesis as a statement about a population parameter. Use conventional symbols.

   (c) (1) Give a point estimate and a 95% confidence interval for the population slope (numeric answers).

   (d) (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 -0.11 and the confidence interval is (-0.21,-0.01) and answer this question accordingly.)

   wer)

   (e) (0.5) What is the estimated value of the standard deviation of points around the regression line? (numeric answer)

9. (1.5) What is the p-value for the one-sided test of no linear relationship between year and time of breakup?

10. (1.5) Use the estimated regression equation to predict the time of breakup for this year (2005). Show your calculation.

11. (1) On the SAS output, circle the numbers that provide the endpoints of the interval in which you are 95% confident that breakup in 2005 time would lie. Be sure to put your name on the SAS output.

12. (1) What proportion of the variability in time of breakup is explained by year? (numeric answer)

13. (0.5) What is the estimated value of the standard deviation of points around the regression line? (numeric answer)

Obs	year	julian
1	1917	120.480
2	1918	131.398
3	1919	123.607
many lines omitted		
77	1993	113.543
78	1994	119.959
79	1995	116.557
80	1996	126.523
81	1997	120.437
82	1998	110.705
83	1999	119.908
84	2000	122.450
85	2001	128.542
86	2002	127.894
87	2003	119.766
88	2005	.

The REG Procedure
Model: MODEL1
Dependent Variable: julian

Number of Observations Read	88
Number of Observations Used	87
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	238.06059	238.06059	7.26	0.0085
Error	85	2787.93204	32.79920		
Corrected Total	86	3025.99262			

Root MSE	5.72706	R-Square	0.0787
Dependent Mean	125.54431	Adj R-Sq	0.0678
Coeff Var	4.56178		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	254.64848	47.92518	5.31	<.0001
year	1	-0.06587	0.02445	-2.69	0.0085

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	159.36037	349.93658
year	1	-0.11448	-0.01726

Output Statistics

Obs	year	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean	
1	1917	120.4795	128.3767	1.2175	125.9560	130.7974
2	1918	131.3983	128.3108	1.1965	125.9320	130.6897
3	1919	123.6066	128.2450	1.1755	125.9077	130.5822
4	1920	132.4490	128.1791	1.1548	125.8831	130.4751
5	1921	131.2795	128.1132	1.1341	125.8583	130.3682
6	1922	132.5559	128.0474	1.1136	125.8331	130.2616
7	1923	129.0837	127.9815	1.0933	125.8077	130.1553
8	1924	132.6323	127.9156	1.0732	125.7818	130.0494
9	1925	127.7726	127.8497	1.0532	125.7556	129.9438
10	1926	116.6691	127.7839	1.0335	125.7291	129.8387
11	1927	133.2378	127.7180	1.0139	125.7021	129.7339
12	1928	127.6844	127.6521	0.9946	125.6747	129.6296
13	1929	125.6538	127.5863	0.9754	125.6468	129.5257
14	1930	128.7941	127.5204	0.9566	125.6185	129.4223
15	1931	130.3913	127.4545	0.9379	125.5896	129.3194
16	1932	122.4274	127.3887	0.9196	125.5602	129.2171
17	1933	128.8128	127.3228	0.9015	125.5303	129.1153
18	1934	120.5885	127.2569	0.8838	125.4997	129.0142
19	1935	135.5642	127.1910	0.8664	125.4685	128.9136
20	1936	121.5406	127.1252	0.8493	125.4365	128.8138
21	1937	132.8365	127.0593	0.8326	125.4039	128.7148
22	1938	126.8434	126.9934	0.8163	125.3704	128.6164
23	1939	119.5601	126.9276	0.8004	125.3362	128.5190
24	1940	111.6441	126.8617	0.7849	125.3011	128.4224
25	1941	123.0767	126.7958	0.7699	125.2650	128.3267
26	1942	120.5615	126.7300	0.7554	125.2280	128.2320
27	1943	118.8073	126.6641	0.7415	125.1899	128.1383
28	1944	125.5892	126.5982	0.7280	125.1507	128.0458
29	1945	136.4038	126.5324	0.7152	125.1104	127.9544
30	1946	125.6948	126.4665	0.7030	125.0688	127.8642
31	1947	123.7455	126.4006	0.6914	125.0259	127.7753
32	1948	134.4677	126.3347	0.6805	124.9817	127.6878
33	1949	134.5274	126.2689	0.6703	124.9361	127.6017
34	1950	126.6767	126.2030	0.6609	124.8890	127.5170
35	1951	120.7462	126.1371	0.6522	124.8403	127.4340
36	1952	133.7115	126.0713	0.6444	124.7900	127.3525
37	1953	119.6628	126.0054	0.6374	124.7381	127.2727
38	1954	126.7510	125.9395	0.6313	124.6844	127.1947
39	1955	129.5927	125.8737	0.6261	124.6289	127.1184
40	1956	122.9753	125.8078	0.6217	124.5716	127.0440
41	1957	125.3962	125.7419	0.6184	124.5124	126.9714
42	1958	119.6226	125.6761	0.6159	124.4514	126.9007
43	1959	128.4767	125.6102	0.6145	124.3884	126.8320
44	1960	123.8003	125.5443	0.6140	124.3235	126.7651
45	1961	125.4802	125.4784	0.6145	124.2567	126.7002
46	1962	132.9747	125.4126	0.6159	124.1879	126.6372
47	1963	125.7677	125.3467	0.6184	124.1172	126.5762
48	1964	141.4872	125.2808	0.6217	124.0446	126.5170

Output Statistics

Obs	year	95% CL Predict		Residual
1	1917	116.7353	140.0181	-7.8972
2	1918	116.6781	139.9436	3.0875
3	1919	116.6206	139.8693	-4.6384
4	1920	116.5630	139.7952	4.2699
5	1921	116.5052	139.7213	3.1663
6	1922	116.4471	139.6476	4.5085
7	1923	116.3889	139.5741	1.1022
8	1924	116.3305	139.5007	4.7167
9	1925	116.2719	139.4276	-0.0771
10	1926	116.2130	139.3547	-11.1148
11	1927	116.1540	139.2820	5.5198
12	1928	116.0948	139.2095	0.0323
13	1929	116.0354	139.1372	-1.9325
14	1930	115.9757	139.0651	1.2737
15	1931	115.9159	138.9932	2.9368
16	1932	115.8559	138.9214	-4.9613
17	1933	115.7956	138.8499	1.4900
18	1934	115.7352	138.7786	-6.6684
19	1935	115.6746	138.7075	8.3732
20	1936	115.6137	138.6366	-5.5846
21	1937	115.5527	138.5659	5.7772
22	1938	115.4914	138.4955	-0.1500
23	1939	115.4300	138.4252	-7.3675
24	1940	115.3683	138.3551	-15.2176
25	1941	115.3065	138.2852	-3.7191
26	1942	115.2444	138.2155	-6.1685
27	1943	115.1821	138.1461	-7.8568
28	1944	115.1197	138.0768	-1.0090
29	1945	115.0570	138.0077	9.8714
30	1946	114.9941	137.9389	-0.7717
31	1947	114.9310	137.8702	-2.6551
32	1948	114.8677	137.8018	8.1330
33	1949	114.8042	137.7335	8.2585
34	1950	114.7405	137.6655	0.4737
35	1951	114.6766	137.5977	-5.3909
36	1952	114.6125	137.5301	7.6402
37	1953	114.5482	137.4626	-6.3426
38	1954	114.4836	137.3954	0.8115
39	1955	114.4189	137.3284	3.7190
40	1956	114.3540	137.2616	-2.8325
41	1957	114.2888	137.1950	-0.3457

42	1958	114.2235	137.1286	-6.0535
43	1959	114.1579	137.0625	2.8665
44	1960	114.0921	136.9965	-1.7440
45	1961	114.0262	136.9307	0.001757
46	1962	113.9600	136.8652	7.5621
47	1963	113.8936	136.7998	0.4210
48	1964	113.8270	136.7347	16.2064

Output Statistics

Obs	year	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean	
49	1965	127.7927	125.2150	0.6261	123.9702	126.4597
50	1966	128.5080	125.1491	0.6313	123.8939	126.4043
51	1967	124.4969	125.0832	0.6374	123.8159	126.3506
52	1968	129.3934	125.0174	0.6444	123.7361	126.2986
53	1969	118.5198	124.9515	0.6522	123.6547	126.2483
54	1970	124.4427	124.8856	0.6609	123.5716	126.1997
55	1971	128.8969	124.8197	0.6703	123.4870	126.1525
56	1972	131.4976	124.7539	0.6805	123.4009	126.1069
57	1973	124.4997	124.6880	0.6914	123.3133	126.0627
58	1974	126.6559	124.6221	0.7030	123.2244	126.0198
59	1975	130.5760	124.5563	0.7152	123.1343	125.9783
60	1976	123.4524	124.4904	0.7280	123.0429	125.9379
61	1977	126.5323	124.4245	0.7415	122.9503	125.8988
62	1978	120.6378	124.3587	0.7554	122.8567	125.8607
63	1979	120.7615	124.2928	0.7699	122.7620	125.8236
64	1980	120.5531	124.2269	0.7849	122.6663	125.7876
65	1981	120.7809	124.1611	0.8004	122.5697	125.7524
66	1982	130.7337	124.0952	0.8163	122.4722	125.7182
67	1983	119.7760	124.0293	0.8326	122.3739	125.6848
68	1984	130.6483	123.9634	0.8493	122.2748	125.6521
69	1985	131.6087	123.8976	0.8664	122.1750	125.6202
70	1986	128.9517	123.8317	0.8838	122.0745	125.5889
71	1987	125.6330	123.7658	0.9015	121.9733	125.5584
72	1988	118.3858	123.7000	0.9196	121.8716	125.5284
73	1989	121.8434	123.6341	0.9379	121.7692	125.4990
74	1990	114.7219	123.5682	0.9566	121.6663	125.4701
75	1991	121.0031	123.5024	0.9754	121.5629	125.4418
76	1992	135.2684	123.4365	0.9946	121.4591	125.4139
77	1993	113.5427	123.3706	1.0139	121.3547	125.3865
78	1994	119.9594	123.3048	1.0335	121.2500	125.3595
79	1995	116.5573	123.2389	1.0532	121.1448	125.3330
80	1996	126.5226	123.1730	1.0732	121.0392	125.3068
81	1997	120.4365	123.1071	1.0933	120.9333	125.2810
82	1998	110.7045	123.0413	1.1136	120.8271	125.2555
83	1999	119.9080	122.9754	1.1341	120.7205	125.2303
84	2000	122.4497	122.9095	1.1548	120.6136	125.2055
85	2001	128.5420	122.8437	1.1755	120.5064	125.1809
86	2002	127.8941	122.7778	1.1965	120.3989	125.1567
87	2003	119.7656	122.7119	1.2175	120.2912	125.1326
88	2005	.	122.5802	1.2600	120.0750	125.0853

Output Statistics

Obs	year	95% CL Predict		Residual
49	1965	113.7602	136.6697	2.5777
50	1966	113.6932	136.6050	3.3589
51	1967	113.6260	136.5405	-0.5863
52	1968	113.5586	136.4761	4.3760
53	1969	113.4910	136.4120	-6.4317
54	1970	113.4231	136.3481	-0.4429
55	1971	113.3551	136.2844	4.0772
56	1972	113.2868	136.2209	6.7437
57	1973	113.2184	136.1576	-0.1883
58	1974	113.1498	136.0945	2.0338
59	1975	113.0809	136.0316	6.0197
60	1976	113.0118	135.9690	-1.0380
61	1977	112.9426	135.9065	2.1078
62	1978	112.8731	135.8442	-3.7209
63	1979	112.8034	135.7822	-3.5313
64	1980	112.7335	135.7203	-3.6738
65	1981	112.6635	135.6586	-3.3802
66	1982	112.5932	135.5972	6.6385
67	1983	112.5227	135.5359	-4.2533
68	1984	112.4520	135.4749	6.6849
69	1985	112.3811	135.4141	7.7111
70	1986	112.3100	135.3534	5.1200
71	1987	112.2387	135.2930	1.8672
72	1988	112.1672	135.2328	-5.3142
73	1989	112.0955	135.1727	-1.7907
74	1990	112.0236	135.1129	-8.8463
75	1991	111.9515	135.0533	-2.4993
76	1992	111.8791	134.9938	11.8319
77	1993	111.8066	134.9346	-9.8279
78	1994	111.7339	134.8756	-3.3454
79	1995	111.6610	134.8168	-6.6816
80	1996	111.5879	134.7581	3.3496
81	1997	111.5146	134.6997	-2.6706
82	1998	111.4411	134.6415	-12.3368
83	1999	111.3674	134.5835	-3.0674
84	2000	111.2934	134.5256	-0.4598
85	2001	111.2193	134.4680	5.6983
86	2002	111.1450	134.4106	5.1163
87	2003	111.0705	134.3533	-2.9463
88	2005	110.9209	134.2394	.

Sum of Residuals	0
Sum of Squared Residuals	2787.93204
Predicted Residual SS (PRESS)	2921.16963