MATH 160G   Introduction to Applied Statistics   Spring 2008   Exam #4

Instructions: This exam is a tool to help me (and you) assess how well you are learning the course material. As such, you should report enough written detail for me to understand how you are thinking about each problem.   (100 points total)

---

1. You have some friends who are considering going to Lewis and Clark College because they think it rains less in Portland than in Tacoma. You decide to get some data to help address this question. The accompanying sheet gives data on annual precipitation (in inches) for Tacoma and Portland over the 41 years from 1966 to 2006 along with various plots and summary statistics.

   (a) Explain why it is best to treat this as a matched pairs design.   (3 points)

   (b) Describe the population and parameter of interest in this situation.   (4 points)

   (c) Explain why the t procedures are reasonable for this situation.   (4 points)

   (d) Set up null and alternative hypotheses for this situation.   (4 points)

   (e) Carry out a significance test on your hypotheses from (d).   (8 points)

   (f) Write a brief e-mail (a few sentences) to your friends giving your conclusions on the issue of whether it rains more in Tacoma than in Portland.   (4 points)

2. A new method for measuring phosphorus levels in soil is described in the paper "A Rapid Method to Determine Total Phosphorus in Soils" (*Soil Sci. Soc Am. J.*, **52**: 1301-1304). To test the accuracy of the new method, it is used on a sample of 11 soil specimens that have been prepared with a known phosphorus content of 548 mg/kg. Phosphorus level measurements for the 11 specimens using the new method have a mean of 587 mg/kg and a standard deviation of 10 mg/kg.

   (a) What is the population and parameter of interest in this situation?   (4 points)

   (b) Compute a 95% confidence interval for the mean phosphorus level measured by the new method.   (8 points)

   (c) What do your results for (b) tells us about the new method of measuring phosphorus levels? Write a non-technical conclusion based on the results of your confidence interval calculation.   (4 points)

   (d) Suppose you could read the original article that reports on this study. What information would you look for in judging whether the calculation you did in (b) is appropriate?   (4 points)

3. You work for a polling company and your supervisor asks you to determine a sample size that will give the current approval rating for Congress with 95% confidence and a margin of error no more that 3%. You want to use as small a sample as possible in order to keep costs down. Historically, the approval rating for Congress is rarely more than 40%. What sample size would you recommend?   (8 points)

4. A recent research article reports on a study that gathered data on the diets of pregnant women and the sex of their babies. The article reports that 721 women "with normal singleton pregnancies ... gave a retrospective report of their usual diet in the year prior to conception ('preconception' data)." Total calories per day was computed for each woman and used to split the group into three smaller groups: high energy intake, moderate energy intake, and low energy intake. Of the 241 women in the high energy intake group, 135 had boys. Of the 240 women in the low energy intake group, 108 had boys.

   Carry out some relevant statistical inference (confidence interval or significance test, your choice) on this data to address the question of whether preconception diet influences sex of the child. Include a check of conditions that the method you choose is appropriate. Give a non-technical conclusion written as a complete sentence (or two).   (15 points)

5. Do piano lessons improve the spatial-temporal reasoning of preschool children? A study designed to address this question measured the change in scores on a spatial-temporal reasoning test given twice, once at the beginning of a six-month period and once at the end. The tests were given to children in two groups. In one group, 34 children took piano lessons during the six-month period. In the other group, 44 children did not take piano lessons. Data, histograms, and summary statistics for these groups are given on the accompanying sheet.

   (a) What are the populations and parameter of interest in this situation?   (4 points)

   (b) Explain why t procedures are appropriate in this situation.   (4 points)

   (c) Compute a 95% confidence interval for the difference in means.   (8 points)

   (d) Write a lead sentence for a newspaper story on this study that gives the main conclusion.   (4 points)

6. (a) Sketch a plot showing the standard normal distribution $N(0, 1)$ and the t distribution $t(3)$ in enough detail to show the essential differences between the two distributions.   (6 points)

   (b) Use your plot from (a) to explain why t^* with $df = 3$ is bigger than z^* for the same confidence level, say 95%.   (4 points)

Data for Problem 1

Annual precipitation (inches)

Year	Tacoma	Portland	Difference
1966	34.95	33.86	1.09
1967	33.65	29.24	4.41
1968	44.97	50.89	-5.92
1969	32.07	37.11	-5.04
1970	36.81	41.35	-4.54
1971	38.78	41.72	-2.94
1972	46.08	38.82	7.26
1973	35.24	41.67	-6.43
1974	38.46	40.28	-1.82
1975	42.96	39.06	3.90
1976	27.70	26.71	0.99
1977	32.53	37.03	-4.50
1978	35.94	30.62	5.32
1979	36.56	35.75	0.81
1980	40.80	42.41	-1.61
1981	40.31	34.29	6.02
1982	27.79	43.04	-15.25
1983	42.86	47.19	-4.33
1984	41.87	37.50	4.37
1985	24.94	22.48	2.46
1986	43.35	35.04	8.31
1987	33.94	29.91	4.03
1988	38.93	31.72	7.21
1989	35.16	30.05	5.11
1990	46.87	32.86	14.01
1991	34.71	33.55	1.16
1992	31.88	29.50	2.38
1993	29.29	30.36	-1.07
1994	38.12	34.96	3.16
1995	43.86	43.48	0.38
1996	53.27	63.20	-9.93
1997	39.59	43.81	-4.22
1998	40.01	46.02	-6.01
1999	47.46	38.88	8.58
2000	20.66	30.20	-9.54
2001	40.71	30.44	10.27
2002	24.65	31.25	-6.60
2003	40.29	37.51	2.78
2004	32.04	27.65	4.39
2005	34.91	36.10	-1.19
2006	48.07	43.00	5.07

Histogram of Difference

Normal Quantile Plot of Difference
Normal - 95% CI
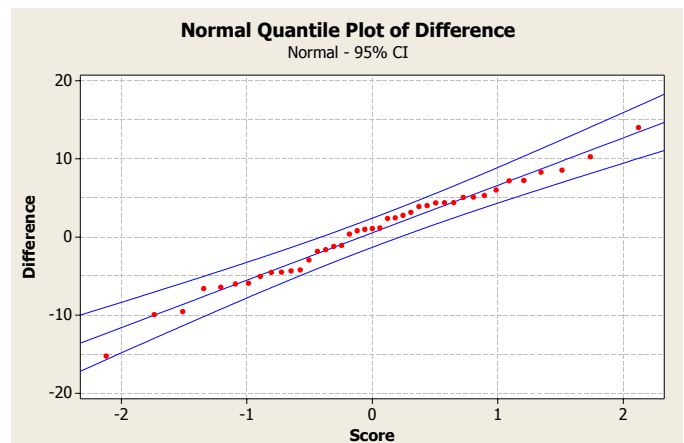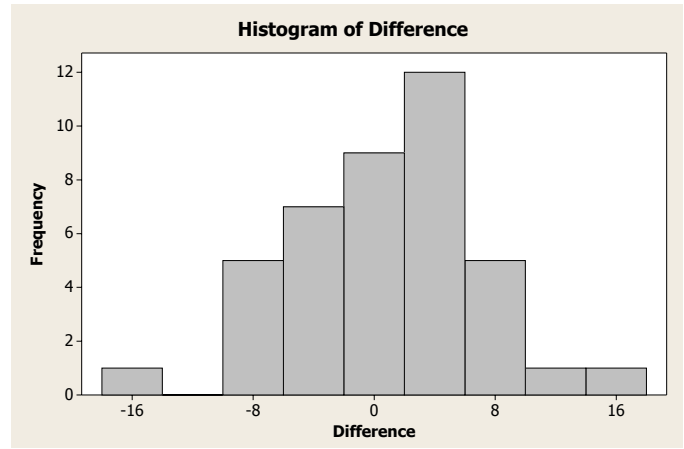
Summary statistics

City	Mean	StDev	Min	Q1	Median	Q3	Max
Tacoma	37.39	6.94	20.66	33.09	38.12	42.37	53.27
Portland	36.84	7.50	22.48	30.53	36.10	41.70	63.20
Difference	0.550	6.07	-15.25	-4.42	1.09	4.74	14.01

Data for Problem 5

Pre-test/Post-test difference for Piano group

2	5	7	-2	2	7	4	1	0	7	3	4	3	4	9	4	5
2	9	6	0	3	6	-1	3	4	6	7	-2	7	-3	3	4	4

Pre-test/Post-test difference for Control group

1	-1	0	1	-4	0	0	1	0	-1	0	1	1	-3	-2
4	-1	2	4	2	2	2	-3	-3	0	2	0	-1	3	-1
5	-1	7	0	4	0	2	1	-6	0	2	-1	0	-2	

Histogram of Piano Group

Histogram of Control Group
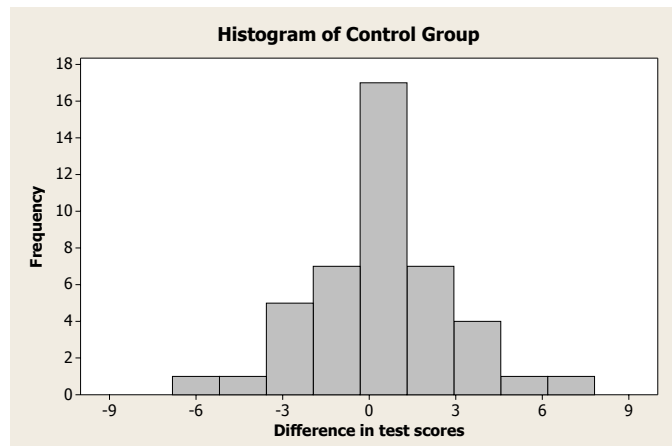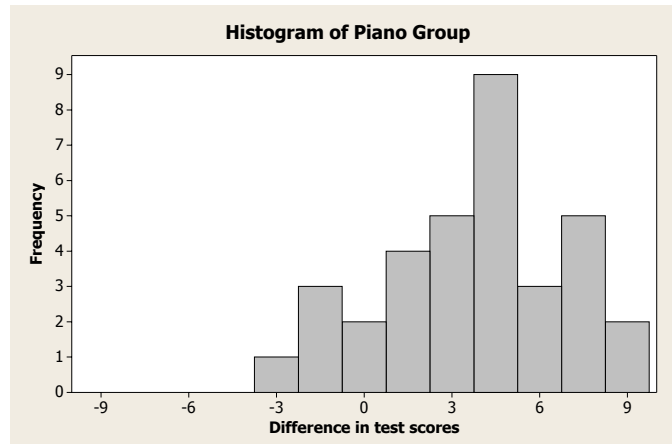
Summary statistics

Group	N	Mean	StDev	Min	Q1	Median	Q3	Max
Piano	34	3.618	3.055	-3	2	4	6	9
Control	44	0.386	2.423	-6	-1	0	2	7