Name _____

MATH 160A Introduction to Applied Statistics Spring 2008 Exam #1

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. (120 points total)

---

1. A researcher plans to gather data on college students in order to get insight on how college students will vote in the 2008 presidential election.

   (a) List two categorical variables and two quantitative variables that might reasonably be considered relevant for this purpose. (4 points)

   (b) For each of the four variables you list in (a), make up one reasonable value (including units if relevant). (4 points)

2. In a study of how much time people who live in Tacoma spend commuting to work, the following commute times (in minutes) are reported:

   10 30 5 25 40 20 10 15 30 20 15 20 85 15 65 15 60 60 40 45

   (a) Create a stemplot for this distribution. (5 points)

   (b) Give the five-number summary for this distribution. (6 points)

   (c) Use the 1.5×IQR rule to identify any potential outliers. (4 points)

   (d) Describe the main features of this distribution. (6 points)

   (e) Write a brief (one or two sentence) summary of what you learn about commute times from this distribution. Give your summary using "everyday" language rather than technical language. (3 points)

3. The heights of women aged 20 to 29 have a distribution that is approximately normal with mean 64 inches and standard deviation 2.7 inches.

   (a) Sketch this normal distribution. Mark relevant values on the horizontal scale. (5 points)

   (b) On your plot from (a), shade the area that corresponds to heights greater than 65 inches. (2 points)

   (c) What proportion of heights for women in this age group are greater than 65 inches? (5 points)

   (d) Without computing a z-score, give the percentage of women in this age group that are shorter than 66.7 inches. (5 points)

   (e) What height corresponds to being taller than 25% and shorter than 75% of women in this age group? (4 points)

Note: The remaining problems relate to the data and other information given on a separate sheet. The data gives annual precipitation (in inches) for each of Tacoma and Seattle for the years from 1966 to 2006.

4. Focus on the distribution of annual precipitation for Tacoma.

   (a) Show how to compute the mean for this distribution. (4 points)

   (b) Show how to compute the standard deviation for this distribution. (4 points)

   (c) Use the given summary statistics and histogram to describe main features of the distribution. (6 points)

   (d) What does the given normal quantile plot tell us about the Tacoma annual precipitation distribution? (3 points)

5. Now compare the distributions of annual precipitation for Tacoma and for Seattle. Using the given information, address the question "Does it rain more in Tacoma than in Seattle?" Give concrete evidence to support any conclusions you make. (8 points)

6. Now consider a possible association between the annual precipitation in Tacoma and the annual precipitation in Seattle. Refer to the given scatterplot for these variables.

   (a) In the scatterplot, which variable is the explanatory (or predictor) variable and which is the response variable? In this case, would it be reasonable to switch these? (4 points)

   (b) Describe any association evident in the scatterplot. (6 points)

   (c) The correlation for Seattle and Tacoma is $r = 0.836$. Show how to compute this correlation. (4 points)

   (d) Describe what this correlation value tells us about any association between the annual precipitation in Tacoma and the annual precipitation in Seattle. (3 points)

7. Let X be the annual precipitation in Tacoma and Y be the annual precipitation in Seattle. For the given data, the formula for the least-squares regression line is

$$\hat{y} = 9.05 + 0.758x.$$

   (a) Show how to compute the slope and intercept of this least-squares regression line. (4 points)

   (b) What fraction of the variation in Seattle's annual precipitation is explained by least-squares regression on Tacoma's annual precipitation? (3 points)

   (c) What does the given residual plot tell us about this least-squares regression? (4 points)

   (d) Use the least-squares regression line to predict the precipitation total in Seattle for a year in which Tacoma receives 36 inches of precipitation. (4 points)

   (e) Explain why it would be unwise to use the least-squares regression line to predict the precipitation total in Seattle for a year in which Tacoma receives 100 inches of precipitation. (3 points)

8. Suppose we get precipitation data for 2007 with a value of 50 inches for Tacoma and 25 inches for Seattle.

   (a) Add a point to the scatterplot for these values. (2 points)

   (b) If we calculate a new correlation with these values included in the data, how would the new correlation compare with the original correlation? Explain how you reach your conclusion. (5 points)

Annual precipitation (inches)

Year	Tacoma	Seattle
1966	34.95	38.23
1967	33.65	35.58
1968	44.97	50.15
1969	32.07	33.73
1970	36.81	37.41
1971	38.78	43.21
1972	46.08	48.36
1973	35.24	35.04
1974	38.46	37.87
1975	42.96	44.48
1976	27.70	26.70
1977	32.53	32.84
1978	35.94	33.99
1979	36.56	32.26
1980	40.80	35.60
1981	40.31	35.40
1982	27.79	39.32
1983	42.86	40.93
1984	41.87	36.99
1985	24.94	25.13
1986	43.35	38.34
1987	33.94	29.93
1988	38.93	32.98
1989	35.16	34.69
1990	46.87	44.75
1991	34.71	35.42
1992	31.88	32.78
1993	29.29	28.80
1994	38.12	34.82
1995	43.86	42.60
1996	53.27	50.67
1997	39.59	43.26
1998	40.01	44.06
1999	47.46	42.11
2000	20.66	28.66
2001	40.71	37.56
2002	24.65	31.36
2003	40.29	41.78
2004	32.04	31.10
2005	34.91	35.44
2006	48.07	48.42

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
Seattle	37.38	6.29	25.13	32.91	35.60	42.36	50.67