**AP Statistics** **Chapter 1-3 (+a little 4) Review**

**1.** Wetlands offer a diversity of benefits. They provide habitat for wildlife, spawning grounds for U.S. commercial fish, andrenewable timber resources. In the last 200 years, the United States has lost more than half its wetlands. *Environmental* *Almanac* gives the percentage of wetlands lost in each state in the last 200 years. For the lower 48 states, the percentage lossof wetlands per state is as follows:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 46 | 37 | 36 | 42 | 81 | 20 | 73 | 59 | 35 | 50 | 87 | 52 |
| 24 | 27 | 38 | 56 | 39 | 74 | 56 | 31 | 27 | 91 | 46 | 9 |
| 54 | 52 | 30 | 33 | 28 | 35 | 35 | 23 | 90 | 72 | 85 | 42 |
| 59 | 50 | 49 | 48 | 38 | 60 | 46 | 87 | 50 | 89 | 49 | 67 |

1. Make a stemplot, boxplot, and histogram.
2. Describe the distribution (the **graph** of the distribution).
3. Find the numerical information: mean, standard deviation, five number summary, outliers, etc.
4. Describe the relationship between the numerical information and the graphical information in context of this problem.
5. I have a data set consisting of 33 whole number observations. Its five number summary is: (16, 20, 22, 30, 46). a) What is the range of the data?

b) How many observations are strictly less than 22?

c) Is it possible that there are no observations equal to 22? Explain.

d) What is the IQR?

e) Show the test for outliers. Are there any outliers?

1. The average playing time of compact discs in a large collection is 35 minutes with a standard deviation of 5 minutes. Assume a normal distribution.
2. Draw the distribution for this problem.
3. What values are at one standard deviation above and below the mean?
4. Assuming a normal distribution, at least what percentage of the times are between 25 and 45 minutes?
5. What value is at the 84th percentile? The 50th percentile?
6. What can be said about the percentage of times that are either less than 20 minutes or greater than 50 minutes?
7. At a ski area in Vermont, the daytime high temperature is normally distributed during January, with a mean of 22 degrees and a standard deviation of 10 degrees (U.S. Department of Commerce, *Environmental Data Services*). You are planning to ski there this January. What is the probability that you will encounter daytime highs of:

a) 42 degrees or higher?

b) 15 degrees or lower?

c) between 29 and 40 degrees?

1. Eleanor scores 680 on the mathematics part of the SAT. The distribution of SAT scores is normal with a mean of 500 and standard deviation 100. Gerald takes the ACT mathematics test and scores 27. ACT scores are normally distributed with mean 18 and standard deviation 6. Assuming both tests measure the same kind of ability, who has the higher score? Explain.
2. According to *The New York Times* (April 2, 1993), the average monthly rate for basic television cable service has increased as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year: 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 |
| Cost($): 11.00 | 13.20 | 13.90 | 15.20 | 16.80 | 18.00 | 20.00 |

1. Draw a scatterplot for the data.
2. Find the equation of the least squares regression line.
3. Slope = \_\_\_\_\_\_\_. Interpret the slope in context.
4. y-intercept = \_\_\_\_\_\_\_. Interpret the y-intercept in context.
5. r = \_\_\_\_\_\_. Interpret the correlation coefficient in context.
6. r2 = \_\_\_\_\_\_. Interpret the coefficient of determination in context.
7. Predict the average monthly rate in 2007.
8. Predict in what year the rate will reach $50.
9. Do you feel confident in the prediction you made in part g? Explain.
10. Find the residuals for each year and make a residual plot. Use the plot to determine whether the LSRL is a good model to predict the cost of basic cable service.
11. Ball players have been signing ever larger contracts. The highest salaries (in millions of dollars per season) for some notable players are given in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Year | Salary (million $) |
| Ryan | | 1980 | 1 |
| Foster | | 1982 | 2.04 |
|  | Puckett | 1990 | 3 |
|  | Canseco | 1990 | 4.7 |
| Clemens | | 1991 | 5.3 |
| Griffey, Jr. | | 1996 | 8.5 |
| Belle | | 1997 | 11 |
| Martinez | | 1998 | 12.5 |
|  | Piazza | 1999 | 12.5 |
| Vaughn | | 1999 | 13.3 |
| Brown | | 1999 | 15 |
|  | Delgado | 2001 | 17 |
| Rodriguez | | 2001 | 25.2 |

1. Draw a scatterplot. Is it linear?
2. The LSRL after taking the logarithms of Salary is: Log(Salary)= -113.47 +0.057(Year) Convert that to an exponential equation.
3. Predict a superstar salary in 2006.
4. Alfonso Soriano just signed a new contract for an average of $16.5 million. Does that match your prediction from part (c)? Explain.