

STATISTICS 101 - Homework 1

Due Friday, January 25, 2008

Homework is due on the due date at the end of lecture.

Readings

- Chapters 3 - 5

Assignment

1. On the night of April 14, 1912, the luxury liner *RMS Titanic* hit an iceberg and sank in the North Atlantic Ocean. In the popular movie about this disaster, first class passengers appeared to be able to get to the life boats, while third class passengers were kept away. Is there truth to this appearance? What effect did the class of ones ticket make on whether or not they were rescued? Below is a table listing the number of people rescued and lost according to the class of ticket. The crew are also included in this tally.

	First	Second	Third	Crew	Total
Rescued	202	118	178	212	711
Lost	123	167	528	673	1490
	325	285	706	885	2201

- (a) Describe the who and what of this data set.
 - (b) Determine the marginal distribution of the Class of Ticket variable.
 - (c) Determine the marginal distribution of the Rescued status variable.
 - (d) What percentage of Rescued passengers held First Class tickets? How does this compare to the percentage of First Class passengers on the ship?
 - (e) What percentage of Rescued passengers held Third Class tickets? How does this compare to the percentage of Third Class passengers on the ship?
 - (f) Based on your answers to parts (d) and (e) above, do you think the perception from the movie is correct? Explain your answer. (We will revisit this problem at the end of the semester).
2. The daily high temperatures recorded at Des Moines International Airport during the months of January 2004 and 2007 are listed in the table below.

Day	Temp 2004	Temp 2007	Day	Temp 2004	Temp 2007	Day	Temp 2004	Temp 2007
1	52	34	12	45	22	22	19	28
2	60	41	13	35	18	23	53	30
3	34	53	14	40	24	24	25	34
4	21	50	15	38	20	25	27	26
5	11	45	16	34	11	26	21	45
6	16	45	17	34	22	27	15	29
7	26	39	18	27	29	28	5	12
8	31	47	19	16	31	29	2	28
9	26	37	20	23	28	30	2	8
10	31	51	21	46	30	31	14	23
11	49	51						

- (a) Describe the who and what of this data set.
 - (b) Make a back-to-back stem-and-leaf display of the high temperatures for the months of January 2004 and 2007.
 - (c) Describe the distribution of the high temperatures for each month. Make sure to mention the shape, center, spread and any outliers for each distribution. Then compare and contrast the distribution of the high temperatures for the two months.
3. **JMP Assignment** This problem requires you to use JMP to analyze the distribution of the total annual rainfall (in inches) for the 100 year period from 1902-2001 for Los Angeles, California. The data set is located on the main course webpage as **Annual Rainfall for Los Angeles from 1902-2001**. Follow the instructions in the JMP Guide to download the data set from the web, open the data set in JMP, and obtain a histogram, stem-and-leaf display and descriptive statistics for this variable. When you are finished, print out the output from JMP and turn it in with your assignment. Use this output to answer the following questions.
- (a) Describe the who and what of this data set.
 - (b) Describe the distribution of annual rainfall in Los Angeles from 1902-2001. Make sure to mention the shape, center, spread and any outliers.
 - (c) In how many years was the annual rainfall in Los Angeles greater than 30 inches?
 - (d) In how many years was the annual rainfall in Los Angeles between 20 and 30 inches?
 - (e) How do you think the distribution of annual rainfall in Des Moines, Iowa from 1902-2001 would compare to the distribution of annual rainfall in Los Angeles? Would it be similar or different? Explain your answer.
4. **JMP Assignment** This problem requires you to use JMP to analyze the distribution of the time (in minutes) between eruptions of Old Faithful in Yellowstone National Park. The data set is located on the main course webpage as **Time between Eruptions of Old Faithful**. Follow the instructions in the JMP Guide to download the data set from the web, open the data set in JMP and obtain a histogram, stem-and-leaf display and descriptive statistics for this variable. When you are finished, print out the output from JMP and turn it in with your assignment. Use this output to answer the following questions.
- (a) Describe the who and what of this data set.
 - (b) In the description of this data set above, some important information is missing. What other information do you need?
 - (c) Does the stem-and-leaf display use split stems? If yes, how were the stems split?
 - (d) Describe the distribution of times between eruptions of Old Faithful. Make sure to mention the shape, center and spread.
 - (e) This data set contains one outlier. Which observation is the outlier? Why is this observation an outlier? Which display did you need to use to find the outlier, the histogram or the stem and leaf display?
 - (f) What are the mean and median times between eruptions of Old Faithful? Which one is larger?