Tutorial 1: ECEN321 Engineering Statistics

(answers are on page 834 of William Navidi)

Section 1.1 Sampling

1. Each of the following processes involves sampling from a population. Define the population, and state whether it is tangible or conceptual.

a. A chemical process is run 15 times, and the yield is measured each time.

b. A pollster samples 1000 registered voters in a certain state and asks them which candidate they support for governor.

c. In a clinical trial to test a new drug that is designed to lower cholesterol, 100 people with high cholesterol levels are recruited to try the new drug.

d. Eight concrete specimens are constructed from a new formulation, and the compressive strength of each is measured.

e. A quality engineer needs to estimate the percentage of bolts manufactured on a certain day that meet a strength specification. At 3:00 in the afternoon he samples the last 100 bolts to be manufactured.

3. True or false:

a. A simple random sample is guaranteed to reflect exactly the population from which it was drawn.

b. A simple random sample is free from any systematic tendency to differ from the population from which it was drawn.

5. A certain process for manufacturing integrated circuits has been in use for a period of time, and it is known that 12% of the circuits it produces are defective. A new process that is supposed to reduce the proportion of defectives is being tested. In a simple random sample of 100 circuits produced by the new process, 12 were defective.

a. One of the engineers suggests that the test proves that the new process is no better than the old process, since the proportion of defectives in the sample is the same. Is this conclusion justified? Explain.

b. Assume that there had been only 11 defective circuits in the sample of 100. Would this have proven that the new process is better? Explain.

c. Which outcome represents stronger evidence that the new process is better: finding 11 defective circuits in the sample, or finding 2 defective circuits in the sample?

Section 1.2 Summary Statistics

1. True or false: For any list of numbers, half of them will be below the mean.

3. Is the sample mean always equal to one of the values in the sample? If so, explain why. If not, give an example.

5. Find a sample size for which the median will always equal one of the values in the sample.

7. Is it possible for the standard deviation of a list of numbers to equal 0? If so, give an example. If not, explain why not.

11. In a sample of 20 men, the mean height was 178 cm. In a sample of 30 women, the mean height was 164 cm. What was the mean height for both groups put together?

Section 1.3 Graphical Summaries

1. The weather in Los Angeles is dry most of the time, but it can be quite rainy in the winter. The rainiest month of the year is February. The following table presents the annual rainfall in Los Angeles, in inches, for each February from 1965 to 2006.

0.2	3.7	1.2	13.7	1.5	0.2	1.7
0.6	0.1	8.9	1.9	5.5	0.5	3.1
3.1	8.9	8.0	12.7	4.1	0.3	2.6
1.5	8.0	4.6	0.7	0.7	6.6	4.9
0.1	4.4	3.2	11.0	7.9	0.0	1.3
2.4	0.1	2.8	4.9	3.5	6.1	0.1

a. Construct a stem-and-leaf plot for these data.

b. Construct a histogram for these data.

c. Construct a dotplot for these data.

d. Construct a boxplot for these data. Does the boxplot show any outliers?

5. A certain reaction was run several times using each of two catalysts, A and B. The catalysts were supposed to control the yield of an undesirable side product. Results, in units of percentage yield, for 24 runs of catalyst A and 20 runs of catalyst B are as follows:

Catalyst A			
4.4	3.4	2.6	3.8
4.9	4.6	5.2	4.7
4.1	2.6	6.7	4.1
3.6	2.9	2.6	4.0
4.3	3.9	4.8	4.5
4.4	3.1	5.7	4.5

Catalyst B			
3.4	1.1	2.9	5.5
6.4	5.0	5.8	2.5
3.7	3.8	3.1	1.6
3.5	5.9	6.7	5.2
6.3	2.6	4.3	3.8

a. Construct a histogram for the yields of each catalyst.

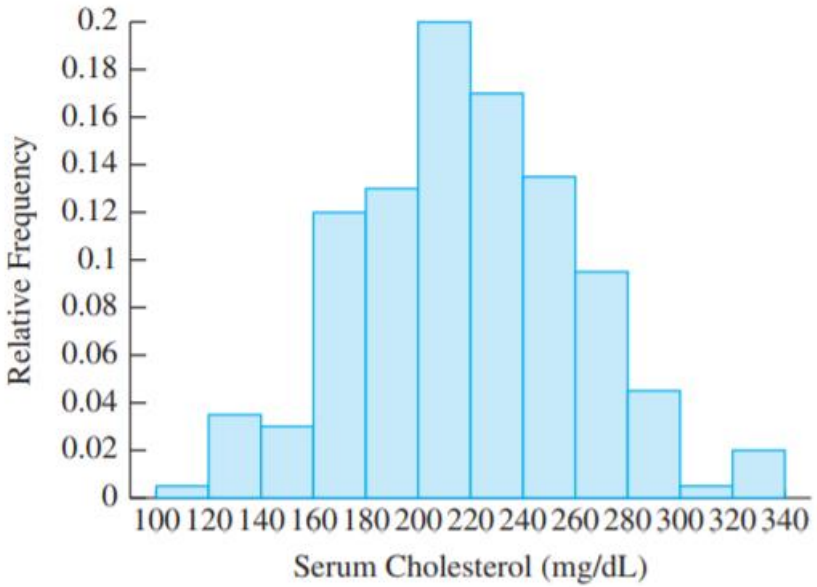
b. Construct comparative boxplots for the yields of the two catalysts.

c. Using the boxplots, what differences can be seen between the results of the yields of the two catalysts?

7. The figure below is a histogram showing the distribution of serum cholesterol level for a sample of men. Use the histogram to answer the following questions:

a. Is the percentage of men with cholesterol levels above 240 mg/dL closest to 30%, 50%, or 70%?

b. In which interval are there more men: 240–260 mg/dL or 280–340 mg/dL?



11. The following table presents the number of students absent in a middle school in northwestern Montana for each school day in January 2008.

Date	Number Absent	Date	Number Absent	Date	Number Absent
Jan. 2	65	Jan. 14	59	Jan. 23	42
Jan. 3	67	Jan. 15	49	Jan. 24	45
Jan. 4	71	Jan. 16	42	Jan. 25	46
Jan. 7	57	Jan. 17	56	Jan. 28	100
Jan. 8	51	Jan. 18	45	Jan. 29	59
Jan. 9	49	Jan. 21	77	Jan. 30	53
Jan. 10	44	Jan. 22	44	Jan. 31	51
Jan. 11	41				

a. Construct a boxplot.

b. There was a snowstorm on January 27. Was the number of absences the next day an outlier?

15. Following are summary statistics for two data sets, A and B.

	A	B
Minimum	0.066	−2.235
1st Quartile	1.42	5.27
Median	2.60	8.03
3rd Quartile	6.02	9.13
Maximum	10.08	10.51

a. Compute the interquartile ranges for both A and B.

b. Do the summary statistics for A provide enough information to construct a boxplot? If so, construct the boxplot. If not, explain why.

c. Do the summary statistics for B provide enough information to construct a boxplot? If so, construct the boxplot. If not, explain why.

19. For the following data:

x	1.4	2.4	4.0	4.9	5.7	6.3	7.8	9.0	9.3	11.0
y	2.3	3.7	5.7	9.9	6.9	15.8	15.4	36.9	34.6	53.2

a. Make a scatterplot of y versus x. Is the relationship between x and y approximately linear, or is it nonlinear?

b. Compute the natural logarithm of each y value. This is known as *making a log transformation of* y. Make a scatterplot of $\ln y$ versus x. Is the relationship between x and $\ln y$ approximately linear, or is it nonlinear?

c. In general, it is easier to work with quantities that have an approximate linear relationship than with quantities that have a nonlinear relationship. For these data, do you think it would be easier to work with x and y or with x and $\ln y$? Explain.