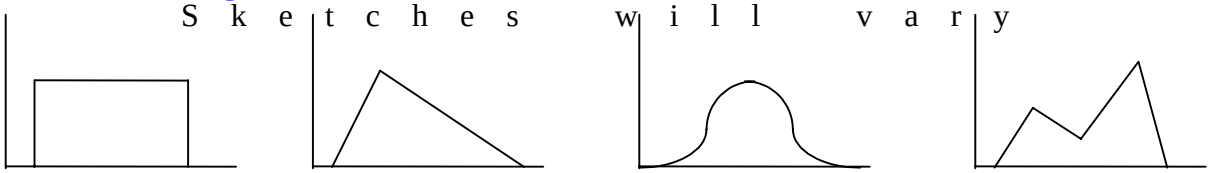
Solutions to Worktext Exercises

Chapter 6

Visualizing Random Samples

Basic Learning Exercises

1. Sketches will vary



a) Five or six intervals best illustrate most samples. b) The sketches will differ in terms of the range, the number of modes, the location of the mode(s), the degree of skewness, and the degree of peakedness (kurtosis). Most of the time you will see a single mode, but sample variation causes the histograms to differ. c) A sample of 25 is not large enough to ensure stability in the histograms.

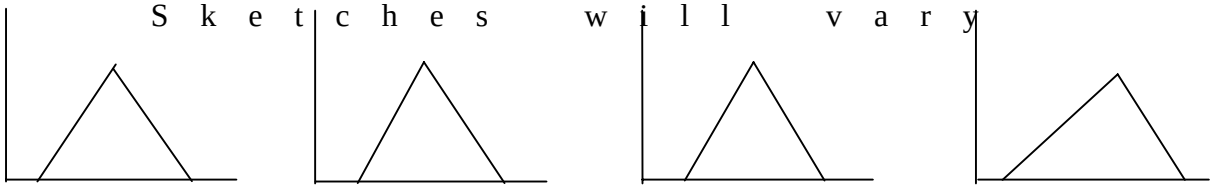
2.

Sample	Mean	Std. Dev.	Skewness	Kurtosis	Median
1	0.996	0.406	-0.525	2.358	1.137
2	0.944	0.397	0.648	2.455	0.871
3	1.144	0.316	0.084	1.956	1.098
4	0.903	0.405	0.459	1.859	0.880
Range	0.241	0.090	1.173	0.596	0.257

The mean will vary between 0.75 and 1.25, standard deviation between 0.22 and 0.58, skewness between ±.6, kurtosis between 1.5 and 4, and median between 0.70 and 1.30.

3. Histograms display the sample with all its variation. The mean and median measure only the center of the distribution, and hence sample variation above and below the center are canceled out.

4. Sketches will vary



All of the histograms suggest a unimodal symmetric distribution. Sample variation is reduced because of the very large sample size.

5.

Sample	Mean	Std. Dev.	Skewness	Kurtosis	Median
1	1.013	0.389	0.145	2.534	1.040
2	1.009	0.411	-0.201	2.333	0.959
3	0.999	0.423	0.125	2.202	0.950
4	0.954	0.404	0.110	2.299	1.005
Range	0.059	0.034	0.346	0.332	0.090

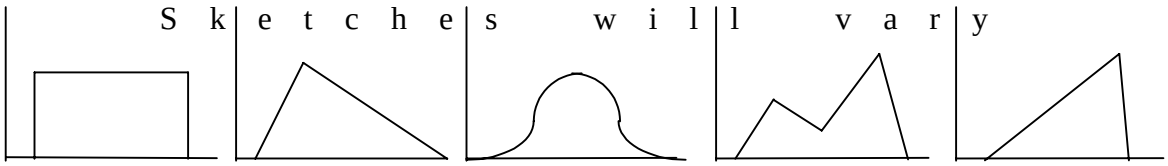
The mean will vary between 0.9 and 1.1, standard deviation between 0.38 and 0.45, skewness between ±0.3, kurtosis between 2 and 2.5, and median between 0.93 and 1.07.

6. a) The range is less in exercise 4 (n = 500) than in exercise 2 (n = 25). b) The sketches are alike in exercise 4 and erratic in exercise 1. c) Relative frequencies in a histogram become more stable as sample size increases, which leads to less variation in histograms and less variability in sample statistics. d) Your confidence would increase because in a large sample it is unlikely that a different sample would change your impressions.

7. Based upon the sample of 25, your impressions may have been inaccurate, while those from the sample of 500 would be very accurate. The population you have been sampling is a triangular distribution with a mode of 1, a minimum of 0, and a maximum 2. Its standard deviation is 0.408, skewness is 0, kurtosis is 2.4, and median is 1.

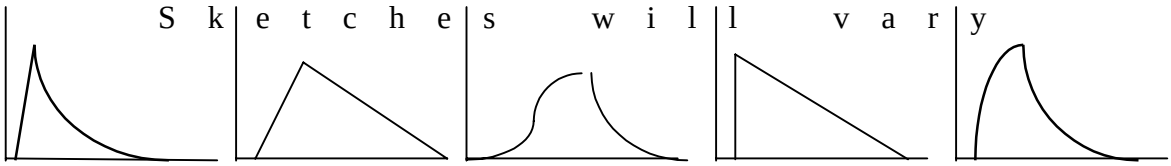
Intermediate Learning Exercises

8. IQ Test Scores: Population Distribution Normal



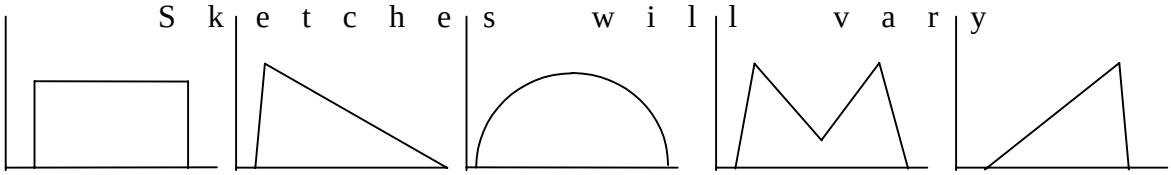
It is likely that some of the sample histograms did not look very much like a normal, though they typically have a single mode and appear somewhat symmetric.

9. Accuracy of an Archer: Population Distribution Skewed Right



Most of the samples will be unimodal and skewed right, like the population.

10. Choosing a Lottery Number: Population Distribution Uniform



It is likely that several of the sample histograms did not look very much like a uniform. Some may have two modes, others may appear asymmetric.

11. The Lottery Number scenario is most likely and Accuracy of an Archer scenario is least likely. In a uniform population (lottery) each data point has an equal chance of being drawn, so any of these histograms could occur. In a right skewed population (archery), large population values are unlikely, so the middle histogram would be unlikely.

12. There is no single correct answer. The expenditure data has N = 500 observations, while the grocery data has N = 200 observations.

13-15. a) For the smaller sample, the ratio is about 0.8. For the larger sample, the ratio is about 0.4. b) The ratio is smaller if the sample size (n) is a larger percentage of the population size (N). c) Ultimately, if n = N the sample is the entire population, so the sample mean is the population mean for every sample and the range would be zero.

Advanced Learning Exercises

16. Answers will vary. The population should be described in terms of its centrality, dispersion, symmetry, peakedness, and number of modes. An example of data that might have this shape should be given. An explanation of why the example would have these same characteristics (skewness, peakedness, dispersion, mean, and modality) should be provided.

17. An outlier is an observation that is more than 3 standard deviations from the sample mean. It can occur if an observation came from a different population than the one you are sampling. It warns the statistician that he/she *may* have a biased sample.

18. An observation or two will be widely separated from the rest of the data displayed in the histogram. A long tail in the boxplot also will suggest the presence of an outlier.

19. The population being sampled ends before the interval containing the outlier.

20. No. An outlier could be from the same population as the other data points. But this is unlikely, because sample items 3 standard deviations from the sample mean are rare.

21. Normal Always bell-shaped distribution.

Chi Square	Positively skewed distribution, more symmetric as DF increases.
F	Positively skewed distribution, more symmetric as both DF increase.
Student's t	Symmetric distribution, less peaked as DF increases.
Gamma	Right skewed distribution, more symmetric as shape parameter increases.
Beta	Skewed left or right based on larger parameter, symmetric if equal.
Uniform	Rectangular distribution.
Exponential	Very positively skewed distribution.
Triangular	Symmetric, skewed right, or skewed left depending on mode.

22. About 25 light bulbs burn out in 250 hours, about 40 in 500 hours, and about 65 in 1,000 hours. The last light bulb usually burns out in 3,500 to 5,000 hours.

23. Student heights would be bimodal because men generally are taller than women. Another example would be exam scores for students who studied and those who didn't. In the student heights scenario, the two modes are usually apparent in the sample. No, a unimodal or uniform population can produce a sample with two modes.