Review Questions on Chapters 1—3

For further practice on questions from Chapters 1 to 3, see Test 1 from Spring 2007, Test 1 from Winter 2007, and Test 1 from Autumn 2003 as available at class website.

I.A. Levels of Data, Types of Samples

I.A.1. Categorize the following data according to level: nominal, ordinal, interval, or ratio.

(a) Time of first class.
(b) Length of time to complete an exam.
(c) Course evaluation scale: poor, acceptable, good.
(d) The length of time it takes for someone to run a marathon.
(e) The time of day the marathon starts.
(f) The name of the city in which the marathon is run.

I.A.2. To estimate the average GPA of all La Sierra Students, President Geraty computed the average GPA obtained in his Advanced Hebrew Grammar class.

(a) Identify the variable?
(b) What is the implied population?
(c) What is the sample?
(d) What type of sample was this?

I.A.3. (a) If your instructor were to compute the class mean of this test when it is graded, and use it to estimate the average for all tests taken by this class this quarter, would this be an example of descriptive or inferential statistics? Explain.

(b) A study on attitudes about smoking is conducted at a college. The students are divided by class, and then a random sample is selected from each class. What type of sampling technique is this (e.g. simple random, convenient, stratified, systematic, cluster)? Explain why this type of sample is not a simple random sample.

(c) A politician wishes to determine the reading level of 5th graders in her State. She does not have funding to test all 5th graders in her state, so she randomly selects some of the schools in her state and tests every 5th grader in those schools. What type of a sample is this?

I.A.4. To compute the average amount of medical insurance its patients had in 2006, a hospital considered taking the following types of samples. For each case classify the sample as simple random, stratified, systematic, cluster, convenient.

(a) The hospital numbered its complete list of patients from 2006 and used a random number generator to select 100 patients from the list from which to collect the data.

(b) The hospital decided to collect the data from the first 50 patients admitted on July 4, 2006.

(c) The hospital randomly chose a patient, and collected data from that patient and every 200th patient admitted thereafter.
I.A.5. Explain how you could use the table of random numbers in your text to help design a true false test of 10 questions so that the pattern of answers is random.

I.A.6. (a) Describe the difference between population data and sample data.
(b) Describe the difference between inferential and descriptive statistics.
(c) Describe the difference between quantitative and qualitative variables.

I.A.7. A medical school is investigating new eye drops as a treatment for glaucoma. Out of 63 volunteers, 35 will get the new eye drops. The others will get the currently used (not new) eye drops. The new eye drops come in a grey plastic bottle and the old eye drops come in a red plastic bottle. Neither the patient nor the doctor knows which color contains which eye drops. After six months, eye pressure on each patient is measured and then a sealed report revealing medication is opened.

(a) Explain how you would make this a randomized two treatment experiment.

(b) Is this a double-blind experiment? Explain.

I.A.8. (Focus Problem: Fireflies) Suppose you are conducting a study to compare firefly populations exposed to normal daylight/darkness conditions with firefly populations exposed to continuous light (24 hours a day). You set up two firefly colonies in a laboratory environment. The two colonies are identical except that one colony is exposed to normal daylight/darkness conditions and the other is exposed to continuous light.

Each colony is populated with the same number of mature fireflies. After 72 hours, you count the number of living fireflies in each colony.

(a) Is this an experiment or observation study? Explain.

(b) Is there a control group? Is there a treatment group? If so which is which?

(c) What is the variable in this study?

(d) What is the level of measurement of the variable?

I.B. Organizing and Presenting Data

I.B.1. Consider the data (which are systolic blood pressures of 50 subjects):

(a) What class width should be chosen if you would like to have 6 classes.

(b) Suppose you dont want a class width of 19, but would like a class width of 15 irrespective of how many classes that would give you. Complete the following table for this data.
(c) Draw a frequency histogram using the table in (b).
(d) Draw a frequency polygon using the table in (b).
(e) Draw an Ogive using the table in (b).
(f) Draw a relative frequency histogram for the table in (b).
(g) Describe in words how to construct each of the above mentioned types of graphs.

I.B.2. Consider the collection of 30 data

\[

data = \begin{array}{cccccccccccccccc}
2 & 3 & 4 & 4 & 5 & 6 & 8 & 9 & 11 & 12 & 14 & 15 & 18 & 19 & 21 \\
\end{array}
\]

Complete the following table given that the first class has limits 1 – 20.

<table>
<thead>
<tr>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Lower Boundary</th>
<th>Upper Boundary</th>
<th>Midpoint</th>
<th>Frequency</th>
<th>Cumulative Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I.B.3. Consider the following table
(a) Construct a relative frequency histogram using the information given in the table above.

(b) Construct an ogive using the information given in the table above.

(c) Construct a frequency histogram using the information given in the table above.

(d) Construct a frequency polygon using the information given in the table above.

I.B.4. Make a stem and leaf display for the following data.

58 52 68 86 72 66 97 89 84 91 91 92 66 68 87 86
73 61 70 75 72 73 85 84 90 57 77 76 84 93 58 47

I.C. Numerical Representations of Data

I.C.1. Consider the following sample consisting of 20 numbers.

20 23 24 24 25 26 28 29 31 32
33 34 35 37 41 46 52 53 58 98

(a) Given that $\sum x = 749$ and $\sum x^2 = 34109$ find the mean, variance and standard deviation for this sample.

(b) Find the $Q_1$, $Q_2$, $Q_3$ and the IQR for the data and then construct a box-and-whisker plot for the data.

I.C.2. Given the data 9,12,15,17,18,19,23,45,52,61,63,63. One has $\sum x = 397$ and $\sum(x - \mu)^2 \approx 5206.917$. Find:

(a) the mean; (b) the sample variance (c) the population variance

(d) the sample standard deviation (e) the median (f) the 50th percentile (g) the mode

I.C.3. The depth of ground water is given in the following grouped data table.

<table>
<thead>
<tr>
<th>Distance from ground to water level (ft), $x$</th>
<th>15 – 19</th>
<th>20 – 24</th>
<th>25 – 29</th>
<th>30 – 34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of wells, $f$</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

(a) Estimate the mean depth of the ground water.

(b) Estimate the sample standard deviation for the depth of the ground water.

(c) Estimate the coefficient of variation for this data.
I.C.4. (Short answer general) (a) Given a collection of data with lowest number 4 and highest number 100. What class width should be chosen if 6 classes are desired? (b) Given a collection of ordered data with 65 numbers, in what position is the median? (c) What percentile is the third quartile $Q_3$? (d) In a set of data, approximately what percentage of the data lie at or above the 63rd percentile? (e) If you are among 8000 people that took a test and you scored at the 79th percentile, approximately how many people scored at your score or lower? Approximately how many people scored at your score or higher? (f) If you are among 4000 students taking the MCAT, and you wish to score at least the 95th percentile, what is the maximum number of students that can score at least as well or better than you?

I.C.5. A population is known to have a mean of 50 and a standard deviation of 6. (a) Use Chebyshev’s theorem to find an interval that contains at least 75% of the data. (b) Use Chebyshev’s theorem to find an interval that contains at least 24/25 of the data. (c) At least what portion of data is contained in the interval from 26 to 74?

I.C.6. Consider the following data of 26 numbers. 8,35,47,48,51,57,60,64,64,65,66,70,72 76,78,80,82,84,85,89,90,90,93,94,96,111 (a) Find the median of the data. (b) Find $Q_1$, $Q_3$ and the IQR. Construct a box and whisker plot for the data. (c) Compute the interval $(Q_1 - 1.5\text{IQR}, Q_3 + 1.5\text{IQR})$. Data outside of this interval are identified as suspected outliers. Are there any suspected outliers in the above data?

I.C.7. (General Questions on Means) (a) A student receives grades of A,A,B,C,C and is surprised to receive of GPA of 3.5 because the average grade of A,A,B,C,C is a B since

$$\frac{4 + 4 + 3 + 2 + 2}{5} = \frac{15}{5} = 3.$$ Explain how the GPA could be 3.5. (b) Bob is pleased to learn from his boss that his annual salary is in the top 10 percent of all salaries in the company. A month later Bob learns that his salary is less than half of the mean salary in the company and suspects his boss lied to him. Explain how Bob’s boss could have told the truth. (c) An army crosses a river whose average depth is 1 foot. Explain how several soldiers could have drowned crossing the river because they cannot swim. (d) Ken had an average of 50% on exams and 90% on assignments in his class, so he computed that his average should be 70% (a C). Why was he shocked when he saw that his grade was a D?

I.C.8. (General Questions on Measures of Data) Consider a data set of 25 distinct measurements with mean $A$, median $B$, and range $R$. 


(a) If the largest number is increased by 250, what is the new mean?
(b) If the largest number is increased by 250, what is the new median?
(c) If the largest number is increased by 250, what is the new range?
(d) If every number is increased by 10, what is the effect on the mean, median, range and standard deviation?
(e) In what position of the ordered data is the median?
(f) Explain why the data does not have a mode.
(g) Is it possible for the mean and median to be equal?

I.C.9. (Weighed Average) Suppose the weightings in a class are such that the tests are worth 60% of the grade, the assignments are worth 10% of the grade, the quizzes are worth 5% of the grade, and the final is worth 25% of the grade.
(a) Calculate the overall grade of a student that has 90% on assignments, 95% on quizzes, 75% on tests and 80% on the final exam.
(b) Is the answer in (a) higher or lower than if you had taken the average of 90%, 95%, 75% and 80%? Explain why in this example you would expect the weighted average to be lower.
(c) Suppose a student has 90% on assignments, 95% on quizzes, 75% on tests going into the final exam. What percentage is needed on the final exam for the student to finish with an overall grade of 82% in the class?

I.C.10. On-time percentages are given for two airlines in Phoenix, Los Angeles and Seattle for 2006.

<table>
<thead>
<tr>
<th>Airline</th>
<th>Los Angeles</th>
<th>Phoenix</th>
<th>Seattle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crashcade Airlines</td>
<td>1000</td>
<td>500</td>
<td>3500</td>
</tr>
<tr>
<td>On time %</td>
<td>90</td>
<td>95</td>
<td>85</td>
</tr>
<tr>
<td>Pacific Worst Airlines</td>
<td>250</td>
<td>4500</td>
<td>250</td>
</tr>
<tr>
<td>On time %</td>
<td>85</td>
<td>90</td>
<td>80</td>
</tr>
</tbody>
</table>

(a) Calculate the on-time percentage average for these three cities for each airline. Do this as a weighted average where the weight for each airline and city is the number of flights.
(b) Given that the on-time percentage for Crashcade Airlines is 5% higher in each city, does the answer in (a) surprise you? Why or why not?

I.C.11. (See the Histogram in figure 3-4, p. 133 of Text) The following data is for hours of sleep of a random sample of 200 subjects. Estimate the mean hours of sleep, standard deviation hours of sleep and coefficient of variation.

<table>
<thead>
<tr>
<th>Hours of Sleep</th>
<th>Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>2</td>
</tr>
<tr>
<td>4.5</td>
<td>2</td>
</tr>
<tr>
<td>5.5</td>
<td>4</td>
</tr>
<tr>
<td>6.5</td>
<td>22</td>
</tr>
<tr>
<td>7.5</td>
<td>64</td>
</tr>
<tr>
<td>8.5</td>
<td>90</td>
</tr>
<tr>
<td>9.5</td>
<td>14</td>
</tr>
<tr>
<td>10.5</td>
<td>2</td>
</tr>
</tbody>
</table>

I.C.12. Consider the following ordered data of 21 numbers

34 36 40 46 46 48 51 54 57 58 59
60 61 62 63 64 66 70 78 85 101
Find $Q_1$, $Q_2$, $Q_3$, the IQR and then construct a box and whisker plot for the data.