**Midterm Exam – IT223**February 8, 2010

**Part A. Multiple Choice Questions.** For each question, you may give a reason or show work for partial credit. You are required to show your work if the question is marked with an asterisk (\*). Questions 3, … all require you to show a reason or work.

1. Who is often known as the father of modern statistics?  
   a. Fisher b. Gauss c. Graunt d. Tukey
2. Who wrote a book that popularized exploratory data analysis, which includes using the boxplot and stemplot.  
   a. Fisher b. Gauss c. Graunt d. Tukey
3. A researcher collects questionnaire responses (response values 1 to 5). This is an example of what type of variable?  
   a. Categorical b. Continuous c. Discrete d. Ordinal
4. \*Which of the following could be most easily be studied using a controlled experiment?  
   a. Whether or not smoking causes lung cancer.  
   b. Who will win the Illinois governor’s election in November, 2010.  
   c. Whether  
   d. ????
5. What type of variable influences one or more variables in a dataset, but is not actually included in the dataset.  
   a. Dependent b. Explanatory c. Independent c. Lurking
6. \*The dataset 31 45 65 69 77 81 85 84 121 contains:  
   a. at least one extreme outlier but no mild outliers.  
   b. at least one extreme outlier but no extreme outliers.  
   a. at least one mild and at least one extreme outlier.  
   d. no outliers.
7. \*Compute SD+ for the following dataset: 3 7 9 11 15  
   a. 4.000 b. 4.472 c. 6.800 d. 20.000
8. If the dataset in Problem 4 is multiplied by 5 and increased by 7, the SD+ of the new dataset will be  
   a. Increased by 7 only. b. Multiplied by 5 only.  
   b. Multiplied by 5 and increased by 7. d. Unchanged.
9. What is the current official definition of the meter?  
   a. 1,650,763.73 wavelengths of a certain type of radiation from the Cesium atom.  
   b. The distance from the earth to the moon divided by 1010.  
   c. The distance from the North Pole to the equator divided by 107.  
   d. The distance that light travels during 1/299,792,458 of a second.

1. \*Why do the x and SD form a parsimonious description of a normal histogram? (Pick the most relevant answer.)  
   a. No single statistic can describe a normal histogram, so two is the smallest number of   
    statistics needed.  
   b. The mean is more efficient for describing the center of a normal histogram than the  
    median.  
   c. The median is only used to describe skewed histograms.  
   d. The SD is easily distorted by outliers.
2. Which of the following is true about the ideal measurement model?  
   a. The true measurement μ is usually unknown.  
   b. The bias is usually larger than SEave.  
   c. The random errors are always zero.  
   d. SD+ is the best estimate of μ if the random errors are normally distributed.
3. Which of these statements is true?  
   a. The sample mean is always larger than the median.  
   b. SD+ can sometimes be negative.  
   c. SD+ is always smaller than SD.  
   d. SEave is always smaller than SD+ if n > 1.
4. \*If heights of a large female population are normally distributed with mean 63.5 inches and SD = 2.5 inches. About how many women out of one million are 73.5 inches   
   (6 feet 4 inches) tall or taller?  
   a. 1 b. 4 c. 18 d. 450
5. \*If heights of a large female population are normally distributed with mean 63.5 inches and SD = 2.5 inches. What is the 85th percentile of height?  
   a. 63.5 b. 64.5 c. 66.1 d. 70.5
6. If the variable x represents the heights of the female population in Problems 11 and 12 and y represents the same heights converted into meters, then the correlation between x and y is  
   a. -1.0 b. 0.0 c. About 0.5 d. 1.0

**Part B. Problems.** For full credit, show all your work.

**Histogram Problem.** Use the data in this table to answer questions about the resulting histogram.

|  |  |
| --- | --- |
| **Bin** | **Percentage** |
| [0,3] | 20% |
| [3,5] | 50% |
| [5,6] | 30% |

a. Draw the histogram.  
  
b. Compute Q0, Q2, Q3, Q4, and the IQR for the histogram. Use linear interpolation.

c. Compute the mean of the histogram using a weighted average.  
  
  
  
d. Estimate what percentage of observation are found in the bin (1.5,5.5].

1. **Regression Problem.** An analysis shows that the midterm (x-variable) and final (y-variable) scores in a large class are bivariate normal. Here are the summary variables:  
   x mean = 70 SDx = 20  
   y mean = 75 SDy = 15 r = 0.6
2. About what percentage of the students have final exam scores over 95?
3. Find the regression line for predicting final score from the midterm score.
4. What is the predicted final exam score for a student that obtains a score of 90 on the midterm.

**Part C. Ideal Measurement Example.** The PingTimes variable represents the number of milliseconds that a “ping” data packet takes to travel from a local PC to a website server and back. Fifty ping times were collected. They are analyzed in the SPSS output entitled *SPSS Output for Problem C: Ideal Measurement Model*. Answer these questions based on this output. No calculations are required.

1. What are the values of these statistics?  
     
   a. Q0 b. Q1 c. Q2 d. Q3 e. Q4   
     
     
   g. Sample Mean h.SD+ i. SEave   
     
     
   j. 95%-confidence interval for μ
2. What do the normal plots (Items 4 and 6) tell you about the data?
3. What do the plots of PingTimes vs. Observation number tell you about the data?

**Part C. Regression Problems.** The SPSS output *entitled SPSS Output for Problem D: Regression Problems* shows the results of trying to predict BrainWeight from BodyWeight for various animals. Because the correlation of BrainWeight with BodyWeight obtained in Item 2 is very small (r = -0.005), logarithm transforms were applied to both BrainWeight and Body Weight to give LogBrainWeight and LogBodyWeight. Answer these questions about the regression equation for predicting LogBrainWeight from LogBodyWeight:

1. What is the correlation between LogBrainWeight and LogBodyWeight?
2. What is the regression equation for predicting LogBrainWeight from LogBodyWeight?
3. What does the residual plot in Item 7 tell you?
4. What does the normal plot of the residuals in Item 8 tell you?