Introduction to Statistical Measurement and Modeling
Homework 1
Due: Tuesday, July 12

1. <u>Data display</u>:

You may access the osteoporosis data on the class website.

a) Create a boxplot of age in the osteoporosis data. Do this by hand and using R.

b) Compute the following for the DPA measure of bone density within (for each of) osteoporosis groups=0, 1:

i) The sample mean
ii) The sample standard deviation
iii) Create a stem and leaf plot (one per group).

Do these by hand, and verify using R.

c) Using the class notes, examine the boxplots of DPA scores by osteoporosis groups.

i) What is the median DPA score for each osteoporosis group?
ii) What is the interquartile range for each osteoporosis group?

d) Using a) and b) comment on the differences in the DPA distributions.

2. <u>Probability calculations</u>: Define a random variable X from the DPA data as follows:

            X=1   if DPA ≤ 0.80
            X=2   if 0.80 < DPA ≤ 0.90
            X=3   if 0.90 < DPA ≤ 0.95
            X=4   if 0.95 < DPA ≤ 1.00
            X=5   if DPA > 1.00

a) Compute the probability mass function of X

b) Compute and graph the cumulative distribution function of X.

c) Compute the mean of X

d) Which do you prefer as a bone density measurement: the original DPA score or the variable you have just defined? Explain in a few sentences.

3. The Food and Drug Administration (FDA) is concerned about the level of "EBDC" pesticides in tomato products. To explore whether levels are safe, it plans to determine EBDC concentrations (in ppm = parts per million) in cans of tomato paste collected from grocery stores. Suppose that concentrations greater than 7 ppm are dangerous.

Denote by "X" the EBDC concentration in a can of tomato paste. Suppose that X has a density which is defined as follows:

$$f_X(x) \quad = \quad .025x \qquad \text{if } 0 \le x \le 8$$
$$= \quad .200-.100(x-8) \quad \text{if } 8 \le x \le 10$$
$$= \quad 0 \qquad \text{otherwise.}$$

a) Sketch $f_X(x)$ and $F_X(x)$
b) Find the proportion of tomato paste cans containing dangerous levels of EBDC.
c) Find the mean and median of the EBDC concentration distribution.
d) On the sketch of F, diagram how to find $Q_X(.5)$.
e) Suppose the FDA argues that since the mean concentration is below the dangerous level, the food supply is safe. Would you be satisfied as a public health official? Explain why or why not in a few sentences, using any of a)-d).