Solution of homework 5

10.24
    (a) Let X be Shelia's measured glucose level. $P(X > 140) = P(Z > (140-125)/10) = P(Z > 1.5) = 0.0668$
    (b) If $\bar{x}$ is the mean of four measurements (assumed to be independent), then $\bar{x}$ has a $N(125, 15/\sqrt{4}) = N(125, 7.5)$ distribution. And $P(\bar{x} > 140) = P(Z > 2) = 0.0228$.

10.26
    (a) $\bar{x}$ is approximately Normal with mean 0.9g/mi and standard deviation 0.15/(125) ≈ 0.01342g/mi.
    (b) $P(Z > 2.326) = 0.01$ if Z is $N(0,1)$, so $L = 0.9 + 2.236*0.01342 = 0.9312$ g/mi.

1.2
    (a) Gender is categorical.
    (b) Age is quantitative.
    (c) Race is categorical.
    (d) Smoker is categorical.
    (e) Blood pressure is quantitative.
    (f) Calcium is quantitative.

1.5 You may use classes 10-14, 15-19, 20-24, etc or 11-15, 16-20, 21-25 etc or other appropriate classes of "width 5miles per gallon". Graph is omitted.

1.17
    (a) The distribution is approximately symmetric – perhaps slightly skewed to the left, but not strikingly so if one ignore the low outlier.
    (b) The center is between 0% and 2%.
    (c) The lowest return was between –16% and –14% if ignoring the outlier; if not, it was between –24% and –22%. The highest return was between 16% and 18%.
    (d) About 40% of these months had negative returns.

1.29 (b) The distribution is clearly skewed to the right, with the District of Columbia a high outlier. The states all have numbers between 155 and 422; D.C. is very different from the states in that it includes very little area that would be considered "rural", where we would expect the density of doctors would drop off considerably. (Observe that the states with large cities tend to have high numbers; D.C. is an extreme case, as it consists mainly of a large city.) The graph is omitted.

2.8
    (a) $\bar{x} = 32.4/6 = 5.4$ mg of phosphate per deciliter of blood.
    (b) The standard deviation is $s = 0.6419$ mg of phosphate per deciliter of blood.

2.11 The median is $53,054 and the mean is $72.674: Income distribution will be skewed to the right, so the mean will be larger.

2.14 The minimum, Q(1), median, Q(3) and maximum for Ruth are 22, 35, 46, 54 and 60. Those for McGwire are 9, 29, 39, 52 and 70. Boxplots are omitted. With the exception of maximum, McGwire's boxplot is lower than Ruth's, suggesting (as did stemplot) that Ruth was somewhat stronger overall, but the boxplots show considerably less detail than the stemplots – e.g., we cannot see that McGwire's minimum seasons are outliers from the rest of his distribution. That's why the stemplots reveal more information about the data sets.

2.20 Either a stemplot or a histogram is a good choice. The stemplot shows the distribution to be fairly symmetrical, with a low outlier of 4.88; $\bar{x}$ and s should be reasonable in this setting. $\bar{x} \approx 5.448$ and s ≈ 0.22095.

2.23
    (a) Either a stemplot or a histogram is a good choice. The expected right skew is clearly evident; the split stems (or bars) emphasize the skewness by showing the gaps. The main peak occurs from 50 to 150 days – the guinea pigs that lived more than 500 days seem to be outliers.
    (b) Five-number summary: 43 82.5 102.5 151.5 598 days. The difference between Q(3) and maximum is much larger than the other differences between successive numbers. This indicates a large "spread" among the high observations – that is, it shows that the data are skewed to the right.