Statistical Methods, Instructors Solutions Manual

CHAPTER 1

EXERCISE 1

a) Mean = 17.00, SD = 5.53, median = 17.00, range = 22, IQR = 20 − 13 = 7.

EXERCISE 2

Variable	Mean	Median	Variance	Standard Deviation	Shape
WATER	7.125	1.500	452.864	21.28	extremely positively skewed
VEG	1.120	0.00	4.327	2.080	extremely positively skewed
FOWL	75.635	11.5	42197.33	205.420	extremely positively skewed

```
WATER

        Stem Leaf                                          #      Boxplot
          14 9                                             1         *
          13
          12
          11
          10
           9
           8
           7
           6
           5
           4
           3 13                                            2         *
           2
           1 05667                                         5         0
           0 00000000000001111111111112222222222345556779  44     +--+--+
             ----+----+----+----+----+----+----+----+----
         Multiply Stem.Leaf by 10**+1
```

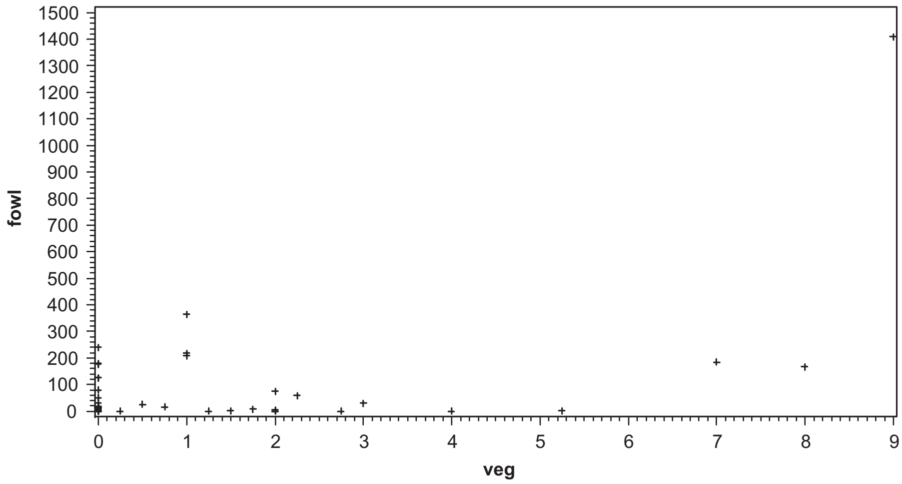
```
VEG

        Stem Leaf                                          #     Boxplot
           9 0                                             1        *
           8
           8 0                                             1        *
           7
           7 0                                             1        *
           6
           6
           5
           5 2                                             1        0
           4
           4 0                                             1        |
           3                                                        |
           3 0                                             1        |
           2 8                                             1        |
           2 00002                                         5        |
           1 58                                            2     +-----+
           1 0002                                          4     |  +  |
           0 58                                            2     |     |
           0 00000000000000000000000000000002             32     *-----*
             ----+----+----+----+----+----+--

FOWL

        Stem Leaf                                          #     Boxplot
          14 1                                             1        *
          13
          12
          11
          10
           9
           8
           7
           6
           5
           4
           3 6                                             1        *
           2 124                                           3        0
           1 227888                                        6        0
           0 00000000000000000000000111111222222333335678 41     +--+--+
             ----+----+----+----+----+----+----+----+-
         Multiply Stem.Leaf by 10**+2
```

b) Frequency distribution for FOWL

Fowl	Frequency	Midpoint
$0 \leq x < 100$	41	50.0
$100 \leq x < 200$	6	150.0
$200 \leq x < 300$	3	250.0
$300 \leq x < 400$	1	350.0
$1400 \leq x < 1500$	1	1450.0

approximate mean = 5500/52 = 105.8
approximate variance = 50746 using 52 in den, per formula in chapter

c) Scatter plots

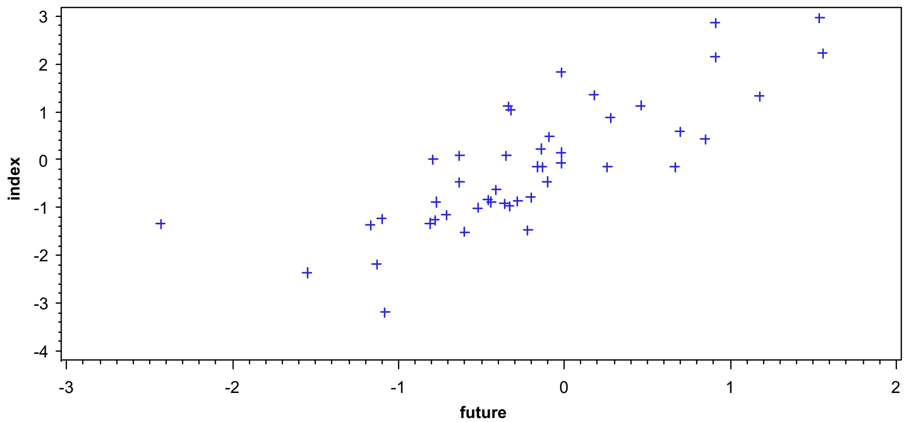


Aside from the single point with extraordinary number of waterfowl, there is little relationship.

EXERCISE 3

a) Index: Mean = −.149348, median = −.155, variance = 1.7707529
Future: Mean = −.208478, median = −.3, variance= .601018

b) The scatter plot is shown below. There is a clear trend for positive changes in the FUTURE contract value to be associated with positive changes in the NYSE INDEX. Yes, FUTURE can be used to help predict changes in the NYSE index.



Yes. The scatter plot shows that, in general, as the NYSE Composite Index increases so does the price movements of the general stock market.

EXERCISE 4

```
Y1                Stem Leaf                                 #           Boxplot
                     8 14                                   2              |
                     7 3                                    1              |
                     6 1237889                              7           +-----+
                     5 338                                  3           *-----*
                     4 025                                  3           |  +  |
                     3 112                                  3           +-----+
                     2 478                                  3              |
                     1 668                                  3              |
                       ----+----+----+----+
```

```
Y2                Stem Leaf                                 #           Boxplot
                     6 5679                                 4              |
                     6 4                                    1              |
                     5 789                                  3           +-----+
                     5 22                                   2           |     |
                     4 678                                  3           *--+--*
                     4 0024                                 4           |     |
                     3 559                                  3           +-----+
                     3 12333                                5              |
                       ----+----+----+----+
```

```
Y3                Stem Leaf                                 #           Boxplot
                     9 4                                    1              |
                     8 0                                    1              |
                     7                                                     |
                     6 777                                  3              |
                     5 444                                  3           +-----+
                     4 000                                  3           |     |
                     3                                                  |  +  |
                     2 7777777                              7           *-----*
                     1 333333                               6           +-----+
                     0 0                                    1              |
                       ----+----+----+----+
```

```
Y4                Stem Leaf                                 #           Boxplot
                     7 4                                    1              |
                     6                                                     |
                     5 0                                    1              |
                     4 578                                  3              |
                     3 99                                   2           +-----+
                     2 123357                               6           *--+--*
                     1 01468                                5           |     |
                     0 1113699                              7           +-----+
                       ----+----+----+----+
```

Y1 is slightly negatively (or left) skewed; Y2 is nearly symmetrically distributed, though without a well defined peak; Y3 is positively (or right) skewed; Y4 is extremely positively skewed.

comparison to empirical rule, actual percentage of observations in interval

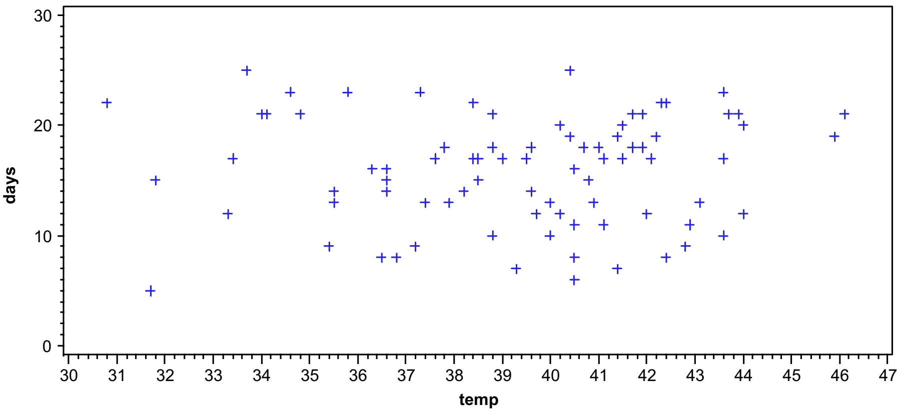
	shape	±1 SD (expect 68%)	±2 SD (expect 95%)	±3 SD (expect all)
Y1	slight left skew	68%	100%	100%
Y2	near symm.	60%	100%	100%
Y3	right skew	76%	96%	100%
Y4	extreme right skew	64%	96%	100%

The empirical rule works reasonably well even for the skewed distributions.

EXERCISE 5

a) Days: Mean = 15.85366, median = 17, variance = 24.324
Temp: Mean = 39.34756, median = 40, variance = 11.15191

b) Scatter plot:

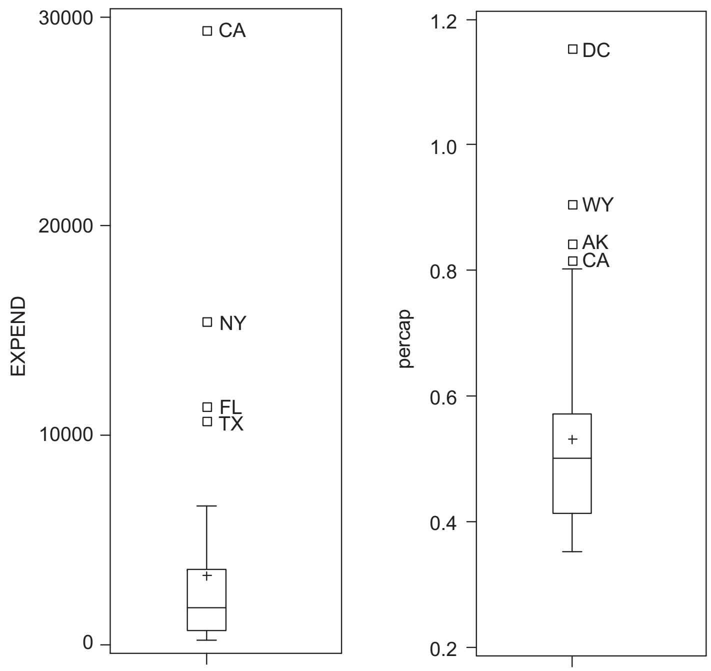


From the scatter plot, there appears to be no definitive relationship between the average temperature and the number of rainy January days.

EXERCISE 6

a)

Variable: EXPEND (EXPEND)			
N	51		
Mean	3302.43137	Median	1757.00
Std Deviation	4802.36473	Variance	23062707



The distribution of expenditures is EXTREMELY positively skewed. The mean is almost twice the median. California spends far more than any other state, followed by New York, Florida and Texas, all of which have relatively large expenditures.

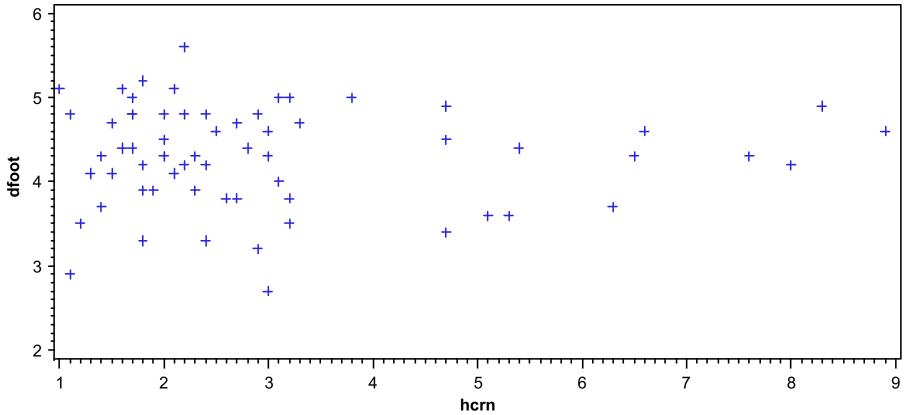
b) PErcap = EXPEND / POP

Location		Variability	
Mean	0.531172	Std Deviation	0.16056
Median	0.501985	Variance	0.02578

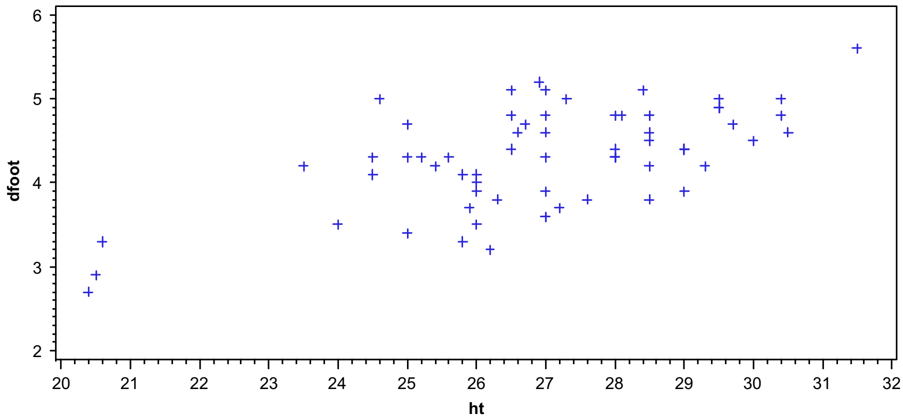
The distribution of PERCAP is still skewed, but less extremely than for EXPEND (see boxplot above). Interestingly, the outlier states are different once you adjust for population sizes. The top states per capita are DC, Alaska and Wyoming, all with relatively small populations. California ranks fourth, even though it was far and away the top in terms of raw expenditures.

EXERCISE 7

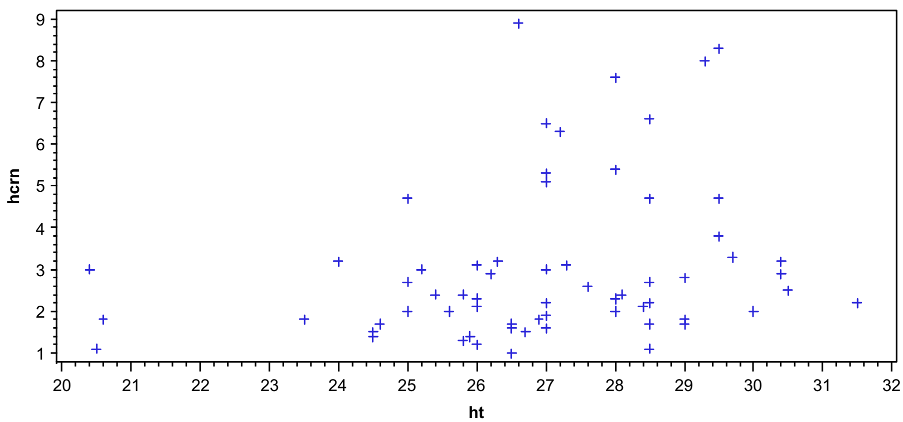
Scatter of dfoot vs. hcrn:



Scatter of dfoot vs. ht:



Scatter of hcrn vs. ht:

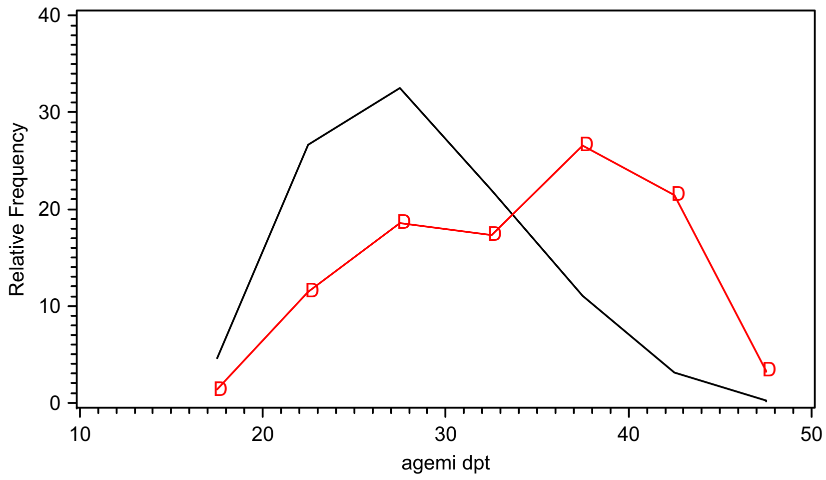


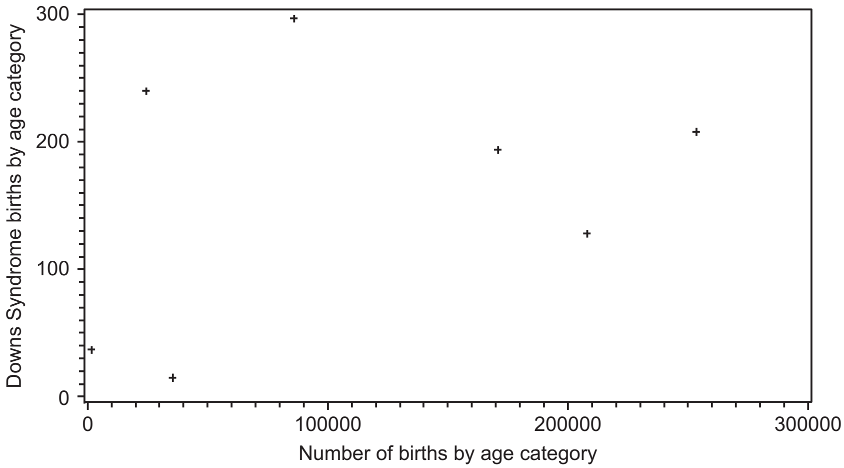
The strongest relationship exists between DFOOT, the diameter of the tree at one foot above ground level and HT, the total height of the tree. (Notice from the scatter plot that it is closest to a linear relationship.) One would expect that as the base of the tree increases in diameter the tree would increase in height as well.

EXERCISE 8

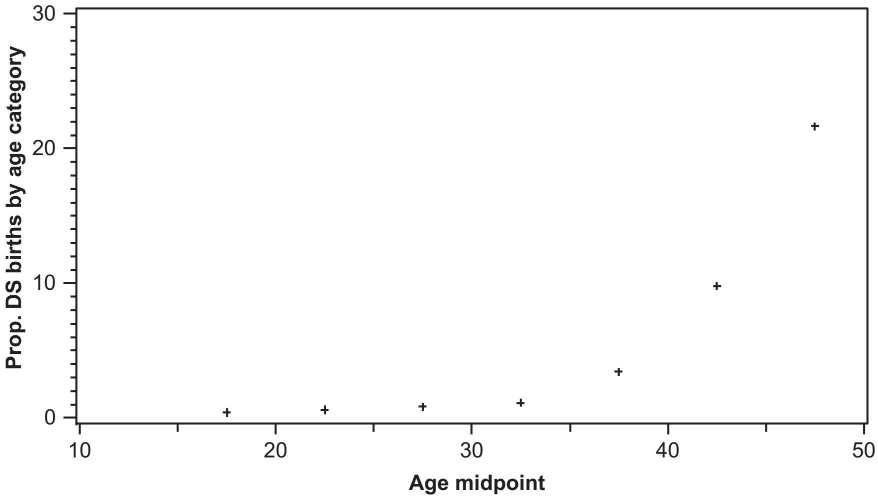
(A, B, and C) To make it easier to compare the distributions, we will display them as relative frequency polygons (connecting the centers of the tops of the bars) rather than as histograms. The open-ended class for mothers 20 or less was assigned a midpoint of 17.5 years, and the open ended class for mothers 45 or over was assigned a midpoint of 47.5. Other class midpoints were in intervals of 5 years. The frequency distribution for the general population of births is in black, that for ages of Downs syndrome births are in red (with a D).

The typical age for a mother of a Downs Syndrome infant is much higher than that of the general population. Age does not completely account for DS, because DS can happen at any age. However, it is a strong component of any explanation, as the proportion of births with DS is much higher in the late 30s and early 40s.





Exactly as specified in the problem, the graph is not very revealing. It does not help draw any conclusions, except to note that the largest number of DS births is in a category with relatively few total births.



The most revealing plot is one which the proportion of DS births per 1000 total births (1000 * number_of_mothers_of_downs_synd/total_number_of_births) versus age midpoint. Now we can clearly see that the risk of a Downs Syndrome birth rises continuously beginning in the late 30's.

EXERCISE 9

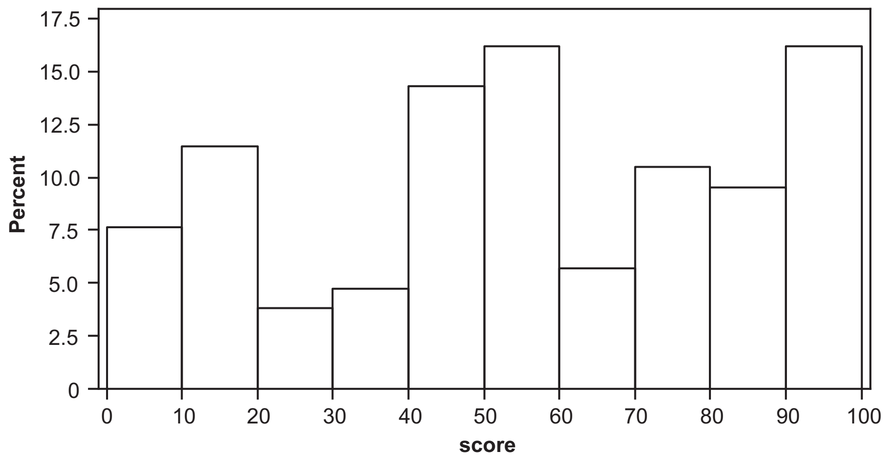
a) The mean is larger than (to the right of) the median, indicating a distribution skewed to the right. Yes, both the stem and leaf plot and the box plot reveal the skewness of the distribution.

b) The outliers 955 and 1160 may have resulted from younger patients or from patients in which the disease was less severe.

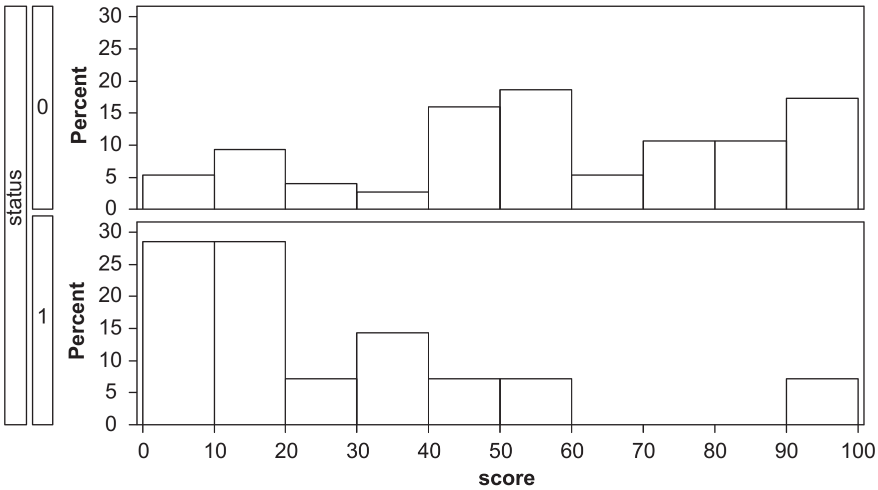
c) 38 out of 51 patients were in remission for less than one year, so approximately 75%.

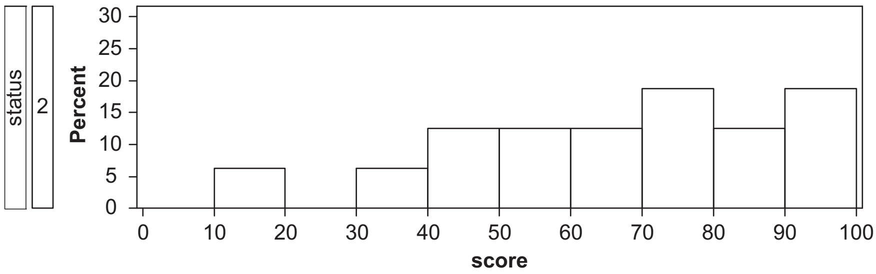
EXERCISE 10

a) For the combined group, scores are almost evenly distributed over the interval from 0 to 100.



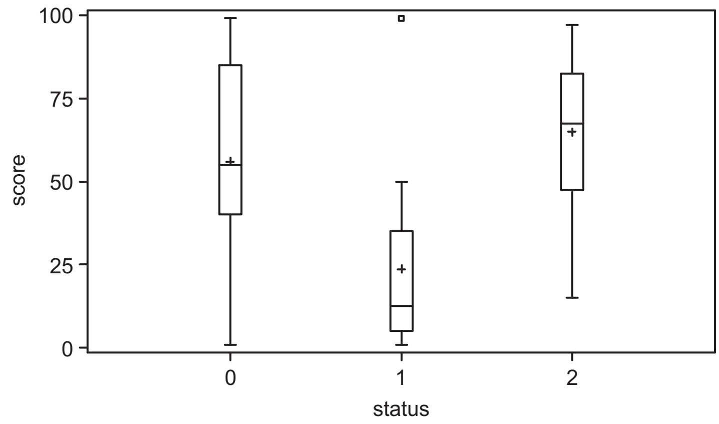
b)





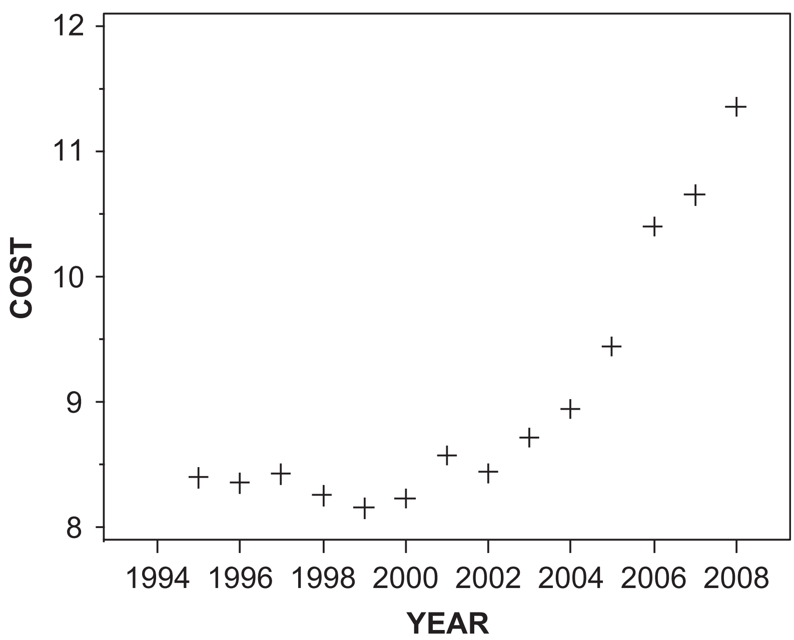
As expected, the placement scores among those who passed the class (Status=0) tend to be higher than the scores of those who failed the class (Status=1). Surprisingly, the scores for those who withdrew (Status=2) are as high as those for those who passed. Perhaps this is because those who withdrew do so for a variety of reasons frequently unrelated to their math ability.

NOTE TO INSTRUCTOR: The differences between the groups might best be displayed with boxplots, though boxplots have a tendency to hide multiple modes.



EXERCISE 11

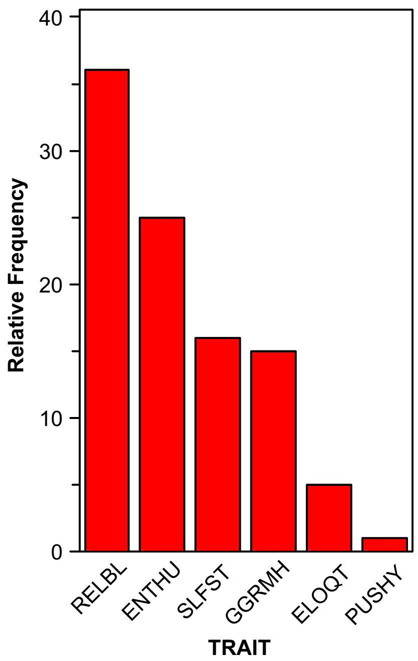
a) Since the data is sequential by year, use a time series plot.



b) Electricity costs were stable (or perhaps even declining slightly) during the mid to late 1990s. However, in 2001 they began to increase, and have been increasing sharply since 2003.

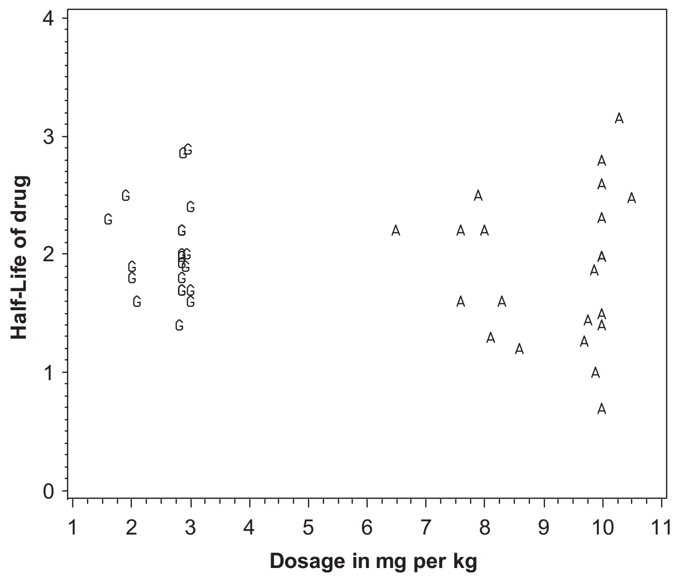
EXERCISE 12

a) Reliability (36.7%) and Enthusiasm (25%) are the traits most frequently thought of as most important.



EXERCISE 13

a) The initial dosage seems to be higher for drug A, but the half-lifes are not much different.



b) There is not any strong relationship between half-life and dosage, either overall or within the individual drugs.

c) Drug A: M = 9.21, SD = 1.14; Drug B: M = 2.67, SD = 0.44. This supports the conclusion in part (a).