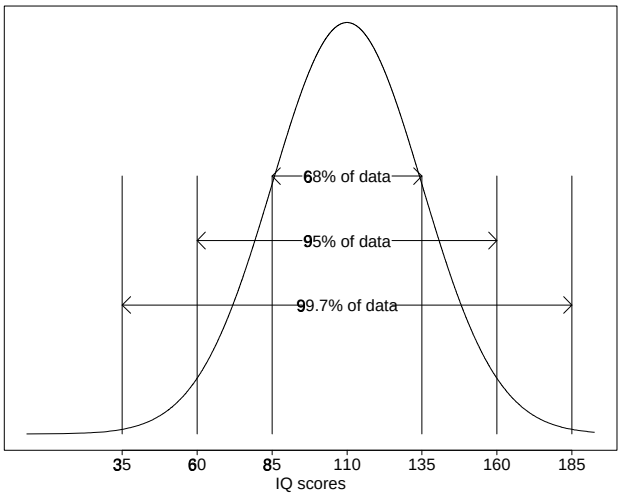
36-201 Spring 1999
Solutions to Homework 3

1. `Moore, 4.71 (p.274).` The 68-95-99.7 rule says that 68% of the distribution is between $110 \pm 25 = (85,\ 135)$, 95% of the distribution is between $110 \pm 2 \times 25 = (60,\ 160)$ and 99.7% of the distribution is between $110 \pm 3 \times 25 = (35,\ 185)$. This can be seen in the figure below.



   (a) In the normal case, the Mean 110 is also the Median. So a 50% of the people have scores above 110.

   (b) 5% of the data are below 60 or above 160. The normal distribution is symmetric, then we know that 2.5% of the people are above 160.

   (c) 32% of the data are below 85 or above 135. The normal distribution is symmetric, then we know that 16% of the people are below 85.

`Moore, 4.74 (p.274).`

   (a) To get Sarah's standard score we must refer to the 20-34 age group described in problem `4.71`. That group has a mean of 110 and a standard deviation of 25, so Sarah's standard score is

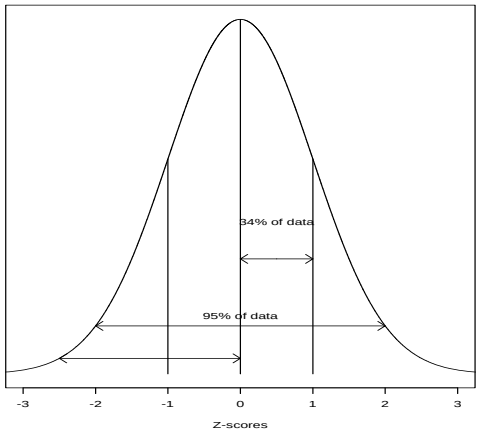
$$\frac{x_{Sarah} - \bar{x}}{SD} = \frac{135 - 110}{25} = 1.$$

   (b) Sarah's mother belongs to the 60-64 age group that has a mean of 90 and a standard deviation of 25. Since she got 120 in her IQ test, her standard score is

$$\frac{x_{mother} - \bar{x}}{SD} = \frac{120 - 90}{25} = 1.2.$$

   (c) Sarah's mother scored higher than Sarah, relative to their age group (the standard score of Sarah's mother is higher than the standard score of Sarah). However, Sarah has a higher absolute level in the IQ test (her score,135, is higher than her mother's, 120).

2. Moore, 4.78 (p.275). In addition, use interpolation to get the percentiles of Z-scores of 1.5 and 1.75. Figure 4.30 looks like



- A point one standard deviation below the mean, $(\bar{x} - SD)$, translates into a standard score of

$$\frac{(\bar{x} - SD) - \bar{x}}{SD} = -1.$$

¿From the chart and using the symmetry of the normal distribution we see that approximately 68% of the data is between -1 and 1, so 16% of the distribution is below -1. That means that the original observation, $(\bar{x} - SD)$, corresponds approximately to the 16th percentile.

- A point two standard deviations above the mean, $(\bar{x} + 2\ SD)$, translates into a standard score of

$$\frac{(\bar{x} + 2\ SD) - \bar{x}}{SD} = 2.$$

¿From the 68-95-99.7 rule we know that 95% of the distribution is between -2 and 2, so (95 + 2.5)% of the distribution is below 2. That means that 2 is the 97.5th percentile of the standard normal distribution, and so is the original observation, $(\bar{x} + 2\ SD)$, with respect to its distribution.

- We can interpolate the percentile of a Z-score of 1.5 using the percentiles of 1 and 2 from the chart. A Z-score of 1 is the (50 + 34) = 84th percentile, while a Z-score of 2 is the 97.5th percentile (as explained above). So a Z-score of 1.5 can be interpolated as

$$\frac{84 + 97.5}{2} = 90.75,$$

that is, the 90.75th percentile.

- In the same way we can interpolate a Z-score of 1.75 as

$$0.25 \times 84 + 0.75 \times 97.5 = 94.125,$$

that is, approximately the 94.1th percentile.

3. Siegel and Morgan, 14 (p.157).

(a) We get $\sqrt{11,900} = 109.0871$, as the table reads.

(b) We check that $\sqrt{11,900} \times \sqrt{11,900} = 109.0871 \times 109.0871 = 11,900$.

(c) We get that $\log_{10} 11,900 = 4.075547$ and that $\ln 11,900 = 9.384294$, as the table reads.

(d) Using the 10^x key we check that $10^{4.075547} = 11,900$. Using the e^x key we check that $e^{9.384294} = 11,900$.

4. Siegel and Morgan, 9 (p.165).

(a) By eye, it looks like the observations are sparse at larger values, so the distribution may be skewed (to the right). In particular there is an extremely high observation: 148,550 which may or may not be an outlier. So, this distribution looks like a good candidate for a transformation.

(b) The stem-and-leaf plot for the Areas is

```
Decimal point is 4 places to the right of the colon

  0 : 333344
  0 : 567789
  1 : 011223
  1 :
  2 : 23
  2 : 57
  3 : 2

High:  148550
```

The distribution is skewed to the right, and there is one outlier.

(c) The stem-and-leaf plot for the Square Root of the Areas is

```
Decimal point is 1 place to the right of the colon

   5 : 5677
   6 : 01
   7 : 29
   8 : 347
   9 : 7
  10 : 057
  11 : 003
  12 :
  13 :
  14 : 9
  15 : 29
  16 : 4
  17 : 8

High: 385.422
```

This distribution seems flatter than the distribution of the untransformed data, so we may say it is less skewed. There is also one outlier.

(d) The stem-and-leaf plot for the Natural Logarithm of the Areas is

```
Decimal point is at the colon

  8 : 00112267899
  9 : 2233444
 10 : 00124
 11 : 9
```

This distribution looks almost as skewed as the original one, but it does not have outliers.

(e) The Square Root transformation makes the distribution most nearly symmetric.

5. A (i) Both, LSAT scores and GPA's are quantitative variables.

(ii) Both, Heights of fathers and Heights of sons are quantitative variables.

(iii) Change in temperature is a quantitative variable while Month can be considered either as quantitative (since we can number months from 1 to 12) or as qualitative (since we can just use month's names).

(iv) The Score in the test of depressive disorder is a quantitative variable while the Treatment (lithium or therapy) is a qualitative variable.

B (i) In this case we think as LSAT scores being the independent variable and GPA's being the dependent variable (GPA scores are actually observed after LSAT scores).

(ii) Heights of fathers is the independent variable while Heights of sons is the dependent one (we would think that the height of the father affects the height of his son and not vice-versa).

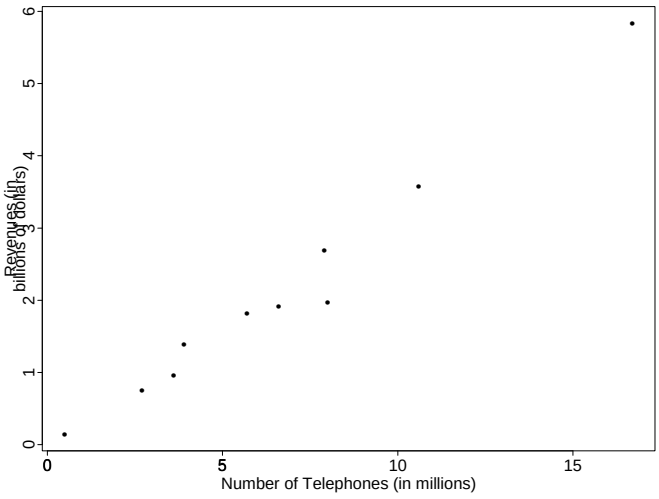
(iii) Month of the year is the independent variable while Change in temperature is the dependent variable.

(iv) The Treatment (lithium or therapy) is the independent variable while the Score in the test is the dependent variable.

6. `Siegel and Morgan, 9 (p.571).`

(a) In this example, it is more natural to think that revenues can be explained by the number of telephones, so the number of telephones are predicting the revenues.

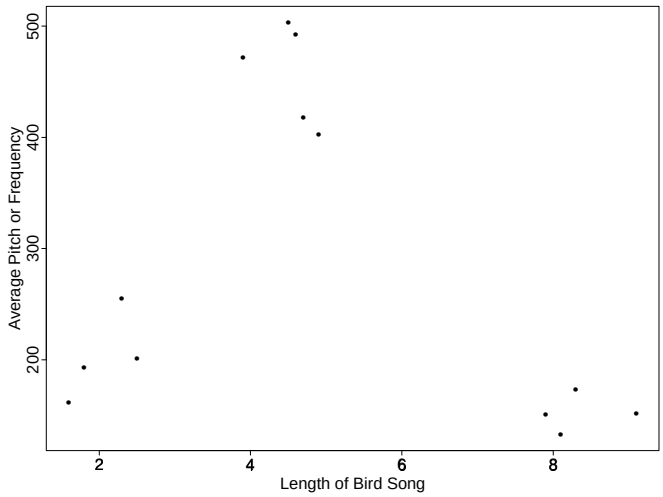
(b) The plot of Revenues vs. Number of Telephones is



(c) The plot shows an increasing linear trend. The more the number of telephones, the more the revenues received. The increase looks like a straight line with some randomness.

7. `Siegel and Morgan, 12 (p.573).`

(a) The plot of Average Pitch vs. Length of Song is



(b) It looks like there are three different groups, where the relationship between Length of Song and Average Pitch is different for each of these groups.

(c) We may think that each cluster in the plot correspond to a different kind of song, i.e., points in the same cluster correspond to songs with similar purposes.