Math 108 **Exam 1** **Spring '98**

1. Consider the data in the minitab worksheet e1-1.
   a) What is the shape of the distribution? Does it have a long tail? If so, on which side?

   SOLUTION: The distribution has a long left tail.

   b) Which values, if any, satisfy the rule for determining outliers?

   SOLUTION: The interquartile range (IQR), given by Q3 − Q1, is 19. Q1 − 1.5 IQR = 39.5, and Q3 + 1.5 IQR = 115.5. Since the minimum value is above 39.5 and the maximum value is below 115.5, there are no outliers.

   c) Recall that for a small data set, the percentile value of a given number is the percentage of the data points that are less than or equal to the given value. What is the percentile value of 74?

   SOLUTION: Sorting the data, we see that the first 7 out of the 19 data elements are less than or equal to 74. Thus, the percentile value of 74 is $\frac{7}{19} \approx 36.84\%$.

2. Consider the data in the minitab worksheet e1-2.
   a) What is the correlation coefficient, r, for c1 and c2?

   SOLUTION: We set C3 equal to the product of the z-values of C1 and C2:

$$\mathrm{C3} = (\mathrm{C1} - \mathrm{mean(C1)})/\mathrm{stdev(C1)} * (\mathrm{C2} - \mathrm{mean(C2)})/\mathrm{stdev(C2)}$$

   Then we set K1 = sum(C3)/count(C3), so r is K1, which computes to .219273 .

   b) Can you detect a pattern in the scatter diagram? Is there greater spread in either direction (right or left)?

   SOLUTION: The data appears to be clustered toward the lower left corner of the plot, and to spread out as you move away from that corner.

3. Supppose given a binormal distribution with

$$\overline{x} = 185 \quad \overline{y} = 120$$
$$S_x = 80 \quad S_y = 45 \quad r = -.75$$

   a) What is the slope of the regression line?

   SOLUTION: The slope is $r\frac{S_y}{S_x} = -.75\frac{9}{16} \approx -.4219$.

b) What is the y-intercept of the regression line?

SOLUTION: The regression line is given by

$$\frac{y - \overline{y}}{x - \overline{x}} = -.75\frac{9}{16}$$

Solving this, we get $y \approx -.4219x + 198.05$, so the y-intercept is approximately 198.05.

c) What is the regression estimate for $x = 80$?

SOLUTION: Substituting 80 into the above equation for the regression line, we get $y \approx 164.30$.

d) Suppose the x-value is in the 70-th percentile. What is the percentile value of its regression estimate?

SOLUTION: The z-value of the regression estimate is r times the z-value of x. And the z-value of the 70th percentile of a normal distribution is approximately .52, from the normal table. Thus, the z-value of the regression estimate is about $(-.75)(.52) = -.39$. Using the normal table, we see the percentile value of the regression estimate is 34.83%.

e) Suppose $y = 170$. What is the regression estimate for predicting the value of x from that of y?

SOLUTION: Here the z-value of the regression estimate is r times the z-value of $y = 170$, i.e.,

$$\frac{x - 185}{80} = -.75\frac{170 - 120}{45}$$

Solving this, we see $x \approx 118.33$.

f) What percent of the x-values are between 200 and 250?

SOLUTION: This concerns the x-distribution alone. The z-value of 250 is $\frac{250-185}{80} \approx .81$, which has a percentile value of .7910. The z-value of 200 is $\frac{200-185}{80} \approx .19$, which has a percentile value of .5753. So the percentage between them is $79.10\% - 57.53\% = 21.57\%$.

4. Suppose given a binormal distribution with

$$\overline{x} = 750 \quad \overline{y} = 950$$
$$S_x = 100 \quad S_y = 120 \quad r = .65$$

Control for $x = 900$.

a) What is the average of the controlled y-distribution?

SOLUTION: The average is the regression estimate for $x = 900$. The z-value of the regression estimate is $.65 \cdot \frac{900-750}{100} = .975$, so the regression estimate is $.975 * 120 + 950 = 1067$.

b) What is the standard deviation of the controlled y-distribution?

SOLUTION: The standard deviation after controlling is

$$\sqrt{1-r^2}S_y \approx 91.19$$

c) What is the 35-th percentile of the controlled y-distribution?

SOLUTION: From the normal table, we see the z-value of the 35th percentile is approximately $-.39$. In the controlled y-distribution, this corresponds to a numerical value of approximately

$$(-.39)(91.19) + 1067 \approx 1031.44$$

d) What is the percentile value of 950 in the controlled y-distribution?

SOLUTION: The z-value of 950 in the controlled y-distribution is $\frac{950-1067}{91.19} \approx -1.28$. The normal table gives this a percentile value of 10.03%.

e) What percent of the controlled y-values are between 950 and 1200?

SOLUTION: The controlled z-value of 1200 is $\frac{1200-1067}{91.19} \approx 1.46$. This has a percentile value of 92.79%. As shown in part d), 950 has a percentile value of 10.03% in the controlled y-distribution. Thus, there is $92.79\% - 10.03\% = 82.76\%$ between them.