Problem 1: The distribution by age in the U.S. in 2004 is given in the table below.

(1) Draw the corresponding histogram. Notice that the class intervals include the left endpoint, not the right one. The interval for “75 and over” can be ended at 85.

(2) Are there more 23-year-olds, or more age 63-year-olds?

(3) Is the percentage of people age 35 or more around 25%, 50% or 75%?

<table>
<thead>
<tr>
<th>Age</th>
<th>% of population</th>
<th>Age</th>
<th>% of population</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–5</td>
<td>7</td>
<td>35–45</td>
<td>15</td>
</tr>
<tr>
<td>5–15</td>
<td>14</td>
<td>45–55</td>
<td>14</td>
</tr>
<tr>
<td>15–20</td>
<td>7</td>
<td>55–65</td>
<td>10</td>
</tr>
<tr>
<td>20–25</td>
<td>7</td>
<td>65–75</td>
<td>6</td>
</tr>
<tr>
<td>25–30</td>
<td>7</td>
<td>75 and over</td>
<td>6</td>
</tr>
<tr>
<td>30–35</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1). See Figure 1 for drawing of the histogram.

(2). There are more 23 year old. Since the age between 20 and 25 occupies 7% of the population and the age between 55 and 65 occupies 10% of the population. So on average, 23 year old amount to 1.4% while 61 year old amount to 1.0%.

(3). The percentage of people age 35 or more is around 50% judging by the histogram.

Problem 2: A study on college students found that the men had an average weight of about 66 kg and an SD of about 9 kg. The women had an average weight of about 55 kg and a SD of 9 kg.

(1) Find the averages and SDs, in pounds (1 kg=2.2 lb).

(2) Just roughly, what percentage of the men weighed between 57 kg and 75 kg?

(3) If you took the men and women together, would the SD of their weights be smaller that 9 kg, just about 9 kg? Why?

(1) Averages are 145.2 lb and 121 lb for men and women separately. SDs are 19.8 lb for men and women respectively.
Figure 1. Histogram of the population distribution by age.

(2). 68%.
(3). Larger than 9 kg. The difference between the averages of men and women is 11 kg. This increases the SD.

Problem 3: Calculate the coefficient of correlation for the data in the following table.

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

Use the following information in your calculations: average of $x=1$, SD of $x=1$, average of $y=8$, SD of $y=5$

In standard units:
$$SD_x \begin{bmatrix} -1 & 0 & 0 & 0 & 2 \\ SD_y \begin{bmatrix} -7/5 \ -1 & 1/5 & 1 & 7/5 \\ SD_x \times SD_y \begin{bmatrix} 7/5 \ 1 & 0 & 0 & 0 & 14/5 \\ \end{bmatrix}$$

Average of $SD_x \times SD_y = 0.87$

**Problem 4:** In a large statistics class, the correlation between midterm scores and final scores is found to be nearly 0.50, every term. The scatter diagrams are football-shaped. Predict the percentile rank on the final for a student whose rank on the midterm is

a) 5% The area to the left is $(1-.90)/2$ and corresponds to a z score of -1.65. Using the regression method: $-1.65 \times r = -1.65 \times 0.5 = -0.825$. This corresponds to an area to the left of $(1 - 0.59)/2 = 0.205$. The percentile rank on the final would be around 21% (2.5 pts)

b) 80% The area to the left is $(1-.60)/2 + .6$, and corresponds to a z score of 0.85. Using the regression method: $0.85 \times r = 0.85 \times 0.5 = 0.425$. This corresponds to an area to the left of $(1 - 0.33)/2 + .33 = 0.665$. The percentile rank on the final would be around 67% (2.5 pts)

c) unknown 50% (1 pt (Bonus point))

**Problem 5:** From a sample of 570 California women age 25-29 in 2005 data can be summarized as follows:

average education $\approx 13.0$ years, $\quad$ SD $\approx 3.4$ years

average income $\approx$ $18,000$, $\quad$ SD $\approx 20,000$, $\quad$ $r \approx 0.37$

(1) Find the r.m.s. error of the regression line for predicting income from education. r.m.s. error is $\sqrt{1 - r^2} \times SD_y = \sqrt{1 - 0.37^2} \times 20,000 = 0.929 \times 20,000 = 18,580.64$ (2.5 pts)

(2) Predict the income of a woman with 8 years of education

Standard units of years of education: $(8 - 13)/3.4 = -1.47$.

Using the regression methods: Standard units of average income: $-1.47 \times 0.37 = -0.54$.

Income will be 0.54 SDs below average.

Multiply by the SD of y: $-0.54 \times 20,000 = -10,882$. Add the average of y: $18,000 - 10,882 = 7,118$

Predicted income will be $7,118$ (2.5 pts)