Stat 3011 (Fall 1995) Midterm 1

Problem 1

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>22334</td>
</tr>
<tr>
<td>7</td>
<td>569</td>
</tr>
<tr>
<td>8</td>
<td>0124</td>
</tr>
<tr>
<td>8</td>
<td>67788888</td>
</tr>
<tr>
<td>9</td>
<td>24</td>
</tr>
</tbody>
</table>

Problem 2

(a) The answers to this part, and two extra columns useful in drawing the histogram are

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Width</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–&lt;30</td>
<td>7</td>
<td>0.175</td>
<td>10</td>
<td>0.0175</td>
</tr>
<tr>
<td>30–&lt;40</td>
<td>9</td>
<td>0.225</td>
<td>10</td>
<td>0.0225</td>
</tr>
<tr>
<td>40–&lt;50</td>
<td>5</td>
<td>0.125</td>
<td>10</td>
<td>0.0125</td>
</tr>
<tr>
<td>50–&lt;75</td>
<td>10</td>
<td>0.250</td>
<td>25</td>
<td>0.0100</td>
</tr>
<tr>
<td>75–&lt;100</td>
<td>6</td>
<td>0.150</td>
<td>25</td>
<td>0.0060</td>
</tr>
<tr>
<td>100–&lt;150</td>
<td>3</td>
<td>0.075</td>
<td>50</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

(b) Because the bin widths are not all equal we must use the rule height = relative frequency/width, giving the heights in the last column of the table in part (a).
The histogram is

![Histogram](image)

**Problem 3**

- The median is better when the population is skewed.
- The median is better when there are outliers or gross errors.

**Problem 4**

The data in sorted order are

5.2 7.2 7.5 7.6 10.2 10.5 11.9 12.9 13.9 14.9 20.0 20.6 21.0

(a) The median is middle value: 11.9

(b) By the rule in book the lower quartile is the median of lower 7 numbers 5.2, … 11.9, which is 7.6, and the upper quartile is the median of the upper 7 numbers 11.9, …, 21.0, which is 14.9. Thus the IQR is 14.9 – 7.6 = 7.3.

(c) 23% of 13 is 3, so trim three off each end of sorted list, and calculate the average of

7.6 10.2 10.5 11.9 12.9 13.9 14.9

which is 81.9/7 = 11.7.
Problem 5

(a) This interval is the mean plus or minus 1 s. d., which is about 68% by the empirical rule.

(b) This interval is the mean plus or minus 2 s. d., which is about 95% by the empirical rule.

(c) This interval goes from 2 s. d. below the mean to 1 s. d. below the mean. We have 95% - 68% = 27% between 1 s. d. and 2 s. d. away from the mean. By symmetry of the normal curve, half of this 27% is below the mean (and half above). We want the half below, that is 13.5%. [This is just like prob 3.31(c) in the textbook.]

Problem 6

(a) Using

\[ b = r \frac{s_y}{s_x} = .5 \times 30/3 = 5 \]

and

\[ a = \bar{y} - b \bar{x} = 160 - 5 \times 70 = -190, \]

the sample regression equation is

\[ y = a + bx = -190 + 5x \]

(b)

\[ y = -190 + 5 \times 66 = 140 \text{ (pounds)} \]

(c) The coefficient of determination is \( r^2 = .5^2 = .25 \), so 25% of the variance of \( y \) is explained by the linear association with \( x \).

Problem 7

(a)

\[ b = \frac{\sum xy - \frac{1}{n} (\sum x)(\sum y)}{\sum x^2 - \frac{1}{n} (\sum x)^2} \]

\[ = \frac{(600 - 35 \times 205/11)/(112 - 35^2/11)}{12.14286} \]

\[ a = \frac{1}{n} \sum y - b \times \frac{1}{n} \sum x \]

\[ = 205/11 - 12.14286 \times 35/11 \]

\[ = -20.00000 \]
Hence the sample regression equation is

\[ y = -20 + 12.14286 \times x \]

(b)

\[ y = -20 + 12.14286 \times x \]
\[ = -20 + 12.14286 \times 3.5 \]
\[ = 22.5 \text{ (feet per second)} \]