STAT2331-802/3, Spring 2003
Key to Practical Exam 1

Multiple Choice

1. c
2. e (Use Table A)
3. d
4. d

Short Answer / Calculation

1. (a) We can see a linear trend in the scatter plot, it is reasonable to use linear regression to summarize these data.

![Scatter plot image]

(b)

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>(x - ( \bar{x} ))</th>
<th>(y - ( \bar{y} ))</th>
<th>(x - ( \bar{x} ))^2</th>
<th>(y - ( \bar{y} ))^2</th>
<th>(x - ( \bar{x} )) (y - ( \bar{y} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>46</td>
<td>-1.5</td>
<td>-11</td>
<td>2.25</td>
<td>121</td>
<td>16.5</td>
</tr>
<tr>
<td>12</td>
<td>72</td>
<td>4.5</td>
<td>15</td>
<td>20.25</td>
<td>225</td>
<td>67.5</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>-4.5</td>
<td>-11</td>
<td>20.25</td>
<td>121</td>
<td>49.5</td>
</tr>
<tr>
<td>11</td>
<td>75</td>
<td>3.5</td>
<td>18</td>
<td>12.25</td>
<td>324</td>
<td>63.0</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>1.5</td>
<td>3</td>
<td>2.25</td>
<td>9</td>
<td>4.5</td>
</tr>
<tr>
<td>8</td>
<td>57</td>
<td>0.5</td>
<td>0</td>
<td>0.25</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>-3.5</td>
<td>-14</td>
<td>12.25</td>
<td>196</td>
<td>49.0</td>
</tr>
<tr>
<td>7</td>
<td>57</td>
<td>-0.5</td>
<td>0</td>
<td>0.25</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>456</td>
<td>0</td>
<td>70</td>
<td>996</td>
<td>250</td>
</tr>
</tbody>
</table>

\( \bar{x} = 7.5 \) \, \bar{y} = 57

S_{xx} = 70, \ S_{yy} = 996, \ S_{xy} = 250

\text{Slope} = b = \frac{S_{xy}}{S_{xx}} = \frac{250}{70} = 3.571

\text{Intercept} = a = \bar{y} - b \bar{x} = 57 - (3.571)(7.5) = 30.21

Least-squares regression line: \( \hat{y} = 30.21 + 3.571x \)

(c) When \( x = 10, \ \hat{y} = 30.21 + (3.571)(10) = 65.93. \)
(d) When $x = 20$, $\hat{y} = 30.21 + (3.571)(20) = 101.64$. The predicted value is greater than 100, which is not reasonable because the exam mark is out of 100. We have a problem because we are using the regression line to predict the value far from the range of $x$ used to calculate the regression line.

2. In this problem we use the 68-95-99.7 rule and Table A.

(a) 90 is two sd's above the mean, so 2.5% of students will score above 90. Since 50% score above 80, this means that 47.5% score between 80 and 90. Or we could say that 95% score between 70 and 90, so between 80 and 90 is half of this, that is 47.5%.

(b) 5% will be either over 90 or below 80, so 2.5% in each group, so the answer is 2.5%.

(c) Plus or minus 3 sd's gives 99.7% of students, so .15% of students will score 3 sd's above the mean or higher. This will be 95, so you need a score of 95 or higher.

(d) 85 is 1 sd above the mean. 16% of people will score 1 sd above mean or higher.

(e) $z = (x - \mu)/\sigma = (74 – 80)/5 = -1.2$. From Table A, The area under the standard normal curve less than the point $z = -1.2$ is 0.1151. Therefore, percentage of students will have exam scores above 74 is $1-0.1151 = 0.8849 = 88.49\%$.

3. Mean = 5, s.d. = 2.
   Min = 3, Q1 = 4, Q2 = Median = 4, Q3 = 7, Max = 8
   IQR = Q3 – Q1 = 7 – 4 = 3.

4. (a) The death rates for good condition patients are; hospital A: 10/360 = 2.8%; hospital B: 50/550=9%. For poor condition patients we have; hospital A 200/1200 = 16.7%; hospital B 25/75 = 33%. We see that for both good condition and bad condition patients we have that hospital A is better.

(b) This is an example of Simpson's paradox. We can explain it as follows;

   (i) The response variable is patient outcome (death or not). The explanatory variable is hospital (A or B). The collapsing variable is condition of patient (good or bad).

   (ii) The collapsing variable is associated with the explanatory variable, hospital. That is hospital A has far more patients of poor condition, 77% of it's total, than hospital B, where 12% are in poor condition. Also the collapsing variable is associated with the response variable, patients who enter in poor condition are much more likely to die than those who enter in good condition.

   (iii) These two facts explain the reversal. Hospital A's death rate is higher, but it's not because hospital A is worse than B, it's because a much greater percentage of A's patients enter in poor condition.