Regression, correlation and hypothesis testing 1B

1 a \( r = 0.9 \) is a good approximation, since the points lie roughly, but not exactly, on a straight line. Remember that the value of \( r \) tells you how ‘close’ the data is to having a perfect positive or negative linear relationship.

b Clearly \( r \) is negative, and the data is not as close to being linear as in part a. \( r = -0.7 \) is therefore a good approximation.

c The data seems to have some negative correlation, but is rather ‘random’. Because so many points would lie far away from a line of best fit, \( r = -0.3 \) is a good approximation.

2 a The product moment correlation coefficient gives the type (positive or negative) and strength of linear correlation between \( v \) and \( m \).

b By inputting the (ordered) data into your calculator, \( r = 0.870 \) (to 3 s.f.).

3 a \( r = -0.854 \) (to 3 s.f.)

b There is a negative correlation. The relatively older young people took less time to reach the required level.

4 a The completed table should read:

<table>
<thead>
<tr>
<th>Time, ( t )</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atoms, ( n )</td>
<td>231</td>
<td>41</td>
<td>17</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>( \log n )</td>
<td>2.36</td>
<td>1.61</td>
<td>1.23</td>
<td>0.845</td>
<td>0.301</td>
</tr>
</tbody>
</table>

b \( r = -0.980 \) (to 3 s.f.)

c There is an almost perfect negative correlation with the data in the form \( \log n \) against \( t \), which suggests an exponential decay curve. (This uses knowledge from the previous section.)

d \( v = 2.4877 - 0.320x \)

\[ \Rightarrow \log n = 2.4877 - 0.320t \]

\[ \Rightarrow n = 10^{(2.4877-0.320t)} = 10^{2.4877} \times 10^{-0.320t} \]

\[ \Rightarrow n = 10^{2.4877} \times (10^{-0.320})^t \]

Therefore \( a = 102.4877 \approx 307 \) (3 s.f.) and \( b = 10^{-0.320} \approx 0.479 \) (3 s.f.).
5  a

<table>
<thead>
<tr>
<th>Width, w</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass, m</td>
<td>23</td>
<td>40</td>
<td>80</td>
<td>147</td>
<td>265</td>
</tr>
<tr>
<td>log w</td>
<td>0.4771</td>
<td>0.6021</td>
<td>0.7782</td>
<td>0.9031</td>
<td>1.041</td>
</tr>
<tr>
<td>log m</td>
<td>1.362</td>
<td>1.602</td>
<td>1.903</td>
<td>2.167</td>
<td>2.423</td>
</tr>
</tbody>
</table>

b  $r = 0.9996$

c  A graph of $\log w$ against $\log m$ is close to a straight line as the value of $r$ is close to 1, therefore $m = kw^n$ is a good model for this data.

d  $y = 0.464 + 1.88x$
  $\Rightarrow \log m = 0.464 + 1.88\log w$
  $\Rightarrow m = 10^{0.464 + 1.88\log w}$
  $\Rightarrow m = 10^{0.464} \times w^{1.88}$
  Therefore $k = 10^{0.464} = 2.91$ (3 s.f.) and $n = 1.88$ (3 s.f.).

6  a  $r = -0.833$ (3 s.f.)

b  $-0.833$ is close to $-1$ so the data values show a strong to moderate negative correlation. A linear regression model is suitable for these data.

7  a  ‘tr’ should be interpreted as a trace, which means a small amount.

b  $r = -0.473$ (3 s.f.), treating ‘tr’ values as zero.

c  The data show a weak negative correlation so a linear model may not be best; there may be other variables affecting the relationship or a different model might be a better fit.

Challenge
Take logs of the data in order to compute all of the required relationships:

<table>
<thead>
<tr>
<th>x</th>
<th>3.1</th>
<th>5.6</th>
<th>7.1</th>
<th>8.6</th>
<th>9.4</th>
<th>10.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>3.2</td>
<td>4.8</td>
<td>5.7</td>
<td>6.5</td>
<td>6.9</td>
<td>7.6</td>
</tr>
<tr>
<td>log x</td>
<td>0.491</td>
<td>0.748</td>
<td>0.851</td>
<td>0.934</td>
<td>0.973</td>
<td>1.03</td>
</tr>
<tr>
<td>log y</td>
<td>0.505</td>
<td>0.681</td>
<td>0.756</td>
<td>0.813</td>
<td>0.839</td>
<td>0.881</td>
</tr>
</tbody>
</table>

Compute the PMCC for $x$ and $y$: $r = 0.999$ (3 s.f.).
Compute the PMCC for $x$ and $\log y$: $r = 0.985$ (3 s.f.).
Compute the PMCC for $\log x$ and $\log y$: $r = 1.00$ (3 s.f.).
Therefore the data indicate that $\log x$ and $\log y$ have a strong positive linear relationship. From the previous section, the data indicate a relationship of the form $y = ax^n$. 