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:

Time, *t*	1	2	4	5	7
Atoms, *n*	231	41	17	7	2
log *n*	2.36	1.61	1.23	0.845	0.301

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 $y = 2.487 - 0.320x$

$\Rightarrow \log n = 2.487 - 0.320t$

$\Rightarrow n = 10^{2.487-0.320t} = 10^{2.487} \times 10^{-0.320t}$

$\Rightarrow n = 10^{2.487} \times \left(10^{-0.320}\right)^t$

Therefore $a = 10^{2.487} = 307$ (3 s.f.) and $b = 10^{-0.320} = 0.479$ (3 s.f.).

5 a

Width, w	3	4	6	8	11
Mass, m	23	40	80	147	265
log w	0.4771	0.6021	0.7782	0.9031	1.041
log m	1.362	1.602	1.903	2.167	2.423

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:

x	3.1	5.6	7.1	8.6	9.4	10.7
y	3.2	4.8	5.7	6.5	6.9	7.6
log x	0.491	0.748	0.851	0.934	0.973	1.03
log y	0.505	0.681	0.756	0.813	0.839	0.881

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 $y = kx^n$.