Fillings – what are the chances?

1) For the data in the column headed **teethfillings**
(a) The variable is discrete

(b) The distribution could be Poisson if we assume the occurrence of the event 'teethfilling' to be random – in particular the length of time between fillings is not predictable. In addition to this, it is assumed that the average number of fillings is the same for any randomly selected student.

(c & e) Calculating the averages in Excel is fairly laborious – the easiest way is to insert an fx column into the table and do it in stages. The pivot table in the workbook will need to be copied in order that it can be manipulated.

Year group	Mean	P(X = 0)	P(1 ≤ X ≤ 3)	P(X > 3)
7	1.696	0.1834	0.7240	0.0926
8	1.541	0.2142	0.7149	0.0709
9	2.433	0.0878	0.6840	0.2282
10	3.59	0.0276	0.4897	0.4827
11	2.364	0.0940	0.6922	0.2138

(d) In order for the distribution to be reasonably considered Poisson, we need to be fairly confident that the mean will be fairly constant in the time period of interest. It is clear from the data that the mean changes according to year group, but it is a reasonable assumption that the mean remains constant throughout the year thus the distribution can be considered to be Poisson.

(f) Year 7 or 8. The average number of fillings increases quite dramatically between year 8 and 9 – and the probability of selecting a pupil with no fillings decreases significantly. A case for intervention prior to this is essential – a case could be made for targeting either year 7 or year 8.

(g) The data used is from a random sample of secondary school pupils through the CensusAtSchool online questionnaire. The questionnaire used is phase 6 – so the majority of responses will be 2 or 3 years old which could mean that patterns have changed since then and it might be worthwhile collecting some more up-to-date information. The sample could be improved by looking at the characteristics of pupils who go to the school and trying to match the sample more specifically – for example selecting pupils from an urban area or from a particular socio-economic group, but a random sample of this type in some cases can be as effective as collecting data in school. Obviously collecting data in school is more likely to be representative of the pupils of interest, however, if the intervention is going to be performed anyway and the conclusion of targeting year 7 and 8 pupils is a logical one, it could be considered that collecting school specific data is a waste of money and time.