Questions 1 to 3 below, on *Chapters 5* to *7*, form Tutor-marked Assignment M246 02. Question 1 is marked out of 30; Question 2 is marked out of 40; Question 3 is marked out of 30. Your overall mark for TMA02 will be the sum of your marks on all three questions.

You are advised to look again at the section entitled *General advice on TMAs* at the start of TMA 01 Part I.

Please send all your answers to your tutor, along with an appropriately completed assignment form (PT3). Be sure to fill in the Assignment Number on this form as

M246 02.

Question 1 (*Chapter 5 The normal distribution*)

(a) Using only your calculator and the printed tables of the normal distribution given in the Handbook, answer the following questions.

(i) The following data list gives the length (in cm) of a random sample of 25 worms. Assuming a normal model for the variation in length, obtain estimates for the parameters μ and σ, and hence state bounds between which 'almost all' recorded lengths might be expected to lie.

11.0	11.1	12.5	8.4	9.3	8.7	6.4	8.5	7.8	9.1
8.2	12.2	11.9	9.6	10.9	9.6	10.1	8.1	6.8	8.9
7.8	8.5	8.9	10.4	11.1					

[5]

(ii) If the random variable W is normally distributed with mean 180 and variance 40, calculate the probability $P(W > 200)$, and find the 80% quantile of W. [4]

(b) In an investigation into the clinical properties of the cholesterol-lowering drug cholestyramine, a sample of men were treated and the decrease in their total plasma cholesterol during the treatment was measured. Measurements varied from −92 to 453 (a negative value means there was an increase observed in total plasma cholesterol) with a mean of 131.2 and a standard deviation of 120.0. Using these estimates as the parameters in a normal model for the variation in cholesterol decrease, estimate (i) the proportion of individuals treated with cholestyramine who would experience a consequent decrease in cholesterol; (ii) the upper 90% point for cholesterol decrease. [6]

(*If you are interested you can discover more about this data set by typing* `??styra` *at the SSC prompt. The situation is not as simple as it is presented here: in fact, not all the men in the trial took the amount of cholestyramine that was intended.*)

(c) The data set **nevada** in your SSC data subdirectory contains data on the annual rainfall in Nevada City, California for 106 years from 1873 to 1978 (rainfall measured in inches). In this question we ignore the temporal properties of the data *though in general the chronological nature of the data set would be highly relevant to any statistical investigation of it.*

(i) By constructing a boxplot of the data, or otherwise, identify the one upper outlier, and state the year in which such heavy rainfall occurred. [4]

(ii) There is more than one way of eliminating particular items from a data set (see, if you are interested, the `purge()` family of routines by trying `ap("purge")` at the SSC prompt); but a particularly straightforward way of removing outliers is to use the routine `trim()`. Remove the upper outlier from the data set **nevada** and then (a) assess the adequacy of a normal model to express the variation in annual rainfall by drawing a