in late decimal places in your estimates are due to numerical rounding.

(d) The version of the logseries distribution that Tripathi and Gupta (1985) fitted to the vehicle occupancy data given in Table 6.6 on page 243 of the course text is a two-parameter extension of the simple model described here. Calculate the maximum likelihood estimate of θ for this simple model using the data of Table 6.6. [6]

Question 3 (*Chapter 7 Estimation with confidence*)

(a) In Exercise 7.9(a) in the course text, you found that a 90% confidence interval for the mean time between serious earthquakes worldwide (assuming an exponential model, and given extensive data) is from 359 days to 546 days. Interpret this finding, without using technical language. [4]

(b) The course text does not explicitly address the question of confidence intervals for the unknown binomial parameter p when the number of successes observed in n trials is 0 (in which case an obvious lower confidence limit for p is $p_- = 0$) or n (in which case an obvious upper confidence limit for p is $p_+ = 1$). In such cases, however, SSC has been trained to respond in a particular way. Use `?bexci` to discover what happens here, and find the 95% lower confidence limit for p, given an observed 10 successes in 10 trials. [4]

(c) In Table 12.18 on page 491 of the course text, one of the more famous classical data sets in statistics is summarized. The table gives the number X of annual deaths from horse-kicks suffered by Prussian army corps over the years 1875–1894. It turns out that the variation in the number of annual deaths is very well modelled by a Poisson distribution.

(i) Estimate the mean of the distribution and find a 95% confidence interval for the mean. [3]

(ii) The sample size is quite large: find an approximate 95% confidence interval for the mean using the central limit theorem and compare it with the interval you found in part (c)(i). [3]

(d) Data were collected on the height, diameter and volume of a sample of 31 black cherry trees in the Allegheny National Forest, Pennsylvania. The data were collected in order to find an estimate for the volume of a tree (and therefore for the timber yield), given its height and diameter. The data are stored in your data subdirectory under the name `trees`.

(i) Comment on whether the normal model seems to you to be a useful representation of the variation in tree height. Estimate the mean height of a tree and its standard deviation. [4]

(ii) Assuming a normal model is adequate for the variation in tree height, calculate a 90% confidence interval for the mean tree height, and a 95% confidence interval for the standard deviation. [5]

(e) The astronomer Simon Newcomb (1835–1909) performed many computational experiments into the speed of light, using mirrors at the US Naval Observatory, the Washington Monument and Fort Myer, Virginia, USA. The 66 readings from the third series of measurements recorded in 1882 are saved in the data set `newcomb` in the SSC data subdirectory.

Making no assumptions about any model underlying the variation apparent in the series, establish 90%, 95% and 99% confidence intervals for the underlying mean of the series; *you should be very clear about what random variable it is that you are investigating.* It may help you in your analysis to know that the 'true' value is now believed to be 33.02. [7]

---

999 278 105 45 20