MTMG37 Example Solution to Class Exercise 5

1. The mean of a large number of random variables should have a distribution that is approximately Normal according to the Central Limit Theorem. Each of the 100 means has expectation 4 and variance 4/100, so the approximating Normal distribution will have mean 4 and standard deviation 2/10. A histogram of the 100 means that I simulated is shown in Figure 1. The corresponding Normal density is superimposed and seems to be a good approximation.

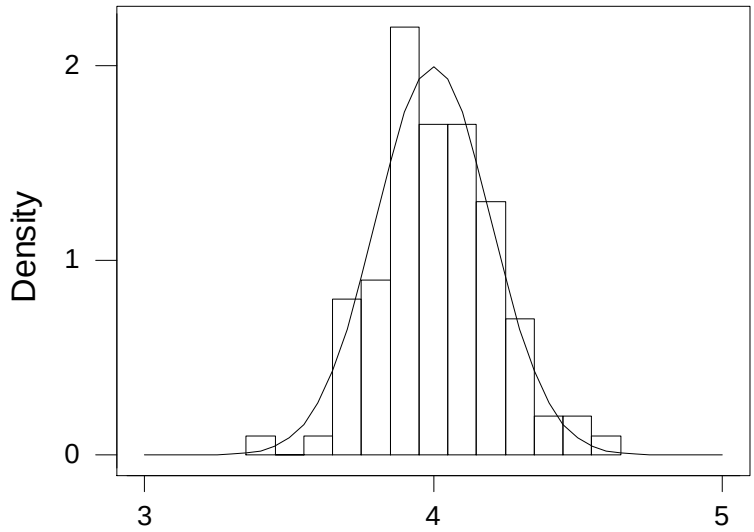

Figure 1. Histogram of 100 means and the approximating Normal density.

2. The usual point estimate is the sample mean, $\bar{x}$, which is 11.22°C for my data. If the true standard deviation, $\sigma = 5\,°\text{C}$, is known and the sample size is n then a $100(1-\alpha)$%-confidence interval is $(\bar{x} - c_\alpha\sigma/\sqrt{n}, \bar{x} + c_\alpha\sigma/\sqrt{n})$. For a 90%-confidence interval, $c_\alpha = 1.645$; for a 95%-confidence interval, $c_\alpha = 1.960$; for a 99%-confidence interval, $c_\alpha = 2.576$. The three confidence intervals for my data are therefore (10.40, 12.04), (10.24, 12.29) and (9.93, 12.51) degrees Celsius. In my case, only the 99%-confidence interval contains the true mean (10°C). The intervals are designed to contain the true mean $100(1-\alpha)$% of the time on average.

3. The standard deviation of my 100 sample means is 0.49. The theoretical standard error of the sample mean is $\sigma/\sqrt{n} = 1/2$. Of my 100 90%-confidence intervals, 90 of them contained the true mean (10°C), which is the proportion that we would expect in the long run.

4. The sampling distributions of my 100 sample means and medians are shown in Figure 2. The boxplots indicate that the two estimators are centred on the true mean but that the median has a greater spread and a skewed distribution. The greater spread is also reflected in the standard deviations: 0.49 and 0.57. Both estimators appear to be unbiased, but the sample mean is preferable because it has lower variance; the skewness has just arisen by chance in my samples. On the other hand, recall that the mean is more sensitive to outliers.

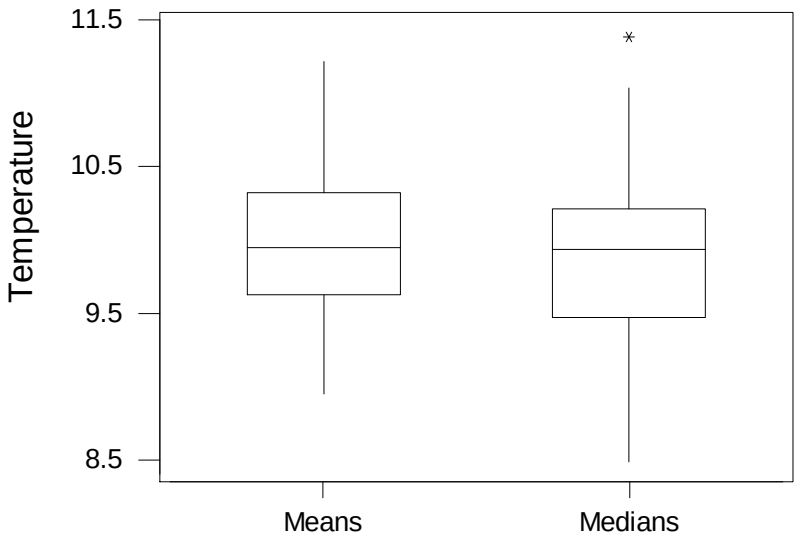

Figure 2. Boxplots of 100 sample means and medians.