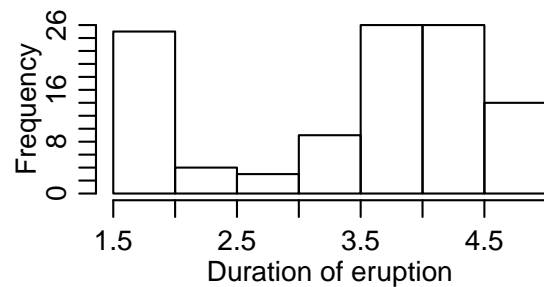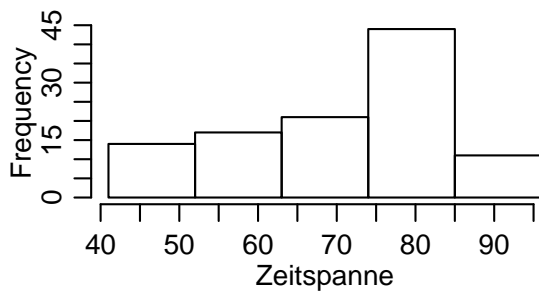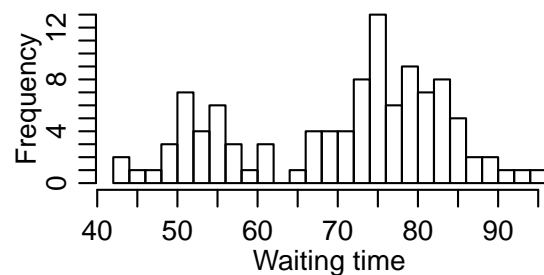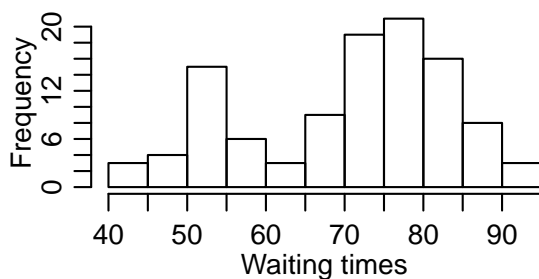# Solution Sheet 4

1. a) b) The first three histograms displayed below show the waiting times between successive eruptions of Old Faithful. One prominent feature of these histograms is that waiting times of about 55 minutes, and ones of about 70 to 85 minutes, are more frequent than waiting times of other lengths. Such a distribution with twin peaks is called *bimodal*.


If the widths of the classes are poorly chosen, this peculiarity of the geysir data remains undetected. In the third histogram, this is precisely what happens. This example illustrates the necessity of correctly specifying the widths – or boundaries – of the classes used.

Finally, a fourth histogram displays the frequencies of various durations of eruptions. In this histogram, the twin peaks are plain to see: „Either the eruption ends again immediately, or it lasts for at least 3 1/2 minutes". However, we *cannot* draw any conclusions on whether or not the length of an eruption is connected to the waiting time before it happened (i.e. whether or not the peaks in the histograms of Part b) correspond in any way to the peaks in the histograms of Part a)).

2. Note: The results will vary slightly, depending on the random numbers generated.

a) b) c) The histograms show relative frequencies. We can see quite clearly that the histogram becomes more regular as the sample size is increased. The relative frequencies in the histogram based on 10000 samples are fairly close to the theoretical probabilities.

The averages we obtain for the data shown here are 10.31 (100 samples) and 10.48 (10000 samples). This suggests that the mean is 10.5. Computing the standard deviations yields 2.977 and 2.979.
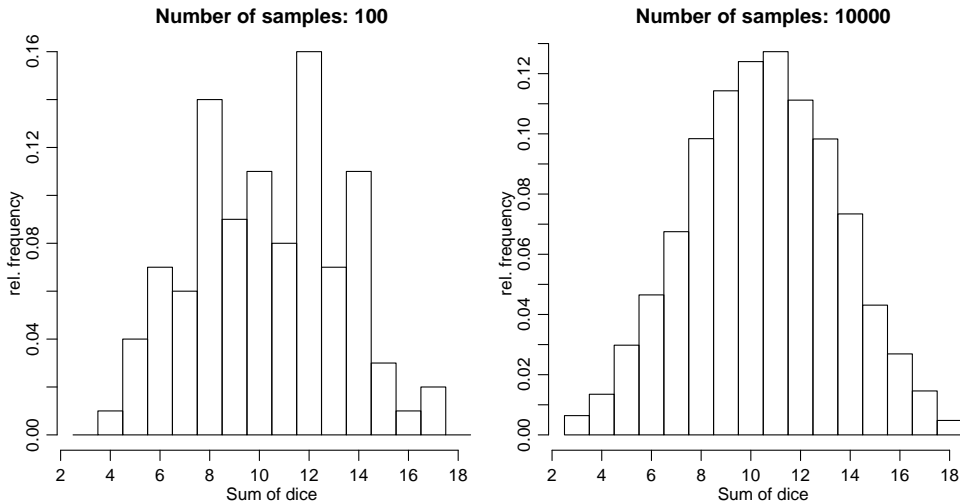
d),e) The average gain for the gambler is -0.22, which indicates that the game is not worthwhile in the long run. The way this gain progresses also indicates that the ultimate winner will be the casino and not the gambler (as usual), with high probability.
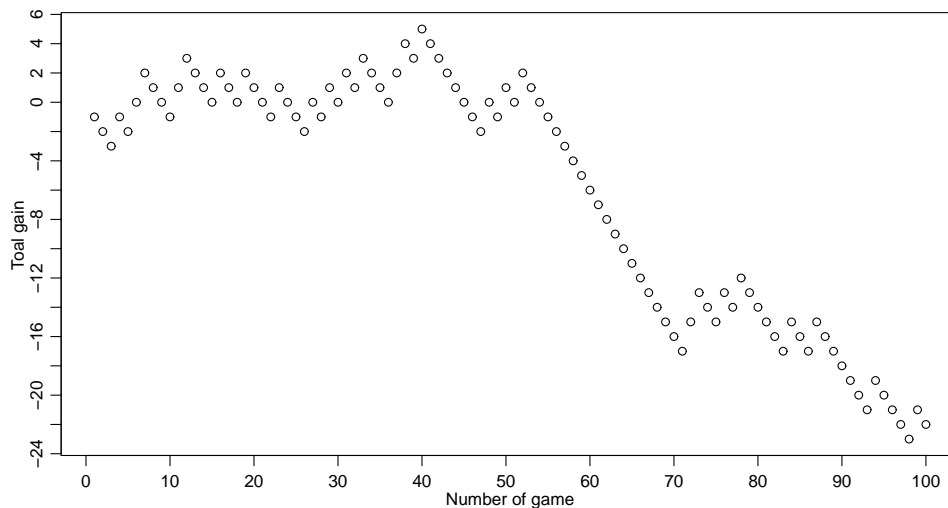
3. A2: Variance very small; median about $0.5 - 0.6$.
B5: Location shifted towards small values; extreme values in the upper range; positive skewness.
C1: Largest variance; median shifted towards small values; no extreme values.
D4: Location shifted towards large values; extreme values in the lower range; negative skewness.
E3: Almost symmetric; median about 0.5; larger variance than A2, but fewer extreme values than B5 and D4.