V31.0018: Statistics (Lab)
Matthias Kredler
June 6, 2005

Solutions for Homework 5 (Chapters 6 and 7)

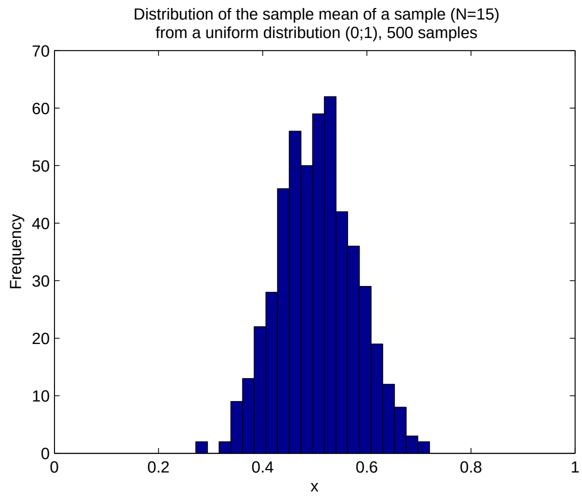
6.1 A *population parameter* is a characteristic of the *population*, a *statistic* is a characteristic of a *sample*. Hence the population parameter is almost always unknown, whereas a statistic is calculated from the sample observations, most often to obtain an estimate of the population parameter.

6.2 The *sampling distribution* of a sample statistic calculated from a sample of n measurements is the probability distribution of the statistic. You can think about the sampling distribution as follows: Take many random draws of size n from the population (*samples*). For each of these samples, calculate the corresponding *sample statistic*. In the end, collect all the sample statistics and draw a histogram. The histogram shows you the *sampling distribution* of the statistic.

6.8 You can generate a uniformly (0;1) distributed random variable in *MS Excel* as follows:

- Click into the cell where you want to generate the random variable.
- Type: `=RAND()`. Press `Enter`.
- Now you can see the random number that Excel has generated. You can copy the formula by clicking on the cell again and typing `Ctrl+c`. Then paste it into the other cells where you want to generate random numbers by marking these cells with the mouse and then typing `Ctrl+v`. Note that Excel will generate random numbers independent from each other in each of the cells where you copy the formula. It is somewhat disturbing that Excel creates new random draws once you perform certain actions, but never mind: they are still uniform (0;1) random numbers!

In the way described above, you can create a big grid of 500 rows and 15 columns containing uniform (0;1) random variables. Each row represents a sample of 15 observations. In the 16th column (e.g. cell P1 for row 1), you can calculate the sample average as follows: for row 1 (e.g.) type `=AVERAGE(A1:O1)`, which gives you the mean over the numbers in cells A1 to O1. Again, you can copy this formula from cell P1 to cells P2 to P500 with the copy-and-paste technique described before to calculate the other 499 sample means. In the end, copy the cells containing the sample means again and paste them into an *SPSS* sheet and create a histogram. The result should look similar to the figure presented below.



7.16 (d) In my opinion, the question should have been stated as follows:

In which way is the confidence interval from part (b) a better estimator than the point estimator of part (a)?

The answer to the modified question is the one given in the solution manual: The confidence interval contains the true value μ with a very high probability. However, the point estimate $\bar{x}$ will never hit the true μ exactly.

[However, point estimates are commonly accepted among statisticians, at least as much as are confidence intervals. Both estimators have their strengths and their shortcomings — either of them might be preferable in a specific situation, depending on the question to be answered. I feel it is not correct to qualify one as 'better' than the other as the solution manual does.]

7.27 (a) x is normally distributed:

- small sample ($n \leq 30$): $\bar{x}$ is t-distributed. Use the table of the t-distribution to create confidence intervals.
- large sample ($n > 30$): $\bar{x}$ is normally distributed. Use z-values from the standard normal distribution to construct confidence intervals.

(b) x follows an unknown distriubution:

- small sample ($n \leq 30$): We should first have a look at the sample distribution of x. If the histogram warrants the use of a normal distribution, we use the t-values to create confidence intervals. If the sample distribution has distinctive non-normal features, we cannot be sure about the distribution of $\bar{x}$; thus we cannot apply the methods introduced so far in this course. [We would have to use the non-parametric methods described in Chapter 14 of the text book.]
- large sample ($n > 30$): $\bar{x}$ is normally distributed since the *Central Limit Theorem* holds. Use z-values from the standard normal distribution to construct confidence intervals.