QUEEN MARY, UNIVERSITY OF LONDON

MTH 4106 Introduction to Statistics

Practical 10 20 March 2012

Today we will

- see how to use Minitab to obtain all the information necessary for estimating a population mean;
- simulate the sampling distribution of the sample mean and the sample variance when the population is standard normal;
- begin to investigate confidence intervals.

1 (Estimating a population mean) Enter the data about the breaking load of cotton yarn into a column (you can find these on the web page if you have not got them with you). Then use

`Stat` → `Basic Statistics` ▶ `Display Descriptive Statistics...`

Make notes on how the output in the **Session window** matches the quantities calculated during lectures.

You may prefer to make these notes in a **Report Pad**.

2 (Fresh Start) Save and close the **Report Pad** and erase the yarn data before starting the next part of the practical. You may prefer to `Exit` from Minitab and start again.

3 (The sampling distribution of the sample mean, from a normal distribution) Generate 100 rows of data from a standard normal distribution into each of columns `C1`–`C9`. We are going to treat each *row* as a random sample of size nine.

Put the sample mean of each row into `C10` by using

`Calc` → `Row Statistics...`

and then choosing `Mean`. Make sure to select of all the first nine columns. Rename `C10` as `SampMean`.

From lectures, we know the theoretical distribution of `SampMean`, including its expectation and variance. Write all of these in the box below before proceeding.

Show the data in `SampMean` on a histogram. Verify that they have the correct distribution by choosing the option `With Fit`. You will find that the histogram is shown on the same picture as a curve of the best-fitting normal distribution, whose mean and standard deviation are shown in the panel. What are their values?

4 (The sampling distribution of the sample variance, from a normal distribution) In a similar way, examine the sample variance of the 100 samples of size nine. Use

`Calc` → `Row Statistics...`

again, but this time choose `Standard deviation`. Put the result into column `C11` and rename it `SampSD`. Then square this, naming the result `SampVar`. Multiply `SampVar` by 8, and store this in a column called `Sxx`.

What is the expectation of the theoretical distribution of `SampVar`?

Now find the mean of `SampVar`. Is it close to the value which you predicted?

Plot `Sxx` on a (simple) histogram. This should look similar to the graph of the probability density function for χ^2 on 8 degrees of freedom. We have not met this distribution yet, but you will meet it next year.

5 (Confidence intervals: variance known) Now we are going to use each sample of size nine to construct a confidence interval for the mean. We admit that we know that the variance is equal to 1. So the standard error of the mean is 1/3, and the 90% confidence interval for the mean is (`L`, `U`), where

$$\texttt{L} = \texttt{SampMean} - 1.645/3 \quad \text{and} \quad \texttt{U} = \texttt{SampMean} + 1.645/3.$$

Make columns containing `L` and `U`.

We know that the true mean is zero. This is in the interval (`L`, `U`) unless both values are negative or both values are positive. Look at each of the 100 confidence intervals, and count how many of them do not contain the true value zero. Make a note of this number.

6 (Points of the z distribution) I know that 1.645 is the correct value so that $\mathbb{P}(Z \leq 1.645) = 0.95$ if $Z \sim N(0, 1)$. If I had forgotten this, I could have used Minitab to look it up. Use

`Calc` → `Probability Distributions` ▶ `Normal...`,

make sure that you have set the mean to 0 and the standard deviation to 1, then choose `Inverse cumulative probability`. Input the constant `0.95`, and do not ask to store the result. You should find that Minitab reports the value 1.645 in the **Session window**.