The role of Minitab in teaching and learning statistics

Nadarajah Ramesh
Department of Mathematical Sciences
University of Greenwich
n.i.ramesh@gre.ac.uk

Abstract

This paper describes how the statistical package Minitab is used in teaching statistics in our undergraduate programmes in Mathematics and Statistics to enhance student learning. How the sophisticated recent versions of Minitab can be used to help students understand statistical concepts, develop their statistical thinking and gain valuable skills in performing statistical analysis are discussed.

Introduction

Statistics is a subject which has developed a rich collection of different techniques and tools to analyse data. Data analysis and inference are an essential part of any statistical investigation or analysis. The analysis may be performed by hand or by calculators or more often nowadays using computers with the help of a statistical package. There are a number of statistical packages available for this task and Minitab is one of them. The earlier versions of Minitab, like many other packages, used a command based approach where you type in commands at the prompt and the output comes out in the same window. Although this command driven approach was very useful at that time and there was good literature on this, see [1] or [2] for example, it required considerable effort to use Minitab in teaching statistics. Since the Windows version came out it has become a lot easier and more convenient to use Minitab in teaching and hence it gained popularity in higher education among staff and students. With its sophisticated recent versions which contain excellent features and graphical facilities, Minitab has now become much simpler to use in teaching and learning statistics. It has evolved greatly over the years and developed into a system that has a vast array of statistical tools, powerful graphical displays, well structured help facility and documentation [3, 4]. The increased sophistication has made Minitab a widely used statistical package in business, industry and higher education.

At the University of Greenwich, all students on mathematics and statistics programmes are required to take a 30-credit statistics course in each of their first two years. Students on combined honours degree programmes involving mathematics have a choice to do either a statistics or a mathematics course at first year and a statistics or computing course at second year. We also have students from other departments (business, computing, economics, education) taking statistics as a minor subject. As they are all taught together and there are overseas and mature students from diverse ethnic backgrounds, the cohort has a varied mathematical background and ability. Teaching statistics in a way that is appealing to the whole audience is always a challenge. One has to find a balance between components of various teaching methods and techniques ranging from traditional classroom-based approach to modern teaching methods. Fortunately, with the help of the modern technology using the projection facilities from multiple sources, staff members are able to use statistical software packages during their usual lectures and tutorial sessions to deliver the lessons successfully and also to enhance student learning.

This paper describes how the package Minitab is used in teaching and learning statistics at our institution for undergraduate students on the mathematics and statistics programmes. The emphasis is placed on showing how Minitab facilitates the learning of statistics. First we focus on the use of Minitab in introductory statistics courses at first and second year levels to help students understand statistical concepts. We then move on to illustrate the use of Minitab with user written macros in learning more advanced computer intensive statistical techniques, before summarising the conclusions.

First and second year statistics modules

Minitab has excellent facilities to help students with their understanding of standard statistical methods like regression, ANOVA and time series. Students taking statistics courses in our programmes regularly use these facilities to assist them in their learning. Our course material for the statistics modules in the first two years has plenty of examples of Minitab analysis to help students develop their knowledge and statistical thinking. These analyses include exploratory data analysis, construction of confidence intervals, parametric and non-parametric tests, linear regression, and analysis of variance. Leaving the standard statistical techniques to one side, Minitab also has a good collection of statistical tools that are very useful in teaching some of the difficult concepts in statistical inference. One concept that some students find difficult to picture in their mind when they learn about statistical inference, especially in the first year of their programme, is the idea behind the central limit theorem and the sampling distribution of the sample mean. We found it useful to illustrate this concept using Minitab as given below by simulating data from various probability distributions.

Sampling distribution of the sample mean

The following outlines how a lesson on this topic is composed with the help of the random data generation facilities in Minitab. First students were introduced to the problem as given below to help them frame the idea in their mind about the topic they were going to learn in that lesson.

> Suppose that we are sampling from a population (not necessarily normal) of measurements which has mean μ and standard deviation σ. Suppose we select a random sample of size *n* from this 'parent' population, and calculate the sample mean $\bar{X}$. If we continue this procedure of taking samples of size *n* and calculate the sample mean on each occasion, we will eventually have a large number of such values. Since the random samples are very unlikely to consist of the same values, these sample means will vary and form a population with a distribution which is called *the sampling distribution of the sample mean*. What do you think this distribution would be?

Before students attempt to answer the question they were asked to put on their thinking hat and discuss the question briefly with a colleague sitting next to them. There were many answers that came from students which included 'it depends on the population used' and 'same as the parent population'. They were then asked to simulate samples from a known parent population and look at some summary statistics to study aspects of the distribution of the sample mean using Minitab. An example is given below.

> To examine the probability distribution of the sample mean of samples of size n=36 taken from a normal population with mean μ=100 and standard deviation σ =12, we take 200 such random samples using Minitab, calculate their mean and obtain summary statistics of the sample means.

```
MTB > RANDOM 200 C1-C36;
SUBC> NORMAL MU=100, SIGMA=12.
MTB > RMEAN C1-C36 C37
MTB > DESCRIBE C37
MTB > HISTOGRAM C37
```

Descriptive statistics: C37

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
C37	200	0	100.02	0.140	1.98	93.43	98.85	100.06	101.49	105.14

The summary statistics allowed them to compare the location and spread of the sampling distribution to that of the parent population. Box plots of the sample means and of a random sample of size 200 from the parent population are also used to help students understand the difference and similarities of the spread and location of the two distributions under consideration, as presented in Fig 1. They were also asked to check visually whether the sampling distribution appears normal by using the histogram.

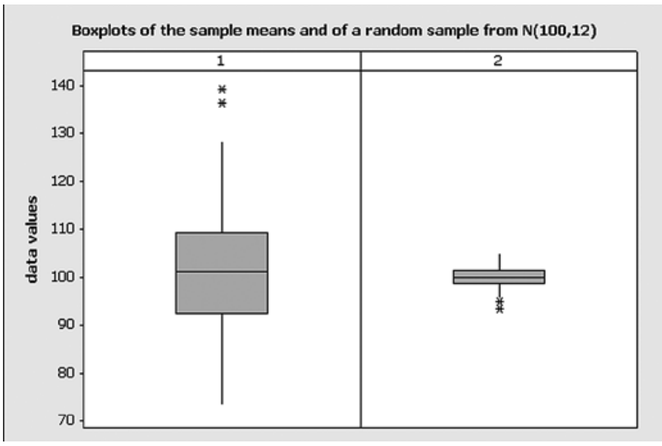

Fig 1 – Box plot of a random sample from a normal distribution with mean=100 and standard deviation=12 (left) together with that of the sample means of 200 samples from the same normal distribution (right)

This approach helped them form a basic understanding of the concept in their mind with some 'hands-on' experience to gain confidence in its application before they go on to learn the theory. At this point the formal theorem is introduced in the usual way, using Fig 2 to assist their understanding.

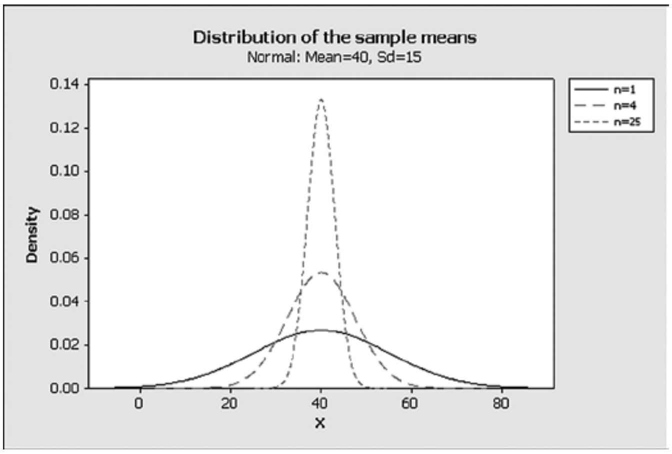

Fig 2 – Probability distribution of the sample means of different sample size (n=1, 4, 25) from a normal distribution with mean=40 and standard deviation=15

Having learnt the theory students would naturally want to experiment and decide whether this theory seems reasonable by doing some simulation. They were then asked to go back to the example and compare the calculated mean and standard deviation of the distribution of the sample mean with its theoretical counter-parts. The theoretical values for the mean and standard deviation of this distribution are 100 and $12/\sqrt{36}=2$. Students were able to see how close they get to these values and encouraged to experiment this with different parent distributions (non-normal) and sample sizes.

Students found it reassuring to find visually that the sampling distribution of the mean follows approximately a normal distribution by using the histogram, regardless of their parent distribution. This was illustrated in Fig 3 and 4, as part of their tutorial exercise, which shows the concept they learnt for a gamma distribution with shape=3 and scale=2 based on 200 random samples of size n=36. Fig 5 and 6 show the same for an exponential distribution with unit mean. These plots clearly illustrate the normality of the distribution of the mean regardless of the skewed nature of the parent populations, reinforcing the theory of central limit theorem they learnt in the lesson. Many students found this illustration of the theory very convincing and appreciated the use of Minitab in their learning. The lesson was concluded with a thought-provoking question which asked students to visualize the distribution of the sample mean when all data from the population were used in the sample.

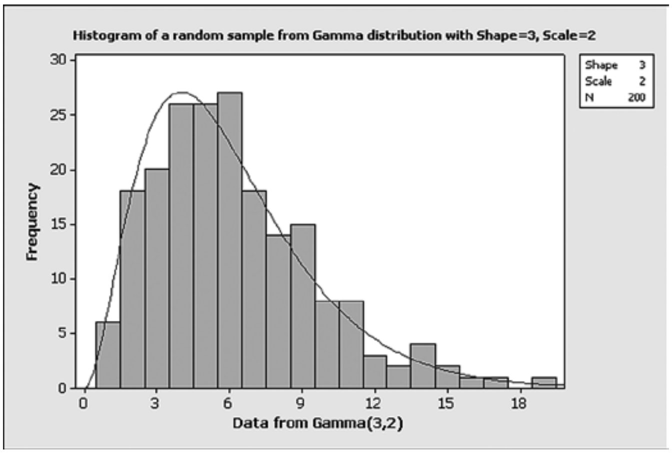

Fig 3 – Histogram of a random sample from a gamma distribution with its probability density function

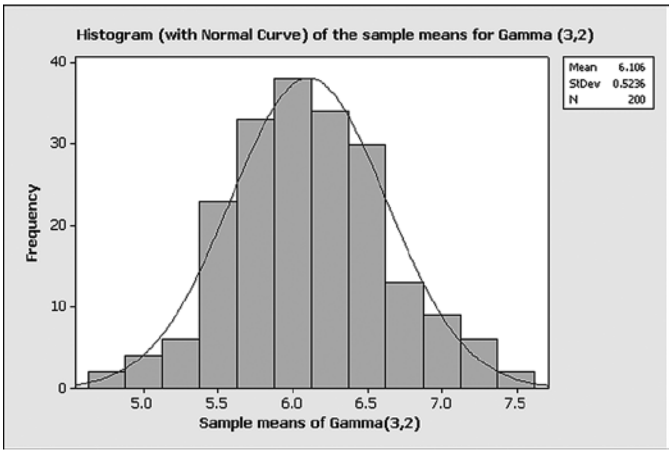

Fig 4 – Sampling distribution of 200 sample means from a gamma distribution with fitted normal distribution

In most statistics lessons, where the main objective is to enable students to understand statistical concepts and make inferences, we use Minitab to empirically verify statistical theory. In view of the varying degree of mathematical sophistication our students have, we use Minitab to help them understand difficult theoretical concepts and results using simulation, as illustrated in this section. With this approach, we are able to provide visual demonstrations that help reinforce some of the difficult or abstract statistical concepts.

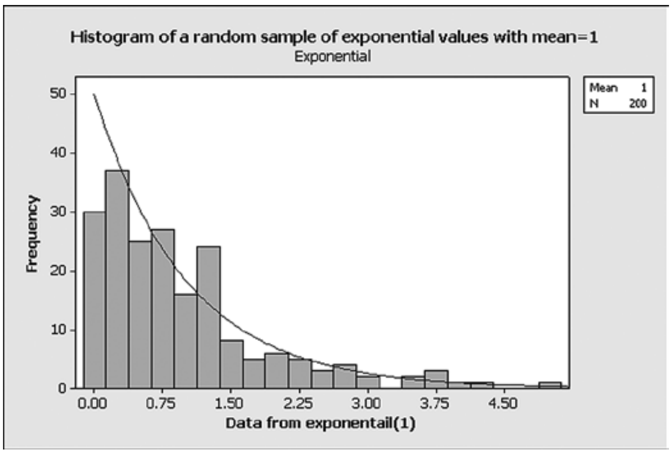

Fig 5 – Histogram of a random sample from an exponential distribution with its probability density function

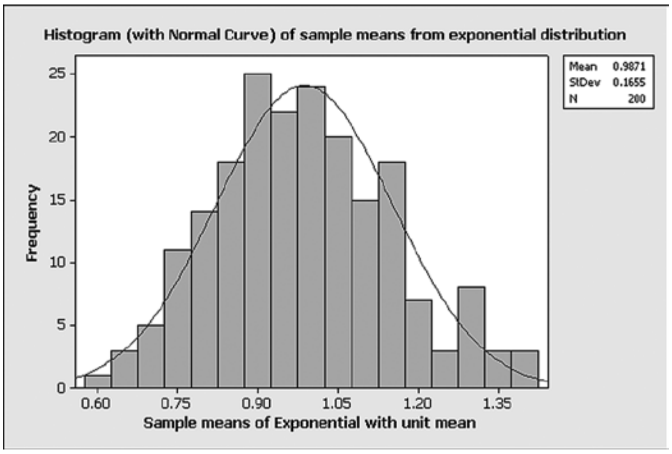

Fig 6 – Sampling distribution of 200 sample means from an exponential distribution with fitted normal distribution

Advanced functionality with user written macros

Minitab's functionality can be extended by user written macros which are like small programmes put together using a set of Minitab commands. The macros can be used to perform a specific more complex statistical analysis on datasets or carry out new statistical procedures or generate and manipulate data in different ways.

One minor drawback with the Windows version of the Minitab, or any other packages for that matter, is that students tend to ignore the command language, as they always use the pull down menu to do the analysis. To develop skills in writing macros one really has to have a good knowledge of the commands used in Minitab, otherwise it becomes a struggle to write good macros. In this regard, it is always a good idea to make students pay attention to the Minitab commands that appear on the session window as they use the pull down menu to do any statistical analysis. For this purpose, we make students enable the command line editor for their session window every time they start a Minitab session. This was emphasised in all lessons that involved students' use of Minitab. In addition, when helping students interpret their Minitab output we draw their attention to the commands used in the analysis to assist them develop a good knowledge of the command language.

We introduced students to this aspect of Minitab in the final year of their programme to enhance their learning in more advanced statistical techniques, possibly related to their final year project, although various other opportunities exist at level 2 with widely used statistical methods. At this stage students may be expected to learn more about topics such as assessing variability of estimates, confidence interval construction using Monte Carlo methods, significance tests etc to name a few. Re-sampling methods are increasingly used for assessment of variability and these are computationally-intensive statistical methods. We used the Minitab macro facility to illustrate the bootstrap techniques and permutation tests to help students gain good practical understanding of these concepts. The following outlines how a bootstrap significance test was carried out using macros in Minitab during a final year lesson when preparing students for their Research project.

Bootstrap significance tests using macros

Without getting too deep into the technical details of bootstrap, which can be found in [5] for example, a gentle introduction to this topic was first given as follows to whet their appetite.

> One of the original role of bootstrap resampling was in testing statistical hypotheses under weak assumptions about the mechanism generating the data. Suppose that increasing values of a statistic T indicate increasing degrees of inconsistency between data and the null hypothesis H_0 concerning data. If the observed value of T is t_{obs} , we aim to calculate its significance level
>
> $p = \text{Prob}(T \geq t_{obs} \mid H_0)$.
>
> The calculation is performed as if the null hypothesis were true.

The test procedure is then summarised in simple steps as given below, in an easy to understand manner, after exposing them to the problem. Students clearly needed this guidance to follow the correct approach to set up and understand the macros.

Procedure for comparing two population means

Let $X_1, X_2, \ldots, X_n$ be a random sample from Population 1 whose mean is $\boldsymbol{\mu_1}$ and let $Y_1, Y_2, \ldots, Y_m$ be a random sample from Population 2 whose mean is $\boldsymbol{\mu_2}$. We want to test the null hypothesis H_0: $\mu_1 = \mu_2$ against the alternative H_1: $\mu_1 > \mu_2$. The proposed test statistic is $T = \bar{x}$ (sample mean). First we calculate the observed value of T and let it be $\boldsymbol{t_{obs}}$ (sample mean of X values). Under the null hypothesis there is no difference between the X and Y populations, we combine the two samples to make a single sample of size (n+m) and follow the steps below.

1. Draw a sample $X^*_1, X^*_2, \ldots, X^*_n$ randomly with replacement from the combined sample.
2. Calculate the corresponding test statistic T^* from $X^*_1, X^*_2, \ldots, X^*_n$
3. Repeat steps (1) and (2), R times independently to get estimates $T^*_1, T^*_2, \ldots T^*_R$
4. Calculate the significance level of T as: $p = \text{Prob}(T^* \geq t_{obs}) = \{\# T^* \geq t_{obs}\}/(R+1)$

This test procedure is then illustrated using Minitab macros in the following example. Once this is done, students gain the confidence to develop similar procedure for a permutation test.

Example

The following table gives body weights (g) of a random sample of diabetic and normal mice. Use a bootstrap significance test to determine whether the mean body

Diabetic	38	39	35	34	37	33						
Normal	30	32	26	31	27	28	33	30	29	28	32	27

Table 1 – Body weight (g) of a random sample of diabetic and normal mice

weight of the diabetic mice is larger than that of the normal mice.

Here n=6, m=12 therefore (n+m) = 18 is the size of the combined sample. The sample mean weight of the diabetic mice is t_{obs} = 36. In R=99 bootstrap samples of size n=6 from the above combined sample, this value t_{obs} = 36 was only exceeded once, giving an estimated significance level of:

p = {1+1}/(1+99) =0.02

Therefore we reject the null hypothesis and conclude that there is evidence of higher mean body weight for the diabetic mice.

A lesson based on the above followed by a lab tutorial kept students actively playing with Minitab macros to explore similar problems. This approach provided students with good introductory knowledge of the topic and skills of developing their own macros for statistical investigations.

The following is a Minitab macro for this Mice Example.

```
GMACRO
# Bootstrap significance Test for Mice data.
# Resampling 99 times to test for difference
in mean body weight of the two groups
# You must have data in Columns C1 (Diabetic)
and C2 (Normal)
BOOT
NOECHO
LET C7=MEAN(C1)
LET K7=C7
ECHO
PRINT K7
NOECHO
STACK C1 C2 C3
LET K1=COUNT(C1)
DO K10=1:99
SAMPLE K1 C3 C4;
REPLACE.
LET K4=MEAN(C4)
STACK  K4 C7 C7
ENDDO
SORT C7 C7
NAME C3 'CombSample'
NAME C4 'BootSample'
NAME C7 'BootT*'
ECHO
PRINT C1-C4
PRINT C7
ENDMACRO
```

Conclusions

This paper discussed how Minitab is used to aid students in their understanding of some of the statistical concepts at our Institution. There are plenty of other instances throughout our programmes where Minitab is used to help students understand standard statistical techniques and to perform data analysis. Nevertheless, this paper focused on some specific topics not usually covered under standard statistical methods to highlight how statistical packages can be used to explain concepts, support students in their learning and help them gain valuable experience in performing data analysis. Use of Minitab certainly facilitates the learning of statistics to a greater extent. Moreover, students can easily employ Minitab to do all the calculations required for their statistical analysis and invest more time on understanding how various statistical procedures work and when to use them. Minitab has a rich collection of tools to help students perform investigations and analyses to reinforce the statistical theory they learn. Feedback from students shows that they like the use of Minitab to assist them in their learning and to better understand statistical concepts and their application. This approach of using Minitab in Statistics lessons is employed at all stages of our undergraduate mathematics and statistics programmes. As a result, a good proportion of students, many for whom statistics is not a favourite subject at the start of the degree, develop an interest in the subject and go on to take more statistics options at later stages of their programme.

References

1. Ryan, B.F, Joiner, B.L and Ryan, T.A (1985) *Minitab Handbook*. Second Edition. Duxbury, Boston.
2. Bond, K.M and Scott, J.P (1988) *Essential Business Statistics: A Minitab Framework*. PWS-Kent.
3. Minitab documentation. Available via: http://www.minitab.com/en-GB/support/documentation/default.aspx [Accessed 18 June 2009].
4. Meet Minitab 15 for Windows (Jan 2007). Available via: http://www.minitab.com/uploadedFiles/Shared_Resources/Documents/MeetMinitab/MeetMinitabEN.pdf [Accessed 18 June 2009].
5. Davison, A.C and Hinkley, D.V (1997) *Bootstrap Methods and their Application*. Cambridge University Press.