Use of Demonstrations and Experiments in Teaching Business Statistics

D. G. JOHNSON   d.g.johnson@lboro.ac.uk
Loughborough University, England

J. A. JOHN[†]   nye@stats.waikato.ac.nz
University of Waikato, New Zealand

Abstract. The aim of a business statistics course should be to help students think statistically and to interpret and understand data, rather than to focus on mathematical detail and computation. To achieve this students must be thoroughly involved in the learning process, and encouraged to discover for themselves the meaning, importance and relevance of statistical concepts. In this paper we advocate the use of experiments and demonstrations as aids to achieving these goals. A number of demonstrations are given which can be used to illustrate and explain some key statistical ideas.

Keywords: activity based learning, statistical thinking, use of spreadsheets

1. Introduction

There is a growing recognition that we need to change the way we teach business statistics. This was identified, at least in the USA, as long ago as 1986 with the first of a series of annual conferences on the theme of *Making Statistics More Effective in Schools of Business*. In the overview to the first conference Easton, Roberts and Tiao [1] note that there is "substantial dissatisfaction with much of current teaching of business statistics, especially as reflected in the poor selection of topics in popular textbooks and the limited opportunity for students to work with real data or to make serious use of statistical computing." As a result, many managers see statistics as simply a collection of complicated techniques that are not relevant to the process of management and decision-making. In John and Johnson [3] we advocate a problem centered approach to teaching statistical thinking based on realistic business examples. We identify three learning concepts found to be particularly useful in helping business students appreciate the importance and value of statistics. They are *interaction*, *discovery* and *experimentation*.

[†] Requests for reprints should be sent to J. A. John, University of Waikato, New Zealand.

Interaction means moving away from a totally passive form of teaching, where the lecturer talks and the students listen, to one that involves the teacher and students in active discussions and exchange of ideas. As with any difficult subject, students learn statistical principles more effectively by active involvement in the learning process. This type of involvement is not uncommon in short courses presented to shop floor workers, supervisors and managers. Lecturing for six hours a day for two, three or more days is not a viable option. Instead we recognize the need to involve the participants in the course, to share their experiences and concerns, and to discuss issues relevant to them and their organizations. We have found that this type of interaction is also very effective with business students, even in a one hour 'lecture' with a very large class of several hundred students.

Instead of telling the students what to do, or what happened, or what the formula is, *discovery* is about stimulating and encouraging the students to find out for themselves. For instance, given a particular set of data, we could tell students what decision the manager took on the basis of this information or how an experienced statistician might analyze the data. Why not challenge the students to say what they might have done if they were the manager concerned, or ask them to consider what information or evidence they might look for in the data? For example, why not let them discover the importance of variability and how to measure it, or fit a regression line or interpret a control chart? Ideas, concepts and methods are more effectively assimilated, and remembered, if discovered rather than imposed.

The third concept is *demonstrations* and *experiments*. These are essentially in-class activities that are designed to create learning that is richer and long lasting. Activity based learning has been widely advocated and used and there are many excellent resources available. For example, CAST is a computer-assisted course in introductory statistics by Sterling (see www-its.massey.ac.nz/CAST/CAST) that uses animations, simulations and other dynamic displays. It follows the same approach as that used in Griffiths, Sterling and Weldon [2]. The Rice Virtual Statistics Lab at www.ruf.rice.edu/∼lane/rvls.html is another comprehensive statistics learning support package. Sowey [6] and Scheaffer et al. [5] provide a number of activities that can be used in the classroom.

The aim of both experiments and demonstrations is to enrich learning by creating a different form of classroom activity, and to encourage participation in order to promote understanding and retention. In a recent book, John, Whitaker and Johnson [4] describe in detail a number of experiments to motivate, illustrate and explain various key concepts. In contrast, a demonstration is a more passive process where the lecturer uses

an appropriate abstraction or simulation to reproduce outcomes that illustrate some process or concept. Demonstrations are often computer-based, and with a little thought and ingenuity it is possible to create meaningful demonstrations for almost any topic that is taught in a statistics course.

2. Use of Spreadsheets

Spreadsheets such as Excel have become an indispensable tool for managers. They are widely used, and managers are comfortable in using them and are familiar with many of the facilities and features they offer. Therefore, we should embrace and build on this expertise and acceptance. While specialist packages, such as Minitab, SPSS and R, are undoubtedly better for any substantial statistical work, we would strongly advocate the use of a spreadsheet package in a first course in business statistics. Many students are already familiar with spreadsheets before they start their statistics course. Also, our experience is that students do not like using specialist packages. The complexity and multitude of tasks that these packages can perform usually means that students are forced to follow a prescribed set of instructions set down by the teacher. They find this tedious and uninspiring.

We acknowledge the fact that Excel functions and data analysis tools have been shown to produce errors in some circumstances. However, we have never found this to be a problem in the relatively straightforward analyses that most business students perform. For the most part, all they need is the ability to input, manipulate and graph data, and to do simple analyses; and these things can be done relatively easily with spreadsheets with very little likelihood of errors arising.

Another powerful capability of spreadsheets is that, with a little imagination, they can be used very effectively to demonstrate many statistical concepts. As such they provide a powerful teaching aid. We have created about 12 Excel demonstrations, making use of macro's and controls such as buttons and sliders, to create a dynamic demonstration. Most take no more than 5 or 10 minutes to run and can create an interesting diversion in the middle of a class, as well as reinforcing a key point. We shall describe below demonstrations that we use in our business statistics course.

3. Standard Deviation

The standard deviation is probably the first conceptually difficult notion that students will face in a business statistics course. It is totally unsatis-
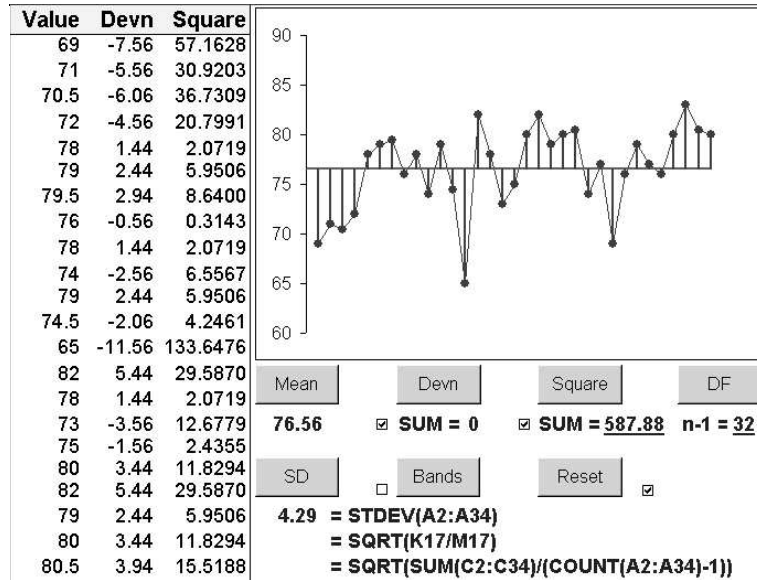
Value	Devn	Square
69	-7.56	57.1628
71	-5.56	30.9203
70.5	-6.06	36.7309
72	-4.56	20.7991
78	1.44	2.0719
79	2.44	5.9506
79.5	2.94	8.6400
76	-0.56	0.3143
78	1.44	2.0719
74	-2.56	6.5567
79	2.44	5.9506
74.5	-2.06	4.2461
65	-11.56	133.6476
82	5.44	29.5870
78	1.44	2.0719
73	-3.56	12.6779
75	-1.56	2.4355
80	3.44	11.8294
82	5.44	29.5870
79	2.44	5.9506
80	3.44	11.8294
80.5	3.94	15.5188

Mean   Devn   Square   DF

76.56   ☑ SUM = 0   ☑ SUM = 587.88   n-1 = 32

SD   ☐ Bands   Reset   ☑

4.29 = STDEV(A2:A34)
= SQRT(K17/M17)
= SQRT(SUM(C2:C34)/(COUNT(A2:A34)-1))

Figure 1. Standard deviation demonstration

factory to rely simply on computing a result from some mystical formula, whether this is done explicitly, or by using a built-in function in a calculator or spreadsheet. It is much more important to understand the rationale behind the standard deviation concept, and be able to interpret meaningfully the results obtained.

The way that we use this demonstration is as an adjunct to a discussion on what we mean by the concept of variation and how to sensibly measure it. Our starting point is to show the data in the form of a run chart, even though we might not be dealing with time series data. The spreadsheet is shown in Figure 1 for the first 22 values from data in John et al. (p2) [4]. At the outset the spreadsheet shows only the data in column A and the graph of the data without the mean line or deviations plotted. The discussion usually starts from the idea of using the range as a measure but the students have already come across the effect of outliers in, for example, boxplots and soon recognize the potentially misleading nature of the range. As the discussion evolves, the notion of deviations arises and the role of the mean (or some other average) becomes apparent. Using the buttons at the bottom of the graph, the progress towards the standard
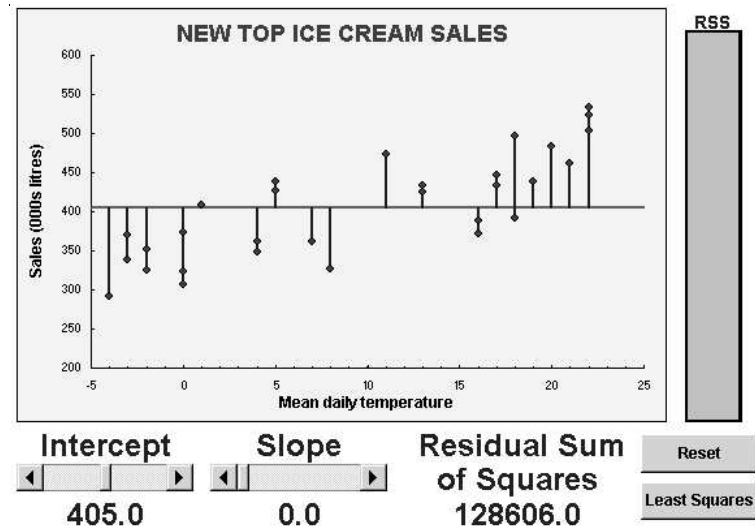
Figure 2. Scatterplot of sales and daily temperatures

deviation is computed by clicking on the buttons in turn to complete both the table (columns B and C) and graph. Formulae for computing the standard deviation in Excel can be revealed using a checkbox (to the right of the *Reset* button), and their use discussed.

There are still difficult issues to deal with, such as the concept of degrees of freedom and the need for the square root, but we have found that the graph in particular helps students to visualize what is happening. It is important to stress the interpretation of the standard deviation, for which we use the idea of a predictable proportion of points falling within 1, 2 and 3 standard deviations of the mean. The *Bands* button puts the $\pm 1\sigma$, $\pm 2\sigma$ and $\pm 3\sigma$ limits on the graph, and displays the proportions falling inside these bands.

4. Least Squares and R^2

The scatterplot in Figure 2 shows the monthly sales of ice cream (in 000's liters) plotted against the mean daily temperature for that month. These data are given in John et al. (p148) [4]. The mean of monthly sales over the 30 months is 405 (000's liters) and is represented by the horizontal line
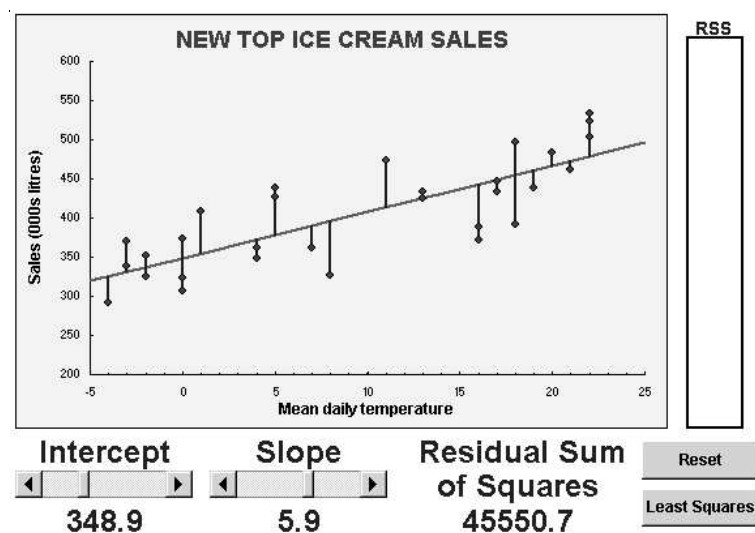
Figure 3. Least squares regression line

in the scatterplot. The residuals are also shown as vertical lines from the data points to this mean line. Hence, this initial plot is the same as that used in Figure 1 to demonstrate the standard deviation. Below the plot are sliders for the intercept and slope of the fitted line. Initially these are the values for the horizontal line, namely an intercept of 405 and a slope of 0. The residual sum of squares for this horizontal line is 128,606.

The aim of the demonstration is to change the values of the intercept and slope so as to give the line that minimizes the residual sum of squares (RSS). In our lectures we ask students to suggest how we should change these values. Initially, of course, the value of the slope should be increased and the intercept decreased. As it is difficult for students to know when they have reached the minimum RSS, a visual indication is provided by the vertical bar to the right of the graph. As the RSS decreases the height of the bar falls to zero corresponding to the least squares regression line. The demonstration shows how the intercept and slope determine the equation of a straight line, and illustrates very clearly the role of the residuals in determining a line of best fit. The least squares line can be obtained by clicking the *Least Squares* button, giving the line shown in Figure 3.

We also use this demonstration to explain the statistic R^2. The total sum of squares of monthly sales is 128,606, as shown in Figure 2. Clicking

Figure 4. One-way table of brand preferences

the *Reset* button will restore this initial plot. We now want to calculate how much of this variation is explained by the explanatory variable (temperature). We find that the most effective approach is to first ask students what the residual sum of squares would be if all the points were on the regression line. At the other extreme, if there were no (linear) relationship between the response and explanatory variables, we would expect the residual sum of squares to be almost unchanged, so that little variation has been explained by the regression. In our example, the residual sum of squares from fitting the regression line is 45,550.7, as given in Figure 3. Hence the proportionate amount of variation explained is

$$R^2 = \frac{128606 - 45550.7}{128606} = 0.6458,$$

which is the *R Square* value given by the Excel regression procedure in the data analysis tools.

5. Chi-Squared Analysis

We find that many students have difficulty in understanding what the chi-squared statistic represents, and what the associated probability means.

This demonstration aims to overcome this difficulty. Separate sheets are given for one-way and two-way tables. In Figure 4 we show the calculations for a one-way table with three categories. The data refer to the results of a market research survey of 150 customers who chose one of three alternative brands of deodorant; see John et al. (p81) [4]. While there might appear to be a greater preference for brand B compared to the other two brands, we need to determine whether this is a real difference or simply due to chance. In Figure 4 a chi-squared value of 4.68 on 2 degrees of freedom has been obtained. From chi-squared distribution tables this value corresponds to a probability of just under 10%. Using the CHITEST function in Excel an exact probability of 0.096 is obtained. But what does this probability mean?

In this spreadsheet we can use the sliders below the table to change the values in cells D3 and D4, the numbers preferring brands A and B. Cell D5 is automatically changed to maintain the total sample size (150) and the chi-squared value and probability are re-calculated continuously. Students can then see that if the actual numbers are almost equal the chi-squared value is close to zero and the probability close to 1. As the actual numbers become more and more unequal, the chi-squared value increases and the probability decreases. They can also see at what point the probability becomes 'small', and so how different the brand purchases have to be to indicate a 'real' difference in brand preference. It is also relatively straightforward to discuss degrees of freedom in terms of the maximum number of sliders that can be used to change the observed frequencies. In this way students can readily get an appreciation of what the chi-squared statistic is and what p-values and degrees of freedom mean.

6. Control Charts and Special Causes

The final example is both a demonstration and a tool that can be made available to students to facilitate the drawing of control charts. Unfortunately, Excel does not have an option to draw control charts, which must be done by adding further data (lines for the mean and control limits) to a run chart. It is, however, relatively straightforward to create a macro that does this, and avoids students getting hung up on the intricacies of graph drawing in Excel. We have created separate sheets to demonstrate the construction and use of X-bar and R charts, individual charts, c-charts and p-charts. Figure 5 shows the sheet for an individuals chart. The time unit (hours, days, month etc.) is selected from a drop-down list box at the top left of the sheet. In this case the unit is *samples* as the data refer to samples taken from a hopper at 20-minute intervals over a single shift,

see John et al. (p267) [4]. The data are entered in the column headed
Value. The *Set-up samples* (in this case 10) is the number of samples used
to calculate the control chart limits. The mean and standard deviation
are calculated and displayed as the data are entered. After the first 10
samples have been entered the limits are set and the subsequent values are
monitored using these limits. The control chart can be viewed and printed
by clicking the *Graph* button.

The control chart is shown in Figure 6. The dotted center line represents
the mean value (27.3) and the continuous lines are the control limits. The
graph will automatically highlight any potential special causes, either by
indicating points outside the control limits or (more importantly) by show-
ing a pattern in the points within the control limits. From experience, it
is easy to overlook patterns in the data that might indicate a special cause
occurring, and the facility to show this automatically is a useful feature.
It is not very clear from the black and white diagram in Figure 6, but on
the spreadsheet different color coding is used to distinguish between the
five different patterns; see John et al. (pp242-244) [4]. In this example, we
suspect that the process may have gone out of statistical control during the
period covered by samples 12 to 21 as there is a sequence of 10 points all
falling above the mean, despite the fact that there were no points outside
the control limits.

7. Conclusions

Our objective in this paper is to make a case for the use of experiments
and demonstrations in teaching business statistics. In a sense, the con-
text of *business* statistics is representative of many other disciplines where
statistics is taught to students who may not have a strong mathematical
background and, irrespective of this, are mostly not interested in the math-
ematical foundations of the subject. Their motivation should be towards
being able to think statistically and, if they understand what they are doing
and why they are doing it, perform relatively straightforward analyses in a
correct and meaningful way. For such students, formulae and algebraic ma-
nipulation will almost certainly deter rather than stimulate an enquiring
approach to data analysis. One way to maintain an interest in and ac-
ceptance of statistics is to use teaching methods to which they can relate,
such as discussion, exploration and an imaginative use of real activities and
computer based resources, in particular experiments and demonstrations.

We have not attempted to survey the very many examples and illustra-
tions reported in the literature. Rather, we have endeavored to give a flavor
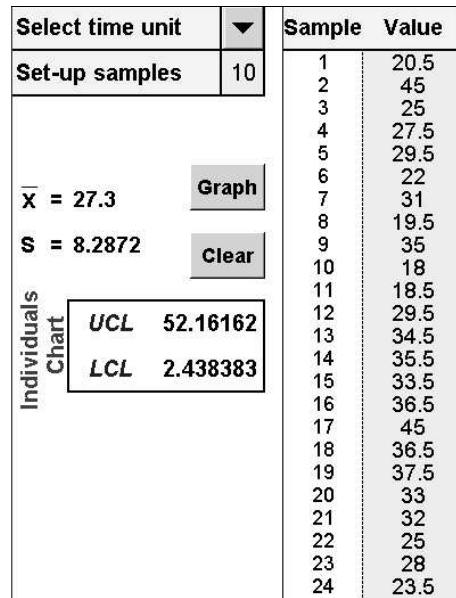of some of the demonstrations that we have found useful. All of our demon-

Select time unit	▼	Sample	Value
Set-up samples	10	1	20.5
		2	45
		3	25
		4	27.5
		5	29.5
		6	22
$\overline{X}$ = 27.3	Graph	7	31
		8	19.5
S = 8.2872	Clear	9	35
		10	18
		11	18.5
		12	29.5
UCL 52.16162		13	34.5
		14	35.5
LCL 2.438383		15	33.5
		16	36.5
		17	45
		18	36.5
		19	37.5
		20	33
		21	32
		22	25
		23	28
		24	23.5

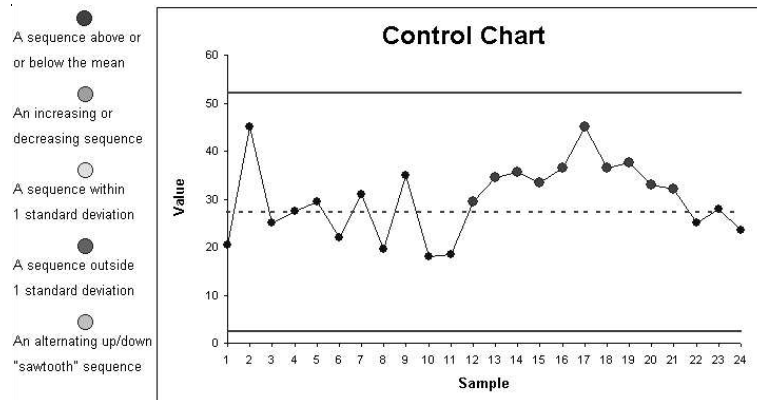*Figure 5.* Samples of data taken during a shift

Figure 6. Control chart of shift data

strations are 'home grown'. They are usually Excel based, and because of its universality they can be made available to students without the need for specialized software.

Experiments are generally more integral to the course and may take a full period to run, such as with the red beads experiment; see John et al. (pp4-12) [4]. For this reason they should be carefully selected to reinforce key ideas that tend to recur throughout the course. Demonstrations on the other hand can be developed and used rather more extensively and may just form a '10-minute interlude' in a class to give visual support to a difficult point, often in the context of a meaningful example or case study.

Demonstrations need not be complicated, although we find that the control facilities that are available on the *Forms* toolbar in Excel can be used to quickly create a powerful and professional looking demonstration. The ability to use Macros in Excel is an advantage, although much can be done by using the *Macro record facility*, without the need to be an expert in Visual Basic. If you are not convinced, try it for yourself. The topics that you pick for demonstration is very much a personal choice, but in our experience you can usually find some way to use Excel to simply reinforce most things. In general, students like to see something different from the usual talk based around a few overhead transparencies, and they tend to remember such things. We believe that it enhances their understanding and performance, and usually reflects well on you, the teacher.

References

1. G. Easton, H.V. Roberts and G.C. Tiao. Overview And Summary. Proceedings of Making Statistics More Effective In Schools Of Business, University of Chicago, 1986.
2. D. Griffiths, W.D. Sterling and K.L. Weldon. Principles and Practice of Statistics. Wiley, New York, 1998.
3. J.A. John and D.G. Johnson. Statistical Thinking for Effective Management. Proceedings of the 6th International Conference on Teaching Statistics (ICOTS6), Cape Town, 2002.
4. J.A. John, D. Whitaker and D.G. Johnson. Statistical Thinking for Managers. Chapman and Hall/CRC, London, 2001.
5. R.L. Scheaffer, J. Witmer, A. Watkins and M. Gnanadesikan. Activity Based Statistics. Springer, New York, 1996.
6. E. Sowey. Striking Demonstrations in Teaching Statistics. *Journal of Statistics Education*, 9(1), 2001.