# TEACHER'S CORNER

In this department *The American Statistician* publishes articles, reviews, and notes of interest to teachers of the first mathematical statistics course and of applied statistics courses. The department includes the Accent on Teaching Materials section; suitable contents for the section are described under the section heading. Articles and notes for the department, but not intended specifically for the section, should be useful to a substantial number of teachers of the indicated types of courses or should have the potential for fundamentally affecting the way in which a course is taught.

# A Design of Experiments Workshop as an Introduction to Statistics

A. J. LAWRANCE

A workshop introduction to statistics is described. Students collect data from themselves in a small factorial experiment to investigate characteristics of short-term memory. With some guidance, student groups both choose the variables to investigate and organize the collection of their data. They learn to tabulate and display the data in ways to understand the ideas of main effects and interactions; the disadvantages of one-factor-at-a-time experiments are made apparent. The data collected lend themselves to later use in binomial regression analysis.

KEY WORDS: Classroom experiments; Data presentation; Factorial designs; Interactions; Statistical workshop.

## 1. INTRODUCTION

This article describes the statistical contribution to a first-semester workshop course for all first-year students in the School of Mathematics and Statistics at Birmingham University. Very little prior statistical knowledge can be assumed at this stage because of nonuniformity of background, and because the first statistics course comes in the second semester of the first year. With the central importance of the design of experiments in statistics, this seemed a good opportunity to introduce it early, on an intuitive basis, and provide students with the opportunity of designing their own experiments, carrying out the data collection, and summarizing their own results. The subject matter chosen for the work was short-term memory, and the factors which affect it. Short memory is measured by recall from lists of words under controlled conditions. Suggestion of this topic is contained in Anderson and Loynes (1987), where the idea ascribed to Dr. A. Bowman (see Bowman 1994). Our experience is that it produces a lively, highly charged workshop, especially when factors of influence and their levels are being discussed. Collection of the data soon identifies the good organizers and the more extrovert, but by being collected from among the students themselves, there is strong motivation for all students to make sense of them. The statistical focus revolves around seeing the sense of varying several factors simultaneously in a small factorial design, and then drawing out the effects of the factors individually, and looking for any combination (interaction) effects. There is some opportunity for deploying any knowledge of the binomial distribution because it can model the number of words recalled. One set of student data is used to exemplify what is required from the workshop. An added bonus is that the data collected can be used in subsequent years in a generalized linear modeling course for logistic analysis of factorial experimental data.

A. J. Lawrance is Professor of Statistics, School of Mathematics and Statistics, University of Birmingham, Birmingham B15 2TT, England.

## 2. SHORT-TERM MEMORY BACKGROUND

Students are given a brief motivating presentation about short-term memory, including a demonstration experiment inflicted on an unsuspecting graduate assistant to the course. Short-term memory (stm) is said to be the ability to remember over periods of up to about 1 minute; it is supposed to be the encoding of up to about 12 items acoustically or visually in the brain, without considering their meanings. Unless passed from stm to long-term memory (ltm), items are forgotten by decay over time and replacement or interference from more recent items. Students are encouraged to find out more background on their own from psychology texts. To assess short-term memory, subjects are shown a list of unconnected words that is then withdrawn; they are then asked to recall the words, with the primary result being the number correctly recalled. The main point is to investigate how recall ability depends on such factors as the length of word list, the length of words, and the time of study.

## 3. DESIGN OF THE EXPERIMENT

This is the most controversial aspect of the workshop as far as the students are concerned—what factors to use and at what levels. Possible factors include the following:

study time of the word list
delay time after viewing to testing of recall
length of words used or number of syllables in words
number of words presented in the list
simultaneous or sequential presentation.

Students will think of others as well. For simplicity fairly strong advice is given that the factors should only be at two levels, and that these two levels be sufficiently different as to

 

Table 1. Exemplary Set of Student Data

| Factor levels | Words recalled by six students | | | | | |
|---|---|---|---|---|---|---|
| $WL^+$ $ST^+$ $LL^+$ | 10 | 9 | 6 | 6 | 7 | 6 |
| $WL^+$ $ST^+$ $LL^-$ | 5 | 6 | 6 | 6 | 6 | 4 |
| $WL^+$ $ST^-$ $LL^+$ | 5 | 7 | 5 | 8 | 6 | 6 |
| $WL^+$ $ST^-$ $LL^-$ | 5 | 4 | 6 | 4 | 4 | 6 |
| $WL^-$ $ST^+$ $LL^+$ | 10 | 9 | 7 | 10 | 8 | 11 |
| $WL^-$ $ST^+$ $LL^-$ | 6 | 6 | 6 | 5 | 6 | 6 |
| $WL^-$ $ST^-$ $LL^+$ | 9 | 4 | 7 | 5 | 4 | 7 |
| $WL^-$ $ST^-$ $LL^-$ | 6 | 6 | 3 | 6 | 6 | 5 |

have an effect—to avoid the experiment being a flop! Ease of administering the tests is also a rather vital concern. Discussion groups of about 25 seem manageable, which then combine with a second group to decide on factors and levels. A certain amount of moderation seems to be required at this stage to ensure law and order! The final grouping of 50 allows enough subjects for the tests in the designed experiment. A pilot run of a few tests is suggested to confirm feasibility and refine experimental technique.

Construction of the word lists is an important feature of the design considerations. Perhaps ideally each subject should be shown a different list of the required character. In practice this would be too time consuming in construction. The guiding requirement is that the subject should have played no part in constructing the list on which he or she is tested. It is preferable that a different set of students is used as the test subjects.

With the use of three factors at two levels each, eight different types of test, a $2^3$ factorial, is suggested in the handout as the experimental design; each type of test is carried out by 5 or 6 students, using 40 or 48 from a group of 50.

Emphasis is placed on the efficiency of investigating all three factors simultaneously, rather than tediously one at a time, and yet being able to investigate the factors individually, as well as jointly in pairs.

## 4. COLLECTION OF THE DATA

The data are collected and checked by the students, and then copies distributed to each of them; they work in pairs to produce reports. One set of student data is chosen here for illustration, and in this the three factors were as follows:

WL = word length; 3 syllables or 1 syllable ($WL^+$, $WL^-$)
ST = study time; 30 seconds or 15 seconds ($ST^+$, $ST^-$)
LL = list length; 12 words or 6 words ($LL^+$, $LL^-$).

This group of students appropriately decided to use everyday words, to present them simultaneously, and to have immediate testing. They also required constant lighting conditions and would have liked silent surroundings, but this was impossible! The data collected are given in Table 1.

## 5. ANALYZING THE DATA

Most of the analysis is to be done on the basis of "guided common sense" rather than explicit knowledge of statistical analysis; in particular, the handling of variation will be

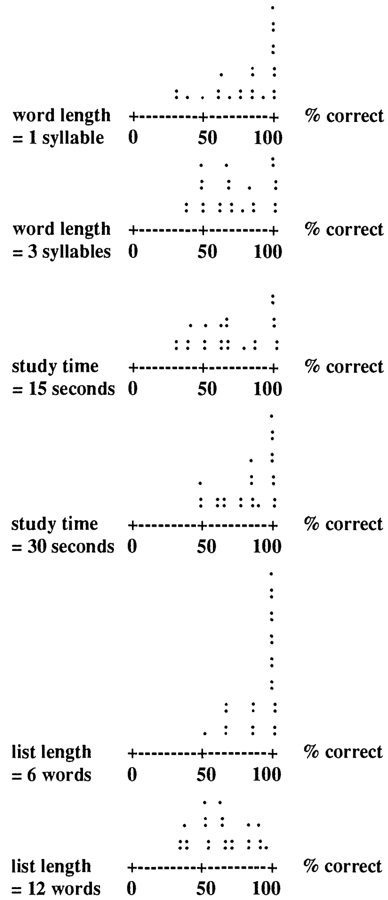


Figure 1. Dot Plot Comparisons of Individual Factor Effects.

very rudimentary. In the data set of Table 1 the direct analysis of the number of words remembered, perhaps averaged over each type of test, will not be appropriate; there are two lengths of test involved, 12 and 6 words. Any student missing this point is of the nonthinking variety. The data are best expressed as percentages of words remembered, with a view to estimating the probability of remembering a word.

As to understanding the data, the main advice is that the data consist of eight groups, and that they could be combined into larger groups, reflecting say $WL^+$ and $WL^-$. For example, the $WL^+$ group would consist of the percentages of all tests with word length of three syllables, and $WL^-$ corresponding to one syllable. Students should also note that this is a fair comparison of the WL levels because the other factors are balanced over their paired levels; the concept of balance will need some understanding. A pair of frequency displays, such as dot plots, followed by aver-

Table 2. Percentages of Words Remembered for the Eight Types of Test

| List length | 6 Words | | 12 Words | |
|---|---|---|---|---|
| Word length | 1 Syllable | 3 Syllables | 1 Syllable | 3 Syllables |
| Study time | | | | |
| 15 sec | 89 | 81 | 50 | 51 |
| 30 sec | 97 | 92 | 76 | 61 |

Table 3. List Length and Study Time Combination Results as Percentages

| List length | 6 Words | 12 Words | Effect of list length |
|---|---|---|---|
| Study time 15 sec | 85 | 51 | −34 |
| 30 sec | 94 | 69 | −25 |
| Effect of study time | 9 | 18 | Difference = 9 |

aging and differencing the plus and minus levels, gives basic inference about the effect of word length. The three pairs of dot plots are given in Figure 1.

It was fairly easy for most students to make the correct inferences = strong effect of list length (30% difference), substantial effect of study time (14% difference), and modest effect of word length (7% difference), with differences being in the expected directions.

A further aspect of understanding the data more widely is summarization, and tabulation was suggested, aggregating words remembered over the eight types of test. The three factors suggest a three-way table, explained as two two-way tables. From the point of view of summarization this also suffers from the effect of the two different list lengths, and so presenting a three-way table in terms of percentages correct is more appropriate. For the student data used here, such a summary is given in Table 2.

This tabular summary points clearly to the effects of the factors individually, as seen previously; the exercise in table construction is useful in itself. The table can also be used in calculating combination (interactive) effects of the factors, and it could be presented geometrically using a cube.

The idea of combination effects was harder to motivate, and an analogy with taking several medicines in combination was used. This notion was introduced in terms of the effect of one factor depending on the level of another factor, using difference of percentages for the factors individually. For instance, from deriving the two-way table of counts for list length and word length, and then converting it to percentages, one obtains Table 3.

The effects of study time are 9% and 18% at the two levels of list length, as shown by Table 3; there is thus a combination effect of 9%. Similar calculations for study time and word length give 10%, but for list length and word length the figure is 0%. This latter figure can be taken at face value, but the other two figures can be presumed approximately equal, and fairly modest.

Overall, as far as the layman's appreciation of short-term memory is concerned, the list length and study time effects seem natural. It is a little surprising that word length makes no difference to short-term memory, at least not at the one and three syllable levels. The combinative effects are perhaps less than one would expect. In retrospect the list length of six words turned out too short because a high proportion of subjects remembered five or six words. Such imperfections must be tolerated!

## 6. FURTHER ASPECTS OF THE WORKSHOP

The main aspects possible to impart over two 2-hour sessions have now been covered; students are required to prepare reports, working in pairs, and later may choose to give their course presentation on this topic. Some students may have prior statistical knowledge, and can perhaps identify the binomial structure of the data, with indices 6 and 12; binomial variation could be investigated over the eight groups of six values. This aspect can be picked up in their subsequent probability course. The distributions underlying the dot plots in Figure 1 are nonstandard mixtures of binomial distributions, again a topic for possible later investigation.

The greatest future opportunity from the workshop data is its analysis in a subsequent generalized linear modeling course, employing the binomial assumption. The standard logit model with additive factors implies the relation,

$$\log\{p/(1-p)\} = \mu + \beta_{\mathrm{WL}} + \beta_{\mathrm{ST}} + \beta_{\mathrm{LL}}$$

where $p$ is the probability of remembering any particular word. Conclusions based on this model are that list length is strongest, followed by study time, with word length having no effect—thus in agreement with the earlier informal analysis. Interactions appear to be weak, with only study time and word length having some effect.

*[Received October 1995. Revised November 1995.]*

## REFERENCES

Anderson, C. W., and Loynes, R. M. (1987), *The Teaching of Practical Statistics*, New York: John Wiley.

Bowman, A. (1994), "Teaching by Design," *Teaching Statistics*, 16, 2–4.