CHAPTER 14

"Students learn more from their own mistakes, than by the correct things they are told."

Tips for Teachers

Although written to accompany our statistics course, we wrote this book to stand on its own. This entailed removing all of the practical exercises and changing the order of the topics slightly. However, students do not learn by being told how to do things correctly, they learn by making mistakes. Our course works because almost all of the class time is spent inducing the students to make mistakes.

This chapter teaches you how to induce students to make mistakes. It is not about how to do it right; it is about how to do it wrong so that students can go into the world and start making their own mistakes. We learned from our blunders, over a period of 20 years. This course condenses 20 years of blunders into a 2-week course so that students will make much more interesting and revealing mistakes in their own work. They need to develop a feel and an understanding for the procedures, rather than just agree that the formulas are logical.

Too much repetition makes a text boring. However, repetition of the same concepts in different situations and with different examples is vital in the classroom. Students do not absorb new concepts, no matter how simple, in the first lesson. Therefore, it is essential that each new level of complexity builds on concepts developed in the previous sessions.

It is not enough that students know of the concepts. In order to be effective, the course must make the students feel them in their bones (Magnusson 1977). Table 14.1 should be presented at the beginning of the course so that students can monitor their progress as they learn. The teacher can also use these concepts and suggestions as the bases for quick quizzes at the beginning of each lesson. In this chapter, we show how the previous chapters fit into the course outline, and we give details of some of the exercises that we have found useful when teaching these lessons. Each session should last about 3 hours, and the course requires about 20 sessions.

We believe that courses should teach both exploratory and confirmatory techniques (Tukey 1980). They should also teach students how to draw conclusions rather than just make decisions (Tukey 1960). In an undergraduate course, rather than a remedial post-graduate discipline, it is worthwhile to teach most of the important techniques without using mathematics (Magnusson 2000b). Remember that there are many forms of intelligence, of which mathematical ability is but one, and that extreme development of any one form of intelligence is not an indication of an individual's capacity to succeed (Goleman 1995). Wilkinson (1999) summarized the opinions of a group of noted psychologists and statisticians as to what they think the most important concepts that apply to all statistical analyses are. Those concepts form a good basis for an undergraduate or post-graduate course.

Our course is repetitive. Statistical concepts do not change; they just expand or become clearer. The course has only two overriding aims: To show what the *P* (probability) values scattered through the literature mean; and to show that people can generally only visualize information in two dimensions (marginally three, when one dimension is categorical), and that the function of the apparently more complicated analyses is to reduce dimensionality so that the answer can be seen graphically in two dimensions. In other words, statistics should simplify, not complicate, biological interpretation.

The rate of presentation is very important. As Tukey (1960) pointed out, we "must educate the client at the proper rate, not too rapidly and not too slowly." If you find that you have exhausted the subject of the 3 hour session in 15 minutes, you have two choices: you can give the students 2 hours and 45 minutes of practical work during which they can try to work out what you were talking about, or you can abandon the course and tell them to read a textbook.

The increase in the rate of learning is exponential and students are capable of incorporating new information at a phenomenal rate towards the end of the course. However, any attempt to increase the pace at the beginning results in the students memorizing, but not internalizing the concepts. By internalizing, we mean having a gut feeling for the concept that does not require, and is not obscured by words. It is like riding a bike or having sex. After a certain point, reading about it, or knowing all the words only makes it more difficult to do.

Do not be tempted to include more information or assumptions about the analyses. This is a basic course. Before students use the techniques, they need to (a) read the literature, (b) read some statistics texts, (c) read the computer program manual, (d) experiment with some of their own data, and (e) (most importantly) create lots of graphs. This is how all researchers learn a new technique. You can distribute references so that students can discover the details when, and only

Table 14.1

Concepts that the student should understand at the end of the course, and before starting to collect their own data.

1. It is not possible to prove something, only disprove it.
2. You cannot test a hypothesis with the data that were used to create it.
3. The observations used in statistical tests must be independent in relation to the question, and collected on the scale of the question.
4. Sources of pseudoreplication (observations that are not independent in relation to the question) may be spatial, temporal, phylogenetic, or technical.
5. Most ecological categories are arbitrary; science advances more quickly when researchers study continuous (or at least ordinal) variables.
6. If you measure a large number of independent or dependent variables, and apply independent statistical tests, you will encounter many low "probabilities" (pseudoprobabilities). This confirms statistical theory but does not contribute to ecology.
7. The logical assumption of an analysis with only one independent variable (that no other variables greatly affect the outcome), is more restrictive than the statistical assumptions of analyses with several independent variables.
8. Statistical analyses with only one independent variable are normally trivial and can be substituted by a dispersion graph (scatterplot).
9. Determination of the number of samples necessary and how they should be distributed is a biological question, and the best method to evaluate the number of independent observations needed is to use hypothetical dispersion graphs.
10. If no information is available about variability in the variables to be measured, the following guidelines usually work for ecological data. To estimate the number of *independent* observations necessary when all the variables are continuous (multiple regression or a GLM analogue), multiply the number of independent variables by 10. When all the variables are categorical (ANOVA, as presented in textbooks), count the number of levels in the factors and multiply the resultant by 4. If you have a mixture of categorical and continuous variables (ANCOVA), sum the number of levels in the categorical variables and multiply the resultant by 10.
11. For regression or for ANOVA with random variables, use at least ten different levels for each random variable.
12. Ecological variables rarely have only direct effects. Often indirect effects are more important than direct effects, in which case, analyses of direct effects alone (ANOVA, ANCOVA, regression) are misleading. For this reason, the definition of questions should be done in relation to a flow chart, even if no formal analysis of the flow chart will be used in the study.

when, they need them. Discourage them from memorizing the details or they will end up not being able to see the forest for the trees. The few concepts that they should get by the end of the course are listed in Table 14.1 (adapted from Table 1 of Magnusson 1997).

It is best not to regurgitate the examples used in this book; instead, substitute them with your own. Students learn statistics more quickly if you make frequent excursions into your experiences and biology in general. It helps keep them focused on why they are using the analyses. If you can not think of examples in your own field, you should not be giving a course in experimental design.

The fundamental concepts of the relevance of the questions and the independence of observations are biological and cannot be presented by a mathematician. It takes a biologist to see the potentials and limitations of biological statistics. If you are a statistician and have been roped into giving a basic course to ecology students, refuse until you have had enough experience working with researchers in that field to understand the limitations of the collectable data. This is not meant to belittle the potential contributions of statisticians and mathematicians; it is the aim of the course to bring the students to a point where they can understand the need for, and make use of, statistical advice.

It is important that students do not confuse computing (or statistics) with sampling design. However, most of the exercises require the use of a computer program. The course must be relevant to the activities of the students, which generally revolve around computers. Statistical packages, unfortunately mostly pirated, are installed in most of their computers.

We have taught the course around the SYSTAT and SPSS statistical packages, but other comprehensive packages, such as SAS, are equally effective. We have found that students who take a 2 to 3 day course on the use of the statistical programs prior to enrolling in our course are better able to avoid computer dazzle, and it also helps prevent them from confusing statistics with computation. If your students are not familiar with the package you intend to use, and you cannot give a preliminary course, you should prepare detailed notes that give all of the keystrokes for each exercise.

The authors are available to help train teachers and install the course at new locations, as long as their transport and lodging costs are covered. We have taught the course in Brazil, Bolivia, Australia, and the U.S. The course is intensive and takes 2 weeks (or 10 working days). Offering it over a longer period is more difficult, and probably involves more teaching hours, though Marc Hero has taught it successfully as an extensive course at the University of California, Berkeley.

Structure of Lectures

Lecture 1 introduces the concept of science as a culture, the importance of defining the question, and shows how experimental design defines which questions can be asked. This lesson must be interactive; if you do not pull students out of passive note-taking now, you will lose them. Make them unsure, make them nervous, make them angry—make them anything as long as they feel something! Hall (1959) gives examples of how culture affects our actions, which students can readily relate to.

To bring home the fact that students may have absorbed some cultural jargon but do not understand the concepts, we present a graph like Figure 2.6 and ask the students to mark one standard deviation either side of the mean. Before the exercise, most will say they have used the concept of standard deviation, or at least understand what it is. However, few come close to the correct position, and almost none of those who do will know why they placed the marks there. Use plenty of irony here so that they begin to be uncomfortable with the culturally accepted jargon that they do not fully understand.

The example in Chapter 1 can then be given to the students for them to do by hand. If it helps you, feel free to divide the class into two groups, and give all members of the same group the same direction of transect. Choosing groups that naturally (or unnaturally if you are a liberal) compete, such as men and women, post-graduate and undergraduate students, persons with and persons without glasses, etc. helps to give students a reason for believing that their direction of transect is the correct one. Much time can be spent showing them that the bar graphs that they can produce hide most of the information (replicates), and the "error" bars that some use are not interpretable by most of their colleagues. Another common error is to put a "plot number" on the x-axis, even though a plot is not a variable and simply represents the sampling unit. Finally, you need to show them that a detail, such as the transect direction, can completely change the question that can be answered.

Lecture 2 deals with Popperian philosophy and pseudoreplication and is based on the first pages of Chapter 5 which discuss Popperian philosophy, and Chapter 4, in that order. The aim of the lesson is to move from the personal to the scientific. Examples should be from everyday experience whenever possible. You cannot give too many examples of spatial, temporal, phylogenetic, and technical pseudoreplication. If time permits, ask students to show examples of potential pseudoreplication in their own work. You must keep this discussion simple. Remember that they still have not had the rest of the course, and that they will not benefit from a demonstration of your scientific knowledge. The students should be asked to present conceptual maps similar to those shown in Figure 4.3. It may help to finish early and ask the students to read Hurlbert (1984) and Platt (1964).

Lecture 3 is summarized in Chapter 3, and introduces the concepts of populations, parameters, and descriptive statistics. Students must become familiar with the concept of deviation in order to describe variability. The teacher should slowly present the concepts on graphs similar to Figure 3.2, and show how each distance can be represented by a simple mathematical formula. If students are not saying, "So that is what the standard deviation (variance, standard error, etc.) is!" you have done something wrong. Always ask the students to define the terms before you give your definition. This will reveal to you how much they do not know. Use these questions to confront communication of culture with communication of objective information.

Most of Lecture 4 is devoted to a "field" exercise. Ask the students to describe the lengths of leaves of a plant species that grows on campus, and have each student investigate a different species. However, instead of one sample, they should collect 30 samples of five leaves. If you want to show the students how knowledgeable you are, you can explain how to collect an unbiased sample of leaves before they do the exercise. They will then learn nothing about experimental design!

When they return, they can use the computer to calculate the 30 means, and the standard deviation (or the standard error) of the means. This standard error, based on 30 samples, can be regarded as very close to the true standard error that would have been obtained by calculating the standard deviation of all means based on a sample size of five. The students then can use the computer to calculate 30 estimates of the standard error by applying the statisticians' magic formula to each of the 30 samples.

In almost all cases, the "true" standard deviation will be much higher than the median of the 30 estimates of the standard error based on the magic formula, and in many cases will be greater than any of the 30. This is because the students did not take random samples of leaves. Each sample was taken from a single bush, or the leaves were collected sequentially across a grass sward, or some other form of pseudoreplication. Make them squirm by relating inadequate experimental design to lying.

You can show that the magic formula is approximately unbiased by having them randomly permute the leaf lengths among samples and re-compute the statistics. This exercise shows that making inferences on samples with $n = 5$ is risky because unbiased estimates are not necessarily precise and, more importantly, it demonstrates that no statistical procedure produces valid results if the sampling was done incorrectly.

Lecture 5 introduces the concepts of a statistic, a null hypothesis, the distribution of results expected when the null hypothesis is true, and regions of rejection. It is summarized in the middle of Chapter 5. The example presented in Chapter 5 can be used as the class exercise. Using the student's measurements—height, weight, etc.,—is better than using data taken from other organisms because the students have expectations about themselves, and they will be looking for the datum, or symbol on the graph that represents them. They can use coins to generate 20 values of DIF expected when the null hypothesis is correct. This allows them to construct histograms and think about regions of rejection.

Lecture 6 uses Manly's RT program (see Manly 1997) to generate 1000 values of DIF when the null hypothesis is true; these values are then used to see how the curve becomes smoother with more simulations and that the precision of the probability values increases. We programmed this in SYSTAT, and you can use any of the better statistical packages, but Manly's program is so cheap and easy to use that it is worthwhile to just use it. Leave plenty of time to review critical values, regions of rejection, types of errors, etc. These are essential for understanding future lessons. Lectures 5 and 6 are usually rushed over because the teacher thinks that the concepts are obvious. They are not, and if you find that you are finishing early, give more examples and exercises.

Lecture 7 presents statistics, null distributions, critical values, and probability again. It also deals with one- and two-tailed tests, parametric approximations to randomization tests, degrees of freedom, and types of errors. This is based on the last part of Chapter 5, "How the textbooks tell the story."

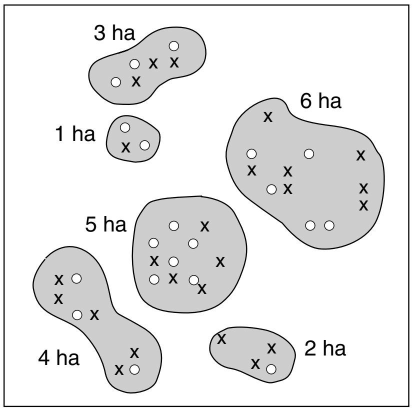
The students use the computer to do *t*-tests on the example investigated in Lecture 4. They also use *t*-tests to compare the samples in their leaf data, comparing sample 1 to 2, 2 to 3, 3 to 4 etc., conducting 29 tests. They will obtain only a few (< 3) apparently significant values at $P < 0.05$ if they collected valid samples. However, the *t*-test will detect their pseudoreplication and give many apparently significant results, reinforcing the importance of sampling correctly.

Repeating the exercise on the randomized data should give only about 5% "significance." Keep reinforcing the concepts of types of errors and pseudoreplication. This exercise also shows that repeated testing is almost guaranteed to give "significant" results for some comparisons. The Bonferroni procedure can be introduced as the simplest means of correcting this distortion. A better procedure for many comparisons will be introduced in Lecture 8.

Lecture 8 deals with one-way ANOVA, and is largely based on Chapter 6. The important concept that needs to be understood is partitioning of variance. Future lessons will return to this concept, but students should get a good introduction to it in this lecture, and time should be spent interpreting the analysis of variance tables produced by computer programs. The concept of compound variances and mean squares should be introduced as well.

If students do not understand why the expected value for *F*-ratios in ANOVA is 1, they will not understand any of the more complicated analyses in future lessons. Students can analyze other examples as well, but they should re-analyze their data on leaves to see that ANOVA considers "sample" to be significant, and therefore recognizes their pseudoreplication. Do not spend much time on multiple comparisons. If you do introduce them, use Tukey (1991) as a philosophical basis for discussion. If students understand the ANOVA principle they will be able to apply it to more interesting problems than those that can be analyzed by one-way

Figure 14.1



ANOVA. Data appropriate for one-way ANOVA can usually be evaluated more meaningfully with a dispersion graph.

Lecture 9 focuses on regression analysis and generally follows Chapter 7. It is important that students take data off a map (for instance, see Figure 14.1) and that it is not given to them in tabular form. The generality of the ANOVA concept is perhaps the most important lesson here, though the ability to recognize parameters that define models is important for the next lesson. Students must understand that the most commonly used models relate only to linear relationships. Although it is important to present computer output, make sure that you interpret the various statistics purely in terms of the concepts from previous lessons.

How much you explain depends on the level of the class. However, a lot can be done with a little explaining. Anscombe (1973) gives a good series of graphs that could be used as class exercises; from these exercises, the students will “discover” that totally different graphs give the same statistical summary.

Lesson 10 is given in Chapter 11 in the book. We changed the order so as not to lose the thread of logic relating to variance partitioning. In the classroom you have more control and can keep reminding students of the basic principles. Although the nonlinear techniques are generally not useful for allocating variance, they are very useful for illustrating the different types of residuals that can be minimized to parameterize a model. In the classroom, it is best to finish all of the analyses with only one independent variable before moving on to the multi-factorial designs.

Lecture 11 uses multiple regression to introduce the concept of variance allocation with more than one dependent variable. One of the most important aspects of this lecture, which follows the first part of Chapter 8, is to show that univariate analyses can be very misleading. It takes a whole session for the students to assimilate the concept of using residuals to uniquely partition variance, so this part should not be rushed. If you can think of an example for which the students can collect their own data, so much the better. However, we contrived an example using data on monkeys, trees, and shrubs that is easy to show on the blackboard. We have included the data here (Table 14.2) in case you want to use it. Give it to the students on a map, such as Figure 14.1.

Lecture 12 completes the discussion of the variance allocation techniques, mainly through an explanation of the tables produced by the computer package.

Important concepts include interactions, the correct *F*-ratios for mixed models, and variable selection. These concepts are explained in the second part of Chapter 8 and in Chapter 9. The exercises consist of the students generating their own random data and seeing what proportions of the factors are "significant" with standard and stepwise procedures. It is important that you require the students to give names to the variables that are relevant to their research. It is common for students to defend the significance of variables such as "predator density" even though they generated the data themselves.

Table 14.2

Shrubs	Trees	Monkeys	Reserves
1	2	1	A
2	1	3	B
3	4	2	C
4	3	5	D
5	6	5	E
6	5	7	F

Lecture 13 provides a brief introduction to multivariate techniques. This lecture is even more superficial than the preceding ones, and you will not even be able to cover all of the material in Chapter 12. However, students should learn to recognize patterns in the data, and to understand that those patterns can be represented as gradients. Although students only get a vague idea of how the techniques work, they can see that the concepts developed in previous classes can still be applied. The most important thing is that the students can see the patterns in the original data in graphs and/or tables. The introductory chapters of Gauch (1982) are a good basis for most of this lecture. With some idea of underlying gradients (phantom variables), students should be able to make some sense of the literature, even if just enough to provide a healthy dose of skepticism.

Lecture 14 brings us to path analyses that show the value of considering direct and indirect effects. All of the principles of previous lessons are used in this presentation, which follows the content of Chapter 10. The purpose of this lecture is not to promote path analysis. Rather, it should show why no non-trivial statistical test can be interpreted without a reference to a flow chart.

The remaining 3 days (Lectures 15–20) are devoted to class participation. Students present and discuss their sampling designs. They should learn that although they are able to criticize the designs of others, they are still quite incapable of evaluating their own work. They will continually try to hide behind jargon, meaningless bar graphs, and other cultural tools. Although the teacher should provide advice on how to communicate better, students should not be evaluated on their ability to present their work. Students should be evaluated on their participation, marks being given for the quantity and quality of constructive criticism. The teacher should emphasize the need for independent review, and the value of constructive criticism in science. The method presented by Brown (1997) combined with community participation, is excellent, if feasible for your class.

The students should be told that the purpose of the course is to teach them to communicate their ideas better, so that they can maximize the constructive criticism they receive. And that is a good summary sentence on which to finish.

Literature Cited

Brown, I. F. 1997. Let the people judge: Community examining boards help students learn to communicate. Bulletin of the Ecological Society of America 78:210–211.

Gauch, H. G. 1982. *Multivariate Analysis in Community Ecology*. Cambridge University Press, Cambridge.

Goleman, D. 1995. *Emotional Intelligence*. Bloomsbury Publishing, London.

Hall, E. T. 1959. *The Silent Language*. Doubleday and Company, New York.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187–211.

Magnusson, W. E. 1997. Teaching experimental design, or how to do statistics without a bikini. Bulletin of the Ecological Society of America 78:205–209.

Magnusson, W. E. 2000b. Statistical iatrogenesis: Cure it or prevent it? Bulletin of the Ecological Society of America 81:198–201.

Manly, B. F. J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman and Hall, London.

Platt, J. R. 1964. Strong inference. Science 146:347–353.

Tukey, J. W. 1960. Conclusions vs. decisions. Technometrics 2:423–433.

Tukey, J. W. 1980. We need both exploratory and confirmatory. American Statistician 34:23–25.

Tukey, J. W. 1991. The philosophy of multiple comparisons. Statistical Science 6:100–116.

Wilkinson, L. 1999. Statistical methods in psychology journals: Guidelines and explanations. American Psychologist 54:594–604.