Questions for Formative and Summative Assessment that Encourage Deep Rather than Surface Approaches to Learning Basic Statistics in a Computer Environment

Ruth Hubbard
Queensland University of Technology

Introduction

Assessment questions must reflect the kind of learning that has been promoted by the instructor. Nevertheless, many instructors who try to instil a deep approach to learning and encourage their students to develop statistical thinking skills, set examination questions, which encourage rote learning of procedures. Why? Because those are the questions that they find in most textbooks and because they don’t know how to create other types of questions, which will effectively test an understanding of statistics.

Students need to be aware of the kinds of questions they will encounter in the examination and if mainly procedures are required, they will do what is most efficient, and concentrate on reproducing the procedures, neglecting all the other aspects of the course.

Many instructors viewed the arrival of statistical packages as a threat to the proper teaching of statistics. I viewed them from the very beginning as a wonderful release from the drudgery of computation and an opportunity for students to explore statistical concepts in more interesting ways.

Other speakers have discussed the merits of using projects for assessment and I believe they are the ideal way for students to demonstrate most aspects of their statistical learning. But, in most institutions the hierarchy still insists on an examination at the end of a course, so we have to comply. My talk is about how we can do this without sacrificing our ideals about statistical learning.

Purpose

When designing new types of questions, the most important questions that the designer must pose and answer are, “What is my purpose? What aspects of the course will the question test?” In fact ideally, the questions should be based on the course objectives, which often incorporate very worthy goals such as:

- Students should develop an understanding of the inherent variability in statistical data and learn to measure and control it.
- Students should be able to produce graphical displays of data and to interpret the information in the graphs.

But in my long experience, the examination questions usually address much more limited goals and have a standard format such as: “Here is the context, here is the data, test the hypothesis.”
I will therefore begin by stating some specific purposes that we might have for producing questions and then give examples of how these could be translated into suitable questions. The purposes are not organized by topic or by importance or difficulty.

- check understanding of formulas and definitions
- make decisions about suitable models
- illustrate understanding by drawing diagrams
- make connections between concepts instead of keeping them in separate compartments
- make connections between graphical and numerical information
- anticipate the result of altering some of the given information without doing any further calculations
- apply a result in a new context
- demonstrate comprehension of a paper or media article

Some words of warning

- Non-standard questions need to be introduced gradually into a course so that students become accustomed to changing the way they think about statistics and mathematics.
- When introducing new questions, there is a temptation to make the questions very wordy, thus putting strains on the reading skills of many students.
- Grading students’ responses to questions in which they reason and explain, is a much more time-consuming business than grading the responses to “Here is the context, here is the data, test the hypothesis” questions.

Examples

1. Check understanding formulas and definitions

Asking students to state definitions is an invitation to rote learn. A simple way to test definitions is to ask students to give examples of what is being defined. For example:

Give an example of:
(a) a categorical ordinal variable;
(b) a discrete random variable.

To test whether students understand a formula we might ask:

Explain in your own words the meaning of the statement:
If \( X \sim N(\mu_x, \sigma_x^2) \) and \( Y \sim N(\mu_y, \sigma_y^2) \) and if the variables \( X \) and \( Y \) are independent, then \( X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2) \).

For students, this is not a trivial exercise, but it could be made a little more difficult by writing the above formula with an error in it, such as using standard deviations instead of variances and asking students to find the errors.
2. Make decisions about a suitable model

Students who rote learn, identify models using statements like, “If there is an n and a p, then it is binomial”. One way of testing whether students can choose the correct model is to describe a number of situations which fit different models, and ask students to identify them. This requires a lot of reading and does not distinguish between students who really understand the assumptions on which the models are based and students who have merely learned the n – p thing. An alternative is the following:

Make up a problem, not one used in class, that could be answered by calculating \( \binom{10}{3} (0.8)^3 (0.2)^7 \)

Here is a less specific question to test understanding of a model:

Describe a problem that could occur in one of the other subjects you are studying, which would require the use of simple regression. Pretend you have collected a small amount of data, write the data down and label it appropriately but do not do any calculations.

3. Illustrate understanding by drawing diagrams

This has already been touched on in the question on sketching histograms and relating them to means and medians. There is plenty of scope for students to draw diagrams to illustrate the calculations involved in regression. For example:

Draw a simple diagram (you do not need scales on the axes, but you need to label important parts of your diagram), that explains how the “total sum of squares” in the ANOVA part of the regression output is calculated.

The same kind of question can be used for one-way ANOVA as a replacement for all those formulae with many subscripts which non-mathematical students find difficult to decipher. Many texts and computer programs use such diagrams to explain statistical procedures but it is not so common to ask students to make the diagrams.

4. Make connections between concepts instead of keeping them in separate compartments

The questions students work on, in class and in assignments, are always related to a particular topic, the one currently being studied. Likewise the questions in textbooks are related to the chapter they follow, and examination questions are frequently also arranged in the order in which the various topics have been studied. This is a perfect recipe for producing compartmentalised knowledge. To help students integrate their knowledge, we need practise and assessment questions that straddle this artificial divide. For example:

What is the relationship between the sign of the correlation coefficient and the slope of the fitted regression line, if both statistics are calculated from the same data set?

An example that cuts across many chapters is:

Which statistical distribution is used to test hypotheses about:
(a) a population variance?
(b) a ratio of two population variances?
(c) independence in a contingency table?
(d) the equality of three or more population means?
(c) the difference between two proportions?

5. Make connections between graphical and numerical data

Students are usually told that they should produce graphs of data before attempting any analysis, so they should have some experience in describing the graphs they produce. To test how well they have understood the relationship between graphs and statistics we could ask:

Sketch histograms for frequency distributions in which the mean is:
(a) greater than the median
(b) about equal to the median
(c) less than the median

We could also ask for examples of populations, which they expect would have graphs of these shapes. Alternatively we could provide the graphs and ask the students to identify which graphs satisfied which of the above conditions.

6. Anticipate the result of altering some of the given information

This technique is very useful for testing whether students have a detailed understanding of what is involved in inference without too much reading or calculation. A simple example first.

An environmental group collects a litre of water from each of 45 locations along a stream and measures the amount of dissolved oxygen in each specimen. (The amount of dissolved oxygen in water decreases when the water is polluted.) The mean oxygen content calculated from their data was 4.62mg per litre with a standard deviation of 0.92 mg per litre.

(a) Calculate a 95% confidence interval for the true oxygen content of this stream assuming $\sigma = 0.85$.

(b) What would happen to the confidence interval in (a) if the sample size was changed to 60 but all other information remained the same. (Do not calculate the new interval.)

(c) What would happen to the confidence interval in (a) if we increased the level of confidence to 99%?

(d) What would happen to the confidence interval in (a) if we made no assumption about $\sigma$.

A more difficult example along the same lines, which also shows that questions about interpreting printouts do not have to be routine or simple-minded is the following:

In order to test a hypothesis about a population mean, the following Minitab printout was obtained.

Z-Test

Test of mu = 5.000 vs mu > 5.000
The assumed sigma = 0.900
If you had used a T-TEST instead of a Z-TEST on the same data, which numbers in the last row of the printout would change? Do you have enough information to decide whether each of the numbers that would change, would increase or decrease? Justify your answers.

7. Apply a result in a new context

An alternative to asking students to explain the outcome of a calculation or a printout in their own words, is to bring in a new idea and ask students how this relates to their previously stated conclusion.

A designer of the interior of a new passenger aircraft has to decide how far apart to put the seats. To help make the decision she chooses random samples of 50 adult females and 50 adult males and makes some relevant measurements. One measurement is the length in mm of the length of the thigh. The thigh measurements are summarised in the table below.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td>40</td>
<td>5.285</td>
<td>1.062</td>
<td>0.142</td>
<td>2.00</td>
<td>0.023</td>
</tr>
</tbody>
</table>

(a) Use the information in the table to calculate a 95% confidence interval for mean thigh length. If you are a female, just calculate the interval for females and if you are a male, calculate the interval for males.

(b) Explain in your own words what these intervals tell you about thigh lengths.

(c) Measure your own thigh length (from your crutch to your knee) and write the measurement down. Does it fall inside the relevant confidence interval? Should it? Why or why not?

(d) Would these confidence intervals actually be useful for making decisions about the distance between airline seats? If not, what other statistics would be useful?

This question, apart from causing some amusement, deals with a very common misconception about the application of confidence intervals. The last part also requires a critical approach to a problem.

8. Demonstrate comprehension of a paper or media article

The ideal is to use technical papers, in students’ major disciplines, which make use of statistics as the texts for testing comprehension. However, students frequently study statistics near the beginning of their university studies, before they have become familiar with the technical vocabulary of their major discipline. If this is the case, it is better to use something less sophisticated. An alternative, are articles especially written for statistics students, such as *Statistics: A Guide to the Unknown*, which covers a wide range of disciplines and statistical techniques.
Articles from the media are plentiful although they frequently omit some of the essential information that is required in order to draw conclusions about the findings. In this case students could be asked what additional information they would require in order to believe the stated conclusion. The part-article below does contain all the necessary information to show that the stated conclusion is wrong because percentages of different quantities have been subtracted.

**Doctors’ incomes rise, survey finds**

*Doctors’ incomes increased by 9.62 per cent over the past year to $152,915 but did not keep pace with inflation, according to the Australian College of General Practitioners. A survey of 511 doctors across Australia showed the median costs of operating a practice rose 7.47 per cent to $88,935 – leaving doctors two per cent better off than the previous year. The President of the College said this meant doctors had received pay increases about one third of the rest of the workforce. “To keep up with the cost of living they needed an increase of about six per cent.”*

A little arithmetic shows that the percentage increase in income minus expenditure is 12.75%.

**Conclusion**

The above categories and the examples illustrating them are by no means comprehensive. They are intended to show that it is possible to design questions which have a specific purpose and which test desired learning outcomes. It is worth noting, that many of the above examples were produced using one basic technique – reversal, ie. asking the question backwards, rather than in the direct, conventional manner. This includes questions in which students are asked to give an example, make up a problem, what would happen if ….

The main point is that questions are a powerful way to encourage deep approaches to learning and to persuade students to adopt this approach, we must include such questions in assessment.