PLANNING A STUDY:
How students answer free response questions on AP Statistics exams.

Fred Djang, Choate-Rosemary Hall and Bob Stephenson, Iowa State Univ.
Fred Djang, P.O. Box 78, Wallingford, CT 06492-0788

Key Words: Design of experiments, observational studies, sampling studies, surveys

Abstract

The AP statistics exam has been offered three times with the number of students sitting for the exam increasing from around 7,500 in 1997 to over 25,000 in 1999. Several of the free response questions in these exams dealt with planning an observational study or a designed experiment. In this paper we look at how students answered these questions and discuss what was good and not so good about the answers. With close to 35,000 students projected to take the exam in May 2000, we will also look at ways that AP statistics teachers can help their students prepare for questions dealing with planning a study. In addition to looking more carefully at the ideas of control, replication and randomization in experiments, we will also look at some basic ideas that are important for every study. These ideas involve the choice of response, the choice of conditions to be investigated, and the choice of material to be used. We will also discuss the role that variability, both the type of variability and the source of variability, plays in the planning of a study. Better understanding of the role of variability can be helpful in deciding how to appropriately randomize a study, complete randomization or restricting randomization within blocks.

1 AP Statistics: The Course

The idea of an Advanced Placement (AP) Statistics exam is over twenty years old. In the early 1980's discussions among the AP Calculus Committee on ways to expand the AP offerings in the mathematical sciences lead to a recommendation to consider a statistics exam. It was not until 1994, when the College Board approved the formation of an AP Statistics Development Committee, did this idea begin to take shape. Over 7,500 students sat for the first offering of the AP Statistics exam in 1997. The number of students sitting for the exam grew to over 15,000 in 1998 and over 25,000 in 1999. Approximately 33,000 took the exam in May 2000. Although the rate of growth is slowing, we can anticipate from 40,000 to 50,000 students taking the AP Statistics exam in the not too distant future. For more information about the AP Statistics exam see Scheaffer (1998), Scheaffer (1999) and Peck (2000)

The interest in the AP Statistics exam and the corresponding AP Statistics course can be attributed to many causes. Among these are the inclusion of topics in statistics throughout the national standards for K-12 mathematics and the accessibility of much of the material in the AP Statistics course.

The AP Statistics course is built upon four major themes.

- Exploring Data
- Planning a Study
- Anticipating Patterns
- Statistical Inference

Although these major themes appear to be four separate areas, each is linked to the other. The exploration of data is tied to statistical inference. A side-by-side box plot can give you some intuition as to whether there may be a sizeable difference in central values given the natural variability in each group. Additionally, checking assumptions necessary for statistical inference often relies on exploratory methods. Statistical inference is also tied to data collection. It is difficult, if not impossible, to do appropriate statistical inference without knowing how data were collected. Many of these links between themes are stated explicitly in the *Advance Placement Program Course Description for Statistics; May 2001, May 2002*. For example, on page 5 is states;

> "Data must be collected according to a well-developed plan if valid information on a conjecture is to be obtained."

2 Planning A Study: Topics

Topics that define the scope of each theme are given in this course description as well. For the theme; Planning a Study, the four major topics are:

- Data Collection Methods
- Sampling Studies (Surveys)
- Experiments
- Making Generalizations

It is worth looking at each of these topics individually in some detail.

2.1 Data Collection Methods

Data collection can be divided into two general activities, one passive and one active. The passive activity is observation; being in the right place at the right time with the right tools. The active data collection method is experimentation; manipulating a variable in a controlled setting in order to see the effect on a response of interest.

2.2 Sampling Studies (Surveys)

The result of passive observation is an observational study. Observational studies run the gamut from a case study (one item or event studied in great detail, e.g., the investigation of an airplane crash) to a census (all items of interest looked at but in not much detail, e.g., the U.S. Census). A longitudinal study, like the Framingham Heart Study, is an examination of a relatively small population (a census) over time. Somewhere in between the case study and the census lies the sampling study (survey).

The sampling study (survey) deals with the examination of a few items or individuals selected from a population of interest. There are many important ideas concerning sampling studies. The distinction between population and sample. How the sample is selected and whether bias is introduced by that selection. Simple random sampling (what it is and how to actually do it). Stratified random sampling.

2.3 Experiments

The active data collection method is experimentation. This is considered active because the researcher must take an active role in manipulating some variable in order to have an experiment. The three basic principles of a well designed experiment are:

- Control of outside variables to prevent confounding.
- Randomization to reduce the chance of bias.
- Replication within the experiment to assess natural variability.

What and how to control outside variables and the effects of control on the ability to generalize results is extremely important. So is randomization. Whether it is complete randomization or randomization within blocks, students should know what randomization is and how to randomize.

2.4 Making Generalizations

The ability to make generalizations depends on the way data are collected. Lurking variables in observational studies and uncontrolled variables in experiments and limit one's ability to attribute cause and effect. Limitations on the individuals or experimental units studied further limits the ability to make generalizations.

In the past three offerings of the AP Statistics exam, planning a study has been an integral part of at least one of the free response questions. In the next section we look at how students were supposed to respond and how they actually responded to these questions.

3 Student Answers

In the 1997 exam one free response question asked the students to design an experiment to compare the weight of salmon raised over a six-month period on new and old types of food. Salmon are in eight tanks in a room with a temperature gradient.

The rubric used to grade students responses focuses on three points. Students must recognize the need to form blocks of tanks because the temperature gradient makes the tanks non-homogeneous. They must form blocks of tanks that make the tanks within blocks the most similar and explain why these blocks are best for accounting for the temperature gradient. Finally students must give a correct randomization scheme for assigning foods (new and old) to the tanks within each block.

In terms of recognizing the need for forming blocks of tanks, many students noted that the temperature gradient was a nuisance factor that could not be controlled. Some students opted for 4 blocks of two tanks. By forming blocks of two tanks they could match tanks not only in terms of distance from the heater but also in terms of placement near windows, doors and walls. It is important to recognize that although the heater might be the main source of the temperature gradient, windows, doors and walls also contribute to this gradient. Students who took only the heater into account and formed two blocks of four tanks failed to recognize that the four tanks were still not homogeneous and so did not get full credit for their answer.

For the four blocks of two tanks, a simple coin flip can be used to randomly assign foods to tanks. Flipping a coin (heads=old food, tails=new food) for the first tank within each block will randomly assign a food to that first tank. The second tank in the block receives the other food. With two blocks of four tanks, flipping a coin will work but a better way would have two chips with "old" and two chips with "new" in a sack. For each tank, the sack is shaken and a chip is drawn (without replacement) the food indicated on the chip is assigned to that tank. This is done until all four tanks in the block have a food assigned.

For this problem, students had to demonstrate that they understood that a completely randomized design would not be appropriate because of differences (due to temperature) in the experimental units (tanks) an d to actually construct blocks to account for this problem.

The 1998 exam had a problem dealing with a study of butterfly migration. The study is designed to evaluate which location on the butterflies' wings is best for marking or if all six marking locations result in the same chances of successful migration. In the study 3,600 butterflies will be marked and released.

The first part of the question asked students to describe a method to assign the marking locations so that there are exactly 600 marked in each location. Students needed to recognize that the six treatments groups should be as nearly alike as possible, except for the marking location. One solution is the "bag-o-chips" method of randomization. This method requires a bag with 3,600 chips, 600 marked A, 600 marked B, etc. As a butterfly is captured, a chip is drawn (without replacement) and the butterfly is marked in the location indicated on the chip. Students had to have a valid randomization to receive full marks on this part. Some students suggested capturing the first 600 butterflies and marking them at location A, the next 600 marked at location B, etc. What these students fail to realize that the ease of capture might have an effect on the ability of the butterfly to have a successful migration. The first 600 butterflies might be the weakest fliers and so the most easily captured. This method of capture and marking introduces a confounding variable and is therefore inappropriate.

The second part of the question asked students to describe a method to assign the marking locations if you wanted the location to be independent from butterfly to butterfly and each location assigned with probability 1/6 each time. If you have the "bag-o-chips" described above, as a butterfly is captured, a chip is drawn (this time **with** replacement) and the butterfly is marked in the location indicated on the chip. Because you return the chip to the bag after each draw, each location has a probability 600/3600=1/6 or being selected at each draw. Alternatively, a six-sided die can be used to randomly assign the location. Assuming it is a fair die and rolls are independent, toss the die each time a butterfly is caught; one spot indicates marking location A, two spots indicates marking location B, etc.

This question is nice because the last two parts of the question ask the students to link the appropriate inference technique to the way the data were collected. In this case, only the marginal totals will be affected by the randomization scheme, they are fixed at 600 in the first part and allowed to vary randomly in the second. A χ^2 test is appropriate for both schemes, however, one is a test of homogeneity of proportions while the other is a test of independence. Students taking the exam were not expected to note this latter differentiation.

In the 1999 exam one of the free response questions looks at a study of the effects of an apple a day on dental health. Fifty dental clinic patients who routinely eat an apple a day are compared to fifty clinic patients who eat fewer than one apple a week. Examination of dental records gives the number of new cavities each patient has had over the last two years.

The question first asked why this is an observational study and not an experiment. Many students might think this is an experiment because of the two groups; apple a day and fewer than one apple a week. It is important to ask: How are the groups formed? In this case it is based on past behavior, and so we are just observing what has happened. It could be an experiment if 100 dental clinic patients are recruited and 50 are assigned at random to an apple a day group and the other 50 are assigned to a group that eats fewer than one apple per week. The patients would be followed for 2 years and the number of new cavities noted. By making the two groups, the researcher is manipulating who will eat apples and who will not.

The second part of the question asked students to explain confounding within the context of the study and to give an example of a possible confounding variable. When explaining confounding, it is important that the student do so within the context of the dental health study. An example of a confounding variable might be general diet. The people who eat an apple a day might have a healthier diet that substitutes fruit for sweets. The other group members may have a diet high in sweets and so don't choose to eat fruit, like apples. If the apple a day group has fewer cavities, on average, is it due to the lack of sweets in their diet compared to the high amount of sweets in the other group's diet or is it due to eating (or not eating) apples. When the effects of two variables (apples and sweets in the diet) can not be separated, those variables are confounded. Because of the possibility of confounding, it is inappropriate to attribute cause to eating an apple a day even if the mean number of cavities is significantly less for the apple a day group compared to the fewer than one apple per week group.

4 Teachers: Preparing Students

It is important to reinforce the fact that observational studies:

- have explanatory and response variables.
- can have a control or placebo group.
- can try to control for outside variables.
- can have blocking or pairing.

What differentiates between an observational study and an experiment is the existence of a variable manipulated by the researcher in an experiment.

Design of experiments is a very important topic. Many teachers of AP Statistics may be uncomfortable with this topic because they do not have the experience with experiments and their design. The book by George Cobb, *Introduction to Design and Analysis of Experiments* provides some nice insights into how to design an experiment. It is worth looking at some of the ideas presented in the first chapters of this book.

Every experiment should have a purpose. What does one intend to discover form the experiment? Defining the purpose of the experiment involves clarifying the question and deciding on how to collect data so that the eventual statistical analysis will be appropriate. Once there is a clearly articulated purpose, there are three decisions each experimenter must make.

- Response: What is going to be measured?
- Conditions: What is the manipulated variable?
- Material: What will be used for the experimental units (eu's)?

Once these decision have been made, the experimenter must face the fact that variability always is and always will be present. Coming to grips with the inevitable variability is an important step in designing an experiment. Cobb suggests three sources of variability.

- Variability due to conditions of interest.
- Variability in the measurement system.
- Variability in the material.

The experimenter wants the manipulation of the variable of interest to produce changes (variability) in the response. She wishes to see what conditions affect the response but she is also interested in conditions that do not affect the response. Variability in the measurement system is unwanted variability. An unstable, highly variable, measurement system can hide or enhance the effects of the conditions of interest. Similarly, variability in experimental material can be a nuisance. Variation in material can obscure the effects of the conditions of interest. However, artificial uniformity of material for the sake of the experiment can lead to trouble later on.

In addition to sources of variability there are also types of variability. Again using Cobb's taxonomy.

- Planned and systematic
- Chance or random
- Unplanned and systematic

Planned and systematic variability is introduced by choosing conditions of interest, manipulating variables. The experimenter plans to introduce systematic variability into the response by her manipulation of the factors of interest. This is the signal that is hidden in the data. Chance or random variability is the variability that is, and always will be, present. Statisticians learn to live with, even exploit, chance like or random variation. This is the background noise against which the strength of the signal is measured. Unplanned and systematic variability can cause big problems. If it enters an experiment it may be impossible to unravel the true effects. Unplanned and systematic variation is the essence of confounding variables.

Most AP Statistics teachers are familiar with the three fundamental principles of experimental design found in Yates, et al.

- Control of outside variables
- Replication within the experiment
- Randomization

By controlling outside variables one reduces the effects of confounding variables. This makes differences due to the conditions of interest easier to see. However, control of outside variables can also limit the ability to generalize. The more one controls, the more limited is one's ability to generalize. For example, a fuel economy experiment done on one car on a dynomometer at a test facility with climate control will not tell you much about the fuel economy of similar cars under normal driving conditions.

Replication is probably the most misunderstood of the three principles. Most students think that replication means repeating the entire experiment at some other time with another group of experimental units. Replication within an experiment is simply having several experimental units run under each condition. Replication within an experiment allows the experimenter to assess the natural variability in the experiment, the experimental error.

Randomization is probably the most important of the three principles. Randomization is the use of chance to assign units to groups. Randomization reduces the chance of bias in the experiment. Randomization does not remove the effects of uncontrolled or lurking variables but it does tend to spread those effects across the groups. Randomization provides the basis for statistical inference.

Randomization can be complete or restricted. The block design restricts randomization to within each block. Blocking is a means of accounting for known variability in the experiment that can not be controlled. If variation in material is a fact of life and can not be controlled, randomization must be adjusted to account for this. By blocking one does not control for or remove the natural variability among experimental units. However, one can account for this variability in the statistical analysis thus making the planned systematic changes induced by manipulating variables easier to see.

When designing an experiment, or teaching others about experimental design, our basic advice is:

> Control what you can,
> block on what you can't control,
> randomize the rest.

5 Summary

Planning both observational studies and experiments is important for better understanding of statistics and statistical analysis. Looking at past free response questions, rubrics, and sample student responses is a good source of information of what is expected on the exam. It is also important, as teachers, to become more familiar with the design and analysis of real experiments. Expect to see free response questions covering planning a study on future AP Statistics exams. Will you, and your students, be ready?

References

Cobb, G., (1998) *Introduction to Design and Analysis of Experiments*, Springer, New York.

Peck, R., (2000) "AP Statistics: A Report on the 1999 Exam," *STATS: The Magazine for Students of Statistics*, No. 27, p 12-13.

Scheaffer, R., (1998) "Making the Grade - AP Statistics, 1997," *STATS: The Magazine for Students of Statistics*, No. 22, p 8-12.

Scheaffer, R., (1999) "Making the Grade Again - AP Statistics, 1998," *STATS: The Magazine for Students of Statistics*, No. 25, p 3-5.

Yates, D., Moore, D. and McCabe, G. (1999) *The Practice of Statistics*, W. H. Freeman and Co., New York.