Probability and Statistics Teacher's Edition - Teaching Tips

CK-12 Foundation

January 8, 2010

CK-12 Foundation is a non-profit organization with a mission to reduce the cost of textbook materials for the K-12 market both in the U.S. and worldwide. Using an open-content, web-based collaborative model termed the “FlexBook,” CK-12 intends to pioneer the generation and distribution of high quality educational content that will serve both as core text as well as provide an adaptive environment for learning.

flexbook
next generation textbooks

Contents

Chapter 1

Probability and Statistics TE - Teaching Tips

1.1 An Introduction to Analyzing Statistical Data

Definitions of Statistical Terminology

This Probability and Statistics Teaching Tips FlexBook is one of seven Teacher's Edition FlexBooks that accompany the CK-12 Foundation's Probability and Statistics Student Edition.

To receive information regarding upcoming FlexBooks or to receive the available Assessment and Solution Key FlexBooks for this program please write to us at teacher-requests@ck12.org.

Probability and Statistics is by far the most text based subject in mathematics. Students new to statistics get a real baptism by total immersion with the early sections as we very quickly attempt to bring them up to speed on all of the terminology that will be used. Because of the significant difference between traditional math classes and stats, some care is going to be needed to teach how to read the textbook, as well as how to read problems.

If your school has a method of taking notes, you should first consider using that method, if you aren't already. If your school doesn't have a prescribed method, then it's worth checking in with the humanities and possibly the science teachers to find out if your students are being ask to use, or had been asked to use, a particular method then that would also be a good choice. If your students have not been asked to learn or use a note taking method, it will be worth your time to teach one, and enforce the use of it. Likely there will be some resistance; AP stats students are likely to be very skilled and resistant to note taking systems. However, even very skilled students can become bogged down and lost quickly with how much language and vocabulary is needed for understanding.

My favorite method is Cornell Notes, or a slightly modified version there of. Many of my past students have hated it at first, but began to believe in the value when they were able to reference and study from a concise source when the topics became more challenging. I don't always have my students attempt to recall entire lectures or pages of reading from their notes, but the recording of key vocabulary and topics and then the summary written at a later time are very important. A PDF from the university of Cornell can be found at: http://lsc.sas.cornell.edu/Sidebars/Study_Skills_Resources/cornellsystem.pdf, as well as many other sources. Other note taking systems can be found in various academic literacy guides and texts.

Another key is being very careful about the language used in class, and making sure that words that have special meaning are reserved. As mathematicians we frequently are sloppy with our language, letting the context define if we mean the special meaning of the word or a more general meaning. One big one is variable. With continuous, discreet, quantitative, qualitative, random and all kinds of variables being very important in statistics, students can get confused quickly, especially as the treatment of these variables is slightly different than it may have been in the past.

An Overview of Data

With the exception of a few theoretical situations, all of statistics is based around data. However, as students will see later, the importance of clear appropriately gathered data can't be overstated. The first step is seemingly always designing the method of gathering the data. Most of the topics presented here are going to be examined in depth in later sections, so there is no need to spend lots of time teaching about the specifics of each at this time.

This is a good time to look at some of the terms and relate them to students' experiences. Especially in contemporary times, studies and experiments are a large part of the news and popular culture. Have students find news articles citing studies and bring them in for discussion. It will especially be powerful if students choose items that have some pertinence to their lives. Studies about school violence, teen health, standardized testing and similar studies will have more meaning for students than ones on heart disease, home prices and other favorites of the media. Current events can be a huge part of the stats classroom, and it helps to make it a rich and memorable class. News articles are wonderful resources because they usually include incomplete or incorrect information when it comes to the math involved. Students can be led in a discussion of each study, what they have included, what they have omitted, and what the omitted details likely are. The examination of sample studies in the text provide the template for such examinations.

Measures of Center

The measures of center are frequently the only statistics that a vast majority of the population will ever use. However, the treatment of these in previous classes is very incomplete, often beginning and ending with mean, median and mode. This is a great place to begin with students, as the task of a first year statistics teacher is frequently to show how common perceptions are not always meaningful in statistics. The association with the "average" being the mean is easily discredited, as is discussed in the text. A fun exercise is to look at alumni lists for either a high school or a university. If there is a super-star athlete, or a top executive on that list, the mean and median income of graduates will be quite different, and outlines clearly why it is sometimes advantageous to use poor choices in statistics to promote a particular idea, in this case saying that the "average" alumnus of a school can expect to earn well above what is realistic.

There are many more measures of center presented and each will have its appropriate place. The text give a brief nod to "It depends" as the mantra of statisticians, but at some point students are likely to struggle with the apparently arbitrary nature of statistics. The purpose of statistics is rarely to nail down truths (although with careful practice, stats can yield surprisingly close results), but to inform and give a clear picture for trends when the data set is frequently too complex to use directly. The different choices about the mean, weighted means and trimmed means shows that statisticians have choices, and there is frequently no clear direction on which to use and that is OK. In the early stages students need to stay calm and just roll with it, as much of the confusion is cleared up with practice and experience.

One key topic here is the difference between the Population Mean and Sample Mean. There isn't much more to say about it at this point, but when we get into continuous distributions the difference between the two becomes really important. Make sure students are aware of the difference and are careful about using the correct label and terms for each.

Measures of Spread

The second of the two major base topics for stats is Spread (the other being Center as discussed in the previous section).These two topics will return over and over again as we begin to look at different distributions. As such, these sections should not be glossed over quickly. Trying to go back and understand where variance comes from is far more difficult when you are also trying to learn about the motivation behind continuous distributions, like the normal distribution.

At the top of page 32 there is a little mention of an important idea:

Even though we are doing easy calculations, statistics is never about meaningless arithmetic and you should always be thinking about what a particular statistical measure means in the real context of the data.

Highlight it! Make a poster of it! Frequently AP stats is derided as not being as tough, or as mathematically intensive as it's brethren AP Calculus. I am not sure where exactly this prejudice comes from, but I suspect it has a lot to do with the fact that very, very few preliminary skills from algebra and geometry are needed for success in statistics, while Calculus will expose every hole in one's high school math experience. This does not, however, directly relate to an easier time in a mathematical sense. Statistics requires a level of attentiveness that other classes do not. There are plenty of "cookie-cutter" problems in calculus that once the correct method is chosen the following steps follow a reliable pattern and it is only a matter of following the algorithm that has been used a hundred times before. Statistics requires a new level of understanding for the nature of the question, the process and then the results. The AP exam rewards such understanding. The free response sections are graded specifically to reward conceptual understanding and deemphasize rote algorithmic procedures. IF a student were to make a small arithmetic mistake, get an incorrect value for the standard deviation, and then interpret the value correctly, the penalty is minimal. However, if a student were to make a small arithmetic mistake and get a probability of 1.17 and the student left that answer as correct the penalty is substantial. The arithmetic mistake is not so bad, but not understanding that a probability greater than 1 necessarily shows an error belies a fundamental lack of understanding of statistics.

You may notice that both this guide, and the text, are constantly mentioning "but later". This is bad form, in my opinion, but speaks to the ease of overlooking details now with the stiff penalty of how complex topics are later. Sometimes having a former student speak at the start of a school year is helpful, not only for alerting students to challenges ahead, but also for time management, formatting of the test and other topics useful to be successful in an advanced placement class. Students will likely get tired of hearing it all from the teacher, but by the time they realize they should have been paying closer attention, it will be too late. I also tend to be ruthless with grading at this point, creating the expectation of extreme attention to detail. Later on I will relax a bit.

1.2 Visualizations of Data

Histograms and Frequency Distributions

Statistics classes are dangerous places for students who try to outsmart math problems. You probably know the type, the kid who will always bring up some fact of why a particular study is nonsense, or provides data, or worse anecdotes, that put the class discussion off track. For example, the bottled water discussion will likely have students bringing up all kinds of strange numbers, reasoning and justification or bottled water use, or lack there of. This is also a regional aspect to this, as my former students living in unincorporated locations of the Santa Cruz mountains with untreated well water, as well as my former students from New Orleans where the Mississippi River is the source, have different considerations than my former students in San Francisco (which pipes in amazing water from Yosemite National Park). Decisions have to be made as an instructor as to how much "digression" is going to be allowed. Sometimes there are teachable moments regarding using the data presented, not bringing in personal bias, or sometimes students have valid questions about how data is collected. Students, especially those who have had success in school, really want to participate. Often in the early chapters they don't have much to contribute to problems as it is a new subject for them. Simply dismissing students' contributions can leave some students with a sour attitude, but certainly not everyone's slightly off topic contributions can be entertained. I tend to try to listen, and then if the comment is not productive to my lesson, I try to let the class know what I would like to hear, and what isn't going to help us with the ultimate goal. This becomes less of a problem as the year progresses, students learn more stats, and the topics are tougher.

Most of these topics are going to be review from previous classes. I wouldn't spend a ton of time in class, but would use these as warm-ups. Give students a topic, like hours of TV watched daily, and have them collect the data and chart it in the first $5-10$ min of class. Making clear and accurate graphical representations is a skill to be practiced without a ton of content. It is also OK to be critical of "artistic" skills here, as presenting data for use by others requires clear charts. Making the histograms and other charts look good is part of the job.

Common Graphs and Data Plots

This is again a chapter to review, practice and make sure that everyone is on the same page. I would make the decision ranking the most important representations of data. This is subject to some debate. I deemphasize pie charts and stem and leaf plots. It is true that the general public loves pie charts, and while I find them easy to read, they are of limited use for extended work. Stem and leaf plots are really difficult to read, and in many cases tough to teach. The idea of using different place values to categorize and split the numbers up seems to confuse students, and is tough to read as an end user of the chart. Scatterplots for bivariate data and dot plots are more useful, especially for later units. I make sure my students are comfortable and proficient with making these charts.

Box and Whisker Plot

Box and whisker plots are great, especially for comparing two or more sets of data in a graphical manner. Students will have the most success with creating these graphs by following a somewhat algorithmic process. The steps outlined in the text is exactly how I work these problems, and would teach and stress the same. The biggest problem that I observe students having is attempting to draw the graph and then add numbers, or not using an even scale for the graph. The best part about the box and whisker plot is the clear representations of center and spread. If an even scale is not used then no quality information can be pulled from the graph. I stress that the number line must be drawn and labeled before any part of the box and whisker graph is drawn. This is even more important when trying to compare two sets of data.

The idea of the middle 50% is an important one. It is worth spending some focused time on, as not only is it useful for finding outliers, but is also an important statistic on its own. One thing students may be unsure of is if the 5 number summary is re-calculated after outliers are "thrown out". The answer is no, as the summary is resistant to outliers. The changes would not be more descriptive; the graph is the only thing changed for clarity's sake. For this reason, and some others, I never say "throw out the outliers" as it implies that they aren't an important part of the data set. Outliers are still important, and have to be treated carefully rather than simply discarded. This is most evident, for example, in analysis of data with an important, but strangely spread data, such as air pollutants by California counties. Los Angeles county is going to be way far out there, but you can't accurately represent the climate situation in the state without considering Los Angeles. List it, but it is most helpful to show it as a point outside of the whisker, because it's important to show how far away that single county is, not because it's not important.

1.3 An Introduction to Probability

Events, Sample Spaces and Probability

Finding the sample space is the key for student success in computing probabilities. Frequently problems arise when possible outcomes are missed, or double counted, or otherwise. That is to say, there are lots of less than intuitive possibilities so tools to support students finding correct answers are helpful. One of those tools is listed in the text, where a table is made with the different outcomes. Lists are also good, especially for sample spaces that only involve one item, like a single die. There is also no shame in drawing pictures or working with manipulative. I frequently visualize a deck of cards in my mind when working with those probabilities, but as a bridge player I'm comfortable with cards, while a student might not be. Don't hesitate to give students cards to work with.

I encourage my students to never work in percentages. I realize that it's kind of picky, but I find it useful to drive home that probabilities must be between zero and one inclusive. The strong restriction is very useful as it frequently provides feedback for mistakes. If you are computing any probability and it ends up outside of that interval, then something is wrong. It's also a great tool to use to eliminate answers on a multiple choice test. I like to have a mantra to really drive it home. Between 0 and 1 and all possibilities add to 1.

Compound Events

Elementary set operations are a fundamental part of mathematics, but one that is taught in inconsistent places. I've always marveled at the fact that nearly all of my college textbooks, including my graduate level texts, start with a preliminary chapter on set theory. This has always led me to believe that set theory is either not taught, or is given an incomplete treatment in a lot of classes. There really isn't any reason to give a deep treatment to it here. Sometimes sticking too much to the strict notation is going to cause more problems than it's worth, since the ideas here are intuitive and students have worked with them in venn diagrams and other problems in the past.

An important thing to remember is that set operations are binary operators. That is, even if there are more symbols, only two sets can be operated on at a time . Due to the fact that the combination of unions and intersections is not associative (that is to say: $A \cup (B \cup C) = (A \cup B) \cap (A \cup C)$) I always include parenthesis if there is more than one operator, even if they are all the same operation, just to avoid confusion. This is, and matrices, are some of the first structures students encounter that do not follow all the classic rules of real numbers that students are used to. It can be an opportunity to push a talented class, but it is really an extra topic that has limited utility to the ultimate goal for a stats class.

The Complement of an Event

The classic example of value here is the birthday problem: In a given room, what is the probability that at least a pair of people have the same birthday? This is a great problem, as it shows how working smarter using some principles of probability makes a seemingly tough problem easy. It also has a result that is fairly counter intuitive; the probability of at least a match is much higher than one would presume. Both are key ideas to drive across to students studying probability for the first time. Asking for "at least one pair" means that if you were to directly calculate the probability it would take a very, very long time. There are just too many ways to get a match once the number of people in the room exceeds 4 or 5. However, asking the question "what is the probability that no two people share the same birthday" is logically equivalent to the first question and much easier to compute. Subtracting this quantity from 1 (finding the complement) then yields the answer.

The key here might be in re-writing the question in terms of the complement. Students probably don't see what the big deal is in calculating complements, and that is really simple. However, making the complement work for you requires seeing where to apply it, and then what exactly you are looking for. The key hints are when a question is asking for a probability where multiple situations are possible. In the birthday problem, asking for exactly a single pair should be calculated directly, but at least a pair dictates that the complement is easier. Students should practice identifying and re-writing questions to make it so that time isn't wasted attempting monumental calculations.

Conditional Probability

Students are going to get tripped up with the order of conditional probabilities. The more intuitive way of thinking of conditional situations is "if then" as opposed to "given". While I am nearly always in favor of understanding the concept and avoiding formulae, in this case the formula is great. Because the order is a little strange, and frequently is mixed up, this is one of the few formulae that I put on a poster and ask students to commit to memory. Application is easy once the spaces are put in their proper place.

Additive and Multiplicative Rules

This is probably student's first exposure to the dreaded problem of double counting. In this case, in contrast to the previous section, I don't emphasize the formula here. Finding values or probabilities that are double counted is a huge part of statistics, and one that most stats students have memories of sitting in a group, having problems with getting the correct answer, pulling hairs out, only to then have someone say "double counting!" For this reason, I really try hard to get my students to understand where double counting occurs and try to train them to always be aware of where the error is likely to occur.

Basic Counting Rules

Combinatorics are the foundation of lots of basic probability questions. Cards are a great way to do look at problems, especially with the recent attention paid to poker. The television broadcasts will show probabilities for winning in "real time". This can be used to practice and find those same percentages, which are relatively simple to compute. The foundation of each probability is finding the different cards that will allow for a win, against the total number of possible cards left. This is a nice extension of finding general probabilities for each type of poker hand.

Whenever combinitorics and probability comes up gambling is not far behind. This can cause problems in some cases considering the ethics of teaching typical gambling games in a classroom to students who are not legally able to wager bets. A couple of thoughts on the subject, and a general defense. First, it may be useful to make the distinction that while games are being taught and talked about, at no point is gambling going to occur. This is similar to a health class where effects of drugs are being discussed, but clearly no endorsement, or use, of drugs is happening. There is also historical context, as the earliest theories and work on probability was, in fact, motivated by gambling. The legacy remains, even in situations where gambling is no longer associated. For example, a hand with no high card points in bridge is called a Yarborough, named in honor of a lord who would offer his opponents a $1000 : 1$ payout if no points were dealt (odds of getting a yarborough are $1827 : 1$, so the Earl made quite a profit on this). Second, I would promote the idea that gambling institutions profit from a lack of knowledge of probability. Like the Earl of Yarborough, the idea of casinos is to present a situation that looks favorable to the gambler where in actuality the advantage is firmly in the direction of the house. Knowing exactly how much of an advantage is disheartening to a gambler, and in many cases will cause a loss of interest in gaming.

While it is useful to compute a couple of combinations and permutations by hand to get a sense for how they work, I quickly pull out the calculators. In fact permutations and combinations might be the functions I use most often on the calculator right behind the trig functions. There is no benefit to spending any extra time with the tedium of not using the calculator functions.

1.4 Discrete Probability Distributions

Two Types of Random Variables

Random variables are sometimes hard for students to understand the nature of the random variable. Because it is a new idea, this may take lots of time to make sure students understand what exactly is going on. I have found that discreet random variables are slightly easier than continuous ones. The problem is that the random variable can take on a number of unknown values, which also tend to be represented as variables. Random variables can be thought of as bins, that hold any number of other variables. This is helpful when talking about distributions later on.

Mean and Standard Deviation of Discrete Random Variables

This is where you need to connect back to the early lesson when we were pretty strict about calling the mean and the standard deviation parameters. When talking about distributions of a random variable, the most important thing is the mean and the standard deviation, and once the method of finding each is established most texts will then give a table of the common ones.

Up to this point there has been lots of theory. This section begins the statistical portion in earnest, and deserves some extra time for practice in finding means and variances. This is especially important before we start to define special distributions where means and variances will frequently be calculated by the rule for each individual distributions.

The Binomial Distribution

This the big one for discreet distributions. The formula is one that needs to be memorize, as so many problems come down to the binomial distribution. The idea of a probability of pass or fail applies in many circumstances. One place that can be fun to look at for students (especially those who are sick of medical studies...) is to look at some of the probabilities that are published about sports. For example, currently Accuscore is famous for running many simulations of each game a team plays to come up with probabilities that each team will make it to the playoffs. Presented in popular media, the mechanics of what goes on behind the scenes is unknown. (Gambling disclaimer! Accuscore attempts to profit off of selling their information to people who use it to place bets. They technically run a monte carlo program with using a binomial distribution as the probability distribution function (my assumption, but I'd be really surprised if I'm wrong.) They try to bamboozle people out of their money with fancy language and guarantees, another reason knowledge is power.) Students are more than capable of running their own simulations, for sporting events or other types of contests. The trick that students will probably discover is that the binomial distribution is great, but there is a huge catch to it. How to find p.

I try to get my students to come up with the questions on their own. Every question they are likely to encounter has something along the lines of "given that p is .6 . . .", so when students are given a more open ended problem, hopefully they will ask how to find the probability of success. They can be guided to begin thinking about it; asking "Does this answer make sense? How do you think they chose p?" after solving some of the books problems. It isn't always easy, or clear, how to choose success. Sometimes it is determined by a specific probability space, but for human studies, it can be a bit of a guessing game. It is an important consideration as students begin to learn about the job of statisticians and what is directly computed and what has to be assumed or chosen by the person doing the study.

The Poisson Probability Distribution

The Poisson distribution, also sometimes called the exponential distribution, is very useful in practice. It is use frequently for risk management scenarios and other applied questions in physics and business. The topic will also be a good one to revisit when confidence intervals are studied, as the two frequently go hand in hand.

However, it is not a topic on the AP examination. In checking a number of textbooks for a first year probability and stats course at the university level, it seldom made an appearance in those as well. Therefore, how you use this section is dependent on what the goals of your class are, how skilled they are, and how badly you may need to speed things along before the examination. Again, it is a great topic, but may not fit with your plans for this course.

The Geometric Probability Distribution

The geometric distribution has a very particular use. That is, how long before a single event will occur. The troubles students are likely to encounter are the geometric distribution's similarity to the binomial distribution and the circumstances that it is applied.

The forms of the two distributions simply must be memorized. There unfortunately no way around it. Something to focus on is what the variable in each distribution represents. In the binomial distribution the variable is how many successes out of a fixed total. The geometric distribution's variable is how many trials. Now if students are really clued in conceptually they will realize that the two can be confused because they are essentially the same. Look at a binomial probability of one success:

$$p(x) = np^1(1-p)^{n-1}$$

The only difference is the n multiplied at the front. This is because the binomial distribution does not care when you have the success. The Geometric distribution only give the probability for a success at a particular point, i.e. the probability of 1 success in 5 trials as opposed to having the first success on the 5^{th} trial.

Because of this choosing when to apply the geometric distribution is sometimes tricky. The thing to focus on is "Before the first success" or "until first success". The number of successes is fixed in the geometric distribution, and that can be the clue pulled from the problem to choose correctly.

1.5 Normal Distribution

The Standard Normal Probability Distribution

The big one. The book does an excellent job of introducing each key part of the normal distribution. I try to keep students away from the equation that describes the distribution for as long as I possibly can. By accessing students prior experiences with things that are normal, and then looking at the key parts of the distribution graph, students will have a better chance of a strong conceptual understanding that can be better applied for a variety of problems. Really, the equation of the distribution is mostly helpful only after calculus.

A key idea that must be stressed is the movement away from discreet distributions and into continuous distributions. This can be tough, because there are many things that we measure in a discreet manner, but model with the continuous curve. SAT scores is one, as there is no possibility of getting half of a question right (well, technically fractional credit is possible with the penalty for errors, but this still does not create a continuous score range) and no one gets half points on the scale after normalizing scores. The key here is the language that I used: we model behavior with the continuous distribution. Students will be well served to remember this little step to avoid thinking that normality requires continuity.

As a personal note, I really get personally confused by $Z-$scores. I suspect this is because I had many calculus classes before my first stats class, and therefore used calculus to solve the problems. The only reason why I bring this up is that reading a standard normal chart, finding $Z-$scores and relating normal distributions that might not be standard to the standard normal curve is a critical skill for a non-calculus based stats class. For my high school and university classes I would spend significant time on the algorithmic process of finding scores and relating them. It's one of the few times in a probability and stats class where a purely algorithmic skill is needed to be practiced (and practiced and practiced!) Plan extra time just for this skill.

The Density Curve of the Normal Distribution

A key idea that I talk about when introducing the standard normal curve, empirical rules and the standard deviation is where the inflection point is placed. This helps with students drawing good curves as well. The inflection point is always going to be a single standard deviation from the mean in each direction. This means that the inflection is going to "pull in" or "spread out" proportionally with the rest of the curve. I often have my students practice drawing generic curves with the same mean, and placing the inflection point in the same place. I don't focus on it yet, but I also want them to understand that the areas (under the whole curve, between two inflection points) must remain equal, therefore establishing the connection between the height of the peak and the spread of the graph.

I'd like to take an additional moment to stress the importance of sketching the graph and the area you are inspecting when working problems involving normal probability density. Because the direction of the "tail" changes the way things are looked at, or the additional steps needed to find an area that is in the middle, or split on both ends, the sketch and shading is critical to staying on track. Another thing that I have my students get in the habit of is labeling the area once they find it. For instance, if they are looking at the area between one standard deviation and the mean, they would shade the area and then label it ".34".

This is one of the key areas where calculators can really help with making things easy and fast. Because the calculator will find the continuous density, regardless of the location of the mean. This is a calculator skill I would make sure all students are comfortable with.

It would be wise to get a copy of the tables that will be provided in the AP examination. I don't have my students' use the charts in the text, but copies of the tables that they have for the examination.

Applications of the Normal Distribution

The lessons at the start of this section are designed primarily as an intermediate step. Rare is the occasion where those specific types of questions are asked. However, trying to solve questions in context requires both the mechanical understanding of how to find the missing information as well as the contextual understanding of what information is given, what is needed and how to set it all up. Trying to teach both at the same time might be too much all at once, hence these initial exercises. If these lessons are used, then I strongly recommend stopping before the section using real data, and having the students work problems 1-3 at the end of the section. Then after everyone is on the same page for those questions you can move on to the real data.

There are a couple of real data problems included in this section, but I would add a bunch more. Interpreting normal data is a huge part of what students will be asked to do. Fortunately, most of the data that is out there is normal, so finding data sets is pretty easy. I have included an appendix of sources from the internet at the end.

The big task here for students is making sure they know what quantity they are being asked to find. Most of the time this is straightforward because it is directly asked for. However, they might not be comfortable "translating" to the variable names that we have been using, like percentile, z-value, mean and standard deviation. For instance the last example asks "How rare is it that we find a female marine iguana..." This isn't using any term that we have before. Students, through the practice of doing many guided problems, need to develop a lexicon of the different ways that quantities can be asked for. In this case, how rare, what is the chance of, and other similar statements are looking for the percentage under the curve beyond a certain marker, in the case of this example below $400g$.

1.6 Planning and Conducting an Experiment or Study

Surveys and Sampling

This lesson is a bit on the thick side. It is reasonable to break it up into smaller parts and move them around as you see fit. Some items are interesting, but not always productive (like the discourse on randomness).

Something I like to do is find more of the bad sampling practices (or bad conclusions) and present them to students and ask them to figure out what probably went wrong. Some classic examples are the Dewey v. Truman presidential election (complete with incorrect newspaper headline!), why Yao Ming has made the NBA all-star game nearly every year, even when he is hurt and not playing, and others. Frequently these are not so cut and dry that one factor can singled out, but often many things combined to make a bad survey or poll. This is partially where students begin to understand the cliché "You can make statistics say anything you want..." which is frequently used by the undereducated to discredit well done surveys. The educated student realizes that there is plenty of opportunity for errors, and because of that, and other factors, absolute certainty is impossible, but a well done survey is a very solid source of information. Students need not memorize each type of bias by name. This is only useful for cocktail parties to stress that you know what you are talking about in terms of studies and surveys.

The randomness idea is lots of fun, but maybe beyond the scope of this class. Since we have not yet looked at the Uniform Distribution, which is how most computers generate random numbers (hence the TI calculators always returning a number between 0 and 1), we really don't have the impetus to work with computer generated randomness. An easy way to show the book's point about seeded generators is to reset the calculator. Don't do this if you have lots of programs in memory. A reset all, will set each calculators seed to the same number, and therefore every TI calculator will return the same "random" number after a reset. Even better, because the seed is incremented the same way for each calculator, the entire sequence of random numbers will all be identical until the seed is changed.

Experimental Design

It may not be critical to the AP examination, but one of the mantras in my classroom, whether it be a stats class or not, appears at the bottom of p235. "Correlation is not causation." For my students who are not going to pursue a future in math or science this is a critical idea. While any student at the end of a year of probability and stats can understand why, the more people who understand, and can communicate the idea, the more intelligent and informed decisions can be made as a community. It's one of the few chances we have as math instructors to teach for social justice.

Experiment design is one of the hardest things to "master" as a professional. Teachers are very familiar with this and I encourage you to share your experiences with educational research. It's tough for any study to give clear results due to the impossibility of true control, many, many confounding variables and the reality that a classroom will never be a lab, nor a lab a classroom. A task that is usually taken on by most AP stats classes after the exam is to design and carry out a large scale study involving their school or peers. I've had some amazing work done by students, including some really salient studies about drug use by students that was in direct conflict with the findings for our site from the CA Healthy Kids survey. The discussions about why the two findings were different were some of the best moments in any of my classes ever. It's maybe good to have the students start thinking about their projects now, if you plan to conduct one after the exam. It will help reinforce the ideas here, as well as set students up for success later on. With the way things go, the month and a half after the exam always seems like a ton of time, until you get there and it runs out in a hurry. Preparations must be made so that surveying can begin shortly after the test.

1.7 Sampling Distributions and Estimations

Sampling Distribution

Sampling distributions are frequently tough for students. Most of what has been studied up to this point are fairly concrete, but now there is a level of abstraction that can be tough to follow. The problem is that we are now talking about sampling a random variable that is itself a function of other random variables. The best way to handle this is to try to present as many different kinds of explanations as possible, using different language. One thing that is nice is that it is very possible to carry out an experiment in the classroom just like the example that is outlined in the test. Another good practice is to make absolutely sure you have been, and continue to be consistent with the notation used, with μ and $\bar{x}$ both representing means, but the population and the sample respectively.

This is the beginning of the some of the interesting facts about the normal distribution. There will be theorae later on that states why, but students should have their attention brought to all of the times that the normal distribution appears.

The formulae for sampling error, the mean and the standard deviation are ones that should be added to the "memorize this" list.

The z-score and the Central Limit Theorem

Using the term $z-$score is unfortunately necessary. It's always preferable to not introduce new terms when they are not needed. In this case the z-score is really just the standard deviations away from the mean, with the small exception of negative or positive signs indicating direction below or above respectively. It is a term that is in common use, however, and will be referred to by that name in both the AP examination as well as later courses students may take. The only possible exception being a calculus based stats class for math majors. It's always a good idea to cycle through some problems regarding reading z-score tables and standardizing values even outside of the chapters that include the topic.

The central limit theorem might be the most important single idea of a first year class. It is critical to know the formulas for sample proportions and sample means, but knowledge of the mechanics of the central limit theorem are not needed. The text does not present a formal proof, nor do most texts for a first year stats class. Because of student's work with various previous problems, students should be familiar with the idea that so many natural occurrences are normal, so a sort of study of a large number of studies should also be normal.

Binomial Distribution and Binomial Experiments

This is the beginning of the set of lessons where a common distribution is discussed, the mean and standard deviation is derived and then practiced. It is a common practice to not determine how to find the mean each time, but rather work from formulae once the distribution is determined. I will give my students a sheet of the common distributions and each of their means and variances. On a timed test it's a major advantage to be able to recall how to find each parameter, and move on with the problem.

Depending on what your technology resources, you may run a computer program that shows an animation of the binomial distribution sampling as n increases. There are instructions available for programs like geometers sketchpad and fathom, as well as some java applets, like [INSERT LINK]. This can give a nice conceptual sense for how, even if the probability of success is off to the side, the sampling approaches the normal distribution.

Confidence Intervals

Statistical inference is a large part of the *AP* examination. Confidence intervals is probably the key topic for the section on inference. It's also sometimes counter intuitive to students, so careful use of language is required.

I force students to use very precise language here. The correct language is "95% confidence level", not 95% chance, or 95% probability of anything else similar. It can be considered picky, but there is no probability here, and that's important. Especially as students are considering different distributions and samples of different distributions, clarity helps.

It's worth having a discussion with students about what they consider to be an appropriate level of confidence for their studies. Obviously 99% is very strong, but students should be aware of how much more "expensive" that level of confidence is. 90% is very easy, but doesn't make you feel very, well, confident. The common levels of confidence tend to be $90\%, 95\%, 97\%$ and 99%. Students should work some problems and get a sense for what they feel the "sweet spot" is for keeping the sample size manageable, but gives a high enough confidence level.

Sums and Differences of Independent Random Variables

I can get confusing for students when trying to figure out when to add probabilities and when to multiply. There are a couple of tricks to help students out. First, I always tell them that choosing one and then another, without replacement, is logically identical to choosing two at the same time. In the case of the book's example of miners, the question can be reformatted to choosing one miner, than another. This clearly implies a multiplication of the two individual probabilities to get the probability for the two together. However, if we then ask about the probability of getting at least one of the two having the illness, then adding the different probabilities is needed. The difference is that the multiplication happens on the "front end", multiple people, multiple coins, multiple dice and so on. The addition happens on the "back end", where a condition is set where multiple outcomes can meet the requirement.

I tend to define expected value for my students as "probability times payout." This isn't always strictly true, like for the television hook-up example, the number of TVs in the house is not directly a payout. Students seem to be able to make the connection fairly easily, and it applies most of the time as expected values are connected to wagering and business more often than not. I would shy away from the contribution to the mean language.

The linearity problem is the first time that students will see why the variance is used frequently in upper level courses. Stress that you can't use the standard deviation in the same way because of the rules of the square root, although I am sure you will come across at least one student who tries.

Student's t Distribution

The story of the Student's t is kind of cool. The person responsible for the invention of the distribution is William Sealy Gosset. Gosset worked for Guinness who was, at the time (1900s), interested in scientifically boosting barley crop production. His work involved small sample sizes, so he had to find a distribution that would work in testing different plots against each other. Due to a previous employee publishing trade secrets of brewing in an academic paper, all employees were barred from publishing, regardless of content. As a consequence, Gosset chose to publish under the name "Student".

One thing that some statistics books will stress is that there is no proving the null hypothesis true. The book dances around this, but if you are aware of the rule then you will see that they specifically say "accept the null hypothesis" or "there is no evidence against the null hypothesis." Strictly speaking, rejecting the null hypothesis is a stronger condition than not rejecting one, or accepting one. It's kind of like how a single counterexample will disprove a theorem definitively, but no amount of examples in support will prove that it is true. The supporting examples will give hints to give you confidence that the theorem is probably true, and that is exactly how a non-rejection should be treated. Sometimes the null hypothesis is written in a manner where rejection is the goal, therefore showing the opposite has a high degree of certainty to be true. Therefore, students need not only the skill to compute the statistic, but writing good null hypotheses.

Degrees of freedom is one of those things that sounds fancy, but really isn't. There are reasons for the name, but they are beyond the scope of the class, and really aren't necessary. The beginning and end of what students need to know is that it's $n-1$.

1.8 Hypothesis Testing

Hypothesis Testing and the p-value.

In this section, you will need to key your students into the precise language being used. While it may not always be completely fair, but the proper language of hypothesis testing is one way that tell the people who have a proper training in statistics apart from the people who are pretending. This is possibly an issue when writing summaries on the open response questions on the AP examination. The use of proper language goes a long ways to promoting the idea that you know what you are talking about. The key phrases are "There is no correlation...", "There is no difference..." for the start of every null hypothesis, " Does not reject", "Does reject" in regards to the results of the test. It might seem strange to students, but this isn't the place for creativity or fancy language (or good grammar... sort of; many standard math phrases that are used over and over again are grammatically suspect, but we plow ahead anyway).

A good tip for helping students to understand hypothesis testing is to look at a couple of examples for each idea. For instance, when you introduce two-tailed testing, give students some problems where the null hypothesis is already written and they only need to evaluate whether or not they are rejecting in based on the level of significance chosen. If the information is not broken up, then students easily get lost in the details.

Calculating error and knowing the different types may seem trivial. It also isn't a large part of later classes and work in statistics. However, it is something that students are expected to know for the AP examination, so they should be made aware of that.

Testing a Proportion Hypothesis

This chapter is about the time that students start having a really tough time keeping all of the test and sample statistics straight. It may be a good use of time to take a break from new information and have students study, re-copy or interact with each of these statistics so they can hopefully keep them straight. The AP examination can be pretty stressful, and if time isn't taken to review throughout the year, it will be hard for students to recall what they need in the test.

Proportion hypothesis testing is a simple instance of hypothesis testing. There are a couple of things that work slightly differently, like calculating the standard deviation, that should be the focus of this unit.

Testing a Mean Hypothesis

It is important for students to have an intuitive sense of why the procedures need to change for small samples. The test outlined here should be tested with the chi-squared distribution, but the idea is the same regardless. Everyone is aware that the theoretical probability for a coin flip is .5. Students can then perform experiments with small numbers of observations, say 1 to 10. Discussing all the experiences from the students, they will see that there is great variation in results even though everyone has a fair coin. If all results are combined on one graph, it should be clear, and intuitively so, that the more flips performed resulted in a better match to the theoretical mean. Going one step further, it's simply impossible to be anywhere close to the theoretical mean with $1, 3$ or 5 samples. Students should be asked "what are the implications of this?" Likely answers include: the need for more samples (which is not always possible), the problem with using binary or discreet outcomes (get used to it, rarely can we depend on continuous empirical results), and most importantly, the uncertainty that is inherent to small sample sizes. What would it take to be certain that a coin is fair with only 5 experiments? Most students will probably indicate that the they could not reject the hypothesis of the coin being unfair regardless of outcome. Now bumping it up to 10, now there will probably be some outcomes that will not be accepted. This is the conceptual foundation for small population, or non-normal tests.

Testing a Hypothesis for Dependent and Independent Samples

It is not standard practice among different texts to assign different symbols for the hypothesis depending on if it is a dependent or independent test. However, this is a good idea, especially as it is easy to get confused or forget a step. The key part of this section is understanding the procedure for testing two different populations, especially when testing for growth. This is the first instance where students will have a chance to see and understand how to show a difference outcome based on procedure. This has to take into account the baseline information for each party, making it a dependent sample. The most interesting consequence for students, as outlined in one of the examples, is the utility of this process to examine various practices in school. If students are working on a project for the end of the year, this may become an integral test for answering many of the questions that students might have about their school.

Students may become a little lost with all the tests. There is a chart in a later section outlining when each test is to be used. It is also advisable at this point to have the students create one of their own, with the creation being an exercise to help them remember.

1.9 Regression and Correlation

Scatterplots and Linear Correlation

Graphical representations are going to be the primary focus of bivariate data in a first year class. Many of the techniques required for statistical analysis of multiple variables requires calculus, so a basic treatment is all students are equipped for at this point. The scatterplot is a very familiar structure to students at this point. In some ways your task is not going to be teaching the principles of a good scatterplot, but rather un-teaching the bad habits or misconceptions that students have developed over the years. Those bad habits are usually focused around poor scaling, sloppy labeling and a general lack of precision. If they are representing data for analytical use, they will need to take extra care in having a graph that is clear and accurate enough to be useful. Also, simply using a computer grapher is not sufficient, as the scaling can still induce poor conclusions.

For many complex measures, like the various correlation coefficients, a table where values are determined step by step, as on page 341, is very useful in keeping all of the variables and values in the correct place. (It's also a great tool for finding standard deviations by hand. If you are careful, you will realize that there is plenty conceptually in common between the standard deviation and correlation coefficients.)

Least Squares Regression

This is a biggie. Not only for the class or the AP exam, but for understanding all kinds of statistical work in the future. If you have students designing statistical projects, some of them will likely need to use least squares regression for their work. Plan a touch of extra time to make sure the class understands this section.

There are two ways to go about presenting this chapter. One is to have students work out their own plan for finding the best fit line. They should be capable of developing a number of different ideas, with a little bit of guidance. I start out by having students, or groups of students, construct a line with a straightedge that they think is the best fit. Inevitably, there will be some students who have different opinions. I ask if anyone can think of a way to test to see analytically who's line is the closest fit for all of the data. Usually students get very close, if not exactly on the correct answer. The drawback to this method is that it takes time, and that it can be confusing for students in the long run. There are many valid methods of finding or confirming a line to fit data. Least squares is the most common in statistics, so students are expected to know it. Some may get confused if a number of ideas are presented.

The other method is to go straight for the table, like the text does for example 1. This is very quick, clear and usually results in a great rate of success for students working these problems. The drawback anytime a strict algorithmic method is applied is that students miss the concept behind the method. This is of lesser importance in this section as opposed to others.

Of additional information presented, all of it is good, but the only part that is really important for the AP examination is the part on calculating residuals. Students may also be asked in a free response question to perform a t-test on residuals to determine if the line is a good enough fit.

Inferences about Regression

The AP examination will usually ask a question where students are required to make an inference about the correlation between two statistics. There are a number of steps to doing so, all of which have been covered somewhere in this text, but not all of them in this section. At this point, students should be reminded of all of the conditions that are required for this inference. Some of it may seem like it's tedious routine, but it is easy to apply rules and tests in statistics in places where they are not going to give meaningful results. Furthermore, the results will seem logical, taking away the logical check system that students usually have.

The sample must be random, as with nearly anything that we are going to be looking at. This is again taken from the design of the experiment as covered in earlier chapters. The errors must be normal, which can be checked through various plotting methods. This is one condition that can frequently be overlooked without consequence, but is technically a requirement. Residual errors must be centered around 0. This can be figured with a plot, or by finding the mean of the errors. Along with the mean of the errors, and another reason to make a residual plot, the standard deviation of errors should be the same for all X. The fact that the errors are independent can also be determined from this plot, provided no trends in plotting are popping up. Once all of these items are checked, then the process can proceed as stated in the examples in the text. Presenting the solution will be a good choice on the AP free response examination for clarity and to show that it is known what the requirements are for the test.

Multiple Regression

Multiple variable regression is tough to visualize. This is compounded by the problem of making 3-D graphs. For this reason, and the fact that the there is no limit as to how many variables can be used, it makes sense to show a single example with a graph, and then move onto making the computations without a visual. There are many instances in mathematics where a two variables, or even three, are used to graphically develop a rule that can be extended beyond what can be represented; linear programming is a classic example. There is nothing lost by having students simply follow steps for a solution now that they have experience with regression for two variables. Another good plan is to use technology in this section for ease of solution or visualizing results. The text mentions SAS and SPSS but there are many stats packages that will perform regression with multiple variables.

1.10 Chi-Square

The Goodness of Fit Test

I have found that students often have an easier time with chi-squared tests than with many of the other tests. Also, the chi-squared distribution can be intuitively constructed from binomial trials. For those reasons, I choose to introduce the chi-squared distribution before the t or the F distributions. The other nice thing about working with the chi-squared test is that the experiments are quick and easy to conduct. After presenting the information I will have a chart up at the front of the class everyday polling the students. Students are now asked to formulate the null hypothesis (which I have already developed to choose the poll question) and then perform the chi-squared test and make a conclusion.

A bit of caution about something in the text. Students are going to want to hold on to a particular chi-squared value to apply to all situations, regardless of the degrees of freedom or level of confidence. The book doesn't help as it almost sounds like it declares that a particular chi-squared value is the threshold for rejection. I always make sure that students have to work different problems so they don't develop a habit.

Test of Independence

It is debatable whether you should keep this topic separate from the previous, or if they should be merged. The purpose of merging the two topics is that they are really the same. The difference in the test for independence is the same as fit, except for the way the null hypothesis is written. A reason to give it its own treatment is because the AP examination treats it more in that manner. I tend to treat them as one section.

Conceptually, the difference is in how the null hypothesis is written and correlation is shown the same way that independence is. Also, while other sections can be helped by being stretched out a little, students do not need as much practice with chi-squared tests as many of the others.

A fun activity I remember from university is that we ran chi-squared tests for homogeneity on various random number generators. Even with different seeds, the random generator on the calculator did not fare well in our experiment. I can be a very fun one for students, especially if human responses are added in.

Testing one Variance

The text mentions that the $F-$test is sensitive to non-normality, which is only partially true. In this chapter it is, as the specific instance of the $F-$test being sensitive is when showing that variances are the same. The F-test is actually guarded pretty well against non-normality in other methods of testing, which is why it is a part of ANOVA, which is specifically mentioned as a robust test. I don't know how much I would bother students with these details, but since a good student is diligent about checking requisite conditions, it may be worth mentioning.

1.11 Analysis of Variance and the F-Distribution

The F distribution and Testing Two Variances

Note that the F distribution and other tests of variance are not topics on the AP examination. This does not make them less valid, but they should be retained until after the examination at topics to fill out the remainder of the year.

This may be a little bit tough for students to understand the motivation behind these tests of variance. Another problem is that the F distribution is complex enough that first year students are not exposed to details behind the distribution. The best thing for students to remember is that while means are often easy to find or approximate, the variance is not. The key use of the F test is going to be when two populations are being examined, or when a comparison can be made to a known variance.

The One-Way ANOVA Test

ANOVA tables are computed frequently in professional statistical analysis. They can be tedious, but in the real world they are nearly always done with the assistance of some computational software. I suspect that it is because of this that ANOVA tables are not a required topic for the AP examination. I think that I have completed a single ANOVA table by hand, ever. This is also what I have my students do and then move quickly on to learning how to use the software program of choice.

ANOVA is a difficult procedure to run on some software packages. At the very least, there is little consistency between any of the software programs. In fact, I was asked to present to the math department when I was an undergraduate the different ANOVA output from different software packages. It took me a very, very long time to prepare the presentation due to the difficulty of learning all of the different syntax and output. I had looked at Mathematica, MAPLE, SPSS, SAS, R, S+ and Excel. I did not have access to Minitab, MATLab (I don't know for sure if the base package will do ANOVA) or Fathom. The most interesting finding was the extreme limitations of Excel, as well as some errors in calculation. Excel errors are well documented, and many have workarounds, but I can't ever recommended it as an accompaniment to the stats class. Maple, SPSS and I have heard Fathom, are pretty user friendly and are popular choices. Whatever your package is, make sure to block out time around these chapters to get comfortable with using the software for these computations.

The Two-Way ANOVA Test

Since there is little difference between a one way and a two way ANOVA test, there shouldn't be much practice necessary. There are a couple benefits to this. First, it is a way to reinforce some of the topics learned in the one way lesson. Second, by contrasting some of the elements that are different about the two way test, the important parts of the one way test become apparent. Remember that the one of the toughest things about seeing a new topic for the first time is that everything seems of equal importance, and everything seems unique. This is especially true as the theory gets tougher and more and more examples are presented. Students will have trouble at times telling what parts of the problem are specific to the example and which parts are always going to be part of the ANOVA table. Looking at two way tests can help with this.

Another consideration is how much time needs to be spent on setting up data for each program. I know that some programs require entry in cells, or imported form comma separated value files. Other programs require the data in the form of a matrix. Students need to be aware of what the requirements of their software is, but it will also help if they know a little about what possibilities are out there, in case they come across different software at a later date.

1.12 Non-Parametric Statistics

Introduction to Non-Parametric Statistics

Non-parametric statistics are a more exotic test. The situations in which data is encountered that is not normal are few, and as a consequence this section is very rarely covered in any statistics course, let alone a high school level AP course. These sections are advisable only for students who are truly exceptional. Furthermore, these tests are almost never run by hand. Every statistics specific package will have tests, and they should be used.

An interesting note is that a common use of the runs test is to test for randomness. This makes sense, as testing for correlation is the opposite, and obviously random numbers will not be randomly distributed.
Appendix A: Online Data Sources

There are a number of sources of real data that is pertinent and interesting for projects and assignments for students. Here is a short list of sources I have used.

Data.gov: `http://www.data.gov/catalog/raw`

The central clearinghouse for raw data for all of the federal government agencies. Data is available in XML or CSV files.

US Census: `http://www2.census.gov/census_2000/datasets/`

Raw data from the 2000 census. There are other tools on the census site, but most of them manipulate the data too much rather than just presenting numbers.

UN Data: `http://data.un.org/`

A clearinghouse for the data collected from various UN organizations including the WHO, WTO, UNICEF and UNESCO.

Baseball Reference: http://www.baseball-reference.com/

While there are other sports that collect statistics, none do it with the verve of baseball. This is the most complete resource for all kinds of baseball statistics.

Appendix B: Statistical Software Packages

This list is not exhaustive, but should give a brief overview as to what software is available to support classroom activities.

Fathom: http://www.keypress.com/x5656.xml

Fathom is published by a textbook publisher, so it has a wide following in schools. It is easy to use, powerful enough for most, and has many activities and lesson plans available for use. It also is one of the cheaper packages. Windows and MacOS

SPSS: http://www.spss.com/

SPSS has outgrown its acronym: Statistical Package for the Social Sciences. It is now a huge data mining and statistics suite owned by IBM. It is easy to get started with, as it uses a spreadsheet type interface for most of its data. Expensive. All platforms.

SAS: http://www.sas.com/

SAS is geared more towards enterprise uses. I have never been in a school lab with SAS installed, but it remains a popular choice for many, with lots of online community help and resources. Expensive. All platforms.

R: http://www.r-project.org/

R is a FREE open-source clone of the functionality from the legendary math and statistics package S and S+ from AT&T. It is huge, the documentation is hard to read, it has few graphical interface items, and is truthfully more of a programming language/environment. It's what I use. It does everything. Did I mention its free? All platforms.

Minitab: http://www.minitab.com/en-US/default.aspx

Minitab is, in my experience, somewhere between SAS and SPSS in usability. It is a favorite of many business schools. There is a significant group of AP teachers using Minitab in their classes, so lesson help should be easy to find. Expensive. Windows only.

Mathematica: http://www.wolfram.com/

An all around math software and programming environment. It's using a cannon to kill a fly for a stats class. Tough to use, but has incredible documentation. Very, very expensive, but massive discounts are available to teachers and students. All platforms.

Excel: http://office.microsoft.com/en-us/excel/default.aspx

You were waiting for this one... Well, the good news is you probably have it already, and everyone is familiar with at least the basics. The bad part is that the range of tests is very limited, and there are an incredible amount of documented errors with various distributions and functions. I can't recommend it, but it has worked for some in the past, and will continue to work for many in the future. You probably don't need to buy it. Windows and MacOS.

Random.org: http://www.random.org/

Not software, but a website that is committed to provide about as good of random numbers as is possible. There is also a ton of info on randomness, why it's elusive and some other techniques for better generating randomness. A great resource for teachers and students. Free!