WHY YOU NEED TO TAKE STATS 101

BY STATS WITH CATS

AUGUST 2020

Contents

Why Do I Have To Take Statistics?................................................................. 3
   Statistics provide a starting point and a course of action.......................... 3
   Statistics give you more ways to analyze data .................................... 4
   Statistics examine both accuracy and precision ................................... 4
   Statistics examine both trends and anomalies ..................................... 4
   Statistics tells you how much information you need ............................. 4
   Statistics provide standardization ...................................................... 4

Five Things You Should Know Before Taking Statistics 101 .......................... 5
   Everything is Uncertain ...................................................................... 5
   Statistics ♥ Models ......................................................................... 6
   Measurement Scales shape Analyses .................................................. 7
   Everything Starts with a Matrix ......................................................... 7
   Statistics is More than Description and Testing .................................. 7
Whether you’re in high school, college, or a continuing education program, you may have the opportunity, or be required, to take an introductory course in statistics, call it Stats 101. It’s certainly a requirement for degrees in the sciences and engineering, economics, the social sciences, education, criminology, and even some degrees in history, culinary science, journalism, library science, and linguistics. It is often a requirement for professional certifications in IT, data science, accounting, engineering, and health-related careers. Statistics is like hydrogen, it’s omnipresent and fundamental to everything. It’s easier to understand than calculus and you’ll use it more than algebra. In fact, if you read the news, watch weather forecasts, or play games, you already use statistics more than you realize.
WHY DO I HAVE TO TAKE STATISTICS?

When you were in school, you probably asked the question, why do I have to take statistics?” Your adviser told you: “because it’s required for the degree.” “But why,” you said “why would I ever need to use statistics?”

Everybody who has completed high school has learned some statistics. There are good reasons for that. Your class grades were averages of scores you received for tests and other efforts. Most of your classes were graded on a curve, requiring the concepts of the Normal distribution, standard deviations, and confidence limits. Your scores on standardized tests, like the SAT, were presented in percentiles. You learned about pie and bar charts, scatter plots, and maybe other ways to display data. You might even have learned about equations for lines and some elementary curves. So by the time you got to the prom, you were exposed to at least enough statistics to read USA Today. In college, you’ll find that most majors require some statistics. Why? Consider the following.

Statistics is an integral part of everyday life in America. Without statistics, there would be no U.S. Census, IRS audits, Nielsen ratings of TV shows, political polls, and consumer preference surveys. Our society couldn’t function without being able to figure out tax brackets, insurance rates, stock prices, and online matchmaking. We couldn’t predict the outcome of elections before the polls close. There would be no standardized tests, no ACT, GRE, TOEFL, MBTI, or CATs (MCAT, LCAT, PCAT, and VCAT). Amazon.com couldn’t tell us what we want to buy. Baseball announcers would have nothing to talk about between pitches. It would be anarchy.

If you’re still not convinced that you need to learn statistics, keep reading.

The use of statistics is common to almost all fields of inquiry—social and natural sciences, sports, business, education, library and information science, and even music and art. Its popularity is attributable at least in part to its applicability to any type of data. Statistical methods can be used for analyzing data based on natural laws, theories, or nothing in particular. If you can measure it, you can analyze it with statistics. If you’re creative enough, you can even analyze things you can’t measure very well.

So why do your advisors want you to take statistics? Here are a few of the reasons.

STATISTICS PROVIDE A STARTING POINT AND A COURSE OF ACTION

If you’re in the natural sciences, you’ll probably have some basic principles, laws, or at least theories to start with in analyzing data. Even some of those were discovered or verified by statistical observation. If you’re in the social sciences, business, economics, or most other fields, though, you’re got little to go on besides statistics. Anecdotes aren’t worth much.

Statistics gives you a place to start by having you focus on the population, so you know what to sample, and the phenomenon, so you know what to measure and how to measure it. Once
you have laid this groundwork, statistics has you define alternative hypotheses to weigh and provide a variety of methods to analyze the data.

**Statistics give you more ways to analyze data**

Statistics is a colossal workshop with more tools than you could ever use in a career. Statistics allows you to describe, correlate, detect differences, group, separate, reorganize, identify, predict, smooth, and model. And it’s not just the variety of tools for doing different things, there are also many tools for doing the same thing in different ways. Want to find the center of a data distribution? You can use the arithmetic mean, the geometric mean, the harmonic mean, trimmed and winsorized means, weighted means, the median, the trimean, or the mode. Each has its own special use, like the variety of types of screwdrivers used by a mechanic. With a statistician’s toolbox, you can gain far more insight from your data than you might from any other type of analysis.

**Statistics examine both accuracy and precision**

Any marksman will tell you that it’s not enough to be able to hit a target. You have to be able to hit it where you aim and do it consistency. That’s accuracy and precision. Many analytical techniques focus on accuracy and forget all about precision. But variability, uncertainty, and risk don’t go away by just ignoring them. Statistics is all about understanding variability.

**Statistics examine both trends and anomalies**

Statistics, in particular, can be used to find linear and nonlinear trends, cycles, steps, shocks, clusters, and many other types of groupings. What’s more, statistics can be used to identify and explore divergent or anomalous cases, which don’t fit general patterns. Sometimes it is these outliers rather than the trends that reveal the information most crucial in an analysis.

**Statistics tells you how much information you need**

In data analysis, more is not always better. It’s not unusual to have too much data to make sense of using only graphs and tables. Statistics provides a variety of ways to help you decide about how many samples you need to achieve a certain objective. Statistics provides ways to judge the quality of the data and compensate for misleading variability. Statistics can also tell you if your data are redundant, and if so, provide ways to reassemble the data more efficiently.

**Statistics provide standardization**

You can usually convince people who are reviewing your work that your data analysis is legitimate because it uses well-known, professionally accepted, statistical procedures. Likewise, it’s easier to use statistics as the basis for any standardized procedures you specify that others use because most people know some statistics. For example, Government regulations frequently require the use of statistics to report and analyze data sets, such as crime rates, pharmaceutical effectiveness, environmental impact, occupational safety, public health, and educational testing.
So you see, statistics has a lot to offer you, whether there is a strong theoretical basis to your field of practice or not. That’s why your advisers want you to learn about it.

**FIVE THINGS YOU SHOULD KNOW BEFORE TAKING STATISTICS 101**

Of the over two million college degrees that are granted in the U.S. every year, including those earned at accredited online colleges nationwide, probably two-thirds require completion of a statistics class. That’s over a million and a half students taking Statistics 101, even more when you consider that some don’t complete the course.

Everybody who has completed high school has learned some statistics. There are good reasons for that. Your class grades were averages of scores you received for tests and other efforts. Most of your classes were graded on a curve, requiring the concepts of the Normal distribution, standard deviations, and confidence limits. Your scores on standardized tests, like the SAT, were presented in percentiles. You learned about pie and bar charts, scatter plots, and maybe other ways to display data. You might even have learned about equations for lines and some elementary curves. So by the time you got to prom, you were exposed to at least enough statistics to read USA Today.

Faced with taking Statistics 101, you may be filled with excitement, ambivalence, trepidation, or just plain terror. Your instructor may intensify those feelings with his or her teaching style and class requirements. So to make things just a bit easier, here are a few concepts to remember.

**EVERYTHING IS UNCERTAIN**

The fundamental difference between statistics and most other types of data analysis is that in statistics, everything is uncertain. Input data have variabilities associated with them. If they don’t, they are of no interest. As a consequence, results are always expressed in terms of probabilities.

Every data measurement is variable, consisting of:
• **Characteristic of Population**—This is the part of a data value that you would measure if there were no variability. It’s the portion of a data value that is the same between a sample and the population the sample is from.

• **Natural Variability**—This part of a data value is the uncertainty or variability in population patterns. It’s the inherent differences between a sample and the population. In a completely deterministic world, there would be no natural variability.

• **Sampling Variability**—This is the difference between a sample and the population that is attributable to how uncharacteristic (non-representative) the sample is of the population.

• **Measurement Variability**—This is the difference between a sample and the population that is attributable to how data were measured or otherwise generated.

• **Environmental Variability**—This is the difference between a sample and the population that is attributable to extraneous factors.

The goal of most statistical procedures is to estimate the characteristic of the population, characterize the natural variability, and control and minimize the sampling, measurement, and environmental variability. Minimizing variance can be difficult because there are so many causes and because the causes are often impossible to anticipate or control. So if you’re going to conduct a statistical analysis, you’ll need to understand the three fundamentals of variance control—Reference, Replication, and Randomization.

**STATISTICS ❤ MODELS**

*Statistics and models are closely intertwined.* Models serve as both inputs and outputs of statistical analyses. Statistical analyses begin and end with models.

Statistics uses distribution models (equations) to describe what a data frequency would look like if it were a perfect representation of the population. If data follow a particular distribution model, like the Normal distribution, the model can be used as a template for the data to represent data frequencies and error rates. This is the basis of parametric statistics; you evaluate your data as if they came from a population described by the model.

Statistical techniques are also used to build models from data. Statistical analyses estimate the mathematical coefficients (parameters) for the terms (variables) in the model, and include an error term to incorporate the effects of variation. The resulting statistical model, then, provides an estimate of the measure being modeled along with the probability that the model might have occurred by chance, based on the distribution model.
You may not hear very much about measurement scales in Statistics 101, but you should at least be aware of the difference between nominal scales, ordinal scales, and continuous scales. Nominal scales, also called grouping or categorical scales, are like stepping stones; each value of the scale is different from other values, but neither higher nor lower. Discrete scales are like steps; each value of the scale has a distinct break from the next discrete value, which is either higher or lower. Continuous scales are like ramps; each value of the scale is just a little bit higher or lower than the next value. There are many more types of scales, especially for time scales, but that’s enough for Statistics 101.

The reason measurement scales are important is that they will help guide which graph or statistical procedure is most appropriate for an analysis. In some situations, you can’t even conduct a particular statistical procedure if the data scales are not appropriate.

EVERYTHING STARTS WITH A MATRIX

You may not realize it in Statistics 101, but all statistical procedures involve a matrix. Matrices are convenient ways to assemble data so that computers can perform mathematical calculations. If you go beyond Statistics 101, you’ll learn a lot about matrix algebra. But for Statistics 101, all you have to know is that a matrix is very much like a spreadsheet. In a spreadsheet you have rows and columns that define rectangular areas, called cells. In statistics, the rows of the spreadsheet represent individual samples, cases, records, observations, entities that you’re making measurements on, sample collection points, survey respondents, organisms, or any other point or object on which information is collected. The columns represent variables, the measurements or the conditions or the types of information you’re recording. The columns can correspond to instrument readings, survey responses, biological parameters, meteorological data, economic or business measures, or any other types of information. You usually have several sets of variables for a given set of samples. Together, the rows and the columns of the spreadsheet define the cells, which is where the data are stored. Samples (rows), variables (columns), and data (cells) are the matrix that goes into a statistical analysis. If you understand data matrices, you’ll be able to conduct statistical analyses even without your Statistics 101 instructor to help you.

STATISTICS IS MORE THAN DESCRIPTION AND TESTING

In Statistics 101, you learn about probability, distribution models, populations, and samples. Eventually, this knowledge will enable you to be able to describe the statistical properties of a population and to test the population for differences from other populations. But these
capabilities, formidable though they are, don’t reveal the truly mind boggling analyses you can do with statistics. You can describe, compare and test, identify and classify, predict, and explain. So, don’t get discouraged if you can’t see how statistics will help you in your career based on Statistics 101. There’s a lot more out there. You just have to take the first step.

THE FIVE PURSUITS YOU MEET IN STATISTICS

When people think about statistical analyses, they often think only of mind-numbing number crunching that creates yet more numbers. But that’s like touring a cabinetmaker’s shop and seeing only the sawdust. A talented cabinetmaker can create beauty and function in his products. In the same way, a creative statistician can create enlightenment and utility if he or she has vision and purpose.

Statistical analyses usually aim at achieving one of five objectives:

- Describe
- Identify or Classify
- Compare or Test
- Predict
- Explain.

DESCRIPTION

Description is relatively straightforward. It involves characterizing populations and samples using descriptive statistics, statistical intervals, correlation coefficients, graphics, and maps. You can do the calculations on spreadsheet software. All you have to be aware of are measurement scales, distributions, sampling schemes, measures of central tendency and dispersion, and methods for dealing with outliers and missing data.
IDENTIFICATION AND CLASSIFICATION

Identification and classification involves the analysis of a known or hypothesized entity or group of entities using descriptive statistics; statistical intervals and tests, graphics, and multivariate techniques such as cluster analysis. Analyses range from simple visual recognition to the exploration of arcane mathematical dimensions where only bold number crunchers venture. It’s like finding Waldo. At a convention of funeral directors, one look would be all you needed. If he were making American flags in a candy cane forest, you might need some non-visual clues. You can determine a person’s sex by looking at him or her but not from a table of eye and hair color. On the other hand, you couldn’t tell who the best players were on a sports team from their pictures, but you could from their performance statistics. However you do it, identification is the gateway to classification. If you can do one, you can probably do the other.

COMPARISON

Comparison involves detecting differences between statistical populations or reference values using simple hypothesis tests, and analysis of variance and covariance. This objective is tougher than describing or classifying even though there is ample software available for most analyses. You need to know what test to run or ANOVA design to use as well as understand probability, effect size, and violations of assumptions. There’s a much greater chance of something going wrong.

PREDICTION

Prediction involves estimating new values using regression and neural networks, forecasting using time-series modeling techniques, and interpolating spatial data. In addition to all the description and comparison techniques, you’ll need to know how to use a variety of model building and assessments methods and understand the morass of prediction error. It’s easy to make a prediction. It’s hard to make an accurate prediction. It’s damn near impossible to make an accurate prediction that is also precise. Even if you did nothing wrong statistically, it’s easy to produce a poor prediction, and a poor prediction will eventually be noticed. One really good prediction and a psychic is famous; one really bad prediction and a statistician is relegated to selling insurance.
EXPLANATION

Explaining latent aspects of phenomena is the toughest of all objectives. Not only do you need to understand some sophisticated statistical methods, using regression, cluster analysis, discriminant analysis, factor analysis, canonical correlation, and data mining techniques, but you also have to understand the conceptual framework of the systems the data come from. Then, you have to have the talent to apply the knowledge creatively. You can’t explain your statistical model of stream contamination without knowing something about stream hydraulics, hydrogeology, meteorology, and environmental geochemistry. You can’t explain customer satisfaction without knowing something about demographics, marketing, business, and psychology. You’ll also probably have to integrate the information and think of it in ways that have never been thought of before.

Explanation can create fundamental wisdom, although most of the time, your results will be humdrum. If you do come up with something truly consequential, though, some people will believe your results are erroneous, coincidental, or faked. Some people will claim that your finding is old news, having discovered it themselves years before, but then post it on Reddit for the karma. Most people, though, will just ignore you.

TOOLS OF THE FIVE OBJECTIVES

The following table provides some examples of data analysis tools that can be used for addressing the objectives.
There are other classification schemes that describe other statistical pursuits, so don’t feel constrained by these five categories. But this classification of statistical aims is a reasonable place to start. It has three features. First, it’s easy to figure out so non-statisticians can decide in which category their project fits. Second, the major statistical techniques tend to be used primarily in just one of the classifications. And third, the scheme can be thought of as an index of the professional peril a statistician could face in doing the analysis.

Creating a finished statistical analysis from raw data requires knowledge, experience, and often a bit of artistry. So when you conduct or review a statistical analysis, don’t let all the numbers obscure the craftsmanship and functionality of the products. And accordingly, don’t neglect to appreciate the talent and the artistry of the numbermaker.

**SIX MISCONCEPTIONS ABOUT STATISTICS YOU MAY GET FROM STATS 101**

When you learn new things, you can develop misconceptions. Maybe it’s the result of something you didn’t understand correctly. Maybe it’s the way the instructor explains something. Or maybe, it’s something unspoken, something you assume or infer from what was said. Here are six misconceptions about statistics you might have gotten from Stats 101.
MISCONCEPTION 1: “STATISTICS IS MATH”

How could you not come to believe this? Even before you took Stats 101, you learned you had to take the course to fulfill a math requirement. It was taught by the Math Department. Then when you took the course, it was all numbers. Homework and exams were almost all about calculations. Stat 101 was all math. Statistics must be all math too.

Reality – Statistics uses numbers but numbers are not the primary focus of statistics, at least to most practitioners. Applied statistics is a form of inductive reasoning that uses math as one of its tools. It also uses sorting for ranks, filtering for classification, and all kinds of graphics. The point of using statistics is to discover new knowledge and solve problems through the use of inductive reasoning involving numbers. It’s not just about doing calculations. That’s why it’s required for college majors in business, social sciences, and many other disciplines. That’s why it’s taught by professors in all those disciplines, too. Yes, it’s required for math degrees and is taught by math professors at many schools. That’s so there will be mathematical statisticians who will invent statistical tools for the applied statisticians to use. You can love statistics and be good at statistical thinking even if you think you hate math.

MISCONCEPTION 2: “STATISTICS REQUIRES A LOT OF DATA”

Stats 101 doesn’t teach you how to work with individual pieces of information, like a solitary measurement, or a picture, or eyewitness testimony. Statistics uses data, lots of data, the more data the better. The number of samples is a term in almost every equation. And anyway, that’s what the law of large numbers says, the more data the better the results.

Reality – The number of samples you really need for a statistical analysis is contingent on how much resolution you want. Think of the resolving power of a telescope or a microscope, or the number of pixels in a computer image. The greater the resolution, the more detail you’ll see. It’s the same way with statistics (https://statswithcats.wordpress.com/2010/07/17/purrfect-resolution/).

What’s more important than the number of data points is the quality of the data points. In statistics, the quality of a set of data point is how well the data points represent the population from which they are drawn. But representative data can be incredibly difficult to generate. How do you decide which registered voters are actually likely to vote in the next election? How do you decide who might use a product you might want to sell?

The number of samples is easy to determine. The quality of the samples is virtually impossible to determine. Nevertheless, what you should remember is that more data may be better but better data are always best.
**MISCONCEPTION 3: “DATA ARE DEPENDABLE”**

In Stats 101, you do a lot of number crunching. You use small datasets and big datasets, real data and fake data, but never were you told to delete data. You figured that data are like facts. You don’t delete them for any reason or you will bias your results.

*Reality* – Data are messy. Most newly generated datasets have errors, missing observations, and unrepresentative samples. Some population properties may be under-represented or over-represented. There may be samples that should not be included in the analysis, like replicates, QA samples, and metadata. All these problems with data require a lot of processing before an analysis can begin ([https://statswithcats.wordpress.com/2010/10/17/the-data-scrub-3/](https://statswithcats.wordpress.com/2010/10/17/the-data-scrub-3/)). In fact, data scrubbing often consumes the majority of a project budget and schedule, but you have to do it anyway.

**MISCONCEPTION 4: “STATISTICS PROVIDES UNIQUE SOLUTIONS”**

In all the problems your Stats 101 instructor solved in class, and all the homework assignments you did, and all the exams you took, there was only one “right answer” to a question. So, any statistical analysis should provide the same results no matter who does it.

*Reality* – Even if two statisticians start with identical data sets, they may not come to identical results, and sometimes, even identical conclusions. This is because they may make different assumptions and scrub the data differently. Furthermore, there may be more than one way, even many ways, to approach a problem ([https://statswithcats.wordpress.com/2010/08/22/the-five-pursuits-you-meet-in-statistics/](https://statswithcats.wordpress.com/2010/08/22/the-five-pursuits-you-meet-in-statistics/)). There may also be different statistical analysis techniques that can be used, or even different options within the same technique ([https://statswithcats.wordpress.com/2010/08/27/the-right-tool-for-the-job/](https://statswithcats.wordpress.com/2010/08/27/the-right-tool-for-the-job/)). It would probably be more surprising for two statisticians to calculate the same results from a dataset than for them to have some differences. Just like most problems in the real world, there may have more than one right answer from a statistical analysis.

**MISCONCEPTION 5: “STATISTICS PROVIDES UNAMBIGUOUS RESULTS”**

Results are either significant or they’re not. That’s pretty unambiguous.
**Reality** – Statistical results are based on data and assumptions about the data. Change the number of samples and you change the resolution of the statistical procedure. Change the data or the assumptions and you change the estimates of variability. Change the resolution or the estimates of variability and you have different results. There is indeed uncertainty in uncertainty. Sometimes uncertainty brings with it ambiguity. Is there really a difference between Type I error rates of 0.049 and 0.051? Many decision makers who never got past Stats 101 think so. But interpretations of these results are based on the assumptions and biases a statistician brings with him. One statistician might take a firm stance and say “significant” and another might say, “maybe not.” Results have uncertainty; interpretations have ambiguity, and decisions have risks. That’s statistics.

**Misconception 6: “It’s Easy to Lie with Statistics”**


**Reality** – It’s hard to do statistics right but it’s even harder to do them wrong in a particular way on purpose. You have to collect data, crunch the numbers, and cook up your story, or perhaps more correctly, cook up your story, make up the data, and call the press conference. If you just want to mislead an audience, it’s much easier to use made up facts, phony anecdotes, and illogical conjectures. So why do so many people, particularly politicians and biased media sites, even bother lying with statistics? It’s because numbers provide credibility. If you have little credibility yourself, using numbers can confer the illusion of expertise. And that is why people use statistics in the first place. Those authors aren’t saying that it’s easy to lie with statistics, they’re saying that it’s easy for you to get fooled if you don’t understand statistics. Big difference.
CONSUMER GUIDE TO STATISTICS 101

Whether you took or are taking an introductory course on statistics, you probably didn’t get to choose from a dozen candidate offerings. You had to take the specific course required for your major. You can, though, evaluate what you got. Did you get your money’s worth from that introduction to statistics class? Here are a few things to think about.

YOUR EXPECTATIONS

Why did you take Statistics 101? Was it a requirement for a degree? Many majors, especially for advanced degrees, require some statistics (https://statswithcats.wordpress.com/2010/06/08/why-do-i-have-to-take-statistics/). Was it a less frightening alternative to other courses? Statistics can be used as a substitute for calculus in some undergraduate programs and for a foreign language in some Ph.D. programs. Or, maybe you didn’t have any expectations other than to learn something new.

I took statistics thirty-five years ago when I was majoring in geology, which doesn’t have a lot of quantitative deterministic theories for describing natural processes. Even today, predicting earthquakes, landslides, volcanic eruptions and other earth phenomena remain elusive goals. I wanted to learn how to develop mathematical equations, models, to explain and predict phenomenon. Regression analysis turned on the light bulb over my head. That’s how I got here.

When you buy an expensive product at a store, you usually have some expectations of what you should get for your money. When you buy a new car, for instance, you may want it to look and handle a certain way. You may not voice your expectations, or even be able to describe what you want, but you do have expectations. Your Stats 101 course is similar. You paid a lot of tuition to take the course; you must have had some expectation, even subconscious, of what you would take from it. This is important because it sets a reference point for what you experience in the course. So ask yourself, did your course give you what you expected, and just as important, were your expectations reasonable?

YOUR INSTRUCTOR

Think of the person who taught you in Statistics 101. How would you rate him or her on these four criteria?

Knowledge—Knowledge may be the first thing you think about when you think of a college professor, and for that reason, it’s probably the least important discriminator of instructor quality.
They all have adequate knowledge, at least from your level of understanding. It’s what the instructors do with their knowledge that makes the differences.

**Communication Skills**—Being able to convey knowledge is a necessity for an instructor. Some instructors communicate information better than others, and unfortunately, some instructors do not communicate well at all. They may be inarticulate, have an accent or a speech defect, be speaking in a second language, or just not be able to explain difficult concepts or answer questions well.

**Engagement**—Instructors usually teach the same courses over and over again. Some instructors add content and try new descriptions with each class. Others use the repetition to become automatons, teaching the same content in the same way year after year, even to the point of reading their past lectures.

**Empathy**—Instructors have an obligation to teach certain content but they also should be sensitive to what their students want to learn and need to learn to further their careers. Empathetic instructors might try to tailor their teaching to the interests of their students, like citing examples from the disciplines of their majors. Oblivious instructors will teach about their latest interests, regardless of the applicability to the students.

The minimum requirements for an instructor are to know the subject and be able to communicate that knowledge. What separates the best instructors from the rest are their level of engagement and their empathy to the needs of the student. So ask yourself, did your instructor convey his excitement over what he was teaching? Did you leave the class curious about what else there might be to learn about statistics and about how you could use statistics yourself?

**Number Crunching**

Artists draw, chefs cook, and statisticians calculate, but they do so in many ways. When I was learning statistics, my choices for doing calculations were a very unfriendly mainframe computer, a hand calculator, or pencil and paper. This choice may be why there are so few old statisticians around today.

How your instructor had you calculate statistics says something not only about the times but also about his level of empathy. Here’s why:

**Pencil and paper**—No professional statistician calculates any serious statistics nowadays by hand, except perhaps on drink-stained cocktail napkins. Still, many instructors want their students to get the hands on feel of number manipulation. That’s valid. If it goes beyond probabilities, descriptive statistics, and simple tests of hypotheses, your instructor is a sadist.

**Calculator**—No professional statistician calculates any serious statistics nowadays with a calculator, unless cocktail napkins are not available. If you plan on being an artist or a chef, manual calculations are fine. If you have any intentions of using statistics in your major, you need to learn software.
**Spreadsheet Software**—Spreadsheet software is probably the best choice for most students. It can be used to set up and edit datasets, calculate statistics, and prepare graphs. Plus, it’s relative easy to use, and likely to be available to the student at school, work, and home.

**Statistical Software**—Statistical software can be another good choice, depending on the learning curve. Simple statistical software can allow students to concentrate on interpreting statistics instead of calculating them. Advanced statistical packages like SAS and SPSS, while necessary for advanced courses, are beyond what introductory students need to learn unless they plan a career in statistics. Because of the cost of these packages, they are not likely to be available to the student at work and at home.

**Programming**—If you plan on a career in statistics, you will probably learn the R language or some other coding tool. If you learn it in Stat 101 and you are not a statistics major, you have to wonder what your instructor is thinking, if he is.

So, was your instructor thinking about your needs when he decided how the class would do calculations? If you can’t use his method of choice in the future, it’s kind of a wasted effort.

**CONCEPTS VS. SKILLS VS. THINKING**

In designing your Stats 101 course, the instructor had to decide how to proportion class time between teaching concepts, skills, and statistical, thinking.

Concepts are the whys of statistics. They are the reasons why statistics work. Examples include populations, probability, the law of large numbers, and the central limit theorem. Instructors tend to devise games and demonstrations to help students remember fundamental concepts. Learning statistical concepts is beneficial to statistics majors and non-majors, both in school and in later life.

Skills are the whats and hows of statistics. They involve calculations, like probabilities, descriptive statistics, and simple tests of hypotheses. Skills are learned by repetition. You learn them by doing the calculations in the homework assignments, at least the even-numbered problems. After Stats 101, skills like designing data matrices for a particular analysis are much more important than the calculations themselves, which are usually carried out by software.

Concepts and skills account for the majority of Stat 101 classes. This is perhaps unfortunate, for the greatest need in society is for people to understand statistical thinking. Statistical thinking involves understanding how to define a problem in light of
some objective, what uncertainty and risk are and how they can be controlled, and the difference between significance and meaningfulness.

So, did the things you learned in Statistics 101 mostly involve concepts, skills, or statistical thinking? What things were you able to take from the class and use in later life?

**WHAT DO YOU THINK?**

Now all of this ignores course content. That’s a BIG topic for another time. For now, think about what your introduction to statistics course was like. Was what you expected? What would have made a better Stat 101 for you?

---

**Add somewhere??**

I have two concerns about using R for a class on introductory statistics, especially for non-majors.

First, statistics is intimidating enough, why add coding on top of it when there are GUIs available? Back in the 1970s, there were no alternatives. Even SAS and SPSS required you to write code. But today, there are GUIs for everything, even R and many programming languages.

Consider the analogy of automobile transmissions. Most people learn to drive on cars with automatic transmissions because they are easier. There’s so much other stuff a novice driver has to learn, learning how to shif gears can be deferred. Automatic transmissions are more expensive to buy and maintain, and don’t perform as well as manual transmissions. Still, most consumers (http://www.metafilter.com/52324/The-decline-of-manual-transmission-cars) buy cars with automatic transmissions, while professionals (truck drivers, NASCAR) use the more efficient manual transmissions.

Is R better than software like Excel? I think it depends on the person. You can make your data jump through R hoops. I like excel and those automatic-transmission statistical packages because I can do everything I need to do with them. They are set up as matrices, so it’s easier for me to visualize, design, and edit data. Furthermore, if I can’t present the results in Excel, my audience at work isn’t likely to understand it. At the same time, I find Excel’s capabilities to configure and annotate basic graphs to be unsurpassed by any statistical software. If I need a Chernoff face, OK, then I have to use Statistica. But the issue isn’t you or me, it’s what would be best for students.

That brings me to my second point. In the business world, workers collaborate on many projects through the use of electronic files. Every business owns some brand of spreadsheet software and every business professional has some level of expertise with spreadsheets. Wouldn’t it be better to polish a student’s ability with something they’ll almost certainly use in the future?
NON-TRADITIONAL CONTENT IN STATS 101

Gap analysis,
Whatever your instructor is working on
If it’s not in an intro book