

Why Should We Even Teach Statistics? A Bayesian Perspective¹

Gudmund R. Iversen
Department of Mathematics and Statistics, Swarthmore College
500 College Avenue, Swarthmore, PA 19081, USA
iversen@swarthmore.edu

Abstract

Statistical methods have an impact on the results of any statistical study. We do not always realize that the statistical methods act in such a way as to create a construction of the world. We should therefore be more aware of the role of statistics in research, and the question is not so much about what we teach researchers but that we train them to be aware of the impact of the methods they use. This becomes particularly important in statistical inference where we have the choice between the classical, frequentist approach and the Bayesian approach. The two approaches create very different views of the world. The word probability carries with it a notion of uncertainty, and it is tempting to think that the uncertainty refers to parameters and not simply data.

1. Who Is Better Off?

The English weekly newsmagazine *The Economist* once showed Figure 1 in an article as part of a series on statistics:

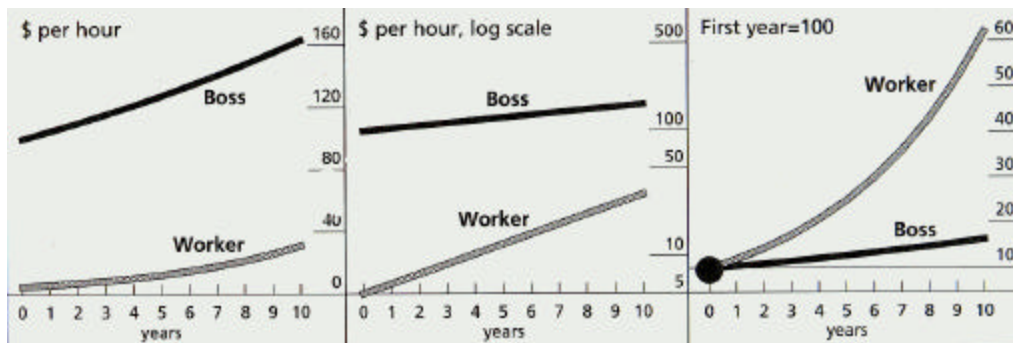


Figure 1. A comparison of wages for bosses and workers. (Source: *The Economist*, May 16, 1998, p. 79)

¹ IASE Round Table Conference on Training Researchers in the Use of Statistics, The Institute of Statistical Mathematics, Tokyo, 7-11 August, 2000 (preliminary paper, May, 2000).

The purpose of the graph was to make a comparison between the wages of Bosses and Workers. The comparison was made with time-series data over a ten-year span, and the graphs plot three aspects of wages against time.

2. Comparing Groups

As statisticians, we are very good at comparing groups. We typically translate the comparison of two groups into a comparison of two means or perhaps a comparison of two percentages. We can even compare two variances, if we have to. We are good at computing the proper test statistic and finding the proper p -value that will make us decide whether the difference between the groups is found to be statistically significant or not.

But are we doing the "right" thing by making such comparisons? What does it mean to say that two groups are different on a statistical variable? A researcher brings me data on two interval/ratio variables for observations in two groups and asks me to help her find out whether the groups are different. She would typically have had a statistics course or two, particularly if she is doing biological research. If she is a social scientist, we are on thinner ice. My immediate reaction is to do a t -test for the difference between two means, assuming no wild departure from normality and that the samples are large enough and that the underlying variances are not strikingly different. I will enter the data in a statistical software package of some kind and ask for the t -test. Based on the p -value returned by the software package, I will tell her whether there is a statistically significant difference between the two groups or not.

I am conditioned to do this from a long career in statistics, and hardly ever do I stop and consider whether I have done the right thing. Does it make any sense to compare the means? My forcing the comparison of two groups into a comparison of two means implies that I have constructed a reality out there of the world for the researcher, whether she wanted it or not. Maybe our comparison of means forces the researchers down paths they have not intended to take. After all, there are many ways in which things can be different.

Different can mean that all the observed values in one group are larger than the observed values in the other group. Different can mean that some of the values in one group are larger than some of the values for the second group. Different can mean that some type of an average is larger in one group than the other group, be it the mean, the median, or whatever our favorite average may turn out to be. So, what is the meaning of "different" when it comes to a comparison between Workers and Bosses? What can we conclude about what that world out there really is like?

I must confess I enjoy taking this graph to class. First I probe the students to see what their thoughts are on the meaning of the statement that two groups are different. The end of the discussion usually consists of an agreement that two groups are different if their means are different. They sometimes go as far as saying that it could be that the two medians are different, since they remember something about skewed distributions. When I make the question more concrete and ask them to think of a comparison between wages for blue-collar worker and white-collar workers, they usually respond that the white-collar workers would be expected to have higher wages, and so the groups are different.

When I push the students, they propose that there is a list somewhere containing wages, and they base their answer on the existence of such data. The implication of their answer is that they base their thinking on the existence of a real fact out there, in the "real world." There are wages out there in the world, and different groups have different wages. That is a fact about the world.

The students think they know what they mean by the study of the difference between two groups, but do they really? Are there facts out there, waiting for statistics to be discovered? I venture to say that the same discussion in this group conferees would not have been very different; perhaps more sophisticated, but in the end not very different.

But when I present the graph about the wages for Workers and Bosses, they are no longer so certain about the factual world. There are three different graphs displaying the same factual world, but the conclusions from the three graphs are very different. According to the first graph, on the left, the wages in dollars per hour are plotted as a dependent variable against time as an independent variable. For ease of comparisons the points for each group have been connected by curves, and we are left with two curves, one for the Bosses and one for the Workers. The top hourly pay is \$160, and I would not have minded that wage myself!

The first graph shows the curve for the Bosses to be considerably higher than the curve for the Workers across the ten years, and from that we draw the conclusion that the Bosses are better off than the Workers. Is that is the way the world is? Is that a fact we have uncovered about the world? Or is this maybe simply a construction of the world we have created and are now, as statisticians, forcing on the researcher and thereby on those who read the research report? And does the researcher recognize that we, the statisticians, have added something to the research finding? The result is not just the data speaking, it is also a particular way of displaying the data that is speaking.

Turning to the middle graph, we see that the dependent variable has been changed. Instead of actual wages per hour, the wages have been transformed into logarithms. This bothers the students right away. They say they do not live on logarithms of money, they spend real dollars and cents. Here in Tokyo we spend real yens, even though some of us may not be used to dealing with such high numbers for simple purchases. The student comments about logarithms may be a sign that as time has passed, not everything has become better. I used to think that such a statement was something old people made, but I actually grew up at a time when we used logarithms to perform multiplications and divisions. After some discussion, however, I usually get the students to accept that the use logarithms of wages makes it possible to see percentage increases over time, and that is what the middle chart shows.

The points for the years are again connected to give us two curves, one for the Workers and one for the Bosses. The curve for the Bosses still lies above the curve for the Workers, but now the curve for the Workers is rising faster than the curve for the Bosses. Somehow, the Workers are gaining on the Bosses. The Workers may not be as badly off, after all. Well, this is a different reality we are now using statistics to construct and paint for the researcher. What is the researcher to do? All of a sudden, maybe statistics is not as helpful as she thought it would be.

The third graph, on the right, again shows two curves. But this time the curve for the Workers lies above the curve for the Bosses. Here, we thought we had shown that the Bosses are better off than the Workers! The graph shows annual wages, and setting both wages equal to 100 at the beginning of the time period compares them over time from a common base. The curve for the Workers now goes up much more rapidly than the one for the Bosses, meaning that the Workers have gained more than the Bosses have.

3. Our Construction Of The World

Each one of these graphs paints a different picture of the world. We use the same data in the three situations. Still, by using statistics, we have constructed three different realities. So, which is the "right picture"? The obvious answer is that none of them is the right picture; it all depends upon how we look at it. This may not be what the researcher wants to learn. And this may be where we have not taught our students, turning into researchers, that *it all depends*.

Lawyers are used to thinking this way. A case before a judge and a jury is not so much about what is true and false, and is there a real world out there? Instead, it is getting a client acquitted, in case of the defense, and getting the accused to be found guilty, by the prosecutor. From time to time both sides even invite statisticians to appear as expert witnesses. It is always amazing to see how two statisticians can use the very same data to arrive at very different conclusions. Statisticians on the two sides of the case construct two very different realities of what the world is like and hope that the jury will accept their constructs.

In answer to the question of what we should teach researchers, maybe we should include in our courses a visit to a courtroom and watch statisticians in action. That will very quickly make anyone realize that there is not one, factual world, for us to discover. It is not that researchers should learn certain, specific statistical methods, and after they do, all is well. Instead, it all depends. Maybe that is the one thing researchers should learn from us: It all depends, and we should not worry about whether they know time series analysis or incomplete two-way analysis of variance or whatever else we teach.

4. Liberal Arts Education

At this point, let me take a small detour into the American system of higher education. Perhaps the greatest contribution to higher learning consists of the American system of liberal arts education. For four years, after high school, students attending a liberal arts curriculum, are not expected to learn a profession, but to immerse themselves in a liberal arts way of approaching life. I teach at such a liberal arts college, and my task is not so much to teach specific statistical methods as it is to convey a way of thinking to the students, based on randomness, variation and statistical regularities. The hope is that such an approach will make the students better citizens. If they need actual training in statistics, they will get that through their graduate studies after college.

5. Effects Of Data And Method

Therefore, in my statistics courses, I go on to discuss the following graph (Figure 2):

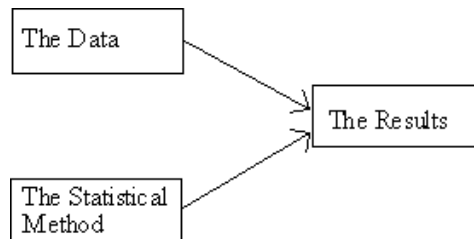


Figure 2. Schematic view of the research process.

The graph illustrates how the results obtained from a particular research project come from two sources. One source is the data, obviously, and the other is the statistical method. This little figure always surprises my students. They like to think that results from the research process somehow are “The Truth” about the topic being studied. The purpose of my teaching these students is not to make them into researchers, being able to do empirical research. Instead, the purpose is for them to be able to understand the role played by statistics in today's society. Hopefully, this graph helps them better to understand the role of statistics. Having seen the three sets of curves from *The Economist*, they begin to appreciate the role played by statistical methods, beyond the data themselves.

We are used to thinking that the data affects the results, and we teach about the presence of sampling variation. This aspect of the research makes sense to the students. What they find discouraging and surprising is that statistics itself somehow can have an effect on the results. When they learn that *the* correlation between two variables equals 0.87, they like to take this as a fact about the relationship between the two variables in the same way as, in physics, a metal has a specific gravity constant. The students get disappointed when I stress that this correlation is the number we get when we base the analysis on least squares. For example, had we used absolute values instead to fit a line, then the result would have been different. Any other measure of the strength of the relationship between two variables is similarly dependent on how it is defined. Again, the impact of the method shows its ugly presence.

So, how are we to look at look at the end result of a research process? What should we, as statisticians, impress upon researchers who make use of our methods? I think that we have the responsibility to stress that any statistical result from a research process represents a construction of the world created by our data and by our methods. Just like the three graphs from *The Economist*, there is no Truth out there in the world with a capital T. The results obtained from the empirical world consist of a *construct of the world* the researchers create.

I think we often forget about that. We are gathered here to discuss what researchers should know about statistics when they go about their tasks. We can make up a wish list of statistical methods that it would be nice if researchers knew how to use in their work. But our

work is not done by simply producing such a list. We have not executed our responsibilities if that is all we do.

6. Results Of The Research Process

We need to do more. We need to make people aware that the result of a research process is a particular construction of the world. This construction comes from the combination of our data and our statistical methods. One implication of this is that is not as important *what* we teach researchers as it is important that they recognize the full implication of using statistics.

It is tempting to think as a parent and ask what tools do we let our children work with as they grow up. We hesitate to let a young child play with a chain saw, and perhaps we should hesitate letting researchers have access to certain statistical methods unless they are fully prepared and ready for such usage.

One good thing about the pre-computer world of statistics was that statistical methods were not as accessible as they are today. It used to be, that to do a ten-variable regression analysis, one put it a good bit of thought about whether it was worth doing before employing several graduate students to compute sums of squares and sums of cross products and invert matrices.

Now this question does not even come up about whether a particular analysis is worth doing. With a few clicks of a computer mouse, the results are there for all to see in a matter of a very short time. Maybe this is not necessarily a good thing. Maybe we should require people to have a license before they are allowed to do multiple regression. I hate to think of all the many misuses that have taken place with such analyses because the tool is so readily available. In the wrong hands, multiple regression software may be as dangerous as the chain saw in the wrong hands.

Clearly, we should teach multiple regression, and clearly we should encourage researchers to use multiple regression. But I think we forget to let people know what they are getting into. I already alluded to the fact that an analysis based on least squares would give results different from an analysis based on absolute values. Just because historically it was computationally more appealing to use squares than absolute values, and the derivative of a square is easier to work with than the derivative of the absolute value function, we should not necessarily continue to use squares. But more than that, it is important that we tell researchers what they are getting into by using something like regression analysis.

We do not tell researchers that their results are partly a consequence of the methods we have taught them. We let people think that the sums of squares they get are *the* measures of the impacts of different variables. We also let people think that *the* p -value they get for a coefficient is a measure of the degree to which we can reject a null hypothesis.

7. Descriptive And Inferential Statistics.

All empirical fields need some way of simplifying their data, and let us distinguish for a moment between descriptive and inferential statistics. Descriptive statistics/exploratory data

analysis is not as controversial as inferential statistics. But both impose something on the research process, and this is not always fully acknowledged. The impact of descriptive statistics is less controversial, and I will limit myself to a discussion of the impact of inferential statistics. Here, of course, we have the major choice between classical and Bayesian statistics. More specifically, my point in this discussion is that through the uses of inferential statistics, we as statisticians impose a major construction of the world based on the results of empirical research.

Statistics is a funny field. Statisticians teach it, but mathematicians, economists, biologists and many other folks also teach it. Statisticians have completely lost control of their field, if they ever had such control, and they have let other people step into the vacuum left by statisticians. I wonder what my sociology department would say if my own department of mathematics and statistics started teaching sociology just because some of us had a couple of courses in sociology in years past.

The goal of most statistics training seems to be to make the students into amateur statisticians who can tell when to use a t -test and when to use a chi-square test, but not have any understanding of why they are doing what they are doing. It seems to me that the danger of doing this is so great that perhaps we should not teach statistics to other than statisticians. Through our inferential methods we imposing a world view that the world perhaps is not ready for.

8. Student Reaction To Inferential Statistics.

The two competing views, classical versus Bayesian, are very different in their approaches to statistical inference. Most discussions of which of the two to use have centered on preferences expressed by statisticians. We have not listened to our customers and heard anything about what view of the world they think should be imposed. Do we realize the heavy hand imposed by statistics, and even more importantly, do the users of statistics realize our heavy hands. And how is it that we have the right to tell the customers what is best for them? If we are to teach statistics, perhaps we should listen more to our customers/users of statistics.

At ICOTS V in Singapore in 1998, I gave a talk on some of these issues (Iversen, 1988). I said, here I am, a most excellent teacher of statistics who can explain the most obscure points with great clarity. I tell my students there will be a question about the interpretation of a specific confidence interval on the next test, and then I get these kinds of answers:

“95% of the intervals would fall between the two values of the parameter.”

“95% of the intervals will lie in this interval.”

“95% chance that the actual value will be contained within the confidence interval.”

“95% sure that μ is between the two numbers.”

“95% of the data will fall within this interval.”

These and other answers probably show that the students have not studied the right part of the book, and I as a teacher have not reached them in my treatment of the material in class, particularly if they did not attend class that crucial day when we talked about confidence intervals. But more than that, I see these answers as cries in the wilderness about how the world view we try to construct for our customers is not a world view our customers are comfortable with.

9. Role Of The Statistical Model.

What I have been discussing here is obviously the role of the statistical model in the research process. We were all taught that the model should be stated in terms of properties of the world we wanted to study. From the model we would then come up with the proper statistical method. Finally, we would examine in what way the data fit the model we had prescribed. I can still remember from my first course in statistics how the professor stated a few assumptions of how telephone calls could be expected to arrive. From that he derived the Poisson process. My trouble is that I cannot point to many such experiences.

How does the issue of a model relate to the construction of the world we come up with after the data have been collected and analyzed? I will argue that for many users of statistics, there is a big difference between the model being used and the construction of the world that arises as a consequence of the statistical analysis by the researcher. We may think that the two are the same, and we may want the two to be the same, but will I argue that they are often different. In the following, I discuss some of the differences between the two. It is in the difference between the two I think we find clues to how we should teach statistics to researchers, particularly statistical inference.

In its most elementary form, a model is something that simply stands for something else. It can be a physical item, such as a model toy car, or it can be a model expressed in some mathematical way. We like to think that the model represents the phenomenon we are studying, and by gathering and analyzing data, we will learn more about the model and thereby the world that it stands for. Statisticians are used to thinking that way, and because they understand the model, their construction of the world will resemble the model. We understand that finding a correlation coefficient of 0.87 implies a model of linear, least square analysis, with its strengths and weaknesses. But I do not believe that all users of statistics think this way. The users often construct their worldview in such a way that the 0.87 becomes a property of the world, as they see it.

How do nonstatisticians select and use statistical models? This raises the question of the origins of our models in relation to the uses of our statistical methods. I think many researchers use our methods with scant thought to why or why not they should use a certain method. Methods are most often used for their convenience and not because the underlying model in some way fits the problem at hand. The choice of model may not be the correct one, and the data may not even satisfy all the requisite assumptions needed for the use of the model.

All this is particularly true in statistical inference. The two major competitors represent very different models, and they construct their own views of reality. The frequentist and

Bayesian constructs are very different, and how does a researcher make a choice between the two? How does a researcher choose between the frequentist and the Bayesian construction of reality? Backing away from that question for a minute, how do researchers view and understand the reality constructed by each of these two approaches?

The attempts of my students to express the idea of a confidence interval in ways that make sense to them, shows a major display of the types of difficulties we are facing in our teaching of statistics to users of statistics. It seems to me that, for better or for worse, if we teach the students statistics, we should teach them methods such that their construct of the world is as close to our model as possible.

This argument takes us directly to Bayesian statistics. There, the model views the world as a random world, full of uncertainties. The best way to model this uncertainty is to use probabilities for what we are uncertain about, and this then takes us to prior and posterior distributions.

10. Misuses Of The Word *Probability*.

As a detour, maybe one reason why we have such difficulties teaching frequentist statistical inference, aside from the fact that it goes about drawing conclusions in strange ways, is our insistence on the use of the word *probability*. The word carries with it the notion of uncertainty, even though in the strict frequentist way it simply is a construct that tells us how often something happens. Suppose we try to explain a significance level by simply saying that it is a number which tells us how often we get the type of data that make us decide to reject a true null hypothesis. Instead, we shorten this and say that the significance level is the probability of rejecting a true null hypothesis. We understand the second way of expressing it, but do researchers understand it in their work? To them, the probability of rejecting the null hypothesis when it is true, becomes a mysterious way of drawing conclusions. Ask your nonstatistical friends and you will get many, strange answers. We cannot blame researchers for constructing a random view of the world since the probability word is so misused that way.

My students have no difficulties asking what is the probability that a population mean μ is larger than 100, say. They ask this question right after we start in on statistical inference and long before any mention has been made of Bayesian statistics. My job is to teach them that the question makes no sense. I typically succeed if I change the question to simply asking, *how often* is μ larger than 100? Then they answer always or never, depending on the value. Maybe we should banish our use of the word *probability* and substitute *how often*, instead, if we stay with the frequentist approach. Then, perhaps we can stay frequentists and still be honest with ourselves.

11. Bringing Model And World Construct Closer

So, where does all this leave us? I have argued that researchers construct their view of reality as a function of the data and the statistical method they use. They often overlook or misunderstand the role of the method, and the statistical model that gave rise to the particular method does not necessarily agree with the researchers' construction of the world. This is

particularly so in statistical inference. There, to bring the model and construct together, we owe it to the world to use Bayesian statistics. That way, we are permitted to deal with the uncertainty we have about population parameters.

This is particularly so for people with only a weak background in statistics. But even well trained researchers who use frequentist methods, very often interpret their results in a Bayesian probabilistic way. Even here, suppose we went around the table and tried to explain what a researcher has learned after rejecting a null hypothesis, we would find traces of Bayesianism. As long as we are uncertain about values of parameters, we will fall into the Bayesian camp. As statisticians, we owe it to researchers using statistics in their research to make clear the impact statistics has on their work and enable them to choose Bayesian methods. We should train researchers well enough to make it possible for them to understand the role Bayesian statistics can play in their work. That way, the worldview they construct may actually be a reflection of their models.

Acknowledgments: I am very grateful to Mary Gergen and Roberta Rehner Iversen for comments on an earlier draft of this paper. I also want to acknowledge my intellectual debt to L. Jimmie Savage. Having worked for him as a research assistant was a unique and very eye opening experience.

References

- Iversen, G. R. (1998). Student perceptions of Bayesian statistics. In L. Pereira-Mendoza, L. Seu Kea, , T. Wee Kee, & W. K. Wong (Eds.) (1998), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 231-238). Singapore: International Statistics Institute.
- Savage, L. J. (1981). *The Writings of Leonard Jimmie Savage. A Memorial Selection*, Washington D.C.: The American Statistical Association and The Institute of Mathematical Statistics.