Statistical Inference for the General Education Student—It Can Be Done

Allen H. Holmes
Walter Sanders
John LeDuc

Allen H. Holmes worked with UICSM, was assistant professor of mathematics at the University of Missouri, St. Louis, and is presently the Schilling Professor of Mathematics at the St. Paul Academy and Summit School. He is mainly concerned with vector geometry, probability and statistics, music and astronomy.

Walter J. Sanders has taught high school in the state of Washington and has been a faculty member at Chico State College, Western Washington State College and the University of Illinois. He is currently Associate Professor of Elementary Education and Director of the Mathematics Curriculum Laboratory at Indiana State University. His interests are in curricular development and teacher preparation.

John W. LeDuc has taught at Maine Township High School and Elgin Community College and is currently Associate Professor of Mathematics at Eastern Illinois University, Charleston, Illinois, where he has been since 1965. His particular interests are teacher preparation and curriculum development.

Many mathematics departments around the country offer an introductory statistics course for the general education student. Typically these students come to the mathematics classroom with minimal skills in arithmetic and algebra. In addition it is not unusual for these students to have very poor attitudes toward mathematics.

With this target population in mind one can design courses of study, called statistics, that will differ radically depending on what priorities are held. Many people choose to teach arithmetic through statistics and thereby build most of the course around descriptive statistics with some combinatorics. Others build most of the course around combinatorics and probabilities with some descriptive statistics. Few courses offered at this level spend much time or effort on statistical inference.

We believe that for the general education student the ideas of statistical inference and the resulting decision rules are of prime importance. This belief is based on the assumption that general education courses are included in the curriculum in order to help students to gain an understanding of their own essence, of their relationship to others, of the world around them, and of how man goes about knowing.

If you inspect most of the texts on the market today, you will find that they generally require that a student spend approximately a semester of study of descriptive statistics and probability theory before attempting statistical inference. This makes it very difficult to get to the general education portion of the subject in the time allotted most general education courses. If you agree with the analysis of the problem to this point the logical question is 'Is there a way to teach statistical inference without the traditional work in descriptive statistics and probability?'. The remainder of this article describes an approach that allows one to answer this question with a yes.

The approach is best understood by looking at it in three parts. The first part involves very concrete experiences with sampling and an intuitive introduction to confidence intervals. The second part develops the concepts of probability via Monte Carlo Simulations and applies these probabilities to statistical decisions and inferences. The third part takes the student on a quick look at traditional statistical treatments.

Description of Part 1

This part of the course is designed to help students get an intuitive grasp of statistical inference through carrying out a variety of laboratory experiments. In order to become familiar with the reliability of sampling estimates of population parameters, the students generate sample distributions using a sampling paddle and colored wooden balls. A sampling paddle is a wooden paddle with holes sunk into one side so that when it is drawn through a population of wooden balls, balls roll into the holes (see Figure 1.)[1] Upon withdrawal the paddle has a sample of the population.

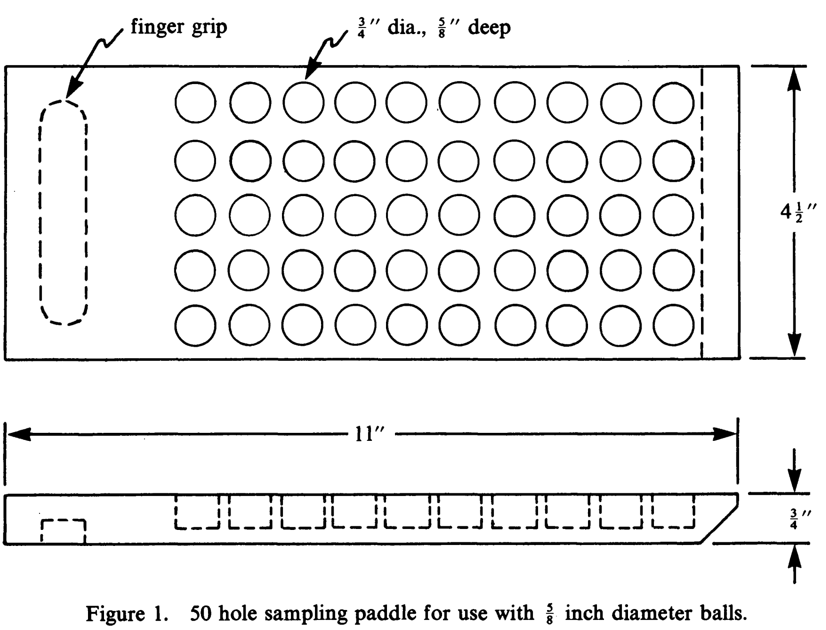

Figure 1. 50 hole sampling paddle for use with $\frac{5}{8}$ inch diameter balls.

[1] Professor Holbrook Working is credited with having popularized the sampling paddle and box of wooden balls as a fast, efficient method for generating sample distributions.

It is an easy matter, for example, to simulate a population of 2,000 light bulbs, 4% of which are defective. Assuming the balls have been sorted by color, the 50 hole and 10 hole paddles can be used to quickly scoop up 80 red balls (4% of 2000) and 1920 white balls. With the population mixed and placed in a cardboard box, and using the 50 hole paddle, a team of three can generate a sample distribution of the "defectives" (number of red balls in each paddle full) for one hundred samples in about fifteen minutes. Thus, two forty-five minute laboratory sessions are more than adequate for six teams to generate sample distributions of one hundred samples each for twenty populations.

Sample distributions are prepared for 1%, 2%, etc., up to 15% populations. Figure 2 shows three typical distributions. One hundred samples are taken for each population so that the per cent of samples with a given number of defectives can be readily determined.

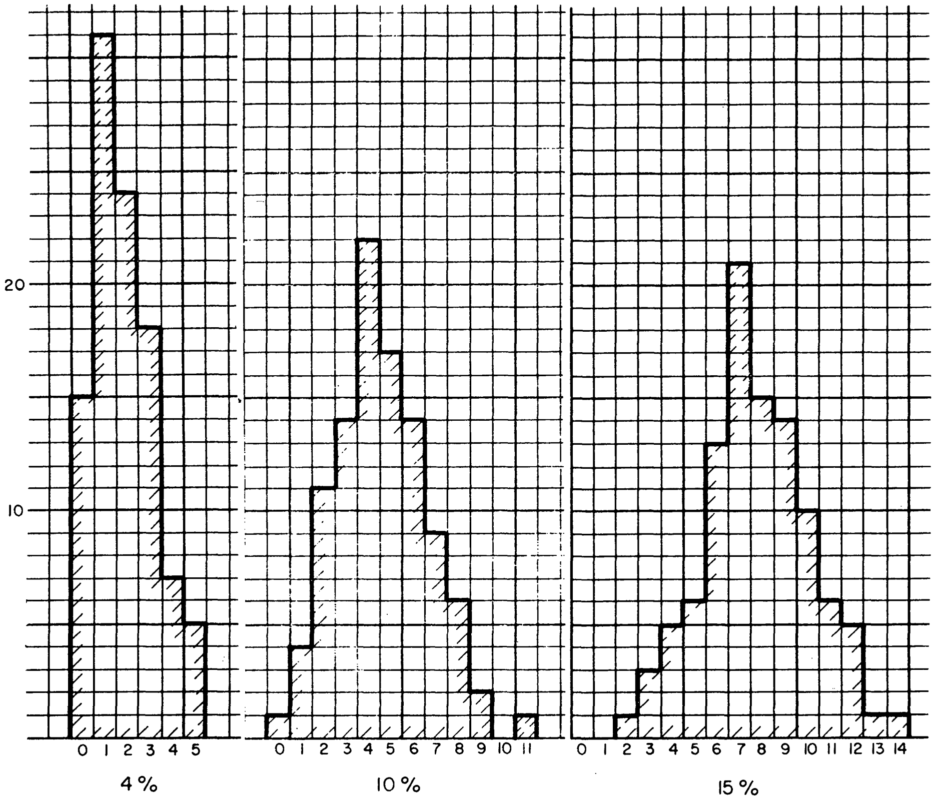

Figure 2. Three sample distributions generated using samples of size 50 from a population of 2000.

These sample distributions exhibit a degree of variability readily noticed by the students. But, common sense usually leads students to believe that samples of 50 from a population with $n\%$ defectives should have $n/2$ defectives more often than other numbers of defectives. And even though "runs" on a particular number occur, continued sampling should result in a smooth, bell-shaped (skewed) curve.

Seven percent of the samples in the 4% population had four defectives, while for the 10% and 15% populations, four defectives per sample occurred 22% and 5% of the time, respectively. One would estimate that the probability of drawing a sample with four defectives from a 4% population is about .07, about .22 for a 10% population, and about .05 for a 15% population.

The curve in Figure 3 was obtained by plotting the probability of drawing a sample with four defectives for each of the sample distributions generated for 1% through 15% populations. For each population between 5.2% and 14.2% the probability of drawing a sample with four defectives is .10 or more.

We are now ready to solve a problem. Suppose a sample of 50 is drawn from a population of 2000 with an unknown percent of defectives. The sample contains 4 defectives. What reasonable estimates can be made of the unknown population? The probability of drawing a sample of 4 defectives is at least .10 for any population with from 5.2% to 14.2% defectives. We can estimate the probability that the unknown population is one of the populations in this interval by comparing the areas under the curve as shown in Figure 4. Since 86% of the area under the curve is in the interval from 5.2% to 14.2%, the probability that the unknown population is one of those in this interval is about .86. We say that the *assurance level* for this interval is 86%.

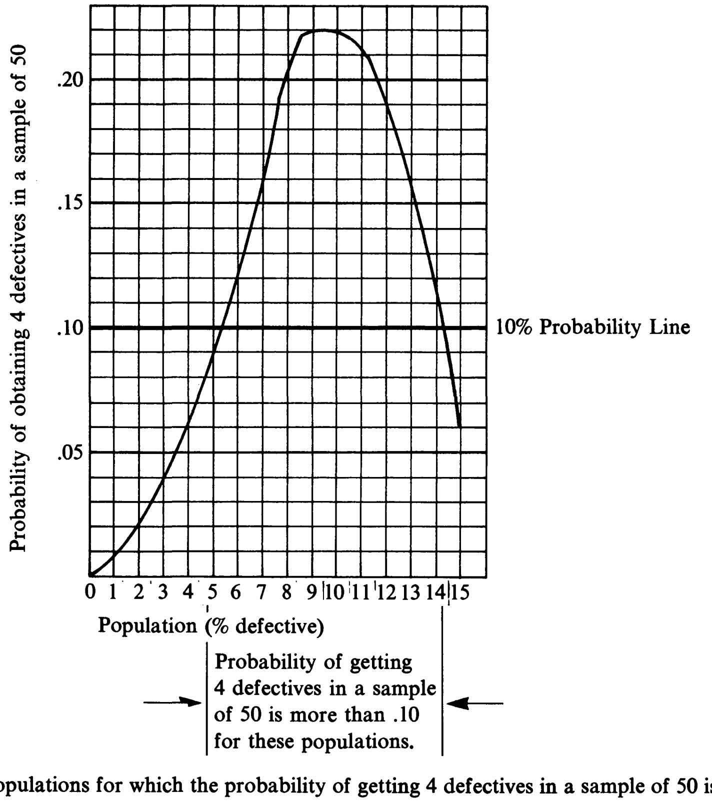

Figure 3. Populations for which the probability of getting 4 defectives in a sample of 50 is greater than .10.

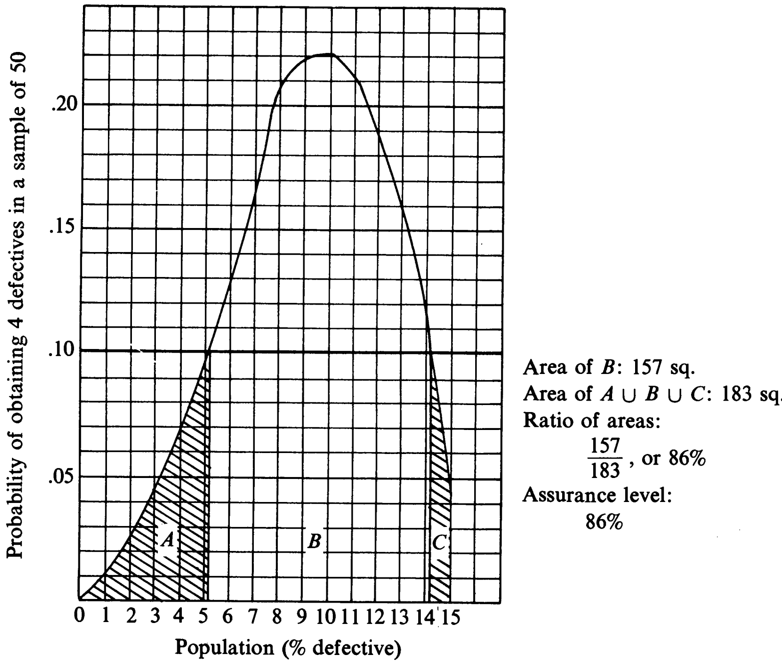

Figure 4. Calculation of the probability that a sample of 50 with 4 defectives was drawn from a population with between 5.2% and 14.2% defectives.

A larger interval of populations will determine a larger portion of the area under the curve in Figure 3, and hence give a greater assurance level, while a shorter interval will be more restrictive and give a lower assurance level.

The assurance level can be tested empirically by sampling a large number of populations at the, say, 85% level, and seeing how often the population lies in the interval determined by the sample. When this has been done by students the results support the theory.

The sample distributions generated can be used to solve a variety of problems, especially those where a test statistic is established before a sample is drawn. Also, the background developed paves the way for solving problems using sample distributions other than those produced through the sampling paddle.

A more thorough discussion of the method and further examples can be found in Sanders, (2).

The transition from part one to part two is accomplished by simulating some of the concrete sampling experiences of part one through use of a table of random numbers and by simulating some coin flipping through use of random numbers. (Usually the class can suggest how the simulations can be accomplished.)

Description of Part II

Frequently a decision must be made on the basis of experimental data. Consider the plight of the golf coach who must choose between two golfers for a college golf match. The coach has the following data from the previous season. (In golf, holes are given *par*; 4, 5, or 3 depending upon length and difficulty. A birdie is one under par, a bogey is one over, and a double bogey is two over.)

Result	Birdie	Par	Bogey	Double Bogey
Golfer X Pct.	.20	.50	.20	.10
Golfer Y Pct.	.05	.75	.15	.05

Which golfer should he choose?

Given enough time, he could have the golfers play each other 100 times and choose the one who won the most matches. He could have them play once, but that might seem inconclusive to him. Unfortunately, the weather is so bad, he can't even get them out to the golf course. His method of choosing must be done by *simulation*.

Simulation can be carried out using any random process. Coins or dice are frequently employed in table games. Spinners are also useful. Most efficient however, is a table of random numbers (or better yet, a computer terminal with access to a random number generator). Essentially, a table of random numbers consists of digits which have been tested for randomness. The digits may be selected two at a time, four at a time, or any other combination. Choosing two digits at a time gives 100 possible combinations—from 01 to 00. To simulate the play of golfer X, consider the matching:

Random Number	Result
01 to 20 (20)%	Birdie
21 to 70 (50%)	Par
71 to 90 (20%)	Bogey
91 to 00 (10%)	Double Bogey

To play a round of golf, select the course first, such as:

Hole	Par	Hole	Par
1	4	10	5
2	4	11	3
3	3	12	4
4	4	13	4
5	4	14	4
6	3	15	3
7	4	16	4
8	5	17	4
9	5	18	5

Then, for each of the 18 holes, select a two digit random number, translate this to a score for golfer X, record it, and get his total score for the 18 holes. (A computer can be programmed to do this quite easily.)

To finish the simulation, the reader should establish an appropriate matching for golfer Y and actually use a table of random numbers to play a round of golf. Who won? Who would you choose in the upcoming golf match?

You should notice that the data were used to make a decision. Simulation was used to help interpret the meaning of the data. Random numbers were used to establish the simulation. These same ideas may now be applied to a more sophisticated idea.

Frequently a researcher needs to compare two sets of test scores. One such comparison occurs when a pretest and postest are given to the same individuals with the hope of discerning significant improvement. A second comparison involves testing two separate groups with the idea of spotting a significant difference in treatments. We can treat each of these cases by carefully setting up our simulation.

Consider two classes with the following grade distributions:

Letter	A	B	C	D	E
Class X	5	12	26	5	2
Class Y	2	3	8	8	4

A question which might be asked here is: Do these data suggest that Class X is doing significantly better, or is the variation the result of a chance occurrence? Imagine that all things are equal except perhaps the teaching technique. Specifically, the researcher is looking for support to a claim that something different is indeed going on

To find out if the difference is due to random fluctuation, suppose that both classes have the same distribution of grades. Use the data to figure out what this distribution is. Then simulate results for a class of 50 and a class of 25. See how often the divergence favors Class X as significantly as the given data does. If it happens rarely in the simulation, the assumption of both classes having the same grade distribution is probably incorrect. For short, there is a significant difference between the classes.

A problem remains: how do you measure the divergence? One such measure is the random variable, class average X minus class average Y. To actually compute such a measure, you might need to assign values to the letter grades—such as 5, 4, 3, 2, and 1 for the five grades in order. Using this scheme, the class average for X is 3.26, while that for Y is 2.64. The difference, 3.26–2.64, suggests that X is better. So would larger differences. Use simulation to estimate the probability that the difference is greater than or equal to the observed value, 0.62, assuming both classes have the same grade distribution. If the probability is small, doubt is cast on the assumption.

As a last example, consider a pretest and a post-test given to the same individuals. Here are the results:

Individual	1	2	3	4	5	6	7	8	9	10
Pretest	B	C	D	B	C	D	E	C	D	E
Post-test	A	C	D	A	A	B	C	D	B	C
Difference	+1	0	0	+1	+2	+2	+2	−1	+2	+2

Since the total difference is +11, there seems to be a significant improvement.

As before, we need to find out if the difference is due merely to random fluctuation. We assume, to be contrary, that there is no improvement to be expected. We may then view the 20 letter grades as values of a single distribution of grades for which the following data has been observed:

Letter Grade	A	B	C	D	E
Number of Occurrences	3	4	6	5	2
Percent	15	20	30	25	10
Cumulative Percent	15	35	65	90	100

We may then set up a random number matching:

Random Number	Letter Grade
01–15	A
16–35	B
36–65	C
66–90	D
91–00	E

(Notice the ease of using cumulative percentages to set up the random number matching.)

Now, carry out several simulated trials, each one requiring 20 random numbers. For each trial, compute the total difference. Keep track of how many times it exceeds $+11$. If this happens rarely, the assumption is probably wrong, suggesting that significant improvement did occur. If it happens frequently, the assumption is probably correct.

We leave to the reader the problem of actually reaching his own conclusions in each of these experiments. In addition, he may be able to apply these ideas to his own every day problems of decision making.

The transition from part two to part three is relatively easy since the last problems studied in part two are very time consuming to actually carry out simulations for, but easy to use a X^2 table on. The rest of part three involves teaching students how to use many of the tables in a typical text. This includes tables for confidence intervals for binomial distributions (which takes them back to part one), and tables of binomial probabilities (where they go back and get new answers to many of the questions of part two).

This approach has many obvious advantages over traditional treatments. It is based on the original work of Sanders (2) and Holmes (1). It should be pointed out that there are some unusual difficulties in this approach to statistics. In part two one trades traditional weakness in arithmetic and algebra for deficiencies in writing since the write-ups of the simulations demand clear and logical exposition on the part of the student. However, if you feel that the importance of 'statistics for the general education student' lies in the areas of inference and decision rules, then you should try this approach. You will like it.

REFERENCES

Allen H. Holmes, Teaching the Logic of Statistical Analysis by the Monte Carlo Approach, Ph.D. Dissertation, University of Illinois, 1968.

Walter Sanders, Teaching Statistical Decision-Making to Junior High Students, Ph.D. Dissertation, University of Illinois, 1971.