



Curriculum Units by Fellows of the Yale-New Haven Teachers Institute
1986 Volume V: The Measurement of Adolescents, II

Size, Error, and Confidence in The Statistics Sampler

Curriculum Unit 86.05.06
by David B. Howell

In "The Statistics Sampler" (Bibliography No. 5), I developed a sequence of activities and lessons to introduce some basic ideas of sampling. The major mathematical underpinning for the lessons was the Central Limit Theorem. Except for counting, tallying, ratio, percent, graphing, and mean, however, NO formal mathematical skills or concepts were required of students. Intuitive ideas of random sampling, of the effects of changing sample size, and of confidence in predictions were the main outcomes expected of students.

In this Unit, I am extending the topic of sampling through activities which make more specific the concepts of, and relationships among, confidence limits, error tolerance, and sample size. Two related formulas are given so that the student, using a calculator, can apply the ideas, developed informally through the activities, to similar situations. It is assumed that the student has the pre-requisite arithmetic skills and the base experience in sampling such as was developed in "The Statistics Sampler." Also required is the ability to substitute values into a formula and to evaluate the result using a calculator. As in "The Statistics Sampler" I would expect the Unit to lie within the grasp and capability of most regular education students from Grade 7 through Grade 12.

Here is the kind of problem I want to solve:

1. Before I buy an ad for my new shoelaces on MTV, I want to know the percent of high school students who watch MTV more than one hour per week. Obviously I can't poll every high school student. In order, then, to predict the percent on the basis of a sample, I first need to know (a) How many high school students do I need in my sample?

(b) If I want my answer accurate to plus or minus 10%, how will that affect the sample size? How will that affect the confidence level?

(c) I want to be very confident about my answer. I can't afford the expense of being 100% confident, though. Maybe 95% "sure" is good enough. How will that affect sample size? How will that affect accuracy?

Here are two more similar problems:

2. Sandra "Dunk-em" Smith is trying to decide whether to run for Student Government president in her high school. It would help her decide if she knew approximately what percent of the over 2000 students in her school know who she is. Three of her friends are willing to take a poll of a random sample of students. To be

95% sure that they have predicted an answer within plus or minus 5%, how many students should they poll?

3. Willie B. Ready works at Awesome Auto Parts. He has noticed that some of the air filters from a particular supplier are ripped. He thinks there may be as many as 2 defective filters out of every 10. He wants to be quite sure about that, however, before he goes to the trouble of changing suppliers. He decides to take a random sample of the air filters received next month to get an approximation to how many are defective. He wants to be 95% confident the percent of defective filters he predicts is accurate to within plus or minus 3%. How many does he need to sample?

The objective of this Unit is for students to be able to answer questions like the three just posed.

The formal mathematics is not really terribly difficult. We're trying to predict P , the percent or proportion of a large population which has a certain characteristic. In the three problems above, the characteristics are (1) watches MTV more than one hour per week, (2) has heard of "Dunk-em" Smith, and (3) is a defective air filter. We will make the prediction of P based on the percent, p , of "successes" (a "success" is a member of the population having the characteristic) in the random sample. In other words, we'll get

$$P \in \{p \pm E\};$$

that is, we will predict P as being within the range of p plus or minus an error tolerance, E . The prediction is one in which we will have some specific degree of confidence. The degree of confidence will depend on E , how wide an error tolerance we will accept, and the size of our sample. The larger the sample, the higher our degree of confidence in the prediction. The wider the error tolerance, the higher our degree of confidence.

E , then, is related both to the degree of confidence and to the sample size. It is also related to the percent or proportion of successes in our sample.

$$E = z \sqrt{pq/n}$$

z = standard normal coefficient at the stated confidence level

p = proportion of successes in the sample of size n

q = $1 - p$ (the proportion of failures in the same sample)

n = number of members in the sample

The more general statistical formula is

(figure available in print form)

where μ = the population mean

\bar{X} = the sample mean

z_{α} = the standard normal coefficient corresponding to the confidence level α

σ (or s) = the standard deviation of the population, if known (or the standard deviation of the sample)

n = the size of the sample

Several references in the Bibliography, especially Devore and Peck (2), Edwards (3), and Mason (7), provide the details connecting the general case to the population proportions of binomial distributions.

Teachers should be aware of several issues I am ignoring here. The issues seem to me to cloud the fundamental concepts we want the students to grasp. Those fundamental concepts are : (1) We can predict the population parameters from sample values. (2) The size of the sample, the error tolerances, and the degree of confidence are interrelated. (3) Smaller error tolerances lower the degree of confidence OR increase the sample size. Higher degrees of confidence require larger error tolerances OR lower degrees of confidence. (4) We can, for a given predicting need, specify two of the values we have been discussing and calculate the third value. There are simple formulas we can use.

“...issues I am ignoring...” Yes. (1) Does the form of the population distribution matter? Maybe. The Central Limit Theorem, however, guarantees us that regardless of the distribution of the original sample, the distribution of sample means (which is what we’re basing prediction on) tends to the normal distribution as the sample size increases. Therefore the binomially distributed cases we are concerned with in this Unit can be treated as normal distributions. (2) Suppose the population is relatively small? Or the sample size is small? Or the ratio of sample size to population is relatively large? Or we do or do not know the population variance? Then we would need to worry about correction factors, t-tests and other hypothesis-testing tools, and whether or not the standard deviation is defined in terms of n or $n-1$. The application problems posed and attacked in this Unit do not get into such detail complications. If your class, or a student, wants to tackle such situations *after* the points in this Unit are understood, then by all means explore the potential problems and solutions. All of the references in the Bibliography agree that

if both $np \geq 5$ and $nq \geq 5$

then the simple procedures presented here are appropriate.

Each of the following Lessons should take from two to five class periods depending on the task efficiency in sampling, sophistication of discussion, the skill levels for charting, graphing, finding percent, rounding off, combining smaller samples into larger ones, etc. The Lesson outlines deal with content, not management or individual differences or testing. All of the Lessons are described in a format similar to that of “The Statistics Sampler:”

- A. Objective
- B. The experimental question — a question which involves statistical sampling
- C. Issues and some possible resolutions — a mini-lecture, a series of questions which should arise, a conversation between teacher and class. There is occasionally a direct comment to the teacher in brackets — or a lesson continued for illustration with *my* data. This section really defines the activity.
- D. Observations and discussion to Objective — more questions or mini-lecture or summary or dialogue relating specifically to the stated Objective or bridging to the next activity.

The lesson numbering continues from “The Statistics Sampler.”

Lesson 9.

- A. Objective—students will be able to describe relationships among sample size, error tolerance, and “confidence level” in a Table.
- B. The experimental question—What percent of the colored cubes in the box are red?
- C. Issues, and some possible resolutions —

Let’s review and extend the activity of Lesson 8 in “The Statistics Sampler.” We were concerned with the percent, or proportion, of the total population of Colsquar which was red (the residents were colored cubic centimeters). We analyzed samples, first, of size 10. Then we combined those into samples of 30, and then of 50. We formed three histograms as follows:

(figure available in print form)

(figure available in print form)

(figure available in print form)

We summarized the graphed data as follows:

Sample size	Range (percent of red)	Mean
10	0-40%	13%
30	0-23%	13%
50	6-22%	13%

Now, based on the information in the graphs, we are going to expand the detail in the Table.

*1 The graph shows a total of 32 cases.

*2 Three cases, those *beyond* the 20% column of the graph, do NOT lie between 0 and 26%.

Therefore, $32-3=29$ cases DO lie in the interval from 0 to 26%. Since the actual population mean is equal to the mean of all possible samples of a given size, we might feel, based on *our* samples, 91% confident that the population mean lies between 0 and 26% red. Or, looked at from the outside, only 3 cases or 9% of our samples, by chance, would cause us to predict smaller than 0 or larger than 26% red.

*3 Our sample size of 10 is too small to show differences between multiples of 10%

(figure available in print form)

D. Observations and discussion to Objective—

The expanded Table on the previous page shows clearly the inter-relationships among sample size, error tolerance, and “confidence level.” For sample size 10, 72% (or 23 of the 32 different samples we made) of our samples have “percent red” within $E = 8$ of the sample mean value (population value) are between $13 - 8 = 5$ and $13 + 8 = 21$ of the population value 13). For sample size 30, 72% of our samples have “percent red” within $E = 6$ of the population value. Or 82% are within $E = 8$. For sample size 50, 69% (almost 72%) of our samples have “percent red” within $E = 4$ of the population value. Or 100% within $E = 8$!

Let’s work with the language a bit. We said “For sample size 50, 69% of our samples have “percent red” within $E = 4$ of the population value.” Said another way, if we took just *one* sample of 50, there would be about a 69% chance that it would have a “percent red” value within $E = 4$ of the actual population value. Or another way, if we took just one sample of 50, we would be 69% confident that it would have a “percent red” value within $E = 4$ of the actual population value.

I can hear you now! “Hold it! Hold it!” you say. “If we take just *one* sample, we won’t know what the population value *is* . So what good does it do us to be 69% “confident” our value is within $E = 4$ of it?”

Well, look at it from the other side. If my value is within 4 of some other value, then isn’t the other value within 4 of mine? If the other value is 13 and my value is 16, we are within 4 of each other. If my value is 9, we are still within 4 of each other. If I get 16, I’ll simply say that the other value, the value I want to predict, is between $16 - 4 = 12$ and $16 + 4 = 20$. 13 qualifies, doesn’t it?! If I get 9, I’ll predict 5 to 13. 13 still qualifies! And if I am 69% confident my value is within 4 of the true one, then I am 69% confident the true value is within 4 of mine.

When we started this experiment with the little cubes, we pretend the cubes were residents of the planet Colsquar, and red residents (cubes) liked the red records we wanted to sell. We wanted to predict the percent of the population which was red. Now pretend something different. Pretend that the colored cubes are air filters. Red cubes are defective. Willie B. Ready of Awesome Auto Parts wants to predict what percent are defective with 95% confidence and error tolerance $E = 3$. How many filters does he need in his sample? He can get to $E = 3$ for a sample of 10, but only at the 44% confidence level. For a sample of 50, he can get to $E = 3$ with 69% confidence. Obviously, he’ll have to sample some number more than 50. But we don’t know, yet, a simple way to find that number.

Before we describe a simple way, however, let’s go through on more model experiment to be sure we have a good idea of this whole sampling process.

Lesson 10 .

- A. Objective — same as Lesson 9.
- B. The experimental questions — What percent of the population is beans? How close can i get? How confident am I?
- C. Issues, and some possible resolutions —

[Materials and procedure. A box of several thousand objects — two different kinds or colors of objects. I used dried beans and peas, less than two small packages in all, which just filled a one-quart container and, I estimated, approached 4000 objects. To mix the beans and peas thoroughly, I dumped them into a large container which I shook vigorously! Sampling was a bit less efficient than for the colored cubic centimeters; since beans and peas are different in size and shape, I couldn't count out the same number of objects for each sample without compromising randomness. I scooped out a level teaspoonful for each sample, getting generally 25 to 30 beans and peas. It became important, then, to have a calculator to find the "percent" of beans in each sample (and later, to find the totals of combined samples).

Theoretically, one should take successive samples *with* replacement to meet a condition of randomness. For such a large population, it shouldn't make a noticeable difference to sample without replacement, however, For the class that raises the issue, and if you have time, it might be worthwhile to try both methods to compare results.

The greatly reduced copies of Worksheets included here illustrate *my* results.

Use full-size Worksheets with the class! The particular combinations used to generate larger samples are, of course, not important.]

Here are several thousand beans and peas. We could use them as models of air filters, with beans (or peas) the defective ones. Or as high school students, with beans (or peas) students who know "Dunk'em" Smith. Or who watch MTV more than one hour per week. In any case, we want to take samples so we can predict the percent of the entire population which is beans.

On your Worksheet, write the Research Question, "What percent is beans?" And in parentheses, record your estimate (guess) right now based on looking at the top layer (these are well-mixed). We'll need to refer to your estimates later.

With the teaspoon, each of you take one sample. Count the total, and the number of beans, and record on your Worksheet. Then we'll list all of your samples on one set of Worksheets.

[Here are my Worksheets for 26 samples]

(figure available in print form)

From these 26 samples, where N averages about 27, what would you be willing to predict about the percent of the total which is beans? Less than 50%? More than 90%? Probably in the 60's or high 50's, or low 70's? How confident are you? What error tolerance will you accept?

Let's graph the data; perhaps it will be easier to see what's happening...

(figure available in print form)

PERCENT OF BEANS

After our experience with the cubes, we would expect that combining the small samples into larger ones would give us a clearer picture and a narrower range. Let's do that here, combining four small samples into new samples averaging about 110 in size.

[Here are *my* results.]

(figure available in print form)

And let's graph these on the same scale we had before.

(figure available in print form)

PERCENT OF BEANS

What are you willing to predict now?

Let's combine again — combining groups of five of the second set of samples into a new set of 26 samples averaging about 550 in size.

[Here are my results.]

(figure available in print form)

(figure available in print form)

And graphing as before, but with one modification since the percents cluster so tightly... Let's keep the same scale, but break each interval in half so we see each percent value.

(figure available in print form)

PERCENT OF BEANS

Now what are you willing to predict?

As we did in Lesson 9, we'll make a Table summarizing these results in terms of N, the percent E, and the "confidence level."

[Here are my results.]

(figure available in print form)

D. Observations and discussion to Objective —

The Table on the previous page makes it clear, for the samples I took, that with a sample size of 27, 85% of the samples lie within $E = 14$ of the population (sample mean) percent of 68. Or, to change the point of view again, 85% of the time the population percent will be within $E = 14$ of whatever my sample percent is. When the sample size is increased to 110, 85% of the time the population percent will be within $E = 14$ of whatever my sample percent is. When the sample size is increased to 110, 85% of the time the population percent will be within $E = 4$ of whatever my sample percent is! And when the sample size gets up to 550, 88% of the time I can predict within $E = 2$ of the population percent.

[You may want to view the “confidence level” in a more technically correct way from the error side. In the last case, for example, one would say that only 12% (100-88) of the time will I have a sample more than $E = 2$ off by chance. Or...the population percent is different from my sample value, say 67%, 67.2, only 12% of the time; therefore I have no reason to reject the hypothesis $P = 67.2$ at the 12% level. But I think such a degree of technicality requires a far more sophisticated and formal background in probability, the normal curve, and hypothesis testing than is appropriate at the level of these lessons and *than is necessary to establish the basic concepts* as we have been doing. For a similar, intuitive, non-technical approach using box and whisker plots for 90% confidence instead of the histogram and Table techniques here, refer to *Information from Samples* by Landwehr et al. (Bibliography reference 6).]

Lesson 11.

- A. Objective — students will be able to apply 2 given formulas to solve problems such as 1, 2, and 3 posed near the beginning of this Unit.
- B. The experimental question—see Problem 1: What percent of the high school population watches MTV more than one hour per week? How large a sample do I need to answer the question within a given error tolerance with 95% confidence?
- C. Issues, and some resolutions —

With the bean/pea population as a model and with some sampling, we essentially worked toward an answer to the experimental question by trial and error. We discovered that *all* our combinations of samples reaching about 550 in size would give us a predicted percent within $E = 3$. So we would be willing to claim 95% confidence! Or, for N approximately 110, our sample gave us a value within $E = 6$, 96% of the time.

I have chosen to concentrate on the 95% confidence level because it is a very common level used by experimenters and pollsters. Other levels sometimes used are 90%, 99%, and 99.9%. [These correspond, of course, to $\alpha = 0.05, 0.10, 0.01, \text{ and } 0.001$ in formal statistics.]

Here is a formula we can use to answer our MTV question:

We will predict that P , the percent of the population we want to know, is p , the percent of the population in our sample, plus or minus E , the error tolerance. In symbols, P is $p \pm E$. Or P is in the interval from $p - E$ to $p + E$. And we will make this prediction with 95% confidence. But how do we know what E is? Or the sample size, N ?

$E = 1.96$ times the square root of p times q divided by N

$$E = 1.96 \sqrt{pq/N}$$

E is the error tolerance.

1.96 is a factor mathematicians calculate from the 95% confidence level we said we'd use. [It is, of course, $z_{0.05}$]; If we wanted only 90% confidence, then the factor would be 1.65; if we wanted 99% confidence, the factor would be 2.58.)

p = the percent of what we want in our sample.

$q = 1 - p$ or the percent of everything else in our sample.

N = the size of our sample, the number of people or answers or objects in our sample.

Let's use this formula with our bean/pea population. For my particular sample B, we had $N = 26$, $p = 62\%$. Then we would predict

(figure available in print form)

P lies within 0.62 ± 1.96 times $0.095 = 0.62 \pm 0.19$.

P lies between $0.62-0.19$ and $0.62+0.19$ OR between 0.43 and 0.81 with 95% confidence.

When we took lots of samples of size about 27, we found P was about 68%, or 0.68. Is that between 0.43 and 0.81? Of course it is!

Let's try this for F. $p = 75\%$ or 0.75

Then $q = 0.25$

(figure available in print form)

P is within 0.75 ± 0.16

P is between 0.59 and 0.91 with 95% confidence.

Let's try it for sample A. $p = 88\%$ or 0.88

Then $q = 0.12$

(figure available in print form)

P is within 0.88 ± 0.11

P is between 0.77 and 0.99 with 95% confidence.

Did you say, “No, P was 0.68. That is NOT between 0.77 and 0.99.”? Well, we didn’t claim 100% confidence, did we?! 95% “confident” means 5% of the time we’re wrong! This was one of those cases where we were wrong!

Let’s try two more. Use samples of about $N = 110$.

For my sample 1, $p = 71\%$ or 0.71

Then $q = 0.29$

$N = 113$

(figure available in print form)

P is within 0.75 ± 0.08

P is between 0.67 and 0.83 with 95% confidence. Notice, since N is larger than before, how much smaller E is.

For my sample 8, $p = 65\%$ or 0.65

Then $q = 0.35$

$N = 111$

(figure available in print form)

P is within 0.65 ± 0.08

P is between 0.57 and 0.73 with 95% confidence.

Let’s go back to the beginning. We wanted to predict what percent of high school students watch more than one hour of MTV a week. We pretended the beans were those students and the beans and peas together were all high school students. Actually, we would conduct a survey, trying to pick students at random, couldn’t we. But how many students should we pick in our sample? There is a way to use the formula we’ve just worked with to answer the question. We’ll stick with the beans/peas model.

P, we said, was within $p \pm E$.

(figure available in print form)

$E = 1.96 \sqrt{pq/N}$. Suppose we decide our error tolerance in advance.

Then we can solve for N as long as we have a guess about p. [If your class can do the solution, do it. Otherwise simply present the following.]

(figure available in print form)

In words, N equals 1.96 divided by E, then squared or multiplied by itself, times p times p.

Remember when we guessed a percent, p, for beans way back at the beginning of Lesson 10? We’ll use that number for p now. And let’s agree we want $E = 0.06$ at the 95% confidence level.

(figure available in print form)

$N = 256$

A sample of 256 should do it.

Suppose we had guessed $p = 0.70$.

(figure available in print form)

$N = 224$. Close, but a little less than the 256 we had before.

Suppose we set $E = 0.10$, and guessed $p = 0.60$. Do you expect N to be larger or smaller? Why? Let's calculate N .

(figure available in print form)

$N = 92$. Did you expect N to be smaller because we made E larger?

Let's make $E = 0.04$, and keep our guess at $p = 0.60$. What do you expect will happen to N now?

(figure available in print form)

$N = 576$. Did you expect N to be larger because we made E smaller?

What do we do if we have no idea at all about what p might be?

The safest solution is to use $p = 0.50$. That will give the largest value of N for a given error tolerance.

D. Observations and discussion to objectives —

[Don't try the foregoing without calculators! And you may have to teach calculator use for the specific formulas, too!]

Here are two more problems. Let's try them to see how to summarize what we've learned. Recall problem 2. Sandra "Dunk-em" Smith may run for Student Government president. First, however, she wants an estimate of what percent of the students in her school know who she is. She'd like to have 95% confidence in a prediction within $E = 0.10$.

How many students should be polled?

(figure available in print form)

"Dunk-em" thinks 75% know who she is. Her campaign manager says to use 50% because it will give a "safer", larger number of students to sample. Try it both ways!

(figure available in print form)

$N = 72$

"Dunk-em" decides to play it safe. Her campaign workers poll a sample of 96 students. 62 of them know who

“Dunk-em” is. What is the prediction for the percent of all students?

(figure available in print form)

(figure available in print form)

(figure available in print form)

P is within 0.65 ± 0.10

P is between 0.55 and 0.75.

“Dunk-em” is now 95% confident that between 55% and 75% of the students at her school know who she is. Now she can decide whether to run for Student Government president. What would you decide?

[One approach to Willie B. Ready’s air filter problem (Problem 3 at the beginning):

For 95% confidence and $E = 0.03$ and a guess of $p = 0.20$, we get

(figure available in print form)

$N = 683$ air filters.

Willie figures it would take two months’s worth of air filters to get that many. So he changes his E to 0.10.

(figure available in print form)

$N = 61$ air filters.

He goes with it. He gets 5 defective ones. So he calculates

(figure available in print form)

P is within 0.08 ± 0.07

P is between 0.01 and 0.15.

Willie has estimated that between 1% and 15% of the air filters are defective. What would you do? Change suppliers? Warn the supplier that you will change if there is no improvement? Ignore it?

Statistics help us predict. But the important decisions we base on the predictions can not be made by the statistics. Human beings make those decisions!

The series of Lessons is concluded. Hopefully, students have met the objectives. The base of understanding in real problems, in concrete experience, should prepare the students both for a clearer understanding of general statistical data as well as for the further study of statistics.

(figure available in print form)

WORKSHEET 1

BIBLIOGRAPHY AND REFERENCES

1. Anderson, David R, Dennis J. Sweeney and Thomas A. Williams.

Statistics: Concepts and Applications . St. Paul, MN. West Publishing Company. 1986.

Chapter 10 bears most directly on this Unit. It will lead to other Chapters.

Readable text for college or senior high school. Wide variety of applications. Superb problems.

2. Devore, Jay and Roxy Peck. *Statistics: The Exploration and Analysis of Data*. St. Paul, MN. West Publishing Company. 1986.

Chapters 7 and 8 apply to this Unit. Slightly more “mathematical” than Anderson (1). Aimed at college. Good problems.

3. Edwards, Allen L., *Statistical Analysis* . New York. Holt, Rinehart and Winston, Inc. 1969.

Chapters 9 and 10 apply to this Unit. Readable and without heavy theory. Good examples.

4. Fehr, Howard F, Lucas N.H. Bunt and George Grossman. *An Introduction to Sets, Probability and Hypothesis Testing* . Boston, MA. D.C. Heath and Company. 1964.

Chapter 4 through 6 apply to this Unit. Good examples are cited. The theory is aimed at Grade 12 or college students.

5. Howell, David B. “The Statistics Sampler.” Unit 4 of *The Measurement of Adolescents* , Curriculum Units by Fellows of the Yale-New Haven Teachers Institute, 1985, Volume VIII.

Pre-requisite for both teacher and students to this Unit.

6. Landwehr, James M, Jim Swift and Ann E. Watkins. *Information from Samples* . A Booklet “prepared for the American Statistical Association—National Council of Teachers of Mathematics Joint Committee on the Curriculum in Statistics and Probability.” Development copyright by the authors. 1984.

An important, readable, doable unit on essentially the same content as this Unit, but from a graphical approach. Wonderful material.

7. Mason, Robert D. *Statistical Techniques in Business and Economics* . Homewood, IL. Richard D. Irwin, Inc. 1978.

Good applications. Chapter II takes a relatively informal approach to sampling and confidence limits.

8. McGhee, John W. *Introductory Statistics* . St. Paul, MN. West Publishing Company. 1985.

Chapter 8 applies to this Unit. Consistent with (1) and (2).

9. Runyon, Richard P. and Audrey Haber. *Fundamentals of Behavioral Statistics* . Reading MA. Addison-Wesley Publishing Company. 1984. Another (usually) readable, non-theoretical college text.

<https://teachersinstitute.yale.edu>

©2019 by the Yale-New Haven Teachers Institute, Yale University

For terms of use visit <https://teachersinstitute.yale.edu/terms>