MARGIN OF ERROR

If you look on pages 205 and 209 of the book, you will notice examples of confidence intervals for population means ... using the standard error of the mean and the appropriate t value from the relevant t distribution.

For the example on page 205, we see that the 95% confidence interval is:

4.9812 +/- 2.1315 (t) * .0502 (stanerr)

This amounts to 4.9812 +/- .107 = 4.874 <--------> 5.088

The value of .107 ... that we add and subtract from the sample mean is called: **THE MARGIN OF ERROR**.

Stated another way ... we have built a confidence interval, at the 95% level, that will be within approximately .1 (in either direction) ... or one tenth of a gallon .... of the true mean gallon value in the larger population.

For the example on page 209, we see that the 95% confidence interval is:

107.22 +/- 1.9734 (t) * 1.03 (stanerr)

This amounts to 107.22 +/- 2.03 = 105.19 <-----> 109.25

The value of 2.03 that we add and subtract to the sample mean is called the **MARGIN OF ERROR.**

Stated another way ... we have built a confidence interval, at the 95% level, that will be within approximately 2 IQ points (in either direction) ... of the true population mean value.

Of course, in both cases, we found out about this margin of error (M) AFTER the fact ... we took a sample of a certain size ... decided on building a 95% CI ... and, the M values of .1 gallon and 2 IQ points are just what we ended up with.

PLANNING A STUDY TO YIELD A CERTAIN MARGIN OF ERROR (M)

But, what if going INTO doing the study, you wanted to plan the study in such a way that, your M or margin of error, would be EQUAL (approximately) to some PRE specified value? For example, in the gasoline gallons study, you had decided AHEAD of time, you would like your CI to be, 95% of the time, within about .05 of a gallon ... not the .1 that we got. Or, in the IQ study, you decided AHEAD of time that being within 5 IQ points of the truth, would be sufficient ... that is, being within 2 IQ points is more PRECISE than you have to be?

Thus, we could reformulate the study question along these lines: how large of a SAMPLE would I need, with 95% confidence, to produce an interval (CI) that would be within .05 gallons ... or, within 5 IQ points? That is, what sort of a sample size (n) taken at random ... would achieve the level of PRECISION that I deem to be important?

Let’s look at the M formula ... and examine it a bit more.

M = t value * standard error

M = t value * (estimate of population SD / sqrt sample size)

M = t value * (s / sqrt n)

What if you want to rearrange this formula ... so that n is on the LEFT ... given some M value which would be on the right?

M = (t value * s)/ sqrt n

If we cross multiply and exchange the places of M with sqrt n, we have:

Sqrt n = (t value * s)/ M

However, we are not looking for the square ROOT of n ... we want n!

So, square both sides ... this would yield n on the left ... which we are solving for.

... n = [(t value * s)/ M]^2

EXAMPLE 1

Think back to the gasoline gallons example ... we found with our sample size of 16 and wanting to build a 95% CI, our M or margin of error was about .1 of a gallon. What if we wanted to be accurate within about .05 of a gallon? Our formula would look like:

... n = [(t value * s)/ M

Now, we can’t know our EXACT t value ... since that is based on n ... which we are looking for and DON’T know ... but, we could use the same t value we did before ... just as an approximation ... and see what happens. So, let’s try that.

On page 206, the s value or estimate of the population SD was given as: .2 ... rounded for simplicity. Thus:

n = [(2.1315 * .2) / .05]^2 = [.4263/.05]^2 = 72 or a sample size of about 70 ...

NOTE: We know that this is a bit high ... since, to achieve GREATER precision, it makes sense that our sample size will have to be larger (to reduce error) ... and, with a large n, our relevant t value will have more degrees of freedom AND thus will be smaller ... so, with a somewhat smaller t value in the numerator ... the overall numerator will be smaller ... and hence, the division by M^2 will yield a smaller overall n value.

The bottom line in this case is that ... if we want to be within about .05 of a gallon, rather than the within .1 that we originally found using n=16 ... our RANDOM sample size would have had to have been approximately 4 times larger ... 70 or so rather than 16.

EXAMPLE 2

For the IQ example, when n=179 in that case, we found M or the margin of error to be about 2 IQ points. What if we were willing or only needed to be accurate to within about 5 IQ points? Since we are saying that we don't have to be as accurate ... we know that we won't need as large of a sample size ... so, how much (approximately) less than 179 would it take ... to build a 95% confidence interval ... that is accurate to within about 5 IQ points rather than 2 IQ points?

Again, as an approximation, we can use the t value for n=178 degrees of freedom (even though with a smaller sample size, we know t will be a bit larger) ... and, we can use the s value of 13.822 as stated on page 209.

.... n = [(t * s) / M]^2 = [27.276/5]^2 = approximately 30 ...

Given that we are willing to tolerate about 5/2 = 2.5 times more error in terms of our precision of our estimate of the population mean IQ value ... we see that our sample size would only need to be about 180/30 = about a 1/6 of the original size.

NOTE: Again, since we have used a t value as an approximation in the numerator that is SMALLER than we know it should be ... that means the numerator is smaller ... and thus, the overall calculation of n is SMALLER than it should be.

The main points from the examples however are that:

IF YOU WANT TO BE **MORE PRECISE** .... ie, HAVE A SMALLER M .... YOUR SAMPLE SIZE **n** WILL HAVE TO BE CONSIDERABLY **LARGER**

IF YOU ARE WILLING TO TOLERATE **LESS PRECISION** ... ie WORK WITH A LARGER M ... YOUR SAMPLE SIZE **n** CAN BE MUCH **SMALLER**

The bottom line is that IF you are able to PRE specify some approximate level of error or M ... that you need or are willing to work with ... and you have some approximation of the SD in the population ... you can examine the kind of n you would need to achieve that goal. Rather than accepting whatever you find ... why not PREPLAN? The benefit of preplanning is to not use more n than you need ... OR know that you need a much bigger n than perhaps you thought you would.

FINAL COMMENTS

All of this seems to be sensible as preplanning strategy but, we have to consider two other elements to better place the above in proper perspective.

1. When finding the standard error, an assumption is made that you have taken a SRS or simple random sample ... ie, error estimates are based on that having been done. Unfortunately, taking REAL random samples is either very hard to do or essentially impossible. Thus, in many cases, the error estimates we obtain (our standard error values) are too small. For many practical data collection problems, this is not a trivial matter. The implication of this is that when we estimate an n that will achieve some desirable level of precision or M ... we might be considerably UNDER estimating the n we really need.

2. Notice in the examples above that, estimates of the population standard deviations are needed. But, what if you are starting out fresh ... that is, you have NO data on which to even have any estimate and there seems not to be any available literature that might give you a clue? What do you do? Well, for this and many other logistical matters, one should do a PILOT study first ... even with a small sample of Ss ... to get some feel for what that standard deviation might be ... and then we could use THAT pilot value as your best estimate at the moment. There are many other reasons for doing a pilot study but, for the purposes here ... providing some approximate estimate for the standard deviation value is key.