

A Sample Size is chosen

In which ... The arbitrary nature of the process is examined, and a few myths are addressed.

*The Nun gets a hearing,
and Dr Smiley sees an opening.*

Vi believed strongly that sample sizes should be calculated using one of the complicated equations that she learned from Dr. Smiley. Why had she spent all that time in statistics class if it was not really important? It was unthinkable. Although she had ignored OH's heretical procedures for the first few weeks of her employment, she decided that it was time to take a stand. Some recent inventory work had been done, and OH had set an arbitrary sample size of 150. Vi sent a memo around the company demanding that this decision should be reviewed. Betsy, the department secretary "mistakenly" left a draft of the memo where OH happened to notice it. OH was intended to be the last one to see the memo. Vi had made a mistake in doing that.

OH's Vice-President was confident that OH could handle the situation. He arranged to preside over the meeting on this topic, and said that Vi would have his full attention. In fact, the VP brought along a bunch of experienced senior managers – none of whom had an advanced degree. Vi thought that gave her an advantage, because *no doubt* they would be impressed with her Masters Degree. She also brought along Dr. Smiley. It was two against one, she reasoned. She had made the classic error of mistaking manpower for firepower. She and Smiley had reviewed sample size theory for a full day before the meeting, and they felt well prepared.

Smiley saw this as a chance for him to grab some consulting work. Although OH had been difficult to deal with on previous occasions, Smiley had been doing extra reading on this subject, and was loaded for bear.

OH, however, was more like a teddy bear – or so it seemed.

“We appreciate your coming, Dr. Smiley”, he said. “We are all prepared to reconsider the arbitrary sample size that I had suggested, and I wonder if the two of you can just lead us through the scientific issues here”, he said – all sweetness and light. Lem knew that this meant he was *really* angry at this put-up job, and at The Nun for organizing it.

Vi, as the good apprentice, started out. “Before Dr. Smiley gets into the complicated optimum forms of sample size estimation”, she said, “I would like to review the most basic logic for the managers present”. They all nodded their appreciation. Truth was, they were just there for the show. You could have sold tickets for this one.

“As you know, the most trivial and simple equation is :”

$$\frac{\text{Standard Deviation}}{\sqrt{n}} = \text{Standard Error}$$

“Now as we move up in complexity ...”

“*Excuse me*, Vi, I have a question here”, said OH. “I take it that you don’t just want to choose ‘n’, the sample size. Is that right? You really want to solve this equation based on the two other variables”.

The Nun thought that was just about the *stupidest* thing she had ever been asked. Of course you calculated sample size. It was in all the books. “Well certainly we want to calculate ‘n’”, she said. “I think it’s apparent that you can’t just *invent* a number in these matters”. *I think I stated that with appropriate academic firmness*, she thought. She was pleased.

“Well, **I** can do that”, replied OH. “We are all adults here. As sampling specialists we have judgment, we know what would be credible, and we have done this sort of thing before¹. **I** have no problem choosing a number – it’s 150 by the way – I just wanted to make sure you felt compelled to calculate it. If it’s not right to simply choose ‘n’ then we have 3 coefficients and 3 unknowns in the equation. I suppose that you can just whittle that down with a little theory. Please continue, Vi”.

¹ It was apparent to all in the room that The Nun and Smiley were not really included in the term “we” here – well, perhaps not obvious to those two.

At this point, Dr. Smiley smelled blood in the water. Perhaps he could score points early with this simpler material that the managers would understand, and he could build on that success later.

“*Excuse me, Vi*”, Smiley interrupted. “I think I can bring some theory to bear upon this situation. As all of you probably know, I teach statistics at the University, and this topic is one I deal with all the time. Now ...”

“*Excuse me*”, said OH. “Perhaps we could just cut to the chase. We know that you need to establish both of these unknowns. Can you tell me, Dr. Smiley, what the Standard Deviation is in this formula?”. OH seemed to give Smiley the perfect opening, and he grabbed for it.

“Well, let’s suppose that it is ...”

“*Whoops* – you misunderstand me, Dr. Smiley, I was asking what the Standard Deviation IS in this instance. What is it for this population?”. Smiley froze for a moment. Nobody had ever asked him such a question. He was used to defining it. You just pulled a number out of the air and ... used it. The answer for an actual population? He was immobilized.

“Well, the point is ...”

“*The point is*, Dr Smiley, that *you don’t know* the Standard Deviation, do you?”, interrupted OH. OH allowed an awkward pause during which it was obvious that this was the truth. “*Troubling*”, said OH, “so I suppose that still leaves us with 3 coefficients and 3 unknowns. Hard to solve that sort of thing, as I am sure your mathematics training tells you. Just to help you out, let’s suppose that after *quite* a bit of work we have an estimate - not the right answer, mind you – but an estimate of $SD = 60$. That only leaves us with 2 unknowns. We seem so close now. Please continue.”

Smiley was not quite sure what had just happened, but it appeared that he had been embarrassed, and had no idea what the answer was for this actual sampling problem. That never happened during his lectures. Not a man to let that kind of thing stop him, he forged ahead. All he needed was one simple coefficient and he was home free. *Perhaps*, he thought for a brief moment, *I should have found out what this sampling project was all about.*

“That seems about right”, he said (after trying to look thoughtful for a few seconds), “Now all we have to determine is the Standard Error that is needed. Let’s suppose ...”

“*Whoops*, pardon me again, Dr. *I don’t know* the amount ‘needed’, and *neither do you*. At this point, I have given you an answer for one of the 2 coefficients. You, however, now need one of your very own. Where are you going to get this Standard Error?”. At this point, Smiley started to lose his confidence a bit. In class, he had simply picked a number out of the air.

In the textbook, the “required” precision was always stated in the exercises. He had never actually thought about this. That didn’t stop him.

“I believe that the standard texts are quite clear about this OH, we simply ask the decision maker what he needs, and then ...”

“*Whoops*. Pardon me again Dr. Smiley, but we are surrounded by decision makers here. In fact, these are exactly the people in the company who would make this decision. You could simply ask them”. OH sat back and waited for Smiley to bleed out.

“Very well”, he said. Choosing a manager at random he said “You over there ... could you give me a number please?”. “What kind of number?” was the reply. “Well, the required precision, of course”, said Smiley. “I have no idea”, was the response. “Well, we have to have one”, said the good Doctor – in the tone one might use with a graduate student.

The manager was not impressed. “Look, Bubba, this sounds like expensive stuff and I am *not about* to grab a number out of thin air. You asked what I ‘needed’ and *I don’t know*. I know of no such requirement. There is no company policy, no legislation, and if it existed it would be arbitrary and not based on any known ‘need’ – it would be as arbitrary as choosing the sample size directly. *I just don’t know*. What are the units for this ‘need’ anyway – dollars, Board Feet, field weeks - what?”. “Well”, Smiley replied, “I don’t think that matters”. This caused a brief silence in the room, reflecting how profoundly detached from reality his response was.

... *Whoops* ...

Smiley now realized his mistake. These people were not students. He was actually expected to answer questions himself, and they were not going to give him a free ride. As always, he thought the best thing to do was to charge ahead. He tried again. “Well let’s assume that you needed ...”

“*Whoa*, Dr. Smiley, *Whoa*”, said OH. “We are still seeking ‘need’ here. It was, after all, your protégé here who insisted that numbers should *not just be made up*. Now I have never personally known anyone who knew what they ‘needed’ for a Standard Error, but maybe you can find one. The first thing we should do is ask if there is anyone in the room who knows. Hands up anyone who knows ... anyone ... anyone?”. No hands went up.

“So, Dr. Smiley, we are still stuck with 3 coefficients and either 2 or 3 unknowns (since you might not even accept my invented number for Standard Deviation). I have a suggestion”. OH nodded at the equation on the board.

$$\frac{60?}{\sqrt{?}} = ?$$

OH reached into his coat pocket, and fanned out ten serially marked \$100 bills on the table in front of him². “Dr. Smiley, I propose to give you this \$1,000, *plus* your daily fee, *plus* a note of apology to be posted on the Internet if you can just solve this *simple* equation before we move on to the ‘more advanced’ methods. If you cannot, all I want is a note of apology from you and Vi here to be posted on my website. The only restriction is that you cannot just ‘invent’ a number as you *rightly* identified that I did – and which led to this expenditure of time and effort. The alternative is that you all walk away and agree that this problem needs no more discussion.”

“Deal - or No Deal ? ... Dr. Smiley ?”

The room went silent. Smiley realized that the game was up. Everyone else had realized that several minutes earlier.

The meeting was at an end. The VP tried to wrap it up politely.



Lem was curious why OH was so annoyed by this encounter, and when the VP took them both to a fancy lunch he had a chance to ask. “This kind of thing usually does not rile you, OH. What was the problem here ?”. “Simple, Lem, this young twit and her jumped-up mentor tried to attack our credibility. In inventory, credibility is *by far* our most important asset. They also went about the issue in an unacceptable way. If Vi had come to talk to me I would have explained this to her, and she might just have understood, but she decided to stage this scene”.

OH put down his fork. “She puts me right off my fresh fried lobster. She is what we call a *regular academic clown*³. The university is too good an institution and too good an influence in the world to be represented by folks who do not understand the rudiments of their craft and have no manners. It’s a double loss. Let’s pray that she does not go straight back and teach. I think that Deming put it well when he said ‘a hack of a statistician should be learning, not teaching’ ”⁴

“Did they have any case at all here ?”, asked Lem. “Nope”, OH replied. “Choosing either a Standard Error or a sample size directly are both arbitrary. The real issue was whether you can live with the Sampling Error you would get with 150 observations. If you were foolish enough you could calculate the Sampling Error that would *require* a sample size of 150, then

² This was not the first time OH had done this to Smiley. Lem had previously heard OH’s story involving World War II, a Swiss bank, and the \$100 bills.

³ Apparently, OH was a fan of the movie “MASH”.

⁴ Deming “Out of the Crisis”, 1986, 24th printing, page 468

pretend you wanted that SE all along, put it back into the equation and find (surprise) a ‘requirement’ of 150 observations as a result. Simple fraud”.

“In practice, the first thing the manager will ask us is ‘what seems like a reasonable goal for the precision?’ – probably in percentage form, or perhaps he would ask ‘what is the industry standard for precision?’. If Smiley did pick his ‘requirement’ out of the air and solve for a sample size, someone would immediately observe that considering the cost, the time limit, the manpower available or the psychology of the situation, some other sample size would be better. At this point they would adjust the ‘required’ SE together until the sample size was comfortable. I have seldom seen it done any another way”.

“The thing that intrigues me here” OH continued, as he sat back in his chair, “is that instinctively I believe this young lady has something worth developing. It’s hard to explain, but I think she has just been criminally misled in the past, and she is worth more than is apparent so far”.

“Do you think this is over now?”, Lem asked “One can only hope”, said OH. “But I doubt it. Wait until she finds out that the plan is to do a systematic sample rather than a random one. That should *really* set them off. Rookies of that degree are mesmerized by random samples”.

Another tangle like this? *Not likely after this embarrassment,* thought Lem – *she would have to be **absolutely bonkers**.*

*Note to reader ... the next chapter title is, naturally ...
“Absolutely Bonkers”, and is about systematic sampling.*