

**STATISTICAL THINKING IN THE KITCHEN:  
SAMPLES, POPULATIONS, SAMPLE SIZE, AND REPRESENTATIVENESS<sup>1</sup>**

*K. P. Mohanan (1 Feb 2010)*

**PART I**

Some of my scientist friends were meeting at my place for potluck dinner and a movie that evening. I was preparing vegetables for a vegetable-cheese dish. Alka, my sister-in-law, who was visiting, walked into the kitchen, saw me cutting the vegetables, and offering to help, picked up the potatoes.

“How do you want these cut?” she said.

“Oh, into about half-centimeter cubes,” I replied.

“Not half-inch, huh?” She laughed. “Half-centimeter! Finicky, aren’t we?”

“Yes, I want them very small.”

As I diced the celery, capsicum, and carrots, I kept thinking of the importance of the way vegetables are cut, how it affects the taste, texture, and soul of a dish, and of the professional pride in cutting vegetables just right. I am not a professional cook, but the way vegetables are cut, I certainly am finicky about.

When Alka finished, she asked, “Does this qualify?” I looked at the bowlful of potato cubes, and said, “It doesn’t really matter, but those look more like three-fourth centimeter cubes to me.”

“No, I think they are half-centimeter cubes,” Alka said emphatically.

Mishti, my fifteen-year-old niece (Alka’s daughter), was watching a TV commercial in the living room. “Mishti, come over here,” I called out. She turned off the TV set, ambled to the kitchen, and stood at the door looking at me.

“Your mother says these are half-centimeter cubes,” I said. “I think they are three-fourths centimeter cubes. What do YOU think?”

She looked at the pile of potato cubes in the bowl, and shrugged. “I don’t know, I can’t tell whether they are half or three fourths. That’s such a small difference.”

“How will you find out?” I said.

“I guess I’ll take a ruler and measure.”

“Go get a ruler, then.”

She got a ruler, picked up a piece of the chopped potato, and measured it.

“It’s half a centimeter.”

“So?”

“So Mom is right. It is half a centimeter, not three-fourths.”

“Pick another piece, Mishti, and measure it.”

The one she measured this time was three-fourths centimeter. “Oh, so you are both right. Some are half and some are three-fourths.”

---

<sup>1</sup> This is a mildly embellished true story. Though only one name appears under the title, it has two authors. A rough draft was made by K P Mohanan and was significantly modified by Tara Mohanan.

“Can I ask you a question, Mishti?”

“Yup?”

“There must be more than a couple of hundred cubes in that bowl, right?”

“Right.”

“What if just one piece is half a centimeter and the rest are all three fourths, and you happened to pick, just by sheer chance, that single half-centimeter one the first time. Who would be right if that were the case?”

“You would be.”

“And what if just one piece is three-fourths and the rest are half, and you just happened to pick, just by sheer chance, that single three-fourths one the second time. Who would be right?”

“Mom would be.”

“And how do you know which is the case?”

“Oh!” Mishti frowned for a moment, lost in thought, then said, “I think I’ll have to measure all the pieces.”

“Suppose you find they are all different sizes. Some are 0.4 centimeters, some are 0.5, some are 0.8, and so on. Then what will you do.”

“I’ll put them in different piles.”

“What kind of piles?”

“I will make a pile of pieces with 0.1 to 0.5, from 0.6 to 1.0. and 1.1 to 1.5.”

“And suppose you found each pile to be equal?”

“Then you will both be equally right, or both equally wrong.” She was beginning to get into the spirit of the question.

“Suppose you found that out of two hundred pieces, a hundred and eighty are in the 0.6 to 1.0 centimeter pile, and ten each in the other two piles.”

“Then Mom would be right. Oh, wait, no, then you would be right. Oh no, ... well, ... I don’t know.”

“So making piles won’t work?”

“Er... I guess not.”

“So what are you going to do?”

“I don’t know.”

“Take your time, Mishti. You can give me an answer when you’re ready.”

“Wait, I know what I will do. I’ll take the average.”

“What is average?”

“Jeez, don’t they teach these professors anything? Average is this number.”

“What number?”

“Well, you measure all the cubes, then you add up all the numbers and then you divide them by the total number of pieces. That would be the average.”

“Very good, Mishti. The kind of average you are talking about is called ‘mean’.”

“Oh yeah! Some mean people like you must have invented it.”

“Ha ha! You can see me cracking up.”

“Okay, bye, Mohanmama.”

“Mishti, wait. Don’t go yet. To do this you said you have to measure all the pieces in the bowl?”

“Yeah, YOU are the one pestering me with questions. How else will I find out who is right, Mom or you?”

“Suppose we’re cutting vegetables for a party of about a hundred people. Instead of a small heap of potato cubes in a bowl, there’s a huge pot of them. More than ten thousand pieces. Are you going to measure ALL of them? Your hair will turn gray by the time you’re done. Is there a more practical way to find a reasonably approximate answer? Not necessarily an exact one.”

“What other way is there?”

“Think about it. Let me know when you come up with a solution.”

“Mohanmama?”

“Yes?”

“My brain is all fried up. Can I go back to my TV?”

“Sorry, go watch your commercials.”

Ten minutes later, I found Mishti sitting in front of the TV but not looking at it. She was lost in thought. And then she jumped up. “Mohanmama, may be I can take a sample and measure the pieces in the sample.”

“What is a sample?”

“A sample is what you take from the pot. Like a handful.”

“What you take from the pot? A handful of peanuts?”

“No. A handful of potato pieces,” she said patiently, in all earnestness. “For a pot of potato pieces a sample is a few pieces from the pot.”

“A sample from the population. Okay.”

“No. Not from a population. We’re talking about potatoes, not people, Mohanmama. Jeez, these profs.”

“Populations are not necessarily populations of people, Mishti. Have you heard of the subject called statistics?”

“Yeah. You showed me a video lecture on the statistics thingie yesterday, remember?”

“Oh yes, I did. Okay, in statistics, a population is a collection of any thing. It can be people, rabbits, tables, marks, prices, just any collection. A sample is a part of the population. So if you pick a handful of potatoes from a pot, the potatoes in the pot is the population and the handful is the sample.”

“Okay, so I pick a sample from the population of potato pieces.”

“Yes. Go ahead, pick your sample from the bowl.”

Mishti picked about ten pieces from the top of the pile. She was going to measure them when I stopped her.

“Can I ask you something, Mishti?”

“I am beginning to get scared about your can-I-ask-you-something questions. Every time you ask that question, you get me into deep trouble.”

“That’s my job, Mishti. Getting kids into trouble.”

“What were you going to ask?”

“You picked all your ten pieces from the top, right?”

“Yes.”

“Suppose you found that those in the sample are mostly half a centimeter cubes. But is it possible that when your Mom started cutting the potatoes, she was cutting them bigger, and when she was about to finish she cut them smaller? She is not a machine, after all, she is a human being. Isn’t that possible?”

“Yes.”

“If that’s what happened, how would you know if what your sample is telling you what is true of the population, or if it’s giving you distorted information?”

“Oh... I think my sample would be distorting. It would say one thing, but the population would be something else.”

“In statistics, they would say that the sample doesn’t adequately represent the population, or is not representative of the population.”

“What’s that?”

“A sample should be representative of the population. Whatever is true of the population should be true of the sample as well. The sample should reflect the properties of the population.”

“I see. So if Mom changed her cutting style towards the end, my sample from the top of the pile would not be representative of the population.”

“That is not unlikely, right? How would you make sure that your sample is representative? At least increase the probability that the sample represents the population?”

Mishti thought for a moment. She dug her fingers into the bowl of potatoes, and did a thorough job of mixing the pieces. Then she closed her eyes, chanted some children verse like “eeni-meeni-maini-mo”, thrust her hand into the pile, and with her eyes still closed, gathered a handful of pieces.

There were eighteen pieces in her sample. Their measurements were:

0.4; 0.7; 0.9; 0.5; 0.4; 0.7; 1.1; 0.7; 0.8; 1.0; 0.6; 0.7; 0.8; 0.8; 0.9; 0.5; 0.6; 0.8.

“Mohanmama, the mean is a little more than 0.7,” she said after her calculations. “So it looks like you were right.”

## **PART 2**

A few minutes later, when Mishti came back to the kitchen to invade the fridge, I said, “Mishti, I want you to take a few potato cubes, say around ten, and find their mean size.” She picked ten pieces randomly, and measured them. The measurements were:

0.4; 0.7; 0.5; 0.5; 0.4; 0.6; 1.1; 0.3; 0.8; 0.4.

Mishti did the calculation on a piece of paper. “Oh, oh!”

“What happened?”

“The mean is 0.57. That means Mom was right. Not you. This is driving me crazy. What is it with these idiotic potatoes any way. Why do they keep switching sides?”

“Let us take a look. Your first sample of eighteen was, where is it, ... okay, the measurements were: 0.4; 0.8; 0.9; 0.5; 0.4; 0.7; 1.2; 0.8; 0.8; 1.1; 0.6; 0.7; 0.8; 0.9; 1.0; 0.5; 0.6; 0.9. And the mean was 0.7.”

“Yes.”

“Let’s pick five from this sample. I am going to pick the first one, 0.4, the fourth, 0.5, the fifth, 0.4, the eleventh, 0.6, and the seventeenth, 0.6. What is the mean?”

“It is 0.5.”

“So even from the first sample, if you took a smaller subset, your conclusion can be different, right?”

“Yeah, the numbers are all over the place.”

“Which sample would you trust more to give you a more accurate conclusion?”

“I don’t know. May be if we took a very large sample, say fifty pieces or so, we can get a more trustable mean. If we took two samples of five pieces each, the mean can fluctuate like crazy, but if we took two samples of fifty pieces each, I have a feeling that they will be much closer.”

“So what you’re saying is that *when the sample size increases, the sample become more representative of the population.*”

“Am I saying that? I didn’t think of it that way, actually, but now that you have put it that way, yes, that’s reasonable. As the sample size increases, the sample becomes more representative of the population.”

“And the more representative a sample is, the more we can trust our conclusion.”

“That looks like it.”

“A little while ago, we figured out that if you pick all your pieces from the top, you won’t get a representative sample: it will be a biased sample. So we made it random to make it representative. But now we find that random selection is not enough. The size of the sample should also be large enough. So the sample has to be both large and random to make it representative.”

“You sound like a teacher, Mohanmaama.”

“I *am* a teacher. Remember, I teach in a university. So here is my question for you. You want to arrive at a reasonably certain conclusion about a population on the basis of a sample. What makes a sample large enough? What would your decision on the sample size depend upon?”

“Erm, ... that depends on the population, right? The sample size should be proportionate to the population.”

“If you take a sample of 10 for a population is 100, you should take a sample of 200 for a population of 2,000, and a sample of 5000 for a population of 50,000?”

“I guess.”

“Let us see. Suppose you see a cup of soup on the table. You want to check if it has enough salt, so you take a teaspoonful and taste it. If the salt is enough, you will conclude that there is enough salt in the soup in the cup, right?”

“Right”

“So a sample of one teaspoon is sufficient for a cup of soup. But what if it is not a cup of soup, it is a big bowl of soup. Do you need to taste more teaspoons of soup to find out if there is enough salt?”

“No, one teaspoon from the bowl is enough.”

“I see. What if it is a huge pot of soup. Or a barrel of soup?”

“It is all the same soup, right? One teaspoon is still enough.”

“But you said a little while ago that the size of the sample should be proportionate to the size of the population. Won't that mean that you take a small sample for a cup of soup and a big sample for a big pot of soup?”

“No... Yes... I mean, I did say that, but that doesn't work here for the soup.”

“So what should the choice of the sample size depend on?”

“I don't know. Mohanmaama, why are you asking me? I am just fifteen, you are the professor, you should know.”

“I am trying to help you to figure it out for yourself, Mishti. It is more fun that way. Don't you feel good when you find out something on your own?”

“Yes, I suppose so.”

“Alright, then. Let us try another example. We were looking at the potatoes that your Mom cut. Suppose the potatoes were cut not by her but a very precise machine. Would you still need a large sample size? When should we use a larger sample size, if something is cut by a machine or by a person?”

“Er... when cut by a person?”

“And suppose you have another population of potatoes cut not by one person but by many different persons. Which one would need a larger sample, when cut by one person or when cut by many?”

“When cut by many people.”

“So when cut by a machine, we use a small sample, when cut by a person we use a medium size sample, and when cut by many different people, we use a large sample. What does this depend on?”

“A machine would cut vegetables uniformly. All pieces would be the same. But when a person cuts it, it will be different. And when there are many people, it will be still more different.”

“What you are pointing to, Mishti, is uniformity and variability. If the property we are investigating is uniform across the population, like in soup, a small sample will do. But if the population is not uniform, if it exhibits variations, then we should take larger samples. The more variability we expect, the larger our sample. Are you falling asleep?”

“No, I am not asleep. I am thinking about what you are saying.”

“Did you understand what I said?”

“I think so, kind of.”

“Alright, tell me then, why is it that in the population of diced potatoes we were talking about, the machine-cut potatoes call for the smallest sample, and the potatoes cut by different people calls for the largest sample?”

“Because machines are mechanical and they cut all the potatoes the same way, so a smaller sample is enough. But humans don’t do everything so uniformly, so human-cut potatoes would be more variable, so we need a larger sample. And if they are cut by different humans, they will be still more variable. So that needs the largest sample.”

“Excellent.”

“Do I pass the test, Sir Professor?”

“You do, Mishti. With distinction.”

“Now can I watch TV to throw out all the potatoes from my brain?”

“I will watch with you.”

“No, you won’t. If you do, you are going to ask me another question in ten minutes. Go away, sit in front of your beloved computer, not in front of the TV.”

“Okay, bye, Mishti.”

“Bye, Mohanmama.’