2 Statistics for Quality

Learning objectives

- Understand main concepts of statistics for quality
- Describe the differences between descriptive and inferential statistics
- Learn main concepts regarding samples, population, standard deviation and sigma level
- Differentiate statistical significance and business significance

Keywords

Statistical process control, sigma, normal curves, sample.

Required skills

A general knowledge of management on the bachelor's degree level.

Time requirements for the study

You will need approximately 90 minutes of your time to study this chapter.

2.1 Introduction

Statistics is valuable to quality management because it gives us a sound basis for making decisions such as:

- Is this batch good enough to meet customer requirements?
- Which of these changes will eliminate the most number of defects?
- What change should we make to our process so future production meets customer requirements?

Statistics is an independent discipline that fits well with the scientific method. It draws from - but is not part of - philosophy, science, and probability. It is applied in science, engineering, psychology, business, and other fields. In the broadest sense, the field of statistics is a structured way of saying things about how the world works. In practice, its two most important benefits are:

- Statistics allows us to view a part of the world - a sample - and conclude things about a larger part of the world - the population - that we can’t see and measure affordably.
- Statistics can be used to take information about the past and forecast the future.

2.2 Measurement for Statistics

In statistical terms, each measurement is taken on one attribute of one subject (for example, our process, our product, or a component), on one occasion (at a particular time and in a particular situation), resulting in a value. We may want to measure many attributes at the same time, as they may affect one another. For example, the amount of humidity in a drying room can affect the dryness, and therefore the weight, of our product. So we would want to record the amount of time the product was in the room, the temperature, the humidity, and the weight of the sliced ham drying on the rack all at once. Every measurement also is associated with a variable that names what we are measuring (weight of ham, weight of cheese, length of hoagie roll) and a unit of measure. The end results of measurement are data - measured values that we record. Data is the input for statistical descriptions and statistical procedures.

What if our measurements aren't good? This brings up the issue of measurement error, which is very different from error in quality management. Error in quality management is about mistakes in production process. Measurement error is about mistakes in the measurement process, which would prevent us from finding the mistakes in the production process. Statisticians talk about two kinds of error: reliability and validity. Reliability is like precision - it is about results being closely clustered together, rather than widely scattered. Validity is about approaching the target, as opposed to being consistently in the wrong direction - validity indicates the absence of bias.

2.3 Samples and populations

The idea of samples and populations is key to statistics. A population is the total group in which we are interested. It might be every ham-and-cheese sandwich we ever made or ever will make, or every one we made in a particular factory in a year, or the entire batch of 10,000 ham-and-cheese sandwiches in the fridge that is due to go to the customer tomorrow. A sample is part of a population. Our goal is to measure a sample and be confident of something about the whole population. This can only happen if we are confident that the sample represents the population reasonably well. There are many types of samples, named according to how we select the sample from the entire population. Of all the types of samples on the following list, only the first two give us a reasonable degree of confidence that our sample will represent the population.

- Comprehensive sample. In a comprehensive sample, we seek to include the entire population in the sample. The difference between a population and a comprehensive sample is due to mistakes such as missed items or lost data.
- Random sample. In a random sample, each item in the population has an equal chance of being included in the sample. Getting a random sample is harder than you think!
- Convenience sample. We want to avoid this kind of sampling, where we simply get the sample in the easiest, cheapest way. If we do this, our sample is unlikely to represent the population. For example, if we want to check the quality of sandwiches in our stores, it would be most convenient just to get sandwiches from the nearest store. But they wouldn't be likely to represent sandwiches in all stores accurately, because they were all stored in the same refrigerator at the same time. Very different things might have happened to sandwiches in other stores.

- Systematic sample. Here, we get a sample in a non-random way. For example, we might pick the ham-and-cheese sandwich in the top right corner of each box. But what if our refrigeration unit is colder on top and warmer on the bottom? Then we'll never see the moldy sandwiches.
- Judgmental sample. This is a term from auditing, actually, rather than from statistics. It means using our own common sense - our expert judgment - to decide how to take our sample. For example, we might choose to look at the sandwiches from only the stores that failed health inspections in the last year.
- Stratified sample. Building a stratified sample is complicated, but essentially, a stratified sample is a combination of a judgmental sample, then random items selected from groups selected on the basis of expert judgment.
- Quota sample. Again, a poor choice, the quota sample is similar to the convenience sample, except that we stop collecting when our sample is large enough.
- Self-selected sample. In this case, the subject has a say in whether or not to be included in the sample. Of course, ham-and-cheese sandwiches can't stand up to be counted, but customers can. A good example of a self-selected sample is customers who choose to answer our customer survey. Unfortunately, we can't be sure that the self-selection doesn't bias the sample.

Just as we can have error in measurement, we can also have sampling error. If we know we didn't use a random sample, then we know that our sample is probably not representative of the population. If we think we have a random sample, then we have more confidence that there is little sampling error and that the sample is similar to the population. However, whenever we can't take a comprehensive sample, there is always some doubt as to the question of whether our sample truly represents the population.

2.4 Descriptive and Inferential Statistics

Once we select a sample and take measurements of attributes of items in the sample, we have our data values, which are the input to our statistical calculation. Descriptive statistics are a summary of the data. Commonly used descriptive statistics include the minimum and maximum values, and any of several types of averages (mean, median, and mode, among others). We can also calculate the variance and the standard deviation, which express how our sample is clustered near or spread out far away from the mean (the average or central tendency of the sample). Descriptive statistics describe our sample, but they can also be extended through a process called estimation to describe the entire population. For example, if we try out an advertisement on 1000 people randomly selected from a list of 100,000, and 20 of them buy a ham-and-cheese sandwich, then we have a 2% purchase rate. If we advertise to all 100,000 people, we should expect to sell about 2% of 100,000, or 2000 sandwiches. A good statistician could tell us more; he could make that "about" more precise. Depending on the sample size, he might say, "There is a 68% chance that you will sell about 1900 to 2100 sandwiches."

Inferential statistics go beyond description. Inferential statistics provide a measure of the relationship between two or more variables along with a second measure that indicates how confident we are that the first measure is correct.

Using inferential statistics, we can do things like forecast likely future events and determine if a particular intervention is likely to have a desired effect.

Although estimation and forecasting have more general meaning in business, in statistics, estimation always means making statements about the population based on statistics from the sample, and forecasting always means predicting expected future measures or results based on past measures or results. In fore-casting, it is essential to remember that we can never truly know the future. Rather, we are saying, "If the future is like the past, then this is what is likely to happen." Also, our past numbers do not cause our future numbers. Rather, our past numbers are a result of past causes. Those numbers show the state and trend of the figures. A forecast says, "if the underlying causes remain similar to the past, and the numbers continue to change in the same direction, then the future is likely to be like this."

2.5 Normal Curves and Standard Deviation (Sigma)

When we gather the data values from our sample, we can plot them on a curve. Often, the results look like Figure 2-1, the normal curve, also called the bell curve. The normal curve is the representation of a population that is randomly distributed around a central point called the mean. Our results might vary from the normal curve in a number of ways. If the curve leans one way or the other, we say it is skewed. If the curve has two peaks, we call it bimodal, and we know we need to examine our data and data collection methods. Skewed curves are shown on the left of Figure 2-2, and a bimodal curve is on the right. Most statistics won't give useful results if our sample population has a bimodal distribution because most statistics operate on the assumption that the curve of sample and population are normal or near normal - and the math doesn't work right when that's not true.

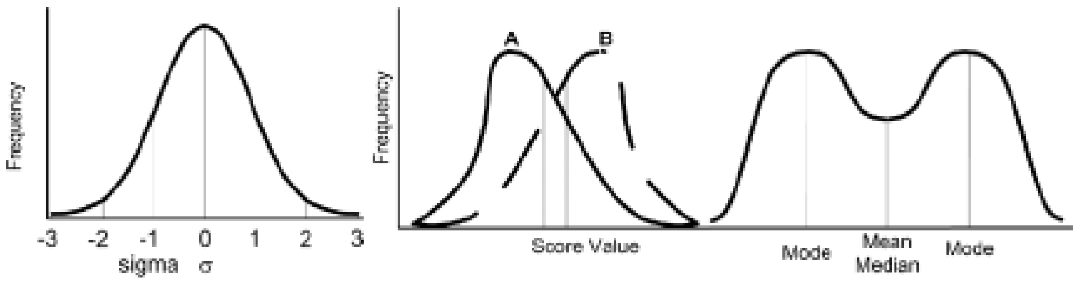

Fig. 2-1 (l) and 2-2 (r). A normal curve and Skewed and bimodal curves.

If we've done our homework - particularly used a good sampling method, chosen an appropriate sample size, and measured well - then the shape of the sample curve will be pretty much the same as the shape of the population curve. Thus, if the sample curve is normal, we can figure the population curve is normal, too. Why does that matter? A statistically normal curve is the result of natural and inevitable variation with no particular cause. A non-normal curve is an indication of the presence of some special cause or causes that are changing the shape of the curve. For example, if we picked a random sample of high school senior boys and measured their heights, we would probably get a normal curve because there is a certain average height, and about equal numbers are taller and shorter than average, but more are close to average than way off average in either direction. But if we picked senior boys playing a pickup game of basketball, and half the kids were on the high school basketball team, we'd be likely to get a skewed curve. If we picked the kids who happened to be in the gym on a particular day, and that turned out to be the basketball team and the wrestling team, we'd be likely to see a bimodal distribution. Now turn all of this around. If we measure a sample or a population, and its distribution curve doesn't match the normal curve, then there's a reason, and it makes sense to go find out what that reason is. We'll see how to do this later in this chapter.

If the curve is a normal curve, then we can identify the degree of variation from the mean average using an idea called sigma. As shown in Figure 2-1, sigma is a quantity or distance-measured outward in both directions from the mean of the curve. One sigma is closest to the central line - the mean; two sigma includes one sigma and more, and so forth. In our sample, more of the values are more clustered near the mean, and that is why the curve is higher at the center. A very important issue in statistics and quality management is the percentage of the total area under the curve within each range of variance, measured in sigma. These values are shown in Table 2-1.

As we become better at quality management, more of our products are within tolerances. Before TQM, 3 sigma (99.75% defect free) was very hard to achieve. However, U.S. business broke through that barrier in the 1980s, and started to move up. Because the curve is nonlinear, percents are no longer meaningful - everything is way over 99% error-free. So we jump from measuring in percent (events per hundred) all the way up to measuring in events per million at 4 sigma. Four sigma quality is 999,936.66 good products per million, or only 63.34 failures. Statistically, 4.5 sigma quality has fewer than 7 failures per million. By the time we reach statistical six sigma, we are actually looking at fewer than two errors per billion measurable events. That's a lot of quality.

If we know the size of the total population, then we can use a ratio to determine likely quantities of any range of values in the whole population. For example, if our sample of 10,000 1-inch screws shows that 9975 screws are between 0.90 inches and 1.10 inches in length, then we know that 3 sigma (99.75%) of our sample fall within that range. If we produce a million screws, we can expect about 997,500 screws to fall within the same range.

Table 2-1. Sigma values and percentages of the sample.

Sigma Value	Range (in sigma)	Unit of Measure	Part Under the Curve and Between the Lines	Part Outside the Range (both sides)
In Quality Management			Defect-free per Unit of Measure	Defects per Unit of Measure
1	−1 to +1	Percent	68.27%	31.73%
2	−2 to +2	Percent	95.45%	4.55%
3	−3 to +3	Percent	99.75%	0.25%
4	−4 to +4	Per million	999,936.66	63.34
4.5	−4.5 to +4.5	Per million	999,993.20	6.80
5	−5 to +5	Per million	999,999.43	0.57
6	−6 to +6	Per billion	999,999,998.03	1.97

2.6 Statistical Significance and Business Significance

When statistics is applied to science, there are very specific rules for statistical significance - for determining if statistical results provide enough information to support a new theory in place of an old one. However, this is not true for business. In business, we have to work closely with definition of business value in relation to statistical results so that the statistics are properly applied in making a business decision. Also statistical significance and business significance may be at odds with one another. For example, one time, an auditor told me how frustrated she was because she was looking for any significant correlation between methods used to do a particular type of work at different locations and effectiveness of the work. She couldn’t find any correlation that was statistically significant. I pointed out that the lack of a significant difference was a very significant business result, because if all methods are about equally effective, then the organization can save money by standardizing on the least expensive method. Statistical analysis showed us no significant difference in effectiveness - which means that the cheapest method is just about as good as the most expensive. A smart businessman will use a less expensive solution if it’s just as good as a more expensive one. Statistical insignificance can be significant for business.

Summary

There are two situations where statistical techniques and related analytical tools are very helpful in the work of quality management. One is when things are very, very bad, and the other is when things are very, very good. When things are very bad, we can feel overwhelmed by the number of problems and their complexity. In that situation, tools like Ishikawa diagrams to find causes and Pareto charts to prioritize our problems help bring error under control. We can tackle the most common or most costly errors first. These cost- and time-saving methods can be fed back into the continuous improvement effort. Iterating our improvement effort, we can bring a very bad situation under control.

We also need statistics if we’re doing a very good job and we want to do bet-ter. As we improve the quality of our processes and products, there are fewer and fewer errors. At a certain point, errors simply become very hard to find. Also, obvi-ous causes are all taken care of, and we need subtle analysis and carefully designed experiments to find subtle causes. At this point, we’re going to need statistical techniques to hunt down the error and their causes so that we can keep improving.

In this section, you will learn about the most important and commonly used tools of statistical quality management. For an in-depth discussion of these tools, their variations and many more statistical tools used in quality management, read Six Sigma Demystified by Paul Keller.

Review questions

1. What is the difference between descriptive and inferential statistics?
2. What does it mean in terms of sigma a defect rate of 3.4 PPM?
3. Describe the concepts of sample and population highlighting their differences.
4. What does it mean a statistically-significant population?

References

Kemp, S., 2005. *Quality management demystified*. McGraw Hill Professional.

Webber, L. and Wallace, M., 2011. *Quality control for dummies*. John Wiley & Sons.