Variation and Its Discontents
Funnel Plots for Fair Comparisons

Stephen Few and Katherine Rowell
Visual Business Intelligence Newsletter
October/November/December 2013

Central to quantitative data analysis is an understanding of variation. When we measure multiple occurrences of things to determine how and to what extent they differ, we're examining variation. Some variation is random and some is caused by factors that we can attempt to identify and perhaps control. Random variation consists of differences in measures that occur routinely, without a specific cause. We should note random variation and move on, because nothing can be done about it. It is noise. It tells us nothing that requires a response. Instances of non-random variation are signals; they tell us something useful and provide opportunities for action. Signals that indicate poor performance—an undesirable state—can perhaps be reduced by controlling the causes. Signals that indicate an especially good state of affairs can provide useful insights and opportunities for improvement.

Despite the significance of variation, relatively few people who work with data in most organizations understand it, especially the nature of randomness. This leads to false conclusions and poor decisions, especially when comparing measures of performance within a set of like entities (e.g., countries or companies). Most organizations spend too much time examining noise: the cacophony of random variation. Learning to distinguish signals from the noise is a fundamental skill of data analysis and performance monitoring. In this article, we'll take a look at a special version of a scatter plot, called a *funnel plot* (not to be confused with a *funnel chart*), which is designed to filter out the noise and shine a spotlight on meaningful variation when we compare performance among entities in a group. Funnel plots address the fact that entities with relatively few occurrences of the thing being measured (a small sample), when compared to entities with many occurrences (a large sample), exhibit a greater degree of random variation, which must be taken into account when comparing them. A little later we'll take a look at this problem and the solution that funnel plots provide in relation to healthcare data, but first let's get more familiar with the effects of sample size on randomness.

Randomness and Sample Size

Randomness is natural and expected. Measures of something routinely vary to a certain degree. In this article we want to emphasize the fact that a statistical measure (e.g., a mean) derived from a sample (subset of a population) can vary from the same statistic derived from the entire population when the sample is small: the smaller the sample, typically the greater the degree of variation.

The average height of a man in the United States today is 5'10" (old guys are slightly shorter on average, but we'll ignore this fact to keep things simple).

A histogram of men's heights in inches looks roughly like this:
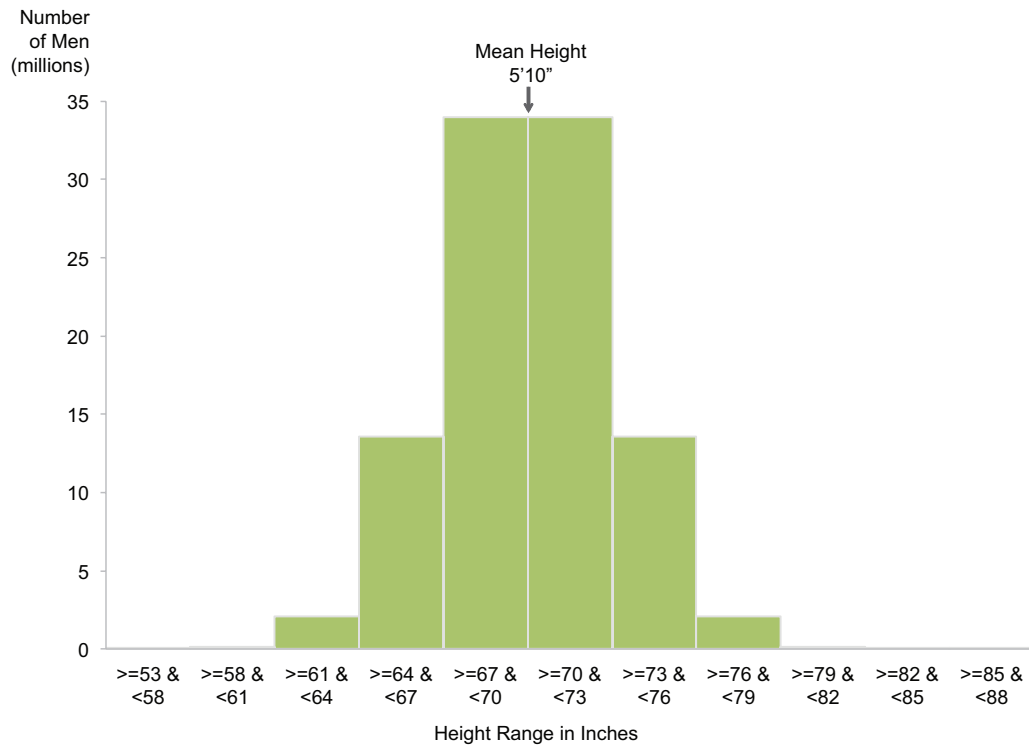
This histogram represents all men in the United States. If I were to pick 10 men at random while roaming the streets of a typical American city, however, their mean height will likely vary somewhat from this typical pattern. Consider the following sample of ten men's heights in inches:

68, 72, 75, 72, 71, 65, 71, 69, 77, 63

Here's a histogram of this small sample:
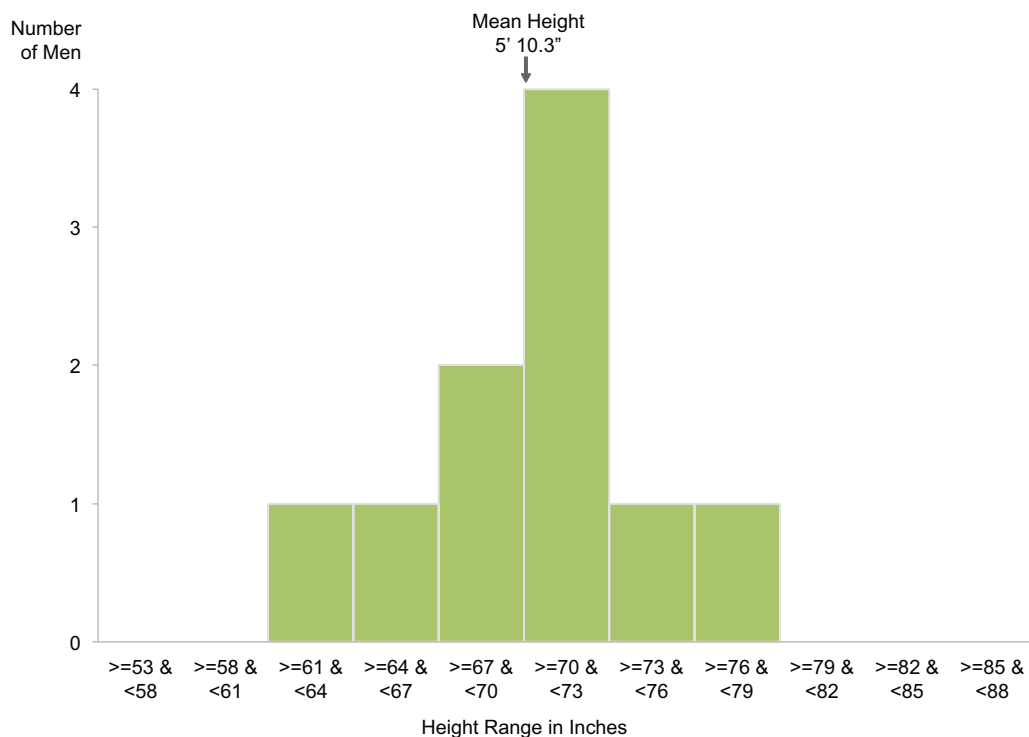
The mean height of this sample is slightly higher than the overall mean and the shape of the distribution is no longer perfectly symmetrical in the form of a normal, bell-shaped curve. This makes perfect sense, doesn't it? With only a few values chosen at random from a range of possibilities, a statistical summary derived from the sample will likely differ from that of the overall population.

Another simple illustration of the effect of sample size involves coin tosses. Imagine that we have performed 10 experiments to find out the percentage of coin tosses that come up heads. Rather than tossing the coin the same number of times in each experiment, such as 100 times, however, imagine that the number of coin tosses ranged from 2 in one experiment to 1,000 in another. Here are the results of the ten experiments:

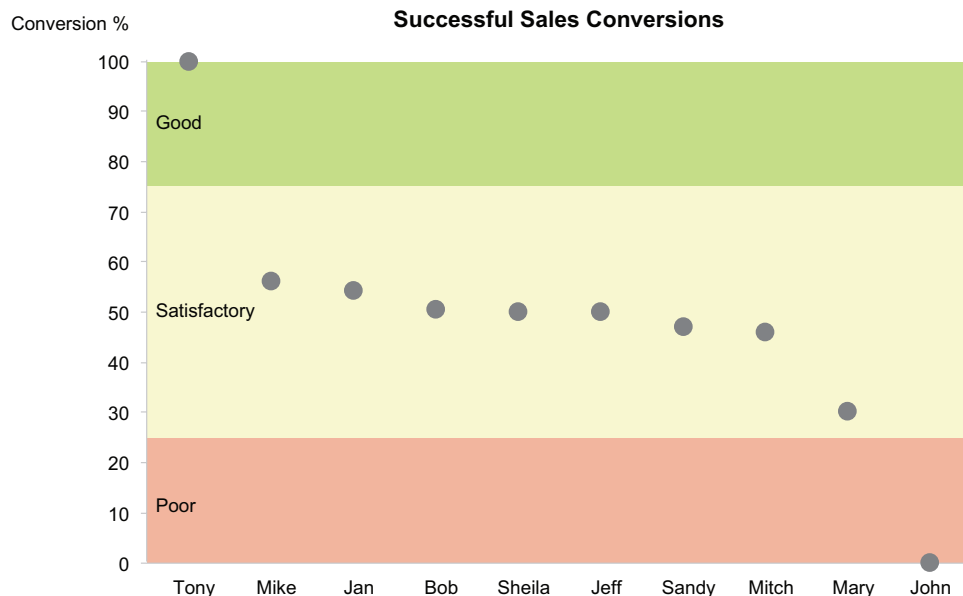| Experiment	Number of Tosses	Number of Heads	Percentage of Heads
1	2	1	50.0%
2	2	0	0.0%
3	2	2	100.0%
4	10	3	30.0%
5	10	5	50.0%
6	100	54	54.0%
7	200	92	46.0%
8	400	224	56.0%
9	500	236	47.2%
10	1,000	505	50.5%
	2,226	1,136	50.4%

It isn't surprising that heads came up 0% of the time in experiment #2 or 100% in experiment #3 when the coin was only tossed twice in each. The slight degree of variation from the expected mean of 50% heads for 2,226 tosses is routine. If we found results of 0% or 100% heads in an experiment involving 10 or more coin tosses, however, we'd recognize that as extraordinary—increasingly so as the number of tosses increases.

Now imagine that, rather than 10 coin-toss experiments, we're examining 10 salespeople to see how successful they've been in converting sales opportunities into actual sales, but the opportunities the salespeople were given varied to the same degree as the coin-toss sample sizes above: three salespeople were given only 2 opportunities each and the others were given from 10 to 1,000 each. Under these circumstances, would it be appropriate to compare their performance in the following manner?
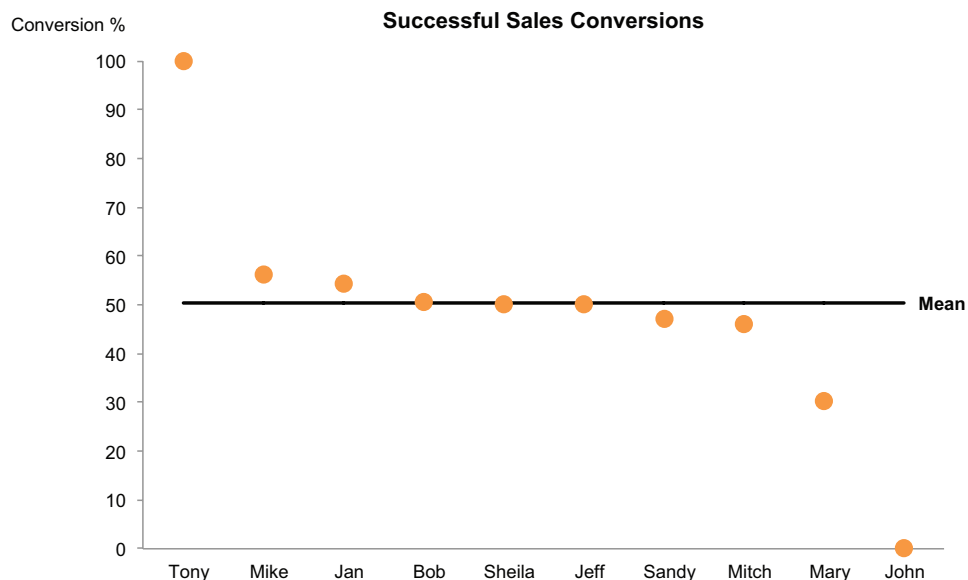
This isn't a fair comparison. Tony's 100% conversion rate and John's 0% conversion rate were both based on only two opportunities each. Mike's 56% conversion rate indicates that he failed 44% of the time: 176 failures out of 400 opportunities. For all we know, Mike might have failed on each of his first 20 attempts, so is he necessarily more successful than John?

The point we're making is that you must be careful when comparing statistical measures from samples that differ significantly in size. Even though it's obvious that the comparison of salespeople above is misleading and harmful, comparisons like this are made all the time. In fact, they're often designed to look something like this:

Successful Sales Conversions

Conversion %

- Good
- Satisfactory
- Poor

Tony Mike Jan Bob Sheila Jeff Sandy Mitch Mary John

Assigning ranges of acceptable variation either arbitrarily or based on a statistical measure such as the mean without taking wide-ranging sample sizes into account assigns significance to data that isn't justified. The nature of random variation must be considered.

Randomness and Unfair Comparisons

I doubt that statistical thinking comes naturally to anyone; it must be learned. Even in the field of healthcare, despite valiant attempts to base practices on evidence-based, statistical reasoning, progress is slow and arduous. You might assume that healthcare professionals—doctors, administrators, policymakers, etc.—are well trained in statistical reasoning, but this is, by and large, not the case.

Although the subject of statistics now appears to some extent in medical school and healthcare focused curricula, most of the courses are introductory and organized around mathematics and calculations rather than the critical appraisal and interpretation of data.

This lack of training, coupled with our collective desire to rank and identify the "best" of everything, is a dangerous combination. Inadequate sample sizes with wide ranges of results often make it difficult and unwise to rank results that we may be inadequately prepared to interpret and explain; results that can lead to undeserved praise, unjust punishment, and bad decisions.

Consider the following display (a caterpillar plot) of actual healthcare data. Each data point represents a hospital, with 260 in total. The values are mortality rates following surgeries and the solid horizontal line represents the mean.
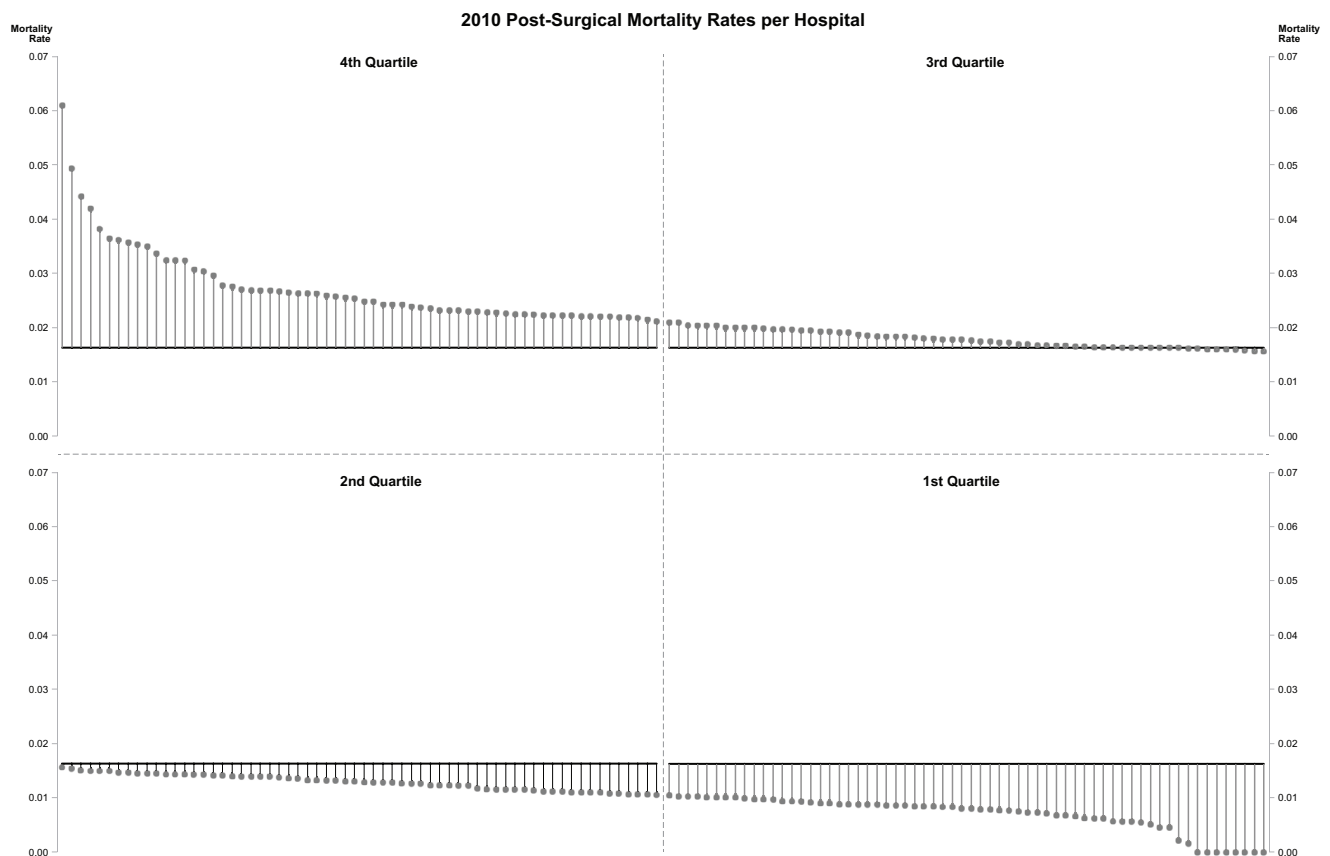
The sample sizes (number of surgeries reported by each hospital) on which this graph is based range from 7 to 3,151. For this reason it isn't appropriate to rank the hospitals by mortality rate. The ranking suggests a relationship of relative performance that cannot be determined by the data.

Imagine that, unlike the anonymous version of the graph above, each data point is labeled with the hospital's name. Can you hear the screams of surgeons from the hospitals with the highest mortality rates in the 4th quartile section of the chart? Can you see the puffed up egos of the surgeons who work at the hospitals with mortality rates of zero in the 1st quartile section? One of the hospitals with a zero mortality rate provided a sample of only seven surgeries.

[Note: You might realize, especially if you work with healthcare data similar to the example above, that a fair comparison of hospitals would require the data to be adjusted not just to account for varying sample sizes but also for varying levels of risk. Some surgeries are more risky than others and some patients, due to varying levels of illness, are more at risk than others. We'll ignore this for now to keep the example simple.]

Funnel Plots to the Rescue

Before anything else, we need to make it clear that a *funnel plot* is not another name for the form of display you might know as a *funnel chart*. The latter, as it is usually designed, is an ineffective chart. It is used to show how values decrease as they proceed through various stages in a process. The purpose is meaningful, but the way it's usually displayed is silly. The most common example is the sales funnel chart, which shows how potential sales revenues decrease as they proceed sequentially through stages in the sales process, such as from initial unqualified sales leads all the way to the final orders. Because revenues decrease at each stage in the sales process, you can think it as having the shape of a funnel that narrows in size from a wide beginning to a narrow finale.

The following example is typical of funnel charts:
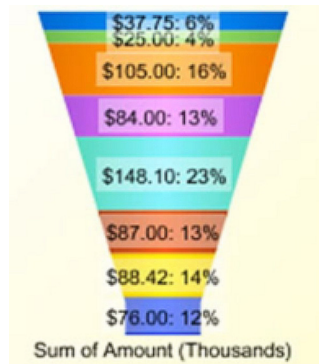
Suffice it to say that there are better ways to display decreasing values at sequential stages in a process. The funnel is a metaphor; the chart need not look like one.

The funnel plot that we're featuring in this article was first introduced by R. J. Light and D. B. Pillemer in 1984 (*Summing Up: The Science of Reviewing Research*, Harvard University Press). It is nothing more than a scatter plot that displays measures of something (e.g., post-surgical mortality rates) collected from various entities (e.g., hospitals) along the Y-axis and sample sizes (e.g., the number of surgeries per hospital) along the other X-axis, with a data point for each entity and lines to mark the boundaries of random variation. These lines start out far apart from each other on the left where sample sizes are small and converge as they proceed to the right where the sample sizes are large, which gives them the shape of a funnel. In other words, the lines mark broader boundaries of random variation for small samples and increasingly restrictive boundaries as sample sizes grow.

Funnel plots strive to distinguish random from non-random variation in distributions much as statistical process control (SPC) charts do this for time series. Typically, a funnel plot includes two sets of boundary lines: one set for 95% confidence intervals (standard error × 1.96), which SPC calls a 2 sigma limit, and one for 99.8% confidence intervals (standard error × 3), called a 3 sigma limit. Here's how a funnel plot might look without data:
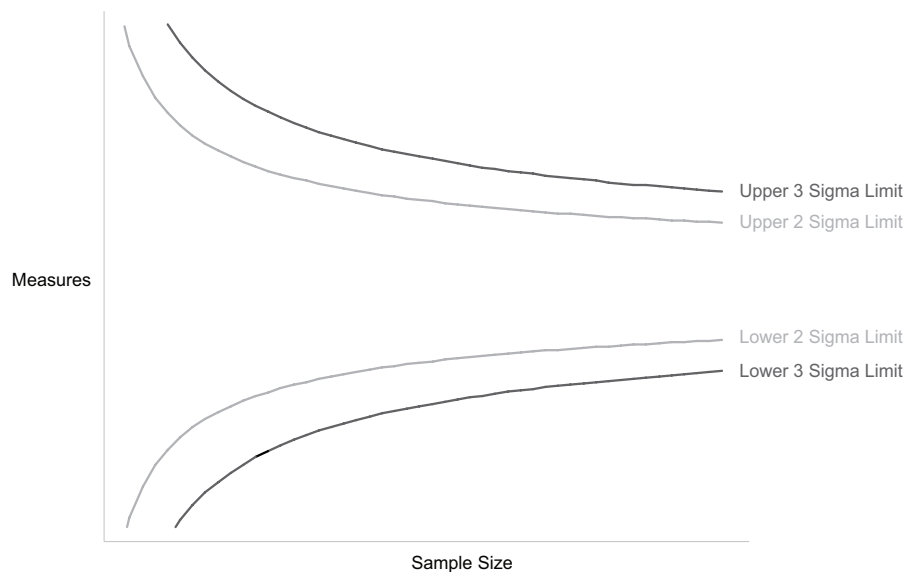

When populated with values, each data point that falls outside of the range defined within the upper and lower 2 sigma limits would represent a value that would be due to randomness only 5% of the time. Any point that falls outside the 3 sigma limits would be due to randomness extremely rarely, only 0.2% of the time.

Let's begin with a simple funnel plot of the sales conversion data that we looked at before.

Successful Sales Conversions
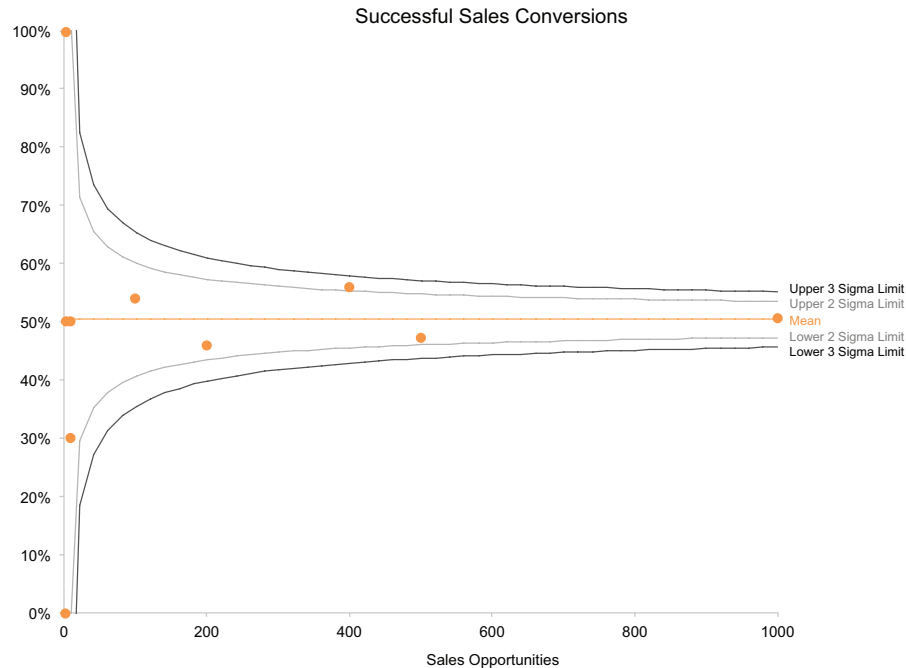
Now that the varying numbers of sales opportunities have been taken into account, we can see that no sales representative had a sales conversion rate that fell outside the range of variation that might be entirely due to randomness. Only one performed better or worse than we would expect due to random variation.

Now let's look at some data that's real and more interesting: the post-surgical mortality rates from before. Unfortunately, we can't just throw the mortality rates that appeared in the earlier chart into a funnel without first doing a little work. Funnel plots calculate the boundaries of expected variation based on the mean and standard error, but the mean as a measure of center and standard errors as measures of confidence around the mean, are statistics that assume a normal (bell-shaped) distribution. Here's what the distribution of mortality rates looks like when displayed in a histogram:
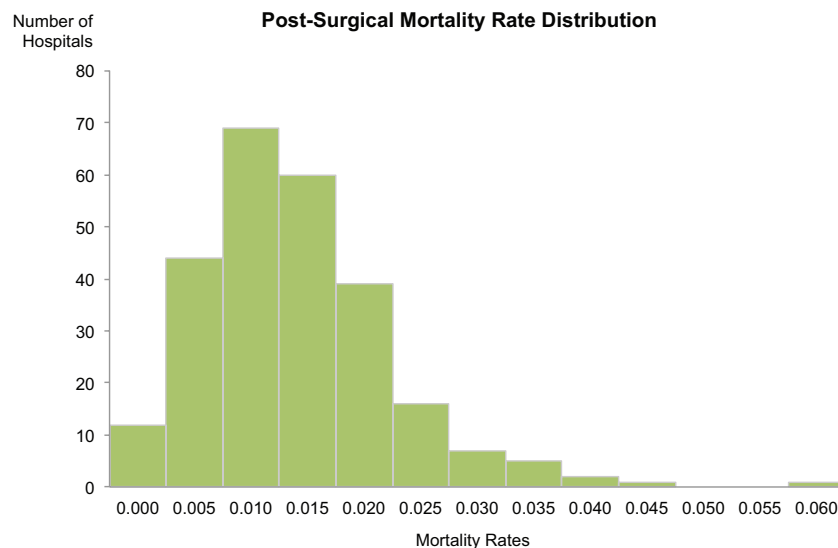

This distribution is skewed to the right toward higher values with a peak on the left. The mean and standard error will not describe the nature of this distribution in a way that will allow us to calculate meaningful boundaries of random variation. This doesn't mean, however, that we can't use a funnel plot. Statisticians routinely use statistics that assume a normal distribution to work with data sets that are not normal in shape

by transforming the values first in a way that results in a normally shaped distribution. One typical way to do this when dealing with a distribution that is slightly skewed toward the high values involves a square root transformation (a.k.a., square root transform). I've taken the liberty of transforming the mortality rates using Excel's SQRT() function, which results in the following distribution:
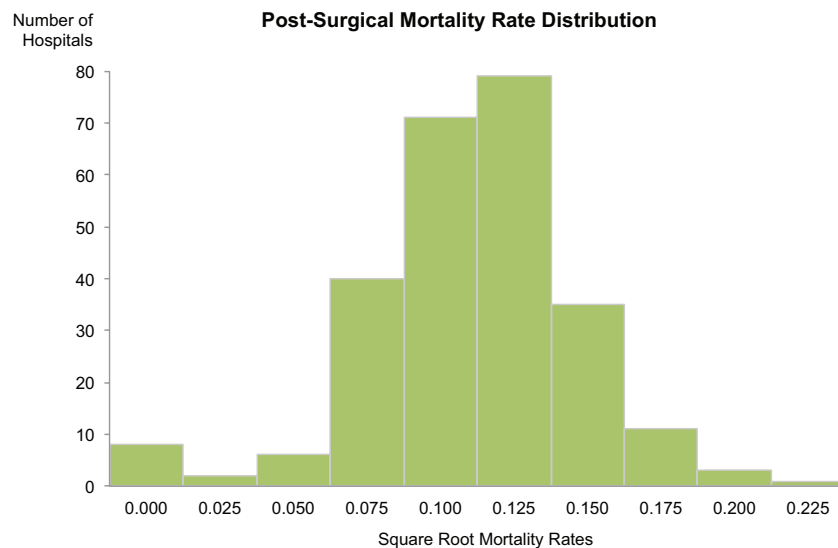

Now we can display these post-surgical mortality rates in a funnel plot, illustrated below:
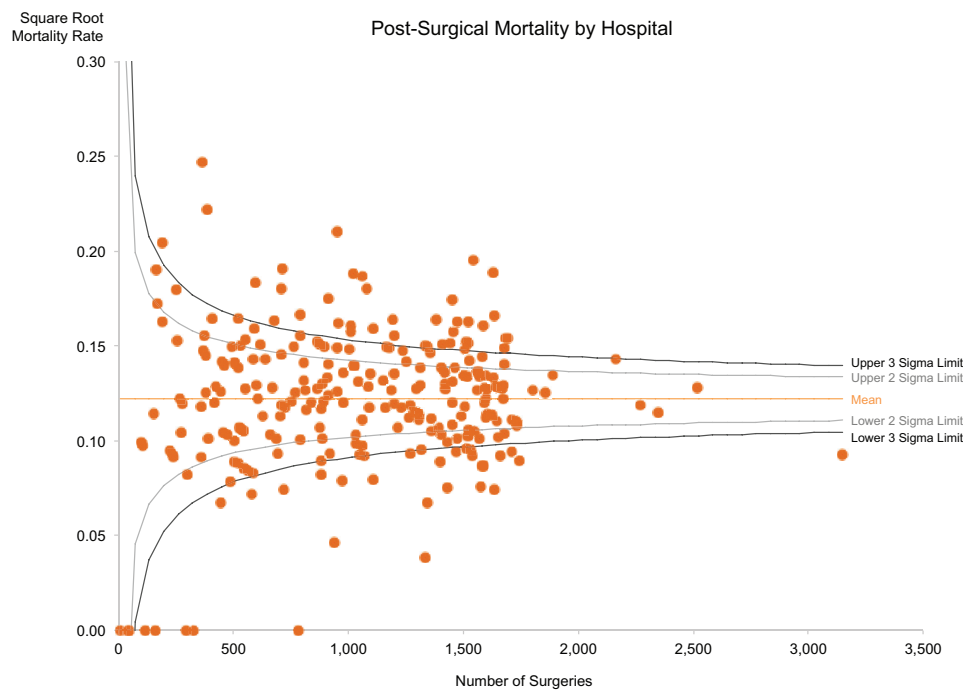
This funnel plot displays one data point per hospital, each of which shows the number of surgeries and the square root transformed mortality rate. The square roots of the mortality rates are not particularly meaningful in and of themselves, but that doesn't matter because we're simply trying to identify the hospitals that exhibit a level of performance that is outside of the boundaries of random variation and thus, according to the language of statistical process control, were probably due to a "special cause." The farther a value falls outside of the boundaries, the more likely it is due to a special cause rather than randomness.

Notice in the lower left-hand corner of the same plot below that two of the hospitals with mortality rates of zero reside within the boundaries, which indicates that we cannot rely on their low rates as significant.
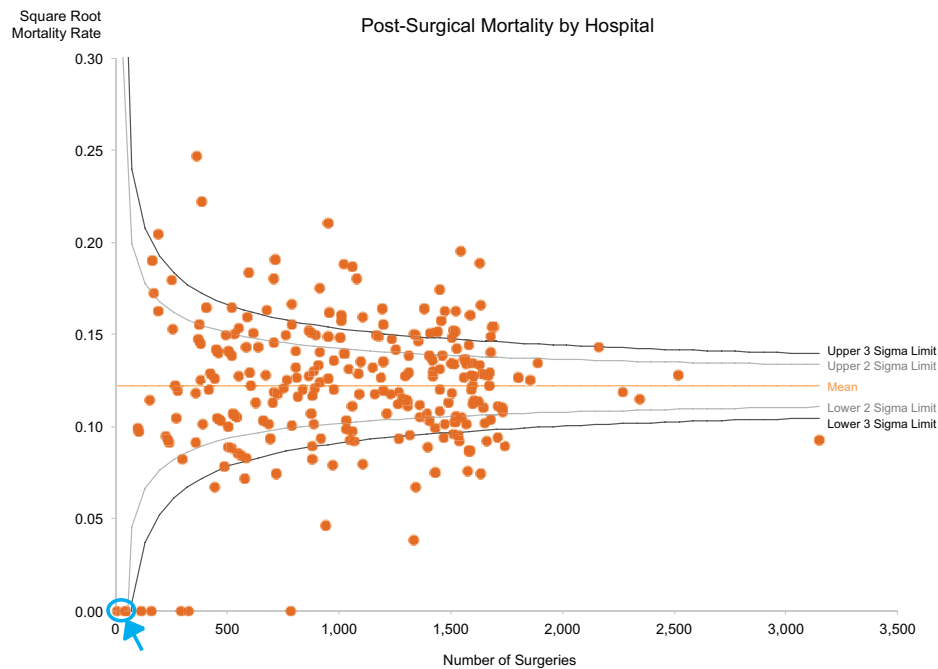
In making this comparison, we are assuming that the hospitals are homogeneous, all part of a single system. By adjusting for the varying numbers of surgeries performed at these hospitals, we are hoping to ask the question, "How is this system of hospitals doing in its efforts to prevent post-surgical mortalities?" No longer are we encouraged to compare hospital performance as a misleading ranking. We are now encouraged to focus exclusively on signals in the data. Unfortunately, however, we have not yet met our objective of comparing post-surgical mortality rates within a homogeneous system. What remains as a problem is the fact that we have not adjusted the mortality rates to account for two factors that keep them from belonging to a homogeneous system:

1. Surgeries of many types are being compared without taking into account the fact that surgeries vary significantly in mortality risk.

2. Patients of various levels of heath are being compared without taking into account that some of them went into surgery much healthier than others and were therefore at lesser mortality risk.

It is routine when dealing with heterogeneous entities such as different types of surgeries and patients of varying levels of health to adjust values to account for these factors. In a case such as post-surgical mortality, typically an expected number of deaths is calculated to account for varying levels of risk and then the observed (actual) number of deaths that occurred is compared to this, resulting in an observed vs. expected (O/E) ratio. For a completely fair comparison of post-surgical mortality among these hospitals, it is this O/E ratio that should appear in the funnel plot. We haven't included this version of the funnel plot, and if we chose to do so we would need to use a different set of calculations for the boundaries than the one that's built into the Excel Funnel Plot Template that we built for distribution with this article. The boundary calculations that we used are designed to handle proportional values: rates from 0 to 1 or percentages from 0% to 100%. Values greater than 1 or 100% won't work. Don't be disheartened, however, for the good folks at the Association of Public Health Observatories (APHO) in the United Kingdom, now part of Public Health England, have provided funnel plot templates in Excel for four types of measures:

1. Proportions and percentages (We based ours on this APHO template)

2. Indirectly standardized ratios (the O/E ratio is one of these)

3. Rates

4. Counts

In addition to the Excel templates, Public Health England provides several sample data sets and papers that describe their use of funnel plots. Take advantage of these freely available resources.

Constructing Funnel Plots

Funnel plots may be used for data of many types. For example, you might want to compare the profit margins of similar products, some of which sell few units and some of which sell many, to identify those with significantly low and high margins. Because differences in sample size (the number of units sold for each product) will have an effect on random variation, a funnel plot might handle this well. Fortunately, once you know how to calculate the boundaries of random variation (the 2 and 3 sigma limits), funnel plots are relatively easy to construct in several charting products, including Excel. As we mentioned previously, a funnel plot is essentially a scatter plot that's designed for a specific purpose. What's necessary is that the product allows you, in a scatter plot, to display several series of values as lines in addition to the one series that's displayed as individual data points.

To show how to construct a funnel plot in Excel and to provide the formulas that are required to calculate the boundary lines as well, we've created an Excel file that you can download with everything that you'll ordinarily need, including instructions. You can copy your own data into this Excel file and by following a few simple steps produce a funnel plot. Keep in mind, however, that this particular template only works for proportions (rates from 0 through 1 or percentages from 0% through 100%).

If you really want to dig into funnel plots more deeply, perhaps the best resources can be found at the website Understanding Uncertainty, which is maintained primarily by David Spiegelhalter, a professor of statistics at the University of Cambridge. On his site you'll find a wealth of material about statistical analysis, including articles and blog posts about funnel plots.

One of the downsides of producing funnel plots in Excel is the fact that Excel has always lacked the ability to attach labels to data points in scatter plots. This is a glaring omission, because the ability to easily identify individual data points in a scatter plot is often necessary. For this reason, when you hover with your mouse over a particular data point in our Excel file, only the quantitative values will appear, not the name of the hospital. Fortunately, to overcome this problem you can download a free Excel add-in called Chart Labeler.

Final Word

We hope that we've done more than just introduce the funnel plot to you, even though that alone is certainly worthwhile. The heart of our message beats more deeply than this practical chart, for it is concerned with something that is fundamental to data analysis: an understanding of variation. Because the nature of variation and of randomness in particular—a central topic of statistics—is not at all intuitive and hasn't been studied by most people who analyze data, its something that you must understand if you want to advance beyond the basics. The "discontents" in your data—items that don't fit into routine variation—are often the most important signals. Until you can separate the signals from the noise, you'll spend most of your time chasing shadows while the voices of discontent remain unnoticed.

Discuss this Article

Share your thoughts about this article by visiting the Variation and Its Discontents thread in our discussion forum.

About the Authors

Stephen Few has worked for nearly 30 years as an IT innovator, consultant, and teacher. Today, as Principal of the consultancy Perceptual Edge, Stephen focuses on data visualization for analyzing and communicating quantitative business information. He provides training and consulting services, writes the quarterly *Visual Business Intelligence Newsletter*, and speaks frequently at conferences. He is the author of three books: *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Second Edition, *Information Dashboard Design: The Effective Visual Communication of Data*, and *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. You can learn more about Stephen's work and access an entire library of articles at www.perceptualedge.com. Between articles, you can read Stephen's thoughts on the industry in his blog.

Katherine Rowell is co-founder and principal of Katherine S. Rowell & Associates (www.ksrowell.com), a Boston firm that specializes in helping healthcare organizations organize, design, and present visual displays of data to inform their decisions and stimulate effective action. She advises providers, payers, policymakers and regulatory agencies how to align systems, design reports, and develop staff to communicate healthcare data clearly. She has published numerous high-profile articles, and *Unleash Your Inner Healthcare Data*, her twice-monthly newsletter, has passionate subscribers all over the world. Her clients, drawn from leading national and international healthcare organizations, include the Cleveland Clinic, Partners Healthcare, Baylor Healthcare, the World Health Organization, and the Children's Hospital Association. They and hundreds of other companies, organizations, and individuals seek out her public and private workshops, in-house training sessions, and private consulting services to learn the best practices of data visualization for healthcare professionals and other stakeholders.