Minitab ®

5 Ways to Make Sure You Can Trust Your Data

Can you trust your data?

This deceptively simple question needs to be answered when you use statistics to improve processes. But many of us forget to ask it, or respond too quickly and confidently.

We can't just assume we have good data—we need to *know* we do. That may require a little bit more work up front, but energy invested in getting good data pays off in better decisions, bigger improvements, and greater confidence.

Here are five ways you can maximize your chances of getting data that you can have confidence in.

1: Plan How, When, and What to Measure—and Who Will Do It

Failing to plan is a great way to get unreliable data. That's because a solid plan is the key to successful data collection.

Asking *why* you're gathering data at the very start of a project will help you pinpoint the data you really need.

A thorough data collection plan is a great first step to getting reliable data.

A data collection plan should clarify:

- What data will be collected
- Who will collect it
- When it will be collected
- Where it will be collected
- How it will be collected

Answering these questions in advance will put you well on your way to getting meaningful data.
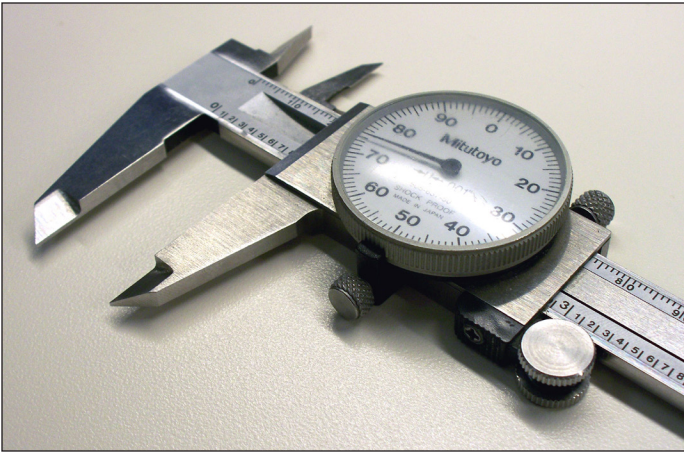
2: Test Your Measurement System

Many quality improvement projects require measurement data for factors like weight, diameter, or length and width. Not verifying the accuracy of your measurements practically guarantees that your data—and thus your results—are not reliable.

A branch of statistics called Measurement System Analysis lets you quickly assess and improve your measurement system so you can be sure you're collecting data that is accurate and precise.

When gathering quantitative data, Gage Repeatability and Reproducibility (R&R) analysis confirms that instruments and operators are measuring parts consistently.

If you're grading parts or identifying defects, an Attribute Agreement Analysis verifies that different evaluators are making judgments consistent with each other and with established standards.

If you do not examine your measurement system, you're much more likely to add variation and inconsistency to your data that can wind up clouding your analysis.

Can you trust your measurements? The only way to know is to test it.

3: Beware of Confounding or Lurking Variables

As you collect data, be careful to avoid introducing unintended and unaccounted-for variables. These "lurking" variables can make even the most carefully collected data unreliable—and such hidden factors often are insidiously difficult to detect.

A well known example involves World War II-era bombing runs. Analysis showed that accuracy *increased* when bombers encountered enemy fighters, confounding all expectations. But a key variable hadn't been factored in: weather conditions. On cloudy days, accuracy was terrible because the bombers couldn't spot landmarks, and the enemy didn't bother scrambling fighters.

Suppose that data for your company's key product shows a much larger defect rate for items made by the second shift than items made by the first. Your boss suggests a training program for the second shift.
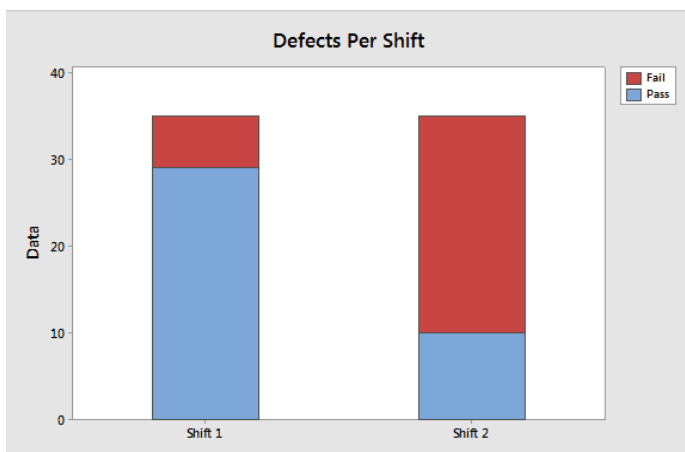
Members of Shift 2 appear to be good candidates for additional training.

But could something else be going on? Your raw materials come from three different suppliers. What does the defect rate data look like if you include the supplier along with the shift?

Now you can see that defect rates for both shifts are higher when using supplier 2's materials. Not accounting for this confounding factor almost led to an expensive "solution" that probably would do little to reduce the overall defect rate.
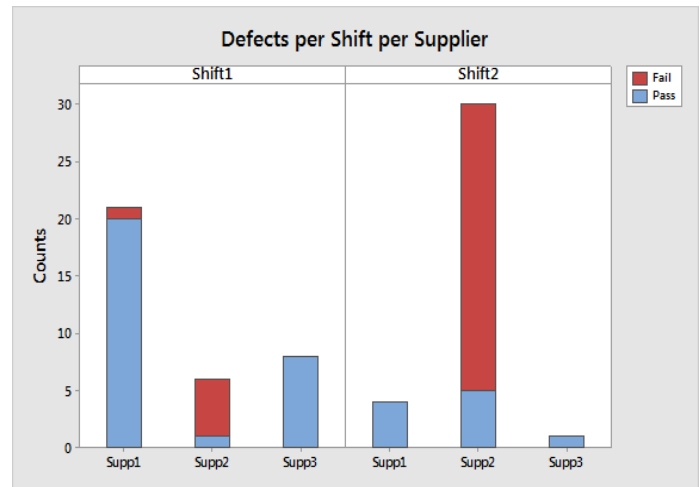
Considering just the shift would hide the influence of a factor that could be even more important to defect rates: the supplier.

4: Get Buy-In from Your Team

Even if you've been diligent about data collection methods and planning, if your team does not understand why and how you're gathering data, you can still get bad information. Fearful employees may focus on "making numbers" by fudging results or other methods, instead of getting accurate data.

If data collection is complicated and demanding, it creates many more opportunities for problems. You can encourage good data collection by making it convenient, and by aligning data-related tasks with other responsibilities wherever possible. Providing adequate training about the data collection process also will reduce the potential for errors.

The easier data is for team members to gather—and the more they understand why it's important—the better the data you'll get.

Graphing your data before you begin in-depth analysis lets you identify outliers and other anomolies in your data.
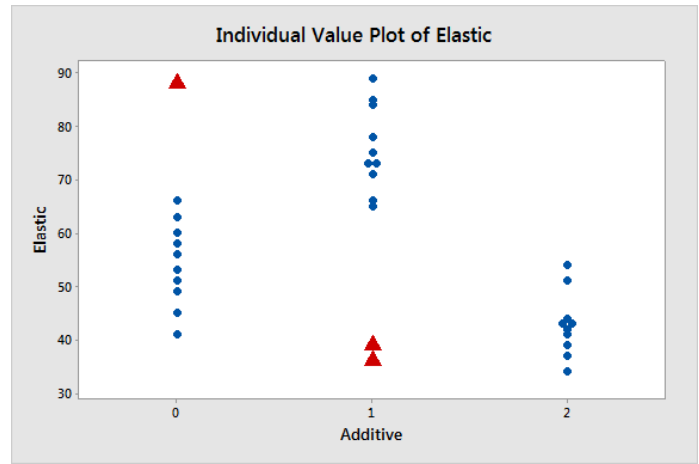
5: Do a Preliminary Data Check

Even if you've been careful when gathering data, you can obtain questionable results if you don't perform an exploratory data analysis. Check the descriptive statistics, including the mean and median values, and the standard deviation. An initial analysis usually also checks to see if data follow the normal distribution, a key assumption in many analyses, or if some other distribution is a better fit.

Graphing your data—in a boxplot, scatterplot, or individual value plot—will reveal outliers and oddities. Extreme values can have a big impact on results, so examine these carefully. If you collected your data in sequence, a time series plot also may show unexpected trends or an unusual series of data points.

Reviewing the raw data in a worksheet also helps. Sorting it by different fields can reveal data entry mistakes and inconsistencies, variations in coding, and other errors.

Failing to check your data before starting the "heavy lifting" with more sophisticated statistics can result in an analysis that requires much more time—or leads to unreliable conclusions!

To Get Data You Can Trust...

Improving quality is not easy, and nobody sets out to waste time or sabotage their efforts by not collecting good data. But as these reminders show, it's all too easy to get problem data even when you're being careful!

When you collect data, be sure to spend the little bit of time required to:

- Create a data collection plan.

- Assess your measurement system.

- Make sure your collection methods do not create bias or other problems.

- Make sure your team understands how gathering good data benefits them.

- Perform a preliminary review of the data before in-depth analysis begins.

Minitab ►®