Print

Data Collection Strategies

In the previous video, you learned about the importance of operational definitions. Another aspect of data collection is to determine how the data on the various characteristics will be collected.

Data collection strategies fall into three general categories: retrospective studies, observational studies, experimental studies. Retrospective studies are based on historical data. These are data that have been previously collected, typically for some other purpose. Historical data might be stored in a database, in a spreadsheet, or in a hand-written log. You can use historical data to describe past process performance. You can also use historical data to identify and prioritize problems.

However, you need to be careful when you use historical data for problem solving, where you are interested in identifying root causes. In order to use historical data with any degree of confidence, you need to evaluate the quality of the data and verify that the data can be used for the intended purpose.

You should be able to answer the following questions: Who collected the data? How were the data collected?

Were the data collected in a consistent manner? Has the process changed since the data were collected? Are the data relevant to the problem you are studying? Are large numbers of records missing from the data?

You might also ask questions about the measurements themselves. For example: Which characteristics were actually measured? Who took the measurements? How was the characteristic measured? Which measurement device, or gauge, was used? Is the gauge calibrated, and is the gauge capable? If you can't answer these questions, or if you find issues with the data, then you might need to collect or compile new data. Note that you learn more about potential data quality issues, and how to prepare data for analysis, at the end of the next module.

Now let's talk about observational studies. In an observational study, data are collected as they become available. Specific data are collected with the intent of using the information to gain insights into the performance of a product or process.

For example, you might measure an important output characteristic over time to study whether the process is stable. You'd plot each of these measurements on a control chart. You learn about control charts in the Quality Methods module.

Or you might measure input and output characteristics of parts as they are produced, and use this information to understand the relationship between the variables. To do this, you might capture the information in a hand-written data collection sheet or a spreadsheet.

Then you'd use statistical methods like correlation and regression to study the relationships between the variables. You learn about these methods in the Correlation and Regression module.

Observational studies tend to be more controlled and more structured than retrospective studies. You determine the data that you need to collect, and you establish operational definitions for the measures before the data are collected. As a result, these data are better suited to your project and the problem you are trying to solve. For these reasons, observational studies can provide data that are of higher quality than retrospective studies, and you can have more confidence in the conclusions that you draw from the data.

The third type of data collection strategy is to use an experimental study. Experimental studies, or designed experiments, produce data that are obtained as a result of active intervention.

You systematically change, or manipulate, a set of experimental factors while controlling other variables.

As you do this, you measure important output characteristics (or responses). Because of the controlled study design, you can use designed experiments to identify, and quantify, cause-and-effect relationships between the factors and the responses.

Here are some examples of experimental studies:

clinical trials, which study the effectiveness of medical treatments or drugs, marketing experimental research, and designed experiments on a manufacturing process. You learn how to design and analyze industrial experiments in the Design of Experiments (or DOE) module.

Statistical Thinking for Industrial Problem Solving

Close