

Part II

Analyzing Data Through Spatial Models

Key Thinking Strategy: Asking Questions

In the last unit, we focused on helping you to understand that the world is made up of information that can be organized into data. These data can then be used to help make real-world decisions. In this unit, we focus on how this data can be used to get the answers to these real-world questions. Think of this as interrogating the data in order to find out what story it tells. If we do not interrogate the data, we may find that we have collected thousands of pieces of information, have it organized nicely into a spreadsheet or a database, but have no idea as to what it all means.

In practical terms, to interrogate the data involves asking questions and finding ways to manipulate the data to get answers. We will model the types of questions you should ask, but remember, every data set comes from a different context, so every data set will need to be asked slightly different questions. In other words, we are hoping that you will start to develop a facility with the kinds of questions that need to be asked, rather than simply going down a list of questions that someone else developed. Some people may not think that this is part of mathematics; in some sense, we agree with them. Asking questions of data is more the purview of a scientist or a detective than a mathematician. However, as a data analyst, you are a detective. Keep in mind also that there are no "right" or "wrong" questions to ask. There are questions that will take you further than others, but once you get in the habit of asking questions, you will see which questions are productive and which questions are not.

After you have asked the questions comes the part most people would consider to be mathematical: getting the answers. Usually, we will need to compute some quantity or quantities. We may carry out these computations by hand, by calculator, or by routines in Excel. This is the part where we have right and wrong answers. For example, if we decide that the mean (a type of average) is the tool we want to use to answer a question about the data, then there is one and only one way to compute the mean; if we do not compute it correctly, our answers from that point forward will be incorrect because they are built on a mistake. We hope to help you avoid these mistakes, but you should always go back and double-check your work. Another useful way to check your work also comes to us from police work: corroborating evidence. Finding the suspect's finger prints at the crime scene is helpful, but does not prove that the suspect was in that location when the crime happened. Finding a witness who saw the suspect enter the location at the same time the crime was committed strengthens the case. Finding that the suspect had a motive for committing the crime pretty much seals the deal. Once the police can establish motive, means, and opportunity, they consider that they have enough evidence to arrest the suspect for the crime. One calculation that supports an answer is okay, but several different quantities or representations derived from the data that lead to the same conclusion (maybe a calculation of the mean, a boxplot of the data, and a histogram of the distribution of the data) make for a much stronger case.

This idea of corroborating evidence is a little different from the way a scientist would approach the problem. A scientist typically collects data from an experiment. This experiment can be recreated and rerun several times. When all the data from each of the runs of the experiment are compared, the scientists can be satisfied they are on the right path if the results are all nearly the same. For business management data, it may be impossible to recreate the data collection method in order to get new data to compare with the old data.

Conditions will change too much between attempts to gather data, or it may be too costly. Thus, rather than looking for multiple sets of data that agree in order to settle on one set of numbers (the scientific approach), you should look for different kinds of evidence that shed light on the same question (the detective approach). This technique is called triangulation and is very common in fields of study such as management, education, and psychology.

It is critical that you also understand that this process is not a linear process. Just because we have answers to the questions we have asked does not mean that we stop. Usually, the answers will lead us to more questions. Sometimes, these answers will lead us back to the beginning and require us to design a method for collecting additional data on the situation.

