* Q13.7 – There is a strong linkage between *statistical data analysis* and data mining. Some people think of data mining as automated and scalable methods for statistical data analysis. Do you agree or disagree with this perception? Present one statistical analysis method that can be automated and/or scaled up nicely by integration with current data mining methodology.

I believe that data mining can be a very useful tool that can be added to a statistical analysis “toolbox”, as well as the inverse with statistical analysis tools and data mining “toolbox”. There are numerous concepts and topics discussed throughout this book that help and aid in statistical analysis tremendously. These tools significantly reduce the time required to collect and analyzing data, while still maintaining the integrity of the original data set. This is vital because statistical analysis typically focuses on a sample of a population’s data; therefore, you want to ensure all of the data that is being reviewed is truly representative of the population you are analyzing.

One statistical analysis method that can be automated and scaled by integrating with data mining methodology is clustering and logistic regression / classification. Data mining can take vast amounts of incoming data and create clusters and nodes based on the concepts described throughout this book. This process can be automated and scaled as more is entered into the dataset. It can also be tweaked as needed relatively easily. This is becoming more important with the increasing popularity of social networks and the further dependence on electronic data storage / networks. As a result, statistical analyses get a reliable source of continuously streaming data sets that are a good representation of the population they are studying.

* Q13.9 – Propose a few implementation methods for *audio data mining*. Can we integrate audio and *visual data mining* to bring fun and power to data mining? Is it possible to develop some video data mining methods? State some scenarios and your solutions to make such integrated audiovisual mining effective.

Audio data mining can be used in conjunction with visual data mining to further explore complex data for unique trends. As long as humans will continue to interact with data and computers, there will need to be new and inventive ways to display complex data sets between human and computers. The more senses that can be used to view data, the better it can be understood by humans. Data can be turned into sound/music via pitches, rhythms, and tunes. This audio representation of data can be used in combination with visual representations/charts to create a video of sorts that humans can view and look for anomalies. This can be useful in a wide range of applications with huge amounts of data. Doctors/scientists can listen to the genetic codes for anomalies/patterns they may not see visually. Additionally, I have read about a blind scientist using audio data to analyze astronomical data sets in regard to exoplanet hunting that contain millions of data points. Hopefully, in the future we can further expand to all human senses for further data analysis.

* Q13.10 – General-purpose computers and domain-independent relational database systems have become a large market in the last several decades. However, many people feel that generic data mining systems will not prevail in the data mining market. What do you think? For data mining, should we focus our efforts on developing *domain-independent* data mining tools or on developing *domain-specific* data mining solutions? Present you reasoning.

I don’t believe that one specific path should be chosen. Both of these choices (specific and independent) can result in breakthroughs to benefit any and all applications. One is looking at the more general, broad view of data mining/collection, which will be vital to some specific applications that may not be prioritizing how to improve data collections for their specific-domain. However, specific domains may come up with some vital breakthroughs as a result of taking a specialized view of a specific data set (i.e. analyzing the human genome). Concepts that are important to this application may also have significant impact across all domains and the data mining field in general. Therefore, I believe that both views are important to ensure consistent innovation is maintained as we move into the future.

