Judge's Perspective

2016 Moody's Mega Math Challenge

Kelly Black[1], Ph.D.

Department of Mathematics, University of Georgia

Introduction

The questions in the 2016 Moody's Mega Math Challenge (M^3 Challenge) asked student teams to examine the economic advantages of different car sharing programs. The first question required the teams to determine what proportions of drivers can be classified as low, medium, or high in terms of the amount of time they use their car and the distance they drive. The second question required the teams to examine four different ways for a company to implement a car sharing program and determine the participation rates for each program. The final question required students to adjust their results for the second question and accommodate the inclusion of self-driving cars as well as cars that make use of alternative energy sources.

The models that student teams submitted this year tended to be simpler compared to previous years. The primary difference was that the teams tended to provide more analysis and insight into their models. This is a welcome development, and it is encouraging to see greater attention paid to this important aspect of modeling.

Modeling is a recursive practice. We generally start with simple models followed by close introspection and analysis, and we then follow up with small changes and additions to our models to account for unforeseen behaviors. This cycle is generally repeated until a model has been more fully developed and something new comes along to catch our eye.

More specific observations with respect to this year's Challenge are given in the commentary that follows. The first three sections focus on each of the three questions in order. After the three sections for each question, an additional section is given that provides some general notes about modeling.

Question One

The first question focused on driver behavior for the overall population of the United States. Students were asked to determine an approximation for the distribution of drivers with respect to how long they drive in terms of both time and distance. The vast majority of teams made use of publicly available data sets and used a straight forward statistical measure to decide the proportions of people within each group. In this section I will discuss the issues that arose with respect to using the different data sets, the difference between commuting and other short trips, deciding limits of the three specified driving categories, and using regression to estimate distributions.

First, a variety of data sets are available that are used to gain insight into the driving habits of people living in the United States. Of the data sets that are available, a data set published by the United Census Bureau (reference one) and the National Household Travel Survey (NHTS) (reference two) were commonly used. A number of teams made use of the summarized results published by the Automobile Association of America (AAA) (reference three). The way in which the teams interpreted the data was influenced by which data set was used. The different resources led to different kinds of insights into driving habits within the United States.

One of the difficult issues with the use of the NHTS survey is that the data set is extremely large. Many teams were not able to read the data into a spreadsheet. Some teams dealt with this difficulty by simply truncating the data set, and other teams constructed a random sample from within the data set. Both of these approaches are valid given limited resources and tight time constraints, and the primary issue from the judges' perspective was how well the teams documented and described their process. I personally did not see any papers that used a bootstrapping method using the data set, which would be one possible refinement.

One important aspect of the data sets is how the teams interpreted the information contained within them. For example, the NHTS data set contained a much wider variety of information. It included a list of trips for each person surveyed, and it contained the time and distance for each trip. The data set also contained a separate file with the distances and times for each person's daily commute. A large majority of teams used the commute distances, but with respect to the use of shared vehicles, the collection of short trips may have been more appropriate.

Regardless of which data set was used, a large number of teams simply used commute times and distances to characterize people's driving habits. This is somewhat problematic in that people driving from suburbs to other areas are likely to make a trip to work, stay in their workplace for a long period, and then return home. On the other hand, if you focus on individual trips that are not associated with commuting to work, then shorter trips appear at a higher frequency. A person who is trying to decide whether to use a car sharing program may make very different decisions when comparing their commute to work versus running errands.

An example of the difference between the two kinds of trips is shown in Illustration 1 and Illustration 2. Both sets of data make use of the data files from the NHTS. The first illustration includes the histograms for both kinds of trips. The second illustration includes the boxplots for both kinds of trips. Both types of trips were truncated to only include trips whose distance was less than 80 miles. The two illustrations demonstrate that the daily trips in the data set include more short distances, while the daily work commutes include more medium length distances. The type and length of the trip may impact a person's choice to use a car sharing program, and the two different distributions may lead to very different conclusions.

Once a team selected a data set and decided which aspect of the data set to use, they faced another decision. They had to make a decision about where to delimit their cut off points between low, medium, and high times and distances. A wide variety of approaches were used to make this important decision. Many teams simply looked at histograms and tried to "eyeball" good cut offs. Other teams used rates from insurance companies to decide the divisions between low and medium times and distances. It was not uncommon for teams to simply use multiples of the standard deviations in the samples or just use simple quantiles such as an even 1/3 of each group.

Any one of these approaches is valid. The primary question was how well the teams described their choices and whether or not they gave a good justification for their choices. For example, some

teams made their decision based on how the distribution looked. Some teams simply said they "used logic" to determine the divide between low, medium, and high levels, and provided little to no insight into their decision process. Teams that discussed the shape of the distribution and explained how certain features such as the locations of certain peaks and low points in the distribution led to different divisions let us know how they made their decisions and also demonstrated an understanding of the information contained in a histogram.

Finally, once the breaks between low, medium, and high levels for the trip times and the distances were identified, the students then had to decide how to determine the relative frequencies in the distribution when the two aspects are brought together. The vast majority of teams simply used the data to determine the frequencies in each group. A small number of teams determined frequencies for particular values and then calculated a regression curve using a high order degree polynomial. Their next step was to use a regression fit to determine percentages based on area under the regression function. This approach is problematic with respect to the behavior of high order polynomials for large values in the domain. It is also a step that required a good deal of time and effort when the data itself is available and offers a more appropriate way to approximate the required ratios.

The final step in the model required the teams to determine an approximation to the percentages of people in each group with respect to both time and distance. The students had to decide how to handle the relationship between time and distance. The large majority of students simply assumed that the two were independent. This is likely not a good assumption, but based on the time restrictions and the difficulties with working with a large data set it is a perfectly valid first step.

The issue that arose is that a large number of teams assumed independence but did not state this assumption. Teams that explicitly acknowledged that their calculations assumed independence between time and distance and noted the potential problem were given wide latitude in this aspect of the model. A small number of teams assumed that the two aspects were independent, presented their results, and then provided an analysis of their data to show why the assumption was problematic. Teams that were able to justify why the independence assumption was problematic and stated that it is something that should be changed in a future update demonstrated a good understanding of the modeling process and showed us the steps that are appropriate in evaluating and testing a model.

Question Two

The second part of the Challenge required the teams to examine the case for a business to establish a ride sharing program in a given city. In particular, four different business models were given. In this section I will comment on the different models that were used, the calculations of costs, the difference between profits and revenues, competition, consumers, and establishing a benchmark for a model.

We first look at the business case for ride sharing. Four different program types were given in the problem statement, and the teams were asked to evaluate each of the four program types. Several teams, however, evaluated only two or three of the program types, rather than all of them. Given the time constraints, some judges felt that it was not unreasonable to focus on fewer programs. However, the judges' reactions to this were mixed and they were more likely to accept this when the decision was well justified. Therefore, of the teams that chose not to evaluate all four program types, those that provided strong arguments for examining a subset of the programs made a better impression than those that didn't.

An additional issue that came up for the different programs is that the way a business might implement one of the approaches may vary. Different teams had subtle changes that they assumed in constructing their model for the second question. The teams that clearly stated and discussed their assumptions made a clear case for the development of their model, and teams that just assumed that the reader understood their assumptions were at a disadvantage.

Another set of issues that came up was how to deal with the business aspect of implementing each program. Some teams focused on the costs to the company. Their assumption was that a program that had lower costs would be a more attractive option for a company to implement. The problem is that simply having lower costs is not necessarily the best option for a business to implement.

Related to the consumer costs, many teams focused on revenue rather than profit. To make matters worse, there were many papers in which revenue and profit were conflated. Those papers could be difficult to examine and required the reader to keep careful track of what was being tracked and calculated. Teams that were more careful to note the difference and maintain a consistent approach made for an easier paper to examine and assess.

The issue of competition was one aspect that a large number of teams ignored. The few teams that recognized that consumers had other options such as taxi service or mass transit immediately stood out. These teams demonstrated a broader understanding of the context and also were better able to calculate a wider variety of options for consumers.

In terms of consumer options, a large number of teams ignored the costs and benefits for the companies implementing the programs and focused instead on the costs to the consumers. The underlying assumption is that consumers are rational and would be able to identify their total costs for each plan and act accordingly. Unfortunately, consumers are not always so smart about how they act. Also, ignoring the costs and revenues to the companies is problematic in terms of their decision as to whether or not to implement a program.

Finally, one last aspect is discussed. The second question required the teams to make predictions for four specific cities. A small number of teams used the model that they constructed and used it to make predictions for cities that already had companies providing ride sharing programs. They compared their results with information they were able to find from the companies themselves and established benchmarks as a test of their model. Regardless of the outcome of the comparison, this is an impressive way to demonstrate how to test a model and decide the efficacy of an approach.

Question Three

Question three required the teams to make a change to the model they developed in question two. For the third question the teams were asked to evaluate the potential impacts for companies that make use of self-driving cars and cars that use alternate fuels. In this section I will comment on the two aspects of this part of the Challenge, how teams responded, and the overall effect with respect to the judging.

A large number of teams only considered one of the two aspects for the third question. That is, many teams just considered the impact of self-driving cars or just considered the impact of alternate fuel vehicles. A smaller number of teams considered both aspects. The question explicitly asked about both aspects, and teams that were able to adapt their model to accommodate self-driving cars

and alternate fuel cars had an advantage for the third part of the Challenge.

In terms of the approaches, the large majority of teams took their models from the second question of the Challenge and made small changes to adapt to the new situation. It was common for teams to make relatively small changes and then use the new model to establish new predictions. On the whole, this was not a part of the Challenge that generated more detailed models or analysis.

This was somewhat of a surprise to the judges since we generally expect that the third question will be the part of the Challenge that will be the most difficult and create the greatest opportunity for diverse and creative solutions. We quickly discovered that this was not the case this year, and we adapted our expectations by placing a greater weight on the teams' approach to the *second* part of the Challenge.

General Notes

In the previous sections my reactions to how students responded to the three parts of the Challenge are given. Here I will examine some of the more general issues about the students' entries, generally focusing on issues in modeling and writing. I will discuss the issue of describing how decisions are made; the use of acronyms; how to present and describe tables; the issue of significant figures; and finally, the role of sensitivity analysis.

First, the process of modeling requires teams to make decisions about what is important and what is not important. This year it was more common to come across papers in which the justification for a decision was simply described as making use of "logic" or some other non-descriptive term. There are many ways to justify a decision, and even a choice that seems obvious or simple should be and can be clearly described. There are essentially four levels of justification: none, the use of an explicit description, the inclusion of a reference, or the use of a description plus a reference. The first level is least desirable, and a reference should always be given if one is available.

Another thing that is seen each year is the use of acronyms, but this year it was more common than usual. Student teams made use of data that was provided by various government organizations, provided by the Automobile Association of America, or provided by a consumer organization. The result was an alphabet soup of acronyms, with each student team having their own preferred set of acronyms. It is easy to understand how students working together and focusing on a problem can fall into the habit of using acronyms, but when writing their report it is important to either define them or simply not use them. It is also important to keep in mind that different people have different views on them. Some people hate them and would rather that they not be used in a report, while others are used to them and accept them as a convenient shortcut.

Another issue that comes up every year is the annotation and description of graphs and tables. In this year's Challenge a common way to present a model was to use a table. For example, the first part of the challenge is fairly explicit in asking for a model that is best described using a table, and it was common to see tables in other parts of the reports this year. The difficulty is that different teams used different formats for their tables, including using colors, different orders, and different arrangements of the rows and columns. Teams that were able to clearly state their table and then clearly describe their table within the narrative had a distinct advantage. It made it much easier for the reader to figure out what information was in the table and how to read the table. A number of teams simply listed the table and assumed that the reader would find it and then understand it.

Tables are used to present a combination of numbers in an organized way, and related to the presentation of numbers is the issue of significant figures. This is another aspect of presenting information that comes up every year. This year many teams presented large tables filled with numbers. Teams should be careful to fully describe each table but should also be careful in presenting the numbers in the table with an appropriate level of precision. A set of tables with each number presented to the nearest 0.000001 is difficult to read, and it is also an inappropriate way to share information.

The last issue discussed here is the use of a sensitivity analysis. Last year was the first time that we saw large numbers of papers that included a sensitivity analysis. This was an exciting and welcome development. It is important to perform a detailed analysis of a model. An analysis should include identifying what is good about a model and what parts are missing (strengths and weaknesses). A sensitivity analysis, however, is different. A sensitivity analysis includes a formal approach to determining which assumptions or which parameters have the greatest impact in the conclusions that can be inferred from a model. For example, a common approach is to make a small change to one of the parameters, determine how the results of the model differ from the original model, and compare the differences for each parameter examined.

It was more common this year for students to include a section titled "Sensitivity Analysis." Unfortunately, in many cases students performed actions that were not a sensitivity analysis. It is worse to mislabel something as a sensitivity analysis then to just not do it. It is also important to recognize that a sensitivity analysis is a comparison between different parts of a model. It is not about deciding whether or not some part of a model is "sensitive," but rather identifying which parts of a model are more sensitive than the other parts.

Conclusions

In this year's Challenge the teams were asked to identify how to classify driving habits and then determine the distribution of people in the United States who fall within the different driving categories. The students were then asked to determine the relative advantages of four different business models and determine which business models would be best for four different cities.

Each of the three parts of the Challenge required teams to overcome different hurdles. The first part required students to examine a complicated data set and make decisions about how to present their interpretation of the information in the data set. The second part required students to bring together the issues of cost, revenue, profit, and consumer behavior to establish the viability of four different business practices in four different cities. The third part required students to adapt their model to include new technologies that have the potential to substantially change the way people get around.

There were a couple of differences in the Challenge problem this year. First, the judges initially underestimated the difficulty of the second part and overestimated the difficulty of the third part of the Challenge. Once we became aware that the second part was actually more difficult than the third part, we changed our weightings to accommodate the teams' responses. Another difference is that we saw more teams submitting simpler models but providing more sophisticated analyses of their models. This latter development is an indication of a growing maturity in the way the teams' advisors support their students, and it is a development that is quite exciting.

Once again, the students did a magnificent job of putting together models and providing an analysis. Every year we are excited to see what the students can do, and once again we were impressed by

what they can achieve in a very short time. This is a reflection on the efforts and dedication not only of the students, but also of the coaches, advisors, and teachers for whom we continue to be grateful and humbled by, as they are often the key to making this event a success every year.

Acknowledgments

Bibliography

"Commuting (Journey to Work)." The United States Census Bureau. Accessed April 2016, http://www.census.gov/hhes/commuting/

"National Household Travel Survey." United States Department of Transportation, Federal Highway Administration. Accessed April 2016. http://nhts.ornl.gov/documentation.shtml

"American Driving Survey Year One." American Automobile Association. Accessed April 2016, https://www.aaafoundation.org/american-driving-survey-year-one
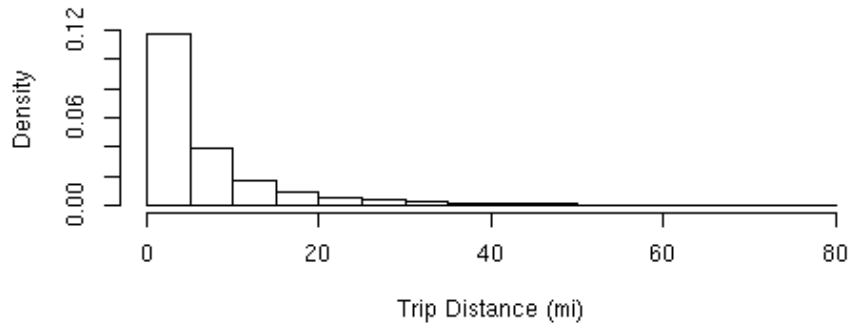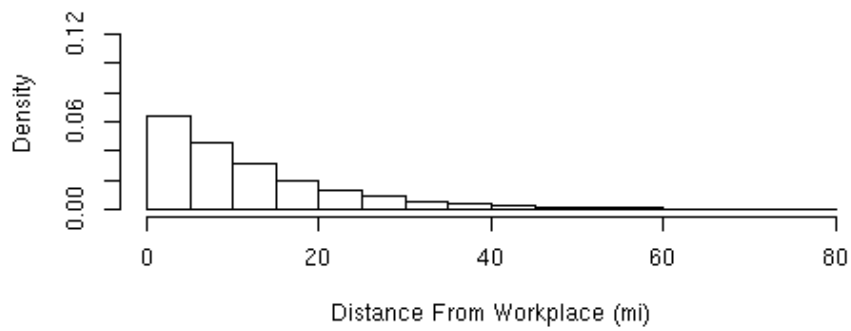
Illustration 1: Comparison of the NHTS data between commuting distance to work versus trip data. The top histogram represents the data for trip distances in the survey. The bottom histogram represents the data for commute distances for each person in the survey.
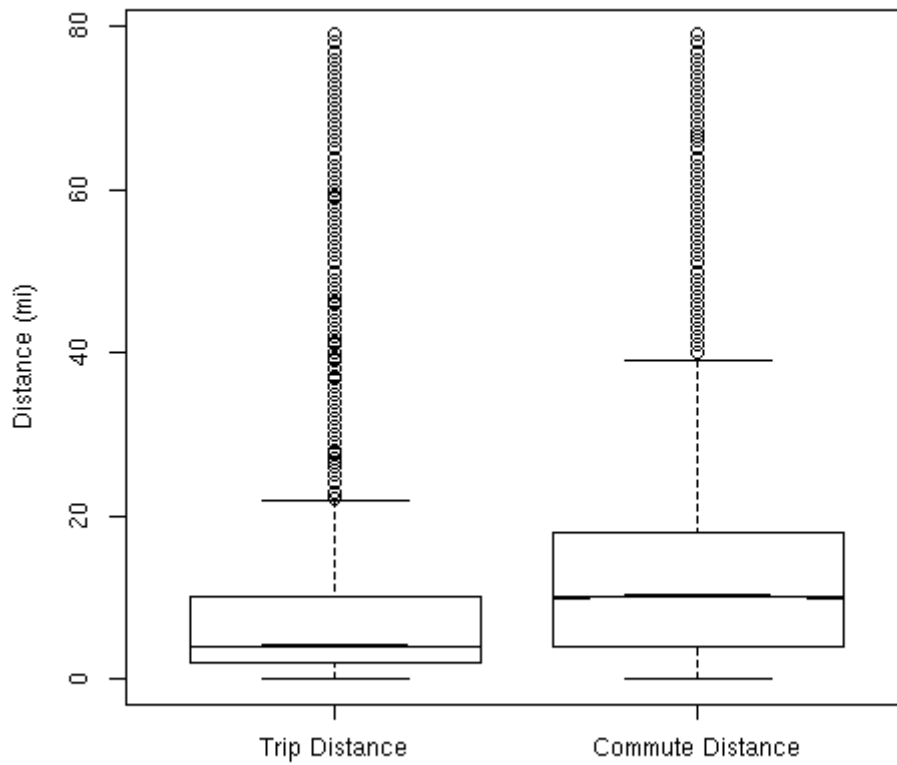
Illustration 2: Comparison of the NHTS data between commuting distance to work versus trip data. The top box plot represents the data for trip distances in the survey. The bottom box plot represents the data for commute distances for each person in the survey.