HIGH SCHOOL MATHEMATICAL CONTEST IN MODELING OUTSTANDING PAPERS

HiMCM

The contest offers students the opportunity to compete in a team setting using applied mathematics in the solving of real-world problems.

November 2013

Additional support provided by the National Council of Teachers of Mathematics (NCTM),
the Mathematical Association of America (MAA),
and the Institute for Operations Research and Management Sciences (INFORMS).

Editor's Comments

This is our sixteenth HiMCM special issue. Since space does not permit printing all eight Outstanding papers, this special section includes abridged versions of two papers and summaries from the other six. We emphasize that the selection of these two does not imply that they are superior to the other outstanding papers. We also wish to emphasize that the papers were not written with publication in mind. Given the 36 hours that teams have to work on the problems and prepare their papers, it is remarkable how much they accomplished and how well written many of the papers are. The unabridged papers from all National and Regional Outstanding teams are on the 2013 HiMCM CD-ROM, which is available from COMAP. ❏

Contest Director's Article

William P. Fox
Department of Defense Analysis
Naval Postgraduate School
Monterey, CA 93943
wpfox@nps.edu

We should all celebrate since the High School Mathematical Contest in Modeling (HiMCM) completed its sixteenth year. It is and continues to be a fantastic endeavor for students, advisors, schools, and judges. We have had schools ask for speakers to come in and discuss the modeling process so that their teams can improve and compete. We hope this trend continues. The mathematical modeling ability of students, and faculty advisors, is very evident in the professional submissions and work being done. The contest is still moving ahead, growing with a positive first derivative, and consistent with our positive experiences from previous HiMCM contests. We hope that this contest growth continues. Figure 1 is a plot of the growth over time. The trend over the last few years has been an exponential increase.

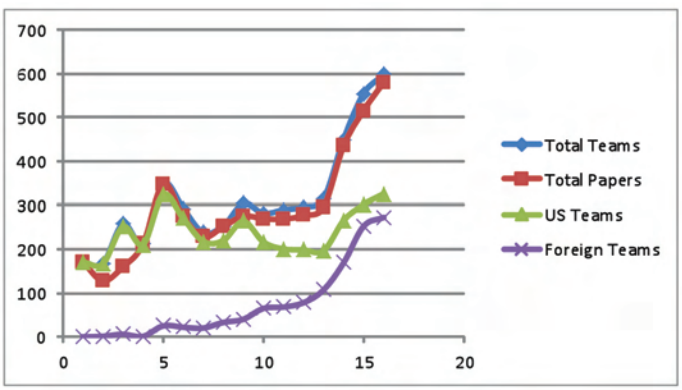

Figure 1: *Number of HiMCM teams vs. contest year*

This year the contest had 598 teams and 581 total submissions. This represents an increase of about 7.5% over last year. We had 326 U.S. teams and 272 foreign teams, representing 11.1% and 6.9% growth, respectively. In the United States, these teams represented 68 schools and 26 states. China represented about 93% of the foreign entries. Of the 2291 students, 835 or almost 36.4% were female students. The breakdown was 835 female, 1448 male students, and 8 unspecified. Since the beginning we have had 17,133 total participants, of which 6,230 or 36.36% have been female. We are proud of the number of female students that HiMCM attracts. We feel this is remarkable and that we hope that all competing students, male and female, continue on to some STEM education.

The teams accomplished the vision of our founders by providing *unique* and *creative mathematical* solutions to complex open-ended real-world problems. This year the students had a choice of two problems both of which represent "real-world issues." (See the Judges' Commentary that follows for the problem statements.)

Commendation: All students and advisors are congratulated for their varied and creative mathematical efforts. Of the 598 registered teams, 581 submitted solutions, which is a 97.2% completion rate. These were broken down as follows: 383 doing problem A and 198 doing Problem B. The thirty-six continuous hours to work on the problem provided for quality papers; teams are commended for the overall quality of their work. Judges frequently comment about the amount of material students put together within the 36 hours. It is amazing.

Teams again proved to the judges that they had "fun" with their chosen problems, demonstrating research initiative and creativity in their solutions. This year's effort was again a success!

Judging: We ran three regional judging sites in December 2013. The regional sites were:

Naval Postgraduate School in Monterey, CA
Francis Marion University in Florence, SC
Carroll College in Helena, MN.

Each site judged papers for problems A and B. The papers judged at each regional site may or may not have been from their respective region. Papers were judged as Finalist, Meritorious, Honorable Mention, and Successful Participant. All finalist papers from the Regional competition were sent to the National Judging in Baltimore. This year's national judging, consisting of eight judges from academia (high school and college) and industry, chose the "*best of the best*" as National Outstanding. We usually discuss between 8-10 papers in the final round so all these papers were awarded "National Finalist." The National Judges commended the regional judges for their efforts and found the results were very consistent. We feel that this regional structure provides a good structure for the future as the contest grows.

National Outstanding Teams

Shanghai Experimental School, Shanghai, China
Hong Kong International School, Hong Kong
University High School, Irvine, CA
Shanghai Foreign Language School, Shanghai, China
Shanghai High School International Division
Shanghai, China
North Carolina School of Science and Mathematics,
Durham, NC
Eastside High School, Gainsville, FL
American Heritage School, Plantation, Coral Springs, FL

Judging Results:

Problem	National Outstanding	National Finalist	Finalist	Meritorious	Honorable Mention	Successful Participant	Unsucessful	Total
A	5	5	23	102	114	132	2	383
%	1%	1	6%	27%	30%	34%	1%	
B	3	5	19	37	58	73	3	198
%	1%	2%	10%	19%	30%	37%	1%	
Total	8	10	42	139	173	205	5	581
%	1%	2%	7%	24%	30%	35%	1%	

National Finalist Teams

East Lansing High School, East Lansing, MI
Illinois Mathematics and Science Academy, Aurora, IL
Hangzhou Foreign Language School, Wenhou, China
University High School, Irvine, CA
China Welfare Institution, Shanghai, China
Shanghai Foreign Language School, Shanghai, China
Practical Workstation for Mathematical Sciences, Shanghai University, Shanghai, China
Changle No. 1 Middle School, Weifang, China
Zhiyuan Middle School, Qingdao, China
Washtenaw International High School, Ypsilanti, MI

Common Core State Standards: The director and the judges asked that we add this paragraph. Many of us have read the Common Core Standards and clearly realize the mapping of this contest to the Common Core mathematics standards. This contest provides a vehicle for using mathematics to build models to represent and to understand real world behavior in a quantitative way. It enables student teams to look for patterns and think logically about mathematics and its role in our lives. Perhaps in a future *Consortium* article we will dissect a problem (paper) and map the standards into it.

General Judging Comments: The Judges' Commentary that follows provides specific comments on the solutions to each problem. As contest director and head judge, I would like to speak generally about solutions from a judge's point of view. Papers need to be coherent, concise, clear, and well written. Students should use spelling and grammar checkers before submitting a paper. Papers should use at least 12-point font. Students need to restate the problem in their own words so that the judges can determine the focus of the paper. Papers that explain the development of the model, assumptions with justifications, and its solutions and then support the findings mathematically usually do well. Modeling assumptions need to be listed and justified, but only those that come to bear on the solution (this can be part of simplifying the model). Laundry lists of assumptions that are never referred to in the context of model development are not considered relevant and deter from a paper's quality. The mathematical model needs to be clearly developed, and all variables that are used need to be well defined. Teams that merely present a model to use without any development do not generally do well. Thinking outside of the "box" is also considered important by judges. This varies from problem to problem but usually includes model extensions or sensitivity analysis of the solution to the team's inputs. Students need to attempt to validate their model even if by numerical example or intuition. A clear conclusion and answers to specific scenario questions are all key components. The strengths and weakness section is where the team can reflect on their solution and comment on the model's strengths and weaknesses. Attention to detail and proofreading the paper prior to final submission are also important since the judges look for clarity and style. Citations are also very important within the paper as well as either a reference or bibliography page at the end. We encourage citations within the paper in sections that deal directly with data and figures, graphs, or tables. Most papers did not use references within the paper, yet we saw many tables or graphs that obviously came from websites. We have noticed an increase in use of Wikipedia. Teams need to realize that although useful, the information might not be accurate. Teams need to acknowledge this fact.

Facts from the 16th Annual Contest:

- A wide range of schools/teams competed including teams from Korea, Singapore, UAE, Finland, Hong Kong, and China.
- The 598 registered teams from U.S. and International institutions represent a 7.2% increase in participation.
- There were 2291 student participants, 1456 (63.56%) male and unspecified and 835 (36.44%) female.
- Schools from twenty-six states participated.

The Future:

The contest, which attempts to give the under-represented an opportunity to compete and achieve success in mathematics, appears well on its way in meeting this important goal.

We continue to strive to improve the contest, and we want the contest to grow. Any school/team can enter, as there are no restrictions on the number of schools or the numbers of teams from a school. A regional judging structure is established based on the number of teams.

These are exciting times for our high school students. Mathematics continues to be more than learning skills and operations. Mathematics is a language that involves our daily lives. Applying the mathematical principles that one learns is key to future success. The ability to recognize problems, formulate a mathematical model, use technology, and communicate and reflect on one's work is key to success. Students gain confidence by tackling ill-defined problems and working together to generate a solution. Applying mathematics is a team sport!

Advisors need only be motivators and facilitators. They should encourage students to be creative and imaginative. It is not the technique used but the process that discovers how assumptions drive the techniques that is fundamental. Let students practice to be problem solvers. Let me encourage all high school mathematics faculty to get involved, encourage your students, make mathematics relevant, and open the doors to success.

Mathematical modeling is an art and a science. Teach your students through modeling to think critically, communicate effectively, and be confident, competent problem solvers for this new century.

International flavor of the contest:

The award format has changed slightly as the contest continues to grow internationally.

Previous Designation and New Designation:

- Successful Participant is still Successful Participant.
- Honorable Mention is still Honorable Mention.
- Meritorious is still Meritorious.
- Regional Outstanding Winner is now designated a Finalist.
- The 8-10 papers in the final discussion are now designated National Finalists.
- National Outstanding Winner is now designated an Outstanding Winner.

Contest Dates: Mark your calendars early: the next HiMCM will be held in November 2014. Registrations are due in October 2014. Teams will have a consecutive 36-hour block within the contest window to complete the problem and electronically submit a solution. Teams can register via the Internet at www.comap.com.

Math Models.org

It is highly recommended that participants in this contest as well as prospective participants take a look at the modeling web site, www.mathmodels.org, which has a wealth of information and resources.

Judges' Commentary

Problem A: Emergency Medical Response

The Emergency Service Coordinator (ESC) for a county is interested in locating the county's three ambulances to best maximize the number of residents that can be reached within 8 minutes of an emergency call. The county is divided into 6 zones and the average time required to travel from one zone to the next under semi-perfect conditions is summarized in the following Table 1.

Average Travel Times (min.)						
Zones	**1**	**2**	**3**	**4**	**5**	**6**
1	1	8	12	14	10	16
2	8	1	6	18	16	16
3	12	18	1.5	12	6	4
4	16	14	4	1	16	12
5	18	16	10	4	2	2
6	16	18	4	12	2	2

Table 1: *Average travel times from Zone i to Zone j in semi-perfect conditions.*

The population in zones 1, 2, 3, 4, 5 and 6 are given in Table 2 below:

Zones	Population
1	50,000
2	80,000
3	30,000
4	55,000
5	35,000
6	20,000
Total	270,000

Table 2: *Population in each Zone*

Goals of your model

1. Determine the locations for the three ambulances which would maximize the number of people who can be reached within 8 minutes of a 911 call. Can we cover everyone? If not, then how many people are left without coverage?

2. We now have only two ambulances since one has been set aside for an emergency call; where should we put them to maximize the number of people who can be reached within the 8 minute window? Can we cover everyone? If not, then how many people are left without coverage?

3. Two ambulances are now no longer available; where should the remaining ambulance be posted? Can we cover everyone? If not, then how many people are left without coverage?

4. If a catastrophic event occurs in one location with many people from all zones involved, could the ESC cover the situation? How do counties or cities design for those rare but catastrophic events?

5. In addition to the contest's format, prepare a short 1-2 page non-technical memo outlining your recommendations from your model and analysis finding for the ESC.

Judge's Comments:
William P. Fox, Naval Postgraduate School
Problem Author: Thomas Fitzkee, Francis Marion University

This problem could be addressed through various mathematical models. It was envisioned that students might use linear programming because of the way the problem was posed. Another method might include integer or mixed-integer programming (if students had ever been exposed to these topics). We found student approaches both rich and robust. Some teams used graph theory and node analysis, trying to minimize time along the nodes. Our goal for time was only within 8 minutes, which would satisfy the need, and we did not need or desire to minimize time especially if that limited coverage. The main goal was to maximize coverage within a constrained time constraint. Some teams used other interesting mathematical models and methods to create and pose an answer. Coverage was possible with 2 or 3 ambulances but not with 1. Few teams, if any, addressed the issue of multiple requirements or requests for medical response in real time for different regions. Some teams obtained 911 call information from the Internet and used those call data in model testing or in a more "realistic" analysis. The disaster model posed more difficult issues and many of these issues were glossed over, as expected in a short duration contest. Judges felt the treatment of a disaster might have been a useful discriminator.

Problem B: Bank Service Problem
The bank manager is trying to improve customer satisfaction by offering better service. Management wants the average customer to wait less than 2 minutes for service and the average length of the queue (length of the waiting line) to be 2 persons or fewer. The bank estimates it serves about 150 customers per day. The existing arrival and service times are given in the tables below.

Time between arrival (min.)	Probability
0	0.10
1	0.15
2	0.10
3	0.35
4	0.25
5	0.05

Table 1: Arrival times

Service Time (min.)	Probability
1	0.25
2	0.20
3	0.40
4	0.15

Table 2: Service times

1. Build a mathematical model of the system.
2. Determine if the current customer service is satisfactory according to the manager guidelines. If not, determine, through modeling, the minimal changes for servers required to accomplish the manager's goal.

3. In addition to the contest's format, prepare a short 1-2 page non-technical letter to the bank's management with your final recommendations.

Judge's Comments:
William P. Fox, Naval Postgraduate School
Problem Author: William P. Fox, Naval Postgraduate School

Most teams used basic queuing theory to attempt to answer the initial question. Teams searched the Internet for anything on the subject, so we saw some Lindley equations from the 1950s that we know are probably not currently taught in schools. We expected that simple queuing theory or a Monte Carlo simulation could answer part 1, and that the arrival and service times would not allow the bank manager to meet the goal. Most teams that attempted to fix the situation immediately tried to add a 2nd server and assumed that the new server would have the same service rates. Although this improves the service times so that the bank manager can meet the goal, the judges felt the cost of a 2nd server might be too great, and very few teams discussed training for the new server. Few teams discussed training the current on-site server to become more efficient and thus lowering service times. Training might be a more cost effective way to achieve better performance. Methods that we saw ranged from Lindley's equation, queuing theory and Monte Carlo simulation. In the simulation modeling few teams provided useful flow diagrams or algorithms. Teams still provide lots of computer code in the appendix. Simulations are useful if a large number of trials is run, and teams should also vary inputs to see the changes in the outputs. This was almost non-existent in student papers.

General Judge Comments
The judges were surprised with the amount of time and effort put into the modeling. Some papers were unnecessarily long, over 70 pages. Teams should be brief in their explanations of models that were considered and then not used. The lengths of the papers are sometimes

artificial because every possible model form has been included in the discussion although rarely used. Teams are encouraged to include only what they used to answer the questions.

Graphs were also an issue. Legends and axes should be labeled and all graphs should make sense. Papers were judged in final hard copy, not electronically. Thus, references to color in a graph were not helpful if the printed copy supplied by the team was not in color. Such teams should restrict their graphs to various gray-scales and perhaps differing markings (line, dashes, dots, etc.)

The executive summaries for the most part are still poorly written, although they appear to be getting a little better. This has been an ongoing issue since the contest began. Faculty advisors should spend some time with their teams and advise them to write a good summary. Many summaries appear to be written before the teams start and only state how they will solve the problem. Summaries need to be written last and should contain the **results** of the model as well as a brief explanation of the problem. The executive summary should entice the reader, in our case the judge, to read the paper. In the real world if the summary is not strong and with "good" results, management may never read the paper.

These comments are also applicable to the non-technical papers or "memos." We do not want equations and methods listed here, but the facts as to why your model is applicable and why your model's results are important to the reader.

Few teams, if any, did sensitivity analysis or error analysis with their models. With the *randomness* in each of the problems this becomes a crucial element.

Assumptions with Justifications: All assumptions should have some justification of their importance in the modeling process. Assumptions should not be listed that do not impact on the model used or developed. We also found many teams that listed initial assumptions and then, as they developed model after model, included new assumptions as they went. It would be better to list all assumptions together even if some assumptions were used much later in the paper. For example, a team could state assumption #7 and note that it is used in model #4.

Variables and Units: Teams must define their variables and provide units for each of them.

Computer generated solutions: Many papers used extensive computer code. Computer code used to implement mathematical expressions can be a good modeling tool. However, judges expect to see an algorithm or flow chart from which the code was developed. Successful teams provided some explanation or guide to their algorithm(s)—a step-by-step procedure for the judges to follow. Code is not read for papers that reach the final rounds unless the code is accompanied by an algorithm that is clearly and logically explained. A complete algorithm, if code is used, is expected in the body of the paper with the code in an appendix.

Graphs: Judges found many graphs that were not labeled or explained. Many graphs did not appear to convey information used by the teams. All graphs need a verbal explanation of what the team expects the reader (judge) to gain (or see) from the graph. **Legends, labels, and points of interest** need to be clearly visible and understandable, even if hand written. Graphs taken from other sources ***should be referenced and annotated***.

Summaries: These are still, for the most part, the weakest parts of papers. These should be written after the solution is found. They should contain results and not details. They should include the "bottom line" and the key ideas used in obtaining the solution. They should include the particular questions addressed and their answers. Teams should consider a brief three paragraph approach: a *restatement of the problem* in their own words, a short description of *their method and solution* to the problem (without giving any mathematical expressions), and the *conclusions* providing the numerical answers in context.

Restatement of the Problem: Problem restatements are important for teams to move from the general case to the specific case. They allow teams to refine their thinking to give their model uniqueness and a creative touch.

Assumptions/Justifications: Teams should list only those assumptions that are vital to the building and simplifying of their mathematical model. Assumptions should not be a reiteration of facts given in the problem statement. Assumptions are variables (issues) acting or not acting on the problem. Every assumption should have a justification. We do not want to see "smoke screens" in the hopes that some items listed are what judges want to see. Variables chosen need to be listed with notation and be well defined.

Model: Teams need to show a **clear link** between the assumptions they listed and the building of their model or models. Too often models and/or equations appeared without any model building effort. Equations taken from other sources should be referenced. It is required of the team to show how the model was built and why it is the model chosen. Teams should not throw out several model forms hoping to WOW the judges, as this does not work. We prefer to see sound modeling based on good reasoning.

In particular for Problem B, many students immediately wrote a queuing model without any discussion as to why it is appropriate. Often the queuing material, the Markov chain material, and even background material was lifted from sources and not referenced. Anything not created by a team should be referenced where it is used.

Model Testing: Model testing is not the same as testing arithmetic. Teams need to compare results or attempt to verify (even with common sense) their results. Teams that use a computer simulation must provide a clear step-by-step algorithm. Lots of trials and related analysis are required when using a simulation. Sensitivity analysis should be done in order to see how sensitive the simulation is to the model's key parameters. Teams that relate their models to real data are to be complimented.

Conclusions: This section deals with more than just results. Conclusions might also include speculations, extensions, and generalizations. This is where all scenario specific questions should be answered. Teams should ask themselves what other questions would be interesting if they had more time and then tell the judges about their ideas.

Strengths and Weaknesses: Teams should be open and honest here. What could the team have done better?

References: Teams may use references to assist in their modeling, but they must also ***reference the source*** of their assistance. Teams are reminded that only ***inanimate resources*** may be used. Teams cannot call upon real estate agents, bankers, hotel managers, or any other real person to obtain information related to the problem. References should be cited where used and not just listed in the back of the paper. Teams should also have a reference list or bibliography in the back of the paper. It is requested that teams use some in-line reference to **graphs, figures**, and direct explanation of materials taken from other sources.

Adherence to Rules: Teams are reminded that detailed rules and regulations are posted on the COMAP website. Teams are reminded that they may use only ***inanimate sources*** to obtain information and that the ***36-hour time limit is a consecutive 36 hours.***

Problem A Summary: Hong Kong International School

Team Members: Christopher Bradsher, Jeffrey He, Andrew Kim, Andrew Kao
Advisor: Kevin Mansell

Based on the given information of time taken to travel from one zone to another, we created a model of the optimum configurations of three, two, and one ambulances, and determined whether the ESC could cover a catastrophe. We started by setting definitions, assumptions, and variables, and then analyzed the given data to see which ambulances could reach which zones within 8 minutes and to see which configurations could actually cover all 6 zones of the county. Following, with the 13 configurations that *could* cover the whole county, we analyzed each one based on the ambulances' coverage of people within time constraints and based on data from emergency calls per day and time spent travelling. With the remaining three ambulances that we determined to cover the same amount of people, we decided the optimum configuration through comparing the spare minutes of each ambulance, with the configuration being placing ambulances in zones 2, 4, and 6, with 97.7% of the people being covered. However at the end, because not everyone was covered, three ambulances would not be able to cover everyone. Next, we repeated the process for two ambulances and one ambulance, and determined the configuration to be in zones 2 and 5, and in zone 2, respectively. Two ambulances covered 77.55% of the people. One ambulance covered 45.83% of the people. We continued by trying multiple methods for three ambulances to tackle disasters, and we discovered that the best way to operate the ambulances in its stations was to use the "disaster-saver" method where an ambulance would go to a closest zone and operate there until all calls were dealt with before returning to its original zone. This method covered up to 94.1% of the people, so we determined that the ESC alone would not be able to cover a catastrophic event. Although as a whole the county may be fine after the catastrophe, it will most likely have received much help from the state or country it is in. In addition, with the information gathered from researching case studies in places including Japan, Cuyahoga, Ohio, and San Diego, we discovered methods for the county or state to be prepared for catastrophes including floods, earthquakes, and hurricanes. Finally, we ended by writing a memo outlining our recommendations from our model and analysis.

Problem A Summary: Shanghai Foreign Language School

Team Members: XiaoChan Yang, Oian XDu, YiLei Han, YiNAn Cai
Advisor: Yue Sun

The first quintessential problem we face is to determine the shortest time it takes to commute between different zones, which we solved by employing both **Dijkstra's Algorithm** and the **Floyd–Warshall Algorithm**, favoring the latter in practice. Having reasoned that an ambulance arrangement that effectively covers the whole county is one that maximizes the population coverage, we applied the **Tree Traversal** programming method to Questions 1 and 2, and obtained solutions that succeeded in covering the entire region. We then analyzed the given data and answered Question 3 by putting one ambulance in the location with the largest population coverage and calculated the number of people left without coverage.

We optimized the basic model from 2 different angles, building matrices and making accurate mathematical calculations to achieve both the most efficient way to cover the population and the fastest way to respond to emergency calls. We concluded that if ambulances are arranged in such a way so that the regions they cover overlap, they can provide more secure care in case one of them is dispatched, while the shorter the average time it takes an ambulance to reach designated areas is, the faster its response. Therefore we came up with two optimization plans, one to increase coverage overlap, and another to ensure fastest contact. We used standard deviation as a method of determining the best solutions among the optimized ones.

To solve the problem concerning catastrophic scenarios we approached it from three perspectives depending on the nature of the catastrophe, and proceeded to discuss three plans on how to station the ambulances so that they most substantially boost rescue rates. In case of sporadic outbreaks in random areas, we applied the aforementioned plan with large area overlaps so that enough ambulances will be on standby once one has been dispatched. Our second plan is to have the ambulances repeatedly dispatched and retrieved all over the county, but we found that this does not provide satisfactory coverage. Most notably, in cases where three ambulances are involved and need to reach all areas as fast as possible, we used the method of overall planning to determine several ways to station ambulances so that they can travel through all six zones within 8 minutes.

We also went into specifics in the catastrophic scenario, and discussed two types of possible calamities in detail: the Radiation Model, in which a crisis erupts in a single location and spreads throughout the county, downgrading in intensity as it spread, and the Diffusive Model, in which a crisis does not decrease in intensity but rather disseminates. We chose earthquakes to represent our Radiation Model and epidemics to represent our Diffusive Model. Using formulas from **Richter's Magnitude Scale** and the **SIR Model** we compiled two constructive plans on how to deal with these disasters.

Having put these specific samples under examination, we concluded with a well-organized plan that can be elaborated to determine the best solution for counties and cities, regardless of their land area.

Problem A Summary: Shanghai High School International Division

Team Members: Allen Chen, Jerry Ding, Matthew Yuan, Sunghun Kim
Advisor: Mingxin Zhang

The goal of our model is to help Emergency Service Coordinator (ESC) maximize the number of residents in the six population zones that can be reached within 8 minutes of an emergency call by determining where the county should station its ambulances.

First, we determined the best location to place one, two, or three ambulances by using the Boolean model, which assigns a "true" or "false" property for an ambulance's ability to reach a zone in 8 minutes. Then by examining the combinations closely, we obtained the maximum value of population coverage: there were 11 solutions for total coverage with three ambulances, a unique solution for total coverage with two ambulances, and a maximum of 160,000 people covered with only one. This model is useful due to its zero-tolerance restrictions on the solutions, but cannot differentiate between acceptable solutions.

Second, we improved our model by minimizing transit time, while taking population into consideration. This model uses weighted averages to find the average amount of time a system of ambulances takes to rescue a random patient from any zone, which are more likely to come from populous zones. We obtained the most optimal locations for different situations: a lone ambulance is best stationed in zone 2, a pair of ambulances are best stationed in zones 2 and 5, and three ambulances are best stationed in zones 1, 2, and 5.

Third, we created two models to examine the optimal solution for dealing with catastrophic situations, where many people from different zones make 911 calls. The first method was the nearest neighbor method, in which we created a near-optimal route shared by three ambulances, passing through all zones. Then in this route, we placed hospitals at every other point. Using this method, we can on average rescue a person every 2.722 minutes.

The second method, a split-cycle method, takes a more complicated approach: it divides the six zones evenly into three areas, with one ambulance responsible for one area. By comparing different combinations produced, we are able to find an optimal solution where ambulances are stationed in zones 2, 4, and 5. This method is both more effective and practical: it averaged only 2.118 minutes to rescue a person.

Problem A Summary: American Heritage School, Plantation

Team Members: Francissco Rivera, Samir Khan, Philip Gaddy, Kwesi Levy
Advisor: Radleigh Santos

In tackling the problem, we considered the ideal distributions of ambulances among the zones depending on the number of ambulances that are available. To do this, we had to develop criteria to be able to compare distributions. To do this, a simplifying assumption was first made and then removed.

If we assume that only one emergency call will happen at a time (two calls will not happen in close succession so that the response to one affects the possibilities of response to the other), we can consider the percent of the population that can be reached by an ambulance in under 8 minutes as a preliminary statistic. Using a computer-based search of all of the possible distributions of algorithms, we found that there are 13 distributions of three ambulances that cover 100% of the population in less than 8 minutes. Therefore, to compare these distributions, we calculated the average response time for each distribution. There was a clearly superior distribution: stationing an ambulance in each of zones 1, 2, and 5.

We repeated this search for an ideal allocation of two ambulances. We found that only one distribution of two ambulances resulted in 100% of the population being reached in under 8 minutes: positioning one in region 2 and another in region 5. Finally, we considered one ambulance, and found that the ideal zone was zone 2.

Additionally, given these ideal distributions, we considered how much time was required to move between them given that an ambulance suddenly became available or unavailable. However, for ambulances that become unavailable for short, predictable periods of time (such as would happen from taking a call) we considered on a case-by-case basis whether it would be beneficial to move the remaining ambulances or simply wait for the ambulance to become available.

However, sometimes multiple calls occur in close succession, and can be modeled as happening simultaneously for purposes of response. Using a Poisson distribution, we calculated estimates as to the likelihood of some number of calls happening within a 10-minute time period. In these situations, if there are less than three calls (which will happen roughly 99.8% of the time), our available resources can still deal with the situation. An event where two emergency calls happen simultaneously was modeled with the 13 distributions previously found, and all things considered, the previous distribution of ambulances was still the best. Although there are three or more incidents extremely infrequently, it is possible, and our resources would be overwhelmed in these cases.

Problem B Summary: North Carolina School of Science and Mathematics

Team Members: Steven Liao, Irwin Li, Zachary Polizzi, Jennifer Wu
Advisor: Daniel Teague

The problem that we are presented involves determining minimal changes to the teller system of a bank to achieve the goals set by the manager, namely to minimize the length of the waiting line and the time customers spend waiting. Although the context of our problem is narrow, the models we developed can be applied to any situation involving queues and improving operational efficiency.

In this project, we wanted to make the smallest change to the bank's teller system while satisfying management goals of having the average queue lengths be 2 people or fewer and the average wait time in queue be 2 minutes or less. To achieve these goals, we constructed a system of probabilistic differential equations that related the average arrival and departure rates to find the likelihood that our bank will be serving a given number of people at any time. From these probabilities, we determined the expected number of customers we will see waiting in line and the average time they will spend waiting. Using our model we determined that hiring an extra teller would decrease average queue lengths from 2.79 to 0.28 people and wait time from 6.84 to 0.69 minutes. We also accounted for factors such as bank peak hours where we would see an increased arrival rate by adding piecewise components with respect to time to our system of differential equations. Again, hiring an extra teller to serve during those peak hours decreased average queue length from 2.32 to 0.98 people and wait time from 5.68 to 2.4 minutes.

We also created a computational model of the system using stochastic processes. This computational model allows for the analysis of more complex scenarios and results in more accurate numerical data than the mathematical model. The computational model agreed with the mathematical model in that the current system does not achieve the manager's goals for queue length and wait time, and that improvements are necessary. Multiple improvements were modeled and tested, including adding a full-time and a part-time additional teller and installing an ATM. All of these changes brought the queue length and wait time within the manager's specifications, so we performed a cost/benefit analysis to choose the optimal change. We found that, although hiring a part-time teller to work peak hours only costs $8,300/year, installing an ATM will be a cheaper and better solution in the long run, with a payback period of only 3 years. Therefore, we state this solution achieves the objectives while having minimal costs.

Problem B Summary: Shanghai Experimental School

Team Members: Jia Li, Yiming Li, Xiaojun Xu, Zehao Zhuang
Adviser: Xiaming Chen

Our goal is to build a mathematical model of a 150-people queuing system that meets the manager's requirement of an average queue length less than 2 people and an average waiting time less than 2 minutes.

To analyze this queuing system, we put forward two models that can verify each other. They are: the mathematical model and the simulation model. In the mathematical model we use the knowledge of probability theory, hoping to figure out a recursion formula about the probability of the state of a queue at a certain time. In order to reach this goal, we described the state of a queue at a certain time by defining several variables. Using this model, we can figure out the two important parameters: the average queue length and the average waiting time. Then we put forward a simulation model. In the given situation, we put forward an algorithm that can simulate and measure the whole process of how the 150 customers come and leave. Using a Monte Carlo method, we simulated the process for many trials and then got the average length of queue as well as the average waiting time.

After a comparison, the average queue length and waiting time from the two models coincided with each other perfectly. The average waiting time is about 4.9 minutes while the average queue length is about 2.4 people. Since the simulation model calculated faster, we decided to use it exclusively in the following discussion.

Results showed that the current situation couldn't meet the manager's requirement, so we put forward several solutions from different perspectives to meet the requirement and improve customer satisfaction, such as improving the service efficiency through training to change the probability distribution of service time; establishing an online bank to lower the everyday customer flow to increase time between arrivals; we can also establish a new service window. For each solution, we discussed the minimal change to meet our goal.

At last, we gathered all our model results, and took several reasonable ones as our recommendation to the manager so that he can choose one according to the actual condition of the bank.

Problem A paper: Eastside High School

Team Members: Sarah Seeger, Maxwell Curtis, Apara Agarwal, Alvaro Valle
Advisor: Carl Henriksen

1. Introduction and Restatement of the Problem

We wanted to determine whether it is possible to reach every zone in a county in eight minutes with different ambulance placements. We addressed situations with one, two, and three ambulances. If possible, we sought to ascertain which placements would result in the lowest average time to reach an individual. We then considered the ability of three ambulances to respond to widespread disaster. The most important factors in determining ambulance placement were the times in which an ambulance would reach different zones and the number of people an ambulance could reach in less than eight minutes.

2. Definitions, Variables, and General Assumptions

2.1 Definitions

- We assume that the data we were given form a basis for our projections of continued population growth in this area.
- The average travel time is calculated as the time taken to get from the center of one zone to the edge of another. The emergency is dealt with once the ambulance reaches the zone.
- We define coverage or success as the ambulance being able to reach every zone within eight minutes.
- We define the probability of an emergency occurring in a zone as the percentage of the population in that zone.

2.2 Variables

- $\bar{t}$ is expected average response time of the ambulance in minutes
- t_n is shortest ambulance response time to a zone in minutes
- p_n is the population of a zone
- P is the total population of 270,000

2.3 Assumptions

Assumption: Zones are internally uniform.
Justification: We were given no information on population density or other factors. There is no way to predict how ambulances are affected in route.

Assumption: Each person is equally likely to face an emergency.
Justification: This is the basis for our definition of an emergency occurring in a zone.

Assumption: Each ambulance can respond to any zone.
Justification: In order to minimize time to reach a zone, each ambulance must be available to every zone that it can reach in 8 minutes. It is reasonable to conclude that the available ambulance that minimizes response time is used.

Assumption: A single emergency will occur at once, and the ambulance sent will have returned by the next call.
Justification: A single ambulance must be responsible for multiple zones in order to ensure coverage within eight minutes. Rutherford County has 274,454 people, just a few thousand above our theoretical county. They get about 62 calls per day, or about one every 23 minutes. We assume that 23 minutes is ample time for an ambulance to reach an emergency and return before the next one.

Assumption: Each ambulance is placed in a different zone.
Justification: The previous assumption removes the need for multiple ambulances in a zone because an ambulance is always available to all zones to which it is assigned.

3. Creating a Model

3.1 Three Ambulances

3.1.1 Determining Optimal Location

We looked at all possible placements for three ambulances. In Figure 1, we see ways for each zone to be reached within 8 minutes.

Average Travel Time (minutes)						
Zone	**1**	**2**	**3**	**4**	**5**	**6**
1	1	8	12	14	10	16
2	8	1	6	18	16	16
3	12	18	1.5	12	6	4
4	16	14	4	1	16	12
5	18	16	10	4	2	2
6	16	18	4	12	2	2

Figure 1: *Average Travel Time in Minutes for an Ambulance to Reach a Zone*

There are ${}_6C_3 = 20$ ways to place three ambulances in six zones with one ambulance per zone. We tabulated all 20 and found that 11 reached all zones in time (Figure 2).

	Ambulance Zone							
Possibility	**1**	**2**	**3**	**4**	**5**	**6**	**Success Zones**	**Uncovered**
1	X	X	X				N	4
2	X	X		X			N	5, 6
3	X	X			X		Y	
4	X	X				X	N	4
5	X		X	X			Y	
6	X		X		X		Y	
7	X		X			X	N	4
8	X			X	X		Y	
9	X			X		X	Y	
10	X				X	X	Y	
11		X	X	X			Y	
12		X	X		X		Y	
13		X	X			X	N	4
14		X		X	X		Y	
15		X		X		X	Y	
16		X			X	X	Y	
17			X	X	X		N	1, 2
18			X	X		X	N	1, 2
19			X		X	X	N	1, 2
20				X	X	X	N	1, 2

Figure 2: *All Possible Locations for Three Ambulances*

To find optimal placement, we created a new unit, people-minutes, referring to the total amount of time required to a reach each individual. We determined which ambulance has the shortest response time for each zone. We then multiplied the population by the response time. To calculate average time to reach an individual, we divided by total population (270,000).

$$\bar{t} = \sum_{n=1}^{6} \frac{t_n \times p_n}{P}$$

We calculated average response time if ambulances are placed in zones 1, 2, and 5. The bold red items in Figure 3 show the shortest response time for each zone.

Average Ambulance Travel Time (minutes)						
Zone	**1**	**2**	**3**	**4**	**5**	**6**
1	**1**	8	12	14	10	16
2	8	**1**	**6**	18	16	16
5	18	16	10	**4**	**2**	**2**
Population	**50,000**	**80,000**	**30,000**	**55,000**	**35,000**	**20,000**

Figure 3: *Average Ambulance Travel Time and Population Per Zone*

If we input these data into the formula, we have:

$$\bar{t} = \frac{1(50,000)+1(80,000)+6(30,000)+4(55,000)+2(35,000)+2(20,000)}{270,000}$$

This gives us an expected average response time of 2.37 minutes.

We then found the other expected average response times (Figure 4).

Ambulance Position	$\bar{t}$ (minutes)
1, 2, 5	2.37
1, 3, 4	4.00
1, 3, 5	3.94
1, 4, 5	3.61
1, 4, 6	3.61
1, 5, 6	4.22
2, 3, 4	3.22
2, 3, 5	3.24
2, 4, 5	2.83
2, 4, 6	2.83
2, 5, 6	3.44

Figure 4: *Expected Average Response Times for Arrangements of Three Ambulances.*

Thus, placing ambulances in zones 1, 2, and 5 is optimal, with an expected average response time of only 2.37 minutes.

3.1.2 Sensitivity Analysis

We tested the effect of a change in average travel times on the time it takes ambulances to reach an emergency.

Average Travel Time (minutes)						
Zones	**1**	**2**	**3**	**4**	**5**	**6**
1	1.1	8.8	13.2	15.4	11	17.6
2	8.8	1.1	6.6	19.8	17.6	17.6
3	13.2	19.8	1.65	13.2	6.6	4.4
4	17.6	15.4	4.4	1.1	17.6	13.2
5	19.8	17.6	11	4.4	2.2	2.2
6	17.6	19.8	4.4	13.2	2.2	2.2

Figure 5: *Increased Shortest Travel Times*

Figure 5 shows that a 10% time increase renders many possible solutions ineffective. An ambulance in zone 1 is the only one to reach zone 1 in time. Similarly, an ambulance in zone 2 is the only one to reach zone 2 in time and can also reach zone 3 in time. The third ambulance must go in zone 5 to cover zones 4, 5, and 6. No placement other than the optimal (zones 1, 2 and 5) works.

Next, we used the adjusted t_n values to calculate new response times (Figure 6). A 10% increase did not change travel times much: all are under eight minutes.

Zone	1	2	3	4	5	6
Response Time (minutes)	1.1	1.1	6.6	4.4	2.2	2.2

Figure 6: *Increased Shortest Travel Times for Ambulances in Zones 1, 2, and 5*

Using our earlier method, we found a new average response time of 2.61 minutes. This is a 10% increase, which we expected because t_n is linearly related to $\bar{t}$. Thus, our optimal solution works with a 10% increase in the travel time.

Currently, zone 3 takes the longest to reach under semi-perfect conditions, at an average of six minutes. We wanted to find the percent change in travel times for an ambulance to no longer reach zone 3 within the eight-minute limit.

$$\frac{8\text{ minutes} - 6\text{ minutes}}{6\text{ minutes}} = 0.\overline{3}$$

Average travel times must increase by over 33% for our optimal solution to fail

3.1.3 Strengths and Weaknesses

Our solution works well because it accounts for the population dispersion among the zones and the shortest travel time for an ambulance from zone 1, 2, or 5 to reach an emergency. Our solution also held up well under a sensitivity analysis. However, our model is too simplistic in some areas. Because we do not know population dispersals, we cannot give a more realistic interpretation. Our model also assumes one emergency at a time, which may not be true.

3.2 Two Ambulances

Editor's Note: Tables in this section are similar to those in section 3.1 and have been removed to conserve space.

3.2.1 Determining Optimal Location

There are ${}_6C_2 = 15$ different ways to place two ambulances. We found one placement that works: zones 2 and 5. Again, this solution relies on the assumption that an ambulance only deals with one emergency at a time.

We used shortest travel times and our earlier method to find expected average response time, , which in this case is 3.67 minutes. Although this is a more than the average with three ambulances, it is small enough.

3.2.2 Sensitivity Analysis

We increased travel times 10%, but found that we no longer have a successful solution. The ambulance traveling to zone 1 cannot arrive in time, and there are 50,000 people who will wait an extra 48 seconds. The new is 4.03 minutes, meaning that the average citizen can still expect help within eight minutes.

3.2.3 Strengths and Weaknesses

Our model predicts that even under conditions that are not semi-perfect, the average citizen can expect to receive aid within eight minutes. However, sensitivity analysis shows how a delay in travelling to zone 1 would cause the ambulance to fail to arrive in time.

3.3 One Ambulance

3.3.1 Determining Optimal Location

With one ambulance, there is no location from which all zones can be reached in time. We found the number of people that can be reached (Figure 7).

Zone	People Reached in 8 Minutes	People Not Reached in 8 Minutes	AverageTime to Reach Patient (minutes)
1	130,000	140,000	9.22
2	160,000	110,000	9.37
3	85,000	185,000	11.24
4	85,000	185,000	12.50
5	110,000	160,000	10.42
6	85,000	185,000	11.59

Figure 7: *People Serviced and Average Service Time for Single Ambulance.*

The difference in average times between zones 1 and 2 is only 0.15 minutes, or 9 seconds. We chose zone 2 because more people can be reached in time.

3.3.2 Sensitivity Analysis

As seen in Figure 8, a 10% increase causes an ambulance starting in zone 1 or 2 to reach less people. It is clear that zone 2 is a better place for the ambulance as it can reach more than twice the number of people within eight minutes, although the average time to reach the entire population is slightly higher.

Zone	People Reached Successfully	People Not Reached Successfully	Average Time to Reach Patient (minutes)
1	50,000	220,000	10.14
2	110,000	160,000	10.31
3	85,000	185,000	12.36
4	85,000	185,000	13.75
5	110,000	160,000	11.46
6	85,000	185,000	12.75

Figure 8: *People Reached and Average Response Time with Increased Travel Time.*

3.3.3 Strengths and Weaknesses

Our solution covers the maximum number of people in time that a single ambulance can. However, it is not the fastest average time and doesn't cover all zones in less than eight minutes.

4. Catastrophic Events

4.1 Multiple Emergencies and Multiple Zones

We considered two simultaneous emergencies and found the minimal time for two of three ambulances in zones 1, 2, and 5 to reach them (Figure 9).

		Second Emergency Zone					
		1	2	3	4	5	6
First Emergency Zone	1	8	1	6	4	2	2
	2	1	8	8	4	2	2
	3	6	8	10	6	6	6
	4	4	4	6	14	10	14
	5	2	2	6	10	10	10
	6	2	2	6	14	10	16

Figure 9: *Minimum Travel Time (minutes) to Deal With Simultaneous Emergencies.*

We then plotted these data on a three-dimensional contour graph (Figure 10).

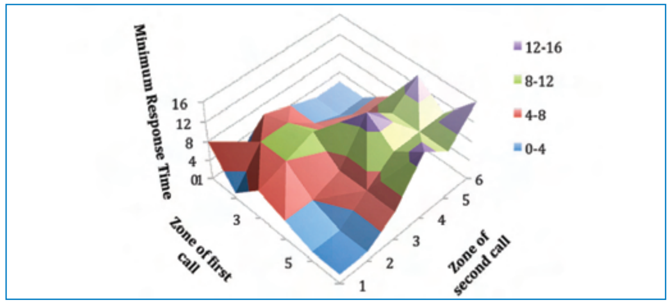

Figure 10: *Contour Plot of Minimum Response Times to Simultaneous Emergencies.*

We can handle all emergencies in zones 1 or 2. We can also handle all emergencies in zone 3 unless both calls are from there. However, we are unable to respond in time if both calls are from zones 4, 5, or 6.

We found that in seven of 21 possible situations ambulances could not respond in time. If we assume that each person is equally likely to be in an emergency, we can calculate the probability of each of those seven situations occurring. We calculated the percentage of the total population each zone (Figure 11).

Zone	Percentage of Population
1	18.5%
2	29.6%
3	11.1%
4	20.4%
5	13.0%
6	7.4%

Figure 11: *Population Divided by Zone*

We found the probability of each unsuccessful situation occurring. We assumed that the probability of an emergency depends only on a zone's population. In essence, each person is independent and equally likely to face an emergency. We multiplied the population portions of the zones involved (Figure 12).

Unsuccessful Situation	Probability
Zones 3, 3	1.23%
Zones 4, 4	4.16%
Zones 4, 5	2.65%
Zones 4, 6	1.51%
Zones 5, 5	1.69%
Zones 5, 6	0.96%
Zones 6, 6	0.55%
Total	12.75%

Figure 12: *Probabilities of Unsuccessful Multiple Emergencies Occurring.*

If two calls come at the same time, the probability that ambulances will not reach both in time is 12.75%. Thus, the probability that ambulances reach both emergencies in time is 87.25%, which shows that our optimal solution can handle more than one call.

Next, we addressed the question of how to maximize coverage when a catastrophe affects all zones. We used ambulances in zones 1, 2, and 5 to find optimal routes. We found two optimal solutions (Figure 13).

	Zone 1 Ambulance		Zone 2 Ambulance		Zone 5 Ambulance		
	Zones Covered	Total Time (minutes)	Zones Covered	Total Time (minutes)	Zones Covered	Total Time (minutes)	Minimum Response Time
#1	1, 2	9	3, 6	10	5, 4	6	**10 minutes**
#2	1	1	2, 3, 6	11	5, 4	6	**11 minutes**

Figure 13: *Minimum Response Time of Two Different Ambulance Configurations.*

In option one, the first ambulance begins in zone 1 and moves to 2. The second ambulance begins in zone 2, moves to 3, and then to 6. The third ambulance begins in zone 5 and moves to 4. The second ambulance takes ten minutes, which is the longest. Option two is similar, with the only difference being that the zone 1 ambulance only covers zone 1, while the zone 2 ambulance covers 2 and then 3. This new route for the ambulance in zone 2 takes 11 minutes.

Although the second option takes an extra minute, it covers more people in time (Figures 14 and 15). For this reason, we chose the second option as optimal.

Zone #	Coverage (minutes)	Under 8 min?	
1	1	Y	Successfully Covered People
2	9	N	
3	6	Y	
4	6	Y	
5	2	Y	170,000
6	10	N	

Figure 14: *Option One's Coverage in Eight Minutes*

Zone #	Coverage (minutes)	Under 8 min?	
1	1	Y	Successfully Covered People
2	1	N	
3	7	Y	
4	6	Y	
5	2	Y	250,000
6	11	N	

Figure 15: *Option Two's Coverage in Eight Minutes*

These figures assume semi-perfect conditions, but it is unlikely they would exist during a major catastrophe. Also, there will be many people to treat in a zone before an ambulance can leave, causing actual times to be more than expected. However, our model provides a basis for assigning ambulances based on travel times. Minimizing travel time gives the quickest medical response.

4.2 Planning for an Emergency

There is no way to cover an event impacting all zones. Although three ambulances could reach three zones in time, there is no way they can address six zones at once. This type of situation is not accounted for in our model.

Additionally, many roads and highways might be closed after a disaster, and ambulance response time would greatly increase.

To maximize survival, the county should institute disaster procedures. Plans in similar-sized counties include awareness at the individual level. For example, people should learn evacuation routes and how to navigate them.

Emergency management teams should coordinate with those of other city departments and agencies such as American Red Cross and the Federal Emergency Management Agency. City departments, such as the police force and the fire department should be prepared to assist in emergency response.

Volunteers can be organized beforehand; for large events, places such as Rutherford County use bicycles for medical treatment as well as traditional ambulances, since they are quicker to cover large crowds and traverse rough terrain. Another important resource would be survivors of a catastrophe, who should be able to rely on each other for help.

5. Conclusion

To decide ambulance placement, we sought to maximize the number of people that are within eight minutes of an ambulance. In situations where more than one ambulance placement was possible, we minimized average time to reach an emergency by weighting the time to reach a zone by the zone's population.

By focusing on minimizing travel time and maximizing the portion of the population reached in eight minutes, we found optimal placements for three, two, and one ambulance. Three ambulances are best placed in zones 1, 2, and 5. The only placement for two is zones 2 and 5. With one ambulance it is impossible to reach every resident in eight minutes. We found that the best placement is zone 2, but this leaves 110,000 people unreachable in the time limit. In a major disaster, even three ambulances are impractical.

We found that out of all possible two-emergency situations, we would be able to successfully reach both 87.25% of the time. If an event affects all zones, assigning each ambulance to two zones at once results in many response times over eight minutes. Instead we suggest sending the ambulance in zone 1 to cover zone 1, the ambulance in zone 2 to cover 2, then 3, then 6, and the ambulance in zone 5 to cover 5 then 4. This covers all zones in 11 minutes and reaches 250,000 people in time.

Problem B Paper: University High School

Team Members: Vinson Luo, Matthew Ku, Daniel Johnson, David Zheng
Advisor: Stephanie Chang

Introduction

In banking, the waiting time for a customer and the length of the line are two factors that affect whether or not a customer has a pleasant experience. Unfortunately, due to the unpredictable nature of customer arrivals and the varying time required to serve each customer, it can be difficult to determine whether a current system is satisfactory. In this paper we provide two methods to model these factors and propose a strategy for a bank to raise customer satisfaction with minimal changes to its current system.

Problem Restatement

The bank management wants to ensure that on average, customers wait no more than 2 minutes, and the wait line has no more than 2 people. We are given a probability distribution of the difference between customer arrival times (from 0-5 minutes) and a probability distribution of the time it takes for the bank to serve a customer (from 1-4 minutes). Using these probabilities and assuming that 150 customers arrive each day to receive service from one server (teller), we are tasked with determining whether the bank's current system is satisfactory. If necessary, we can then determine minimal changes to accomplish management's goal.

Assumptions

- Customers only arrive at the bank and are served at exact minute intervals.
  Justification: The given data is by the minute.

- Customers are served in the order they arrive.
  Justification: Most queues work this way. This is necessary in order to properly count the waiting time of customers.

- Servers work continuously until all 150 customers have been served (no breaks), and the time difference between serving customers is negligible.
  Justification: As soon as one customer is done being served, the next customer should immediately begin receiving service in order to keep times on the minute.

- The service time data provided correspond to the rate of service of a single server, and this server serves all customers in the original service system.
  Justification: The provided data imply that there is only one line and one server, and the rate of service for a single server must be consistent to create a model.

- Multiple servers work at the same rate.
  Justification: Servers need to work at the same rate given to us so that we can predict the outcome of waiting times and the length of the queue.

- The time for an "emergency" (back-up) server to begin serving customers from when he or she is called is two minutes.
  Justification: It is impractical for someone to immediately begin working when they are called, so we added a two-minute transition delay.

For the purposes of our model, we will also define the following:

- Customers are numbered from 1 to 150 in the order that they arrive.

- The queue is the line in which customers stand waiting to be served.

- A server is a person who is capable of providing service to customers

- All times are measured in minutes unless otherwise specified.

- An emergency server is a server that only begins working when the queue length exceeds a predetermined limit and stops working when the queue is empty.

- q_{enter} is the length of the queue before an emergency server begins work.

- A probability mass function (PMF) is the discrete-time equivalent of a probability density function. A PMF gives the probability that a random variate is exactly equal to a given integer value. For example, if F is a PMF, then $F(x) = P(f = x)$.

Designing the Mathematical Model

We use two methods: one mathematical that yields exact theoretical results and one using a simulation that yields approximate results for more complex situations.

Purely Theoretical Approach

The problem can be interpreted as a discrete-time version of a G/G/1 queue (a queue with two separate non-exponential probability distributions that determine when people enter and leave the queue, and with a single server). In general, a continuous-time G/G/1 queue can be modeled using Lindley's integral equation, given by

$$W(x) = \int_0^\infty U(x-y)\,dW(y)\ , \ x \geq 0$$

where

- *W(x)* is the probability that the *n*th customer waits no more than *x* minutes as *n* tends to infinity
- *U(x)* is the probability that the difference between the previous customer's service time and the *n*th customer's arrival time is less than or equal to *x* minutes as *n* tends to infinity
- *dW(y)* is the infinitesimal probability that the *n*th cus tomer waits for exactly *y* minutes as *n* tends to infinity.

We derive a discrete-time equivalent of this equation to find a theoretical wait time distribution. We decided to compute the discrete-time version using probability mass functions instead of cumulative density functions, as we are given tables that match discrete time intervals. Also, as we were given an explicit estimate of number of customers, we decided not to take the limit as customer number approaches infinity, but instead calculate each customer's wait-time distribution separately. We found that the distribution of waiting times for the *n*th customer can be found solely on the basis of the distribution of waiting times for the previous customer and the data in the Potential Changes section below. This formula summarizes our relation:

$$W_n(y) = \begin{cases} \displaystyle\sum_{i=0}^{\max(W_{n-1})} W_{n-1}(i)\,U(y-i), \text{ if } y>0 \\ \displaystyle\sum_{i=0}^{\max(W_{n-1})} \left(W_{n-1}(i) \sum_{j=0}^{j_{\max}} U(-i-j) \right), \text{ if } y=0 \end{cases}$$

where

- $max\{w_{n-1}\}$ is the maximum possible wait time of the $(n-1)$th customer
- $W_n(y)$ is the probability that the *n*th customer waits exactly *y* minutes
- *U(x)* is a probability mass function that gives the proba bility that $s_{n-1} - t_n = x$, where s_{n-1} is the service time of the $(n-1)$th customer and t_n is the time interval between arrivals of the $(n-1)$th and *n*th customers.
- $j_{\max}$ is a constant indicating maximum time in minutes between the end of the *n*th customer's service and the arrival of the $(n+1)$th customer. Given the provided data, we can set $j_{\max} = 4$.

Given the initial condition that $W_1(0) = 1$, we can use these two recurrence relations to generate $W_n(x)$ for each $n > 1$. This allows us to construct probability distributions of the wait times for an arriving customer. Once we have computed $W_1(x)$, … $W_{150}(x)$, we can compute average wait time of the *n*th customer by treating W_n as weights for a weighted average of the integer waiting times:

$$\bar{w}_n = \sum_{i=o}^{\max(w_n)} iW_n(i)$$

We can then find average waiting time of all customers by taking the mean of $\bar{w}_1, \ldots, \bar{w}_{150}$.This theoretical approach provides an exact mathematical formulation for average wait time and can be extended to cases with an arbitrary number of people or different probability distributions. However, the complexity of evaluating $W_n(x)$ escalates as *n* increases, making it unfeasible for large *n*. This approach also cannot easily incorporate multiple servers. Also, this method only finds average waiting time for each customer, which cannot easily be converted into the mean queue length.

Computational Approach

To simulate the effects of adding more servers and estimate mean queue length, it is more feasible to analyze the results of multiple simulation trials. We designed a computer model based on state transitions at each discrete time step: at a given minute, customers arrive, customers finish being served, other customers begin being served, and emergency servers transition between serving as tellers and other tasks.

We differentiate between two types of servers: regular servers, who are ready to accept customers at all times, and *emergency servers*, who usually perform other tasks but can act as regular servers if needed, but only after some constant transition period.

We simulate the bank queue with this algorithm:

1. All 150 customers are initialized, each with random *t* (time between arrival of previous customer and arrival of current customer) and *s* (service time).
2. Using t_1, t_2, … t_{150}, arrival time (*a*) is computed for all customers using cumulative summation.
3. An internal clock variable *time* is set to 0 minutes. All customers are placed into a queue, and a list of servers is created.
4. Perform the following procedure until all consumers are served, the waiting queue is empty and no customers are in service:
   a. All customers who were calculated to arrive at this time step (*a* = *time*) are removed from the customer queue and added to the waiting queue.

b. All servers decrease their service counters by 1. If the counter reaches 0, remove the currently served customer and mark the server as inactive.

c. For each emergency server in the server list, increment the emergency-server-time-spent counter.

d. For each server that is currently inactive:

i. If this server is an emergency server and the waiting queue is smaller than the emergency server's exit queue length, mark the emergency server for removal from the server list.

ii. Otherwise, if there is anyone in the waiting queue, remove the next customer from the queue and record their waiting time (the difference between *time* and *a*). Set this server as active and set their service counter to *s* (the time this customer will be served for).

iii. If there is no one in the queue and this server is not an emergency server, increment the idle-time counter.

e. If there are emergency servers not on duty and the queue is larger than or equal to the emergency enter queue length, add an emergency server to the server list and decrement the number of not-on-duty emergency servers. Give this emergency server a service time equal to the transition time, as this server won't accept a customer until after the transition.

f. Record current values of queue length.

At the end of each run, average queue length for that "day" is found by summing all recorded queue lengths and dividing by the final value of *time*. Similarly, the average wait time for each customer is found by averaging the *w*s for all customers. In this way, our simulation accurately models the proceedings of a random day, and its versatility allows it to be easily extended to new circumstances, such as addition of new servers or changes in server efficiency. However, because our algorithm produces a single random outcome during each run, it can only approximate the true average wait time. Multiple trials thus become essential to increasing the confidence of our results.

By comparing our experimental distribution with our theoretical one for cases involving one server, we can determine the veracity and accuracy of our experimental results, allowing us to proceed in cases that the theoretical approach cannot handle well.

Model Data and Testing

We used Wolfram Mathematica to evaluate recurrence relations 2a and 2b from the starting point $W_1(0) = 1$, and found average wait time $\overline{w}$ to be 4.92761 minutes. We also computed a probability mass distribution $\overline{W}(x)$ (average probability that a customer waits x minutes). Interestingly, with the given probabilities for differences in arrival time and service time, there is a $\overline{W}(0) = 25.08\%$ chance that a given customer is served immediately: over a quarter of customers do not wait even with one server is present.

We ran our simulation 10000 times, recording wait times for each customer and average queue length for each trial. Plotting experimental wait times against theoretical wait times ($\overline{W}(x)$) , we see that the two distributions match almost exactly (Figure 1).

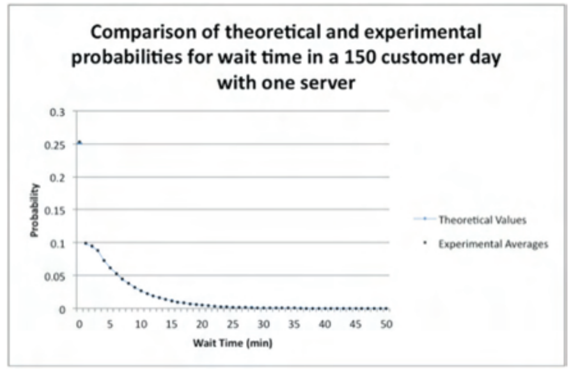

Figure 1

We found average wait time for the theoretical distribution to be 4.92761 min., whereas average wait time for the experimental distribution (10000 trials) was 4.9195 min., a difference of 0.164%. With this level of accuracy, we can safely extend our simulation to cases with multiple servers without worrying about insufficient confidence levels.

Because our theoretical model can't evaluate average queue times, we assumed that our simulation's accuracy in average wait times is indicative of overall accuracy, especially accuracy in average queue length. Though this assumption is not entirely justified, strong correlation between the theoretical and experimental wait times still helps to support the validity of our computation model.

With one server, the experimental average queue length was 1.84773, a value already less than the manager's goal of 2 minutes. Because our average wait time of 4.92761 minutes is well over 2 minutes, however, a strategy must be found to reduce this time. Also, the standard deviation of average wait times on a given day is 3.30591 minutes, indicating that the wait times vary greatly among days.

Potential Changes

We tested two changes in server structure to reduce the average customer wait time:

Increase the number of servers
Add "emergency" servers when queue length exceeds a certain number

Strategy 1: Increase the number of servers

By introducing one new server, we can reduce average queue length and average wait time. Using our simulation, we got an average wait time of 0.11285 minutes and an average queue length of 0.043093, well within the 2-minute and 2-customer goals. Furthermore, the standard deviation of the daily average wait times is 0.06283 minutes, indicating that wait time is consistently small. However, when two servers are assigned there is a large period of time with one or both idle (on average 429.8945 server-minutes, or 7.165 server-hours), resulting in reduced efficiency.

From the manager's perspective, adding a full-time server would reduce the average queue length and wait time to values below the goals, but at the same time this strategy would waste money on a full-time server who would only work some of the time.

Strategy 2: Add an emergency server

Unlike a dedicated server, an emergency server only works when queue length exceeds a predetermined limit (q_{enter}) and continues to work until no customers are in the queue. Because it is unreasonable for an emergency server to switch between tasks instantly, we added a two-minute transition before the emergency server could begin work.

We tested various values of q_{enter} to minimize the time the emergency server works while still keeping average wait time below two minutes. Using our simulation, we found average wait times and average queue lengths for the two-minute and two-customer goals. While adding an emergency server also creates time for which the dedicated server is idle, this idle time is much less than in strategy one. By setting q_{enter} to 3, we can minimize average server downtime to 61.7 minutes while still satisfying the goal. Our results are summarized in Table 1:

q_{enter}	Mean emergency server use time	Mean customer wait time	Mean queue length	Average number of of times emergency server is called	Mean server idle time
2	64.7958	0.927046667	0.350106628	10.7512	75.2787
3	38.5827	1.464951	0.552342	4.6354	61.7095
4	25.0496	2.034833	0.7677	2.3729	53.0709

Table 1: *Comparison of three values of q_{enter}. The emergency server with $q_{enter} = 3$ meets the goal while minimizing mean server idle time and time worked by the emergency server.*

Comparison of Strategies

Adding an emergency server with $q_{enter} = 3$ achieves the goal while limiting the time the additional server works to an average of 38.5 minutes per day and keeping server idle time to 61.7 minutes. This makes adding an emergency server the minimal change needed to accomplish the goal. A comparison of the approaches is shown in Table 2.

	Mean total server idle time	Mean additional server work time	Mean customer wait time	Mean queue length (customers)
Single server	36.8365 min	n/a	4.91953 min	1.847728
Emergency server ($q_{enter} = 3$)	61.7095 min	38.5827 min	1.464951 min	0.552342
Full-time second server	429.8945 min	397.6912 min	0.112851 min	0.043093

Table 2: *A comparison of three service systems. Note that the emergency server meets the goals required by the manager while minimizing the mean total server idle time.*

Sensitivity Analysis

To determine how our model responds to fluctuations in the original data, we ran our model using sets of slightly different probabilities. Our model proved to be very sensitive to small changes. For the single-server case, after increasing the arrival time distribution by 5%, average customer wait time increased to 8.2 minutes; after a similar decrease, wait time decreased to 3.1 minutes. Modifying service times in a similar fashion caused wait time to fluctuate between 6.0 minutes and 4.0 minutes.

When we tested these different distributions with the presence of an emergency server, however, average wait time stayed between 1.3 and 1.6 minutes.

These results show that a single-server system depends greatly on the distribution of customer arrival times and service durations. This makes sense: if the single server cannot keep up with customer arrivals, the queue quickly grows and increases waiting times. When an emergency server is present, however, this server is ready to step in once the queue increases in length. This demonstrates that our recommended emergency server system can adapt to random variation.

Strengths of Our Model

Our model uses both theoretical and computational methods. The theoretical approach uses purely mathematical methods, and its results are exact. The computational method uses a computer program, making it easily adaptable different entry and exit rates and addition of emergency servers.

Our computational method produces results nearly identical to the theoretical method. This verifies the precision and accuracy of our data and allows us to apply our computation model to different situations with high confidence.

We consider two methods of increasing customer satisfaction. Our model allows us to generate concrete data and easily determine which change is better using metrics including server idle time, average queue length, and average waiting time. This ensures that our suggested changes are both practical and effective.

Our model is easily adapted to different probability distributions, which makes it applicable other queues such as at other banks or other businesses.

Because our model produces a probability distribution for average customer wait time, businesses can extract other information about wait times, including, for instance, the probability that a customer waits more than 10 minutes.

Weaknesses of Our Model

We do not have a way to find exact theoretical values for the queue length distribution, so we have to approximate them with our simulation.

Due to the discrete data given us, our model cannot account for events that occur between minutes. This time gap is considerable, especially compared to the manager's two-minute goal, as many different events that significantly affect the inputs of our model can occur in these gaps.

Our model can't account for different rates of customer influx, such as more frequent arrivals during rush hour.

We do not consider other methods of increasing customer satisfaction, such as training workers to reduce service times.

Conclusion

We model a bank service queue with two approaches: theoretical mathematics and computational simulation. By using a series of recurrence relations on several probability mass distributions derived from the given data, we can compute exact average wait time for queues with a single server. However, because this method cannot account for multiple servers and cannot compute average queue length, we also used a computer program to simulate the activity of any one day. By comparing the theoretical and computational results for a queue with a single server, we are able to verify the reliability and accuracy of our computational model from the nearly identical wait time distributions and average wait times. With this increased confidence, we can extend our computational model to incorporate cases that our theoretical approach cannot handle, such as the addition of new full-time servers or emergency servers, while also estimating the average queue length by averaging over multiple trials.

Our computational model suggests that the best method is to implement a single emergency server who would only begin work when the queue length exceeds 3. If this is not feasible, adding an additional server would also reduce average queue length and average wait time significantly below the manager's requirements, though this approach would result in a significant amount of server idle time. The emergency server would only need to work for an average of about 40 minutes a day. This makes using a single emergency server more economical for the bank. Additionally, during the time when the emergency server is not interacting with customers, he/she could be performing other tasks. Through a sensitivity analysis, we also found that, when an emergency server was present, wait times fluctuated only slightly given different arrival and service rates.

In the future, we could apply our computational method to larger banks with more servers and more customers. It would be interesting to consider different probability distributions of customer arrivals depending on time of day. We could also look at the economic impacts of adding extra servers. Additionally, we could attempt to construct a computational method with continuous intervals between arrivals and continuous service lengths in order to model the situation with greater accuracy.

Editors note: Appendices A, B, C and D are omitted to conserve space.