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 2010

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 thirteenth HiMCM special issue. Since space does not permit printing all nine National Outstanding papers, this special section includes the summaries from seven papers and abridged versions of two. 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 2010 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

The High School Mathematical Contest in Modeling (HiMCM) completed its thirteenth year in excellent fashion. The mathematical and modeling ability of students, and faculty advisors, is evident in the professional submissions and work being done. The contest is still moving ahead slowly, growing with a positive first derivative, and consistent with our positive experiences from previous HiMCM contests.

This year the contest consisted of 295 teams with a total of 1131 students from 52 schools. These institutions represented twenty-two states and three different countries. Of the 1131 students, almost 40% were female students. The breakdown was 439 female and 692 male students. There were 31 all-female teams this year. This year, we again charged a registration fee of $75.

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.

Commendation: All students and advisors are congratulated for their varied and creative mathematical efforts. Of the 295 teams, 161 submitted solutions to Problem A and 134 to Problem B. The 36 continuous hours available to work on the problem provided for quality papers; teams are commended for the overall quality of their work.

Many teams had female members. There were 439 female participants on the 295 teams. There were 1131 total participants, so females made up over 38.8% of the total participation, showing this competition is for both genders. This percentage is almost triple the percentage of woman in other math competitions. There was at least one female on most of the teams and 10.5% of the teams were all female (31 teams).

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 a success!

Judging: We ran three regional sites in December 2010. They are:
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 Outstanding, Meritorious, Honorable Mention, and Successful Participant. All finalist papers for the Regional Outstanding award were sent to the national judging. For example, eight papers may be discussed at a Regional Final and only four selected as Regional Outstanding, but all eight papers are judged for the National Outstanding award. Papers receive the higher of the two awards.

The national judging chooses the best of the best to receive the National Outstanding award. 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 continues to grow.

Judging Results:

Problem	National Outstanding	Outstanding	Meritorious	Honorable Mention	Participant	Total
A	3	21	42	66	29	161
%	2%	13%	26%	41%	18%	
B	6	9	34	62	23	134
%	5%	7%	25%	46%	17%	
Total	9	30	76	128	52	295
%	3%	10%	26%	43%	18%	

National Outstanding Teams

Evanston Township High School, Evanston, IL
Glenbrook North High School, Northbrook, IL
Illinois Mathematics and Science Academy, Aurora, IL
Hanover High School, Hanover, NH
Mills Godwin High School, Richmond, VA
Shanghai Foreign Language School Affiliated to SISU, Shanghai, China
Hong Kong International School, Hong Kong (3 Teams)

Regional Outstanding Teams

Albuquerque Academy, Albuquerque, NM
Central Academy, Ankeny, IA
Charter School of Wilmington, Wilmington, DE (3 Teams)
Chesterfield County Mathematics and Science High School, Midlothian, VA (2 Teams)
Evanston Township High School, Evanston, IL (2 Teams)
Hanyoung Foreign Language High School, Seoul, Korea
Hong Kong International School, Hong Kong (4 Teams)
Illinois Mathematics and Science Academy, Aurora, IL (3 Teams)
Maggie Walker Governor's School, Richmond, VA (5 Teams)
Middlesex School, Concord, MA
Mills Godwin High School, Henrico, VA (2 Teams)
NC School of Science and Mathematics, Durham, NC
NO. 2 High School of East China Normal University, Shanghai, China
North Springs Charter School, Atlanta, GA
Shanghai Foreign Language School, Shanghai, China
The Ellis School, Pittsburgh, PA

NCTM Standards: The director and the judges asked that we add this paragraph. Many of us have read the NCTM standards and clearly realize the mapping of this contest to the NCTM 9–12 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' commentaries provide 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, and clear. 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, and its solutions and then support the findings mathematically generally do quite well. Modeling assumptions need to be listed and justified, but only those that come to bear on the solution (and that can be part of simplifying the model). Laundry lists of assumptions that are never referred to in the context of the model development are not considered relevant and detract from a paper's quality. The mathematical model needs to be clearly developed, and all variables that are used need to be well defined. Thinking outside of the "box" is also considered important by the 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 weaknesses 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 very important within the paper as well as either a references or bibliography listing at the end. We encourage citations within the paper in sections that deal directly with data and figures, graphs, or tables. We have noticed an increase in the use of Wikipedia. Teams need to realize that although useful, the information in Wikipedia might not be accurate; teams need to acknowledge this.

FACTS FROM THE 13TH ANNUAL CONTEST:

- Wide range of schools/teams competed including teams from Hong Kong and China.

- The 295 teams representing U.S. and international institutions constitute a 6.4% increase in participation.

- There were 1131 student participants, 692 (61.2%) male and 439 (38.8%) female. There were 31 all-female teams.

- Schools from only twenty-two states participated in this year's contest.

THE FUTURE:

The HiMCM, which aims to give the under-representative 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 it 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 that can compete. 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 a key to future success. The abilities to recognize problems, formulate a mathematical model, use technology, and communicate and reflect on one's work are keys 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. What is fundamental and matters most is not the technique used but the process that discovers how assumptions drive the techniques. Let students practice being 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.

Contest Dates: Mark your calendars early: the next HiMCM will be held in November 2011. Registrations are due in October 2011. Teams will have a consecutive 36-hour block within the contest window to complete the problem. Teams can register via the Internet at www.himcm.org.

Math Models.org

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

Judges Commentary, 13th Annual HiMCM

Problem A: Bicycle Club

Several cities in the United States are starting bike-share programs. Riders can pick up and drop off a bicycle at any rental station. These bicycles are typically used for short trips within the city center, either one-way or roundtrip. The idea is to help people get around town on a bike instead of by car. Those making longer trips (such as commuting to work) are likely to use their own bikes.

Some of the challenges are how to determine where to locate the rental stations, how many bikes to have at each station, how/where to add new locations as the program grows, and how many bikes to move to another location and when (time of day, day of week).

The downtown city maps, the bike rental locations and the number of bikes at each location for Chicago, Denver, and Des Moines are available from the following websites:
http://chicago.bcycle.com/
http://denver.bcycle.com/
http://desmoines.bcycle.com/

You have been asked to develop an efficient bike rental program for these cities.

- List the traffic/bike usage and other information that you would need to collect in order to plan the bike rental program for these cities.

- Develop a mathematical model that the city could use to plan the program, including the location of new rental stations for the next 5 years.

- Assume that the bike usage in the program will grow by 30% per year.

In your analysis consider the existing bike paths in the city center, attractions such as museums, theaters, etc., in the city center, and the other transportation hubs in the city center. When your analysis is complete, prepare a short letter to the mayor explaining the benefits and recommendations of your analysis.

Judge's Comments:

Author: Veena Menderatta
William P. Fox, HiMCM Contest Director

This problem is of interest to the author. Originally proposed as a problem for the college level MCM, the problem was selected for the HiMCM by the contest directors because it can be addressed with high school mathematics.

This raises the question: What are reasonable expectations for a high school team, in 36 hours, to model this situation? As a regional head judge and one of the national judges, and as the problem author, I offer the following insights.

First, and this comment is made each year, if the problem asks for a letter, news article, or position paper, the student group must provide one if they hope to be recognized. Additionally, the letter should be written after the problem is solved, as the letter must contain facts and results that will excite the reader to look at the entire analysis. Most of the letters did not report their facts and thus, would not motivate a mayor or anyone else to examine the analysis more closely.

Second, the problem statement explicitly called out three components to be addressed (see problem statement above). Addressing all three greatly increases the chance of recognition.

Although not explicitly asked for it would be hard to address these three issues without considering costs.

The better papers this year attempted to present frameworks for choosing solutions. The mathematics required to do this are very accessible at the high school level. The approaches for traffic density were mostly good. Few teams, if any, considered the impact of sharing the road or riding in darkness as most teams had the bike kiosks open 8–11 p.m.

This year's papers have many strong points. Almost all the papers did a reasonable job of estimating the demand for bikes but few discussed the issue of weather and how that might impact the use of bikes. Several of the papers did excellent jobs modeling bike usage and comparing it to other modes of traffic.

There were a wide variety of approaches used from simple algebra or statistics through simulation models. We found many simulation models were not well explained, nor were flow charts presented and used. It was as if these techniques were a black box. As models, they should be explained as to what they do and why they could be used in the scenario.

A few of the papers did outstanding jobs of representing their strategies graphically.

There were some notable patterns of weakness. Many papers never considered foul weather, like snow and ice in cities such as Chicago. Many did not look at all three required facets of the problem. Others forced the use of calculus in their solution, although it was not really appropriate. However, some papers offered no mathematical treatment of the problem at all.

Student groups should remember that the problems posed in these contests are not going to have a unique solution—they are designed **not** to have one, in fact. Students also should remember that general high school mathematics are adequate to the task at hand—what we are looking for is evidence of good modeling of the problem with these tools, and then discussion of the implications of the model and its solution(s). We are looking for the quality of creative modeling and a thorough job of implementing the modeling process.

Problem B: Curbing City Violence

A regional city has had lots of problems with gangs and violence over the years. The mayor, chief of police, and city council need your help. Data are available for the following: Incidents of violence, homicides, and assaults, regional population numbers (from U.S. Census data), unemployment rate, high school enrollment numbers, high school dropout rate and graduation rate, prison population numbers, numbers of prisoners released on parole, parole violations, percent of parole violations, and juvenile inmates.

Analyze and model these data to give the city a plan to reduce violence. After you complete your analysis and model, prepare a news release for the mayor briefly outlining your proposals that recommend a campaign strategy to curb the violence.

Some real data was provided to the students.

Judge and Author's Comments

William P. Fox, HiMCM Contest Director

This problems statement is concise but clearly has elements for the students to consider. Students should have clearly defined what they considered "violence" and how they were going to measure it.

Most students completed most of the required tasks. Many did not pick the variables that impacted violence the most and therefore did not discuss how to control those variables. Nearly all teams wrote the letter, but few did what we would call an excellent job of concisely telling their story and relating the facts in the letter. Thus, teams should ensure that that they complete and include all the required tasks in their submission.

The executive summaries for the most part were either absent or poorly written. This has been ongoing since the beginning of the contest. Faculty advisors should spend some time with teams on how to write a good summary. Many summaries read like technical reports or were too vague to be helpful. Summaries need to contain the results of the model as well as brief explanation of the problem. A summary should entice the reader, in our case the judge, to read the paper.

The letter to the mayor (news release) should have been a concise explanation of the modeling results to include (1) defining violence and why controlling it is important, (2) listing the variables that most influence violence, and (3) a brief description or statement of the potential impacts and changes to reduce violence. Again, many teams failed to do this in their submission.

The judges felt the first critical task was to define violence and define measures that could be used to measure such violence.

The modeling seen was not based on the first principles of the modeling process. Students obtained scatter plots to look for correlations and built linear regression models. Some built multiple regression models. Almost all teams used the variables as presented without forming maybe ratios or creating new variables. The data were integer data, yet models often had so many decimals points it was absurd. Some teams built regression models of higher order (8 data pairs and a 7^{th} order polynomial). Teams did not graph their polynomials (lower or higher order) to check to see if the trends were always captured. Few teams looked at residuals, percentage errors, or anything other than r^2 to determine the adequacy of the model.

Students should review this example:

Consider the following 4 sets of data:

I

x	y
10.0	8.04
8.0	6.95
13.0	7.58
9.0	8.81
11.0	8.33
14.0	9.96
6.0	7.24
4.0	4.26
12.0	10.84
7.0	4.82
5.0	5.68

II

x	y
10.0	9.14
8.0	8.14
13.0	8.74
9.0	8.77
11.0	9.26
14.0	8.10
6.0	6.13
4.0	3.10
12.0	9.13
7.0	7.26
5.0	4.74

III

x	y
10.0	7.46
8.0	6.77
13.0	12.74
9.0	7.11
11.0	7.81
14.0	8.84
6.0	6.08
4.0	5.39
12.0	8.15
7.0	6.42
5.0	5.73

IV

x	y
8.0	6.58
8.0	5.76
8.0	7.71
8.0	8.84
8.0	8.47
8.0	7.04
8.0	5.25
19.0	12.50
8.0	5.56
8.0	7.91
8.0	6.89

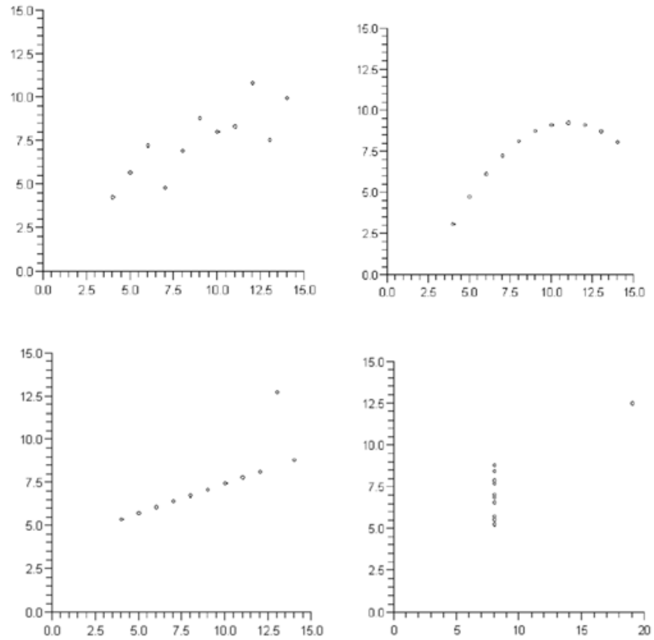
Suppose we fit the model $y = ax + b$ to each data set using the least-squares criterion. In each case the following model results:

$$y = 3 + 0.5x$$

The correlation coefficient in each case is 0.82, and $r^2 = 0.67$. The sum of the squared deviations between observed and predicted values is also the same. In particular,

$$\sum_{i=1}^{11}[y_i-(3+5x)]^2 = 13.75$$

These two numerical measures imply that for each case $y = 3 + 0.5x$ does about the same job explaining the data, and that it is a reasonable fit ($r^2 = 0.67$). However, the following scatter plots



convey a different story:
A point to consider is how well the model $y = 3 + 0.5x$ captures the trend of the data. (This example is adapted from F. J. Anscombe, "Graphs in Statistical Analysis," *Amer. Stat.*, 27, 1973, 17–21.)

Few teams, if any, did any sensitivity analysis on the data or the model.

We found that many of the assumptions and research were very good. Teams did research into the history of violence and data on violence, but none used their modeling efforts to see if the new data followed the same trends. We encourage teams who take data from other sources or graphics to ***include the reference at that point*** as well as a reference page at the end. We saw the use of data from blogs and Wikipedia—such information can be suspect, and we encourage teams to obtain data and information from reliable sources.

There were a wide variety of approaches used from simple algebra through simulation models. We found many simulation models were not well explained, nor were flow charts used. It was as if these techniques were a black box. As models, they should be explained as to what they do and why they could be used in the scenario.

One issue concerned significant digits. The models built by the teams were in number of violent acts of some magnitude. Yet numerical values were presented to (at times) many (up to 20) decimal places. Clearly, this was not necessary.

General Comments from Judges:

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

Computer-generated Solutions: Many papers used computer code. Computer code used to implement mathematical expressions can be a good modeling tool. However, the 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 may only be read for the papers that reach the final rounds, but not unless the code is accompanied by a good algorithm in words. The results of any simulation need to be well explained and sensitivity analysis preformed. For example, consider a flip of a fair coin. Here is a general algorithm

INPUT: Random number, number of trials

OUTPUT: Heads or tails

Step 1: Initialize all counters.

Step 2: Generate a random number between 0 and 1.

Step 3: Choose an interval for heads, like [0, 0.5]. If the random number falls in this interval, the flip is a head. Otherwise the flip is a tail.

Step 4: Record the result as a head or a tail.

Step 5: Count the number of trials and increment: Count = Count + 1.

An algorithm such as this is expected in the body of the paper with the code in the 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. They 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 vital to the building and simplifying of their mathematical model. Assumptions should not be a reiteration of facts given in the problem statement. 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 appear 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 impress the judges, as this does not work. We prefer to see sound modeling based on good reasoning.

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 runs 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. However, 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.

Adherence to Rules: Teams are reminded that detailed rules and regulations are posted at 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: Shanghai Foreign Language School

Advisor: Yue Sun

Team Members: Zemin Yu, Honghao Zhang, Meng'ou Zhu, Lingzi Zhuang

Our model is 5-year plan for the development of bicycle-sharing programs in Chicago, Denver and Des Moines, aiming to popularize a new means of transport and to relieve the city center of traffic overflow.

The basis of the model is a database of roads, bike lanes, buildings, facilities and existing bike-rental stations in the three cities. All sites of buildings, facilities and existing rental stations are marked with accurate longitudes and latitudes. With this database, we define the city center for every city. According to our analysis, we should set up an initial network of bicycle-sharing program (including bicycle rental stations and bike lanes) in Des Moines and Chicago, where the existing bicycle service is far from enough. Especially, in Chicago, a metropolis covered with a developed subway system, we should place bicycle rental stations near subway stations. In Denver, a city already covered with a bicycle-sharing network, we may directly embark on improvements.

The five-year program comes in three parts. The first part is a two-year initial plan for Chicago and Des Moines, with which we build an initial network of bicycle rental stations and lanes in these two cities. When looking for ideal positions to place new stations, we divide the buildings/facilities in the city center into several types and then apply the analytic hierarchy process (AHP) to quantize the different significance of these types. Afterwards, we weigh the significance values of buildings with the distance between the buildings and the new station so as to find the ideal location of the rental station. When building bike lanes, we apply linear regression to find a line such that the total distance between stations and this line is minimized. Then we fit the line into actual streets with appropriate linkage to existing lanes.

The second part is an additional plan to improve the initial bicycle-sharing program in the cities and thus includes all three cities. We deal with addition of new stations and lanes. We apply the Central Place Theory and find out the initial service range of every existing bicycle rental station. With the increase in demand for bicycles, the service range of each station will decrease year by year, and some “blind areas” will be left out of the coverage of the program. New stations should be built in these areas to maintain coverage of service and new lanes should be built if these new stations are far away from existing lanes.

The third part is the discussion on allocation of bicycles and regulation of bicycles according to the change in demand. We apply the Markov Chain Model to decide the different status change of bicycle stations on working days and work-free days, and distribute bicycles according to the steady state of the Markov Chain.

Finally, we put all the results together and make annual plans for each city, and prepare a letter to the mayor recommending our plan.

Problem A Summary: Hong Kong International School

Advisor: Edgar Fong
Team Members: Janet Kwan, Chloe Nguy, Anthea Wong

Based on existing bike stations in Chicago, Denver and Des Moines, our group aims to develop an efficient bike rental program and determine the locations in which new bike stations should be constructed.

We first determined the factors that would impact bike usage. By categorizing existing bike stations according to zip codes, we found a general positive relationship between the number of bike stations and each of the various factors we studied, such as population density, civic center popularity, and the number of tourist attractions in within the zip code. After deriving and combining the functions, we were able to arrive at a final rating system:

$$R = 3.1964e^{3.6396C} + T + (3.926285282 \cdot 10^{-5})(1.0000709285)^{P} + \frac{0.00141}{\sqrt{0.0067125S}}$$

This rating system considers the relative weightings of each variable, and can be used to evaluate the need for new biking stations at certain locations.

To locate the different sites for upcoming bike stations that would efficiently serve the increase in bike usage and demand, we substituted data in our model to obtain ratings of the zip codes near city center. By comparing these different rankings and considering the maximum number of stations an area can carry, we found the ideal locations for the expansion of the bike program in the next five years in zip codes: 60611, 60605, 60616, 50309, 50312, 80204, 80205, 80206, 80209, 80210, 80211, 80218, 80223, and 80207.

Problem A Paper: Hanover High School

Advisor: William Hammond
Team Members: Scott Collins, Leah Eickhoff, Diksha Gautham, Jake van Leer

Restatement of Problem.

One of the most effective ways to reduce carbon emissions and increase good health is to switch from using cars to riding bicycles. Several U.S. cities have bicycle rental programs through which members can rent cycles from automated kiosks. These programs attempt to provide a cheap, convenient, and healthy alternative for short-distance transportation. Program members can pick up and drop off bicycles at any of the kiosks, which are scattered at convenient locations. Several cities, including Chicago, Des Moines, and Denver, are starting such programs. Chicago’s pilot program, starting with 100 bikes dispersed among six kiosks, ran from July 30 through November 1. During the first 17 days, more than 1,500 bicycles were rented and 80 memberships in were sold. Due to its overwhelming success, the Chicago program will restart on March 31. As membership increases, the number of bikes needs to increase, as does the number of kiosks. New kiosks should be added in locations where they are most helpful, taking into account attractions such as shopping, museums, theaters, and parks, as well as transportation hubs such as train stations.

The Des Moines program has an 18-bicycle squadron and four stations. The Denver program has 1,000 bikes and almost 50 kiosks for the 554,636 residents.

Assumptions and Justifications

- People use bike kiosks that are nearest to their current locations. People do not walk or drive long distances to rent bikes.
- People rent bikes at a random rate. There is no general formula as to exactly when people rent bikes.
- People drop bikes off at kiosks nearest their final destinations. People do not have to return the bikes to the kiosks from which they were rented. If they do return the bikes to the same kiosk, the values in our model balance out, in essence negating each other.
- Once a bike is rented, it is returned by the end of the day. This is one of the rules of rental. Each bike has its own GPS system, so if bikes are stolen, rental companies can quickly find them.
- Each attraction generates the same amount of interest. Admittedly faulty, this assumption allows us to simplistically model the situation. However, in future improvements to our model, we can easily weight attractions.
- The number of bikes currently in use is sufficient for current needs; i.e., there is no pressing demand for more bikes, or there is no excess of unused bikes.
- The biking area is limited to the area defined by maps on the B-Cycle website. People do not bike beyond the boundaries a map outlines.
- Population distribution in each city is equal. For the purposes of our model, an area covering 37% of the city contains 37% of its inhabitants.
- In Des Moines, the lake is a popular point of interest. When we added new kiosks to our Des Moines model, we assumed that many people would go to the lake and bike on the trails. Thus, we weighted the lake as 2 points of interest instead of one.
- Although kiosks open at 5 a.m., we assume that B-cycle members do not begin renting bikes until 8 a.m.

MODEL AND APPLICATIONS

Des Moines

In Des Moines, the B-Cycle program currently has 4 kiosks. There are 18 bikes in circulation, which per our assumptions are sufficient for the city's current needs. Kiosk A is at the intersection of 13th and Grand, Kiosk B is at 7th and Grand, Kiosk C is at the Brenton Skating Plaza, and Kiosk D is at Principal Park. We created a Voronoi diagram (Figure 1) to divide the map given on the B-Cycle website into the domains of the kiosks. We found that the area is divided as follows: Kiosk A = 37%, Kiosk B = 13%, Kiosk C = 21%, and Kiosk D = 29%.

To determine where bikes are left, we calculated how many points of interest are in each region. We found the percentage in each area, and figured that those numbers would be the proportion of

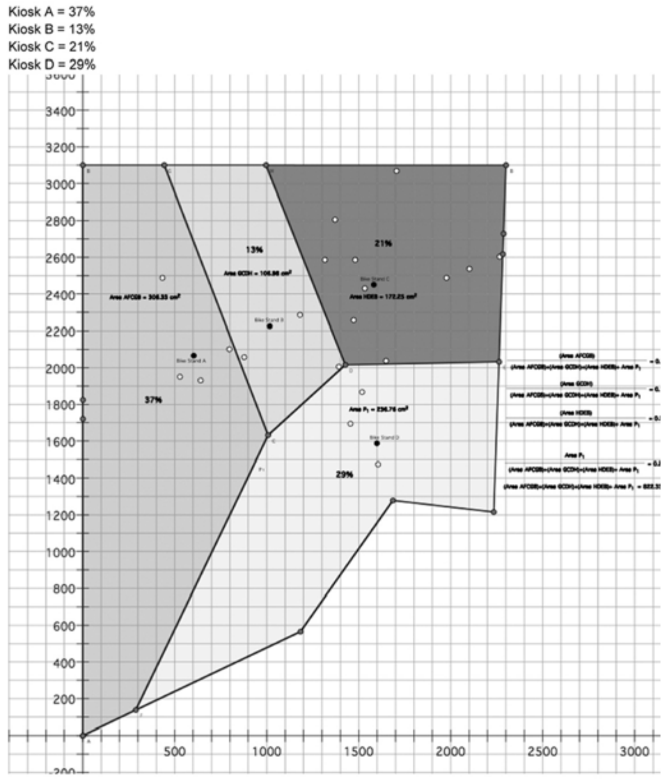


Figure 1.

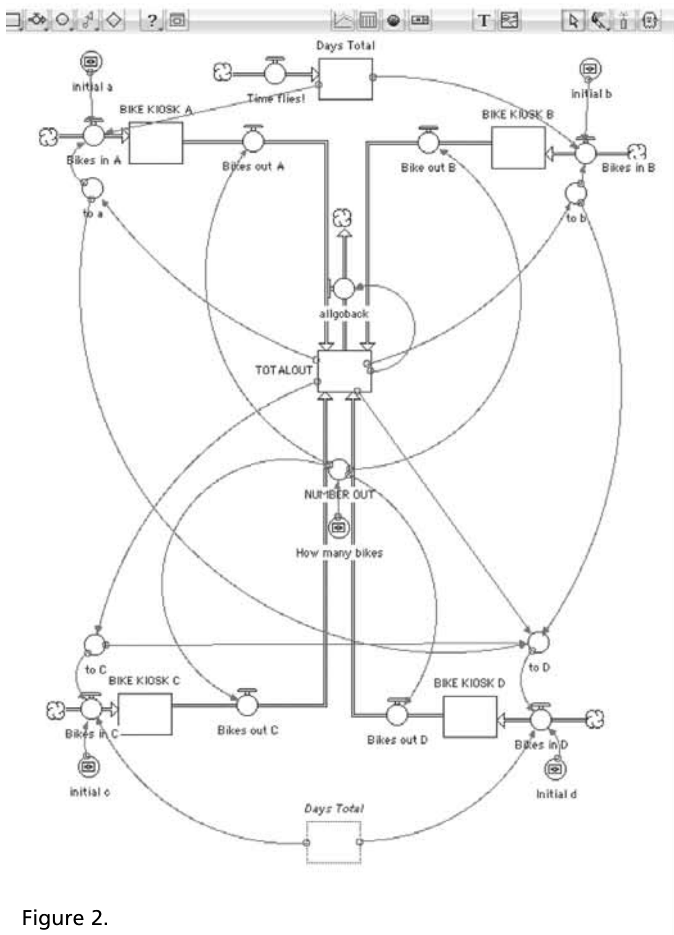


Figure 2.

```
BIKE_KIOSK_A(t) = BIKE_KIOSK_A(t - dt) + (Bikes_in_A - Bikes_out_A) * dt
INIT BIKE_KIOSK_A = 0
 INFLOWS:
   Bikes_in_A = IF(Days_Total=0) THEN (initial_a)
      ELSE(to_a)
 OUTFLOWS:
   Bikes_out_A = ROUND(.37*NUMBER_OUT)
BIKE_KIOSK_B(t) = BIKE_KIOSK_B(t - dt) + (Bikes_in_B - Bike_out_B) * dt
INIT BIKE_KIOSK_B = 0
 INFLOWS:
   Bikes_in_B = IF(Days_Total=0) THEN (initial_b)
      ELSE(to_b)
 OUTFLOWS:
   Bike_out_B = ROUND(.13*NUMBER_OUT)
BIKE_KIOSK_C(t) = BIKE_KIOSK_C(t - dt) + (Bikes_in_C - Bikes_out_C) * dt
INIT BIKE_KIOSK_C = 0
 INFLOWS:
   Bikes_in_C = IF(Days_Total=0) THEN (initial_c)
      ELSE(to_C)
 OUTFLOWS:
   Bikes_out_C = ROUND(.21*NUMBER_OUT)
BIKE_KIOSK_D(t) = BIKE_KIOSK_D(t - dt) + (Bikes_in_D - Bikes_out_D) * dt
INIT BIKE_KIOSK_D = 0
 INFLOWS:
   Bikes_in_D = IF(Days_Total=0) THEN (initial_d)
      ELSE(to_D)
 OUTFLOWS:
   Bikes_out_D = ROUND(.29*NUMBER_OUT)
Days_Total(t) = Days_Total(t - dt) + (Time_flies!) * dt
INIT Days_Total = 0
 INFLOWS:
   Time_flies! = 1
TOTALOUT(t) = TOTALOUT(t - dt) + (Bikes_out_A + Bike_out_B + Bikes_out_D + Bikes_out_C - allgoback) *
dt
 INFLOWS:
   Bikes_out_A = ROUND(.37*NUMBER_OUT)
   Bike_out_B = ROUND(.13*NUMBER_OUT)
   Bikes_out_D = ROUND(.29*NUMBER_OUT)
   Bikes_out_C = ROUND(.21*NUMBER_OUT)
 OUTFLOWS:
   allgoback = TOTALOUT
How_many_bikes = 18
initial_a = 4
initial_b = 4
initial_c = 5
Initial_d = 5
NUMBER_OUT = ROUND(RANDOM(0,How_many_bikes))
to_a = ROUND(.2*(TOTALOUT))
to_b = ROUND(.15)*TOTALOUT
```

Figure 3.

bikes returned to each kiosk. Our logic is based on the assumption that once someone visits a point of interest, they are likely to leave the bike at the nearest kiosk. These percentages are: Kiosk A = 20%, Kiosk B = 15%, Kiosk C = 50%, and Kiosk D = 15%.

We used this information to create a model with Stella software (Figures 2 and 3)

This model represents the flow of bikes in and out of kiosks over time. Each station starts with a fixed number of bikes, determined by our group. The number of bikes in use at any time is randomly determined as a number between zero and the total number of available bikes. This simulates the fact that not all bikes are in use at all times—essentially, it allows for variability. From where these bikes leave is determined by the percentages of area covered by kiosks (per our Voronoi diagram). The number of bikes leaving Kiosk A is 0.37*(number of bikes out). For Kiosk B, the number leaving is 0.13*(number of bikes out). For Kiosk C this is 0.21*(number of bikes out) and for Kiosk D this is 0.29*(number of bikes out). All bikes are then returned after 1 hour. Where the bikes are returned is determined by the percentage of points of interest in the kiosk's area. The number of bikes out in the previous step is split so that 20% return to Kiosk A, 15% return to B, 50% return to C, and 15% return to D, at which point the number of bikes out has been depleted and the stock is regenerated randomly. Of course, people may keep bikes for more than an hour, but this means that one of the bikes taken out can be assumed to have never gone back in the next step. For example, if 6 bikes are out, and the next hour 8 are out, it is possible that only 5 bikes went back and 7 new bikes came out, but one bike was not returned.

Simplified, the number of bikes leaving is random, but the number of bikes returning is always equal to the bikes that left during the previous unit of time. We feel that this gives elegance to our model, keeping it both simple and realistic. It's impossible to know how many people will want to use a bike at a given time, something that we sidestepped using our random system.

YEAR 1

We used this model to determine what a possible outcome of the B-Cycle setup might be. We found that bikes tend to accumulate at Kiosk C, which makes sense, as it has half the points of interest in its domain. Kiosks A, B, and D tend to run out quickly—after 3–5 hours. (See Figure 4.) We found that optimal starting values for bikes (using the B-Cycle website's value of 18 bikes in use) are: 7 in A; 5 in B; 1 in C; 5 in D. This runs successfully for 6 hours, sometimes more, without depletion of a kiosk's bikes. Since most people won't begin using bikes until 8 a.m., by 2 p.m. bikes will have to be shipped to other kiosks from Kiosk C.

Something we will consider in placing new kiosks is how to divert bikes away from Kiosk C.

9:26 PM 11/9/10 Table 1 (Untitled Table)

Time	BIKE KIOSK	BIKE KIOSK	BIKE KIOSK	BIKE KIOSK	TOTALOUT
0	0.00	0.00	0.00	0.00	0.00
1	6.00	4.00	0.00	4.00	4.00
2	6.00	4.00	2.00	4.00	2.00
3	5.00	4.00	3.00	4.00	2.00
4	5.00	4.00	4.00	5.00	0.00
5	2.00	3.00	2.00	2.00	9.00
6	2.00	2.00	6.00	3.00	5.00
7	0.00	0.00	6.00	0.00	12.00
8	0.00	0.00	10.00	1.00	7.00
9	0.00	0.00	13.00	2.00	3.00
Final	0.00	0.00	14.00	1.00	3.00

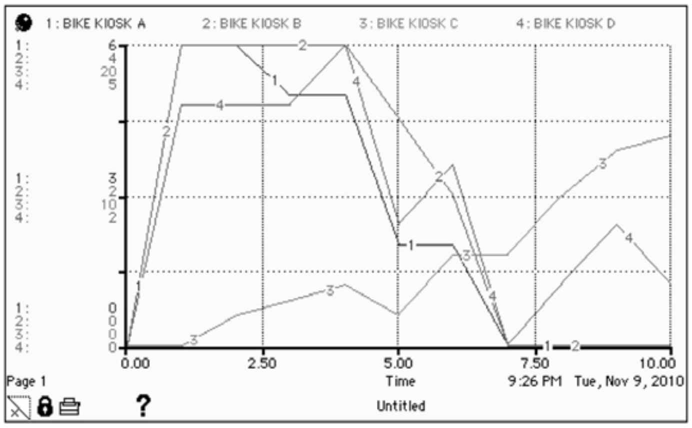


Figure 4.

Growth of Bike Use

Over 5 years, the program is projected to grow by about 30%, and there will be 51 bikes total in the system after 5 years. We have modeled this growth with a different Stella model (Figures 5 and 6).

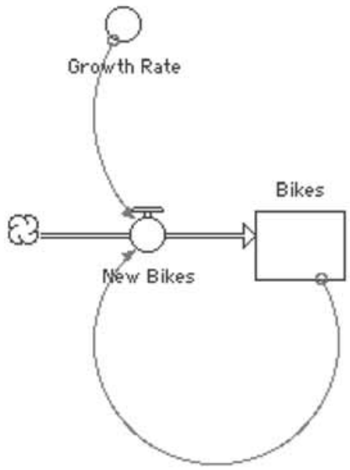


Figure 5

12:41 PM 11/9/10 Table 1 (Basic Bike Growth Model Test)

Time	Bikes
0	18.00
1	23.40
2	30.42
3	39.55
4	51.41
Final	66.83

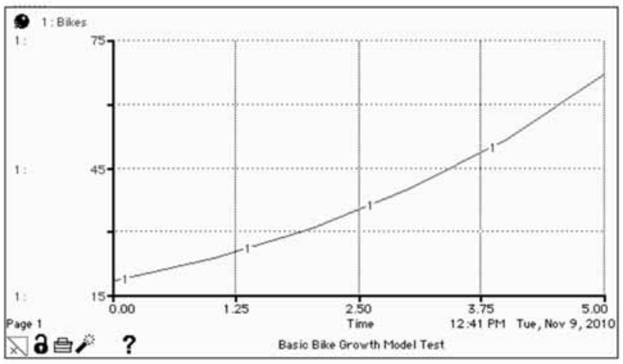


Figure 6

Editor's note: Figures 1–6 are representative of those in the Hanover team's paper, and we have omitted the remainder due to space limitations.

Without doubt, as the number of bikes increases from 18 to 51 over 5 years, the city will need more kiosks.

YEAR 2

In the second year, the model will have grown by 30%, or 5 bikes. Using the same Stella model, we found new optimal values. By placing 9 bikes at Kiosk A, 7 bikes at Kiosk B, 1 bike at Kiosk C, and 6 bikes at Kiosk D, we can make the program run for 7 hours without incident. Some kiosks run for even longer, but once again, after 7 hours (by 3 p.m.) bikes have to be shipped away from Kiosk C.

First Expansion (Year 3)

Where is the best place for the first new kiosk? Looking at our Voronoi diagram and the map of the region, the best place is in the lower left-hand corner, where there is a large lake and many bike trails. People will rent bikes from any kiosk, ride to the lake and around its trails, and return the bike to the kiosk by the lake. Since the lake is such a great point of interest, we weighted it as 2 points of interest.

When we make a new Voronoi diagram, which factors in this new kiosk, the distribution of area among regions is more balanced; region A no longer has so much space. The new area distribution is Kiosk A =24%, Kiosk B = 13%, Kiosk C = 21%, Kiosk D = 21%, and Kiosk E = 20%.

We adjusted our STELLA model to accommodate the new kiosk (E). What this entails is essentially changing the in and out rates of all the kiosks to accommodate the new one (see above for new percentages).

Using our model, we found new optimum values. By starting Kiosk A at 9 bikes, B at 7, C at 1, D at 6, and E at 7, we managed to let the model run for about 4 hours. This should be taken with a grain of salt, as people will likely rent bikes at the lake and return them there. This creates a loop, and Kiosk E may not run out as quickly as it seems, and the model would run longer.

Second Expansion (Year 4)

We should place a second new kiosk to accommodate the return of the new bikes. When we run our model, Kiosk C fills up quickly. So, we should put this kiosk in region C. The new area distribution is: Kiosk A =24%, Kiosk B = 13%, Kiosk C = 10%, Kiosk D = 21%, Kiosk E = 21%, and Kiosk F = 11%.

This evens out the land and population distributions nicely, and, more important, splits up the points of interest in the upper right-hand corner. Our next step was to re-form our STELLA model. Much like the other years, this reenters the rates and accommodates a sixth kiosk (F). We proceeded to find optimal values: 10 bikes should go to A, 7 to B, 1 to C, 9 to D, 11 to E, and 1 to F. This allows the system to run for somewhere between 5 and 6 hours on average and accommodate the growth.

Third Expansion (Year 5)

Where to put a third new kiosk? With a 30% growth rate, a second kiosk will accommodate 9 more bikes after the fourth year, and there will be 51 bikes in the system.

Running our model, we find that now region C ends up with a lot of bikes. It's not excessive or uncontrollable, but it would be helpful if there were another kiosk in the area to collect bikes. Thus, we can split up the points of interest and add a third kiosk near regions F and C. The new area distribution is: Kiosk A = 24%, Kiosk B = 9%, Kiosk C = 7%, Kiosk D = 21%, Kiosk E = 21%, Kiosk F = 9%, and Kiosk G = 9%.

We need 8 bikes at A, 5 at B, 2 at C, 15 at D, 14 at E, 2 at F, and 5 at G. This system is successful for 6 hours, but with the exception of Kiosk E, by the lake, which runs for 9 hours. The more kiosks and bikes, the more complex the model, but with more bikes to circulate, the model tends to last longer.

As time continues, the model can be easily adapted to include additional kiosks and rising bike numbers by simply adding a new kiosk and adjusting the rates. This is the benefit of our model. Despite its apparent complexity, it is relatively simple to adapt.

Chicago

Due to time constraints, we created a single, in-depth model for Des Moines, rather than three separate models for each city. This Des Moines model serves as a template for the other cities; it is easy to replicate and adjust it. The same mathematics applies to this city, as well as any other that our model is used for.

Chicago was reported to have 100 bikes in circulation over 6 kiosks. The great thing about our model for Des Moines is that it is easy to apply to another city. The only requirements are a Voronoi diagram (to show percentage of area covered by each kiosk) and the points of interest in the city (to apply to our weighting system).

Conclusion

We developed a strong model for the development of bike-sharing programs in Des Moines, Chicago, and Denver. Our relatively simple model uses a Voronoi diagram, making portions based on proximity to bike kiosks, as well as the points of interest in the regions. We use the percentages of area and destinations to determine how many bikes enter and leave each kiosk every hour. Using our Stella program, which models movement of bikes, we optimized the initial number of bikes at each kiosk, even as the number of bikes and the number of kiosks grows over five years. Our model takes into account many important factors, but is easily adaptable to any city, regardless of layout, bike use, and kiosk locations.

STRENGTHS AND WEAKNESSES

- Since we were severely limited time-wise, we decided to focus on an in-depth model of only one city, with the assumption that our process could be replicated (with more time) for the other two.

- Many of our assumptions are inaccurate, but due to time constraints, we were forced to use them. We assumed that the parameters of the biking area are limited to the area defined by maps on the B-Cycle website. In reality, as program membership increases, the regions covered by the system will expand outside current boundaries. Also, we assumed that population distribution is equal throughout each city. This assumption is inaccurate because any city has areas of varying population density. Another faulty assumption is that all attractions generate the same amount of interest. To improve our model, we would need to weight the attractions based on member interest.

- We focused only minimally on the roads and paths themselves. Biking on a busy road is unpleasant, and bicyclists tend to avoid those routes in favor of less busy streets and pathways, regardless of directness.

Problem B Summary: Illinois Mathematics and Science Academy

Advisor: Steven Condie

Team Members: Aditya Karan, Nilesh Kavthekar, Abhinav Reddy, Nishith Reddy

Violence is a prevalent issue in every society. Although in some cities, this is a worse problem than others. The particular city we were given was plagued by a high rate of violent incidents and people in jail. The problem was so bad that there were more people in the city's jail than there were out of jail. Problems stemmed from the lack of high school enrollment, low rates of graduation, high dropout rates, and unreasonable unemployment rates, all leading to a high incidence of violent crimes. The height of violent crimes led to more incarcerations of both adults and juveniles. Our model took into account these factors by relating the causes to the incidence of violent crimes that the mayor and others were seeking help on.

Our approach to this dilemma is to emphasize ways to reduce the problems in the lack of high school enrollment, low rates of graduation, high dropout rates, and unreasonable unemployment rates. In order to do this we proposed instituting after-school programs that have been shown to decrease high school dropout rates and increase graduation rates. We also proposed vocational education in schools, which has been shown to decrease dropout rates. The last part of our solution was to build a new school in the city. Similar endeavors have been shown to increase enrollment rate. When all of these improvements were taken into consideration, there was a direct 10% decrease in incidence of violent crimes. This decrease of violent crimes from 711 a year to 641 a year is a very great accomplishment. Our model illustrates that this approach is the most efficient solution to this crisis that will allow for the safety of the citizens of Honeycomb.

Problem B Summary: Evanston Township High School

Advisor: Mark Vondracek

Team Members: Paul Barnes, David Lenz, Oliver Manheim

While the problem did not give any specific goal for decreasing the incidence of crime in the city, we determined that an effective and practical goal for the city is to reduce violent crime by 10% by the year 2020. In order to reduce the amount of crime in the city, we first created a model that calculated incidence of violence as a function of different data that we were given. We determined that some of the provided data had little to no impact on the amount of violent crime. Trying to incorporate this unhelpful information into our model would make it less accurate, so we developed a way to decide if a set of data was worth considering. When we applied this method to all of the sets of data, we were left with five that had a noticeable correlation with violence. These five values were City Population, High School Graduation Rate, High School Dropout Rate, Prison Population, and Percent of Parole Violations. With these five variables determined, we checked to make sure that each relationship with violence was roughly linear, then ran a multivariate linear regression to create a function that outputs an approximate incidence of violence given the City Population, High School Graduation Rate, High School Dropout Rate, Prison Population, and Percent of Parole Violations. When we tested this model against data that we already had, there was relatively low error, with discrepancies in incidence of violence ranging from 15 crimes to less than 1 crime.

Once we could predict an approximate amount of violent crime for any set of data, we found which variable(s) affected the function most. We found that High School Graduation and Dropout Rates had the biggest effect, which supports the comment that the city had problems with gang violence (we assumed high school dropouts were much more likely to join gangs). Because these variables held the most weight in our model, we set out to create a relation which could model increases in graduation rate and decreases in dropout rate as a function of the percent increase in school funding. We felt that this model would be beneficial in our proposal to the mayor, as it allows us to state specific and concrete changes that the city can make, instead of trying to generally increase the number of graduates and stop students from dropping out. There was error in this model as well, which compounded the error of our original model. However, we felt that it allowed us to make a stronger and more realistic proposal to the mayor.

As part of our report, we created a computer program that could accept proposed changes in funding, convert it to an approximate change in graduation and dropout rates, produce the multivariable linear regression function, and input the modified data (modified by the change in school funding) to output a projected incidence of violence in a given year, and the amount eliminated. By using this program, we were able to test different values for the proposed change in funding and assess how well it reduced crime. From this testing, we were confident that by increasing school funding by 275% by the year 2020, the crime would decrease by 10%. However, in the creation of our models we consistently used conservative estimates, so we believe that the true percent increase of funding necessary to reduce the incidence of violence by 10% may be less than that. In addition, our proposal includes qualitative measures for the city to introduce to continue to reduce crime. These measures could not be included in our computer program, so the percent increase in funding may be even less.

Problem B Summary: Mills Godwin High School

Advisor: Todd Phillips

Team Members: Alison Haulsee, Ashley Hedberg, Thomas Hunold, Andreas Keller

Crime is an inevitable part of human society. In order to maintain public safety, there have been many attempts at reducing crime rate throughout history through the implementation of policies on both the national and local level. However, Felonia, a small city with an alarming crime rate that has been rising for several years, has no effective crime reduction policies in place. Therefore, it is necessary to determine a method of reducing the dangerous level of crime in Felonia.

Four models were constructed that analyzed the given Felonian public data. Model 1 used multiple linear regression and only the given data to create a representation of the crime rate in Felonia. Models 2 and 3 used the "broken windows" policy and the "three strikes" law, which have been implemented in other regions of the country, respectively to display their potential positive effects on Felonian violent crime rate. Model 4 combined these two policies to show their joint effect on the reduction of violence in Felonia. A method was mathematically modeled to successfully decrease the Felonian crime rate and violence.

Problem B Summary: Hong Kong International School

Advisor: Edgar Fong

Team Members: Enoch Chan, Nicholas Chan, Victoria Chen, Sasha Huynh

We were set the task of curbing the recent increase in incidences of violence in a regional city given statistics on a variety of indicators. Our analysis of the data showed that there was a strong correlation between increased high school graduation rates and decreased violent crime rates, as well as increased parole violation rates and increased violent crime rates. Given that the unemployment rate was found to have no correlation with violent crime rates, we deduced that the recent rise in violent crime was perpetrated by youth. Our model is designed to provide three components of decreasing high school dropout rates as well as one component of decreasing parole violation rates. The first component to decreasing high school dropout rates involves the implementation of school-to-work programs in underprivileged high school. The second component involves implementing youth mentoring programs for all high school students. The third component involves developing and implementing quality after school programs for all high school students. The component to decreasing parole violation rates involves improving parole and rehabilitation programs.

Problem B Summary: Hong Kong International School

Advisor: Kevin Mansell

Team Members: Brendan Hung, Kristofer Siy, Jessica Tan, Daniel Zhu

In this problem, we were asked to assist a city in reducing its violent crime rate by developing a mathematical model given several statistics about the city. We were asked to write a letter to the mayor of this city addressing a solution to this problem.

The assumptions we considered were that unemployment is cyclical in the short run, that the hypothetical city is located in the USA, that all violent crimes are reported, that the government cannot intervene with families, and that property crimes are disregarded. In addition, the majority of our solution was based on a sixth assumption that violent crime in the city was caused solely by the factors for which we were given statistics: population, unemployment, high school enrollment, dropout rate, graduation rate, juvenile delinquent population, prison population, prison population released on parole, parole violators, and parole violation rate. Out analysis of the problem produced two models, which were used in conjunction to form a solution to the problem of violent crime.

Our first model involved using multiple regressions. We first divided our variables into four categories labeled population, unemployment, education and incarceration, and then selected the one variable from each category that had the highest correlation in a linear regression, to ensure accurate and independent data. Unemployment was disregarded because there proved to be no correlation when a least-squares line was put to the graph. We then used a semi-elastic model to portray the incidence of violence as a composition of functions of county population, graduation rate and parole violation. To reduce this to the incidence of violence as a function of time, we used cubic and quartic regression functions as determined by our calculators to form a final multiple regression equation. We used the same technique to form a fairly accurate model of homicides.

Our second model involved deriving data from historical records of the United States and comparing it to the data provided. We did this through converting some of the variables to their socioeconomic counterparts, and analyzed the situation through lenses of law enforcement, education, and socioeconomic status. Within each lens, we looked at historical cases in California, Texas, Illinois and New York as we felt these four states represented the US as a whole quite well. For the law enforcement aspect, we looked at the New York Police Department and its efforts in using CompStat to reduce crime rates. For the education aspect, we looked at juvenile delinquency over time and various states' efforts to resolve the issue. Finally, for the socioeconomic status aspect, we looked at the recession of 1981-1982 and the subsequent curbing of crime rates.

With both models in mind, we came to several conclusions that we feel will benefit the city in reducing its violent crime rate. Firstly, the only way to ensure a reduction in crime rates is to increase the size of the police force. Secondly, people who are jobless or have zero income are the most likely to commit crime. Finally, a lack of education has a direct impact on violent criminals. From this, our recommended course of action is threefold: To increase the size of the police force, to limit unemployment by creating jobs, and to reform the education system to better target potential delinquents.

Problem B Paper: Glenbrook North High School

Advisor: Brad Benson

Team Members: Sam Aber, Robert Krebs, Harrison Marick, Charlie Tokowitz

Introduction

The mayor of a local city wants to eradicate gang-related violence and violent crimes in general. After analyzing the given data, we came to the conclusion that there are several factors that contribute to the level of violence, most prevalently the unemployment rate and high school dropout rate. By working to lower these rates, we expect a decrease in incidents of violence. However, before conclusions can be drawn regarding how to curb these trends, we must first examine the extent of these issues and more specifically their impact on the occurrence of incidents of violence.

Variables

- Before doing our analysis and condensing our data to the most significant variables, we assumed every variable to be relevant. We considered the following variables.

- Incidents of violence: The number of violent crimes in one year. It is composed of the number of homicides and the number of assaults. In our model, this figure is contingent on the graduation rate and the unemployment rate; it is a dependent variable.

- Rate of incidents of violence: The number of incidents of violence per 100,000 people. We derived this from the given statistics.

- Homicides: The number of homicides in a given year. We did not include this in our model.

- Rate of homicides: The number of homicides per 100,000 people. We did not include this in our model.

- Assaults: The number of assaults in a given year. We did not include this in our model.

- Rate of assaults: The number of assaults per 100,000 people. We did not include this in our model.

- City population: The number of people living in the city. Although this variable was not directly included in our model, the two independent variables are ratios that depend on it.

- County population: The total population of the county in which the city is located. We did not include this in our model.

- Unemployment: The number of unemployed people seeking a job. This variable is not directly included in our model.

- Unemployment rate: A ratio of the number of unemployed to the city population. In our model, this is an independent variable.

- High school enrollment: The number of students in the city who are enrolled in high school in a given year. We did not include this in our model.

- High school dropouts: The number of students who legally drop out of high school. We did not include this in our model.

- High school graduation rate: The percentage of enrolled high school students that graduate from high school. This is an independent variable in our model.

- High school dropout rate: The percentage of enrolled students who are assumed to have dropped out of high school. This is not included in our model. Note that dropout rate and graduation rate do not add to 100% because some students drop out before the legal age or without documentation. Also, some students might move to another city.

- Juvenile inmates: Juvenile inmates are under the age of 18 and are kept at juvenile detention facilities, as opposed to standard prisons. We did not include this in our model.

- Prison population: The total number of inmates including juvenile inmates. We did not include this in our model.

- Released on parole: The number of inmates including juvenile inmates who are released on parole in a given year. We did not include this in our model.

- Parole violation: The number of parolees who violate terms of parole. It includes juvenile parolees. We did not include this in our model.

- Rate of parole violation: A ratio of the number of parole violators to the total number of parolees.

Assumptions

1. The city does not have any of our programs already in place. We assume this because we have nothing to indicate otherwise. Without this assumption, we would have no way of making predictions because there would be no baseline.

2. Our city is comparable to a city of the same size (e.g., Dayton, Ohio—see Case Study). We assume this because of the comparable populations, and this is the only information we were given that could relate it to another city.

3. Any factors other than the given variables are inconsequential. Without this assumption, all variables could be affected by one another and other variables that were not accounted for. This would make it difficult to perform an analysis.

4. It was unclear to us whether some of the given data were representative of the entire county or only the city. We determined that, based on comparisons to similarly sized cities, the data was from the city and not the county.

5. Dropouts recorded in the dropout rate are only those who have legally dropped out. We assumed this in order to explain why the dropout rate and graduation rate did not total 100%.

6. The number of incidents of violence is the sum of the numbers of homicides and assaults. We assume this because the sum of the homicides and assaults in a given year is equal to the number of violent crimes in that year.

7. Only those who are actively searching for a job are considered unemployed. We assume this because this is the definition of unemployment.

8. All independent variables (graduation rate, unemployment rate, etc.) are independent of crime, which is the dependent variable. We assume this because we need a way to quantify the dependent variable.

Model and Analysis

We determined that the best method for deriving our model would be to use linear regression analysis, which takes a set of known independent variables and their values and finds the variables that have the most predictive capacity in terms of a single, dependent variable (in our case, the number of violent crimes in a year). We wanted to explain our dependent variable (violent crime rate) using the other, independent variables. Regression assigns coefficient values to all of the qualifying independent variables, and puts them on one side of the equation along with a constant. Equations with fewer variables are considered stronger; more variables with insignificant levels of statistical relevance to the dependent variable only weaken the model, as more variables mean a greater likelihood for outlying or irrelevant data points that could skew the prediction. Because performing a regression with a large number of variables can be complex, we used the statistics program Statistics Package for the Social Sciences (SPSS). Not only did this save time, it also provided us with a number of different means for measuring the effectiveness and accuracy of our model.

We used the same software to calculate Pearson correlations between every possible pair of variables. These gave us a basis for comparing the relative strengths of the variables in terms of their correlation with the dependent variable. Moreover, these correlations shed light on some interesting phenomena in our data. The number of violent crimes could be broken into two subcategories: homicides and assaults. Specifically, we saw significant correlations between high school graduation rate and number of assaults at -0.742, as well as between unemployment rate and number of homicides at 0.524. However, neither of these had a large effect on the other dependent variable (graduation rate did very little for murder rate just as unemployment was all but inconsequential to assault rate). We also found the data regarding prisons (juvenile inmates, prison population, number of parolees, percentage of parole violators, and parole violators) to be largely irrelevant to our independent variables, as the correlations were weak.

Based on the correlations, we determined that it did not make sense to simply throw every variable into the regression. Instead, we chose the variables that had the strongest correlations with our dependent variables: graduation rate and unemployment rate. These variables also stood out as ones that would be immune to fluctuations in the city's population. There were other variables that had notable levels of correlation, notably the dropout rate, though we felt that it was too closely related to the graduation rate and therefore would not be useful. Regardless, we tested a number of different combinations of variables and confirmed our hunches—the prison-related statistics were useless, as was the implementation of the redundant pairs of variables. In the end, the simplest solution prevailed, and our model was:

violent crimes per 100,000 people = 1216.159 + 6.979(unemployment rate) –915.862(graduation rate)

Our model's usefulness was substantiated by various statistical values that measure the strength of such models and were calculated by SPSS. In particular, our R-value, which indicates collective correlation between independent and dependent variables, was fairly high at 0.751. Our R2 value, a measure of total variance in the data explained by our equation, was a solid 0.563. The model is statistically significant to a level of 0.083, which tells us that there is at most an 8.3% chance that the null hypothesis would be true and that our model would be no more useful than simply predicting the mean of the data for every year in the future. Because we were able to economize on the number of variables used as predictors, we can be sure that there is quite a small chance of our equation being compromised by outlying or extraneous data. However, this chance is certainly smaller than if we had used many variables that made only marginal contributions. We were also able to avoid falling victim to fluctuations in the city's population by normalizing our crime data and calculating the number of violent crimes, assaults, and homicides per 100,000 people using this equation:

$$\frac{\textbf{Given number of violent crimes}}{\textbf{City population for that year}} = \frac{\textbf{Adjusted number of violent crimes}}{\textbf{100,000 people}}$$

The graph in Figure 1 represents both the actual values, normalized per 100,000 people, and the values we derived from our model.

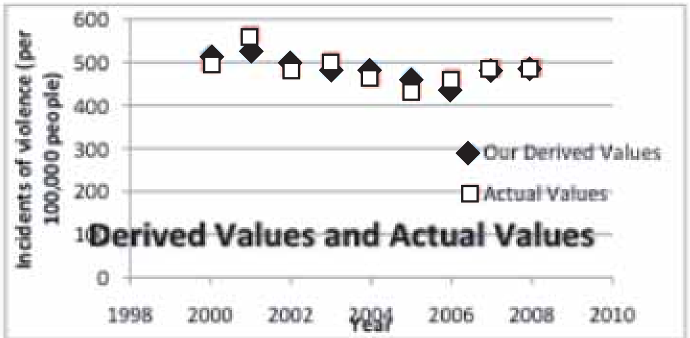


Figure 1

In terms of our model's weaknesses, we had an insubstantial amount of data. A single eight-year block of time could harbor economic microcosms that could influence unemployment or funding of schools. We were also forced to make a number of assumptions in order to achieve data that would work with our regression analysis. Among other things, we assumed this city to be representative of others of similar size. The fundamental weakness of our model is that its usefulness as a predictor is limited by our own ability to foresee the future effectiveness of the programs that we intend to put into place. That is to say, one can't use the future to predict the future.

Inability to project into the future notwithstanding, our model is still useful. It provides us with an idea of what kind of issues need to be addressed, as well as being useful as a means of setting goals. If the desired crime rate is plugged into the equation, one can solve for the other two variables and determine what benchmarks need to be reached to achieve the desired result. For example:

violent crimes per 100,000 people = 1216.159 + 6.979(unemployment rate) – 915.862(graduation rate)

300 = 1216.159 + 6.979(unemployment rate) – 915.862(graduation rate)

-916.159 = 6.979(unemployment rate) – 915.862(graduation rate)

0 = 916.159 + 6.979(unemployment rate) – 915.862(graduation rate)

This is a standard Cartesian equation of a line. There are infinitely many solutions, but we can pick one variable and solve for the other, thus allowing us to set goals for achieving a lower number of violent crimes

High school dropouts plague communities nationwide, and our city is no exception. We found correlations between both high school graduation rate and dropout rate, those being -0.721 and 0.644, respectively. Correlation between graduation rate and incidence of violence is negative, indicating that a higher graduation rate yields less violent crime, and, conversely, the correlation with dropout rate is positive, showing that more dropouts tend to result in more violent crime.

We believe that when teens drop out of school, there is greater opportunity for them to become involved with gangs and illicit activities, which in turn can lead to violence. According to Ernest Chavez of Colorado State University, "20 percent [of dropouts] had cut someone with a knife." A study by Laurie Drapela of Washington University (2006) also finds that high school dropouts have "higher arrest rates, greater involvement in violence, and higher usage of drugs." But why do teenagers drop out of school? "Youth crime, gang involvement and violence are additional reasons students drop out," according to Kathryn Barr, an eHow News correspondent (2010). However, many teens drop out of school to help support their families. It can be hard for one or even two parents to earn a sufficient income to support a family, and many former students have cited this as their reason for dropping out of high school (CEC.org 2010).

Based on our research, we came up with several suggestions to curb the dropout rate. Youth centers provide children with a safe environment and a pathway to a productive future, according to the White Light Group, an inner-city youth center foundation. Instituting more youth centers will instill in children a desire to remain in school. Another example of youth services providing outreach and opportunities to children is the Boys and Girls Clubs of America (BGCA), which have done studies showing that violence peaks between 3:00 pm and 6:00 pm. Thus, by providing after-school activities we can reduce the number of violent crimes committed during this block of time. A 2009 study conducted by Amy Arbretren of Public/Private Ventures confirms that high levels of attendance in BGCA meetings leads to higher attendance in school, increased academic success, and decreased numbers of negative peers as friends. A Columbia University study found that those who attended BGCA meetings compared with their peers had grade point averages 11% higher and missed 87% less school. In addition, BGCA create a sense of community, which can be a deterrent to the territoriality of gang culture.

If children require extrinsic motivation, programs such as Chicago's Cash for Grades can provide a tangible reason to remain in school. The Perfect Attendance Incentive Program, also enacted by Chicago Public Schools (CPS), rewards students with laptops, gift cards, and even family vacations. CPS saw a quadrupling of the number of students who did not miss a single day of school.

Our plan for students who drop out of school to support families is to set up career-specific vocational schools. Juliet Miller and Susan Imel of the U.S. Department of Education state: "students with low motivation to attend school have shown improvement in school attendance and retention after participating in career education." This makes sense because students who choose to attend career and vocational schools are more motivated to graduate. Students also may be more inclined to stay in vocational schools because the information they learn is pertinent to their career path.

Another way to cut incidents of violence is to reduce unemployment. Our statistics have shown us that there is a correlation between unemployment and incidence of violence, though in a different and intriguing way. While there was not a strong correlation between unemployment rate and number of assaults, which composed the majority of instances of violent crime, there was a strong correlation between unemployment and the number of murders. This correlation was 0.524, one of the strongest that we saw. But why does this link exist? A study conducted at Hebrew University in Jerusalem by Eric Gould suggests that unemployment causes people to turn to violence. While wages fell and the economy worsened between 1979 and 1997, violent crime went up 35%. This led us to believe that unemployment can devastate a person financially and psychologically, which can lead to violent crime. If we can reduce unemployment, we will reduce violent crime.

Another tried and true method to lower unemployment rates is to create public works jobs. One of the best historical examples of this is the Works Progress Administration (WPA) enacted by President Franklin D. Roosevelt in 1935. Between 1935 and 1943, the WPA created nearly eight million jobs. Public works projects could include building the vocational schools we discussed earlier.

To verify our model's effectiveness, we searched for a city of comparable size and with comparable statistics. Dayton, Ohio, with a population of about 154,000, is slightly larger than our city. After researching Dayton, we concluded that this city offers many programs similar to the ones we proposed. First, Dayton has an extensive BGCA program with three chapters at various locations. These clubs reach out to youth by providing them with a safe place to interact with each other. In addition, these clubs help build strong community ties. We are proposing implementation of a similar system in our city in an attempt to prevent children from joining gangs, which will reduce violent crime.

Dayton has an unemployment rate of 10.2%, which is similar to our city's average rate of 11. 3 percent; this average is taken from the rates between 2000 and 2008. Dayton has adopted numerous measures to try to reduce unemployment. One solution was launching a large public works program to build roads and sewage systems, maintain streets, collect waste, and other projects. This generates jobs for many of the unemployed. Dayton's unemployment rate should not result in more crime incidents than our city's; if anything, the low unemployment rate should correlate to a decrease in violent crime.

We hypothesize that the clubs will help decrease crime. We found that Dayton's average number of assaults per year between 2003 and 2008 was about 657 while our city averaged about 708 assaults per year. Despite the lower average number of assaults, Dayton averages more homicides than our city does. Dayton averaged about 34 homicides per year between 2003 and 2008, but our city only averaged 16 homicides per year between 2000 and 2008. Dayton's higher rate is due to the gang presence in the city. This gang culture has trickled down to the youth of Dayton, and this is a way in which the Boys and Girls Clubs have not been implemented to their full extent.

Our research of Dayton tells us to what degree we must implement each program. The lack of success of programs such as the Boys and Girls Clubs proves that it is imperative that all children have a safe place to go after school. Moreover, Dayton's public works system works well in terms of employing relatively uneducated people. Therefore, our city will mirror Dayton's program to give people a steady source of income and a reason to not resort to violence. Through implementation of these programs we can expect a reasonable decrease in the amount of crime.

Conclusion

Despite the fact that we can figure out the effects of graduation and unemployment rates on the amount of crime, we cannot predict future crime statistics. This is because we are not aware of future unemployment and graduation rates. Both rates are pertinent to our model, and not knowing them hinders our ability to predict. We can predict future unemployment and graduation rates, yet those will only be estimates based on research and data trends. Thus, our model is unable to predict very accurately. However, the most value the model has is in helping to reduce future violent crimes as it will provide rough sketches of future data and create goals, along with providing plans of action that will decrease the rate of violent crime.