Dear $E=mc^2$ readers,

My name is Corey Wald, from North Bellmore, New York. I was selected as a semifinalist in the 2013 Intel Science Talent Search my senior year at Wellington C. Mepham High School. The title of my paper was The Sky's the Limit- An Investigation of Cloud Cover on Major League Baseball Performance. My research project was inspired by a genuine passion for the game of baseball and my desire to learn more about its subtle nuances. I often wondered how much weather variables such as sun, clouds and shadows affected the outcome of a game or individual player performance. My curiosity prompted me to do some preliminary research to identify whether these questions were previously investigated. This served as the impetus for my project and interestingly the results only led me to formulate more questions.

Dr. David Kommor, a science research and AP Biology teacher at Wellington C. Mepham High School along with my personal interest in baseball sparked my interest in researching this topic. During my literature search, I found a study by Wes Kent investigating weather variables and baseball player performance. Kent's study captivated my interest and I emailed him to obtain additional information about his study. This was the beginning of my three year research journey that culminated in the paper that I submitted to the Intel Science Talent Search. Dr. Kommor has been the most influential person in the development of my scientific career. He has been my project mentor for three years, but his hard work and dedication reached well beyond the school day. Dr. Kommor was instrumental in helping me complete this project and understanding the importance of having a strong work ethic. I can still hear his voice in the back of my mind saying, "go big or go home". More importantly, Dr. Kommor helped me understand the value of using my mistakes and setbacks as learning opportunities and maintaining a positive attitude.

The majority of my research was completed at home. The remaining aspects of this study were completed at Wellington C. Mepham High School 2401 Camp Avenue Bellmore NY 11710. Dr. David Kommor, my advisor/teacher who works at Wellington C. Mepham High School supervised my research and assisted in experimental design and statistical data interpretation. Bernard Gorman, Ph.D., at Hofstra University assisted in statistical data analysis. Wes Kent, graduate student at Kent State University also assisted in experimental design, variable creation and obtaining data.

At first glance, one might think that the statistical analysis for this project was simplistic, but that could not be further from the truth. As a result, I had to teach myself the mathematical constructs to accurately analyze and interpret the data. The data collection for this study was absolutely massive as data from over 3,000 baseball games was collected by hand and put into Excel spreadsheets. The statistical analysis was when the "fun" really started. The only statistical test that I really understood before this research project was making a graph from an equation of a line and the chi-square test. The rest was all new to me. Dr. Bernard Gorman introduced me to the SPSS, Version 16 procedure GENLIN, a series of Generalized Estimation Equations (GEE), to test my hypotheses. My project also utilized Cramer's V, Estimated Marginal Means and Pairwise Comparison of Estimated Marginal Means, which I was introduced to and learned as a result of this research project.

Interestingly, my project definitely made the melding of science and mathematics more of a reality. As a result, I've come to realize that scientific investigations and mathematics are like solving puzzles; another one of my passions. You open the box and there are 1000's of seemingly unrelated pieces that will eventually fit together and through a systematic approach, the solution is at hand. However, science and mathematics is more complicated than a boxed

puzzle. There are far more pieces and the solution may not be simply finding the corners and the straight edge pieces. This explains my passion for math and science. Often times, the solutions to problems lead to new questions and the puzzle is never truly complete. Just when you think you know it all, there is always someone discovering something different or a student asking a question that leaves the teacher wondering the answer of the question. I've always found this very exciting and I'm more certain than ever that I will pursue a career as a biomedical engineer. Spending my life attempting to unravel mysteries will be a challenge that will make getting up in the morning so much easier. My research teacher has always said, "If you love what you do, you will never "work" a day in your life".

My advice for other high school students who might like to undertake a project combining science and mathematics is simple. Be yourself and let nothing hold you back. Research something that genuinely interests you and the process becomes an extension of your intellectual growth. High school students have an urge to follow in others footsteps, which works in most situations, but not when beginning a research project. Not all research projects initially show results so work through the bleak moments that occur throughout the course of a research project. It makes the final product that much sweeter. Last but not least, as in everything in life, give your research project your all because research can truly make a worldwide difference for the world now and in years to come. Not all research projects create breakthroughs but you never know when your breakthrough is going to come so you must consistently work at it with the utmost effort.

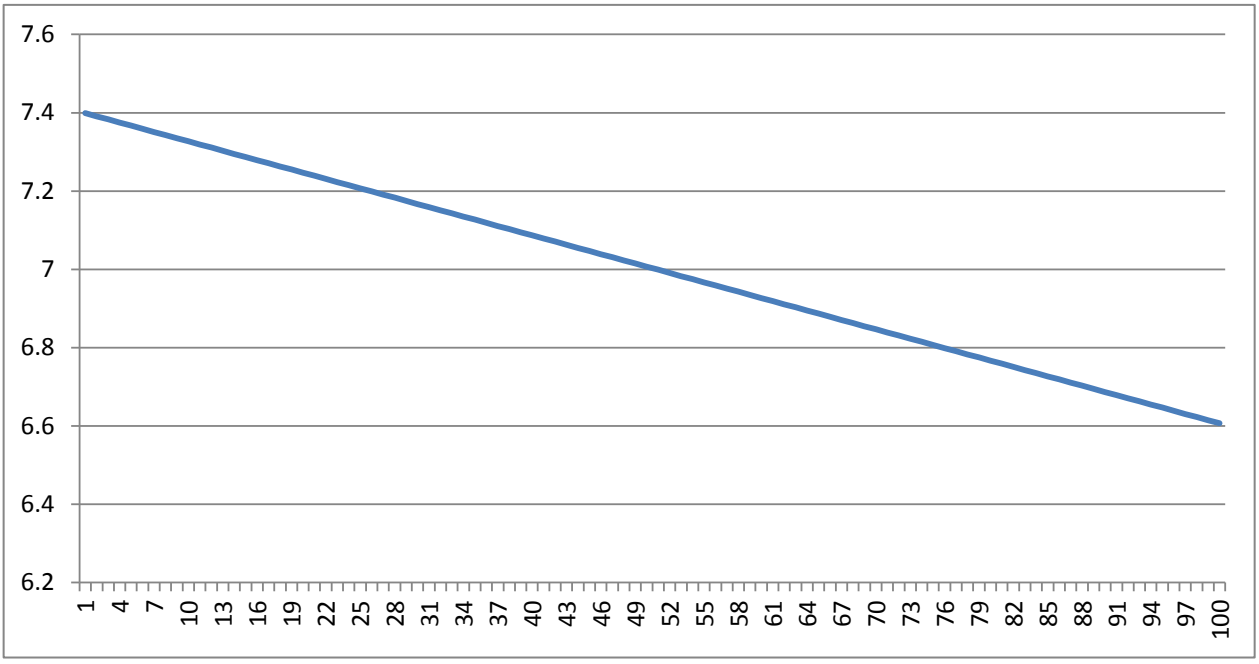
The focus of my project was to investigate whether different percentages of cloud cover during a baseball game favored batters or pitchers. Additionally, what effect cloud cover percentages had on fielders during a Major League Baseball game? The data collected for this

investigation was obtained from the years 2007 to 2010. Baseball data gathered during day conditions were also compared to baseball data gathered during night conditions to see if any significant differences existed. Baseball data was collected from www.baseball-reference.com and weather data was gathered from the National Climatic Data Center. Seven variables were compared across three categories of cloud cover, 0-29%, 30-79% and 80-100% cloud cover during day games and night games, which served as the control for my study. The overall results suggest that clearer conditions tend to favor the pitcher in each day game while cloudier day games tend to favor the batter.

The key components to my project are the massive amounts of data collected, the mathematics behind my conclusions and how my research has only left more questions for future researchers than answered. Over 3,000 baseball games were hand entered into an Excel spreadsheet over the course of the first two years of my study. The third year of my study consisted of statistical analyses.

A series of Generalized Estimation Equations (GEE) regression analyses, employing the SPSS, Version 16 procedure, GENLIN, were employed to test the hypotheses. In Ordinary Least Squares regression (OLS), it is assumed that each observation unit is independently observed. For the present study, a team's performance on a given day constituted the observation unit. However, as the same teams repeatedly played over seasons and within the same ball parks, the assumption of independence could not be assumed. GEE overcomes this problem by correcting for correlated, non-independent events All of the outcome variables: strikeouts, earned run average, home runs allowed, batting average, on base plus slugging percentage, walks and errors were examined, but only strikeouts turned out to be statistically significant so that will be the

only variable I will go in detail with. All other variables went through the same data tests. GEE produced the following regression analysis in which



This represents a continuous test for all percentages of cloud cover, with 7.407 being the y-intercept and constant while -.008 represents the slope of the equation. The equation tells us that as the cloud cover increases by 1%, the number of strikeouts drops by .008. These results were statistically significant, $\chi 2(1, N=3142) = 6.377, p = .012$. This finding suggests that increased cloud cover tends to favor the batter. The strength of association between the two categorical variables was measured using Cramer's V. This test represents a measure of association between two nominal variables, giving a value between 0 and +1 (inclusive). It is

based on Pearson's chi-squared statistic. The effect size for this finding, Cramer's V, was moderate, .47 value (Cohen, 1988).

The next test that the data was subjected to was an Estimated Marginal Means which is necessary to compare the means of unequal sample sizes. As a result, one can obtain a mean response for each categorical variable, while adjusting for other variables. In this study, the cloud cover variable was subdivided into 4 additional categories; 0-29% cloud cover, 30-79% cloud cover, 80-100% cloud cover and night. This approach was implemented to identify statistically significant differences in the number of strikeouts dependent on percent cloud cover.

Table 1. Cloud Cover Percentage and Mean Strikeouts

Cloud Cover Conditions	Mean (M)
0-29% cloud cover	7.51
30-79% cloud cover	6.49
80-100% cloud cover	6.46
Night	6.48

Table 1 indicates that the mean number of strikeouts is greatest for games played with 0-29% cloud cover (M=7.51). As cloud cover increases the trend moves towards a decrease in mean strikeouts, indicating that low percent cloud covers tends to favors the pitcher.

Table 2. Pairwise comparison of estimated marginal means based on strikeouts

0-29% cloud cover (I)	Other cloud conditions (J)	Mean difference (I-J)	Df	Sig
	30-79% cloud cover	1.02	1	.000
	80-100% cloud cover	1.05	1	.000
	Night (control)	1.03	1	.000

Using pairwise comparisons, one can see that the 30-79% cloud cover, 80-100% cloud cover and the night category were all significant with games played under 0-29% cloud cover as p=0. A chi-square test was performed to examine the relationship amongst cloud cover variables and strikeouts, $\chi 2(3, N=3142) = 33.743$, p=0, Cramer's V= .955.

Shadows, while not investigated in this study, can also have a major impact on the performance of a pitcher. Future research could focus on which stadiums frequently have shadows in between the pitcher's mound and home plate to identify whether those stadiums have better pitching performances, as measured by the highest number of strikeouts and lowest number of home runs allowed and ERA. Additional studies could investigate whether day conditions place a fielder at a disadvantage due to shadows and these shadows effect offensive and defensive performance. These studies can also incorporate the home field advantage variable. Future research might yield more significant results if ceilometers, an instrument for measuring the percent cloud cover, were installed into each Major League Baseball stadium. Percent cloud cover is an ever changing variable; having a single number representing cloud cover, which was the case in this study, may not be appropriate for all games. The ceilometers would allow researchers to measure changing cloud cover conditions in "real-time" throughout

the game. There are also many different types of clouds that can affect a baseball game. High and thin cirrus clouds, found above 18,000 feet could create a brighter playing environment, potentially benefiting the pitcher, while lower stratus clouds, found below 6,000 feet could result in less sunlight, providing potential advantages to the batter and fielder. Some games can also have a combination of various types of clouds which can impact the environment of the baseball game being played beneath them. One last objective that future research should take into account is trying to control for talent level of each particular baseball team. Not all baseball teams are created equal and therefore, some baseball teams have much more talent on paper than other baseball teams.