**Weight Gain during the Freshman Year of College: A Negative Binomial Regression**

Elaine Perrin, Travis Sellers, Lorena Hernandez Maxwell, Malachi Morgan, Sherod-Malik Davis, Cynthia Cook, Dr. Monica C. Jackson

**Miami University**

**Department of Mathematics and Statistics**

**Oxford, OH, 45056, U.S.A.**

**Maryville College**

**Department of Mathematics**

**Maryville, TN, 37804, U.S.A**

**Rose-Hulman Institute of Technology**

**Department of Mathematics**

**Terre Haute, IN, 47803, U.S.A**

**Morehouse College**

**Department of Mathematics**

**Atlanta, Georgia, 30314, U.S.A**

**American University**

**Department of Mathematics and Statistics**

**Washington, DC, 20016, U.S.A.**

**Abstract**

A common fear for incoming freshman in college is the dreaded “freshman fifteen”. The combination of being in a new environment away from home, a high level of stress, and the consumption of alcohol and dining hall food causes weight gain in college students. This study examines the weight gained during the first year of college and what factors contributed to it. A zero-inflated negative binomial regression analysis was performed to investigate which characteristics of a student's freshman year, along with some characteristics of a student prior to his/her freshman year impact the odds of a student gaining weight. In our study it was found that major, gender, the numbers of hours of sleep one usually had, and the relationship one had with his/her family all helped predicting whether an individual was likely to gain weight or not. Humanities and Arts majors were more likely to gain weight than Business majors. Also our study shows that those who got on average five to eight hours of sleep a night were less likely to gain weight than those who got less than four or more than nine hours of sleep. Our study also showed that females who had a better relationship with their parents had better chances of weight gain.

**I. Introduction**

Significant weight gain is a common problem many college students experience during their freshman year [Levitsky, Halbmaier, and Mrdjenovic]. Many studies have investigated the “Freshman Fifteen” and found this to be an overstatement [Vella-Zarb and Elgar]. A vast majority of the studies showed there was a modest weight gain, around three to seven pounds, in college freshmen during their first semester [Anderson, Shapiro, and Lundgren]. Many researchers also found that by the second semester, the freshman’s weight had stabilized [Anderson, Shapiro, and Lundgren]. According to Holm-Denoma, Heatherton, Vohs, and Joiner, a student’s relationship with their parents was a significant factor for weight gain in both males and females. However, it was found that a good relationship with parents led to an increase in weight for females, while a troublesome relationship with parents led to one in males.

Other variables have also been studied to see their connection with weight gain. A few of them were campus setup, a student’s amount of physical activity, and their eating habits. It was found that the layout of a campus had an effect on whether or not students were eating healthier or working out more than others [Kapinos and Yakusheva]. Research also showed that students who lived on-campus gained more weight than those who lived off-campus [Vella-Zarb and Elgar]. Another study reported that those who gained weight were less physically active in college than in high school and more likely to eat breakfast and sleep more than those who did not gain weight [Wengreen and Moncur]. Ironically, Lowe, Annunziato, Tuttman Markowitz, Didie, Bellace, Riddell, Maille, McKinney, and Stice found that traditional methods such as restraint, disinhibition, and emotional eating were not predictive of weight gain, but a history of weight loss, dieting, and weight suppression were.

We are interested in testing a few of the previously mentioned variables, along with other factors that many studies have not examined. We are curious to see if there are any unexpected or overlooked variables that could trigger weight gain. The survey we administered considered the following variables: current age and freshman age, college’s distance from hometown, major, race, eating habits, participation in extracurricular activities, gender, alcohol intake, if they smoked, how often they studied, their grade point average (GPA), their stress level, the amount of exercise they did, whether or not they were a vegetarian, if they had previously lived away from home, if they ate breakfast, if they had a job during their freshman year, their relationship with their parents, the amount of financial support they received from their parents, their parents’ marital status, amount of time spent playing video games, how healthy they viewed their lifestyle, average amount of sleep they received, and whether or not they consumed caffeine. Many of these variables are influenced by the change in environment students experience during their freshman year and our goal is to see if any of these factors have an impact on weight gain.

1. **Data**

We obtained our data by performing a twenty-seven question survey[Appendix]. The survey was created in Qualtrics Survey Software provided by Miami University.

The questions on the survey provided us with one response variable, weight gain in pounds, and twenty-four possible explanatory variables. The quantitative explanatory variables were stress level, GPA, freshman age, and relationship with parent(s). Current age was also a quantitative value, however wasn't treated as a variable; it was collected to see the age difference between freshman age and present age. The remainder variables were treated as categorical variables, they were, distance of school from hometown, major, race, eating habits, number of extracurricular activities, gender, alcohol consumption, smoker or not, hours spent sleeping per day, hours spent watching TV or playing video games per week, hours spend studying per week, days a week spent exercising, vegetarian or not, whether a student lived away from home during high school, ate breakfast regularly, employment status, marital status of parents, financial support from parent(s), caffeine increase or not, and how well health was promoted by family. The participants gave text responses for the qualitative variables (six questions), as well as the schools they attended. The remainder of the questions gave the participants options for a response, hence, the remaining variables were categorical.

In order to perform a Poisson regression, the response values must be integers, so we decided to round up, using the ceiling function in R, and replaced any decimal numbers with integers. Also the weight gain responses are all estimates made by the participants, and typically weight estimations are whole numbers.

Since majority of our variables were categorical with multiple levels, this resulted in over-parametrization. To help fix this problem we decided to shorten the list of majors one could fall under. We combined the number of arts students and the number of humanities students together, as well as the number of science students engineering students together, and formed a categorical variable with five response options instead of seven. We combined arts with humanities, and sciences and engineering, because we believe they were the most similar majors. Also we collapsed the race into five races instead of the six that were options. We originally gave Indian as an option and Asian as an option despite the fact Indian is a sub-race of Asian. So it was logical to combine Indian and Asian into one response option.

1. **Exploratory Analysis**

The distribution of the response variable was right skewed and fifty-two percent of the response values were zero, therefore we used zero inflated negative binomial regression. The minimum, 25% quartile, median, 75% quartile, maximum, and mean were found for each of the explanatory variables that were quantitative. For the categorical variables, the percentage of each category (?) occurring was calculated. It is important to note that seventy-two percent of the survey participants were female. Also since the range for freshman age is so small, we will exclude this as a possible explanatory variable because it will appear significant in our regression however it really isn't. Whether a participant was a smoker or was a vegetarian should be excluded since over ninety percent claimed they were not at the time.

1. **Results**
2. **Preliminary Analysis**

To begin our analysis, we observed the mean and the variance of the response variable. After observation, we found our response data to be over-dispersed. We decided to compare a negative binomial regression model with all of the explanatory variables and a zero-inflated negative binomial regression model with all of the explanatory variables since the response data contained an excess amount of zeros. Many of our responses said that they did not experience weight gain. Therefore, we denoted their answer with a zero and we found our response data to have a large amount of zeros. After comparison, we found that the zero-inflated negative binomial regression model was a significantly better model. After choosing the zero-inflated negative binomial model, we decided that we must remove some of the categorical variables from our model since it was over parameterized.

1. **Zero-inflated Negative Binomial Model**

Our final model consisted of the explanatory variables major: education (p-value=.98372, 95% CI -1660.46-1626.24), major: Humanities/Arts (p-value=0.00223, 95% CI -1.91- -.418), major: pre-med (p-value = 0.22153, CI -2.35 – 0.544), major: engineering (p-value = 0.03784, 95% CI -1.61- -0.053982), family relation (p-value = .00757, 95% CI -0.2130172- -0.1833428), sleep: 5 or 6 hours (p-value = 0.08583, 95% CI -0.119 – 1.80), sleep: 7 or 8 hours (p-value = 0.17411, 95% CI -0.304 – 1.68), sleep: >9 hours (p-value = .05179, 95% CI -0.012 – 3.01), and gender (p-value = 0.17500, 95% CI -0.187 – 1.03). The variables that we use for the baseline cases for major and sleep are business and sleep with less than 4 hours. Our final model is given by the following equation:

Weight gain = exp (β0) \* exp (β1\*education) \* exp (β2\*Humanities/Arts) \* exp (β3\*pre-med)\* exp (β4\*Engineering) \* exp (β5 \*family relation) \* exp (β6 \* sleep ( 5 or 6 )) \* exp (β7 \* sleep (7 or 8)) \* exp (β8\*sleep(>9)) \* exp (β9\*gender),

The following table shows that the variables major: humanities and arts, major: engineering, family relation, sleep with 5 or 6 hours, and sleep with more than 9 hours have a significant relationship with the weight gained by a freshman at the .1 significance level:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Estimates | Standard Error | Z values | Pr(>|z|) |
| Intercept | 2.83279 | 0.65193 | 4.345 | 1.39E-005 |
| Major 3 (Education) | -17.10724 | 838.44469 | -0.02 | 0.98372 |
| Major 4 (Humanities/Arts) | -1.16349 | 0.38046 | -3.058 | 0.00223 |
| Major 5 (Pre-med) | -0.90233 | 0.73811 | -1.222 | 0.22153 |
| Major 6 (Engineering) | -0.82975 | 0.39958 | -2.077 | 0.03784 |
| Family relation | -0.19818 | 0.07421 | -2.671 | 0.00757 |
| Sleep 1 ( 5 or 6 hrs.) | 0.84197 | 0.49013 | 1.718 | 0.08583 |
| Sleep 2 ( 7 or 8 hrs.) | 0.68729 | 0.50569 | 1.359 | 0.17411 |
| Sleep 3 ( more than 9) | 1.49773 | 0.77010 | 1.945 | 0.05179 |
| Gender | 0.42016 | 0.30978 | 1.356 | 0.17500 |

The next table gives the incident ratio rates for each variable:

|  |  |
| --- | --- |
| Variable | Incident Rate |
| Intercept | 16.99 |
| Major: Education | 3.70E-008 |
| Major: Humanities/Arts | 0.312 |
| Major: Pre-med | 0.4056 |
| Major: Engineering | 0.4362 |
| Family Relation | 0.8202 |
| Sleep (5 or 6 hours) | 2.3209 |
| Sleep (7 or 8 hours) | 1.988 |
| Sleep (more than 9 hours) | 4.472 |
| Gender | 1.5222 |

Furthermore, we found the 5 number summary for the Pearson’s residuals to be:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | 1st Quart | Median | 3rd Quart | Max |
| -0.8240 | -0.7894 | -0.1689 | 0.5839 | 4.8855 |

The next thing we referred to with our model was the theta value, which was 4.9572. The final test that we ran was Vuong Non-Nested Hypothesis, which tested the standard negative binomial model with these variables against this zero-inflated negative binomial model with these variables. This method is used to see if the zero-inflated negative binomial model is significantly better than the standard negative binomial model. After running this test, we find that the zero-inflated negative binomial model was significantly better than the standard binomial model with a p-value of 4.445505E-06.

1. **Conclusion**

The resulting model includes the variable major which was split into 5 factors. The factor business major was used as the baseline case. From the model The incident rate for a freshman with a Humanities or Arts major is .312 times the incident rate for a freshman with a business major. When all other variables in the model are held constant if a freshman was classified as a Humanities or Arts major, then the probability that they will experience weight gain decreases by a factor of .312. The factor Humanities or Arts Major was significant in the variable major. It was also found that The incident rate for a freshman with an Education major is 3.74x10-8 times the incident rate for a freshman with a business major. When all other variables in the model are held constant if a freshman was classified as an Education major, then the probability that they experience weight gain decreases by a factor of 3.74x10-8, but Education was not a significant factor in the variable major. The Pre- Med factor was also not significant in the variable major, but the incident rate associated with this major is .4 times the incident rate for a freshman with a business major. When all other variables in the model held constant if a freshman was classified as a Pre- Med major, then the probability that they will experience weight gain decreases by a factor of .4. The other significant major in the model was Engineering and Science. The incident rate associated with this major is .44 times the incident rate for a freshman with a business major. This implies that when all other variables in the model are held constant if a freshman is classified with a major in Engineering or the Sciences, then the probability that they will experience weight gain decreases by a factor of .44.

The model also included the variable sleep which was also split into factors. The baseline case for this variable was sleeping for less than 4 hours a night. The incident rate for a freshman that slept for 5 or 6 hours a night is 2.31 times the incident rate for freshman sleeping less than 4 hours a night. This infers that when all other variables in the model are held constant, if a freshman slept 5 or 6 hours a night then the probability that they will experience weight gain decreases by a factor of 2.31. the incident rate for a freshman that slept for 7 or 8 hours is 1.97 times the incident rate for a freshman that slept less than 4 hours a night. This implies that when all other variables in the model are held constant the probability that a freshman who sleeps 7 or 8 hours a night will experience weight gain decreases by a factor of 1.97. The incident rate for a freshman who sleeps for more than 9 hours a night is 4.44 times the incident rate for freshman who sleep less than 4 hours. When all variables are held constant the probability that a freshman who sleeps for more than 9 hours a night will experience weight gain decreases by a factor of 4.44. The model showed that students who sleep for 4 or 5 hours and freshman who sleep for more than 9 hours were the only factors that were significant in the variable sleep.

The model also shows that if a college freshman had a positive relationship with their family then they were more likely to lose weight than one who had a negative relationship with their family. The model also showed that the gender variable had a positive relationship with weight gain. This means that Females are more likely to gain weight their freshman year than males, but this variable was not significant in our model.

1. **Discussion**

Like most statistical studies, we had limitations. One would be the number of questions we could put on a survey, and furthermore, the number of variables we could study. In order for our survey to look desirable to participants, we could not make it too long. In other words, there are hundreds of factors that could impact weight gain during one's freshman year, but we were only able to ask participants a chosen number of questions about a chosen number of possible explanatory variables. The survey was distributed through a social media site (Facebook), and e-mail. This choice of distribution may have caused under coverage due to the lack of schools we received responses from. Also majority of participants attended school in Eastern United States. Another limitation was the estimations the participants made about the weight they gained, and the remainder of the questions in which they had to recall, or approximate their answers. As the current age increased it is likely the responses became less accurate. If we were to perform this survey again, we would construct our survey in such a way that there would be less categorical variables as a product. The categorical variables made it easy for participants to be more accurate in their approximations, however throughout the regression a more precise answer may have resulted in a better model. For example, instead of breaking the responses to the number of hours of sleep into categories, we could ask the participant for a text response.

Our study was a continuation on many previous studies with additional possible explanatory variables we thought would impact weight gain. Future studies could look at variables such as the characteristics of the school's surrounding city. For example, the healthiness of the common citizen, the number of fast food restaurants within a given radius, or the number of health food stores within a given radius. It would be interesting to see if these demographics affect the weight gain of a school's freshman class.

**Acknowledgements**

The authors would like to thank the SUMSRI program at Miami University of Ohio for funding their project. We would also like to thank Dr. Kimberly F. Sellers for helping with the analysis.