# The Effect of Commuting on Semester Grade Point Average: A Statistical Analysis System Interpretation


By:

Tulsi Shrivastava

Ashley Shriver


Topics in Applied Statistics

Ms. Dawn Cable

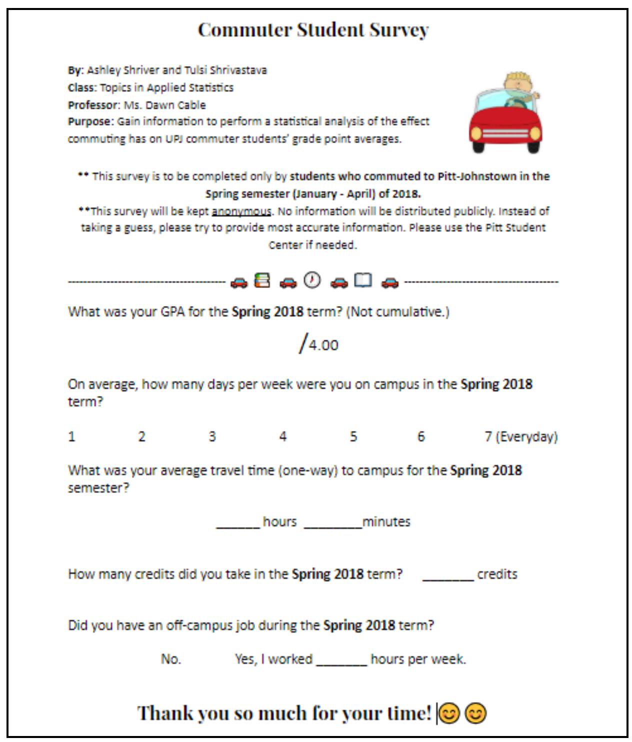
December 14, 2018

## I. Introduction

Grade point average (GPA) is often used as a good indicator of a student's academic standing throughout college. Maintaining a high GPA is a common goal for many undergraduate students, one that requires hard work, consistent dedication, and motivation. It can be especially difficult when a student must factor the additional worries of commute time and job status into their schedule on top of other factors influencing their study. The purpose of this analysis was to gain insight into how a UPJ student's commuting situation affects their grade point average for a semester. Our target audience consisted of students who attend the University of Pittsburgh at Johnstown. Specifically, these students must have commuted to the campus in the spring term of 2017-2018. We restricted it to this set of students because students do not have access to their current term (Fall 2018) GPA. We defined a commuter student to be any individual who must drive to campus because he or she does not have sidewalk access to walk to class. This document details our full study, in which we used the Statistical Analysis System to conduct our analysis.

## II. Data Collection

From November 12th to December 5th, 2018, we administered a survey through both physical and online versions, which allowed us to expand our reach of commuter students. The figure on the right displays the paper version of the survey we administered to students in the UPJ Commuter Lounge, located in the Student Union. We chose this location because many commuter students utilize this room and its amenities throughout the week. Through the online survey software Survey Monkey, we were able to easily share the survey link to members of the commuter student body who may not frequently visit the lounge.


**Commuter Student Survey**

**By:** Ashley Shriver and Tulsi Shrivastava
**Class:** Topics in Applied Statistics
**Professor:** Ms. Dawn Cable
**Purpose:** Gain information to perform a statistical analysis of the effect commuting has on UPJ commuter students' grade point averages.

** This survey is to be completed only by **students who commuted to Pitt-Johnstown in the Spring semester (January - April) of 2018.**
**This survey will be kept anonymous. No information will be distributed publicly. Instead of taking a guess, please try to provide most accurate information. Please use the Pitt Student Center if needed.

What was your GPA for the **Spring 2018** term? (Not cumulative.)

/4.00

On average, how many days per week were you on campus in the **Spring 2018** term?

1 2 3 4 5 6 7 (Everyday)

What was your average travel time (one-way) to campus for the **Spring 2018** semester?

_____ hours _______minutes

How many credits did you take in the **Spring 2018** term? ______ credits

Did you have an off-campus job during the **Spring 2018** term?

No. Yes, I worked ______ hours per week.

**Thank you so much for your time!** 😊😊

The following five questions were included in the survey:

1. What was your GPA for the Spring 2018 term? (Not cumulative.)
2. On average, how many days per week were you on campus in the Spring 2018 term?

3. What was your average travel time (one-way) to campus for the Spring 2018 semester?
4. How many credits did you take in the Spring 2018 term?
5. Did you have an off-campus job during the Spring 2018 term? If so, how many hours did you work per week?

Our aim was to gather at least 30 survey results, 10 times the number of variables we initially planned to include in the model. To predict GPA, we chose variables that may influence commuting situation the most. The variable names and their reasoning of use in our SAS data is detailed below.

1. GPA (dependent variable) = grade point average for the Spring Semester 2017-2018 term only (number of points earned per term based on the letter grades / total number of credits completed that term)
2. DaysWeek (independent variable) = number of days on campus per week
3. TravelTime (independent variable) = time spent driving to campus in minutes
4. Credits (independent variable) = number of credits taken during the Spring Semester 2017-2018 term only
5. OC_Job (independent qualitative variable) = whether or not the student has an off-campus job
6. HoursWeek (independent variable) = the number of hours worked per week for students that did have an off-campus job

At the conclusion of the survey administration, we gained 36 observations. However, due to missing information and possible misentry, we had to omit two of them. This will be discussed later in Section VI. Our full data set is displayed in the image on the next page.

| | GPA | DaysWeek | TravelTime | Credits | OC_Job | HoursWeek |
|---|---|---|---|---|---|---|
| 1 | 3.2 | 5 | 10 | 18 | Y | 40 |
| 2 | 3.56 | 5 | 40 | 12 | Y | 40 |
| 3 | 3.5 | 5 | 15 | 17 | N | 0 |
| 4 | 3.58 | 4 | 20 | 13 | Y | 20 |
| 5 | 3.85 | 3 | 20 | 15 | Y | 32 |
| 6 | 3.6 | 2 | 10 | 18 | Y | 24 |
| 7 | 4 | 5 | 10 | 15 | N | 0 |
| 8 | 2.9 | 5 | 10 | 14 | Y | 55 |
| 9 | 3.518 | 5 | 10 | 14 | Y | 20 |
| 10 | 3.6 | 4 | 7 | 15 | N | 0 |
| 11 | 3.4 | 3 | 20 | 16 | N | 0 |
| 12 | 3.71 | 4 | 9 | 18 | Y | 40 |
| 13 | 4 | 2 | 10 | 21 | Y | 35 |
| 14 | 3.9 | 5 | 25 | 17 | Y | 15 |
| 15 | 4 | 5 | 40 | 15 | Y | 2 |
| 16 | 3.109 | 5 | 20 | 16 | Y | 20 |
| 17 | 3.7 | 5 | 2 | 18 | N | 0 |
| 18 | 3.67 | 5 | 11 | 21 | N | 0 |
| 19 | 2.8 | 4 | 5 | 13 | N | 0 |
| 20 | 3.2 | 5 | 30 | 15 | Y | 10 |
| 21 | 3.25 | 5 | 25 | 12 | Y | 25 |
| 22 | 3 | 7 | 20 | 15 | N | 0 |
| 23 | 2.75 | 3 | 10 | 13 | Y | 20 |
| 24 | 2.5 | 5 | 20 | 4 | Y | 40 |
| 25 | 4 | 5 | 15 | 17 | Y | 14 |
| 26 | 3.766 | 5 | 10 | 17 | Y | 23 |
| 27 | 3.441 | 5 | 20 | 17 | Y | 25 |
| 28 | 4 | 5 | 7 | 15 | Y | 7 |
| 29 | 2 | 5 | 10 | 16 | N | 0 |
| 30 | 3.8 | 4 | 50 | 15 | Y | 20 |
| 31 | 4 | 5 | 65 | 18 | N | 0 |
| 32 | 3 | 5 | 90 | 15 | N | 0 |
| 33 | 3.2 | 5 | 60 | 17 | Y | 16 |
| 34 | 4 | 5 | 30 | 18 | N | 0 |

## III. Data Analysis

### A. Model Creation

In order to determine which variables were the most significant in predicting GPA, we ran the backward elimination model selection method. All the variables initially collected were TravelTime, DaysWeek, HoursWeek, Credits, and OC_Job. Since OC_Job is the only qualitative variable and has 2 levels, we utilized $2 - 1 = 1$ dummy variable, referred to as Job:

$$Job = \{1 \; if \; Yes \; (Y), \; 0 \; if \; not\}$$

Out of the three model selection methods discussed in class, we chose this method in order to keep the largest number of variables possible and to force TravelTime as a variable into our model. The variables that the backwards elimination method deemed useful were TravelTime, DaysWeek, HoursWeek, Job.

| Summary of Backward Elimination | | | | | | | |
|---|---|---|---|---|---|---|---|
| Step | Variable Removed | Number Vars In | Partial R-Square | Model R-Square | C(p) | F Value | Pr > F |
| 1 | TravelTime | 4 | 0.0027 | 0.3172 | 4.1116 | 0.11 | 0.7408 |
| 2 | DaysWeek | 3 | 0.0039 | 0.3133 | 2.2735 | 0.17 | 0.6858 |
| 3 | HoursWeek | 2 | 0.0433 | 0.2699 | 2.0571 | 1.89 | 0.1791 |
| 4 | Job | 1 | 0.0424 | 0.2275 | 1.8026 | 1.80 | 0.1894 |

These variables gave us the following general form for our model:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

After gaining the general form of our model, we ran the regression procedure to obtain the beta estimates and fit the model to our data.

| | Parameter Estimates | | | | | |
|---|---|---|---|---|---|---|
| | Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| | Intercept | 1 | 3.80244 | 0.45432 | 8.37 | <.0001 |
| $x_1$ | TravelTime | 1 | 0.00065282 | 0.00457 | 0.14 | 0.8874 |
| $x_2$ | DaysWeek | 1 | -0.08871 | 0.09028 | -0.98 | 0.3339 |
| $x_3$ | Job | 1 | 0.41402 | 0.27336 | 1.51 | 0.1407 |
| $x_4$ | HoursWeek | 1 | -0.01405 | 0.00853 | -1.65 | 0.1104 |

By plugging in our parameter estimates, we were able to gain the following equation:

$$E(y) = 3.8024 + 0.0007x_1 - 0.0887x_2 + 0.4140x_3 - 0.0140x_4$$

B. Model Analysis

<u>Beta Interpretations</u>

$\beta_0 = 3.8024$(y-intercept) is the GPA when TravelTime =0minutes, DaysWeek =0 days, HoursWeek =0hours, and OC_Job status is N "base" ($\beta_3 = 0$). This is not a meaningful interpretation because there is no commuter student with a travel time of 0 minutes who does not come to campus any day of the week. This point would not exist in the data set.

$\beta_1 = .00065282$ is the point change in the GPA ($y$) for every one minute increase in travel time ($x_1$), holding all other variables constant.

$\beta_2 = -.08871$ is the point change in the GPA ($y$) for every one day increase in days on campus per week ($x_2$), holding all other variables constant.

$\beta_3 = .41402$ is the difference in mean response for the not working off-campus job status N "base" and the mean response for the working off-campus job status Y

$$Job = \{1\ if\ Yes\ (Y),\ 0\ if\ No\ (N)\}$$

Base level = no off-campus job $(N) = 0$

$$\beta_0 = \mu_N$$

$$\beta_3 = \mu_Y - \mu_N$$

$\beta_4 = -.01405$ is the point change in the GPA ($y$) for every one hour increase in hours worked per week ($x_4$) if a student had an off-campus job, holding all other variables constant.

Beta Confidence Intervals

| Parameter Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | 95% Confidence Limits | |
| Intercept | 1 | 3.80244 | 0.45432 | 8.37 | <.0001 | 2.87326 | 4.73162 |
| TravelTime | 1 | 0.00065282 | 0.00457 | 0.14 | 0.8874 | -0.00869 | 0.01000 |
| DaysWeek | 1 | -0.08871 | 0.09028 | -0.98 | 0.3339 | -0.27336 | 0.09593 |
| Job | 1 | 0.41402 | 0.27336 | 1.51 | 0.1407 | -0.14507 | 0.97310 |
| HoursWeek | 1 | -0.01405 | 0.00853 | -1.65 | 0.1104 | -0.03151 | 0.00340 |

These confidence intervals were produced with an alpha level of 0.05 (95% confidence): The interpretation for each variable goes as follows:

1. TravelTime: (-0.00869, 0.01000) With 95% confidence, we can say the mean GPA changes by -0.00869 points to 0.01000 points for each additional minute of travel time.
2. DaysWeek: (-0.27336, 0.09593) With 95% confidence, we can say the mean GPA changes by -0.27336 points to 0.09593 points for each additional day per week spent on campus.

3. Job: (-0.14507, 0.97310) With 95% confidence, we can say the mean GPA difference from students without an off-campus job (N) to students with an off-campus job (Y) is -0.14507 points to 0.97310 points. This may indicated that the model is more likely to predict job status (of 1=Y: students with an off-campus job) as a positive change in GPA.
4. HoursWeek: (-0.03151, 0.00340) With 95% confidence the mean GPA changes by -0.03151 points to 0.00340 points for each hour increase in hours worked per week.

Hypothesis Tests

To test the quality of our model and its accuracy, we conducted the following hypothesis tests.

1. Overall/Global F-test

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 4 | 0.92237 | 0.23059 | 0.95 | 0.4474 |
| Error | 29 | 7.00975 | 0.24172 | | |
| Corrected Total | 33 | 7.93212 | | | |

$$H_0\text{: } \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_A\text{:at least one } \beta \neq 0$$

Since p-value = $.4474$, which is > $\alpha = .05$, we fail to reject $H_0$. There is not sufficient evidence to conclude that this model is statistically useful for predicting the grade point average of a commuter student. The F value is .95, which indicates that an extremely low proportion of variability is accounted for by the model.

2. Individual βTest

We conducted a individual βtest to check whether job status is a useful predictor of GPA. We tested theβfor the $x_3$term, which is $\beta_3$.

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | 1 | 3.80244 | 0.45432 | 8.37 | <.0001 |
| TravelTime | 1 | 0.00065282 | 0.00457 | 0.14 | 0.8874 |
| DaysWeek | 1 | -0.08871 | 0.09028 | -0.98 | 0.3339 |
| Job | 1 | 0.41402 | 0.27336 | 1.51 | 0.1407 |
| HoursWeek | 1 | -0.01405 | 0.00853 | -1.65 | 0.1104 |

$$\beta_3 = \mu_Y - \mu_N = 0$$
$$\mu_Y = \mu_N$$

$$H_0: \beta_3 = 0$$
$$H_A: \beta_3 \neq 0$$

Since p-value = $.1407$, which is > $\alpha = .05$, we fail to reject $H_0$. There is not sufficient evidence to conclude that the job status alone is statistically useful for predicting the grade point average of a commuter student.

3. Comparison of p-values

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | 1 | 3.80244 | 0.45432 | 8.37 | <.0001 |
| TravelTime | 1 | 0.00065282 | 0.00457 | 0.14 | 0.8874 |
| DaysWeek | 1 | -0.08871 | 0.09028 | -0.98 | 0.3339 |
| Job | 1 | 0.41402 | 0.27336 | 1.51 | 0.1407 |
| HoursWeek | 1 | -0.01405 | 0.00853 | -1.65 | 0.1104 |

We did not want to conduct too many individual $\beta$tests because these would not be a good way to determine the overall model effectiveness. However, we wanted to make note of the p-values for each of the variables because higher p-values indicate weak evidence against the null hypothesis. Although each of our variables would fail to reject the null hypothesis, it is worth examining that Job and HoursWeek were slightly more significant predictors than TravelTime and DaysWeek. The greater influence of off-campus job status and hours worked per week will be discussed in further sections as well.

## IV. Inferential Statistics

$\sqrt{MSE} = 0.49165$

| | | | |
|---|---|---|---|
| Root MSE | 0.49165 | R-Square | 0.1163 |
| Dependent Mean | 3.45600 | Adj R-Sq | -0.0056 |
| Coeff Var | 14.22586 | | |

A. S interpretation

1. At least 95% of the observed term GPAs in the data set will fall within $\pm$ 2 standard deviations = 2(0.49165) = 0.9833 points of the predicted GPA using the model.

B. R-Square interpretation

1. 11.63% of the variability in GPA is explained by the model. Since adjusted R-Square is negative the variables in the model may not be the best predictors of GPA when used in combination.

$$R^2 = 0.1163 \quad R_a^2 = -0.0056$$

C. Partial R-Square

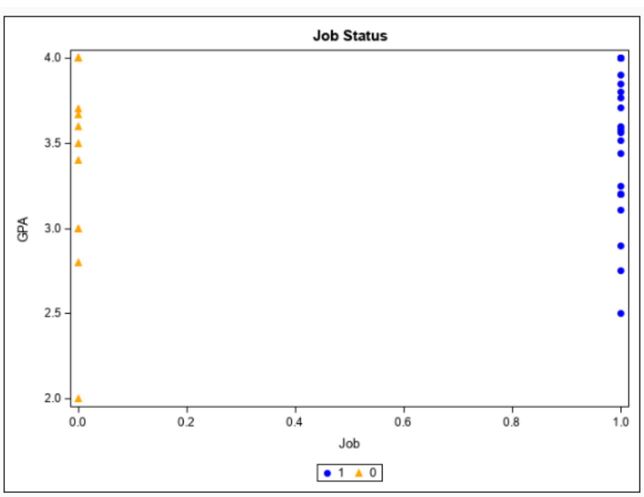
| Summary of Backward Elimination | | | | | | | |
|---|---|---|---|---|---|---|---|
| Step | Variable Removed | Number Vars In | Partial R-Square | Model R-Square | C(p) | F Value | Pr > F |
| 1 | TravelTime | 4 | 0.0027 | 0.3172 | 4.1116 | 0.11 | 0.7408 |
| 2 | DaysWeek | 3 | 0.0039 | 0.3133 | 2.2736 | 0.17 | 0.6858 |
| 3 | HoursWeek | 2 | 0.0433 | 0.2699 | 2.0571 | 1.89 | 0.1791 |
| 4 | Job | 1 | 0.0424 | 0.2275 | 1.8026 | 1.80 | 0.1894 |

Partial R-Square values can be examined to determine which predictors are most useful.

1. TravelTime: $R^2 = 0.0027$. This shows that 0.27% of all the variation in GPA is explained with only travel time.
2. DaysWeeks: $R^2 = 0.0039$. This shows that 0.39% of all the variation in GPA is explained with only the number of weeks a student was on campus.
3. HoursWeek: $R^2 = 0.0433$. This shows that 4.33% of all the variation in GPA is explained with only the number of hours per week a student worked.
4. Job: $R^2 = 0.0424$. This shows that 4.24% of all the variation in GPA is explained with only whether or not a student has an off-campus job.

Overall, off-campus job status and hours worked per week are more significant predictors than the travel time of a student and the number of days per week a student was on campus. Recall that our p-value comparison also yielded the same conclusion.

D. Job Status Scatterplot




To further compare the job status variable, we created a scatterplot which sorted the qualitative variable by status. We also found the mean GPAs of students with a job and without a job.

Overall mean GPA (34 data points): 3.456
Mean GPA for 'N' job (12 yellow data points): 3.389
Mean GPA for 'Y' job (22 blue data points): 3.492

The average GPA for students with an off-campus job is slightly higher than the overall GPA and the GPA for students with no off-campus job.

E. Estimation and Prediction Intervals

In our original data set, we noticed observation 21 was missing a DaysWeek value, so it was omitted. Thus, we chose to utilize this point in our estimation and prediction intervals because the actual GPA was known (3.85). Since this student had 18 credits, we estimated that they may come to campus 5 days a week. We chose to estimate and predict the GPA of a student with a 10-minute travel time who is taking 18 credits, comes to campus 5 days a week, and has no off-campus job.

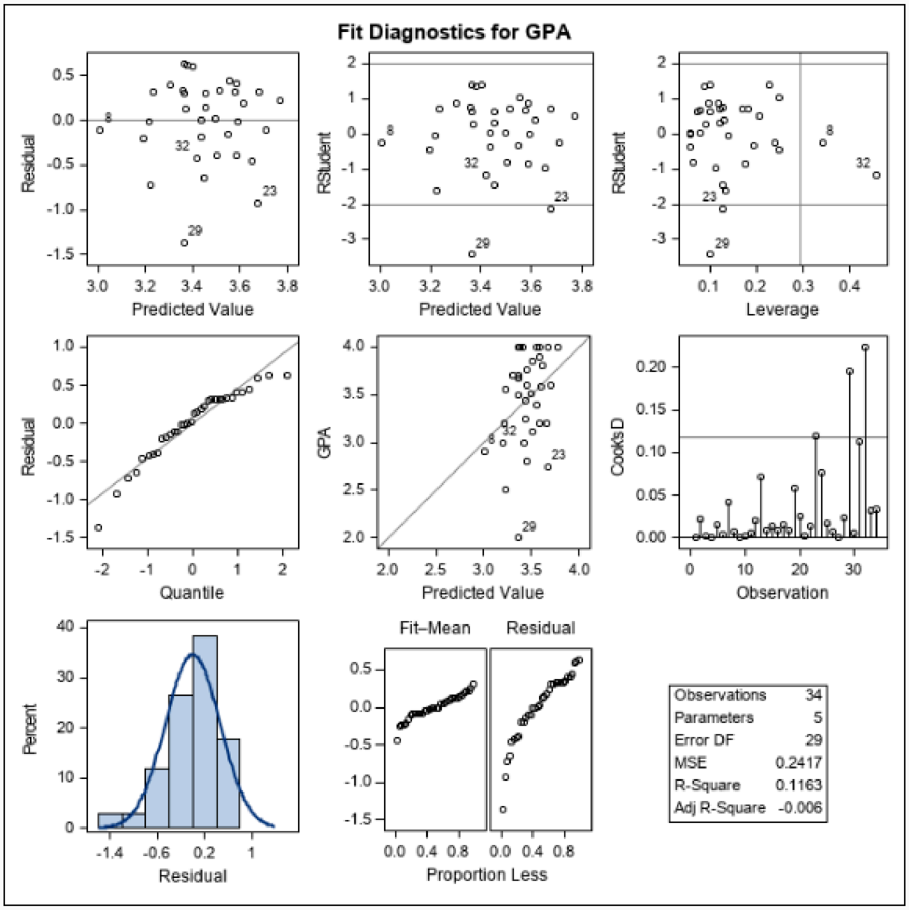
| Output Statistics | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Obs | TravelTime | DaysWeek | Credits | Job | Dependent Variable | Predicted Value | Std Error Mean Predict | 95% CL Mean | | 95% CL Predict | | Residual |
| 35 | 10 | 5 | 18 | 0 | . | 3.3654 | 0.1570 | 3.0444 | 3.6864 | 2.3099 | 4.4209 | . |

The confidence interval for $E(y)$is (3.0444, 3.6864). With 95% confidence, we can say that the mean GPA of students with a 10-minute travel time who is taking 18 credits, comes to campus 5 days a week, and has no off-campus job is between 3.0444 and 3.6864 points. The actual mean GPA for students without a job who were surveyed is 3.389, which falls into our confidence interval.

The prediction interval for $(y)$is (2.3099, 4.4209). With 95% confidence, we can say that the GPA of a student with a 10-minute travel time who is taking 18 credits, comes to campus 5 days a week, and has no off-campus job is between 2.3099 and 4.4209 points. The actual GPA of this student (3.85) does fall in this interval. However, since the prediction interval is extremely wide, it is not very useful.

## V. Assumptions

Our assumptions are that errors at each x are normally distributed with a mean of zero. Another assumption is that there is a constant variance at each of the x values indicated by similar spreads. Lastly, we assume that the distribution at each x is independent of the other x values. We use the following fit diagnostics produced by SAS to determine whether these assumptions hold true.




<u>Graph 1: Residuals vs Predicted Values</u>

No trends or patterns are present.

<u>Graph 2: Externally Studentized Residuals vs Predicted Values</u>

There are no patterns presented. Observations 23 and 29 are potential outliers, which are residuals that are leads than -2 standard deviations away from the mean residual. Observation 23 is a student who earned a GPA of 2.75, was on campus an average of 3 days a week, and worked 20 hours a week on average. Observation 29 is a student who earned a GPA of 2.00, was on campus an average of 5 days a week, and did not have an off-campus job. These GPAs are 2 of the lowest in the data set. There are only a total of 4 points that are less than 2.90 GPA, because of this the model is skewed to predict for a higher GPA value.

<u>Graph 3: Externally Studentized Residuals vs Leverage</u>

Observation 8 and 32 have high leverage indicating that they have an extreme value on a predictor. Observation 8 is a student who worked 55 hours per week on average, the highest hours worked out of all data points. Observation 32 is a student whose average travel time was 90 minutes, the highest value of all data points.

Graph 4: Residuals vs Quantile

Residuals normally distributed with most points falling very close to the diagonal line. A possible pattern is present. This could be further examined, but for our purposes it is not major enough to alter the model.

Graph 5:  GPA vs Predicted Value

Most of the estimates fall between 3.0 and 3.7 indicating that the model may be bias to predict a GPA in this range. Several outliers are presented.

Graph 6: CooksD vs Observation

Observations 23, 29, and 32 are once again shown to be possibly influential.

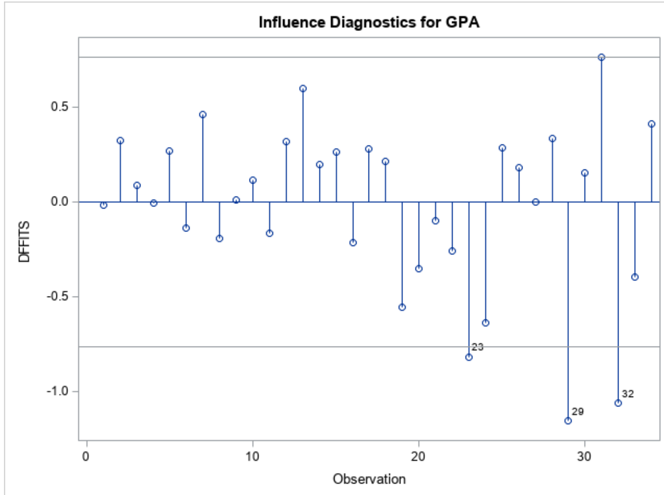
Graph 7:Histogram of the Residuals

The residuals show no major departure from normality. It could be perceived as slightly skewed to the left (negatively skewed), indicating that the model is more likely to overpredict GPA.

Graph 8: Residual Fit Plot

The majority of the residual spread falls in line with the spread of the fit with the exception of about 6 points.

To measure the level of influence of the above mentioned observations, we examined the fit difference. These DFFITS values show that observations 23, 29, and 32 are influential and if removed would significantly change the predicted values.




The model is a satisfactory predictor but is not exceptional. A few influential points are present. The assumptions did not hold for the estimators we used.
Based on this data set and analysis of S value, R values, and fit diagnostics (Sections IV and V), the model created explains little variation in GPA. The most relevant variables included in the model is the number of hours a student worked during the average week and whether or not a student had an off-campus job. Travel time and days per week on campus are almost irrelevant to the GPA of a commuter student.

## VI. Flaws of the Study

A. Measurement and Human Errors

1. Students who were taking the survey may have provided inaccurate or incorrect information.
    a. Fabrication of data may have occurred if students were unwilling to disclose their actual spring term GPA.
    b. Students may also have not made the effort to look up the correct information and just put an estimate.
2. Students could have possibly misunderstood some questions. In retrospect, we should have made the questions more specific and clear. In Survey Monkey, there is an option to make a question mandatory and limit the answer choices, so the participant cannot skip it or enter an extreme value. We failed to employ these techniques. In addition, all of the information was provided anonymously, so we were unable to verify the collected data.
3. Due to the above mentioned reasons, we had to omit 2 of the observations.
    a. One student filling out the online survey failed to give the number of days he or she is on campus during the week. Including this value

would have given us an incorrect prediction since a variable value is missing.

b. Another student filling out the online survey entered a commute time of 22 hours and 30 minutes. This is a highly unlikely scenario, so it is possible the student may have misunderstood the question or made an error in entering information. Including this data point would have extremely skewed our results.

B. Lurking Variables

1. Backwards elimination only left out the number of credits, which may have contributed significant information to student's GPA.
    a. If we chose to use the stepwise model selection, the only variable that SAS deems necessary is Credits.

| Summary of Stepwise Selection | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Step | Variable Entered | Variable Removed | Number Vars In | Partial R-Square | Model R-Square | C(p) | F Value | Pr > F |
| 1 | Credits | | 1 | 0.2275 | 0.2275 | 1.8026 | 9.43 | 0.0043 |

    b. If we chose to use forward model selection, the variables that SAS deems necessary are Credits, Job, and HoursWeek.

| Summary of Forward Selection | | | | | | | |
|---|---|---|---|---|---|---|---|
| Step | Variable Entered | Number Vars In | Partial R-Square | Model R-Square | C(p) | F Value | Pr > F |
| 1 | Credits | 1 | 0.2275 | 0.2275 | 1.8026 | 9.43 | 0.0043 |
| 2 | Job | 2 | 0.0424 | 0.2699 | 2.0571 | 1.80 | 0.1894 |
| 3 | HoursWeek | 3 | 0.0433 | 0.3133 | 2.2735 | 1.89 | 0.1791 |

    c. In both of these selection methods, Credits is the only variable with a p-value which makes it significant (using an alpha of 0.05). Credits is shown to be a good predictor (out of the variables we collected) of GPA, but it is not one that is exclusive to commuter students nor is it affected by commuting.
2. Many other variables which were not collected can also affect a student's GPA.
    a. High school GPA
    b. Living situation (children, married, who they lived with, etc)
    c. Major/Minor
    d. Commuting Method (walk, public transport, carpool, personal car)

C. Small Sample Size

1. Relatively small sample size skewed our model to predict GPAs closer to 3.00. Model predicts higher than actual GPAs. Since we ended up using more than 3 variables, we would have definitely benefited from more data points.

D. Sample Was Not Truly Random
   1. Students who may have initially decided to fill out the survey may have declined after seeing the questions. The survey we conducted was voluntary. Like we mentioned earlier, those students with lower GPAs may have been unwilling to provide the information (or fabrication of data may have occurred). Some may have decided that they did not want to put forth the effort in order to answer the questions (look up GPA, look up number of credits).
   2. This differs from an experiment or survey in which randomly chosen participants are forced to fill out the questions.

## VII. Conclusion/Interpretation

Time and time again, it has been shown that this model (using job status, hours worked per week, travel time to campus, and number of days on campus) is not an effective or useful predictor of a student's GPA. Based on this data set, a possible conclusion is that commuting by itself has minimal to no effect on GPA. The time spent commuting to campus and days per week does not affect a commuter student's academic performance. The commuter students sampled who have an off-campus job achieve slightly higher GPAs than their counterparts who do not have off-campus jobs. This possibly shows that students with jobs have a higher work ethic than those without. Our model has shown little to no indication that a student's travel time and days on campus affect their GPA.

## VIII. Possible Improvements

A. Below are some additional variables we could have included.
   1. High School GPA
   2. Major/Minor
   3. Living situation (dummy variable: by self, with family, with friends, with spouse/significant other)
   4. Commuting method (own car, dropped-off, public transport). This may also reflect other factors such as financial situation, class schedule, and other time restrictions. If a student is limited by the public transport availability, it may hinder their ability to use resources provided on campus (computer labs, study rooms, library, tutor session, and study rooms).

B. We would be better off with a wider sample.
   1. This data set includes many GPAs above 3.00 and few from 2.00 to 2.90. A wider sample set could help determine if this is the normal distribution of GPAs across all UPJ commuters or if our survey has reached more average and higher achieving students than lower achieving students.
   2. A wider sample set may also provide more data points from students without an off-campus job (only 35.29% of data points were from students without jobs). If this is an inaccurate representation of the overall percentage of commuter students who do not work off-campus, our data is skewed to commuter students who do have off-campus jobs and may make model creation less accurate.

C. Inclusion of Resident Students
   1. The inclusion of resident students and a second dummy variable (for commuting or not) would help determine if there is a significant difference in GPAs between commuter students and resident students. This could show whether or not commuting, by itself, has an effect on GPA.