Department of Management, UTSC

MGEB12H3S Quantitative Methods in Economics II

Case Study – Winter 2015

Factors Affecting GPA and GPA Performance Between Males &Females

Students: Isaac Aktam (998986575), Wajahat Ali (999817328)

Professor: Dr.Vinh Quan

Due Date: Thursday, 6 PM, April 2, 2015

Lecture: L30

Table of Contents

Summary/Abstract

The purpose of this project is to determine what factors have the most significant effect on student's grade point average (GPA) and subsequently determine if there is a significant difference in academic achievement between males and females. To do such analysis, we have surveyed 103 University of Toronto Scarborough students from various departments. From our analysis, we gather the following results:

- For all of the 103 students, GPA is affected by hours studied, courses taken, and classes missed.
- For 50 male students, GPA is affected by classes missed and hours taken
- For 53 female students, GPA is affected by classes missed.
- Additionally, our hypothesis analysis showed that on average females perform better academically than males.

Introduction

The purpose of this project is to determine what factors have the most significant effect on student's grade point average (GPA) and subsequently determine if there is a significant difference in academic achievement between males and females. Students often question themselves where they should concentrate the most to improve their GPA. This is the primary reason we decided to do this research.

Therefore, our research question is:

- What factors affect GPA the most?
- Who performs better academically, males and females?

To do such analysis, we have surveyed 103 University of Toronto Scarborough students from various departments. There are many factors affecting students GPA, however we decided to focus on factors like Extracurricular Hours, Amount of Courses Taken, Amount of Classes Missed, Hours Studying, Happiness Level, Hours Spent Working, Hours Spent on Commuting, and Hours Spent on Relationships.

Factors affecting GPA and GPA performance between males and females

Refer to Appendix A to see the sample survey we used. The survey questions cover both personal, academic, and social life of students thus encompassing a great range of activities that we hypothesize to have a great effect on student's GPA. We divide our analysis into two parts:

1. Linear regression analysis of gathered data from 103 surveys for the purpose of determining factors influencing student's GPA. Additionally, we employ the method of Backward Elimination.
2. Hypothesis analysis of GPA of 50 male students (Sample 1) and GPA of 53 female students (Sample 2) for the purposes of determining who performs better academically.

For this study, we hypothesize the following results to be true:

1. Studying hours and happiness level to have high positive correlation with GPA.
2. Work hours, classes missed, and relationship commitments have strong negative correlation with GPA.
3. Amount of courses taken have either weak positive correlation or weak negative correlation with GPA.
4. Commute hours, classes missed, and extracurricular activities have weak negative correlation with GPA.
5. Equal academic performance between male students and female students.

Similar Studies

In statistics research, there exist many academic papers that discuss which variables affect the students' GPA. In a research study, Urien (2003) found that personal factors like family background and lifestyle has an effect on how well students perform at school. In supporting this research, a study by Fertig and Schmidt (2002) stated that there is a positive correlation between students' personal life and their academic performance. Studies in the United States by authors like Robst & Keil (2000) and Stinebrickner & Stinebrickner (2003) have established a good correlation between students' GPA and their past academic and extracurricular activities.

Additionally, we are interested in looking at our problem from a different perspective; between males and females, who performs better academically vise? There have been many studies attempting to figure out the significant factors that relate to the gender differences in academic performance. The studies conducted around the world present different outcomes. A study by Warrington and Williams (1999) in the UK, focused on the performance level between girls and boys in the secondary schools. Girls were believed to perform better because of their maturity and great learning techniques. It was also because of the differences in the students' attitude and behavior towards their academic life. In another research study, findings stated that females work harder and attend their classes more often than the males, thus receiving better grades at school. (Wainer and Steinberg, 1992). In analysis of achievement tests by authors like Lindberg, Pomerantz, Mikelson & Greene, age was seen as an important factor. There seemed to be a male advantage in certain achievement tests (i.e. math tests) at early age in elementary schools. However, findings between genders varied at the university level.

Methodology and Analysis

To gather the necessary data, we surveyed 150 University of Toronto Scarborough students. This group of 150 people includes students from IC and BV. Each of them received a single survey. Our survey included short and concise questions so as to let students finish them quickly and not to discourage students from not answering them. We were not guaranteed to receive a completely answered survey. As a result, after gathering all of the surveys, we had to drop 47 faulty questionnaires (i.e. people would leave some of the questions blank or give a vague answer). As a result we were left with 103 surveys. Next, we divided 103 remaining surveys into a Male group of 50 surveys and a Female group of 53 surveys. Lastly, we gathered all of the necessary results for the further analysis. Since the size of each group/sample is bigger than 30, we are guaranteed to have a normally distributed data. All of the hypothesis and confidence interval calculations, Excel output, and related plots are included in Appendix.

Part 1.0: Regression Analysis – General Information

In general, population GPA can be described by the following regression model Y = β_0 + $\beta_1 X_1$+$\beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8$ + ε. But, since we are dealing with samples instead of populations, we can only estimate the GPA results. Therefore, we use the regression estimation line$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 + b_6 X_6 + b_7 X_7 + b_8 X_8$э:

- X_1 = *Extracurricular hours per week*
- X_2 = *Number of courses taken this semester*
- X_3 = *Classes missed so far into the semester*
- X_4 = *Hours per week spent on studying*
- X_5 = *Happiness level(from 1 to 5)*
- X_6 = *Hours per week spent on working*
- X_7 = *Hours per week spent on commuting*
- X_8 = *Hours per week spent on relationship*

To build the estimated regression equation, we employ the method of Backward Elimination.

Part 1.1.0: Regression Analysis– Backward Elimination, Males and Females Grouped Together

First of all, let's us analyze the relationship between each independent variable and dependent variable GPA. The reason we are doing this is to get an idea what independent variables have the most effect on GPA. Therefore, we have the following results:

Correlation table	
	GPA
Extracurricular Hours	0.001643479
Course #t	0.215309578
Classes missed	-0.320595634
Studying Hours	0.274095319
Happiness	0.119464834
Employment Hours	-0.139038422
Commute Hours	0.016810694
Relationship Hours	0.01155027

- Studying hours, happiness level, and commute hours have weak positive correlation with GPA.
- Work hours and classes missed have weak negative correlation with GPA. Whereas, relationship commitment hours have weak positive correlation with GPA.
- Amount of courses taken and extracurricular activities have weak positive correlation with GPA.

We build our estimated regression equation using the following algorithm:

1. For the male sample of 103 students, we start with 8 independent variables.
2. We compute *p-value* for each independent variable in the model.
3. Look for the independent variable with the maximum *p-value* > $\alpha = 0.05$.

4. Independent variable with the largest *p-value* is removed from the model.
5. Stop if *p-value* for all the remaining independent variables is less than or equal to $\alpha = 0.05$. Else, return to step 3 and continue on with the process.

Therefore, we will have K ≥ 0 runs so as to reach our final estimated regression equation. Additionally, once we get the final regression equation, we will be able to run a hypothesis test on each single independent variable to determine if it has any effect on GPA.

Variables in red text have the greatest *p-value* and thus dropped in the next run. Therefore, we have the following results:

Run	Regression equation
0	Yhat = 2.4809-0.00938(Extracurricular hours)+0.117457(Number of course taken)-0.117457(Classes missed)+0.00681(Studying hours)+0.04849(Happiness)-0.0075(Employment hours)+0.01449(Commute hours)+0.00161(Relationship hours)
1	Yhat = 2.65091-0.0008(Extracurricular hours)+0.11614(Number of course taken)-0.0259(Classes missed)+0.0069(Studying Hours)+-0.0061(Employment Hours)+0.011(Commute hours)+0.00166(Relationship hours)
2	Yhat = 2.64489+0.10941(Number of course taken)-0.0246(Classes missed)+0.00692(Studying Hours)-0.0061(Employment Hours)+0.00863(Commute hours)+0.00152(Relationship hours)
3	Yhat = 2.69718+0.10687(Number of course taken)-0.0245(Classes missed)+0.00692(Studying Hours)-0.005(Employment Hours)+0.00133(Relationship hours)
4	Yhat = 2.76022+0.09601(Number of course taken)-0.0237(Classes missed)+0.00694(Studying Hours)-0.0054(Employment Hours)
5	Yhat = 2.69023+0.09601(Number of course taken)-0.10404(Classes missed)+0.00714(Studying Hours) $R_a^2 = 0.1943$.

Therefore, we run a hypothesis test to determine if the set of variables; Number of courses taken, Classes missed, and Studying hours as a whole have effect on GPA.

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

H_a: $at\ least\ one\ \beta_i \neq 0$

Critical value approach shows that we reject the hypothesis and conclude that overall set of variables; Number of courses taken, Classes missed, and Studying hours as a whole have effect on GPA since $F_0 \geq F\alpha \rightarrow 9.2037 \geq 2.6964$. Additionally, p-value approaches supports the above the result since $P[F \geq F_0] = 1.99*10^{\wedge}(5) < 0.05$.

But, does there exist a relationship between Number of courses taken and GPA? How about Classes missed and GPA? And, lastly what about Studying hours and GPA? To answer these questions, we run the following hypothesis tests. But, is correlation between independent variables low? Yes, as can be seen in Excel output:

	Classes missed	*Studying Hours*
Classes missed	1	
Studying Hours	0.0354	1

	Course amount	*Studying Hours*
Course amount	1	
Studying Hours	0.0322	1

	Course amount	*Classes missed*
Course #t	1	
Classes missed	-0.0697	1

We care about correlation because *t-test* tends to give faulty answers when correlation between independent variables is high. But, in our case it is low. This means that our following *t-test* hypothesis results should reinforce the *F-test* hypothesis result above.

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

We reject the hypothesis and conclude that Course amount has effect on GPA since *p-value*$< \alpha = 0.05 \rightarrow 0.004102 < 0.05$.

H_0: $\beta_2 = 0$

H_a: $\beta_2 \neq 0$

We reject the hypothesis and conclude that Classes missed has effect on GPA *p-value*<α = 0.05 → 0.000566 < 0.05

H_0: $\beta_3 = 0$

H_a:$\beta_3 \neq 0$

Lastly, we reject the hypothesis and conclude that Studying hours has effect on GPA since *p-value*<α = 0.05 → 0.00226 < 0.05

Therefore, given above *t-test* hypothesis results reinforce our *F-test* hypothesis results.

Part 1.2.1: Regression Analysis – Backward Elimination, Males

Run	**Regression equation for male sample**
0	y = 2.478-0.0061(Extracurricular Hours) +0.1262(Course Amount)-0.0176(Classes Missed) +0.0067(Studying Hours) +0.0368(Happiness) – 0.0085(Employment Hours) + 0.0013(Commute Hours) + 0.001(Relationship Hours).
1	y = 2.4869-0.0059(Extracurricular Hours) +0.1267(Course Amount)-0.0176(Classes Missed) +0.0067(Studying Hours) +0.0351(Happiness) – 0.0084(Employment Hours) + 0.00099(Relationship Hours).
2	y = 2.4989+0.1183(Course Amount)-0.017 (Classes Missed) +0.0067(Studying Hours) +0.0333(Happiness) – 0.0088(Employment Hours) + 0.001(Relationship Hours).
3	y = 2.5004+0.11723(Course Amount)-0.0159 (Classes Missed) +0.0066(Studying Hours) +0.0368(Happiness) – 0.0091(Employment Hours)
4	y = 2.6017+0.11904(Course Amount)-0.0171 (Classes Missed) +0.0069(Studying Hours) – 0.00856(Employment Hours)
5	y = 2.5317+0.1229(Course Amount)-0.0165 (Classes Missed) +0.00714(Studying Hours).
6	y = 3.0809-0.0179 (Classes Missed) +0.0071(Studying Hours) $R_a^2 = 0.2055$

Therefore, we run a hypothesis test to determine if the set of variables Classes missed and Studying hours as a whole have effect on GPA.

H_0: $\beta_1 = \beta_2 = 0$

H_a: $at\ least\ one\ \beta_i \neq 0$

Critical value approach show that we reject the hypothesis and conclude that overall set of variables; Classes missed and Studying hours as a whole have effect on GPA since $F_0 \geq F\alpha \rightarrow 7.3399 \geq 3.19$. Additionally, p-value approach supports our above result since $P[F \geq F_0] = 0.001682 < 0.05$.

But, does there exist a relationship between Classes Missed and GPA? How about Studying hours and GPA? To answer these questions, we run the following hypothesis tests. But, is correlation between independent variables low? Yes, as can be seen in Excel output:

	Classes missed	*Studying Hours*
Classes missed	1	
Studying Hours	-0.017318299	1

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

Reject the hypothesis and conclude that Classes missed has effect on GPA since *p-value*$< \alpha = 0.05 \rightarrow 0.024 < 0.05$

H_0: $\beta_2 = 0$

H_a: $\beta_2 \neq 0$

Reject the hypothesis and conclude that Studying hours has effect on GPA since *p-value*$< \alpha = 0.05 \rightarrow 0.0042 < 0.05$.

Part 1.2.2: Regression Analysis – Backward Elimination, Females

Run	**Regression equation for female sample**
0	y = 2.4762-0.0106(Extracurricular Hours) +0.1278(Course Amount)-0.0326(Classes Missed)

	+0.0069(Studying Hours) +0.0404(Happiness) – 0.0041(Employment Hours) + 0.0212(Commute Hours) + 0.002(Relationship Hours)
1	y = 2.4526-0.0096(Extracurricular Hours) +0.1355(Course Amount)-0.033(Classes Missed) +0.0072(Studying Hours) +0.0312(Happiness) + 0.0185(Commute Hours) + 0.0021(Relationship Hours)
2	y = 2.4526-0.0096(Extracurricular Hours) +0.1355(Course Amount)-0.033(Classes Missed) +0.0072(Studying Hours) +0.0185(Commute Hours) + 0.0021(Relationship Hours)
3	y = 2.5798+0.1233(Course Amount)-0.0331(Classes Missed) +0.0074(Studying Hours) +0.0141(Commute Hours) + 0.0018(Relationship Hours).
4	y = 2.7785+0.0856(Course Amount)-0.0327(Classes Missed) +0.0083(Studying Hours) +0.011(Commute Hours)
5	y = 2.9053+0.0679(Course Amount)-0.0333(Classes Missed) +0.0091(Studying Hours)
6	y = 3.212-0.0333(Classes Missed) +0.0101(Studying Hours).
7	y = 3.3835-0.009(Classes Missed) $R_a^2 = 0.0975$

Therefore, we run a hypothesis test to determine if the variable Classes missed has effect on GPA.

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

Critical approach show that we reject the hypothesis and conclude the variables; Classes missed has effect on GPA since $F_0 \geq F\alpha \rightarrow 6.62 \geq 4.034$. Additionally, p-value approach reinforces our conclusion since $P[F \geq F_0] = 0.013 < 0.05$.

<u>Part 1.2.3: Hypothesis Testing and 100(1-0.05)% Confidence Intervals for Independent Variables of Male Regression Model and Female Regression Model</u>

<u>Male regression model</u>:Yhat = 3.0809-0.0179 (Classes Missed) +0.0071(Studying Hours)

<u>Female regression model</u>:Yhat = 3.3835-0.009(Classes Missed).

Since Classes Missed is a common variable between 2 regression equations, it is appropriate to run various hypothesis tests and build 95% Confidence Intervals for Classes Missed so we can answer following questions: Who on average misses more classes, males or females? Who has greater variation for classes missed, males or females? Who has bigger confidence interval for mean Classes Missed, males or females? Who has bigger

We are interested in who, on average, misses their classes the most, males or females. First, we check if there exist a difference between the mean classes missed by males and females. Therefore, we run the following hypothesis test:

H₀: $\mu_{Males} - \mu_{Females} = 0$

Hₐ: $\mu_{Males} - \mu_{Females} \neq 0$

Critical-value approach shows we do not reject hypothesis and concluded that there exists a difference in average number of classes missed since $\begin{cases} t_0! \leq -t_{\frac{\alpha}{2}} \\ t_0! \geq t_{\frac{\alpha}{2}} \end{cases} \rightarrow \begin{cases} 1.143 ! \leq -1.98 \\ 1.143 ! \geq 1.98 \end{cases}$. P-value approach reinforces our above conclusion since $2P[t \geq t_0]$ = $2P[t_{90} \geq 1.143]$ = 2*0.128 !< $\alpha = 0.05$.

Our next hypothesis test is about who actually misses the most classes on average, males or females?

H₀: μ_{Males} - $\mu_{Females} \leq 0$

Hₐ: μ_{Males} - $\mu_{Females} > 0$

Critical-value approach shows that we do not reject the hypothesis and conclude that, on average, females miss more classes than males since $t_0 \geq t_{\alpha}$ →1.143 !≥ 1.662. Additionally, p-value approach reinforces our above conclusion since $P[t \geq t_0]$ = $P[t_{90} \geq 1.143]$ = 1.143 !< $\alpha = 0.05$.

Additionally, we can run the hypothesis to test to determine who out of two samples has the greatest Classes Missed variance. To do so we apply *F-test* for testing the ratio of 2 population variances.

H₀ : $\sigma^2_{Females} \leq \sigma^2_{Males}$

Hₐ : $\sigma^2_{Females} > \sigma^2_{Males}$

Critical-value approach shows us that we reject hypothesis and conclude that males' Classes Missed has a greater variance than that of females since $P[F_{52,49} \geq F_{0.05}] = 0.05 \rightarrow F_{0.05} \approx 1.6 \rightarrow F_0 = 1.7146 \geq 1.6$. Additionally, p-value approaches reinforces our above conclusion since $P[F \geq F_0] = P[F_{52,49} \geq 1.7146] = 0.029 < 0.05$.

The following table contains the summary our 95% Confidence Intervals and necessary conclusions:

95% Confidence Intervals for Classes Missed	
Difference in mean Classes Missed for the 2 genders	$-1.1588 \leq \mu_{Males} - \mu_{Females} \leq 4.302$
Ratio of $\sigma^2_{Females}$ and σ^2_{Males}:	$0.9835 \leq \frac{\sigma^2_{Males}}{\sigma^2_{Females}} \leq 3.003$
Male Classes Missed population variance, σ^2_{Males}	$42.5 \leq \sigma^2_{Males} \leq 94.58$
Female GPA population variance, $\sigma^2_{Females}$	$25.0264 \leq \sigma^2_{Females} \leq 54.38$
Male Classes Missed population mean: μ_{Males}	$4.8815 \leq \mu_{Males} \leq 9.3184$
Female Classes Missed population mean: $\mu_{Females}$	$3.8852 \leq \mu_{Females} \leq 7.1714$

We can make two main conclusions regarding above 95% Confidence Intervals:

1. Since the difference between male population mean and female population mean can be positive and negative, we cannot make a true conclusion regarding who on average misses classes the most.
2. Since the ratio of male population variance and female population variance can be less than 1 and greater than 1, we cannot make a true conclusion regarding whose Classes Missed varies the most.

Part 2: Analysis of GPA between males and females

To determine who performs better in terms of GPA, males or females, we run the following hypotheses tests.

Our first hypotheses test checks if there exists a significant difference between the mean GPA of males and mean GPA of females.

H_0: $\mu_{Males} - \mu_{Females} = 0$

H_a: $\mu_{Males} - \mu_{Females} \neq 0$

Critical-value approach shows us that we reject the H_0 and conclude that there is a difference between mean Male GPA and mean Female GPA since

one of the conditions for rejection of null hypothesis holds true: $\begin{cases} t_0 \leq -t_{\frac{\alpha}{2}} \\ t_0 \geq t_{\frac{\alpha}{2}} \end{cases} \rightarrow$

$\begin{cases} -5.627 \leq -1.984 \; Reject \\ -5.627 \geq 1.984 \; Do\ not\ reject \end{cases}$. Additionally, p-value approach reinforces above conclusion since $2P[t \geq -t_0] < \alpha \rightarrow 2P[t_{100} \geq -5.627] = 2(1 - P[t_{100} \leq 5.627]) \approx 2(1-1) = 0 < 0.05$

Therefore, we now need to test who performs better, Males or Females

$H_0: \mu_{Males} - \mu_{Females} \leq 0$

$H_a: \mu_{Males} - \mu_{Females} > 0$

Critical-value approach shows that we do not reject the hypothesis and concluded that on average females perform better academically than males since $t_0 \geq t_\alpha \rightarrow -0.888 \,!\geq 1.645$. P-value approach reinforces the above statement since $P[t \geq t_0] \rightarrow P[t_{100} \geq -0.888] = 0.188 \,!< 0.05$.

Now, we are interested in who has greater GPA variance, males or females? Thus, we run the following hypothesis test:

$H_0: \sigma^2_{Females} \leq \sigma^2_{Males}$

$H_a: \sigma^2_{Females} > \sigma^2_{Males}$

Critical-value approach shows that we do not reject the hypothesis and conclude that males have a higher GPA variance compared to females since $P[F \geq F_\alpha] = \alpha \rightarrow P[F_{52,49} \geq F_{0.05}] = 0.05 \rightarrow F_{0.05} \approx 1.6 \rightarrow F_0 = 1.3165 \,!\geq 1.6$. Additionally, p-value approach reinforces our above conclusion since $P[F \geq F_0] = P[F_{52,49} \geq 1.3165] = 0.17 \,!< 0.05$.

Now, we can proceed with the construction of 95% Confidence intervals:

100(1-0.05)% Confidence Intervals

Difference in mean GPA for the 2 genders	$-0.2878 \le \mu_{Males} - \mu_{Females} \le 0.1098$
Ration of $\sigma^2_{Females}$ and σ^2_{Males}	$0.7516 \le \frac{\sigma^2_{Females}}{\sigma^2_{Males}} \le 2.2951$
Male GPA population variance, σ^2_{Males}	$0.1563 \le \sigma^2_{Males} \le 0.3478$
Female GPA population variance, $\sigma^2_{Females}$	$0.2077 \le \sigma^2_{Females} \le 0.4514$
Male GPA population mean: μ_{Males}	$2.9892 \le \mu_{Males} \le 3.2583$
Female GPA population mean: $\mu_{Females}$	$3.063 \le \mu_{Females} \le 3.3625$

We can make two major conclusions regarding above confidence intervals:

1. We see that the difference between male population mean and female population mean is between negative and positive numbers. Therefore, we cannot make a conclusion regarding who performs better academically.
2. We see that that the ratio of male variance and female variance is between the values that are less than one and greater than one. Therefore, we cannot make a conclusion regarding whose GPA is spread more around the mean.

Conclusion

Our main results are as follows:

- General regression equation: Yhat = 2.69023+0.09601(Number of course taken)-0.10404(Classes missed)+0.00714(Studying Hours). $R_a^2 = 0.1943$.
- Male regression equation: Yhat = 3.0809-0.0179 (Classes Missed) +0.0071(Studying Hours). $R_a^2 = 0.2055$
- Female regression equation: Yhat = 3.3835-0.009(Classes Missed). $R_a^2 = 0.0975$
- Mean GPA performance of females is higher than that of males.

For all of the above regression equations, adjusted coefficient of determination is quiet low. This means our regression equations explain/predict less than 21% of the actual data. Does this mean that the actual regression model Y = β_0 + $\beta_1 X_1$+$\beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8$ + ε can explain the real life GPA data with less than 21% accuracy? No, it does not mean that. Our regression equations are built only for our sample data and sample data results should not be generalized to explain population data. What can be done to improve our results? First of all, we can increase sample data to more than 103 data points. Second, we can use other methods of regression equation building such as Stepwise Regression, Forward Selection, and Best-Subset Regression. Third, to make our hypothesis analysis and confidence intervals more precise, we can use α = 0.025, 0.01, 0.001. Fourth, we can use different sampling methods such as online surveying or phone surveys. Lastly, we can increase the number of independent (i.e. add such variables as Leisure Hours, Socializing Hours, and Sleeping Hours) variables to make our regression equations more precise.

Appendix

SAMPLE SURVEY QUESTIONS

- What is your gender?
- How many hours per week do you work?
- How many hours per week do you spend on commuting to and from university?
- If you are in a relationship, how many hours per week do you spend on your partner?
- How many hours per week do you spend on extracurricular activities?
- How many courses are you taking this semester?
- How many lectures, in total, have you missed during this semester?
- How many hours per week do you study?
- What is your GPA?
- One a scale from 1 to 5, specify your happiness level for this semester: 1 2 3 4 5

Part 1.1.0: Regression Analysis – Backward Elimination, Males and Females Grouped Together

Run a hypothesis test to determine if the set of variables Number of courses taken, Classes missed, and Studying hours as a whole have effect on GPA.

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

H_a: $at\ least\ one\ \beta_i \neq 0$

Reject H_0 if $F_0 \geq F\alpha$ or $P[F \geq F_0] < \alpha = 0.05$ such that $P[F \geq F_\alpha] = \alpha$

$P[F \geq F_\alpha] = \alpha \rightarrow P[F_{3,103-3-1} \geq F_{0.05}] = 0.05 \rightarrow F_{0.05} = 2.6964$

Critical value approach

$F_0 \geq F\alpha \rightarrow 9.2037 \geq 2.6964 \rightarrow$Reject the hypothesis and conclude that overall set of variables; Number of courses taken, Classes missed, and Studying hours as a whole have effect on GPA.

<u>P-value approach</u>

$P[F\geq F_0]=1.99*10^{\wedge}(5)< 0.05$. Therefore, reject the hypothesis and conclude that overall set of variables; Number of courses taken, Classes missed, and Studying hours as a whole have effect on GPA

	Course #t	*Classes missed*
Course #t	1	
Classes missed	-0.0697	1

	Classes missed	*Studying Hours*
Classes missed	1	
Studying Hours	0.0354	1

	Course #t	*Studying Hours*
Course #t	1	
Studying Hours	0.0322	1

H_0: $\beta_1 = 0$

H_a:$\beta_1 \neq 0$

Reject H_0if $p\text{-}value<\alpha = 0.05 \rightarrow 0.004102 < 0.05 \rightarrow$ Reject the hypothesis and conclude that Course # has effect on GPA.

H_0: $\beta_2 = 0$

H_a:$\beta_2 \neq 0$

Reject H_0if $p\text{-}value<\alpha = 0.05 \rightarrow 0.000566 < 0.05 \rightarrow$ Reject the hypothesis and conclude that Classes missed has effect on GPA.

H_0: $\beta_3 = 0$

H_a:$\beta_3 \neq 0$

Reject H_0if $p\text{-}value<\alpha = 0.05 \rightarrow 0.00226 < 0.05 \rightarrow$ Reject the hypothesis and conclude that Studying hours has effect on GPA.

Therefore, given above *t-test* hypothesis results reinforce our *F-test* hypothesis results.

<u>Part 1.2.1: Regression Analysis – Backward Elimination, Males</u>

Run a hypothesis test to determine if the set of variables Classes missed and Studying hours as a whole have effect on GPA.

H_0: $\beta_1 = \beta_2 = 0$

H_a: $at\ least\ one\ \beta_i \neq 0$

Reject H_0 if $F_0 \geq F\alpha$ or $P[F \geq F_0] < \alpha = 0.05$ such that $P[F \geq F_\alpha] = \alpha$

$P[F \geq F_\alpha] = \alpha \rightarrow P[F_{2,50-2-1} \geq F_{0.05}] = 0.05 \rightarrow F_{0.05} = 3.19$

<u>Critical value approach</u>

$F_0 \geq F\alpha \rightarrow 7.3399 \geq 3.19 \rightarrow$ Reject the hypothesis and conclude that overall set of variables; Classes missed and Studying hours as a whole have effect on GPA.

<u>P-value approach</u>

$P[F \geq F_0] = 0.001682 < 0.05$. Therefore, reject the hypothesis and conclude that overall set of variables; Classes missed and Studying hours as a whole have effect on GPA.

But, does there exist a relationship between Classes Missed and GPA? How about Studying hours and GPA? To answer these questions, we run the following hypothesis tests. But, is correlation between independent variables low? Yes, as can be seen in Excel output:

	Classes missed	*Studying Hours*
Classes missed	1	
Studying Hours	-0.017318299	1

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

Reject H_0 if $p\text{-}value < \alpha = 0.05 \rightarrow 0.024 < 0.05 \rightarrow$ Reject the hypothesis and conclude that Classes missed has effect on GPA.

H_0: $\beta_2 = 0$

H_a: $\beta_2 \neq 0$

Reject H_0 if *p-value* $<\alpha = 0.05 \rightarrow 0.0042 < 0.05 \rightarrow$ Reject the hypothesis and conclude that Studying hours has effect on GPA.

Part 1.2.2: Regression Analysis – Backward Elimination, Females

Run a hypothesis test to determine if the variable Classes missed has effect on GPA.

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

Reject H_0 if $F_0 \geq F\alpha$ or $P[F \geq F_0] < \alpha = 0.05$ such that $P[F \geq F_\alpha] = \alpha$

$P[F \geq F_\alpha] = \alpha \rightarrow P[F_{1,53-2-1} \geq F_{0.05}] = 0.05 \rightarrow F_{0.05} = 4.034$

Critical value approach

$F_0 \geq F\alpha \rightarrow 6.62 \geq 4.034 \rightarrow$ Reject the hypothesis and conclude the variables; Classes missed has effect on GPA.

P-value approach

$P[F \geq F_0] = 0.013 < 0.05$. Reject the hypothesis and conclude the variables; Classes missed has effect on GPA.

Part 1.2.1: Hypothesis Testing and 100(1-0.05)% Confidence Intervals for Independent Variables of Male Regression Model and Female Regression Model

Male regression model: y = 3.0809-0.0179 (Classes Missed) +0.0071(Studying Hours)

Female regression model: y = 3.3835-0.009(Classes Missed).

Since Classes Missed is the only independent variable in common, our hypothesis analysis is based solely on it.

$\bar{x}_{Males}$	$\bar{x}_{Females}$	s^2_{Males}	$s^2_{Females}$	n_{Males}	$n_{Females}$	α
7.1	5.5283	60.908	35.523	50	53	0.05

We interested in who, on average, misses their classes the most, males or females. First, we check if there exist a difference between the mean classes missed by males and females. Therefore, we run the following hypothesis test:

H_0: $\mu_{Males} - \mu_{Females} = 0$

H_a: $\mu_{Males} - \mu_{Females} \neq 0$

Reject H_0 if $\begin{cases} t_0 \leq -t_{\frac{\alpha}{2}} \\ t_0 \geq t_{\frac{\alpha}{2}} \end{cases}$ such that $t_0 = \frac{\bar{x}_{Males} - \bar{x}_{Females} - D}{\sqrt{\frac{s^2_{Males}}{n_{Males}} + \frac{s^2_{Females}}{n_{Females}}}}$. Or, reject H_0 if $\begin{cases} t_0 \geq 0: 2P[t \geq t_0] < \alpha \\ t_0 \leq 0: 2P[t \geq -t_0] < \alpha \end{cases}$

э: $P[t \geq t_\alpha] = \alpha, P[t \geq t_{\alpha/2}] = \frac{\alpha}{2}, \alpha = 0.05$

Therefore, $t_0 = \frac{7.1 - 5.5283 - 0}{\sqrt{\frac{60.908}{50} + \frac{35.523}{53}}}$ = 1.143. *d.f.* = $\frac{\left[\frac{s^2_{Males}}{n_{Males}} + \frac{s^2_{Females}}{n_{Females}}\right]^2}{\frac{\left[s^2_{Males}/n_{Males}\right]^2}{n_{Males} - 1} + \frac{\left[s^2_{Females}/n_{Females}\right]^2}{n_{Females} - 1}} = \frac{\left[\frac{60.908}{50} + \frac{35.523}{53}\right]^2}{\frac{\left[60.908/50\right]^2}{50-1} + \frac{\left[35.523/53\right]^2}{53-1}} \approx 90$

$P\left[t \geq t_{\frac{\alpha}{2}}\right] = \frac{\alpha}{2} \rightarrow P\left[t \geq t_{\frac{0.05}{2}}\right] = \frac{0.05}{2} \rightarrow t_{\frac{0.05}{2}} = t_{90,0.025}$=1.98

- Critical-value approach

  $\begin{cases} 1.143 \,! \leq -1.98 \; Do\ not\ reject \\ 1.143 \,! \geq \; 1.98 \; Do\ not\ reject \end{cases}$ → Do not reject the hypothesis and concluded that there exists a difference in average number of classes missed.

- P-value approach

  Since t_0= 1.143 > 0, we use the following method. $2P[t \geq t_0]$ = $2P[t_{90} \geq 1.143]$ = 2*0.128 !<$\alpha =$ 0.05 → Do not reject the hypothesis and concluded that there exists a difference in average number of classes missed.

Our next hypothesis test is about who actually misses the most classes on average, males or females?

H_0: $\mu_{Males} - \mu_{Females} \leq 0$

H_a : $\mu_{Males} - \mu_{Females} > 0$

Reject H_0 if $t_0 \geq t_\alpha$ or $P[t \geq t_0] < \alpha$. t_0 =1.143. *d.f.* = 90

$$P[t \geq t_{\alpha}] = \alpha \rightarrow P[t_{90} \geq t_{0.05}] = 0.05 \rightarrow t_{0.05} = 1.662$$

- Critical-value approach

  $t_0 \geq t_{\alpha}$ →1.143 !≥ 1.662 → Do not reject the hypothesis and conclude that, on average, females miss more classes than males.

- P-value approach

  $P[t \geq t_0] = P[t_{90} \geq 1.143] = 1.143$!< $\alpha = 0.05$ → Do not reject the hypothesis and conclude that, on average, females miss more classes than males.

Additionally, we can run the hypothesis to test to determine who out of two samples has the greatest Classes Missed variance. To do so we apply *F-test* for testing the ratio of 2 population variances.

$H_0 : \sigma^2_{Females} \leq \sigma^2_{Males}$

$H_a : \sigma^2_{Females} > \sigma^2_{Males}$

Reject H_0 if $F_0 \geq F_{\alpha}$ or $P[F \geq F_0] < \alpha$. $F_0 = \frac{s^2_{Females}}{s^2_{Males}}$ ϶: *d.f. of numerator* = 52 and *d.f. of denominator* = 49.

$F_0 = \frac{60.908}{35.523} = 1.7146.$

- <u>Critical-value approach</u>: $P[F \geq F_{\alpha}] = \alpha \rightarrow P[F_{52,49} \geq F_{0.05}] = 0.05 \rightarrow F_{0.05} \approx 1.6 \rightarrow 1.7146 \geq 1.6$. Therefore, do reject H_0 and conclude that females' Classes Missed has a greater variance than that of males.

- <u>P-value approach</u>: $P[F \geq F_0] = P[F_{52,49} \geq 1.7146] = 0.029 < 0.05$. Therefore, do reject H_0 and conclude that females' Classes Missed has a greater variance than that of males.

Now, we can proceed with the construction of 95% Confidence intervals:

<u>100(1- 0.05) % 2-sided C.I. for the difference in mean Classes Missed for the 2 genders</u>:

$(\mu_{Males} - \mu_{Females}) \in (\bar{x}_{Males} - \bar{x}_{Females}) \pm t_{\frac{\alpha}{2}}\sqrt{\frac{s^2_{Males}}{n_{Males}} + \frac{s^2_{Females}}{n_{Females}}}$ ϶: $P[t \geq t_{\frac{\alpha}{2}}] = \frac{\alpha}{2} \rightarrow P[t_{90} \geq t_{\frac{0.05}{2}}] = \frac{0.05}{2} \rightarrow t_{90,0.025} = 1.9867 \rightarrow (7.1 - 5.5283) \pm 1.9867\sqrt{\frac{60.908}{50} + \frac{35.523}{53}} \rightarrow -1.1588 \leq \mu_{Males} - \mu_{Females} \leq 4.302$

- <u>100(1- 0.05) % 2-sided C.I. for the ration of $\sigma^2_{Females}$ and σ^2_{Males}</u>:

$$\frac{s^2_{Males}}{s^2_{Females}}\frac{1}{F_{\alpha/2,n_{Males}-1,n_{Females}-1}} \le \frac{\sigma^2_{Males}}{\sigma^2_{Females}} \le \frac{s^2_{Males}}{s^2_{Females}} F_{\alpha/2,n_{Females}-1,n_{Males}-1}$$

$$\ni:\begin{cases} P\left[F_{n_{Males}-1,n_{Females}-1} \ge F_{\alpha/2,n_{Males}-1,n_{Females}-1}\right] = \dfrac{\alpha}{2} \to P\left[F_{49,52} \ge F_{0.025,49,52}\right] = 0.025 \to F_{0.025,49,52} = 1.7433 \\ P\left[F_{n_{Females}-1,n_{Males}-1} \ge F_{\alpha/2,n_{Females}-1,n_{Males}-1}\right] = \dfrac{\alpha}{2} \to P\left[F_{49,52} \ge F_{0.025,52,49}\right] = 0.025 \to F_{0.025,52,49} = 1.7515 \end{cases}$$

$$\to \frac{60.908}{35.523}\frac{1}{1.7433} \le \frac{\sigma^2_{Males}}{\sigma^2_{Females}} \le \frac{60.908}{35.523}1.7515 \to 0.9835 \le \frac{\sigma^2_{Males}}{\sigma^2_{Females}} \le 3.003$$

- <u>100(1-0.05)% Confidence Interval for Male Classes Missed population variance,</u>σ^2_{Males}:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2$$

$$\le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \ni:\begin{cases} P\left[\chi^2_{n-1} \ge \chi^2_{\alpha/2}\right] = \alpha/2 \to P[\chi^2_{49} \ge \chi^2_{0.025}] = 0.025 \to \chi^2_{0.025,49} = 70.2224 \\ P\left[\chi^2_{n-1} \ge \chi^2_{1-\alpha/2}\right] = 1-\alpha/2 \to P[\chi^2_{49} \ge \chi^2_{0.975}] = 0.975 \to \chi^2_{0.975,49} = 31.5549 \end{cases}$$

$$\to \frac{49 * 60.908}{70.2224} \le \sigma^2_{Males} \le \frac{49 * 60.908}{31.5549} \to 42.5 \le \sigma^2_{Males} \le 94.58$$

- <u>100(1-0.05)% Confidence Interval for Female GPA population variance,</u>$\sigma^2_{Females}$:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2$$

$$\le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \ni:\begin{cases} P\left[\chi^2_{n-1} \ge \chi^2_{\alpha/2}\right] = \alpha/2 \to P[\chi^2_{52} \ge \chi^2_{0.025}] = 0.025 \to \chi^2_{0.025,52} = 73.8098 \\ P\left[\chi^2_{n-1} \ge \chi^2_{1-\alpha/2}\right] = 1-\alpha/2 \to P[\chi^2_{52} \ge \chi^2_{0.975}] = 0.975 \to \chi^2_{0.975,52} = 33.9681 \end{cases}$$

$$\to \frac{52 * 35.523}{73.8098} \le \sigma^2_{Females} \le \frac{52 * 35.523}{33.9681} \to 25.0264 \le \sigma^2_{Females} \le 54.38$$

- <u>100(1-0.05)% Confidence Interval for Male Classes Missed population mean</u>: μ_{Males}

$$\bar{x}_{Males} - t_{n_{Males}-1,\frac{\alpha}{2}}\frac{s_{Males}}{\sqrt{n_{Males}}} \le \mu_{Males} \le \bar{x}_{Males} + t_{n_{Males}-1,\frac{\alpha}{2}}\frac{s_{Males}}{\sqrt{n_{Males}}} \ni: P\left[t_{n_{Males}-1} \ge t_{n_{Males}-1,\frac{\alpha}{2}}\right] = \frac{\alpha}{2}$$

$\rightarrow t_{n_{Males}-1,\frac{\alpha}{2}}$=$t_{49,0.025}$ $= 2.01$ → 7.1-2.01$\frac{\sqrt{60.908}}{\sqrt{50}} \leq \mu$ ≤7.1+2.01$\frac{\sqrt{60.908}}{\sqrt{50}}$ → $4.8815 \leq \mu_{Males}$ ≤9.3184

- <u>100(1-0.05)% Confidence Interval for Female GPA population mean</u>:$\mu_{Females}$

$$\bar{x}_{Females} - t_{n_{Females}-1,\frac{\alpha}{2}} \frac{s_{Males}}{\sqrt{n_{Females}}} \leq \mu_{Females}$$

$$\leq \bar{x}_{Females} + t_{n_{Females}-1,\frac{\alpha}{2}} \frac{s_{Females}}{\sqrt{n_{Females}}} \ni: P\left[t_{n_{Females}-1} \geq t_{n_{Females}-1,\frac{\alpha}{2}}\right] = \frac{\alpha}{2}$$

$\rightarrow t_{n_{Females}-1,\frac{\alpha}{2}}$=$t_{52,0.025}$ $= 2.007$ →5.5283-2.007$\frac{\sqrt{35.523}}{\sqrt{53}} \leq \mu_{Females}$ ≤5.5283+2.007$\frac{\sqrt{35.523}}{\sqrt{53}}$

$\rightarrow 3.8852 \leq \mu_{Females}$ ≤7.1714

Part 2: Analysis of GPA between males and females

As an offshoot to our analysis, we decided to examine who performs better academically, males or females?

$\bar{x}_{Males}$	$\bar{x}_{Females}$	s^2_{Males}	$s^2_{Females}$	n_{Males}	$n_{Females}$	α
3.1238	3.2128	0.224	0.2949	50	53	0.05

H_0: μ_{Males} - $\mu_{Females}$= 0

H_a :μ_{Males} - $\mu_{Females}$ ≠ 0

Since we are dealing with samples instead of actual populations, we have to work with *t-test*. Therefore, rejectH_0 if $\begin{cases} t_0 \leq -t_{\frac{\alpha}{2}} \\ t_0 \geq t_{\frac{\alpha}{2}} \end{cases}$such that $t_0 = \frac{\bar{x}^2_{Males} - \bar{x}^2_{Females} - D}{\sqrt{\frac{s^2_{Males}}{n_{Males}} + \frac{s^2_{Females}}{n_{Females}}}}$. Or, reject H_0 if$\begin{cases} t_0 \geq 0: 2P[t \geq t_0] < \alpha \\ t_0 \leq 0: 2P[t \geq -t_0] < \alpha \end{cases}$

э: $P[t \geq t_{\alpha}] = \alpha, P\left[t \geq t_{\alpha/2}\right] = \frac{\alpha}{2}, \alpha = 0.05$

Therefore, t_0 = $\frac{3.1238^2 - 3.2128^2 - 0}{\sqrt{\frac{0.224}{50} + \frac{0.2949}{53}}}$ = -5.627. *d.f.* = $\frac{\left[\frac{s^2_{Males}}{n_{Males}} + \frac{s^2_{Females}}{n_{Females}}\right]^2}{\frac{\left[s^2_{Males}/n_{Males}\right]^2}{n_{Males} - 1} + \frac{\left[s^2_{Females}/n_{Females}\right]^2}{n_{Females} - 1}} = \frac{\left[\frac{0.224}{50} + \frac{0.2949}{53}\right]^2}{\frac{\left[0.224/50\right]^2}{50 - 1} + \frac{\left[0.2949/53\right]^2}{53 - 1}} \approx 100$

$P\left[t \geq t_{\frac{\alpha}{2}}\right] = \frac{\alpha}{2} \rightarrow P\left[t \geq t_{\frac{0.05}{2}}\right] = \frac{0.05}{2} \rightarrow t_{\frac{0.05}{2}} = t_{100,0.025} = 1.984$

- Critical value approach:
- $\begin{cases} -5.627 \leq -1.984 \; Reject \\ -5.627 \geq 1.984 \; Do \; not \; reject \end{cases}$ →*Reject the H_0 and conclude that there is a difference between mean Male GPA and mean Female GPA. Therefore, we now need to test who performs better, Males or Females*
- P-value approach: Since, t_0 = -5.627 < 0, we use the following method

  $2P[t \geq -t_0] < \alpha \rightarrow 2P[t_{100} \geq -5.627] = 2(1 - P[t_{100} \leq 5.627]) \approx 2(1 - 1) = 0 < 0.05$

  → *Reject the H_0 and conclude that there is a difference between mean Male GPA and mean Female GPA.*

H_0: μ_{Males} - $\mu_{Females} \leq 0$

H_a : μ_{Males} - $\mu_{Females} > 0$

Reject H_0 if $t_0 \geq t_{\alpha}$ or $P[t \geq t_0] < \alpha$. t_0 = $\frac{3.1238 - 3.2128 - 0}{\sqrt{\frac{0.224}{50} + \frac{0.2949}{53}}}$ = -0.888. *d.f.* = $\frac{\left[\frac{0.224}{50} + \frac{0.2949}{53}\right]^2}{\frac{\left[0.224/50\right]^2}{50 - 1} + \frac{\left[0.2949/53\right]^2}{53 - 1}} \approx 100$.

$$P[t \geq t_{\alpha}] = \alpha \rightarrow P[t_{100} \geq t_{0.05}] = 0.05 \rightarrow t_{0.05} = 1.645$$

- Critical value approach: $t_0 \geq t_{\alpha}$ →-0.888 ! ≥ 1.645. Therefore, do not reject the H_0 and conclude that on average females perform better academically than males.
- P-value approach: $P[t \geq t_0] \rightarrow P[t_{100} \geq -0.888] = 0.188$ *!< 0.05*. Therefore, do not reject the hypothesis and conclude that on average females perform better academically than males.

Additionally, we can test whether females' GPA is spread out more around the mean compared to that of male sample. To do so we apply *F-test* for testing the ratio of 2 population variances.

$H_0 : \sigma^2_{Females} \leq \sigma^2_{Males}$

$H_a : \sigma^2_{Females} > \sigma^2_{Males}$

Reject H_0 if $F_0 \geq F_\alpha$ or $P[F \geq F_0] < \alpha$. $F_0 = \frac{s^2_{Females}}{s^2_{Males}}$ э: *d.f. of numerator* = 52 and *d.f. of denominator* = 49.

$F_0 = \frac{0.2949}{0.224} = 1.3165$.

- Critical-value approach: $P[F \geq F_\alpha] = \alpha \rightarrow P[F_{52,49} \geq F_{0.05}] = 0.05 \rightarrow F_{0.05} \approx 1.6 \rightarrow 1.3165 \,!\!\geq 1.6$

  Therefore, do not reject H_0 and conclude that males' GPA is more spread out around the mean compared to females.

- P-value approach: $P[F \geq F_0] = P[F_{52,49} \geq 1.3165]=0.17 \,!\!< 0.05$. Therefore, do not reject H_0 and conclude that males' GPA is more spread out around the mean compared to females.

Now, we can proceed with the analysis of 95% Confidence intervals:

- Explanation regarding the examination of 95% Confidence intervals is provided at the end of the 95 C.I. construction
- 100(1- 0.05) % 2-sided C.I. for the difference in mean GPA for the 2 genders:

  $(\mu_{Males} - \mu_{Females}) \,\epsilon\, (\bar{x}_{Males} - \bar{x}_{Females}) \pm t_{\frac{\alpha}{2}}\sqrt{\frac{s^2_{Males}}{n_{Males}} + \frac{s^2_{Females}}{n_{Females}}}$ э: $P[t \geq t_{\frac{\alpha}{2}}]=\frac{\alpha}{2} \rightarrow P[t_{100} \geq t_{\frac{0.05}{2}}]=\frac{0.05}{2} \rightarrow t_{100,0.025}=1.984 \rightarrow (3.1238-3.2128) \pm 1.984\sqrt{\frac{0.224}{50} + \frac{0.2949}{53}} \rightarrow -0.2878 \leq \mu_{Males} - \mu_{Females} \leq 0.1098$

- 100(1- 0.05) % 2-sided C.I. for the ration of $\sigma^2_{Females}$ and σ^2_{Males}:

$$\frac{s^2_{Females}}{s^2_{Males}}\frac{1}{F_{\alpha/2, n_{Females}-1, n_{Males}-1}} \leq \frac{\sigma^2_{Females}}{\sigma^2_{Males}} \leq \frac{s^2_{Females}}{s^2_{Males}} F_{\alpha/2, n_{Males}-1, n_{Females}-1}$$

$$э: \begin{cases} P\left[F_{n_{Females}-1, n_{Males}-1} \geq F_{\alpha/2, n_{Females}-1, n_{Males}-1}\right] = \frac{\alpha}{2} \rightarrow P\left[F_{52,49} \geq F_{0.025,52,49}\right] = 0.025 \rightarrow F_{0.025,52,49} = 1.7515 \\ P\left[F_{n_{males}-1, n_{Females}-1} \geq F_{\alpha/2, n_{Males}-1, n_{Females}-1}\right] = \frac{\alpha}{2} \rightarrow P\left[F_{49,52} \geq F_{0.025,49,52}\right] = 0.025 \rightarrow F_{0.025,49,52} = 1.7433 \end{cases}$$

$$\rightarrow \frac{0.2949}{0.224}\frac{1}{1.7515} \leq \frac{\sigma^2_{Females}}{\sigma^2_{Males}} \leq \frac{0.2949}{0.224} 1.7433 \rightarrow 0.7516 \leq \frac{\sigma^2_{Females}}{\sigma^2_{Males}} \leq 2.2951$$

- 100(1-0.05)% Confidence Interval for Male GPA population variance,σ^2_{Males}:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2$$

$$\le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \ni: \begin{cases} P\left[\chi^2_{n-1} \ge \chi^2_{\alpha/2}\right] = \alpha/2 \to P[\chi^2_{49} \ge \chi^2_{0.025}] = 0.025 \to \chi^2_{0.025,49} = 70.2224 \\ P\left[\chi^2_{n-1} \ge \chi^2_{1-\alpha/2}\right] = 1-\alpha/2 \to P[\chi^2_{49} \ge \chi^2_{0.975}] = 0.975 \to \chi^2_{0.975,49} = 31.5549 \end{cases}$$

$$\to \frac{49 * 0.224}{70.2224} \le \sigma^2_{Males} \le \frac{49 * 0.224}{31.5549} \to 0.1563 \le \sigma^2_{Males} \le 0.3478$$

- 100(1-0.05)% Confidence Interval for Female GPA population variance,$\sigma^2_{Females}$:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2$$

$$\le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \ni: \begin{cases} P\left[\chi^2_{n-1} \ge \chi^2_{\alpha/2}\right] = \alpha/2 \to P[\chi^2_{52} \ge \chi^2_{0.025}] = 0.025 \to \chi^2_{0.025,52} = 73.8098 \\ P\left[\chi^2_{n-1} \ge \chi^2_{1-\alpha/2}\right] = 1-\alpha/2 \to P[\chi^2_{52} \ge \chi^2_{0.975}] = 0.975 \to \chi^2_{0.975,52} = 33.9681 \end{cases}$$

$$\to \frac{52 * 0.2949}{73.8098} \le \sigma^2_{Females} \le \frac{52 * 0.2949}{33.9681} \to 0.2077 \le \sigma^2_{Females} \le 0.4514$$

- 100(1-0.05)% Confidence Interval for Male GPA population mean: μ_{Males}

$$\bar{x}_{Males} - t_{n_{Males}-1,\frac{\alpha}{2}} \frac{s_{Males}}{\sqrt{n_{Males}}} \le \mu_{Males} \le \bar{x}_{Males} + t_{n_{Males}-1,\frac{\alpha}{2}} \frac{s_{Males}}{\sqrt{n_{Males}}} \ni: P\left[t_{n_{Males}-1} \ge t_{n_{Males}-1,\frac{\alpha}{2}}\right] = \frac{\alpha}{2}$$

$\to t_{n_{Males}-1,\frac{\alpha}{2}} = t_{49,0.025} = 2.01 \to$ 3.1238-2.01$\frac{\sqrt{0.224}}{\sqrt{50}} \le \mu \le$3.1238+2.01$\frac{\sqrt{0.224}}{\sqrt{50}} \to 2.9892 \le$ $\mu_{Males} \le$3.2583

- 100(1-0.05)% Confidence Interval for Female GPA population mean:$\mu_{Females}$

$$\bar{x}_{Females} - t_{n_{Females}-1,\frac{\alpha}{2}} \frac{s_{Males}}{\sqrt{n_{Females}}} \le \mu_{Females}$$

$$\le \bar{x}_{Females} + t_{n_{Females}-1,\frac{\alpha}{2}} \frac{s_{Females}}{\sqrt{n_{Females}}} \ni: P\left[t_{n_{Females}-1} \ge t_{n_{Females}-1,\frac{\alpha}{2}}\right] = \frac{\alpha}{2}$$

$\to t_{n_{Females}-1,\frac{\alpha}{2}} = t_{52,0.025} = 2.007 \to$ 3.2128-2.007$\frac{\sqrt{0.2949}}{\sqrt{53}} \le \mu_{Females} \le$3.2128+2.007$\frac{\sqrt{0.2949}}{\sqrt{53}}$

$\to 3.063 \le \mu_{Females} \le$3.3625

References

Dayioglu, Meltem, and Serap Türüt-Asik. "Gender Differences in Academic Performance in a Large Public University in Turkey."Higher education 53.2 (2007): 255-77. ProQuest. Web. 31 Mar. 2015.

Voyer, Daniel, and Susan D. Voyer. "Gender Differences in Scholastic Achievement: A Meta-Analysis." Psychological bulletin 140.4 (2014): 1174-204. ProQuest. Web. 31 Mar. 2015.

Decore, Anne Marie. "Vive La Difference: A Comparison of Male-Female Academic Performance." Canadian Journal of Higher Education 14.3 (1984): 35-58. ProQuest. Web. 31 Mar. 2015.

E., Cumhur. "Factors Affecting Grade Point Average of University Students." *Empirical Economics Letters* 6.5 (2007): 359. *Academia.edu*. Web. 31 Mar. 2015. <http://www.academia.edu/6569423/Factors_Affecting_Grade_Point_Average_of_University_Students>.

Birch, Elisa Rose, and Paul W. Miller. "Student Outcomes at University in Australia: A Quantile Regression Approach." Australian Economic Papers 45.1 (2006): 1-17. ProQuest. Web. 31 Mar. 2015.