Determining Factors of GPA

Among full-time undergraduate Stevens students,
number of credits and which school a student belongs to
were the most important factors in determining GPA.

Natalie Arndt
Allison Mucha
MA 331
12/19/07

We pledge our honor that we have abided by the Stevens Honor System.

Introduction

There is much debate surrounding the effects of various factors on a student's performance in the academic world. Sophisticated studies take in dozens of variables and years' worth of data in attempt to predict how successful a particular student will be during his or her academic career. Determining which factors most greatly affect academic performance would not only allow individuals to predict future GPAs or grades, but could possibly help students make lifestyle choices that facilitate success.

This study aims to deduce which factors most significantly impact a full-time Stevens student's GPA. Other points of interest included making use of methods and analytic techniques discussed in class, as well as observing any significant differences between engineering and science students at Stevens. The ultimate goal of this study would be to be able to predict a Stevens student's GPA based on the factors that are determined to be significant.

Data Collection

Initially, the data being considered was notably less focused than it would become later in the study. Factors such as average hours of non-studious work per week, average hours of sleep per night, or which SAT score was higher for the student in question were proposed. Ultimately, these factors were dropped in favor of variables that did not require loose estimation or guess-work. The list of factors was refined to the following: gender, major, number of semesters at Stevens, credit load per semester, the corresponding GPA for that semester, cumulative number of credits, and cumulative GPA.

The data was gathered by reaching out to several subsets of the general student body. These students were sent a survey, and were asked to return it voluntarily with full anonymity. This survey can be seen in *Figure 1*. Only full-time (at least 12 credits), undergraduate Stevens students were considered in the study. Recent alumni who satisfied these conditions during their time at Stevens were also considered.

Due to the nature of the study, there were numerous lurking variables that had to be accounted for. These factors include: the influence of extracurricular activities on a student's performance, changes in curriculum from year to year, personal issues, medical problems, any stressful situation that could impact a student's ability to work, and differences in professors and grading. Perhaps in a more sophisticated study with a longer duration, an appropriate method for removing these lurking variables could be developed. For example, the study could focus on students with similar extracurricular involvement, or eliminate students who have had medical problems or similar traumas. In favor of receiving the greatest number of usable responses, the survey was not designed to do so.

Data Preparation

After a three week period, the data collection process was terminated. The completed data table can be viewed in *Appendix 1*. Combined, 28 students participated in the study, which yielded 154 semester's worth of data for analysis. Among the participants were 18 males and 10 females, with 19 engineering majors, 8 science majors, and 1 art major. One noteworthy fact was that the breakdown of these categorical variables generally reflected the ratios of the entire Stevens student body, providing a roughly stratified sample. The range of GPA for one semester was 2.317 to 4.000. The number of credits taken in any one semester ranged from 12.0 to 25.5. The number of cumulative credits taken thus far ranged from 33.0 to 177.0.

After the data was collected, a number of measures were taken to make it easier to manipulate and analyze. All students' names were removed from the corresponding data, and replaced with an identification number ranging from 1 to 28. Cumulative data was entered as semester number zero, to distinguish it from individual semesters. The students' primary majors were used to create the "school" category, taking one of the two values "engineering" or "science." Art majors were not represented significantly enough to be considered in the study. The number of credits per semester was used to create the category referred to as "load." Beginning with 12 credits and increasing in increments of 3 (the equivalent of one Stevens class), load categories A through E were thus created.

Data Analysis

A primary analysis was performed to determine the normality of the data that had been collected. As evidenced by *Figure 2a*, the distribution of the number of credits per semester was fairly normal, with a slight tail to the right. *Figure 2b* shows that the normal Q-Q plot is strongly linear. The distribution of GPA, however, was not as regular. *Figure 3a* shows a strongly skewed distribution with a notable left tail. The normal Q-Q plot shown in *Figure 3b* does not suggest linearity, implying that the distribution of GPA is not normal for the data collected.

The GPA by semester data was combined, and a linear regression performed. The data with the fitted regression line can be seen in *Figure 4*. The slope of the line is 0.01799, or approximately 0.02, which is statistically significant and meaningful in a school where GPAs are carried out to three decimal places. Unfortunately, the R^2 value for this regression was only 0.01623, showing the need for further analysis and the search for a better fitting model. The cumulative GPA data was then combined and fitted with its own linear regression, together seen in *Figure 5*. The slope of this line was statistically zero, with an R^2 value of less than 0.001. As such, this data was no longer considered, as a zero slope implies no relation between cumulative credits taken and cumulative GPA.

Residual plots of the GPA by semester data were then considered. The residuals did appear to be centered about zero, with apparent random scatter about that center, as seen in *Figure 6a*. However, the displacements above and below the zero line were not equal, meaning the data was not normally distributed, and that the relationship between the explanatory and response variables was not necessarily linear. This is also evidenced by the non-linearity of the normal Q-Q plot, which can be seen in *Figure 6b*. Attempts to find a better-fitting model were made using a box-cox transformation. However, no better model could be found.

Comparisons

The GPA breakdown according to gender was examined first. It was found, as can be seen in *Figure 7a*, that the mean GPA earned in a given semester, regardless of credit load, was statistically equivalent between males and females. When cumulative data was considered, this mean was also statistically equivalent, as is visible in *Figure 7b*. It should also be noted that the minimum and maximum values of these ranges were consistent with each other when comparing males to females.

The number of credits taken in a given semester broken down by school was then considered. A minimum credit load of 12 credits per semester was imposed, as discussed previously. However, as hypothesized, the maximum credit load differed greatly between the two considered schools. However, despite this difference in range, the average credit load for both science and engineering students was statistically equivalent, as seen by *Figure 8a*. It was also interesting to find that from the data gathered, the average cumulative number of credits taken by engineering and science students was essentially equivalent, as seen in *Figure 8b*, even though the range of engineering cumulative credits spread farther both above and below the mean.

Lastly, the GPA earned based on the created variable of credit load was studied. An interesting trend was observed, as seen by *Figure 9*. The highest mean GPAs correspond to load A (12-15 credits) and load D (21-24 credits). The lowest mean GPAs correspond to load B (15-18 credits) and load E (24-27 credits). It is to be noted that load C (18-21 credits) lies in the middle of the range for average GPA, while containing the average credit load for both engineering and science students.

The trends viewed to the left and right of Load C find explanation in a real-world context. A student taking few credits (load A) theoretically has more time to devote to his or her studies, and will therefore do well. As a student takes on more credits (load B), he or she may struggle with the additional work load. A student taking an average number of credits (load C) earns an average GPA. Students who are especially talented or hard-working taking more credits than average (load D) will still succeed. But there comes a time when a student takes on too many credits (load E), and the GPA suffers accordingly.

Multivariable Regressions

A stepwise regression analysis was performed on the data, with GPA as the response variable, and credits, school, gender, and semester as explanatory variables. Both forward and stepwise regression returned the fact that number of credits and school are the most important variables, as were noted earlier during comparisons. This confirmed the notion that gender and semester were statistically insignificant in this study. A summary of the regression returned that credits were significant within 10% significance, while an ANOVA analysis of the same regression returned that school is significant within the same 10% significance. Provided the high variability of the data, it will be maintained that these factors are significant.

Observations and Conclusions

Through numerous analytical methods, it was determined that school and credit load were the most important factors in determining a student's GPA. If a base GPA of 2.96 is taken, values can be added based on the student's school and his or her credit load to approximate the GPA he or she will earn for that semester. If the student is an engineering major, add 0.09 to the base GPA value; if the student is a science major, add 0.27. For each credit taken during the semester in question, add 0.02 to the base value plus the value associated with school. This can be described by the following equation:

$$\boxed{\text{GPA} = 2.96 + 0.02*\text{credits} + V_s}$$

where V_s is the value associated with the student's school. This equation has an R^2 value of 0.05626, which while low, is certainly an improvement from the initial linear fit of GPA vs. credits in one semester, which had an R^2 value of 0.01623, as previously mentioned.

Overall, it can be concluded that the science majors represented by the study have an average GPA 0.18 points higher than their engineering major counterparts. It was also determined that credit load does have a significant effect on the GPA earned by a student in a particular semester, though the effect is relatively low in comparison.

Recommendations

If this study were to be continued or performed again, there are a number of alterations that could be made that would be beneficial. Of course, drawing data from a significantly larger sample of the student population would improve the distributions of the data collected, thereby strengthening the integrity of the results. Appropriate methodology could be found to remove the effect of lurking variables, such as only considering students who reported no major stressful situations or circumstances outside of their scholarly activities during a given semester. Also, the study could be refined to allow focus on a specific area, rather than manipulating several variables at once without knowledge of their impact on the study.

Overall, the study provided insight into the important factors related to a Stevens student's GPA, as was the initial objective. A definite difference between science and engineering students was observed and determined, as was originally hypothesized. Multiple statistical techniques were used to arrive at these conclusions, and generally solidified the importance and validity of such analytical methods to real-world applications. It would be interesting to carry out this study on the entire Stevens population, and see how well this single study reflected the results for the entire campus. The next step would be to extend the study to other colleges and compare the patterns among different schools. It might also be interesting to analyze some of the variables that were initially considered for this study, especially those variables over which a student has control. By doing so, students could learn which lifestyle choices might help them succeed in their academic endeavors.

Appendix 1: Data

obs	Gender	major	school	sem	credits	gpa
1	Male	Civil Engineering	E	0	50.0	3.732
1	Male	Civil Engineering	E	1	17.5	3.536
1	Male	Civil Engineering	E	2	20.5	3.847
2	Male	Engineering Management	E	0	101.0	3.947
2	Male	Engineering Management	E	1	17.0	3.938
2	Male	Engineering Management	E	2	18.0	4.000
2	Male	Engineering Management	E	3	18.5	3.829
2	Male	Engineering Management	E	4	17.5	4.000
2	Male	Engineering Management	E	5	20.0	4.000
3	Male	Engineering Physics	E	0	40.0	3.207
3	Male	Engineering Physics	E	1	15.0	3.177
3	Male	Engineering Physics	E	2	22.0	3.263
4	Male	Computer Science	S	0	71.0	3.455
4	Male	Computer Science	S	1	18.0	3.260
4	Male	Computer Science	S	2	14.0	3.240
4	Male	Computer Science	S	3	19.0	3.421
4	Male	Computer Science	S	4	20.0	3.791
5	Female	Biomedical Engineering	E	0	131.0	3.841
5	Female	Biomedical Engineering	E	1	19.0	4.000
5	Female	Biomedical Engineering	E	2	25.0	4.000
5	Female	Biomedical Engineering	E	3	23.5	3.630
5	Female	Biomedical Engineering	E	4	23.0	3.900
5	Female	Biomedical Engineering	E	5	22.0	3.747
5	Female	Biomedical Engineering	E	6	17.5	3.772
6	Female	Civil Engineering	E	0	59.0	2.870
6	Female	Civil Engineering	E	1	16.0	3.044
6	Female	Civil Engineering	E	2	17.0	2.466
6	Female	Civil Engineering	E	3	20.0	3.433
7	Female	Biomedical Engineering	E	0	48.0	3.956
7	Female	Biomedical Engineering	E	1	16.0	3.938
7	Female	Biomedical Engineering	E	2	23.0	3.951
8	Male	Civil Engineering	E	0	177.0	3.711
8	Male	Civil Engineering	E	1	20.0	3.648
8	Male	Civil Engineering	E	2	21.0	3.176
8	Male	Civil Engineering	E	3	19.5	3.622
8	Male	Civil Engineering	E	4	19.0	3.896
8	Male	Civil Engineering	E	5	12.0	4.000
8	Male	Civil Engineering	E	6	21.5	3.753
8	Male	Civil Engineering	E	7	21.0	3.840
8	Male	Civil Engineering	E	8	18.0	3.557
8	Male	Civil Engineering	E	9	17.0	3.882
9	Male	Computer Science	S	0	52.0	4.000
9	Male	Computer Science	S	1	17.0	4.000
9	Male	Computer Science	S	2	19.0	4.000
9	Male	Computer Science	S	3	19.0	4.000

10	Male	Mechanical Engineering	E	0	125.0	2.842
10	Male	Mechanical Engineering	E	1	18.0	2.740
10	Male	Mechanical Engineering	E	2	20.0	2.929
10	Male	Mechanical Engineering	E	3	15.5	3.129
10	Male	Mechanical Engineering	E	4	18.5	3.225
10	Male	Mechanical Engineering	E	5	18.0	2.317
10	Male	Mechanical Engineering	E	6	19.0	2.982
11	Male	Engineering (Undecided)	E	0	33.0	3.544
11	Male	Engineering (Undecided)	E	1	16.0	3.599
11	Male	Engineering (Undecided)	E	2	17.0	3.490
12	Male	Mechanical Engineering	E	0	47.0	3.830
12	Male	Mechanical Engineering	E	1	16.0	3.688
12	Male	Mechanical Engineering	E	2	18.0	3.944
13	Female	Mechanical Engineering	E	0	126.5	3.968
13	Female	Mechanical Engineering	E	1	20.0	4.000
13	Female	Mechanical Engineering	E	2	20.5	4.000
13	Female	Mechanical Engineering	E	3	22.0	4.000
13	Female	Mechanical Engineering	E	4	22.0	4.000
13	Female	Mechanical Engineering	E	5	23.0	3.895
13	Female	Mechanical Engineering	E	6	19.0	4.000
14	Male	Mathematics	S	0	112.5	3.911
14	Male	Mathematics	S	1	18.0	3.778
14	Male	Mathematics	S	2	18.0	4.000
14	Male	Mathematics	S	3	19.5	3.893
14	Male	Mathematics	S	4	20.5	3.952
14	Male	Mathematics	S	5	12.5	3.911
15	Female	Computer Engineering	E	0	146.0	3.033
15	Female	Computer Engineering	E	1	15.0	3.418
15	Female	Computer Engineering	E	2	16.0	2.584
15	Female	Computer Engineering	E	3	18.5	2.432
15	Female	Computer Engineering	E	4	18.5	2.756
15	Female	Computer Engineering	E	5	12.0	2.665
15	Female	Computer Engineering	E	6	16.0	3.732
15	Female	Computer Engineering	E	7	18.0	3.647
16	Male	Computer Engineering	E	0	130.0	3.355
16	Male	Computer Engineering	E	1	16.5	3.010
16	Male	Computer Engineering	E	2	20.5	3.215
16	Male	Computer Engineering	E	3	18.5	3.715
16	Male	Computer Engineering	E	4	25.0	3.121
16	Male	Computer Engineering	E	5	24.5	3.519
17	Male	Mechanical Engineering	E	0	75.0	3.931
17	Male	Mechanical Engineering	E	1	15.0	3.934
17	Male	Mechanical Engineering	E	2	19.0	3.882
17	Male	Mechanical Engineering	E	3	20.0	4.000
17	Male	Mechanical Engineering	E	4	18.0	3.945
18	Female	Mathematics	S	0	101.0	3.413
18	Female	Mathematics	S	1	18.0	3.357
18	Female	Mathematics	S	2	17.0	3.195
18	Female	Mathematics	S	3	15.5	3.291

18	Female	Mathematics	S	4	20.5	3.707
19	Male	Computer Engineering	E	0	35.0	2.991
19	Male	Computer Engineering	E	1	15.0	2.991
19	Male	Computer Engineering	E	2	17.0	2.736
20	Male	Computer Science	S	0	69.0	3.884
20	Male	Computer Science	S	1	15.0	3.866
20	Male	Computer Science	S	2	19.0	3.948
20	Male	Computer Science	S	3	13.0	3.769
20	Male	Computer Science	S	4	13.0	3.845
21	Male	Computer Science	S	0	96.5	3.501
21	Male	Computer Science	S	1	18.0	3.391
21	Male	Computer Science	S	2	21.0	3.117
21	Male	Computer Science	S	3	20.5	3.555
21	Male	Computer Science	S	4	12.0	3.918
21	Male	Computer Science	S	5	19.0	3.632
22	Female	Mathematics	S	0	129.0	3.313
22	Female	Mathematics	S	1	20.0	3.755
22	Female	Mathematics	S	2	19.0	3.335
22	Female	Mathematics	S	3	16.0	2.876
22	Female	Mathematics	S	4	12.0	3.455
22	Female	Mathematics	S	5	16.0	3.625
22	Female	Mathematics	S	6	16.0	3.063
23	Female	Art & Technology	A	0	60.0	3.334
23	Female	Art & Technology	A	1	13.0	3.077
23	Female	Art & Technology	A	2	16.0	3.312
23	Female	Art & Technology	A	3	16.0	3.563
23	Female	Art & Technology	A	4	15.0	3.331
24	Male	Civil Engineering	E	0	115.0	3.118
24	Male	Civil Engineering	E	1	18	3.520
24	Male	Civil Engineering	E	2	21	2.348
24	Male	Civil Engineering	E	3	18.5	2.666
24	Male	Civil Engineering	E	4	17.5	3.086
24	Male	Civil Engineering	E	5	18	3.502
25	Male	Chemical Engineering	E	0	169.0	3.645
25	Male	Chemical Engineering	E	1	20.0	3.896
25	Male	Chemical Engineering	E	2	22.5	3.938
25	Male	Chemical Engineering	E	3	14.0	3.571
25	Male	Chemical Engineering	E	4	22.0	3.777
25	Male	Chemical Engineering	E	5	25.5	3.518
25	Male	Chemical Engineering	E	6	20.5	3.496
25	Male	Chemical Engineering	E	7	20.5	3.414
26	Female	Electrical Engineering	E	0	69.0	3.592
26	Female	Electrical Engineering	E	1	15.0	3.222
26	Female	Electrical Engineering	E	2	14.0	3.668
26	Female	Electrical Engineering	E	3	20.0	3.651
26	Female	Electrical Engineering	E	4	20.0	3.773
27	Female	Mathematics	S	0	99.5	3.815
27	Female	Mathematics	S	1	18.0	3.704
27	Female	Mathematics	S	2	20.5	3.812
27	Female	Mathematics	S	3	13.0	3.924

27	Female	Mathematics	S	4	14.5	3.654
27	Female	Mathematics	S	5	13.5	3.851
27	Female	Mathematics	S	6	16.5	3.936
28	Male	Computer Engineering	E	0	130.0	3.537
28	Male	Computer Engineering	E	1	19.0	3.537
28	Male	Computer Engineering	E	2	23.0	3.634
28	Male	Computer Engineering	E	3	18.5	3.601
28	Male	Computer Engineering	E	4	21.0	3.334
28	Male	Computer Engineering	E	5	20.5	3.398
28	Male	Computer Engineering	E	6	16.0	3.688

Appendix 2: Figures

Figure 1

Gender: ____________		Major: _________________
Semester	Credits	GPA for Semester
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
Total Credits Earned:		Cumulative GPA:

Figure 2a

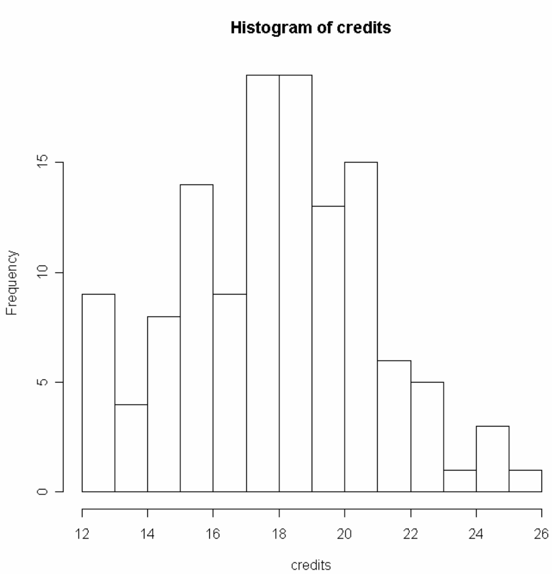


Figure 2b

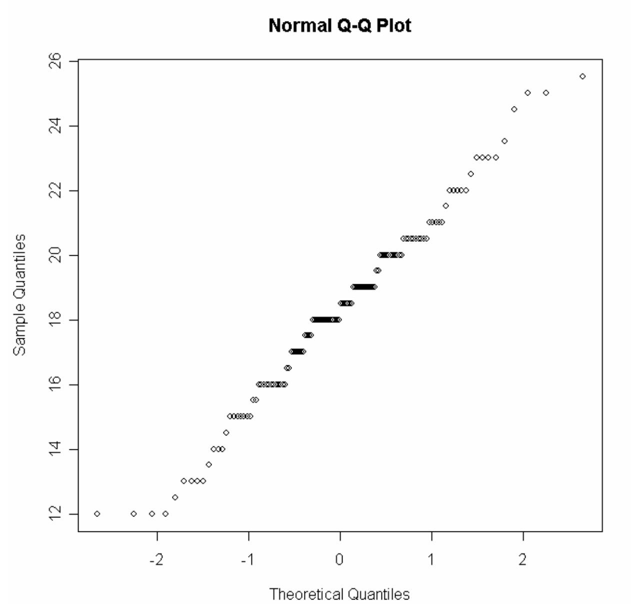


Figure 3a

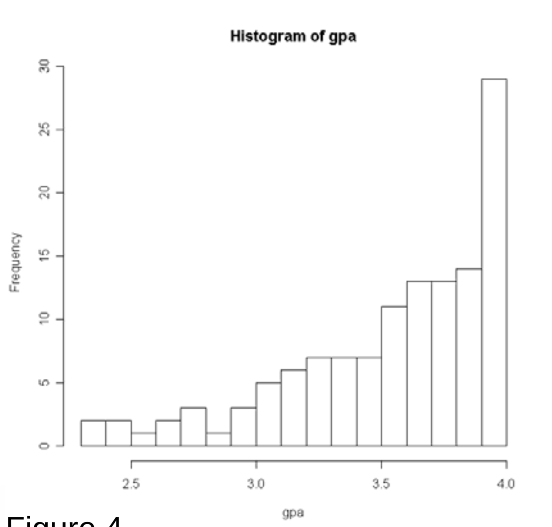


Figure 3b

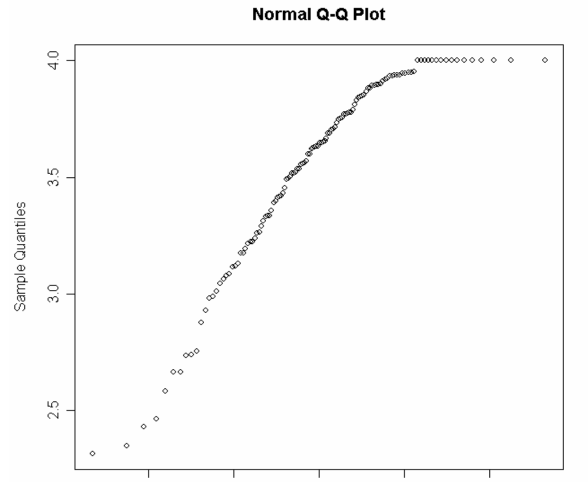


Figure 4

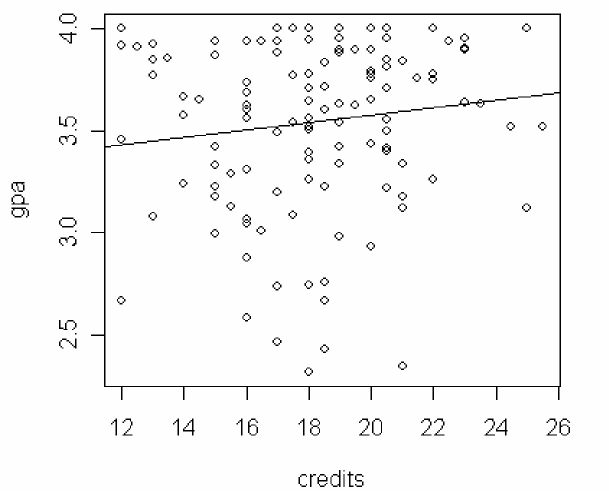


Figure 5

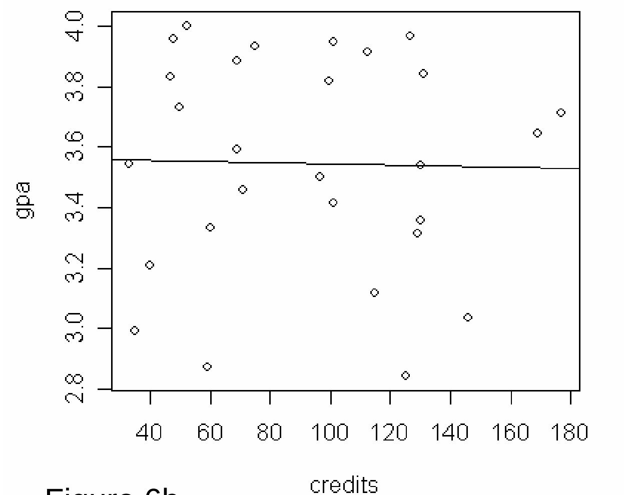


Figure 6a

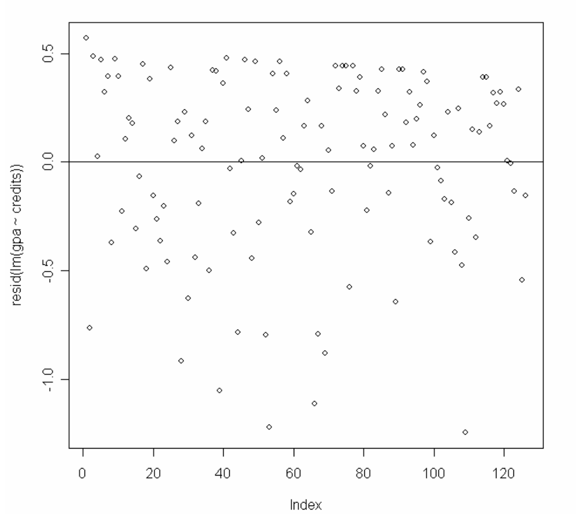


Figure 6b

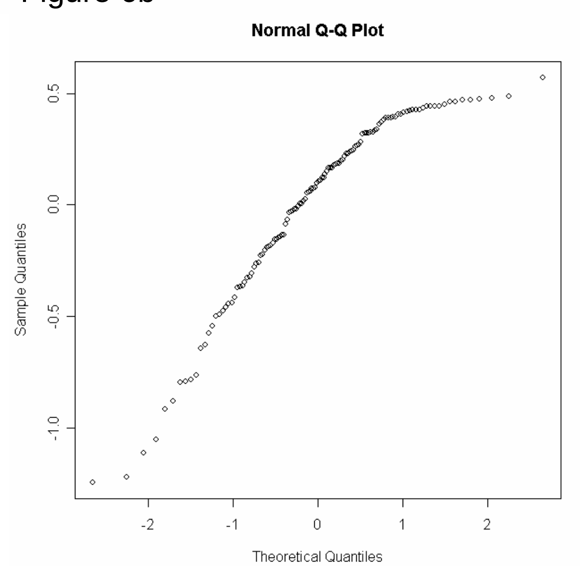


Figure 7a

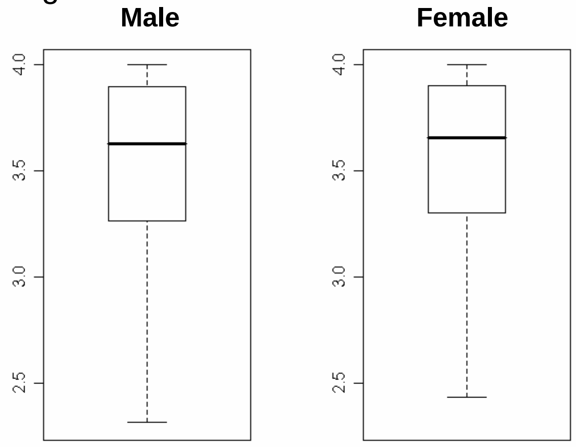


Figure 7b

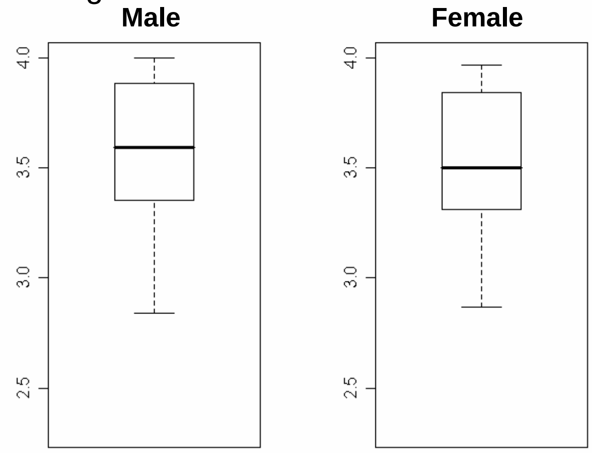


Figure 8a

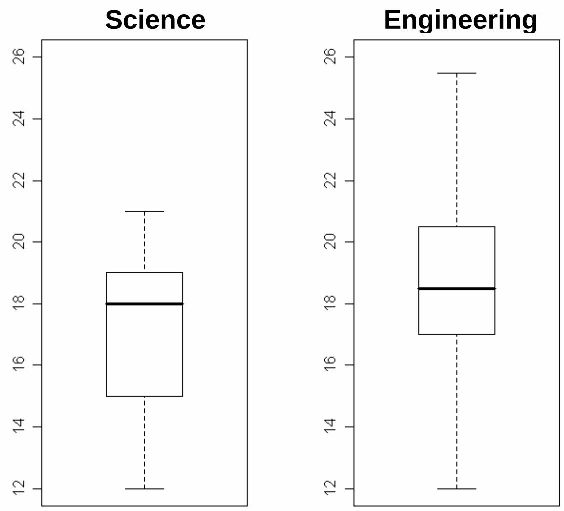


Figure 8b

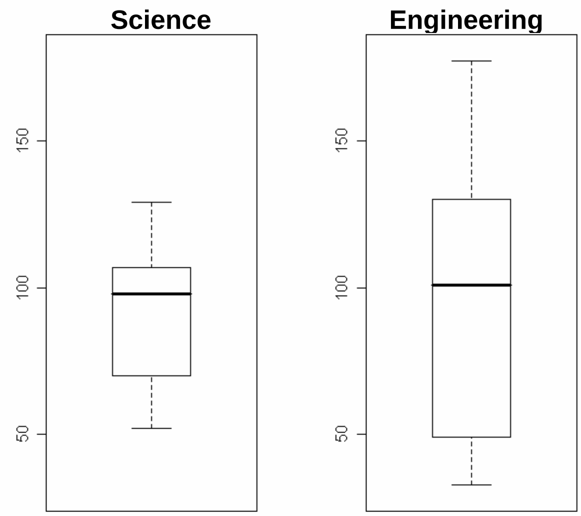
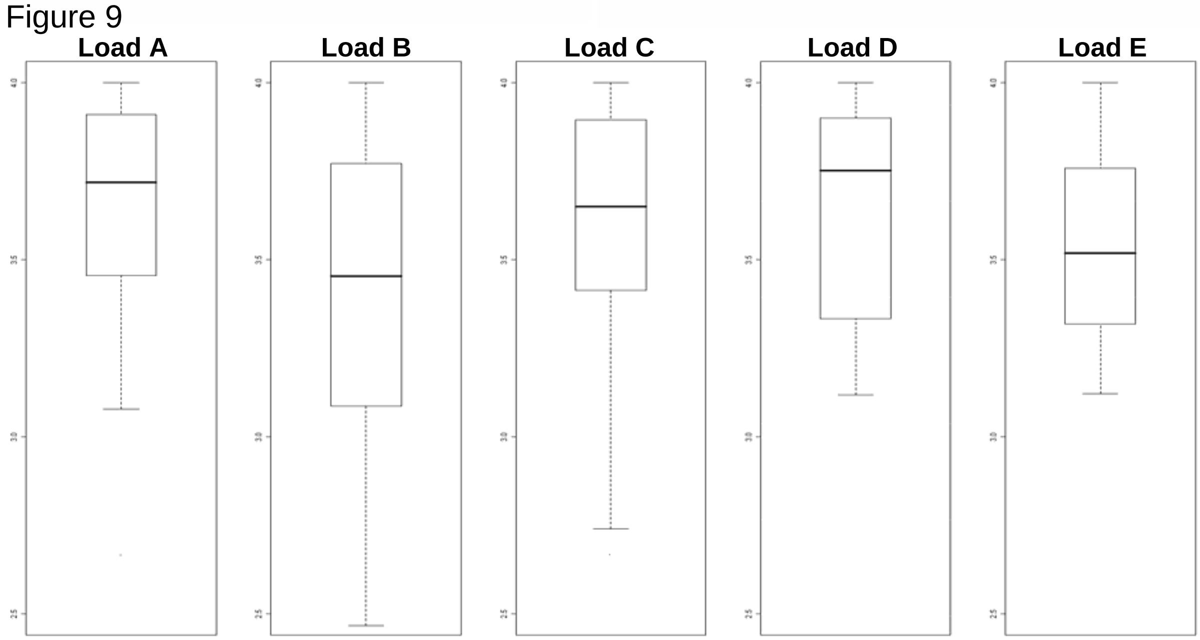


Figure 9



Appendix 3: Code

```
> stepwise = step(lm(gpa~credits+school+gender+sem),direction="both")
Start:  AIC=-217.77
gpa ~ credits + school + gender + sem
          Df Sum of Sq      RSS      AIC
- gender   1     0.017   20.359 -219.667
- sem      1     0.198   20.541 -218.549
<none>                   20.342 -217.772
- credits  1     0.524   20.866 -216.568
- school   2     0.907   21.250 -216.273
Step:  AIC=-219.67
gpa ~ credits + school + sem
          Df Sum of Sq      RSS      AIC
- sem      1     0.194   20.553 -220.472
<none>                   20.359 -219.667
- credits  1     0.530   20.889 -218.427
- school   2     0.905   21.264 -218.189
+ gender   1     0.017   20.342 -217.772
Step:  AIC=-220.47
gpa ~ credits + school
          Df Sum of Sq      RSS      AIC
<none>                   20.553 -220.472
+ sem      1     0.194   20.359 -219.667
- school   2     0.872   21.425 -219.238
- credits  1     0.556   21.109 -219.108
+ gender   1     0.013   20.541 -218.549
Call:
lm(formula = gpa ~ credits + school)
Coefficients:
(Intercept)      credits      schoolE      schoolS
    2.95972      0.02407      0.09478      0.27379

 > summary(stepwise)
Call:
lm(formula = gpa ~ credits + school)
Residuals:
    Min      1Q  Median      3Q     Max
-1.2119 -0.2735  0.0806  0.3038  0.6567
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.95972    0.28566  10.361   <2e-16 ***
credits      0.02407    0.01325   1.817   0.0717 .
schoolE      0.09478    0.21630   0.438   0.6620
schoolS      0.27379    0.21774   1.257   0.2110
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.4104 on 122 degrees of freedom
Multiple R-Squared: 0.05626,    Adjusted R-squared: 0.03305
F-statistic: 2.424 on 3 and 122 DF,  p-value: 0.06899

> anova(stepwise)
Analysis of Variance Table
Response: gpa
           Df  Sum Sq Mean Sq F value  Pr(>F)
credits     1  0.3536  0.3536  2.0987 0.14999
school      2  0.8717  0.4359  2.5872 0.07936 .
Residuals 122 20.5532  0.1685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```