

MARKING KEY

Assignment #3 Solutions

Due: 1 pm, Friday, March 6th

TOTAL = 42 PTS

Issue: Student success in their first year is a strong predictor of subsequent degree success. Students experiencing difficulty in first year have a high likelihood to not return for their second year. Interventions to assist at risk students must occur early in the first year if they are to be effective. Ideally, they should start in the first two months of the fall term. This is before the first midterm results are available. Can we identify who will be unsuccessful (or successful) before classes even start?

Questions: *What is success for a new student?* Getting a GPA above 2.00? Keeping your entrance scholarship? Performing above average compared to similar peers?

What data is available? We will limit attention to data that is captured in Banner. In recent years, we have received electronic transcripts for all Nova Scotia high school applicants raising the possibility of knowing student performance, by course, as far back as grade 10. However, for this study, we will look at admissions data from 2015, before these transcripts were available.

For applicants from Canadian high schools, the **high school average** (out of 100) was recorded at the time of the admission decision. The average was based upon the 5 courses used for the admission decision. If the admission decision was made in the fall while the student was in grade 12, the average was based upon final grade 11 grades. If the decision was made later, it could include some final or mid-year grade 12 course grades.

Exclusion: High school average data for international students is inconsistent, so we will restrict analyses to Canadian students only. Many new students are transferring from colleges or universities. This is a very diverse student population for whom high school data may not be available or may not be a good predictor of success. We will not include these applicants in this study. Identifying potentially unsuccessful international or transfer students should be addressed in separate studies.

Other admission data includes the **date of application** and the date of the admission decision, the name of the **high school** and the **province** in which it is located, and the student's **program** of study. Note that admission requirements vary among Arts, Business, Science and Engineering, so the courses included in the high school average will be different. Similarly, the recommended mix of first year courses varies by program.

The extract occurred in January 2016 and included only those students admitted for September 2015. The **fall GPA** was recorded for each student.

Exclusion: Some new students drop out during the fall term. Some register but never actually attend. These students will show with no earned credits and a GPA of zero. We will consider them as a distinct population of students that should not be included in this study. We will delete all students with a fall GPA of zero. This action deletes students who were active in the fall but failed all courses. As such, some valid records have been deleted, but there was no way to identify these students separately from those who simply disappeared.

In addition to the above "data cleaning", several variables have been recoded. It is impossible to look at individual high schools, but this information was used to distinguish between students from the Halifax

region and those from elsewhere in Nova Scotia. Sample sizes from many provinces outside Nova Scotia are small, so they have all been grouped together as Out of Province (OOP). Similarly, because of the small sample sizes, Engineering students have been merged with Science students and Environmental Studies with Arts.

Although significant data preparation has been done, you will need to do more data preparation to complete assignments 3 and 4. Some will be necessary to answer the questions posed and some will be needed to conform to the way Excel does data analysis.

1. a) Rather than simply define "success" in terms of GPA, let us define it in terms of the consequence of high or low GPA. If GPA exceeds 3.70, then a student's scholarship will be renewed or the student will be eligible for a scholarship. However, if the GP is below 2.00, then the student may be placed on probation. Construct the look up table below in the Look Up Table tab of the file **GPA and High School assignment 3.xlsx**.

GPA	Success
0	Probation
2.0	Satisfactory
3.7	Scholarship

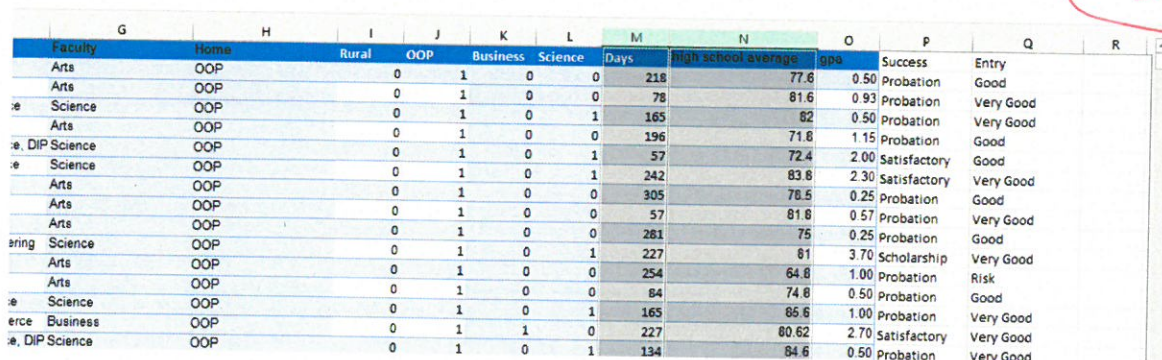
Insert a column to right of GPA and call it Success. Use VLOOKUP to assign success values based upon GPA.

- b) Similarly, let us recode the high school average according to the following look up table:

HS Average	Entry
60	Risk
70	Good
80	Very Good
90	Excellent

Put the above table in the Look Up Table tab. Insert a column to the Right of Success and call it Entry. Use VLOOKUP to assign values to Entry based upon HS Average.

- c) Print a screen shot showing the new columns.



The screenshot shows an Excel spreadsheet with columns G through R. The data is organized by Faculty (Arts, Science, Business, DIP Science) and then by location (Home, Rural, OOP). The new columns added are 'Success' (column P) and 'Entry' (column Q). The 'Success' column contains values like 'Probation', 'Satisfactory', and 'Scholarship'. The 'Entry' column contains values like 'Good', 'Very Good', 'Risk', and 'Excellent'. A red circle with the text '2 PTS' is drawn over the top right corner of the spreadsheet.

Faculty	Home	Rural	OOP	Business	Science	Days	High school average	GPA	Success	Entry
Arts	OOP		0	1	0	0	218	77.6	0.50	Probation
Arts	OOP		0	1	0	0	78	81.6	0.93	Probation
Science	OOP		0	1	0	0	165	82	0.50	Probation
Arts	OOP		0	1	0	1	196	71.8	1.15	Probation
DIP Science	OOP		0	1	0	1	57	72.4	2.00	Satisfactory
Science	OOP		0	1	0	1	242	83.8	2.30	Satisfactory
Arts	OOP		0	1	0	0	305	78.5	0.25	Probation
Arts	OOP		0	1	0	0	57	81.8	0.57	Probation
Arts	OOP		0	1	0	0	281	75	0.25	Probation
Science	OOP		0	1	0	1	227	81	3.70	Scholarship
Arts	OOP		0	1	0	0	254	64.8	1.00	Probation
Arts	OOP		0	1	0	0	84	74.8	0.50	Probation
Science	OOP		0	1	0	1	165	85.6	1.00	Probation
Business	OOP		0	1	1	0	227	80.62	2.70	Satisfactory
DIP Science	OOP		0	1	0	1	134	84.6	0.50	Probation

- d) Save your file and insert a Pivot Table.
2. Let us look at what "success" is like by program. In your Pivot Table, pull Faculty to the Rows and Success to the columns. Pull ID to the Values box. Change the summary to Count and Show Values to Percent of Row Total. This will show you the percentage of students in each program who achieve each level of success.

Row Labels	Probation	Satisfactory	Scholarship	Grand Total
Arts	30.41%	55.85%	13.74%	100.00%
Business	14.98%	62.80%	22.22%	100.00%
Science	10.67%	59.55%	29.78%	100.00%
Grand Total	21.18%	58.73%	20.08%	100.00%

- a) What percentage of Science students are "Scholarship" students? What percentage of Arts students are Scholarship students?

30% of Science students are scholarship students compared to only 14% of Arts students.

2 PTS

- b) What percentage of Science students are at risk of probation? What percentage of Arts students are at risk of probation?

11% of Science students are at risk of probation compared to 30% of Arts students.

2 PTS

- c) You would think that Science is more demanding than Arts. Why do you think Science is so easy?

Is Science easy or are Science students different? We cannot tell what the reason for the performance difference is. But Science students must take specific courses in high schools and their course mix in university is different. Science students must plan their courses in high school, whereas Arts students do not. So, admission to Science is intentional, whereas admission to Arts may not be. The point is that the two groups of students are different. Whatever the difference is may be the reason for different performance.

0 PTS

3. Let us look at what students look like on admission. In the Pivot Table, drag Success out from the columns and drag Entry in. The table likely shows entry performance in a sequence that is not best to worst. Rearrange the columns by moving labels. Print a copy of your table with columns arranged correctly.

Count of ID	Column Labels				
Row Labels	Excellent	Very Good	Good	Risk	Grand Total
Arts	15.79%	50.29%	30.99%	2.92%	100.00%
Business	15.46%	49.76%	34.78%	0.00%	100.00%
Science	25.28%	52.81%	19.66%	2.25%	100.00%
Grand Total	18.02%	50.76%	29.30%	1.93%	100.00%

MUST MOVE VERY GOOD TO LEFT

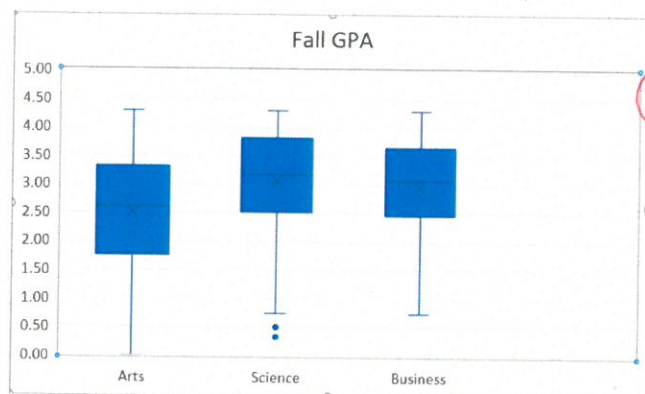
3 PTS

4. What percentage of science students were excellent or very good? What percentage of Arts students were excellent or very good?

78% of Science students were very good or excellent in high school compared to 66% in Arts.

2 PTS

5. It looks like Science attracts stronger students and they do better than Arts students. Let us look at the distributions as a whole. Construct a Box and Whisker chart for GPA. In Chart Design, and Select Data, Edit the horizontal axis to show Faculty. **Print the Box and Whisker chart for Arts, Business and Science. This should be a single chart showing all three faculties.** Comment on the differences. Are they as clear as what you found in 3?

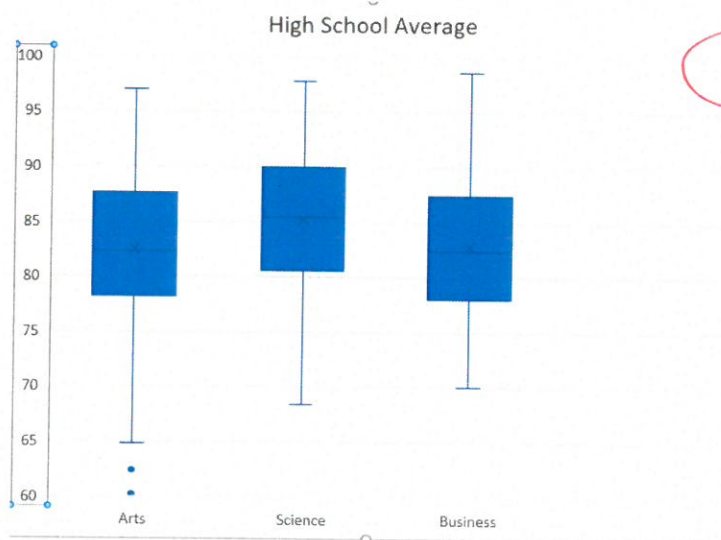


3 PTS

Although Both Arts and Science have comparable overall spread, the middle 50% of Science is clearly higher than the corresponding middle of Arts. It does look like Science has a larger group of the best and a smaller group of the worst.

2 PTS

6. And what about how they look on admission. Construct a Box and Whisker chart of HS Average and then break it down by Faculty as done in 5. You may want to format the vertical axis so that it ranges from 60 to 100 so that you can see detail more clearly. **Print the Box and Whisker chart for Arts, Business and Science.** Are the differences as clear as you saw in 4?

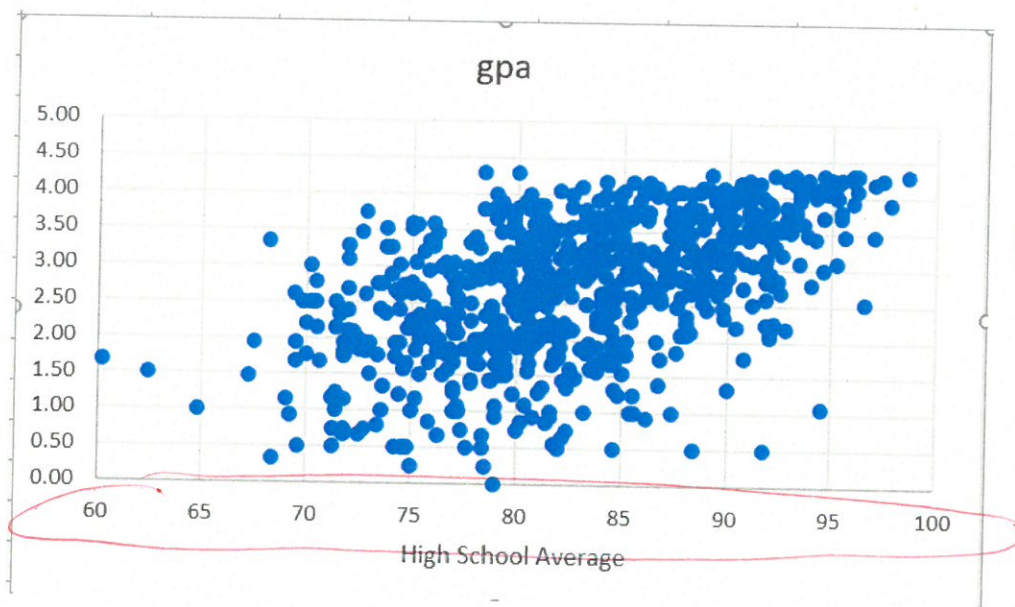


2 PTS

Although it appears that Science students were stronger in high school than Arts students, the difference is less profound than we saw in the GPA data. Science students may have been relatively more successful in high school, but they were much more successful in university. We were not asked about Business students, but they look like Arts students in high school and perform like Science students in university. I guess this is one of the bird courses that business students take to boost their GPA!

2 PTS

7. Box and whisker charts hide unnecessary clutter but can still mask patterns that can emerge from grouping data – less detail can show more. Nonetheless, we feel that Science attracts better students and Science students do better. Let us look at high school grades versus GPA directly, without looking at program. Construct a Scatter Chart of GPA versus HS Average. Excel puts the variable in the left column on the vertical axis and the one in the right column on the horizontal axis. Right click on the horizontal axis and Format Axis such that the minimum is 60 and the maximum is 100. **Print the Scatter chart.**



3 PTS

Maybe this is not a surprise to you, but it is always a surprise to faculty and university administrators. High school and university are two different worlds and the transition from one to the other can have unexpected positive and negative outcomes.

8. As you can see, there is a lot of noise. Although GPA is higher (on average) for those who did better in high school, there are obviously many other factors in play. How strong is this correlation? Use the Correlation function in the Data Analysis add in in Excel to measure the correlation. **What is the correlation?**

	high school average	gpa
high school average	1	
gpa	0.544596	1

The correlation is 0.54. Looks ok when we think that the correlation ranges from -1.0 to + 1.0. The correlation is positive meaning that better high school grades correspond to better university grades.

2 PTS

9. Some believe that the best applicants apply early and late applicants are the weakest. The variable Days measures the number of days between the start of classes and the date of application. Find the correlation between Days and HS average and GPA. To use the Correlation function in Excel, the variables must be in neighboring columns. This should be the case in your file. If not, then you must move the Days column so that it is adjacent to the other two. Is there evidence that stronger students apply earlier? Justify your conclusion.

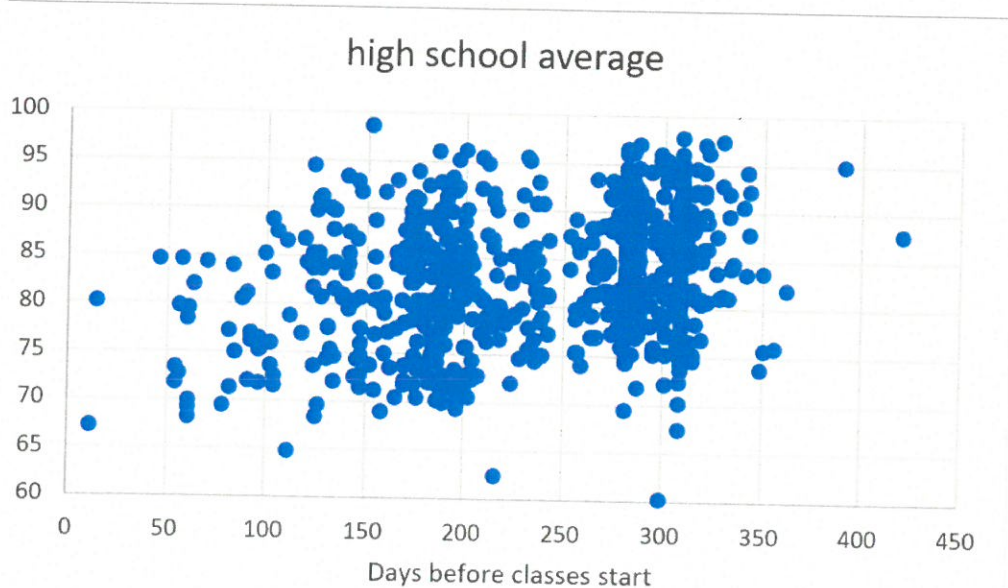
	Days	high school average	gpa
Days	1		
high school average	0.31880994	1	
gpa	0.2552367	0.544596	1

2 PTS

The correlation between HS average and Days is 0.32. This suggests as the number of days increases, HS average increases. Students who apply in October have Higher grades than those who apply in August. If the correlation coefficient had been negative, then the best students would be those who applied late!

2 PTS

10. Construct a scatter chart of days versus HS average. Right click on the vertical scale and Format the axis such that the minimum is 60 and the maximum is 100. You will get some anomalies on the horizontal axis with some unusual application dates. Many would feel that the correlation coefficient is quite large. Comment on how strong you feel the pattern is. Print your scatter chart.



2 PTS

From the chart it is hard to say that the high school average is much higher for those that apply a long time before classes start. This looks like a very weak pattern. Positive correlation is on a scale of 0 to 1. Our correlation is 0.32. Not super high but one would think that the pattern would be stronger. In fact, strong visual patterns only occur for very high correlations. 2 PTS

11. To date, we have still been "exploring" our data. As we move to "modelling" we want to know how to predict student success (value estimation) and to be able to explain why they are successful (Causal modelling). Use the Regression function in Data Analysis in Excel to fit the linear relationship between GPA and HS Average. What is the equation of the line?

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.544596
R Square	0.296585
Adjusted R Square	0.295615
Standard Error	0.810406
Observations	727

ANOVA

	df	SS	MS	F	Significance F
Regression	1	200.762	200.762	305.6863	2.23E-57
Residual	725	476.1498	0.656758		
Total	726	676.9118			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-3.77452	0.375904	-10.0412	2.66E-22	-4.51251	-3.03652	-4.51251	-3.03652
high school average	0.078726	0.004503	17.48389	2.23E-57	0.069886	0.087566	0.069886	0.087566

Excel gives lots of numbers. For the equation of the line, we need the numbers from the last table that give intercept and high school average. The estimated line is

Est GPA = -3.775 + 0.0787HS Avg 3 PTS.

12. The equation of the line has a slope and an intercept. Explain the meaning of the slope.

The slope is 0.0787. This means that if a student's HS average is 1 point higher, then the student's GPA will be 0.0787 points higher. Suppose that you have two students, one with a HS average of 75 and another with an average of 85 (10 points higher). On average, the GPA of a students with an 85 average will be $0.0787(10) = 0.79$ points higher than a student with a 75 average. That is like an A- instead of a B. 2 PTS

13. The intercept is more challenging to explain. It represents the expected GPA when a student has a high school average of zero. This doesn't happen. The lowest entering average cannot be below 60 for a student on special admission terms. What is this student's expected GPA?

So a student with a 0 HS average will have a GPA of -3.775. Makes no sense because no such student exists. With a 60 average, we would expect the student to have a GPA of $-3.775 + 0.0787(60) = 0.95$. Not very good. Why would we admit a student who is likely to fail? Look at the chart. Many weak students succeed and strong students fail. Although this student may be likely to fail, if given a chance, the student may surprise us and succeed if given the opportunity. If not given the opportunity, we will never know.

2 PTS

14. If we wish to admit no student whose expected GPA was below 2.00, what average should we set as a minimum for admission?

Math question. Target GPA is 2.00, so $2.00 = -3.775 + 0.0787(HS)$

$$2.00 + 3.775 = 5.775 = 0.0787(HS)$$

$$HS = 5.775 / 0.0787 = 73.3$$

The minimum HS average for admission should be 73.

2 PTS

15. If your institutional mission includes accessibility (Saint Mary's College was created because Irish Catholics were not allowed access to education), should we set the minimum entry average at the value you found in 14?

This is a difficult question and there are many different points of view. Some will say that university is ridiculously expensive so we should not admit students unless we are confident that they will succeed. Others will argue that the economic returns to a university degree are significant. If you are from a low income family, the best path to the middle class is through education. Take a chance. If the institution sees accessibility as part of their mission, they can reduce financial barriers to access and lower the financial risk to targeted populations.

0 PTS