

DATA STORY TWO

Yesterday, NUT Management approached the NUT Baccalaureate Analysis and Reporting Unit (NUT-BAR) with a request for statistical support and recommendations based on its understanding of NUT's undergraduate programs. In particular, NUT Management wanted clarity on whether students' junior college (JC) or polytechnic academic achievement was associated with their future academic performance at the university.

NUT-BAR's team leader, Angela, conveyed the request to her team. The team quickly discussed how to first analyse the dataset. After quickly reviewing the dataset NUT Management provided the team, one member calculated the means, standard deviations and medians of students' pre-entry scores, first-term GPAs and final-year GPAs. At first glance, the data looked to be prepared well enough to work with and sufficient enough to answer the question.

Today, Angela has called the team together for one morning meeting to work out the team's next steps.

Angela: "Good morning, team! I hope you had a chance to rest up yesterday. Today, we're digging deep into students' academic records before we head on to our other duties. Where would you like to start?"

Jennifer: "Given yesterday's briefing, I would think it's very important to identify the strength of the relationship between students' pre-entry scores and their academic performance at NUT.

Luke: "Jenn is right. If we want to know if students' JC or polytechnic courses may have an impact on their academic performance at NUT, we should look at the relationship between the two variables. It seems straightforward to check whether the past performance is **associated** with the academic **outcomes**."

Jayce: "Yes, totally. Then, if there is an **association**, we should check whether it is positive or negative."

Luke: "I'm going to guess it will be positive. Good grades before leads to good grades later, right? Give me a minute and I'll visualise the variables to make things easier for us to view the strength. I'll start with a basic **scatterplot**. Hopefully, that will surface a possible relationship between the changes observed in the values of two different variables."

Jayce: "Hang on there, Luke. While you were getting ahead of yourself, I made this scatterplot showing the relationship between Pre-Entry Scores and First-Term GPAs. I'll share it on the screen."

Figure 1 shows Jayce's scatterplot. Each point in the scatter plot corresponds to one student, and it is of the form (x,y), where x stands for the Pre-entry Score and y, the Term 1 GPA of the student. The values for Pre-Entry Scores mostly cluster between 80 and 100. A few outliers lower than 70 are also observed.

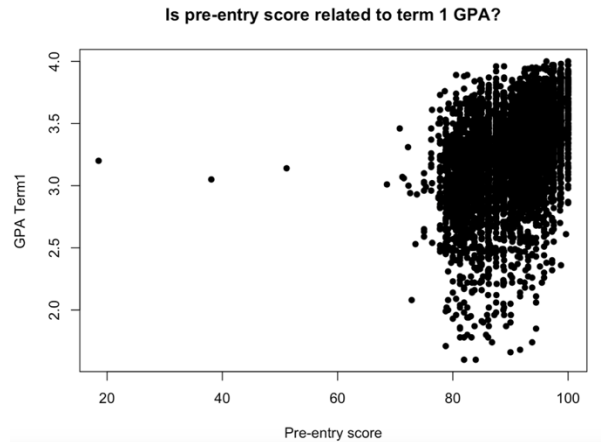


Figure 1

Jayce: “Well, this is interesting. We’ve got three or four points outside hmm maybe the Pre-Entry score of 68. Most values are clustered between 80 and 100.”

Jennifer: “Well, currently I don’t see any clear patterns on the visualisation. I suspect there might be some **confounders** in the data. Some data, but in a small amount, may not be valid enough to be included. My guess is there is something going on with those low Pre-Entry Scores. How do students with scores that low get into NUT? Is there someone we can check with about them?”

Angela: “The admissions team would know. I’ll call them now.”

In her office, Angela called the Office of Admissions (OAM) data manager. The two had a five-minute conversation before Angela returned to her team.

OAM data manager confirmed that this batch of students were admitted via different programs. Some students were through the Special Program, where students excelled in sports or arts.

Angela: “Just talked to OAM. NUT does admit students with low Pre-Entry Scores, but only in special circumstances where a student has a special talent for a sport or an art. These students were admitted through a “Special Program.”

The team seemed to be confused and was waiting for a more complete answer to the mystery, and she knew it.

Angela: “I’ll get to the point. NUT’s admission procedures actually go through each application twice. On the first pass, students’ academic performance, represented by their Pre-Entry Scores, are looked at. The Pre-Entry Score is a single number that represents all of a student’s accumulated academic performance records from their JC or polytechnic studies. For this particular cohort, the minimum Pre-Entry Score for admission to NUT was 79 out of 100.”

Luke: “What happens on the second pass?”

Angela: “During the second pass, the admissions team looks to see if an application with a Pre-Entry Score below 79 can still be admitted through the ‘Sports Talent Program’ or ‘Arts Talent

Program'. For these students, their past academic records were not the priority criterion; instead, NUT considered their potentials to excel in sports or arts."

Jayce: "Oh I see! Do we have a list of these students?"

Angela sent an updated version of the dataset to the team. The newer version contained students' admission program information. Luke got to work.

Luke: "Ok, I think I have some quick counts of how many were admitted via the first pass and the second pass... Hmm, I see 35 students admitted via the 'Sports Talent Program' and 91 via the 'Arts Talent Program'. The remaining 5450 students were admitted through the 'Standard Program'. Wait, NUT only admitted 126 students with special talents?"

Jennifer: "Well, when compared to 5450 students, 126 students may not be so influential. Not to second guess the Management, but only 35 students were admitted through the sports program?? Maybe we can just exclude the students from the 'Special Program'?"

Jayce: "Wait a second, I don't think we should jump to conclusion so quickly. It might make more sense if we see the percentage for each category. Let me show you this figure."

Figure 2 shows the chart Jayce shared with the team.

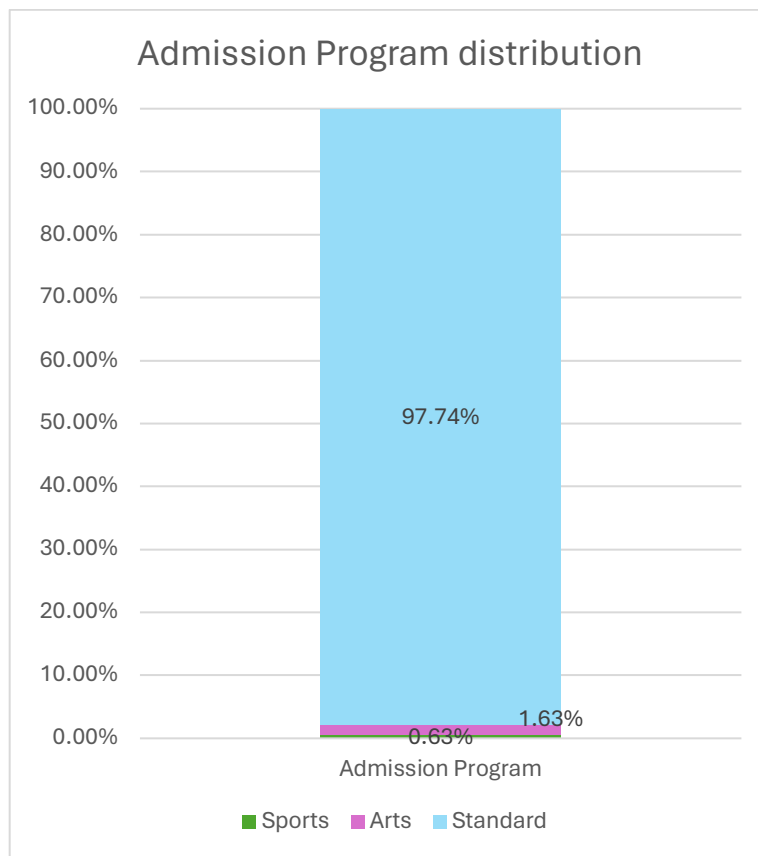


Figure 2

The chart indicates that 97.74% of the students were admitted through the standard procedure of earning a Pre-Entry Score above the cut off value for the cohort. 1.63% of students entered NUT due to their talents in arts, and 0.63% were admitted because of their sports talents.

Luke: “Now it’s quite clear that there are much more students under the standard admission procedure than through the special programs. It’s like 43 to 1 in favour of the standard process! Can we filter out the special program students now?”

Jayce: “In this case, as currently we’re discussing the impact of Pre-Entry Scores on students’ future GPAs, maybe we can exclude data from these students for now because NUT’s admission office looked at their talents instead of their academic records, right? It feels like they thought these students would make their mark on the university in non-academic ways. My guess is if we exclude these students who were selected outside the normal process, we will have a better chance of seeing patterns more clearly.”

Jennifer: “Well, I agree with Jayce. If we know the cut off for the first pass was 79 and the upper limit is 100, maybe we can include Pre-Entry Scores between 79 and 100. Records with Pre-Entry

Scores falling outside this range should be considered invalid as the corresponding students were admitted to NUT via considerations very different from the other 98% of students.”

Everyone turned to Angela.

Angela stared at the chart, and then looked over her team.

Angela: “Show me how many data points we would filter out.”

Figure 3 shows the scatterplot of Pre-Entry Scores and Term 1 GPAs Jayce shared with the team.

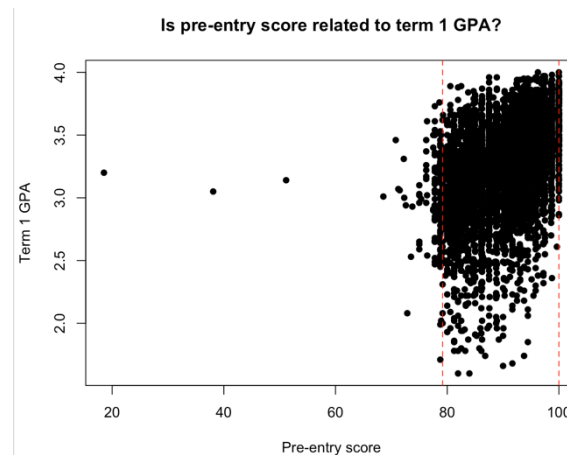


Figure 3

The red lines indicate the suggested minimum and maximum Pre-Entry Score values Jennifer suggested for retaining the data in the dataset. Some of the Pre-Entry Score values fell below the stated minimum threshold provided by OAM.

Angela stared at the chart for a while. Her eyes moved back and forth between the boundary lines.

Angela: “If that was for Term 1 GPA, show me the same thing for Final GPA.”

Jayce: “Saw that coming. Here you go.”

Figure 4 shows a similar chart as Figure 3 with minimum and maximum Pre-Entry Score values demarked in red when considering the relationship between Pre-Entry Scores and Final GPAs.

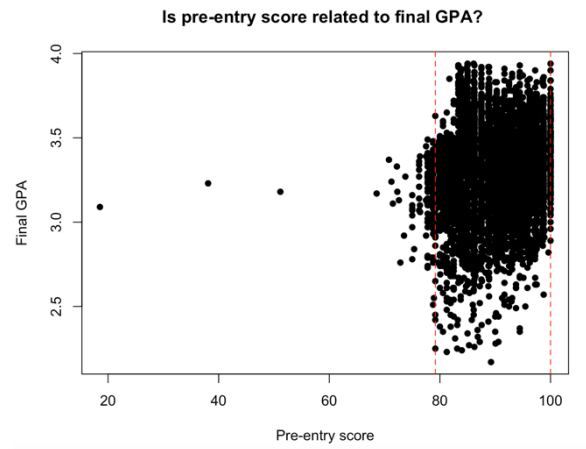


Figure 4

Angela looked at the new chart and again focused along the red lines. Then, she solemnly nodded to Jayce.

Jayce: “Say no more! Consider them gone! Applying minimum and maximum Pre-Entry Score filters, now! And re-rendering Term 1 GPA!”

Figure 5 shows Jayce’s resulting scatterplot of Pre-Entry Score and Term 1 GPA after filtering out invalid Pre-Entry Scores.

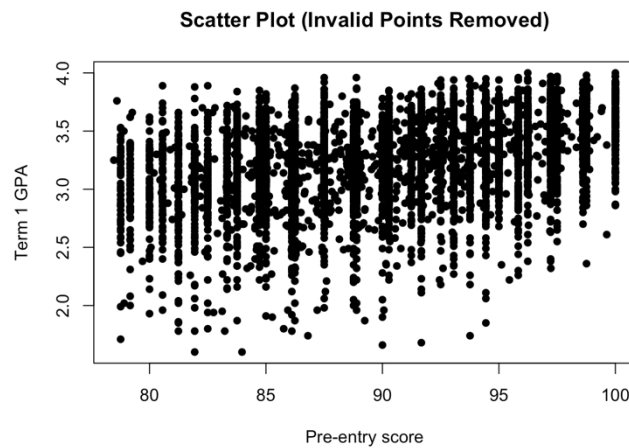


Figure 5

Figure 6 shows the resulting scatterplot for Pre-Entry Scores and Final GPA after the removal of invalid Pre-Entry Scores.

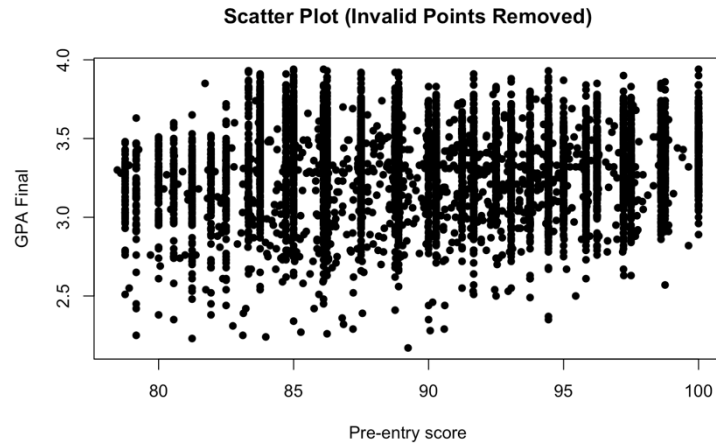


Figure 6

Jennifer: “Well, after removing the invalid data, I can see some interesting patterns on the plots. It’s definitely worth exploring more!”

Luke: “Agreed. We should go deeper for the association of the variables.”

Angela: “Looks good, team! Good job for today. Now I think we get a gist of the data. The plots are definitely telling us something could be there. We will continue our work tomorrow and see if we can use these patterns to see how strong the relationships are.”

Team: “Thank you everyone for the hard work today!”