

DATA STORY THREE

Last week, the NUT Baccalaureate Analysis and Reporting Unit (NUT-BAR) consulted the Office of Admissions (OAM) data manager who informed the team about NUT's Special Programs admissions process. The "Sports Talent Program" and "Art Talent Program" selected applicants by giving more priority to their excellence in sports or arts rather than solely focusing on their past academic performance distilled into a single score. For the cohort described in the dataset NUT-BAR received, there were 35 students admitted via the "Sports Talent Program" and 91 via the "Arts Talent program." The remaining 5450 students were admitted to the university by reaching a minimum pre-entry score of 79 out of 100.

After considering how the Special Program admissions process affected the data questions, the team decided to filter out the data of students admitted to NUT through the Special Programs. Filtering out data from those students ensured that the team was analysing data from students who matriculated at NUT through the same selection process.

With a harmonised dataset, the team now had two nicely distributed scatterplots showing some interesting patterns. Based on the scatterplots, the team agreed to look deeper at the data and explore the relationships between students' pre-entry scores and the two sets of GPAs when they next met.

As usual, Angela initiated the discussion: "Team, now that the surface looks to be telling us something, we can test for something hidden deep inside the data. I know everyone is aware that we need to submit the results to the bosses as soon as possible, but since we have a bit of time to breathe let's discuss our approach before our other duties."

Jennifer: "Definitely! So now we've already removed the invalid data, I think we can focus on understanding the strength of the relationships."

Luke: "Like using a correlation coefficient to quantify the level of linear association between two variables? And we can fit a line or curve through the points in the scatterplots and enable us to make predictions on the value of one numerical variable when given the other numerical variable."

Jennifer: "That's so textbook but exactly what I was thinking, Luke. We can use our existing scatterplots to measure the association between pre-entry scores and GPAs."

Angela: "That sounds like a plan. Let's proceed with the correlation coefficient and regression analysis approaches. Luke, can you take the lead on this?"

Luke: "Sorry, Angela. I have a meeting at 1pm..."

Angela: "I see. No worries. Jennifer, can you take ownership of this and get it done asap? I'm curious about what you'll find."

Jennifer: "Sure, I'll get started on it right away!"

Jennifer conducted a regression analysis at the university level to assess the relationships between the variables while the team went about other tasks. As shown in Figures 1 and 2, vertical banding lines appeared in the plot with few values between bands.

Scatter Plot of Term 1 GPA vs. Pre-Entry Score

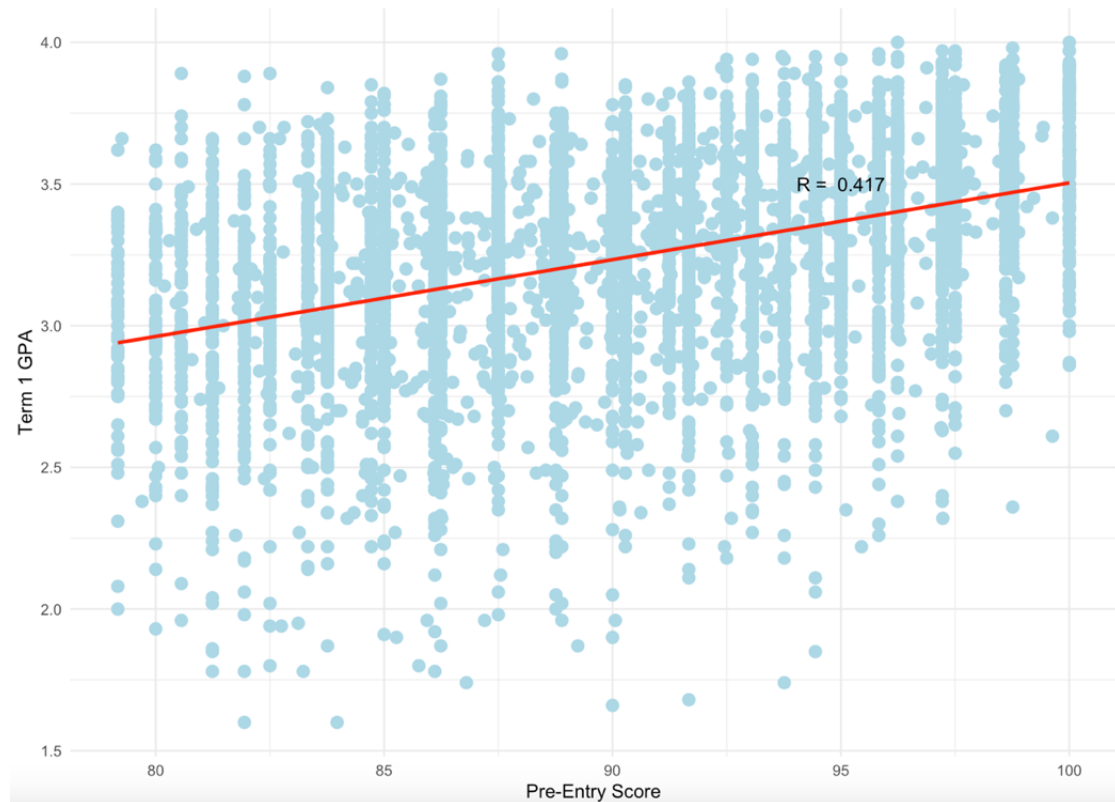


Figure 1

Scatter Plot of Final Term GPA vs. Pre-Entry Score

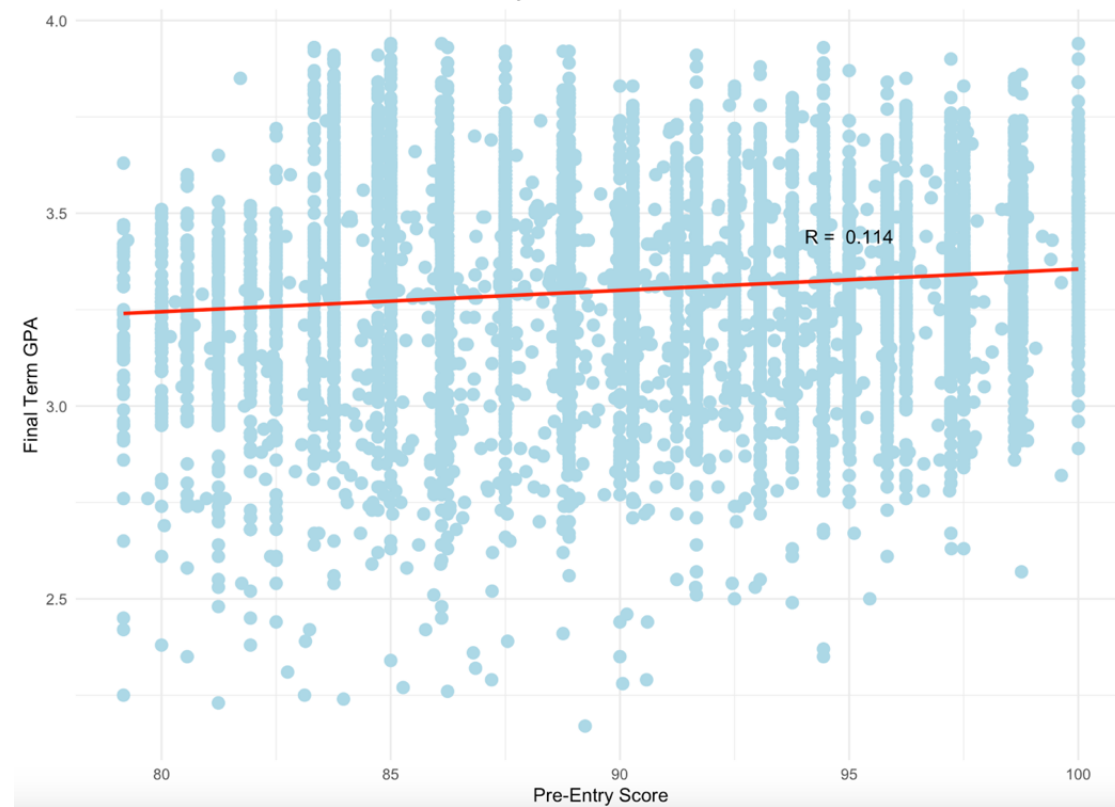


Figure 2

Jennifer: “Hey guys, I think I’ve done the work. Did you notice anything interesting going on in the data?”

Most of the team nodded.

Luke: “Sorry, I just got out of my meeting. I’m looking at the graphs now. I’ll catch up.”

Jennifer: “To help you out, Luke, the red lines indicate the direction of the linear association between the two variables. I also calculated the **correlation coefficient (R value)** for these two. Correlation coefficient R shows the strength of the linear relationship between an independent and a dependent variable.”

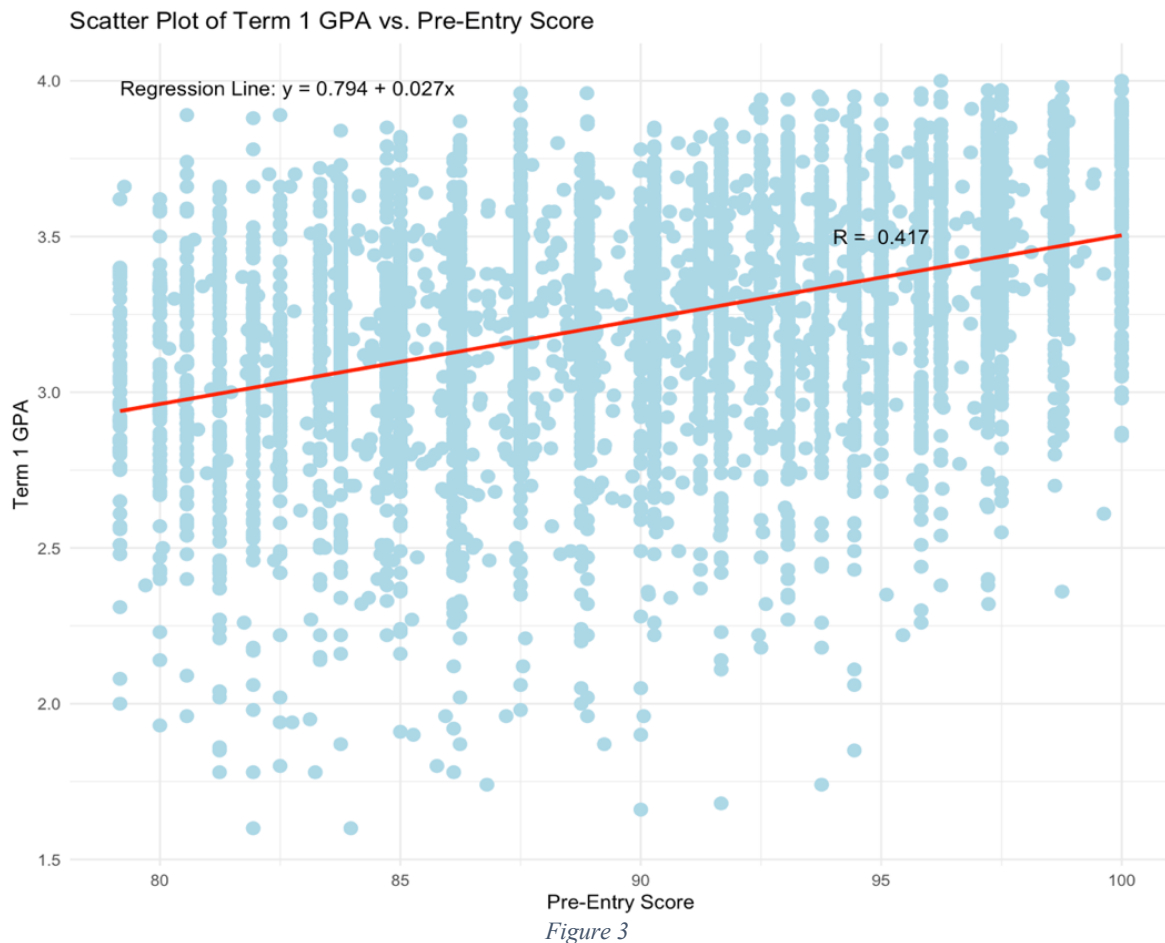
Jayce: “Interesting... I think it appears that the strength of the linear relationship between Pre-Entry Score and Term 1 GPA is different from that between Pre-Entry Score and Final GPA.”

Luke: “That sounds right... It also looks like the Term 1 GPA analysis indicates a stronger relationship between variables than the Final GPA analysis.”

Jennifer: “Agreed. I think the Final Term GPA analysis indicates that Pre-Entry Score and Final Term GPA are mostly unrelated.”

Jayce: “Hmm I would say the relationship between Pre-Entry Score and Term 1 GPA is moderate as the R value is at 0.417, but I wouldn’t say the relationship between Pre-Entry Score and Final Term GPA is mostly unrelated. Maybe it’s weak?”

Angela: “Well, since the correlation coefficient R is 0.114, I wouldn’t say it’s mostly unrelated. Also, it’s worth noting that with the data we have right now, we could potentially predict a student’s future GPAs based on their Pre-Entry Score. I guess the prediction will go better if we use the regression equation for Term 1 GPA since the R value is higher? I personally added an equation to show the tendency. See Figure 3, everyone.”



Jayce: “That’s quite interesting. However, I guess the equation doesn’t really answer the data questions we’re trying to solve here, but we can see the relationship between variables and their trend better.”

Jennifer: “Based on the two graphs, I think we can safely say:

1. **Past academic performance is moderately associated with a student’s first term GPA.**
2. **Past academic performance is weakly associated with a student’s final cumulative GPA.”**

Luke: “That sounds safe and accurate. Should we lock that in as a conclusion and wrap this up?”

Jennifer: “Wait, hold on first. We shouldn’t do something so simple, just looking at the students across the university. NUT has around 5600 students per cohort. I think we may need to look deeper into the data. Considering the diversity among schools, it might be beneficial to use separate regression lines for each school when predicting outcomes for students within their respective schools. I think we should explore the data further with a deeper analysis rather than jump at the first conclusion. What does everyone think?”

Angela: “I think you make a solid point, Jennifer. Considering how the dynamics of each school differ from each other, I would also propose we look at different schools to offer a more specific analysis.”

Jennifer: “This is exactly my point! Thank you for the support. I think we should focus on that!”

Angela: “Great! Can you work on that with Luke?”

Luke and Jennifer nodded their heads.

Working together, Luke and Jennifer calculated the **regression coefficient (gradient of regression line)** for each of the eight schools. Tables 1 and 2 show the regression coefficient values for the two GPA variables overall for all schools and for each individual school.

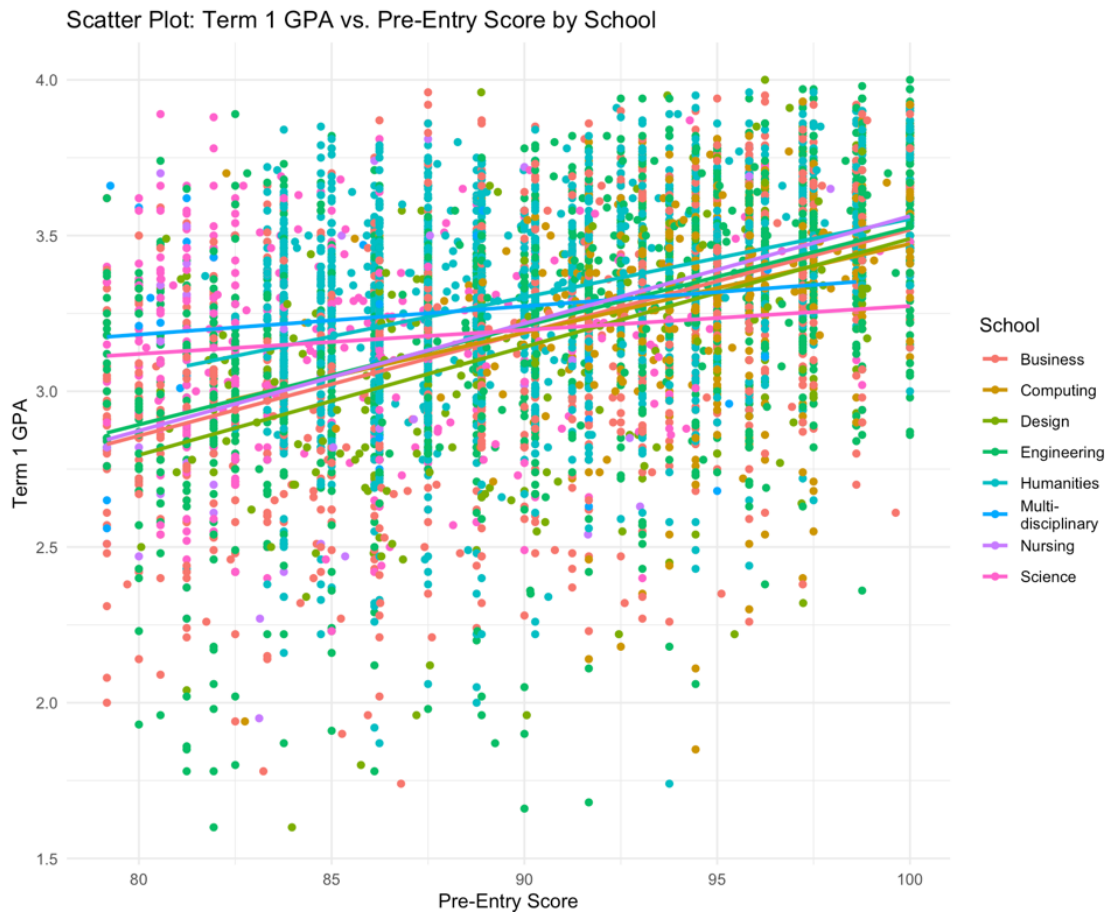
Table 1. Regression coefficient value for Term 1 GPA regressed on Pre-entry Score overall and by school

SCHOOL	REGRESSION COEFFICIENT
ALL SCHOOLS	0.027
BUSINESS	0.033
COMPUTING	0.028
DESIGN	0.035
ENGINEERING	0.032
HUMANITIES	0.025
MULTI-DISCIPLINARY	0.009
NURSING	0.034
SCIENCE	0.008

Table 2. Regression coefficient value for Final Term GPA regressed on Pre-entry Score overall and by school

SCHOOL	REGRESSION COEFFICIENT
ALL SCHOOLS	0.006
BUSINESS	0.013
COMPUTING	0.014
DESIGN	0.016
ENGINEERING	0.011
HUMANITIES	0.000
MULTI-DISCIPLINARY	0.002
NURSING	0.016
SCIENCE	0.002

Jennifer tried to draw eight lines for the eight schools. Figure 4 shows the scatterplot of Term 1 GPA vs. Pre-Entry Score by school.



Jennifer: “Hey, Luke. I don’t think we can fit so many lines into one graph if we intend to compare the regressions by school.”

Luke: “Yeah... We can’t do that. Now we’ve got eight schools. Although these lines show us the predictable linear relationships between variables, it’s impossible for us to even see all the lines in one plot, not to mention using them to compare schools!”

Jennifer: “And if we make eight separate scatterplots for the eight schools, we still won’t be able to compare between them!”

Overhearing the conversation, Jayce walked over to Jennifer and Luke: “How’re you two doing? Everything good?”

Jennifer: “We’ve hit a wall, Jayce. I don’t think we can visualise the school-level patterns the same way we did the overall pattern because we can’t fit all eight lines into one graph to compare their linear associations. It’s chaotic!”

Luke: “Take a look!”

Luke held up his screen directly in front of Jayce’s face.

Jayce: “Ah, I can see. That is a problem. You are comparing the strength of association between the past and future academic performance across eight schools and do it all at the same time

while still wanting to know the relationship between variables within each school, right? That's a bit tough."

Luke: "Understood. What if we measure the strength of the association by calculating the correlation for each school? Correlation determines the association or relationship between two variables. And if we want to know how the increase or decrease of students' pre-entry scores may influence their future GPAs, we can use the regression coefficients we calculated. Regression describes how to numerically relate an independent variable to the dependent variable. This helps us see the magnitude of the association."

Sensing the activity in the office, Angela listened from afar.

Jayce: "If I am following you, I believe this means that a one-point increase in Pre-Entry Score, would be associated with an average increase of 0.027 in Term 1 GPA and a 0.006 increase on average in Final Term GPA. Am I understanding it correctly?"

Luke: "Yes, exactly!"

Noticing the opportunity to convert enthusiasm and common understanding into action, Angela jumped into the conversation.

Angela: "Interesting point, Luke. Will you help us get the correlation results out?"

Luke: "Sure, with pleasure!"

Jennifer and Luke decided to conduct the correlation analysis across the eight schools. Table 3 and Table 4 present the correlation coefficient values.

Table 3. R value for Term 1 GPA vs. Pre-entry Score overall and by school

SCHOOL	R VALUE
ALL SCHOOLS	0.417
BUSINESS	0.483
COMPUTING	0.288
DESIGN	0.473
ENGINEERING	0.479
HUMANITIES	0.362
MULTI-DISCIPLINARY	0.197
NURSING	0.463
SCIENCE	0.122

Table 4. R value for Final Term GPA vs. Pre-entry Score overall and by school

SCHOOL	R VALUE
ALL SCHOOLS	0.114
BUSINESS	0.341
COMPUTING	0.249
DESIGN	0.384
ENGINEERING	0.303
HUMANITIES	0.000
MULTI-DISCIPLINARY	0.100

NURSING	0.436
SCIENCE	0.063

Jennifer: “Ok, so now we have a R value for each school that take Pre-entry Score into consideration. It’s good to compare between schools.”

Luke: “Yes! For the R values we calculated, we can then categorise them by strength. I can assign labels to the values to indicate their strengths with words. I’ll add a column to the tables for that. I believe this will help the bosses understand what we’re seeing. For example, a value of ‘0.4’ is equivalent to ‘a moderate relationship’. You can all refer to Table 5 for the mappings.”

Table 5 Conversion of Correlation values to strength of relationship labels

STRENGTH OF ASSOCIATION	CORRELATION VALUE RANGE
WEAK	>0.000 – 0.300
MODERATE	>0.300 – 0.700
STRONG	>0.700 – 1.000

Tables 6 and 7 show the correlation values and strength of relationship labels.

Table 6 Term 1 GPA vs. Pre-entry score

SCHOOL	R VALUE	STRENGTH
ALL SCHOOLS	0.417	Moderate
BUSINESS	0.483	Moderate
COMPUTING	0.288	Weak
DESIGN	0.473	Moderate
ENGINEERING	0.479	Moderate
HUMANITIES	0.362	Moderate
MULTI-DISCIPLINARY	0.197	Weak
NURSING	0.463	Moderate
SCIENCE	0.122	Weak

Table 7 Final Term GPA vs. Pre-entry score

SCHOOL	R VALUE	STRENGTH
ALL SCHOOLS	0.114	Weak
BUSINESS	0.341	Moderate
COMPUTING	0.249	Weak
DESIGN	0.384	Moderate
ENGINEERING	0.303	Moderate
HUMANITIES	0.000	Weak
MULTI-DISCIPLINARY	0.100	Weak
NURSING	0.436	Moderate
SCIENCE	0.063	Weak

Jennifer: “That does help to make it clear!”

Jayce: “Thanks, Luke!”

Jennifer: “The only thing I’m concerned about is how can we present all these analyses to the bosses... There is so much information to process!”

Jayce: “Let’s leave that discussion for tomorrow. I think we’ve done enough work today.”

Jennifer and Luke: “Agreed. Thanks, team!”

Angela: “I am so proud of how you are handling this assignment! Communication is key. Let’s continue with keeping each other updated on our progress. Regular check-ins, whether scheduled or impromptu, will ensure we stay on track and can quickly address any challenges that may arise!”