

# PREDICTING ACADEMIC SUCCESS

- A REGRESSION ANALYSIS OF STUDENT PERFORMANCE FACTORS
- PREPARED BY DERRICK NYAGESUKA AND BAYSA WAKJIRA
- STAT 311 REGRESSION ANALYSIS FALL 2025



# INTRODUCTION



**Goal:** To identify the most significant predictors of a student's final exam score



**Question:** how do individual habits such as study hours vs environmental factors such as parental involvement impact success?

# DATA DESCRIPTION

Source: student performance factors dataset

Sample: 6529 rows after removing 78 rows

List key variables: exam\_score(dependent). Hours\_studied,  
Attendance, parental\_involvement, access\_to\_resources,  
sleep\_hours, previous\_scores, motivation\_level, teacher\_quality



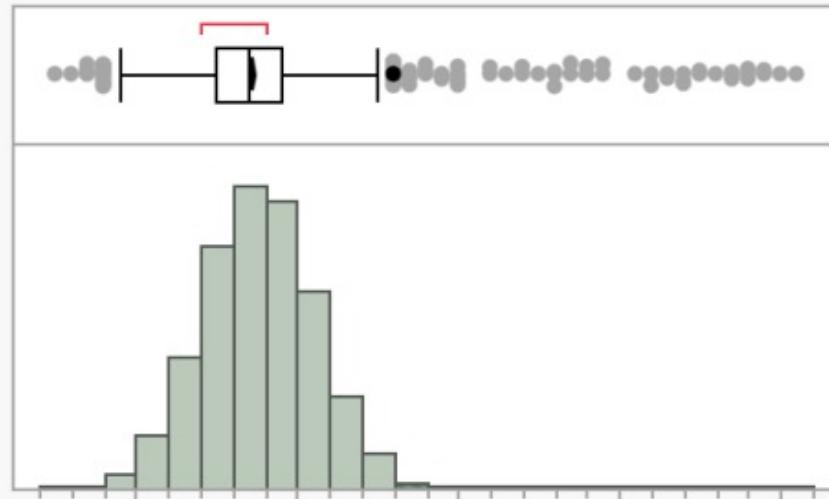
StudentPerformanceFactorsClean...



## Distributions

Student Edition

### Exam\_Score



### Quantiles

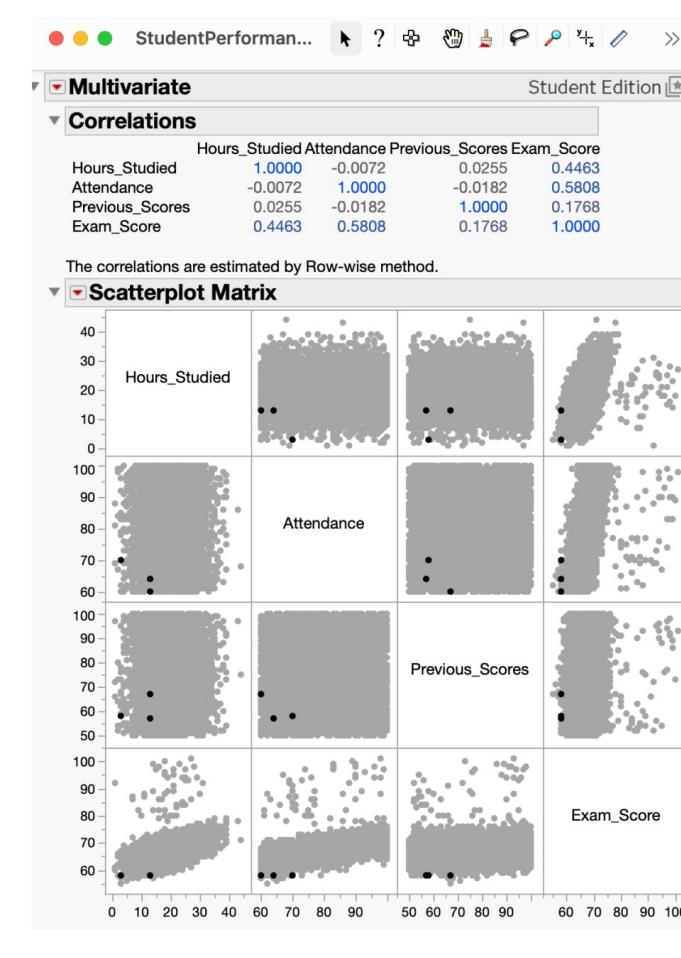
|        |          |     |
|--------|----------|-----|
| 100.0% | maximum  | 101 |
| 99.5%  |          | 86  |
| 97.5%  |          | 74  |
| 90.0%  |          | 72  |
| 75.0%  | quartile | 69  |
| 50.0%  | median   | 67  |
| 25.0%  | quartile | 65  |
| 10.0%  |          | 63  |
| 2.5%   |          | 61  |

### Summary Statistics

|                |           |
|----------------|-----------|
| Mean           | 67.242763 |
| Std Dev        | 3.9007333 |
| Std Err Mean   | 0.0482751 |
| Upper 95% Mean | 67.337398 |
| Lower 95% Mean | 67.148128 |
| N              | 6529      |
| N Missing      | 0         |

# EXPLORATORY DATA ANALYSIS

# EXPLORATORY DATA ANALYSIS



# METHODS

| Summary of Fit             |  |          |  |  |  |
|----------------------------|--|----------|--|--|--|
| RSquare                    |  | 0.666048 |  |  |  |
| RSquare Adj                |  | 0.665271 |  |  |  |
| Root Mean Square Error     |  | 2.238398 |  |  |  |
| Mean of Response           |  | 67.23898 |  |  |  |
| Observations (or Sum Wgts) |  | 5172     |  |  |  |

| Analysis of Variance |      |                |             |          |          |
|----------------------|------|----------------|-------------|----------|----------|
| Source               | DF   | Sum of Squares | Mean Square | F Ratio  | Prob > F |
| Model                | 12   | 51553.846      | 4296.15     | 857.4432 |          |
| Error                | 5159 | 25848.775      | 5.01        | <.0001*  |          |
| C. Total             | 5171 | 77402.622      |             |          | <.0001*  |

| Lack Of Fit |      |                |             |          |          |
|-------------|------|----------------|-------------|----------|----------|
| Source      | DF   | Sum of Squares | Mean Square | F Ratio  | Prob > F |
| Lack Of Fit | 5158 | 25848.275      | 5.01130     | 10.0226  |          |
| Pure Error  | 1    | 0.500          | 0.50000     | Prob > F |          |
| Total Error | 5159 | 25848.775      |             | 0.2479   |          |
|             |      |                | Max RSq     |          |          |
|             |      |                | 1.0000      |          |          |

| Parameter Estimates          |           |           |         |         |  |
|------------------------------|-----------|-----------|---------|---------|--|
| Term                         | Estimate  | Std Error | t Ratio | Prob> t |  |
| Intercept                    | 41.555851 | 0.332098  | 125.13  | <.0001* |  |
| Hours_Studied                | 0.2965202 | 0.005247  | 56.51   | <.0001* |  |
| Attendance                   | 0.1996928 | 0.002685  | 74.38   | <.0001* |  |
| Parental_Involvement[Low]    | -1.005362 | 0.052144  | -19.28  | <.0001* |  |
| Parental_Involvement[Medium] | -0.009794 | 0.041873  | -0.23   | 0.8151  |  |
| Access_to_Resources[Low]     | -1.029709 | 0.052408  | -19.65  | <.0001* |  |
| Access_to_Resources[Medium]  | 0.0374599 | 0.042052  | 0.89    | 0.3731  |  |
| Sleep_Hours                  | 0.0022601 | 0.021337  | 0.11    | 0.9156  |  |
| Previous_Scores              | 0.0472414 | 0.002176  | 21.71   | <.0001* |  |
| Motivation_Level[Low]        | -0.548298 | 0.047263  | -11.60  | <.0001* |  |
| Motivation_Level[Medium]     | 0.0113584 | 0.041909  | 0.27    | 0.7864  |  |
| Teacher_Quality[Low]         | -0.481919 | 0.068914  | -6.99   | <.0001* |  |
| Teacher_Quality[Medium]      | -0.03986  | 0.046172  | -0.86   | 0.3880  |  |

| Effect Tests         |       |    |                |          |          |
|----------------------|-------|----|----------------|----------|----------|
| Source               | Nparm | DF | Sum of Squares | F Ratio  | Prob > F |
| Hours_Studied        | 1     | 1  | 15998.507      | 3193.045 | <.0001*  |
| Attendance           | 1     | 1  | 27721.139      | 5532.694 | <.0001*  |
| Parental_Involvement | 2     | 2  | 2557.763       | 255.2442 | <.0001*  |
| Access_to_Resources  | 2     | 2  | 2577.791       | 257.2428 | <.0001*  |
| Sleep_Hours          | 1     | 1  | 0.056          | 0.0112   | 0.9156   |
| Previous_Scores      | 1     | 1  | 2362.267       | 471.4706 | <.0001*  |
| Motivation_Level     | 2     | 2  | 741.679        | 74.0136  | <.0001*  |
| Teacher_Quality      | 2     | 2  | 509.039        | 50.7980  | <.0001*  |

## Crossvalidation

| Source         | RSquare | RASE   | Freq |
|----------------|---------|--------|------|
| Training Set   | 0.6660  | 2.2356 | 5172 |
| Validation Set | 0.5881  | 2.5799 | 1357 |

# METHODS

| Summary of Fit               |       |                |                |          |          |
|------------------------------|-------|----------------|----------------|----------|----------|
| RSquare                      |       | 0.648914       |                |          |          |
| RSquare Adj                  |       | 0.648321       |                |          |          |
| Root Mean Square Error       |       | 2.313232       |                |          |          |
| Mean of Response             |       | 67.24276       |                |          |          |
| Observations (or Sum Wgts)   |       | 6529           |                |          |          |
| Analysis of Variance         |       |                |                |          |          |
| Source                       | DF    | Sum of Squares | Mean Square    | F Ratio  |          |
| Model                        | 11    | 64455.464      | 5859.59        | 1095.036 |          |
| Error                        | 6517  | 34872.756      | 5.35           | Prob > F |          |
| C. Total                     | 6528  | 99328.221      |                |          | <.0001*  |
| Lack Of Fit                  |       |                |                |          |          |
| Source                       | DF    | Sum of Squares | Mean Square    | F Ratio  |          |
| Lack Of Fit                  | 6503  | 34842.256      | 5.35787        | 2.4594   |          |
| Pure Error                   | 14    | 30.500         | 2.17857        | Prob > F |          |
| Total Error                  | 6517  | 34872.756      |                |          | 0.0264*  |
| Parameter Estimates          |       |                |                |          |          |
| Term                         |       | Estimate       | Std Error      | t Ratio  | Prob> t  |
| Intercept                    |       | 41.605645      | 0.271271       | 153.37   | <.0001*  |
| Hours_Studied                |       | 0.2936872      | 0.004787       | 61.35    | <.0001*  |
| Attendance                   |       | 0.1999374      | 0.002481       | 80.60    | <.0001*  |
| Parental_Involvement[Low]    |       | -0.981939      | 0.047972       | -20.47   | <.0001*  |
| Parental_Involvement[Medium] |       | -0.033394      | 0.038496       | -0.87    | 0.3857   |
| Access_to_Resources[Low]     |       | -1.035845      | 0.048204       | -21.49   | <.0001*  |
| Access_to_Resources[Medium]  |       | 0.0520001      | 0.0386         | 1.35     | 0.1780   |
| Previous_Scores              |       | 0.0472077      | 0.001991       | 23.71    | <.0001*  |
| Motivation_Level[Low]        |       | -0.54198       | 0.043396       | -12.49   | <.0001*  |
| Motivation_Level[Medium]     |       | -0.005421      | 0.038579       | -0.14    | 0.8883   |
| Teacher_Quality[Low]         |       | -0.501605      | 0.063942       | -7.84    | <.0001*  |
| Teacher_Quality[Medium]      |       | -0.022482      | 0.042676       | -0.53    | 0.5983   |
| Effect Tests                 |       |                |                |          |          |
| Source                       | Nparm | DF             | Sum of Squares | F Ratio  | Prob > F |
| Hours_Studied                | 1     | 1              | 20142.392      | 3764.198 | <.0001*  |
| Attendance                   | 1     | 1              | 34765.235      | 6496.906 | <.0001*  |
| Parental_Involvement         | 2     | 2              | 3171.934       | 296.3846 | <.0001*  |
| Access_to_Resources          | 2     | 2              | 3197.787       | 298.8002 | <.0001*  |
| Previous Scores              | 1     | 1              | 3007.959       | 562.1256 | <.0001*  |

# METHODS

| Summary of Fit |      |                |             |          |          |
|----------------|------|----------------|-------------|----------|----------|
| Source         | DF   | Sum of Squares | Mean Square | F Ratio  | Prob > F |
| Model          | 12   | 64456.821      | 5371.40     | 1003.689 |          |
| Error          | 6516 | 34871.400      | 5.35        |          | <.0001*  |
| C. Total       | 6528 | 99328.221      |             |          |          |

| Analysis of Variance |      |                |             |         |          |
|----------------------|------|----------------|-------------|---------|----------|
| Source               | DF   | Sum of Squares | Mean Square | F Ratio | Prob > F |
| Lack Of Fit          | 6502 | 34840.900      | 5.35849     | 2.4596  |          |
| Pure Error           | 14   | 30.500         | 2.17857     |         | <.0264*  |
| Total Error          | 6516 | 34871.400      |             |         |          |

| Lack Of Fit |      |                |             |         |          |
|-------------|------|----------------|-------------|---------|----------|
| Source      | DF   | Sum of Squares | Mean Square | F Ratio | Prob > F |
| Lack Of Fit | 6502 | 34840.900      | 5.35849     | 2.4596  |          |
| Pure Error  | 14   | 30.500         | 2.17857     |         | <.0264*  |
| Total Error | 6516 | 34871.400      |             |         |          |

| Parameter Estimates          |  |           |           |         |         |
|------------------------------|--|-----------|-----------|---------|---------|
| Term                         |  | Estimate  | Std Error | t Ratio | Prob> t |
| Intercept                    |  | 41.711674 | 0.343435  | 121.45  | <.0001* |
| Hours_Studied*Hours_Studied  |  | 0.0002839 | 0.000564  | 0.56    | 0.6147  |
| Attendance                   |  | 0.1999344 | 0.002481  | 80.60   | <.0001* |
| Parental_Involvement[Low]    |  | -0.981263 | 0.047993  | -20.45  | <.0001* |
| Parental_Involvement[Medium] |  | -0.033446 | 0.038498  | -0.87   | 0.3850  |
| Access_to_Resources[Low]     |  | -1.035359 | 0.048217  | -21.47  | <.0001* |
| Access_to_Resources[Medium]  |  | 0.0520605 | 0.038603  | 1.35    | 0.1775  |
| Previous_Score               |  | 0.0471838 | 0.001992  | 23.69   | <.0001* |
| Motivation_Level[Low]        |  | -0.541709 | 0.043402  | -12.48  | <.0001* |
| Motivation_Level[Medium]     |  | -0.005639 | 0.038584  | -0.15   | 0.8838  |
| Teacher_Quality[Low]         |  | -0.501331 | 0.063948  | -7.84   | <.0001* |
| Teacher_Quality[Medium]      |  | -0.022678 | 0.042681  | -0.53   | 0.5952  |
| Hours_Studied                |  | 0.2823147 | 0.02309   | 12.23   | <.0001* |

| Effect Tests                |       |    |                |          |          |
|-----------------------------|-------|----|----------------|----------|----------|
| Source                      | Nparm | DF | Sum of Squares | F Ratio  | Prob > F |
| Hours_Studied*Hours_Studied | 1     | 1  | 1.357          | 0.2535   | 0.6147   |
| Attendance                  | 1     | 1  | 34764.001      | 6495.932 | <.0001*  |
| Parental_Involvement        | 2     | 2  | 3164.876       | 295.6912 | <.0001*  |
| Access_to_Resources         | 2     | 2  | 3192.439       | 298.2663 | <.0001*  |
| Previous_Score              | 1     | 1  | 3003.211       | 561.1740 | <.0001*  |

# METHODS

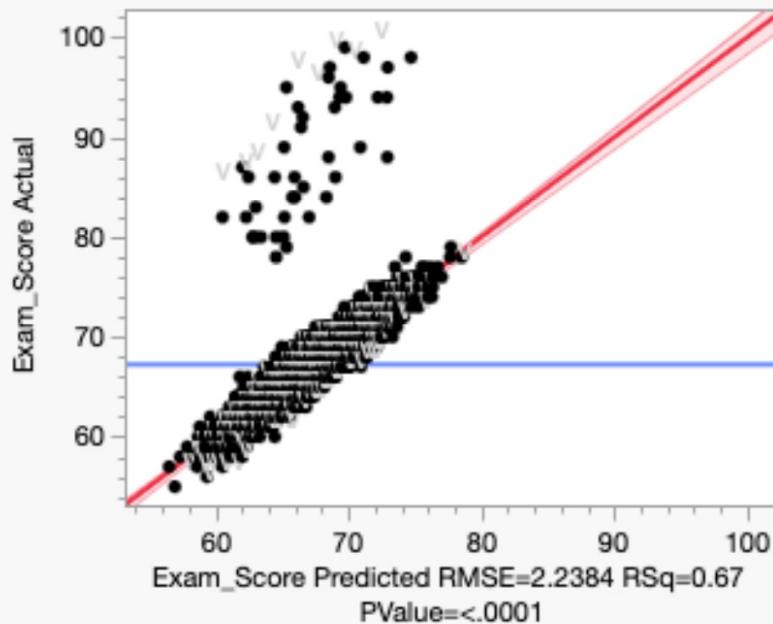
| Summary of Fit                         |       |                |                |          |          |
|--|-------|----------------|----------------|----------|----------|
| Source                                 | DF    | Sum of Squares | Mean Square    | F Ratio  | Prob > F |
| Model                                  | 13    | 64481.962      | 4960.15        | 927.3703 | <.0001*  |
| Error                                  | 6515  | 34846.258      | 5.35           |          |          |
| C. Total                               | 6528  | 99328.221      |                |          | <.0001*  |
| Analysis of Variance                   |       |                |                |          |          |
| Source                                 | DF    | Sum of Squares | Mean Square    | F Ratio  | Prob > F |
| Lack Of Fit                            | 6501  | 34815.758      | 5.35545        | 2.4582   |          |
| Pure Error                             | 14    | 30.500         | 2.17857        |          | 0.0265*  |
| Total Error                            | 6515  | 34846.258      |                |          | 0.0265*  |
|  |       |                | Max RSq        |          | 0.9997   |
| Parameter Estimates                    |       |                |                |          |          |
| Term                                   |       | Estimate       | Std Error      | t Ratio  | Prob> t  |
| Intercept                              |       | 41.627273      | 0.273621       | 152.13   | <.0001*  |
| Hours_Studied                          |       | 0.2928024      | 0.005161       | 56.74    | <.0001*  |
| Attendance                             |       | 0.1999085      | 0.00248        | 80.61    | <.0001*  |
| Parental_Involvement[Low]              |       | -0.982243      | 0.047966       | -20.48   | <.0001*  |
| Parental_Involvement[Medium]           |       | -0.030903      | 0.038504       | -0.80    | 0.4222   |
| Access_to_Resources[Low]               |       | -1.033768      | 0.048205       | -21.45   | <.0001*  |
| Access_to_Resources[Medium]            |       | 0.0514313      | 0.038592       | 1.33     | 0.1827   |
| Previous_Scores                        |       | 0.0471589      | 0.001991       | 23.69    | <.0001*  |
| Motivation_Level[Low]                  |       | -0.858106      | 0.152214       | -5.64    | <.0001*  |
| Motivation_Level[Medium]               |       | 0.0249383      | 0.134055       | 0.19     | 0.8524   |
| Teacher_Quality[Low]                   |       | -0.503521      | 0.063951       | -7.87    | <.0001*  |
| Teacher_Quality[Medium]                |       | -0.021231      | 0.042671       | -0.50    | 0.6188   |
| Hours_Studied*Motivation_Level[Low]    |       | 0.0158802      | 0.007316       | 2.17     | 0.0300*  |
| Hours_Studied*Motivation_Level[Medium] |       | -0.001454      | 0.006436       | -0.23    | 0.8213   |
| Effect Tests                           |       |                |                |          |          |
| Source                                 | Nparm | DF             | Sum of Squares | F Ratio  | Prob > F |
| Hours_Studied                          | 1     | 1              | 17216.733      | 3218.911 | <.0001*  |
| Attendance                             | 1     | 1              | 34753.932      | 6497.738 | <.0001*  |
| Parental_Involvement                   | 2     | 2              | 3161.993       | 295.5896 | <.0001*  |
| Access_to_Resources                    | 2     | 2              | 3185.165       | 297.7557 | <.0001*  |

# RESULTS

## Response Exam\_Score

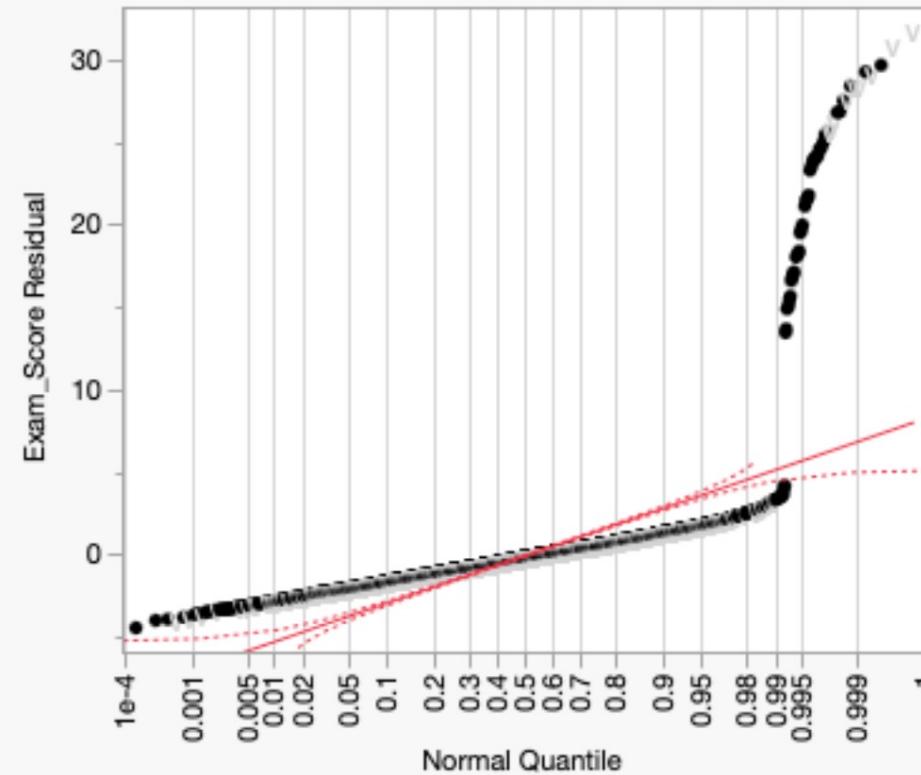
Validation: Validation

### Actual by Predicted Plot



# RESULTS

## ▼ Residual Normal Quantile Plot



# REFERENCES

- [HTTPS://GITHUB.COM/KOZAH34/PREDICTING-ACADEMIC-SUCCESS](https://github.com/kozah34/predicting-academic-success)
- QUESTIONS?