

SAS Assignment 1 – Regression

GPA Prediction Regression Model based on student's age and GRE Score

JOO YONG YOON

MI 353

Dr. CEVIKPARMAK

11 Feb 2023

Answers to the questions presented (#6~12)

* Please refer to the figure(s) provided for the question. The whole regression output is attached at the end of the report.

6. Is the regression model significant?

| | | | | | | |
|----|-----------------|----|----------------|-------------|---------|--------|
| 49 | | | | | | |
| 50 | | | | | | |
| 51 | | | | | | |
| 52 | | | | | | |
| 53 | Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| 54 | | | | | | |
| 55 | Model | 2 | 10.385763 | 5.192881 | 80.50 | <.0001 |
| 56 | Error | 82 | 5.289348 | 0.064504 | | |
| 57 | Corrected Total | 84 | 15.675111 | | | |

Figure 1. Analysis of Variance

The regression model is significant because 1) the F value is large enough and 2) the p-value is less than the significance level (.05).

7. What does the F test tell us about this model?

The associated p value of that f value is less than .05. This means that the regression model has statistical utility; overall, the model statistically makes sense.

8. Are both predictors significant?

| | | | | | | |
|----|-----------|----|----------|----------------|---------|---------|
| 67 | | | | | | |
| 68 | | | | | | |
| 69 | | | | | | |
| 70 | Parameter | DF | Estimate | Standard Error | t Value | Pr > t |
| 71 | | | | | | |
| 72 | Intercept | 1 | -0.5438 | 0.2796 | -1.94 | 0.0552 |
| 73 | Age | 1 | 0.1343 | 0.0140 | 9.58 | <.0001 |
| 74 | GRE | 1 | 0.00144 | 0.000369 | 3.90 | 0.0002 |
| 75 | | | | | | |

Figure 2. Analysis of Maximum Likelihood Estimates

Both predictors are significant. The $\text{Pr} > |t|$ column represents the p-value associated with the value in the t value column. Because the p-values of both predictors are less than a significance

level of .05, the predictor variables are said to have a statistically significant relationship with the response variable in the model.

9. Which predictor has the most influence on GPA?

Based on the output, the regression model is $\text{GPA} = -.5438 + .1343 \cdot \text{Age} + .00144 \cdot \text{GRE}$. Since the largest standardized regression coefficient is Age's, it has the most influence on GPA in this model.

10. What are the R2 and Adjusted R2 values for this model?

| | | | | |
|----|----------------------|-----------|----------|-----------|
| 60 | Model Fit Statistics | | | |
| 61 | | | | |
| 62 | R-Square | 0.6626 | Adj R-Sq | 0.6543 |
| 63 | AIC | -230.0413 | BIC | -227.8245 |
| 64 | SBC | -222.7133 | C(p) | 3.0000 |
| 65 | | | | |

Figure 3. Model Fit Statistics

R-Square is .6626, and Adjusted R-Square is .6543.

11. What does the R2 tell you?

The closer the R2 is to 100%, the better the regression model is. Adj R-Sq of .6543 means that input variables explain about 65% of the variation in the target variable.

12. Using the output of your regression, what would you predict as the final GPA of a student that was 25 years old and scored 680 on the GRE?

$$\text{GPA} = -.5438 + .1343 \cdot \text{Age} + .00144 \cdot \text{GRE}$$

$$\text{GPA} = -.5438 + .1343 \cdot 26 + .00144 \cdot 680 = 3.9272$$

Since our original data on excel only uses 2 decimal points for GPA, I rounded the GPA to 3.93.

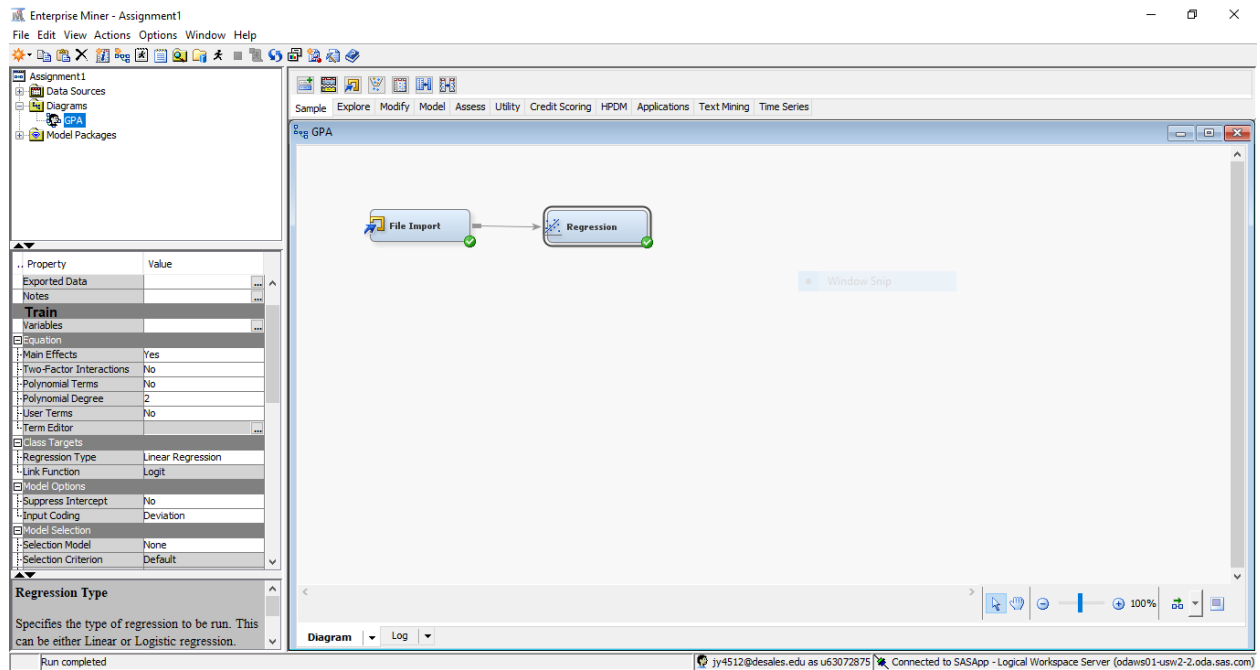


Figure 4. Diagram

The 'Variables - FIMPORT' dialog box is shown, allowing users to configure variable roles and levels for the 'FIMPORT' node. The 'Columns' section is active, displaying a table with variable details.

| Name | Role | Level | Report | Order | Drop | Lower Limit | Upper Limit |
|---------|----------|----------|--------|-------|------|-------------|-------------|
| Age | Input | Interval | No | | No | . | . |
| GPA | Target | Interval | No | | No | . | . |
| GPAPass | Rejected | Binary | No | | No | . | . |
| GRE | Input | Interval | No | | No | . | . |
| Obs | ID | Nominal | No | | No | . | . |

The dialog includes options for 'Columns' (Label, Mining, Basic, Statistics) and buttons for 'Apply', 'Reset', 'Explore...', 'OK', and 'Cancel'.

Figure 5. Edit Variables

Figure 6. Output

```

1  *-----*
2  User:                u63072875
3  Date:                12 February 2023
4  Time:                01:20:31
5  *-----*
6  * Training Output
7  *-----*
8
9
10
11
12 Variable Summary
13
14           Measurement      Frequency
15 Role           Level        Count
16
17 INPUT          INTERVAL      2
18 REJECTED       BINARY        1
19 TARGET         INTERVAL      1
20
21
22
23
24 Predicted and decision variables
25
26 Type           Variable      Label
27
28 TARGET         GPA           GPA
29 PREDICTED       P_GPA        Predicted: GPA
30 RESIDUAL        R_GPA        Residual: GPA
31
32
33
34
35
36 The DMREG Procedure
37
38           Model Information
39
40 Training Data Set              WORK.EM_DMREG.VIEW

```

```

41 DMDb Catalog          WORK.REG_DMDB
42 Target Variable       GPA (GPA)
43 Target Measurement Level Interval
44 Error                 Normal
45 Link Function         Identity
46 Number of Model Parameters 3
47 Number of Observations 85
48
49
50                         Analysis of Variance
51
52                         Sum of
53 Source                DF      Squares      Mean Square      F Value
54   Pr > F
55 Model                  2       10.385763      5.192881      80.50
56   <.0001
57 Error                  82       5.289348      0.064504
58 Corrected Total       84      15.675111
59
60                         Model Fit Statistics
61
62 R-Square              0.6626      Adj R-Sq      0.6543
63 AIC                  -230.0413      BIC          -227.8245
64 SBC                  -222.7133      C(p)          3.0000
65
66
67                         Analysis of Maximum Likelihood Estimates
68
69                         Standard
70 Parameter      DF      Estimate      Error      t Value      Pr > |t|
71
72 Intercept      1      -0.5438      0.2796      -1.94      0.0552
73 Age            1       0.1343      0.0140      9.58      <.0001
74 GRE            1       0.00144     0.000369     3.90      0.0002
75
76
77 *-----*
78 * Score Output

```

```

79 *-----*
80
81
82 *-----*
83 * Report Output
84 *-----*
85
86
87
88
89 Fit Statistics
90
91 Target=GPA Target Label=GPA
92
93     Fit
94 Statistics      Statistics Label              Train
95
96  _AIC_          Akaike's Information Criterion    -230.041
97  _ASE_          Average Squared Error             0.062
98  _AVERR_        Average Error Function            0.062
99  _DFE_          Degrees of Freedom for Error      82.000
100 _DFM_          Model Degrees of Freedom          3.000
101 _DFT_          Total Degrees of Freedom          85.000
102 _DIV_          Divisor for ASE                   85.000
103 _ERR_          Error Function                     5.289
104 _FPE_          Final Prediction Error              0.067
105 _MAX_          Maximum Absolute Error             0.894
106 _MSE_          Mean Square Error                  0.065
107 _NOBS_         Sum of Frequencies                 85.000
108 _NW_          Number of Estimate Weights          3.000
109 _RASE_         Root Average Sum of Squares        0.249
110 _RFPE_         Root Final Prediction Error        0.258
111 _RMSE_         Root Mean Squared Error            0.254
112 _SBC_          Schwarz's Bayesian Criterion      -222.713
113 _SSE_          Sum of Squared Errors              5.289
114 _SUMW_         Sum of Case Weights Times Freq     85.000
115
116
117
118

```

119 Assessment Score Rankings

120

121 Data Role=TRAIN Target Variable=GPA Target Label=GPA

122

| | | Number of | Mean | Mean |
|-------|--------------|-----------|-----------|------|
| Depth | Observations | Target | Predicted | |
| 5 | 5 | 3.6720 | 3.59754 | |
| 10 | 4 | 3.4150 | 3.53142 | |
| 15 | 4 | 3.4575 | 3.45822 | |
| 20 | 4 | 3.3100 | 3.37115 | |
| 25 | 5 | 3.1980 | 3.32868 | |
| 30 | 4 | 3.3875 | 3.24734 | |
| 35 | 4 | 2.9875 | 3.12454 | |
| 40 | 4 | 3.1450 | 3.05465 | |
| 45 | 5 | 3.0860 | 2.96940 | |
| 50 | 4 | 2.9300 | 2.91852 | |
| 55 | 4 | 3.0700 | 2.88399 | |
| 60 | 4 | 2.8975 | 2.84814 | |
| 65 | 5 | 2.7880 | 2.81111 | |
| 70 | 4 | 2.6850 | 2.78094 | |
| 75 | 4 | 2.5800 | 2.74521 | |
| 80 | 4 | 3.1575 | 2.69681 | |
| 85 | 5 | 2.4700 | 2.61939 | |
| 90 | 4 | 2.4000 | 2.54760 | |
| 95 | 4 | 2.4450 | 2.47218 | |
| 100 | 4 | 2.3675 | 2.41412 | |

146

147

148

149

150 Assessment Score Distribution

151

152 Data Role=TRAIN Target Variable=GPA Target Label=GPA

153

| Range for | Mean | Mean | Number of | Model |
|---------------|---------|-----------|--------------|---------|
| Predicted | Target | Predicted | Observations | Score |
| 3.610 - 3.677 | 3.70000 | 3.67733 | 1 | 3.64378 |
| 3.543 - 3.610 | 3.66500 | 3.57760 | 4 | 3.57670 |

| | | | | | |
|-----|---------------|---------|---------|---|---------|
| 159 | 3.476 - 3.543 | 3.43200 | 3.52573 | 5 | 3.50962 |
| 160 | 3.409 - 3.476 | 3.44333 | 3.44331 | 3 | 3.44253 |
| 161 | 3.342 - 3.409 | 3.32200 | 3.36691 | 5 | 3.37545 |
| 162 | 3.275 - 3.342 | 3.19400 | 3.31744 | 5 | 3.30836 |
| 163 | 3.208 - 3.275 | 3.40000 | 3.26506 | 2 | 3.24128 |
| 164 | 3.141 - 3.208 | 3.34000 | 3.16465 | 2 | 3.17419 |
| 165 | 3.074 - 3.141 | 3.00600 | 3.09896 | 5 | 3.10711 |
| 166 | 3.006 - 3.074 | 3.08750 | 3.01928 | 4 | 3.04003 |
| 167 | 2.939 - 3.006 | 3.14500 | 2.94969 | 2 | 2.97294 |
| 168 | 2.872 - 2.939 | 2.98250 | 2.90969 | 8 | 2.90586 |
| 169 | 2.805 - 2.872 | 2.90250 | 2.84190 | 8 | 2.83877 |
| 170 | 2.738 - 2.805 | 2.65444 | 2.77168 | 9 | 2.77169 |
| 171 | 2.671 - 2.738 | 3.05200 | 2.70418 | 5 | 2.70460 |
| 172 | 2.604 - 2.671 | 2.47667 | 2.63939 | 3 | 2.63752 |
| 173 | 2.537 - 2.604 | 2.41600 | 2.57553 | 5 | 2.57044 |
| 174 | 2.470 - 2.537 | 2.63333 | 2.48463 | 3 | 2.50335 |
| 175 | 2.403 - 2.470 | 2.28800 | 2.44943 | 5 | 2.43627 |
| 176 | 2.336 - 2.403 | 2.35000 | 2.33564 | 1 | 2.36918 |