Membandingkan kebaikan model dari 4 macam pre-processing dengan menggunakan *Phyton* pada analisis regresi dengan variabel sebagai berikut :

Y = Final

X1 = Assignment

X2 = Tutorial

X3 = Midterm

X4 = Take Home

1. ***MISSING VALUE***

**Regression Analysis: Final versus Assignment; Tutorial; Midterm; TakeHome**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 4 18616,9 4654,2 29,24 0,000

Assignment 1 19,2 19,2 0,12 0,729

Tutorial 1 332,4 332,4 2,09 0,152

Midterm 1 10821,4 10821,4 67,98 0,000

TakeHome 1 426,6 426,6 2,68 0,105

Error 94 14963,8 159,2

Total 98 33580,7

Model Summary

S R-sq R-sq(adj) R-sq(pred)

12,6170 55,44% 53,54% 50,49%

Regression Equation

Final = 2,3 + 0,044 Assignment + 0,1363 Tutorial + 0,6055 Midterm + 0,1100 TakeHome

Berdasarkan hasil analisis regresi linear berganda dengan mengatasi *missing value* menggunakan program minitab, didapatkan bahwa nilai *R-sq* sebesar 55,44% artinya bahwa variasi keragaman total *interaction* (Y) yang mampu dijelaskan oleh variabel prediktor dalam model sebesar 55,44% dan 44,56% sisanya dijelaskan oleh variabel lain diluar model dengan nilai *p-value* sebesar 0,000 dan . Karena nilai , sehingga dapat disimpulkan bahwa minimal ada satu variabel prediktor yang signifikan terhadap model. Namun, untuk uji parsial didapatkan terdapat nilai *p-value* yang lebih dari α = 0,05 yaitu pada variabel *assignment* (X1), *tutorial* (X2) dan *take home* (X4), hal ini berarti ketiga variabel tidak berpengaruh signifikan terhadap model, sedangkan variabel *midterm* berpengaruh signifikan terhadap model.

1. ***MISSING VALUE, OUTLIER***

**Regression Analysis: Final\_1 versus Assignment\_1;**

**Tutorial\_1; Midterm\_1; TakeHome\_1**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 4 18263,3 4565,8 27,89 0,000

Assignment\_1 1 55,3 55,3 0,34 0,563

Tutorial\_1 1 269,9 269,9 1,65 0,202

Midterm\_1 1 10354,6 10354,6 63,25 0,000

TakeHome\_1 1 392,8 392,8 2,40 0,125

Error 91 14897,2 163,7

Total 95 33160,5

Model Summary

S R-sq R-sq(adj) R-sq(pred)

12,7948 55,08% 53,10% 49,35%

Regression Equation

Final\_1 = -0,9 + 0,084 Assignment\_1 + 0,140 Tutorial\_1 + 0,6004 Midterm\_1 + 0,1062 TakeHome\_1

Berdasarkan hasil analisis regresi linear berganda dengan mengatasi *missing value* dan *outlier* menggunakan program minitab, didapatkan bahwa nilai *R-sq* sebesar 55,04% artinya bahwa variasi keragaman total *interaction* (Y) yang mampu dijelaskan oleh variabel prediktor dalam model sebesar 55,04% dan 44,96% sisanya dijelaskan oleh variabel lain diluar model dengan nilai *p-value* sebesar 0,000 dan . Karena nilai , sehingga dapat disimpulkan bahwa minimal ada satu variabel prediktor yang signifikan terhadap model. Namun, untuk uji parsial didapatkan terdapat nilai *p-value* yang lebih dari α = 0,05 yaitu pada variabel *assignment* (X1), *tutorial* (X2) dan *take home* (X4), hal ini berarti ketiga variabel tidak berpengaruh signifikan terhadap model, sedangkan variabel *midterm* berpengaruh signifikan terhadap model.

1. ***MISSING VALUE, TRANSFORMATION***

**Regression Analysis: Final\_2 versus Assignment\_2; Tutorial\_2; Midterm\_2; TakeHome\_2**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 4 0,40151 0,100377 70,17 0,000

Assignment\_2 1 0,23825 0,238254 166,55 0,000

Tutorial\_2 1 0,22844 0,228443 159,69 0,000

Midterm\_2 1 0,09879 0,098792 69,06 0,000

TakeHome\_2 1 0,20838 0,208382 145,67 0,000

Error 94 0,13447 0,001431

Total 98 0,53598

Model Summary

S R-sq R-sq(adj) R-sq(pred)

0,0378224 74,91% 73,84% 68,94%

Regression Equation

Final\_2 = 2,123 - 1,1946 Assignment\_2 - 1,0999 Tutorial\_2 - 0,7426 Midterm\_2- 0,7598 TakeHome\_2

Berdasarkan hasil analisis regresi linear berganda dengan mengatasi *missing value* dan *transformation* menggunakan program minitab, didapatkan bahwa nilai *R-sq* sebesar 74,91% artinya bahwa variasi keragaman total *interaction* (Y) yang mampu dijelaskan oleh variabel prediktor dalam model sebesar 74,91% dan 25,09% sisanya dijelaskan oleh variabel lain diluar model dengan nilai *p-value* sebesar 0,000 dan . Karena nilai , sehingga dapat disimpulkan bahwa minimal ada satu variabel prediktor yang signifikan terhadap model. Untuk uji parsial tidak terdapat nilai *p-value* yang lebih dari α = 0,05 sehingga, semua variabel prediktorberpengaruh signifikan terhadap model.

1. ***MISSING VALUE, OUTLIER, TRANSFORMATION***

**Regression Analysis: Final\_3 versus Assignment\_3; Tutorial\_3; Midterm\_3; TakeHome\_3**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 4 0,3977 0,099431 85,98 0,000

Assignment\_3 1 0,2514 0,251386 217,37 0,000

Tutorial\_3 1 0,2504 0,250420 216,54 0,000

Midterm\_3 1 0,1146 0,114644 99,13 0,000

TakeHome\_3 1 0,2372 0,237204 205,11 0,000

Error 91 0,1052 0,001156

Total 95 0,5030

Model Summary

S R-sq R-sq(adj) R-sq(pred)

0,0340071 79,08% 78,16% 75,13%

Regression Equation

Final\_3 = 2,365 - 1,3724 Assignment\_3 - 1,2542 Tutorial\_3 - 0,8194 Midterm\_3- 0,8653 TakeHome\_3

Berdasarkan hasil analisis regresi linear berganda dengan mengatasi *missing value, outlier* dan *transformation* menggunakan program minitab, didapatkan bahwa nilai *R-sq* sebesar 79,08% artinya bahwa variasi keragaman total *interaction* (Y) yang mampu dijelaskan oleh variabel prediktor dalam model sebesar 79,08% dan 20,92% sisanya dijelaskan oleh variabel lain diluar model dengan nilai *p-value* sebesar 0,000 dan . Karena nilai , sehingga dapat disimpulkan bahwa minimal ada satu variabel prediktor yang signifikan terhadap model. Untuk uji parsial tidak terdapat nilai *p-value* yang lebih dari α = 0,05 sehingga, semua variabel prediktorberpengaruh signifikan terhadap model.

Setelah dilakukan analisis regresi berganda dari 4 macam *pre-processing*, maka didapatkan model terbaik adalah hasil analisis regresi dengan *pre-processing* yang mengatasi *missing value, outlier* dan *transformation* karena memiliki nilai *R-sq* yang paling besar (79,08%)dengan model sebagai berikut.

Final = 2,365 - 1,3724 Assignment- 1,2542 Tutorial-0,8194 Midterm- 0,8653 TakeHome

Kesimpulan

Sebelum melakukan pengolahan data perlu dilakukan *pre processing* data. Hal ini bertujuan untuk memastikan data yang akan diolah adalah data yang tepat (tidak terdapat *missing value*, *outlier*, dll). Karena jika pada data masih terdapat *missing value, outlier*, dll bisa jadi model yang didapatkan kurang baik.