**TUGAS 3**

1. **Bandingkan kebaikan model yang dihasilkan**
2. Missing value

**Regression Analysis: distance versus consume; speed; ...**

The regression equation is

distance = - 108 + 8,44 consume + 2,27 speed - 0,28 temp\_inside

+ 0,532 temp\_outside + 10,3 AC - 19,7 rain - 21,8 sun

+ 0,704 refill liters

52 cases used, 336 cases contain missing values

Predictor Coef SE Coef T P

Constant -108,21 38,51 -2,81 0,007

consume 8,440 3,067 2,75 0,009

speed 2,2718 0,3309 6,87 0,000

temp\_inside -0,275 1,149 -0,24 0,812

temp\_outside 0,5323 0,7242 0,74 0,466

AC 10,32 12,19 0,85 0,402

rain -19,67 11,96 -1,64 0,107

sun -21,84 12,97 -1,68 0,100

refill liters 0,7036 0,3612 1,95 0,058

S = 20,2405 R-Sq = 55,5% R-Sq(adj) = 47,2%

1. Missing value, outlier

**Regression Analysis: distance versus consume; speed; ...**

The regression equation is

distance = 18,4 - 4,47 consume + 0,004 speed + 0,88 temp\_inside

- 0,303 temp\_outside + 4,05 AC + 2,03 sun

30 cases used, 303 cases contain missing values

Predictor Coef SE Coef T P

Constant 18,38 23,36 0,79 0,439

consume -4,472 1,222 -3,66 0,001

speed 0,0044 0,1498 0,03 0,977

temp\_inside 0,884 1,038 0,85 0,403

temp\_outside -0,3031 0,2479 -1,22 0,234

AC 4,055 3,467 1,17 0,254

sun 2,031 3,317 0,61 0,547

S = 3,88755 R-Sq = 62,5% R-Sq(adj) = 52,7%

1. Missing value, transformasi

**Regression Analysis: distance versus consume; speed; ...**

The regression equation is

distance = 0,0165 - 0,315 consume - 0,959 speed - 2,08 temp\_inside

- 1,26 temp\_outside + 0,124 specials

Predictor Coef SE Coef T P

Constant 0,0165068 0,0003177 51,95 0,000

consume -0,31472 0,02703 -11,64 0,000

speed -0,95926 0,01409 -68,07 0,000

temp\_inside -2,08488 0,02396 -87,00 0,000

temp\_outside -1,26186 0,04737 -26,64 0,000

specials 0,1240 0,1529 0,81 0,418

S = 0,0000826102 R-Sq = 96,4% R-Sq(adj) = 96,3%

1. Missing value, outlier, transformasi

**Regression Analysis: distance versus consume; speed; ...**

The regression equation is

distance = 0,0165 - 0,315 consume - 0,959 speed - 2,08 temp\_inside

- 1,26 temp\_outside + 0,124 specials

Predictor Coef SE Coef T P

Constant 0,0165068 0,0003177 51,95 0,000

consume -0,31472 0,02703 -11,64 0,000

speed -0,95926 0,01409 -68,07 0,000

temp\_inside -2,08488 0,02396 -87,00 0,000

temp\_outside -1,26186 0,04737 -26,64 0,000

specials 0,1240 0,1529 0,81 0,418

S = 0,0000826102 R-Sq = 96,4% R-Sq(adj) = 96,3%

1. **Bandingkan kebaikan model yang dihasilkan**

Semakin kecil nilai koefisien determinasi (R-square) maka pengaruh variabel bebas terhadap variabel terikat semakin lemah. Sebaliknya, jika nilai R Square semakin mendekati 1, maka pengaruh tersebut akan semakin kuat. Dari perhitungan regresi diatas didapatkan hasil r-square dari 4 data yaitu sebagai berikut :

|  |  |
| --- | --- |
| Data | R-Sq |
| missing value | 55,5% |
| missing value, outlier | 62,5% |
| missing value, transformasi | 96,4% |
| missing value, outlier, transformasi | 96,4% |

Dari tabel berikut diketahui bahwa nilai R square yang mendekati 1 berada pada data missing value, transformasi dan data missing value, outlier, transformasi maka semakin baik data semakin baik juga kebaikan model yang dihasilkan.

1. **Kesimpulan**

Kesimpulannya yaitu sebelum menganalisis data lebih jauh perlu dilakukan pengecekan preprocessing data seperti mendeteksi missing value dan outlier, agar data tersebut baik tidak terdapat missing value dan outlier/ kecacatan. Karena semakin baik data semakin baik juga kebaikan model yang akan dihasilkan.