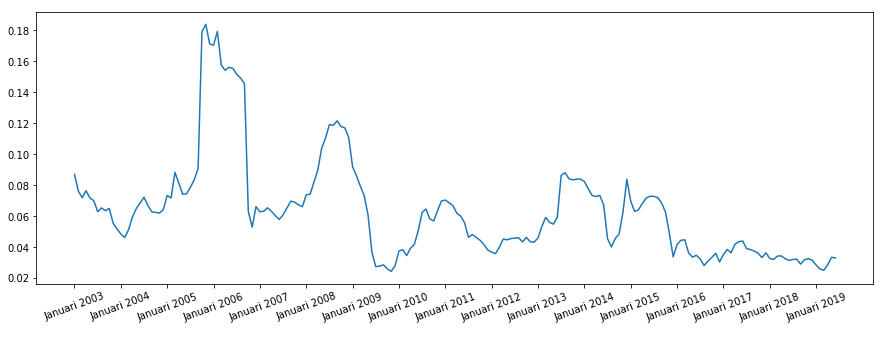
**HASIL ANALISIS DATA**

**SOAL UJIAN DATA ANALYST**

**JAGOAN HOSTING**

1. Untuk analisis data inflasi akan digunakan analisis time series. Analisis ini digunakan untuk menyelidiki pergerakan data berdasarkan waktu. Selain itu, analisis juga dapat digunakan untuk memprediksi data pada beberapa waktu ke depan. Berikut adalah grafik dari data inflasi.



Berdasarkan grafik di atas, dapat dilihat bahwa data inflasi memiliki trend dan tidak stasioner terhadap mean. Untuk lebih jelasnya, akan dilakukan ADF Unit Root Test dari data inflasi.

* Hipotesis:

: Data inflasi tidak stasioner terhadap mean.

: Data inflasi stasioner terhadap mean.

* Tingkat Signifikansi:

: .

* Statistik Uji:

P-value : 0.38987669092034716

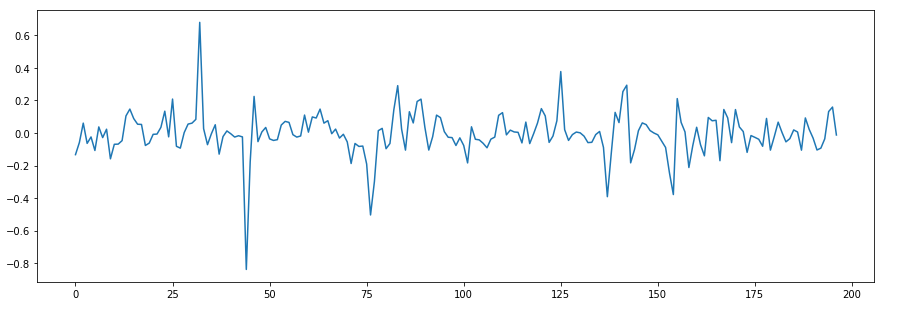
* Daerah Kritik:

ditolak jika p-value .

* Kesimpulan:

Karena p-value , maka tidak ditolak sehingga dapat disimpulkan bahwa data inflasi tidak stasioner terhadap mean.

Karena data inflasi tidak stasioner terhadap mean, maka akan dilakukan differencing dan transformasi log dari data inflasi. Untuk ke depannya, yang dimaksud dengan transformasi adalah differencing dan transformasi log dari data. Berikut grafik dari data inflasi yang sudah ditransformasi.



Berdasarkan grafik di atas, dapat dilihat bahwa transformasi data inflasi sudah stasioner terhadap mean. Untuk lebih jelasnya, akan dilakukan ADF Unit Root Test dari transformasi data inflasi.

* Hipotesis:

: Transformasi data inflasi tidak stasioner terhadap mean.

: Transformasi data inflasi stasioner terhadap mean.

* Tingkat Signifikansi:

: .

* Statistik Uji:

P-value : 9.412082670843295e-09

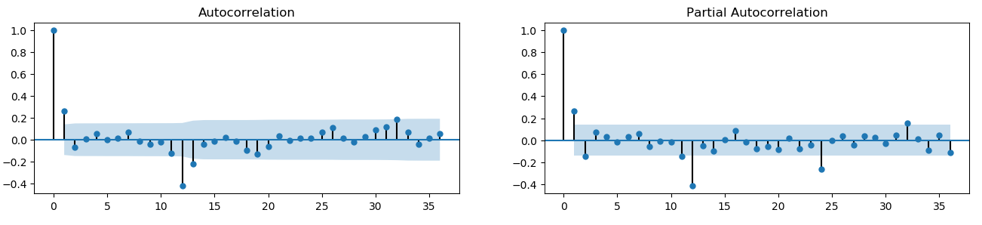
* Daerah Kritik:

ditolak jika p-value .

* Kesimpulan:

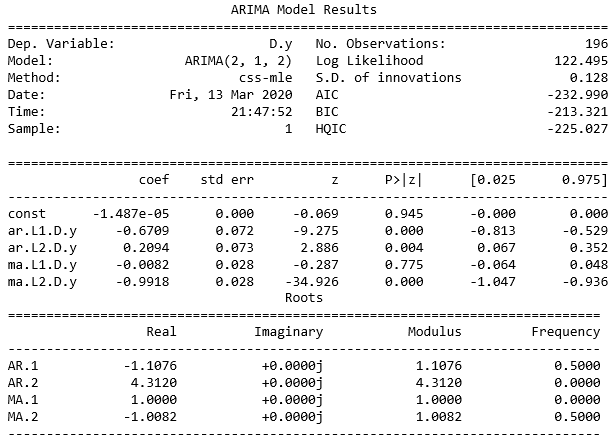
Karena p-value , maka ditolak sehingga dapat disimpulkan bahwa transformasi data inflasi stasioner terhadap mean.

Karena data sudah stasioner terhadap mean, akan ditentukan model terbaik yang dapat menggambarkan data inflasi. Untuk analisis ini akan digunakan model ARIMA. Berikut adalah plot ACF dan PACF dari transformasi data inflasi.

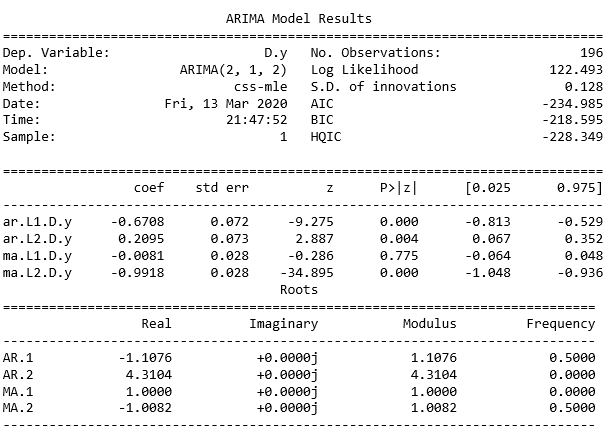


Diperhatikan 4 lag pertama dari ACF dan PACF. Karena lag ke-2 dari ACF dan PACF melebihi interval konfidensi (daerah yang diarsir biru), maka dipilih model ARIMA(2,1,2) sebagai model awal pada analisis ini. Selanjutnya, akan diselidiki mana saja model ARIMA yang signifikan. Model ARIMA dikatakan signifikan jika p-value dari semua variable model kurang dari (0.05). Berikut adalah output dari tes beberapa model ARIMA.

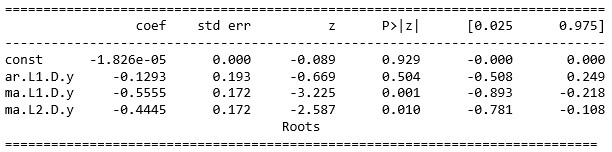
1. ARIMA(2,1,2) dengan konstan.



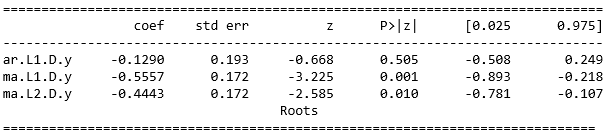
1. ARIMA(2,1,2) tanpa konstan.



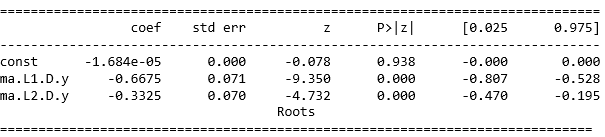
1. ARIMA(1,1,2) dengan konstan.



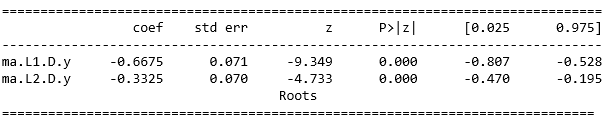
1. ARIMA(1,1,2) tanpa konstan.



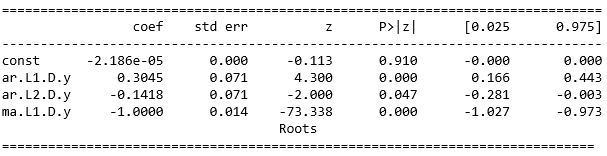
1. ARIMA(0,1,2) dengan konstan.



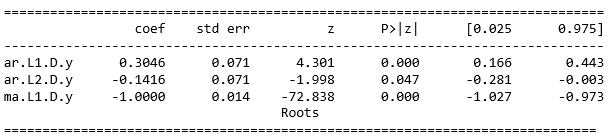
1. ARIMA(0,1,2) tanpa konstan.



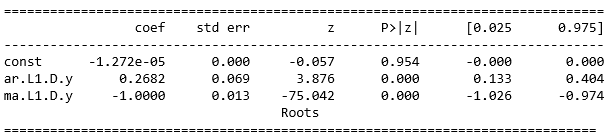
1. ARIMA(2,1,1) dengan konstan.



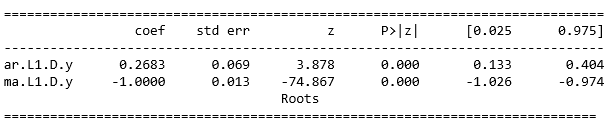
1. ARIMA(2,1,1) tanpa konstan.



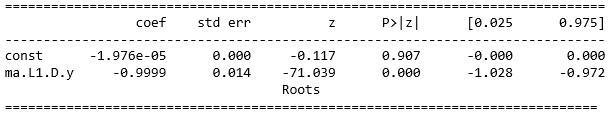
1. ARIMA(1,1,1) dengan konstan.



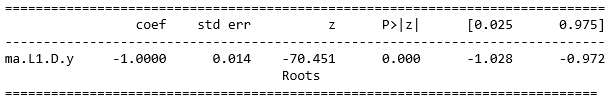
1. ARIMA(1,1,1) tanpa konstan.



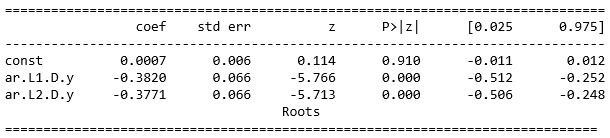
1. ARIMA(0,1,1) dengan konstan.



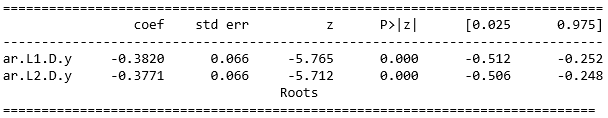
1. ARIMA(0,1,1) tanpa konstan.



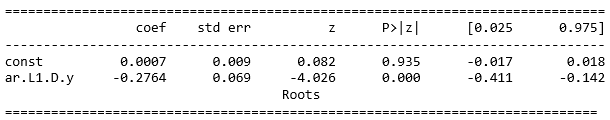
1. ARIMA(2,1,0) dengan konstan.



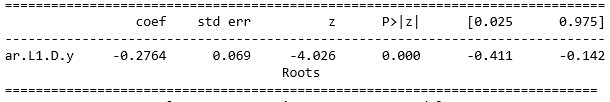
1. ARIMA(2,1,0) tanpa konstan.



1. ARIMA(1,1,0) dengan konstan.

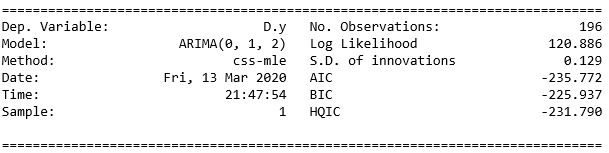


1. ARIMA(1,1,0) tanpa konstan.

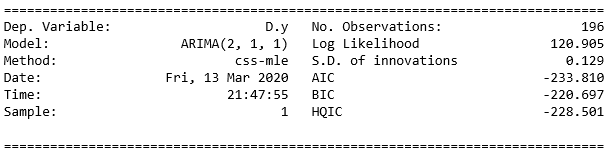


Berdasarkan output-output di atas, dapat dilihat bahwa model ARIMA(0,1,2) tanpa konstan, ARIMA(2,1,1) tanpa konstan, ARIMA(1,1,1) tanpa konstan, ARIMA(0,1,1) tanpa konstan, ARIMA(2,1,0) tanpa konstan, dan ARIMA(1,1,0) tanpa konstan merupakan model-model yang signifikan, karena p-value dari semua variabelnya kurang dari . Dari keenam model tersebut, akan dicari model yang terbaik dengan membandingkan nilai loglikelihood, AIC, dan BIC setiap model. Model terbaik adalah ketika ketiga nilai tersebut paling rendah dibandingkan dengan model lainnya. Diperhatikan ketiga nilai tersebut pada masing-masing model berikut.

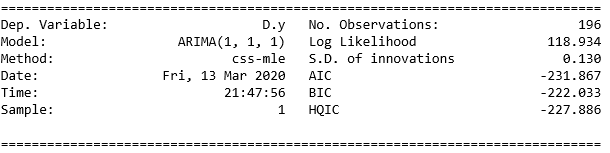
1. ARIMA(0,1,2) tanpa konstan.



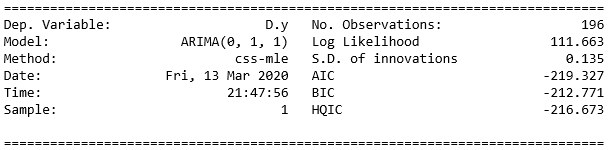
1. ARIMA(2,1,1) tanpa konstan.



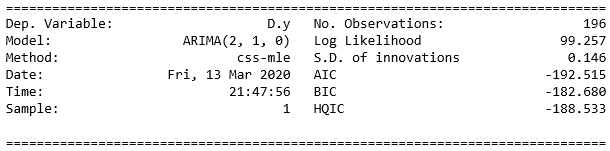
1. ARIMA(1,1,1) tanpa konstan.



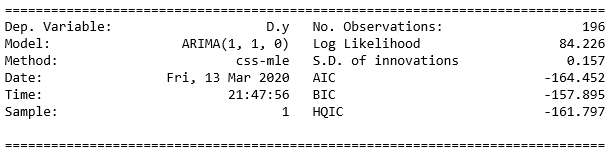
1. ARIMA(0,1,1) tanpa konstan.



1. ARIMA(2,1,0) tanpa konstan.



1. ARIMA(1,1,0) tanpa konstan.



Berdasarkan output di atas, dapat dilihat bahwa nilai loglikelihood, AIC, dan BIC terendah adalah model ARIMA(1,1,0) tanpa konstan sehingga dipilih model ARIMA(1,1,0) tanpa konstan sebagai model terbaik. Dengan demikian, diperoleh model data inflasi

dengan

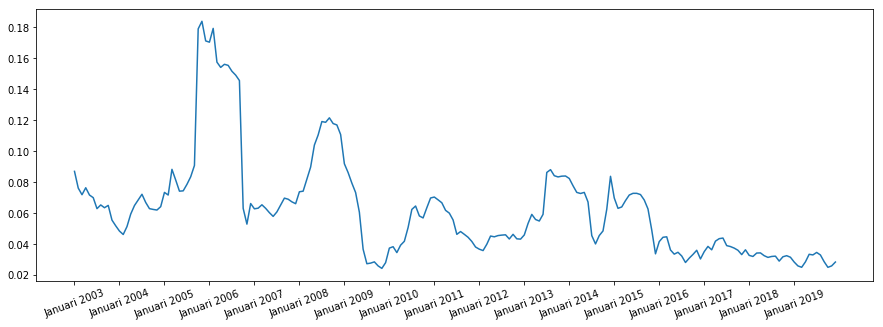
: inflasi pada bulan ke - .

: error pada waktu ke – .

Selanjutnya, akan dilakukan forecasting untuk bulan Juli 2019 sampai Desember 2019. Diperoleh inflasi selama bulan Juli 2019 sampai Desember 2019

1. Juli 2019 : 3.4395%.
2. Agustus 2019 : 3.2767%.
3. September 2019 : 2.8403%.
4. Oktober 2019 : 2.4775%.
5. November 2019 : 2.5707%.
6. Desember 2019 : 2.8198%.

Jika data inflasi (beserta hasil forecast) diplot Kembali, diperoleh grafik data inflasi



dimana hasil forecast tidak terlalu menyimpang dari data sebelumnya.

1. Untuk analisis mtcars, akan digunakan analisis regresi linear berganda. Analisis ini digunakan untuk mengetahui apakah ada hubungan antara variable dependen (mpg) dengan variable independent lainnya. Selain itu, analisis juga dapat digunakan untuk memprediksi nilai mpg jika diketahui nilai dari variable-variabel independent. Terlebih dahulu akan diselidiki asumsi normalitas dari variable mpg menggunakan tes Shapiro-Wilk.

* Hipotesis:

: Data mpg berdistribusi normal.

: Data mpg tidak berdistribusi normal.

* Tingkat Signifikansi:

: .

* Statistik Uji:

P-value : 0.12288373708724976

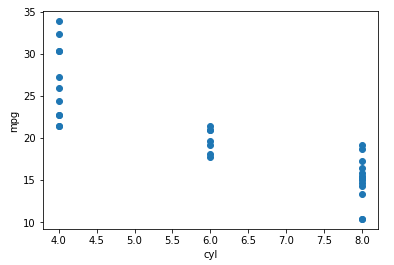
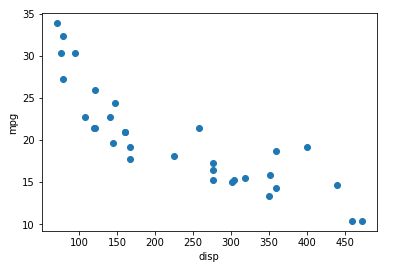
* Daerah Kritik:

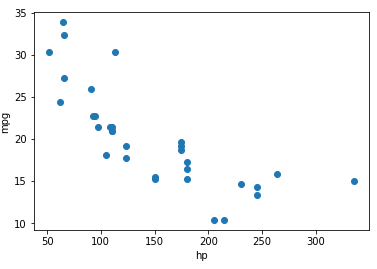
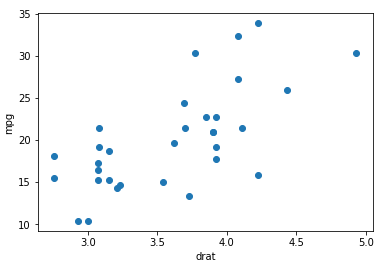
ditolak jika p-value .

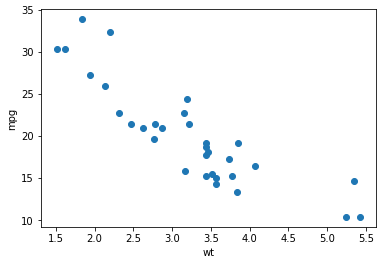
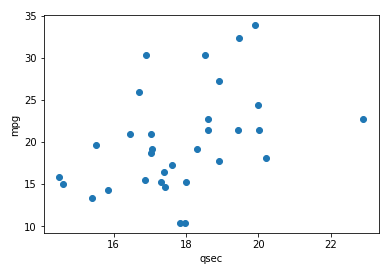
* Kesimpulan:

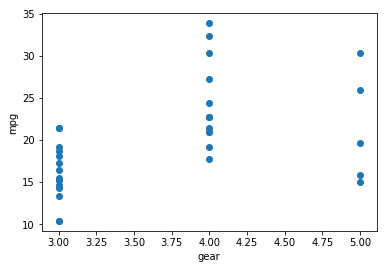
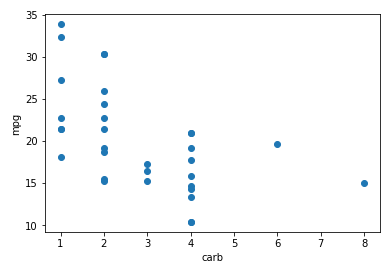
Karena p-value , maka tidak ditolak sehingga dapat disimpulkan bahwa data mpg berdistribusi normal.

Selanjutnya, diperhatikan scatter plot antara variable mpg dengan variable lainnya.

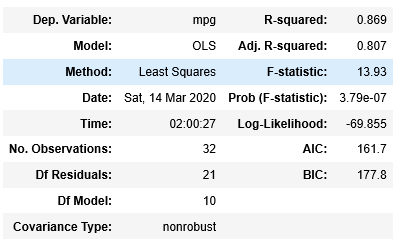
 

Berdasarkan scatter plot di atas, diperoleh hubungan antara variabel mpg dengan variabel lainnya sebagai berikut:

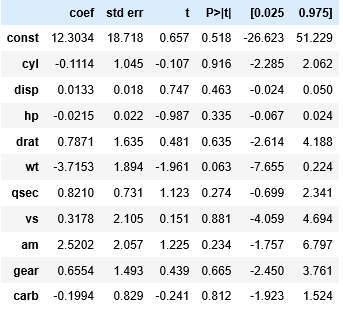
* Terdapat hubungan linear negative antara variable mpg dengan variable cyl. Artinya semakin besar nilai cyl, maka semakin kecil nilai mpg.
* Terdapat hubungan linear negative antara variable mpg dengan variable disp. Artinya semakin besar nilai disp, maka semakin kecil nilai mpg.
* Terdapat hubungan linear negative antara variable mpg dengan variable hp. Artinya semakin besar nilai hp, maka semakin kecil nilai mpg.
* Terdapat hubungan linear positif antara variable mpg dengan variable drat. Artinya semakin besar nilai drat, maka semakin besar nilai mpg.
* Terdapat hubungan linear negative antara variable mpg dengan variable wt. Artinya semakin besar nilai wt, maka semakin kecil nilai mpg.
* Terdapat hubungan linear positif antara variable mpg dengan variable qsec. Artinya semakin besar nilai qsec, maka semakin besar nilai mpg.
* Terdapat hubungan linear positif antara variable mpg dengan variable gear. Artinya semakin besar nilai gear, maka semakin kecil nilai mpg.
* Terdapat hubungan linear negative antara variable mpg dengan variable carb. Artinya semakin besar nilai carb, maka semakin kecil nilai mpg.

Karena terdapat hubungan linear antara variable mpg dengan variable lainnya, maka analisis regresi linear berganda dapat dilakukan. Selanjutnya, akan dilakukan uji overall dan uji parsial. Uji overall digunakan untuk menyelidiki apakah model yang dibentuk signifikan atau tidak, sedangkan uji parsial digunakan untuk menyelidiki mana saja variable yang signifikan terhadap model. Pertama, semua variable dimasukkan ke dalam model. Kemudian tentukan variable yang tidak signifikan (p-value ) dan memiliki nilai p-value terbesar. Pada iterasi berikutnya, variable tersebut akan dikeluarkan dari model, kemudian dianalisis kembali. Diperhatikan model-model yang terbentuk berikut.

1. Model 1

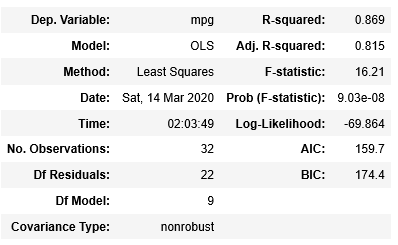


Karena nilai Prob (F-statistics) adalah 3.79e-07 , maka model signifikan. Selanjutnya, diperhatikan output berikut.



Karena variable cyl tidak signifikan dan p-value nya adalah yang terbesar, maka variable cyl dikeluarkan dari model 1.

1. Model 2

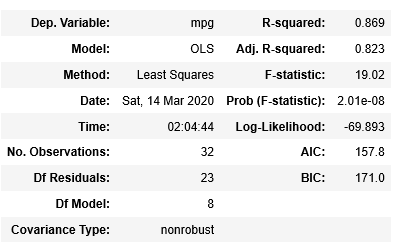


Karena nilai Prob (F-statistics) adalah 9.03e-08, maka model signifikan. Selanjutnya, diperhatikan output berikut.

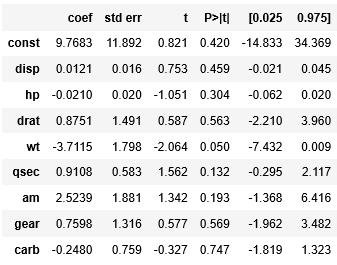


Karena variable vs tidak signifikan dan p-value nya adalah yang terbesar, maka variable vs dikeluarkan dari model 2.

1. Model 3

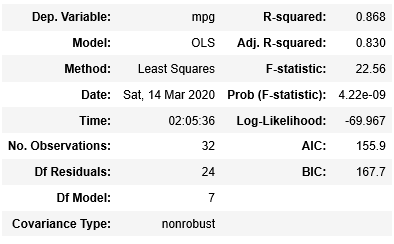


Karena nilai Prob (F-statistics) adalah 2.01e-08, maka model signifikan. Selanjutnya, diperhatikan output berikut.

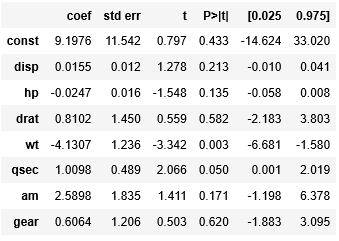


Karena variable carb tidak signifikan dan p-value nya adalah yang terbesar, maka variable carb dikeluarkan dari model 3.

1. Model 4

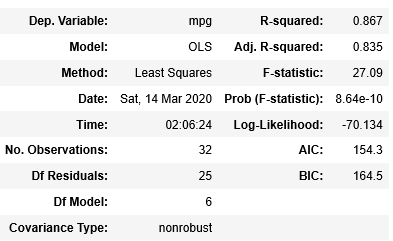


Karena nilai Prob (F-statistics) adalah 4.22e-09, maka model signifikan. Selanjutnya, diperhatikan output berikut.

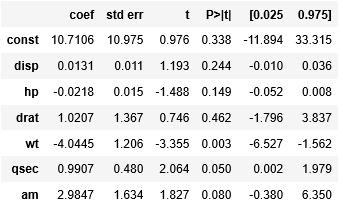


Karena variable gear tidak signifikan dan p-value nya adalah yang terbesar, maka variable gear dikeluarkan dari model 4.

1. Model 5

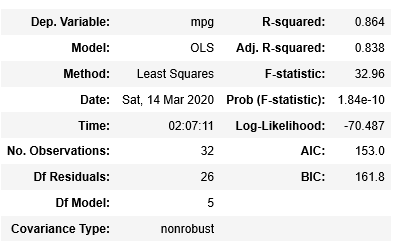


Karena nilai Prob (F-statistics) adalah 8.64e-10, maka model signifikan. Selanjutnya, diperhatikan output berikut.

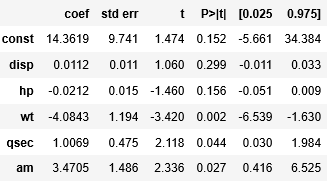


Karena variable drat tidak signifikan dan p-value nya adalah yang terbesar, maka variable drat dikeluarkan dari model 5.

1. Model 6

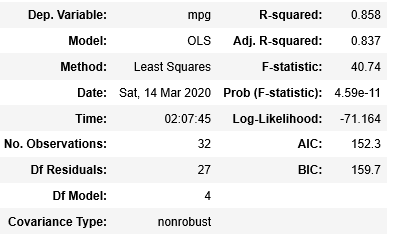


Karena nilai Prob (F-statistics) adalah 1.84e-10, maka model signifikan. Selanjutnya, diperhatikan output berikut.



Karena variable disp tidak signifikan dan p-value nya adalah yang terbesar, maka variable disp dikeluarkan dari model 6.

1. Model 7

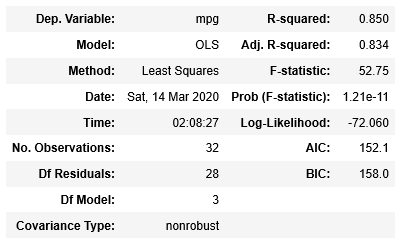


Karena nilai Prob (F-statistics) adalah 4.59e-11, maka model signifikan. Selanjutnya, diperhatikan output berikut.

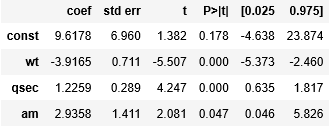


Karena variable hp tidak signifikan dan p-value nya adalah yang terbesar, maka variable hp dikeluarkan dari model 7.

1. Model 8

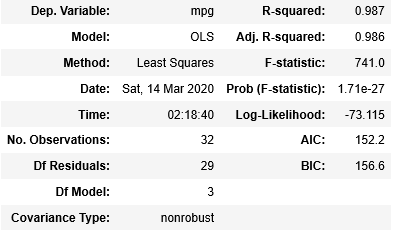


Karena nilai Prob (F-statistics) adalah 1.21e-11, maka model signifikan. Selanjutnya, diperhatikan output berikut.

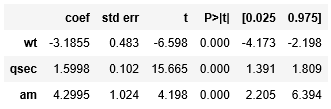


Karena konstanta tidak signifikan dan p-value nya adalah yang terbesar, maka konstanta dikeluarkan dari model 8.

1. Model 9



Karena nilai Prob (F-statistics) adalah 1.71e-27, maka model signifikan. Selanjutnya, diperhatikan output berikut.



Karena semua variable sudah signifikan, maka iterasi dihentikan.

Selanjutnya, akan dipilih model yang terbaik dengan membandingkan nilai dari . Model terbaik adalah model yang memiliki nilai terbesar dan terkecil. Diperhatikan table berikut.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 0.869 | 0.807 | 161.7 | 177.8 |
| 0.869 | 0.815 | 159.7 | 174.4 |
| 0.869 | 0.823 | 157.8 | 171.0 |
| 0.868 | 0.830 | 155.9 | 167.7 |
| 0.867 | 0.835 | 154.3 | 164.5 |
| 0.864 | 0.838 | 153.0 | 161.8 |
| 0.858 | 0.837 | 152.3 | 159.7 |
| 0.850 | 0.834 | 152.1 ✓ | 158.0 |
| 0.987 ✓ | 0.986 ✓ | 152.2 | 156.6 ✓ |

Karena model ke 9 memenuhi kriteria pemilihan model terbaik terbanyak, maka model 9 merupakan model terbaik. Dengan demikian, diperoleh model untuk mtcars adalah

dengan

: variable mpg

: variable wt

: variable qsec

: variable am

Artinya, nilai variable mpg hanya dipengaruhi oleh variable wt, qsec, dan am, dan:

* Setiap penambahan 1 satuan wt, mengakibatkan pengurangan mpg sebesar 3.1855 dengan mengasumsikan variable lain konstan.
* Setiap penambahan 1 satuan qsec, mengakibatkan penambahan mpg sebesar 1.5998 dengan mengasumsikan variable lain konstan.
* Setiap penambahan 1 satuan am, mengakibatkan penambahan mpg sebesar 4.2995 dengan mengasumsikan variable lain konstan.

Akan tetapi, menurut saya model ini belum tentu merupakan model yang baik karena dalam pembuatan model ini belum mempertimbangkan variable mana saja yang memang benar-benar penting sehingga perlu penelitian lebih lanjut.