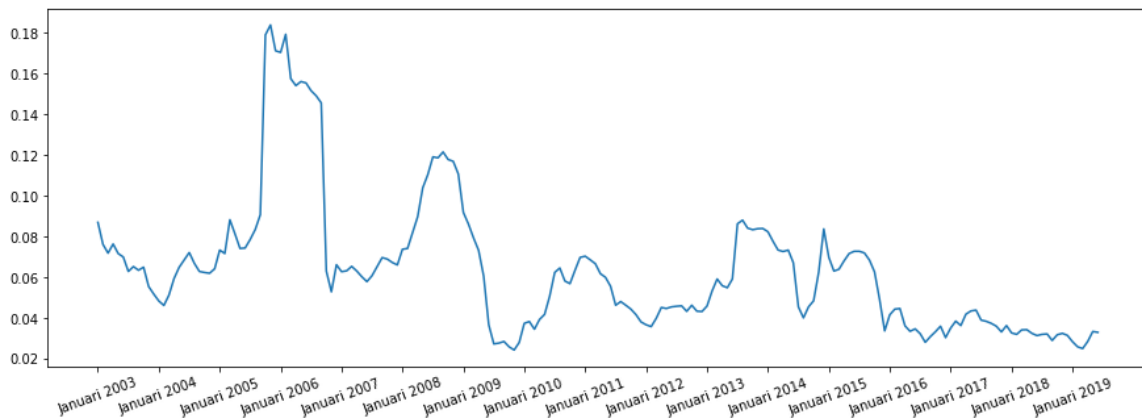


HASIL ANALISIS DATA
SOAL UJIAN DATA ANALYST
JAGOAN HOSTING

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:

H_0 : Data inflasi tidak stasioner terhadap mean.

H_1 : Data inflasi stasioner terhadap mean.

- Tingkat Signifikansi:

α : 0.05.

- Statistik Uji:

P-value : 0.38987669092034716

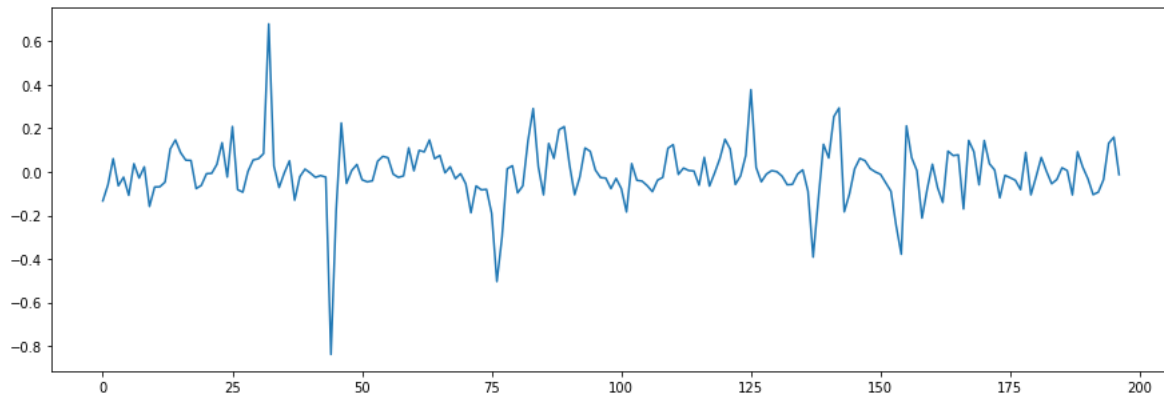
- Daerah Kritik:

H_0 ditolak jika $p\text{-value} < \alpha$.

- Kesimpulan:

Karena $p\text{-value} > \alpha$, maka H_0 tidak ditolak sehingga dapat disimpulkan bahwa data inflasi tidak stasioner terhadap mean.

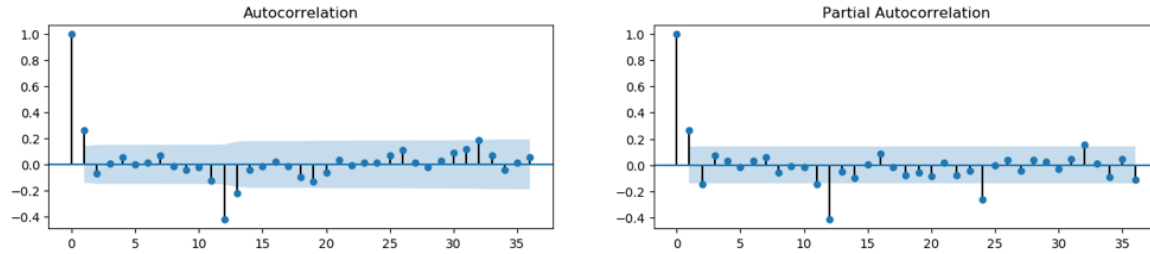
Karena data inflasi tidak stasioner terhadap mean, maka akan dilakukan differencing dan transformasi log dari data inflasi. Untuk ke depannya, yang dimaksud dengan transformasi adalah differencing dan transformasi log dari data. Berikut grafik dari data inflasi yang sudah ditransformasi.



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:
 H_0 : Transformasi data inflasi tidak stasioner terhadap mean.
 H_1 : Transformasi data inflasi stasioner terhadap mean.
- Tingkat Signifikansi:
 α : 0.05.
- Statistik Uji:
P-value : 9.412082670843295e-09
- Daerah Kritik:
 H_0 ditolak jika $p\text{-value} < \alpha$.
- Kesimpulan:
Karena $p\text{-value} < \alpha$, maka H_0 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 α (0.05). Berikut adalah output dari tes beberapa model ARIMA.

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

[illegible]

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

[illegible]

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

[illegible]

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

	coef	std err	z	P> z	[0.025	0.975]
ar.L1.D.y	-0.1290	0.193	-0.668	0.505	-0.508	0.249
ma.L1.D.y	-0.5557	0.172	-3.225	0.001	-0.893	-0.218
ma.L2.D.y	-0.4443	0.172	-2.585	0.010	-0.781	-0.107
Roots						

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

[illegible]

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

	coef	std err	z	P> z	[0.025	0.975]
ma.L1.D.y	-0.6675	0.071	-9.349	0.000	-0.807	-0.528
ma.L2.D.y	-0.3325	0.070	-4.733	0.000	-0.470	-0.195
Roots						

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

[illegible]

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

	coef	std err	z	P> z	[0.025	0.975]
ar.L1.D.y	0.3046	0.071	4.301	0.000	0.166	0.443
ar.L2.D.y	-0.1416	0.071	-1.998	0.047	-0.281	-0.003
ma.L1.D.y	-1.0000	0.014	-72.838	0.000	-1.027	-0.973
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
const	-1.272e-05	0.000	-0.057	0.954	-0.000	0.000
ar.L1.D.y	0.2682	0.069	3.876	0.000	0.133	0.404
ma.L1.D.y	-1.0000	0.013	-75.042	0.000	-1.026	-0.974
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
ar.L1.D.y	0.2683	0.069	3.878	0.000	0.133	0.404
ma.L1.D.y	-1.0000	0.013	-74.867	0.000	-1.026	-0.974
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
const	-1.976e-05	0.000	-0.117	0.907	-0.000	0.000
ma.L1.D.y	-0.9999	0.014	-71.039	0.000	-1.028	-0.972
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
ma.L1.D.y	-1.0000	0.014	-70.451	0.000	-1.028	-0.972
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
const	0.0007	0.006	0.114	0.910	-0.011	0.012
ar.L1.D.y	-0.3820	0.066	-5.766	0.000	-0.512	-0.252
ar.L2.D.y	-0.3771	0.066	-5.713	0.000	-0.506	-0.248
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
ar.L1.D.y	-0.3820	0.066	-5.765	0.000	-0.512	-0.252
ar.L2.D.y	-0.3771	0.066	-5.712	0.000	-0.506	-0.248
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
const	0.0007	0.009	0.082	0.935	-0.017	0.018
ar.L1.D.y	-0.2764	0.069	-4.026	0.000	-0.411	-0.142
Roots						

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

	coef	std err	z	P> z	[0.025	0.975]
ar.L1.D.y	-0.2764	0.069	-4.026	0.000	-0.411	-0.142
Roots						

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 α . 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.

```
=====
Dep. Variable:          D.y    No. Observations:          196
Model:                ARIMA(0, 1, 2)    Log Likelihood          120.886
Method:                css-mle    S.D. of innovations          0.129
Date:                  Fri, 13 Mar 2020    AIC          -235.772
Time:                  21:47:54    BIC          -225.937
Sample:                1    HQIC          -231.790
=====
```

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

```
=====
Dep. Variable:          D.y    No. Observations:          196
Model:                ARIMA(2, 1, 1)    Log Likelihood          120.905
Method:                css-mle    S.D. of innovations          0.129
Date:                  Fri, 13 Mar 2020    AIC          -233.810
Time:                  21:47:55    BIC          -220.697
Sample:                1    HQIC          -228.501
=====
```

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

```
=====
Dep. Variable:          D.y    No. Observations:          196
Model:                ARIMA(1, 1, 1)    Log Likelihood          118.934
Method:                css-mle    S.D. of innovations          0.130
Date:                  Fri, 13 Mar 2020    AIC          -231.867
Time:                  21:47:56    BIC          -222.033
Sample:                1    HQIC          -227.886
=====
```

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

```
=====
Dep. Variable:          D.y    No. Observations:          196
Model:                ARIMA(0, 1, 1)    Log Likelihood          111.663
Method:                css-mle    S.D. of innovations          0.135
Date:                  Fri, 13 Mar 2020    AIC          -219.327
Time:                  21:47:56    BIC          -212.771
Sample:                1    HQIC          -216.673
=====
```

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

```
=====
Dep. Variable:          D.y    No. Observations:          196
Model:                ARIMA(2, 1, 0)    Log Likelihood          99.257
Method:              css-mle    S.D. of innovations          0.146
Date:                Fri, 13 Mar 2020    AIC          -192.515
Time:                21:47:56    BIC          -182.680
Sample:              1    HQIC          -188.533
=====
```

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

```
=====
Dep. Variable:          D.y    No. Observations:          196
Model:                ARIMA(1, 1, 0)    Log Likelihood          84.226
Method:              css-mle    S.D. of innovations          0.157
Date:                Fri, 13 Mar 2020    AIC          -164.452
Time:                21:47:56    BIC          -157.895
Sample:              1    HQIC          -161.797
=====
```

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

$$\Delta \log(X_t) = -0.2764 \Delta \log(X_{t-1}) + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

dengan

X_t : inflasi pada bulan ke - t .

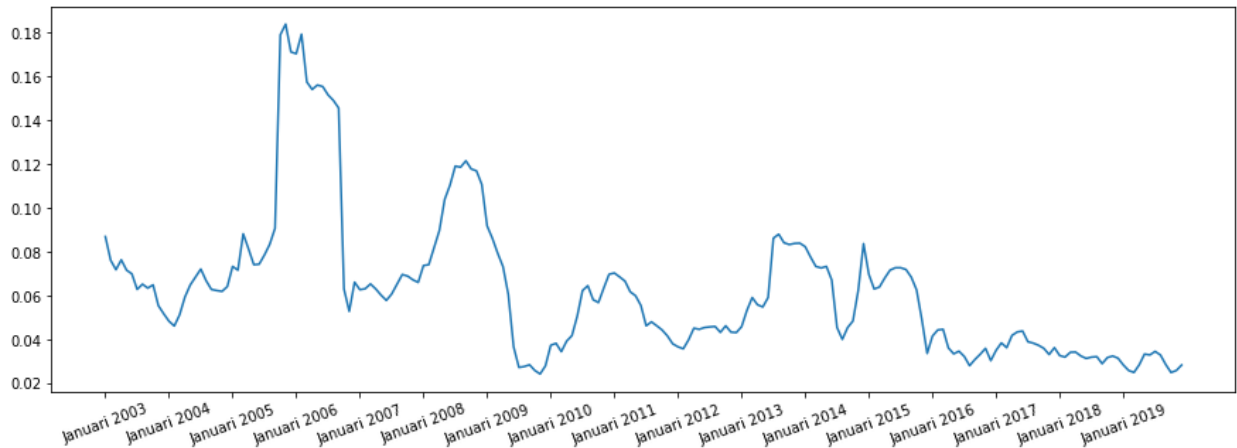
ε_t : error pada waktu ke - t .

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.

2. 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:

H_0 : Data mpg berdistribusi normal.

H_1 : Data mpg tidak berdistribusi normal.

- Tingkat Signifikansi:

α : 0.05.

- Statistik Uji:

P-value : 0.12288373708724976

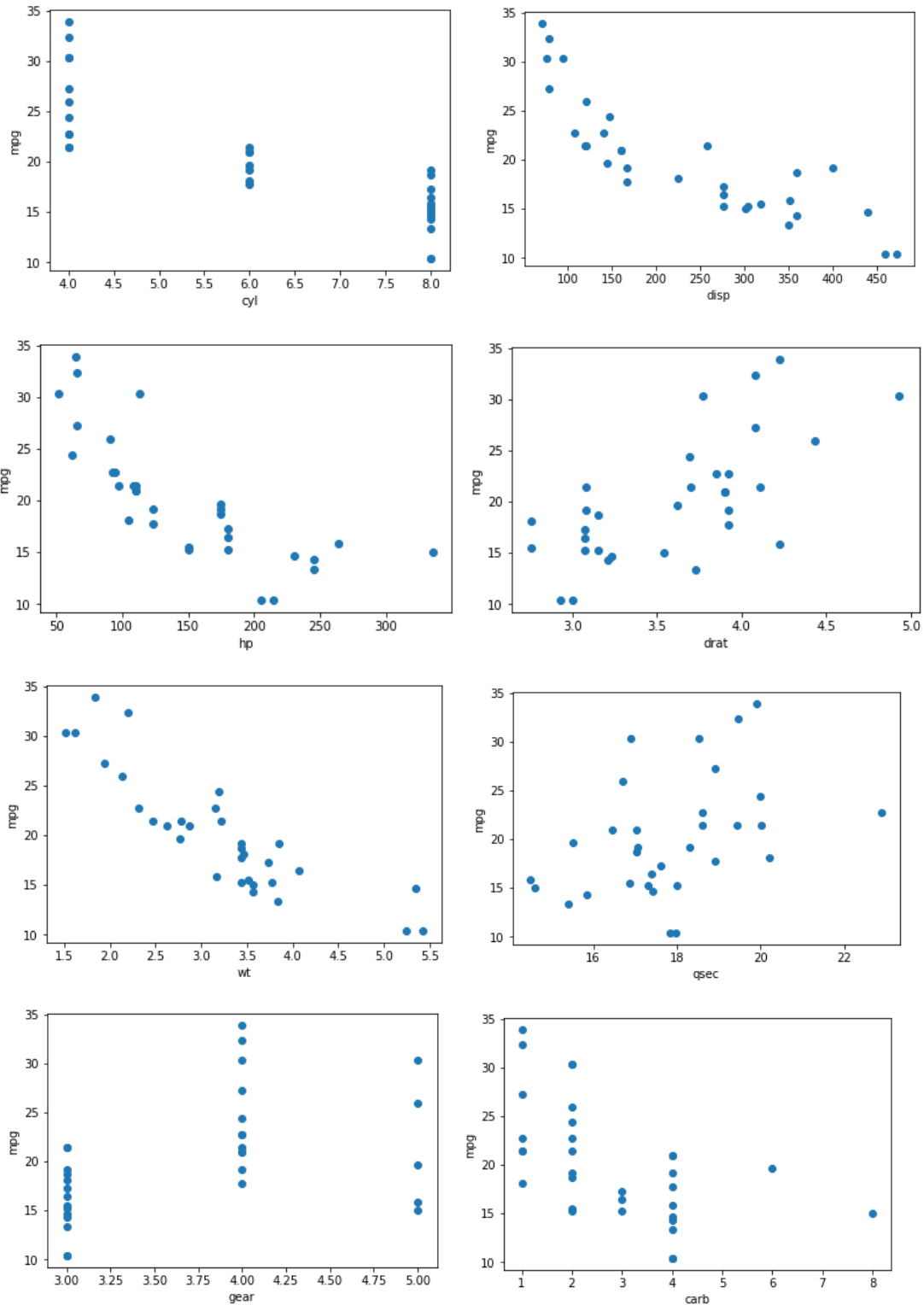
- Daerah Kritik:

H_0 ditolak jika $p\text{-value} < \alpha$.

- Kesimpulan:

Karena $p\text{-value} > \alpha$, maka H_0 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 ($p\text{-value} > \alpha$) 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

Dep. Variable:	mpg	R-squared:	0.869
Model:	OLS	Adj. R-squared:	0.807
Method:	Least Squares	F-statistic:	13.93
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	3.79e-07
Time:	02:00:27	Log-Likelihood:	-69.855
No. Observations:	32	AIC:	161.7
Df Residuals:	21	BIC:	177.8
Df Model:	10		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	12.3034	18.718	0.657	0.518	-26.623	51.229
cyl	-0.1114	1.045	-0.107	0.916	-2.285	2.062
disp	0.0133	0.018	0.747	0.463	-0.024	0.050
hp	-0.0215	0.022	-0.987	0.335	-0.067	0.024
drat	0.7871	1.635	0.481	0.635	-2.614	4.188
wt	-3.7153	1.894	-1.961	0.063	-7.655	0.224
qsec	0.8210	0.731	1.123	0.274	-0.699	2.341
vs	0.3178	2.105	0.151	0.881	-4.059	4.694
am	2.5202	2.057	1.225	0.234	-1.757	6.797
gear	0.6554	1.493	0.439	0.665	-2.450	3.761
carb	-0.1994	0.829	-0.241	0.812	-1.923	1.524

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

2. Model 2

Dep. Variable:	mpg	R-squared:	0.869
Model:	OLS	Adj. R-squared:	0.815
Method:	Least Squares	F-statistic:	16.21
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	9.03e-08
Time:	02:03:49	Log-Likelihood:	-69.864
No. Observations:	32	AIC:	159.7
Df Residuals:	22	BIC:	174.4
Df Model:	9		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	10.9601	13.530	0.810	0.427	-17.100	39.020
disp	0.0128	0.017	0.763	0.454	-0.022	0.048
hp	-0.0219	0.021	-1.048	0.306	-0.065	0.021
drat	0.8352	1.536	0.544	0.592	-2.351	4.021
wt	-3.6925	1.840	-2.007	0.057	-7.507	0.122
qsec	0.8424	0.687	1.227	0.233	-0.582	2.267
vs	0.3897	1.948	0.200	0.843	-3.650	4.430
am	2.5774	1.940	1.328	0.198	-1.447	6.601
gear	0.7116	1.366	0.521	0.608	-2.121	3.544
carb	-0.2196	0.789	-0.278	0.783	-1.855	1.416

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

3. Model 3

Dep. Variable:	mpg	R-squared:	0.869
Model:	OLS	Adj. R-squared:	0.823
Method:	Least Squares	F-statistic:	19.02
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	2.01e-08
Time:	02:04:44	Log-Likelihood:	-69.893
No. Observations:	32	AIC:	157.8
Df Residuals:	23	BIC:	171.0
Df Model:	8		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	9.7683	11.892	0.821	0.420	-14.833	34.369
disp	0.0121	0.016	0.753	0.459	-0.021	0.045
hp	-0.0210	0.020	-1.051	0.304	-0.062	0.020
drat	0.8751	1.491	0.587	0.563	-2.210	3.960
wt	-3.7115	1.798	-2.064	0.050	-7.432	0.009
qsec	0.9108	0.583	1.562	0.132	-0.295	2.117
am	2.5239	1.881	1.342	0.193	-1.368	6.416
gear	0.7598	1.316	0.577	0.569	-1.962	3.482
carb	-0.2480	0.759	-0.327	0.747	-1.819	1.323

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

4. Model 4

Dep. Variable:	mpg	R-squared:	0.868
Model:	OLS	Adj. R-squared:	0.830
Method:	Least Squares	F-statistic:	22.56
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	4.22e-09
Time:	02:05:36	Log-Likelihood:	-69.967
No. Observations:	32	AIC:	155.9
Df Residuals:	24	BIC:	167.7
Df Model:	7		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	9.1976	11.542	0.797	0.433	-14.624	33.020
disp	0.0155	0.012	1.278	0.213	-0.010	0.041
hp	-0.0247	0.016	-1.548	0.135	-0.058	0.008
drat	0.8102	1.450	0.559	0.582	-2.183	3.803
wt	-4.1307	1.236	-3.342	0.003	-6.681	-1.580
qsec	1.0098	0.489	2.066	0.050	0.001	2.019
am	2.5898	1.835	1.411	0.171	-1.198	6.378
gear	0.6064	1.206	0.503	0.620	-1.883	3.095

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

5. Model 5

Dep. Variable:	mpg	R-squared:	0.867
Model:	OLS	Adj. R-squared:	0.835
Method:	Least Squares	F-statistic:	27.09
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	8.64e-10
Time:	02:06:24	Log-Likelihood:	-70.134
No. Observations:	32	AIC:	154.3
Df Residuals:	25	BIC:	164.5
Df Model:	6		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	10.7106	10.975	0.976	0.338	-11.894	33.315
disp	0.0131	0.011	1.193	0.244	-0.010	0.036
hp	-0.0218	0.015	-1.488	0.149	-0.052	0.008
drat	1.0207	1.367	0.746	0.462	-1.796	3.837
wt	-4.0445	1.206	-3.355	0.003	-6.527	-1.562
qsec	0.9907	0.480	2.064	0.050	0.002	1.979
am	2.9847	1.634	1.827	0.080	-0.380	6.350

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

6. Model 6

Dep. Variable:	mpg	R-squared:	0.864
Model:	OLS	Adj. R-squared:	0.838
Method:	Least Squares	F-statistic:	32.96
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	1.84e-10
Time:	02:07:11	Log-Likelihood:	-70.487
No. Observations:	32	AIC:	153.0
Df Residuals:	26	BIC:	161.8
Df Model:	5		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	14.3619	9.741	1.474	0.152	-5.661	34.384
disp	0.0112	0.011	1.060	0.299	-0.011	0.033
hp	-0.0212	0.015	-1.460	0.156	-0.051	0.009
wt	-4.0843	1.194	-3.420	0.002	-6.539	-1.630
qsec	1.0069	0.475	2.118	0.044	0.030	1.984
am	3.4705	1.486	2.336	0.027	0.416	6.525

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

7. Model 7

Dep. Variable:	mpg	R-squared:	0.858
Model:	OLS	Adj. R-squared:	0.837
Method:	Least Squares	F-statistic:	40.74
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	4.59e-11
Time:	02:07:45	Log-Likelihood:	-71.164
No. Observations:	32	AIC:	152.3
Df Residuals:	27	BIC:	159.7
Df Model:	4		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	17.4402	9.319	1.871	0.072	-1.681	36.561
hp	-0.0176	0.014	-1.247	0.223	-0.047	0.011
wt	-3.2381	0.890	-3.639	0.001	-5.064	-1.412
qsec	0.8106	0.439	1.847	0.076	-0.090	1.711
am	2.9255	1.397	2.094	0.046	0.059	5.792

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

8. Model 8

Dep. Variable:	mpg	R-squared:	0.850
Model:	OLS	Adj. R-squared:	0.834
Method:	Least Squares	F-statistic:	52.75
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	1.21e-11
Time:	02:08:27	Log-Likelihood:	-72.060
No. Observations:	32	AIC:	152.1
Df Residuals:	28	BIC:	158.0
Df Model:	3		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
const	9.6178	6.960	1.382	0.178	-4.638	23.874
wt	-3.9165	0.711	-5.507	0.000	-5.373	-2.460
qsec	1.2259	0.289	4.247	0.000	0.635	1.817
am	2.9358	1.411	2.081	0.047	0.046	5.826

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

9. Model 9

Dep. Variable:	mpg	R-squared:	0.987
Model:	OLS	Adj. R-squared:	0.986
Method:	Least Squares	F-statistic:	741.0
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	1.71e-27
Time:	02:18:40	Log-Likelihood:	-73.115
No. Observations:	32	AIC:	152.2
Df Residuals:	29	BIC:	156.6
Df Model:	3		
Covariance Type:	nonrobust		

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

	coef	std err	t	P> t	[0.025	0.975]
wt	-3.1855	0.483	-6.598	0.000	-4.173	-2.198
qsec	1.5998	0.102	15.665	0.000	1.391	1.809
am	4.2995	1.024	4.198	0.000	2.205	6.394

Karena semua variable sudah signifikan, maka iterasi dihentikan.

Selanjutnya, akan dipilih model yang terbaik dengan membandingkan nilai dari R^2 , $Adj R^2$, AIC , BIC . Model terbaik adalah model yang memiliki nilai R^2 , $Adj R^2$ terbesar dan AIC , BIC terkecil. Diperhatikan table berikut.

R^2	$Adj R^2$	AIC	BIC
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

$$Y = -3.1855X_1 + 1.5998X_2 + 4.2995X_3 + \varepsilon$$

dengan

Y : variable mpg

X_1 : variable wt

X_2 : variable qsec

X_3 : 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.