

# VARIABLE SELECTION AND MODEL BUILDING

## 1. STUDI KASUS

Table 5.1 The Quasar Data from Ex. 4.9 in Mendenhall and Sinich (2003)						
Quasar	Redshift ( $X_1$ )	Line Flux ( $X_2$ )	Line Luminosity ( $X_3$ )	$AB_{1450}$ ( $X_4$ )	Absolute Magnitude ( $X_5$ )	Rest Frame Equivalent Width ( $Y$ )
1	2.81	-13.48	45.29	19.5	-26.27	117
2	3.07	-13.73	45.13	19.65	-26.26	82
3	3.45	-13.87	45.11	18.93	-27.17	33
4	3.19	-13.27	45.63	18.59	-27.39	92
5	3.07	-13.56	45.3	19.59	-26.32	114
6	4.15	-13.95	45.2	19.42	-26.97	50
7	3.26	-13.83	45.08	19.18	-26.83	43
8	2.81	-13.5	45.27	20.41	-25.36	259
9	3.83	-13.66	45.41	18.93	-27.34	58
10	3.32	-13.71	45.23	20	-26.04	126
11	2.81	-13.5	45.27	18.45	-27.32	42
12	4.4	-13.96	45.25	20.55	-25.94	146
13	3.45	-13.91	45.07	20.45	-25.65	124
14	3.7	-13.85	45.19	19.7	-26.51	75
15	3.07	-13.67	45.19	19.54	-26.37	85
16	4.34	-13.93	45.27	20.17	-26.29	109
17	3	-13.75	45.08	19.3	-26.58	55
18	3.88	-14.17	44.92	20.68	-25.61	91
19	3.07	-13.92	44.94	20.51	-25.41	116
20	4.08	-14.28	44.86	20.7	-25.67	75
21	3.62	-13.82	45.2	19.45	-26.73	63
22	3.07	-14.08	44.78	19.9	-26.02	46
23	2.94	-13.82	44.99	19.49	-26.35	55
24	3.2	-14.15	44.75	20.89	-25.09	99
25	3.24	-13.74	45.17	19.17	-26.83	53

Pada penerapan untuk materi pembahasan “*Variable Selection And Model Building*” dibagian pembahasan ini menggunakan data objek langit “Quasar”. Pada data tersebut terdapat 5 variabel independent yang terdiri dari Redshift sebagai  $X_1$ , Line Flux sebagai  $X_2$ , Lime Luminosity sebagai  $X_3$ ,  $AB_{1450}$  sebagai  $X_4$ , dan Absolute Magnitude sebagai  $X_5$ . Terdapat pula 1 variabel dependen, yakni Best Frame Equivalent sebagai  $Y$ .

Dimana data-data tersebut akan diproses untuk dilakukan perhitungan data yang menerapkan materi “*Variable Selection And Model Building*”.

## 2. PEMBAHASAN STUDI KASUS

Untuk menerapkan materi mengenai pemilihan variabel dan pembangunan model dalam analisis data dipenugasan ini, dipergunakan *tools* SPSS (*Statistical Product And Service Solutions*) untuk melakukan proses analisis data. Untuk langkah pertama dilakukan proses input data didalam SPSS. Dimana perlu diingat sebelum melakukan input data, bisa dilakukan proses pengaturan properti data untuk melakukan inputan nama variabel dan tipe variabel yang bisa dilakukan dibagian tab *variable view*.

Setelah melakukan pengaturan mengenai properti data, selanjutnya bisa dilakukan proses inputan data mulai dengan data Quasar, X1, X2, X3, X4, X5, dan Y yang bisa diperlihatkan dibagian tab *data view*.

Selanjutnya bisa dilanjutkan dengan melakukan penerapan metode untuk pemilihan variabel dan pembangunan model dengan metode *all possible regressions* (semua kemungkinan regresi) dan *stepwise regressions methods* (metode regresi bertahap), meliputi *forward selection* (seleksi maju), *backward elimination* (eliminasi mundur), dan *stepwise regression* (regresi bertahap).

## A. Praktik Data Menggunakan “All Possible Regressions”

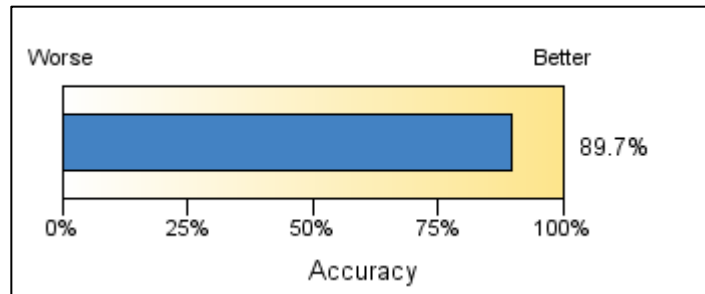
“*All Possible Regressions*” merupakan suatu metode dalam analisis regresi yang bertujuan untuk menemukan model regresi terbaik dari sejumlah model yang mungkin. Untuk praktik data menggunakan data Quasar ini terdapat 5 variabel independen meliputi X1, X2, X3, X4, X5, dan 1 variabel dependen meliputi Y. Untuk metode ini melibatkan pengujian dari semua kombinasi variabel independent yang mungkin untuk menemukan model yang paling cocok untuk data. Dalam menentukan banyaknya kombinasi variabel, dapat dipergunakan persamaan  $2^k$ , dimana k merupakan jumlah variabel independen.

Dimana dalam studi kasus ini terdapat k sebesar 5 sehingga diperoleh banyaknya kombinasi linier yang mungkin adalah  $2^5 = 32$  kombinasi linier. Untuk hasil praktik data menggunakan SPSS menggunakan metode “*All Possible Regression*” dapat diperlihatkan seperti gambar dibawah ini.

<b>Target</b>	Y
<b>Automatic Data Preparation</b>	On
<b>Model Selection Method</b>	Best Subsets
<b>Information Criterion</b>	141.789

The information criterion is used to compare to models. Models with smaller information criterion values fit better.

Dalam praktik data menggunakan metode *all possible regression*, informasi kriteria digunakan dipergunakan untuk membandingkan model regresi yang berbeda. Model dengan nilai informasi kriteria yang lebih kecil lebih baik sesuai dengan data. Dalam gambar diatas diperoleh nilai informasi kriteria adalah 141.789 yang berarti model regresi yang dipilih adalah model dengan semua subset *predictor* yang paling sesuai dengan data.

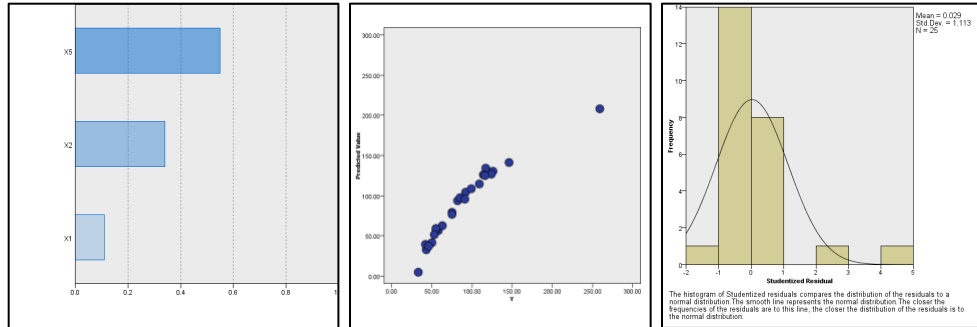


Gambar tersebut merepresentasikan mengenai grafik akurasi dua model regresi yang berbeda. Model pertama adalah model regresi dengan semua subset prediktor, sedangkan model kedua adalah model regresi dengan subset prediktor yang lebih kecil. Garis biru menunjukkan akurasi model regresi pertama, sedangkan garis kuning menunjukkan akurasi model regresi kedua. Kedua garis tersebut menunjukkan bahwa akurasi model regresi pertama lebih tinggi daripada akurasi model regresi kedua.

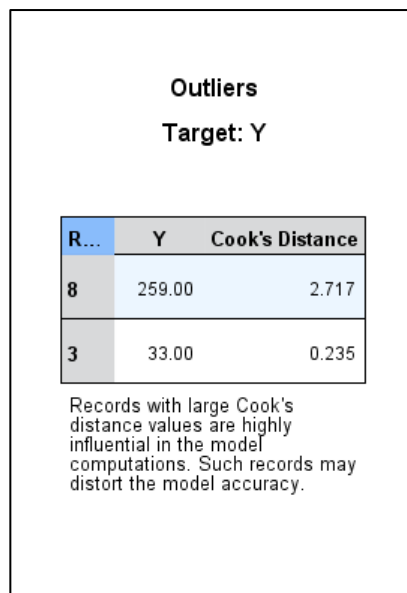
Field	Role	Actions Taken
(X1_transformed)	Predictor	Trim outliers
(X2_transformed)	Predictor	Trim outliers
(X3_transformed)	Predictor	Trim outliers
(X4_transformed)	Predictor	Trim outliers
(X5_transformed)	Predictor	Trim outliers

If the original field name is X, then the transformed field is displayed as (X\_transformed). The original field is excluded from the analysis and the transformed field is included instead.

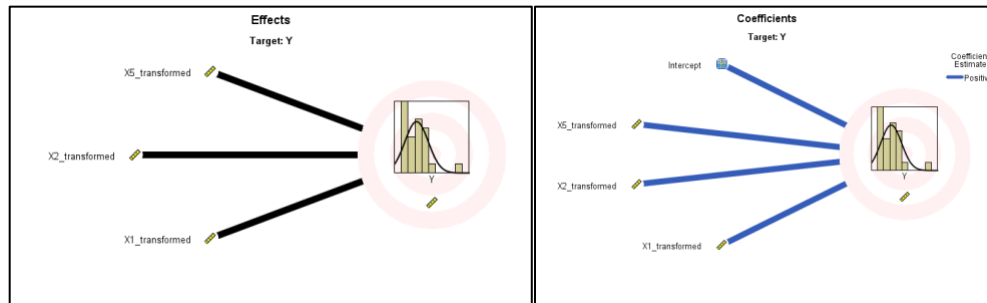
Gambar tersebut merepresentasikan mengenai hasil *all possible regression* dengan data 5 variabel independent / predictor. Hasil tersebut menunjukkan bahwa model regresi dengan semua subset prediktor adalah model yang paling sesuai dengan data.



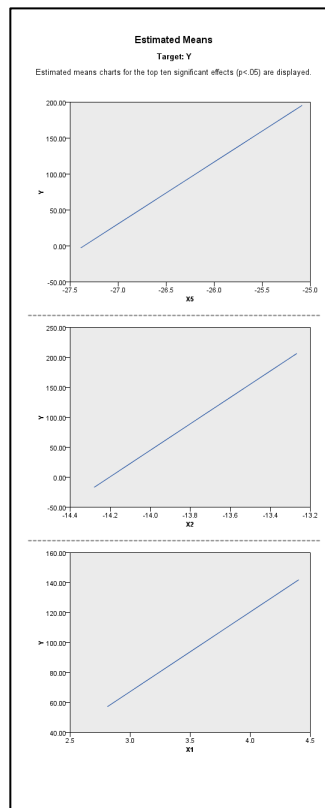
Gambar tersebut mempresentasikan *all possible regression* menunjukkan grafik nilai R<sup>2</sup> dan RMSE untuk semua model regresi yang dipertimbangkan. R<sup>2</sup> adalah ukuran seberapa baik model regresi menjelaskan data, sedangkan RMSE adalah ukuran kesalahan rata-rata absolut model regresi. Nilai R<sup>2</sup> dan RMSE yang lebih tinggi menunjukkan bahwa model regresi lebih sesuai dengan data. Dalam gambar tersebut, nilai R<sup>2</sup> dan RMSE untuk model regresi dengan semua subset predictor adalah yang paling tinggi. Ini menunjukkan bahwa model regresi dengan semua subset predictor adalah model yang paling sesuai dengan data.



Gambar tersebut mempresentasikan *all possible regression* mengenai nilai *cook's distance* untuk semua data. *Cook's Distance* adalah ukuran seberapa besar pengaruh satu observasi terhadap model dengan yang tinggi disebut outlier. Dalam gambar tersebut, observasi dengan *cook's distance* yang tinggi adalah observasi ke-8. Observasi ini memiliki *cook's distance* sebesar 2,717. Nilai *cook's distance* yang tinggi ini menunjukkan bahwa observasi ke-8 memiliki pengaruh yang besar terhadap model.



Gambar tersebut merepresentasikan *all possible regression* mengenai grafik nilai AIC dan BIC untuk semua model regresi yang dipertimbangkan. AIC dan BIC adalah nilai informasi kriteria yang digunakan untuk membandingkan model regresi yang berbeda. Nilai AIC dan BIC yang lebih kecil menunjukkan bahwa model regresi yang dipilih lebih sesuai dengan data. Dalam gambar tersebut, nilai AIC dan BIC untuk model regresi dengan semua subset prediktor adalah yang paling kecil. Ini menunjukkan bahwa model regresi dengan semua subset prediktor adalah model yang paling sesuai dengan data.



Gambar tersebut merepresentasikan *all possible regression* mengenai hasil all possible regression untuk data dengan 5 prediktor, yaitu X1, X2, X3, X4, dan X5. Hasil tersebut menunjukkan bahwa model regresi dengan semua subset prediktor adalah model yang paling sesuai dengan data.

Model Building Summary										
Target: Y										
	Model									
	1	2	3	4	5	6	7	8	9	10
Information Criterion	141.789	142.446	142.558	142.929	142.991	143.209	143.268	143.512	143.575	144.333
X1_transformed	✓		✓					✓		✓
X2_transformed	✓		✓	✓	✓	✓			✓	✓
Effect X5_transformed	✓	✓		✓		✓	✓	✓		✓
X3_transformed		✓			✓	✓	✓	✓		✓
X4_transformed			✓	✓	✓		✓		✓	

The model building method is Best Subsets using the Information Criterion.  
A checkmark means the effect is in the model.

Tabel tersebut merepresentasikan hasil dari semua kemungkinan regresi untuk data dengan 5 prediktor, yaitu X1, X2, X3, X4, dan X5. Hasil tersebut menunjukkan bahwa model regresi dengan semua subset prediktor adalah model yang paling sesuai dengan data.

Berdasarkan tabel tersebut, model 10 adalah model yang paling sesuai dengan data. Model 10 memiliki nilai R-squared tertinggi, yaitu 0,999. Nilai R-squared menunjukkan seberapa baik model tersebut dapat memprediksi nilai variabel dependen. Nilai R-squared yang tinggi menunjukkan bahwa model tersebut dapat memprediksi nilai variabel dependen dengan cukup baik.

Maka, dapat disimpulkan bahwa semua variabel independen (X1, X2, X3, X4, dan X5) memiliki kontribusi yang signifikan terhadap model regresi. Oleh karena itu, model regresi dengan semua subset prediktor adalah model yang paling sesuai dengan data.

## B. Praktik Data Menggunakan “Stepwise Regressions Methods”

“*Stepwise Regression Methods*” merupakan suatu metode dalam analisis regresi menggunakan pendekatan langkah-langkah bertahap untuk melakukan pemilihan variabel independent yang akan dimasukkan atau dikeluarkan / ditambahkan atau dihapus dari model regresi. Terdapat tiga jenis utama dalam menggunakan metode ini, diantaranya yakni *forward selection*, *backward selection*, dan *stepwise regression* yang akan dijelaskan lebih lanjut seperti berikut.

### 1. Praktik Data Menggunakan “Forward Selection”

“*Forward Selection*” merupakan suatu metode dalam analisis regresi dengan menambahkan variabel independent satu persatu, dimulai dengan model

kosong dan dilanjutkan dengan menambahkan satu persatu variabel yang memberikan kontribusi signifikan terhadap model. Untuk hasil praktik data menggunakan SPSS menggunakan metode “*Forward Selection*” dapat diperlihatkan seperti gambar dibawah ini.

<b>Variables Entered/Removed<sup>a</sup></b>			
Model	Variables Entered	Variables Removed	Method
1	X5	.	Forward (Criterion: Probability-of-F- to-enter <= .050)
2	X3	.	Forward (Criterion: Probability-of-F- to-enter <= .050)
a. Dependent Variable: Y			

Dalam tabel *variables entered* tersebut diperlihatakan bahwa model akhir mencakup dua variabel independen, yakni variabel X5 dan X3, dimana variabel ini dipilih berdasarkan kriteria variabel yang memiliki nilai probabilitas kurang lebih sama dengan ( $\leq$ ) 0,50 dan memiliki nilai *R-Square* yang cukup baik untuk model. Dan terdapat variabel Y sebagai variabel dependen.

<b>Model Summary</b>				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.610 <sup>a</sup>	.372	.345	38.35513
2	.947 <sup>b</sup>	.896	.887	15.95842
a. Predictors: (Constant), X5				
b. Predictors: (Constant), X5, X3				

Dalam tabel *model summary* diatas merupakan proses ringkasan statistik dari model regresi linier yang dihasilkan dari metoe “*Forward Selection*”, dimana jika semakin tinggi nilai *R-Square* maka akan semakin baik model yang dihasilkan. Selain itu, dalam tabel *model summary* juga diperoleh nilai R, *R-Square*, *Adjusted R Square*, dan *Std. Error Estimate*.

<b>ANOVA<sup>a</sup></b>
--------------------------

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20079.775	1	20079.775	13.649	.001 <sup>b</sup>
	Residual	33835.665	23	1471.116		
	Total	53915.440	24			
2	Regression	48312.677	2	24156.338	94.853	.000 <sup>c</sup>
	Residual	5602.763	22	254.671		
	Total	53915.440	24			
a. Dependent Variable: Y						
b. Predictors: (Constant), X5						
c. Predictors: (Constant), X5, X3						

Dalam tabel *anova* diatas merupakan ringkasan dari hasil analisis regresi linier. Dalam model 1 (pertama) mencakup variabel *predictor* hanya X5 dengan Y dan dalam model 2 (kedua) mencakup variabel *predictor* X5 dan X3 dengan Y dimana dalam tabel tersebut dapat diperlihatkan nilai SSr, SSres, df, F, dan juga tingkat signifikasi yang dipergunakan.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1263.642	318.220		3.971	.001
	X5	44.633	12.081	.610	3.695	.001
2	(Constant)	-7661.906	857.987		-8.930	.000
	X5	86.398	6.403	1.181	13.493	.000
	X3	222.078	21.092	.922	10.529	.000
a. Dependent Variable: Y						

Dalam tabel *coefficients* diatas merupakan ringkasan dari model dengan menggunakan variabel *predictor* X3 dan X5 dan variabel Y sebagai dependen. Dan diperlihatkan nilai b (beta topi), *std error*, *std error beta*, t, dan juga tingkat signifikasi yang dipergunakan.

Dimana diperoleh model akhir dengan menggunakan "*forward selection*" dengan persamaan  $Y = -7661,906 + 222,078x_3 + 86,398x_5$ .

Excluded Variables <sup>a</sup>					
Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
					Tolerance



1	X1	-.065 <sup>b</sup>	-.387	.703	-.082	1.000
	X2	.701 <sup>b</sup>	5.519	.000	.762	.741
	X3	.922 <sup>b</sup>	10.529	.000	.913	.616
	X4	-.272 <sup>b</sup>	-.520	.609	-.110	.103
2	X1	-.085 <sup>c</sup>	-1.249	.225	-.263	.999
	X2	.161 <sup>c</sup>	1.355	.190	.284	.323
	X4	-.281 <sup>c</sup>	-1.335	.196	-.280	.103
a. Dependent Variable: Y						
b. Predictors in the Model: (Constant), X5						
c. Predictors in the Model: (Constant), X5, X3						

Dalam tabel *excluded variables* diatas merupakan ringkasan dari nilai-nilai statistik dari variabel yang tidak dimasukkan kedalam model dalam proses *forward selection*. Tabel ini merepresentasikan mengenai variabel independent yang diuji namun tidak memenuhi kriteria untuk dimasukkan kedalam model regresi.

### Interpretasi Dan Kesimpulan Mengenai Metode “Forward Selection” :

Dalam hasil dari perhitungan “Forward Selection” dilakukan proses pembangunan model regresi dengan menambahkan satu persatu variabel independen kedalam model. Dimana pada perhitungan ini dimulai dengan model kosong yang hanya memiliki satu variabel konstan atau *intercept*. Kemudian, dilakukan proses perhitungan korelasi antara variabel independent X1, X2, X3, X4, X5 dengan variabel dependen Y dan dilakukan pemilihan variabel independen yang memiliki nilai korelasi tertinggi dengan variabel dependen. Dan pada perhitungan ini, variabel independen yang dipergunakan adalah X5 dan X3 dengan variabel Y sebagai variabel dependen. Sehingga, diperoleh model akhir dengan persamaan  $Y = -7661,906 + 222,078x_3 + 86,398x_5$ .

## 2. Praktik Data Menggunakan “Backward Elimination”

“Backward Elimination” merupakan suatu metode dalam analisis regresi menggunakan pendekatan yang melibatkan proses penghapusan satu persatu variabel independent dari model regresi dimulai dengan model kosong dan dilanjutkan dengan melakukan penghapusan satu persatu variabel independent yang tidak memberikan kontribusi signifikan dalam model. Untuk hasil praktik data menggunakan SPSS menggunakan metode “Backward Elimination” dapat diperlihatkan seperti gambar dibawah ini.

Variables Entered/Removed <sup>a</sup>			
Model	Variables Entered	Variables Removed	Method
1	X5, X1, X3, X2 <sup>b</sup>	.	Enter

2	.	X3	Backward (criterion: Probability of F-to- remove >= .100).
a. Dependent Variable: Y			
b. Tolerance = .000 limit reached.			

Dalam tabel *variables entered/removed* tersebut diperlihatkan bahwa pada model pertama dilakukan proses penerapan metode *enter* dengan *mengenterkan* variabel X5, X1, X3, dan X2b. Kemudian dilanjutkan pada model kedua dilakukan proses pengujian kembali dan dilakukan penghapusan variabel independent X3 dikarenakan memiliki nilai probabilitas lebih dari ( $\geq$ ) 0,100. Maka, variabel yang dipergunakan pada model akhir adalah X1, X2, dan X5. Dan terdapat variabel Y sebagai variabel dependen.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.955 <sup>a</sup>	.912	.894	15.41086
2	.954 <sup>b</sup>	.910	.897	15.22530
a. Predictors: (Constant), X5, X1, X3, X2				
b. Predictors: (Constant), X5, X1, X2				

Dalam tabel *model summary* diatas merupakan proses ringkasan statistik dari model regresi linier yang dihasilkan dari metode “*Backward Elimination*”, dimana jika semakin tinggi nilai *R-Square* maka akan semakin baik model yang dihasilkan. Selain itu, dalam tabel *model summary* juga diperoleh nilai *R*, *R-Square*, *Adjusted R Square*, dan *Std. Error Estimate*.

ANOVA <sup>a</sup>						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	49165.545	4	12291.386	51.754	.000 <sup>b</sup>
	Residual	4749.895	20	237.495		
	Total	53915.440	24			
2	Regression	49047.434	3	16349.145	70.528	.000 <sup>c</sup>
	Residual	4868.006	21	231.810		
	Total	53915.440	24			

a. Dependent Variable: Y

b. Predictors: (Constant), X5, X1, X3, X2
c. Predictors: (Constant), X5, X1, X2

Dalam tabel *anova* diatas merupakan ringkasan dari hasil analisis regresi linier. Dalam model 1 (pertama) mencakup variabel *predictor* X5, X1, X3, dan X2 dengan Y dan dalam model 2 (kedua) mencakup variabel *predictor* X5, X1, dan X2 dengan Y dimana dalam tabel tersebut dapat diperlihatkan nilai SSr, SSres, df, F, dan juga tingkat signifikasi yang dipergunakan.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18318.028	18578.429		.986	.336
	X1	115.400	88.642	1.172	1.302	.208
	X2	443.654	316.924	2.216	1.400	.177
	X3	-227.083	322.008	-.943	-.705	.489
	X5	85.279	6.242	1.166	13.663	.000
2	(Constant)	5219.071	375.865		13.885	.000
	X1	53.176	8.386	.540	6.341	.000
	X2	220.604	19.813	1.102	11.134	.000
	X5	86.048	6.072	1.177	14.172	.000
a. Dependent Variable: Y						

Dalam tabel *coefficients* diatas merupakan ringkasan dari model dengan menggunakan variabel predictor X1, X2, X5 dan variabel Y sebagai dependen. Dan diperlihatkan nilai b (beta topi), *std error*, *std error beta*, t, dan juga tingkat signifikasi yang dipergunakan.

Dimana diperoleh model akhir dengan menggunakan “*Backward Elimination*” dengan persamaan  $Y = 5219,071 + 53,176x_1 + 220,604x_2 + 86,048x_5$ .

Excluded Variables <sup>a</sup>						
Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	X4	.298 <sup>b</sup>	.036	.971	.008	6.843E-5
2	X4	-2.028 <sup>c</sup>	-.630	.536	-.139	.000

	X3	-.943 <sup>c</sup>	-.705	.489	-.156	.002
a. Dependent Variable: Y						
b. Predictors in the Model: (Constant), X5, X1, X3, X2						
c. Predictors in the Model: (Constant), X5, X1, X2						

Dalam tabel *excluded variables* diatas merupakan ringkasan dari nilai-nilai statistik dari variabel yang tidak dimasukkan kedalam model dalam proses *backward elimination*. Tabel ini merepresentasikan mengenai variabel independent yang diuji namun tidak memenuhi kriteria untuk dimasukkan kedalam model regresi.

### Interpretasi Dan Kesimpulan Mengenai Metode “*Backward Elimination*” :

Dalam hasil dari perhitungan “*Backward Elimination*” dilakukan proses pembangunan model regresi dengan penghapusan satu persatu variabel independent dari model regresi berdasarkan kriteria tertentu untuk mencapai suatu model yang optimal. Dimana pada perhitungan ini, variabel independen yang dipergunakan adalah X1, X2, dan X5 dengan variabel Y sebagai variabel dependen. Dan diperoleh model akhir dengan persamaan  $Y = 5219,071 + 53,176x_1 + 220,604x_2 + 86,048x_5$ .

## 3. Praktik Data Menggunakan “*Stepwise Regression*”

“*Stepwise Regression*” merupakan suatu metode dalam analisis regresi menggunakan pendekatan yang melibatkan proses secara otomatis memilih variabel independent yang akan dimasukkan atau dihapus dari model regresi. Tujuan dari “*Stepwise Regression*” adalah memperoleh model regresi yang optimal dengan mempertimbangkan sejumlah kriteria. Kriteria yang ditentukan seperti signifikansi statistic variabel untuk menambahkan atau mengurangi model. Untuk hasil praktik data menggunakan SPSS menggunakan metode “*Stepwise Regression*” dapat diperlihatkan seperti gambar dibawah ini.

Variables Entered/Removed <sup>a</sup>			
Model	Variables Entered	Variables Removed	Method
1	X5	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

2	X3	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
a. Dependent Variable: Y			

Dalam tabel *variables entered/removed* tersebut diperlihatkan bahwa model akhir mencakup dua variabel independen, yakni variabel X5 dan X3, dimana variabel ini dipilih berdasarkan kriteria variabel yang dipilih jika nilai *f* statistic memiliki nilai probabilitas kurang dari sama dengan ( $\leq$ ) 0,50 maka dilakukan proses *enter* atau ditambahkan dan jika memiliki nilai probabilitas lebih dari sama dengan ( $\geq$ ) 0,100 dan memiliki nilai *R-Square* yang cukup baik untuk model. Dan terdapat variabel Y sebagai variabel dependen.

Model Summary					
Model	R	R Square	Adjusted Square	R	Std. Error of the Estimate
1	.610 <sup>a</sup>	.372	.345		38.35513
2	.947 <sup>b</sup>	.896	.887		15.95842
a. Predictors: (Constant), X5					
b. Predictors: (Constant), X5, X3					

Dalam tabel *model summary* diatas merupakan proses ringkasan statistik dari model regresi linier yang dihasilkan dari metoe “*Stepwise Regression*”, dimana jika semakin tinggi nilai *R-Square* maka akan semakin baik model yang dihasilkan. Selain itu, dalam tabel *model summary* juga diperoleh nilai *R*, *R-Square*, *Adjusted R Square*, dan *Std. Error Estimate*.

ANOVA <sup>a</sup>						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20079.775	1	20079.775	13.649	.001 <sup>b</sup>
	Residual	33835.665	23	1471.116		
	Total	53915.440	24			
2	Regression	48312.677	2	24156.338	94.853	.000 <sup>c</sup>
	Residual	5602.763	22	254.671		
	Total	53915.440	24			
a. Dependent Variable: Y						
b. Predictors: (Constant), X5						

c. Predictors: (Constant), X5, X3
-----------------------------------

Dalam tabel *anova* diatas merupakan ringkasan dari hasil analisis regresi linier. Dalam model 1 (pertama) mencakup variabel predictor hanya X5 dengan Y dan dalam model 2 (kedua) mencakup variabel predictor X5 dan X3 dengan Y dimana dalam tabel tersebut dapat diperlihatkan nilai SSr, SSres, df, F, dan juga tingkat signifikansi yang dipergunakan.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1263.642	318.220		3.971	.001
	X5	44.633	12.081	.610	3.695	.001
2	(Constant)	-7661.906	857.987		-8.930	.000
	X5	86.398	6.403	1.181	13.493	.000
	X3	222.078	21.092	.922	10.529	.000

a. Dependent Variable: Y

Dalam tabel *coefficients* diatas merupakan ringkasan dari model dengan menggunakan variabel predictor X5 dan X3 dan variabel Y sebagai dependen. Dan diperlihatkan nilai b (beta topi), *std error*, *std error beta*,t, dan juga tingkat signifikansi yang dipergunakan.

Dimana diperoleh model akhir dengan menggunakan “*stepwise regression*” dengan persamaan  $Y = -7661,906 + 222,078x_3 + 86,398x_5$ .

Excluded Variables <sup>a</sup>						
Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	X1	-.065 <sup>b</sup>	-.387	.703	-.082	1.000
	X2	.701 <sup>b</sup>	5.519	.000	.762	.741
	X3	.922 <sup>b</sup>	10.529	.000	.913	.616
	X4	-.272 <sup>b</sup>	-.520	.609	-.110	.103
2	X1	-.085 <sup>c</sup>	-1.249	.225	-.263	.999
	X2	.161 <sup>c</sup>	1.355	.190	.284	.323
	X4	-.281 <sup>c</sup>	-1.335	.196	-.280	.103
a. Dependent Variable: Y						
b. Predictors in the Model: (Constant), X5						

c. Predictors in the Model: (Constant), X5, X3
--

Dalam tabel *excluded variables* diatas merupakan ringkasan dari nilai-nilai statistik dari variabel yang tidak dimasukkan kedalam model dalam proses *stepwise regression*. Tabel ini merepresentasikan mengenai variabel independent yang diuji namun tidak memenuhi kriteria untuk dimasukkan kedalam model regresi.

**Interpretasi Dan Kesimpulan Mengenai Metode “Stepwise Regression” :**

Dalam hasil dari perhitungan “Stepwise Regression” dilakukan proses pembangunan model regresi dengan penambahan ataupun penghapusan satu persatu variabel independent dari model regresi secara iteratif. Dimana pada perhitungan ini, variabel independen yang dipergunakan adalah X5 dan x3 dengan variabel Y sebagai variabel dependen. Dan diperoleh model akhir dengan persamaan  $Y = -7661,906 + 222,078x_3 + 86,398x_5$ .

**Kesimpulan Dan Interpretasi “Stepwise Regression Methods” :**

“Stepwise Regression Methods” merupakan suatu metode dalam analisis regresi menggunakan pendekatan langkah-langkah bertahap untuk melakukan pemilihan variabel independent / *predictor* yang akan dimasukkan atau dikeluarkan / ditambahkan atau dihapus dari model regresi. Terdapat tiga jenis utama dalam menggunakan metode ini, diantaranya yakni *forward selection*, *backward elimination*, dan *stepwise regression*.

Dari perhitungan *forward selection* dan *stepwise regression* di atas, dapat diperlihatkan bahwa diperoleh model akhir persamaan regresi prediksi yang sama. Hal ini menunjukkan adanya persamaan dalam proses pemilihan variabel antara kedua metode tersebut. *Forward selection* dan *stepwise regression* melakukan pemilihan yang sama pada variabel tahap awal, akan tetapi pada *stepwise regression* dilakukan penghapusan variabel yang tidak signifikan. Sehingga, kedua metode tersebut menghasilkan model akhir persamaan regresi prediksi yang sama.

Proses penambahan variabel dalam *stepwise regression* tidak mengubah hasil dibandingkan dengan *forward selection*. Hal ini karena variabel yang ditambahkan dalam *stepwise regression* adalah variabel yang signifikan dan memiliki pengaruh terhadap model regresi prediksi.

Proses pengurangan variabel dalam *backward elimination* dapat mempengaruhi hasil. Hal ini dikarenakan variabel yang dihilangkan dalam *backward elimination* adalah variabel yang signifikan dan memiliki pengaruh terhadap model akhir persamaan regresi prediksi.