

# Pemanfaatan Software MINITAB Untuk Regresi PLS (Partial Least Square)

Retno Subekti<sup>1</sup>

## Abstrak

Dalam analisis regresi klasik terdapat asumsi tidak adanya multikolinieritas yaitu tidak adanya korelasi yang cukup tinggi diantara variabel bebasnya. Karena dengan adanya multikolinieritas akan menyebabkan kesalahan interpretasi model regresi yang terbentuk. Salah satu solusi untuk menangani kasus multikolinieritas tersebut adalah regresi PLS. Dalam proses olah data dengan menggunakan analisis regresi linier dapat digunakan software statistik, beberapa diantaranya adalah SPSS, MINITAB, MATLAB, Program R. Untuk pilihan regresi PLS, sudah dapat dijumpai pada software MINITAB versi 14 dan SPSS 17.0 Untuk MATLAB, diperlukan tool tambahan untuk PLS. Sedangkan pada program R menyediakan *package* untuk PLS. Software SPSS dan MINITAB seringkali dipilih karena lebih mudah dalam penggunaannya.

*Kata kunci : Regresi PLS, Statistical Software, MINITAB, SPSS*

## Pendahuluan

Multikolinearitas tidak hanya terjadi akibat dari korelasi yang cukup tinggi diantara variable bebas pada analisis regresi, tetapi dapat juga terjadi karena kurangnya informasi dan variabilitas dalam data/bias atau banyaknya observasi lebih sedikitnya dibandingkan dengan banyaknya variable yang terlibat. Ada beberapa cara untuk mengatasi hal ini, salah satunya adalah regresi PLS. Regresi PLS yang awalnya di tahun 1966 oleh Herman Wold khususnya pada bidang ekonomi menjadi populer pertama kalinya dalam *chemometrics* dan evaluasi sensor (Marten dan Naes, 1989). Metode PLS sendiri dapat dikembangkan tidak hanya untuk regresi linear berganda, beberapa pengembangan generalized linear regresi PLS lainnya adalah pada regresi logistic, regresi cox dan regresi multivariate.

Dalam pengolahan data dengan metode PLS ini, banyak software yang sudah dapat digunakan seperti SIMCA-P dan Unscrambler (Bastien, 2004). Pada dasarnya, prosedur PLS sangat sederhana. Prosesnya dapat dilihat untuk regresi univariat, regresi PLS ini memodelkan variable dependen dengan tiap-tiap variable bebasnya. Dengan menggunakan metode OLS (Ordinary Least Square), prinsip kerja PLS regresi dapat dilihat pada [7]. Saat itu penulis menggunakan bantuan MINITAB versi lama yaitu MINITAB 11.0, dan secara sederhana

---

<sup>1</sup> Dosen Jurdik Matematika FMIPA UNY

dilakukan analisis regresi Y terhadap masing-masing variable bebasnya sampai menghasilkan persamaan regresi PLS.

Beberapa software bantu statistic yang cukup dikenal luas oleh para peneliti, praktisi akademis/non akademis adalah SPSS dan MINITAB karena lebih *user friendly*. Pada beberapa versi lama kedua software ini belum memiliki menu PLS pada pilihan regresinya. Saat ini untuk MINITAB mulai versi 15 dan SPSS 16.00 sudah menyediakan menu PLS regresi, sehingga untuk mengolah data dengan analisis regresi PLS menjadi semakin mudah. Walaupun di satu sisi ada kekurangan dalam menggunakan software ini karena perlunya *license* yang seringkali diabaikan oleh pengguna. Sedangkan untuk saat penulisan makalah ini, paket PLS dalam SPSS, masih mengharuskan terhubung dengan internet karena harus mendownload secara terpisah. Sehingga dalam penulisan makalah ini digunakan MINITAB yang sudah memasukkan menu PLS regresi ke dalam menu statistics-nya.

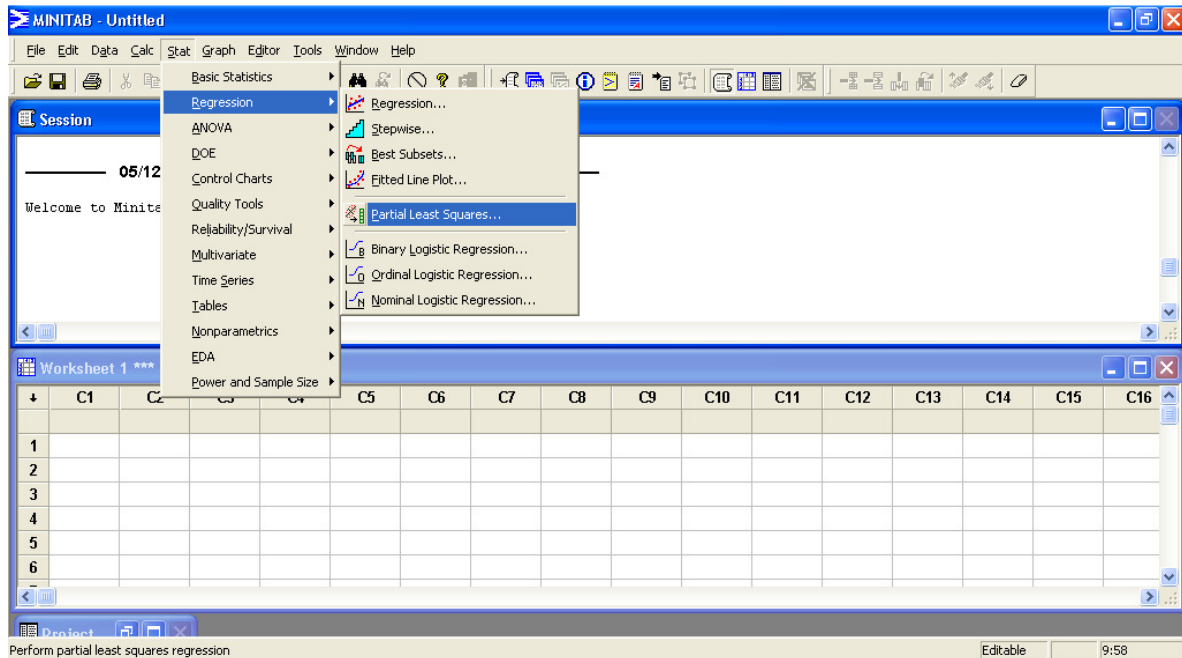
## **Pembahasan**

Dalam analisis regresi klasik dengan beberapa asumsinya mensyaratkan tidak adanya masalah multikolinieritas atau terjadi korelasi yang cukup tinggi diantara variable prediktornya. Dan dianggap mempunyai masalah seperti adanya multikolinieritas, karena akan menyebabkan kesalahan pada interpretasi model yang terbentuk. Selain melihat korelasi antar variable bebasnya, indikasi yang menunjukkan adanya kasus multikolinieritas dapat dilihat dari nilai VIF (Variation Inflation Factor). Jika nilai VIF  $>1$  maka data menunjukkan adanya indikasi antar variable bebasnya saling berkorelasi cukup tinggi. Montgomery dan Peck menyebutkan bahwa jika VIF diantara 5 - 10, maka koefisien regresinya sangat lemah untuk diestimasi.

Solusi yang dapat diambil adalah mengeluarkan salah satu variable bebas yang saling berkorelasi cukup tinggi. Solusi ini merupakan solusi yang dapat menimbulkan masalah baru bagi peneliti, andaikan peneliti harus memilih variable mana yang harus dikeluarkan dari model. Karena variable yang akan dikeluarkan adakalanya mempunyai peranan yang cukup penting atau pengaruhnya terhadap variable respon sangat besar. Solusi alternative lain adalah menggunakan regresi ridge, principal komponen, dan regresi PLS.

## **PLS Regresi dalam MINITAB**

Pada MINITAB versi 14 sudah dapat dijumpai menu untuk PLS regresi, tetapi saat ini sudah beredar MINITAB versi 15. Berikut ini tampilan pada MINITAB 14.



PLS regresi dalam MINITAB ini dapat digunakan untuk regresi linier berganda dimana menghubungkan antara sekumpulan variable bebas dengan satu variable respon dan juga untuk mengakomodir regresi linier multivariat, dimana sekumpulan variable bebas dihubungkan dengan sekumpulan variable respon.

Langkah PLS regresi dalam MINITAB adalah Stat > Regression > Partial Least Squares.

### PLS regresi dalam Software lain

Beberapa software bantu statistic yang telah mengakomodir PLS regresi dalam menunya adalah SIMCA, MATLAB, SAS, UNSCRAMBLER, SmartPLS. Tetapi ada yang memerlukan tambahan beberapa ketentuan lain, misalkan harus mendownload secara terpisah paket untuk PLSnya.

✎ Klasifikasi dengan PLS regresi dalam SIMCA

([http://www.umetrics.com/default.asp/pagename/software\\_simcap/c/3/](http://www.umetrics.com/default.asp/pagename/software_simcap/c/3/)).

✎ Prosedur PLS dalam SAS

(<http://support.sas.com/rnd/app/da/new/dapls.html>)

Software SAS sudah mengimplementasikan beberapa metode reduksi dimensi seperti Principal Component Regression, Reduced Rank Regression dan Partial Least Squares.

✎ PLS toolbox dalam MATLAB

([http://software.eigenvector.com/toolbox/3\\_5/index.html](http://software.eigenvector.com/toolbox/3_5/index.html))

✎ Unscrambler

(<http://www.camo.com/rt/Products/Unscrambler/unscrambler.html>)

## Data dan Analisis

Sebagai ilustrasi penggunaan PLS regresi dengan MINITAB, berikut ini penulis menggunakan data Heart Catheterization dari Weisberg 1980 dan Chambers 1983[Rosseuw, p 103].

*A catheter is passed into major vein or artery at femoral region and moved into heart. The catheter can be maneuvered into specific regions to provide information concerning the heart function. This techniques is sometimes applied to children with congenital heart defects. The proper length of the introduced catheter has to be guessed by the physician. For 12 children , the proper catheter length(y) was determinig by checking with fluoroscope that the catheter tip had reached the right position. The aim is to be describe the relation between the catheter length and the patient's height (x1) and weight(x2).*

index	Height (inches)	Weight (pounds)	catheter length (cm)
1	42,8	40,0	37
2	63,5	93,5	50
3	37,5	35,5	34
4	39,5	30,0	36
5	45,5	52,0	43
6	38,5	17,0	28
7	43,0	38,5	37
8	22,5	8,5	20
9	37,0	33,0	34
10	23,5	9,5	30
11	33,0	21,0	38
12	58,0	79,0	47

## Output Regression Analysis: catheter lenght versus height; weight

The regression equation is  
catheter lenght = 20,4 + 0,211 height + 0,191 weight

Predictor	Coef	SE Coef	T	P	VIF
Constant	20,376	8,386	2,43	0,038	
height	0,2107	0,3455	0,61	0,557	13,1
weight	0,1911	0,1583	1,21	0,258	13,1

S = 3,77828    R-Sq = 82,5%    R-Sq(adj) = 78,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	607,19	303,59	21,27	0,000
Residual Error	9	128,48	14,28		
Total	11	735,67			

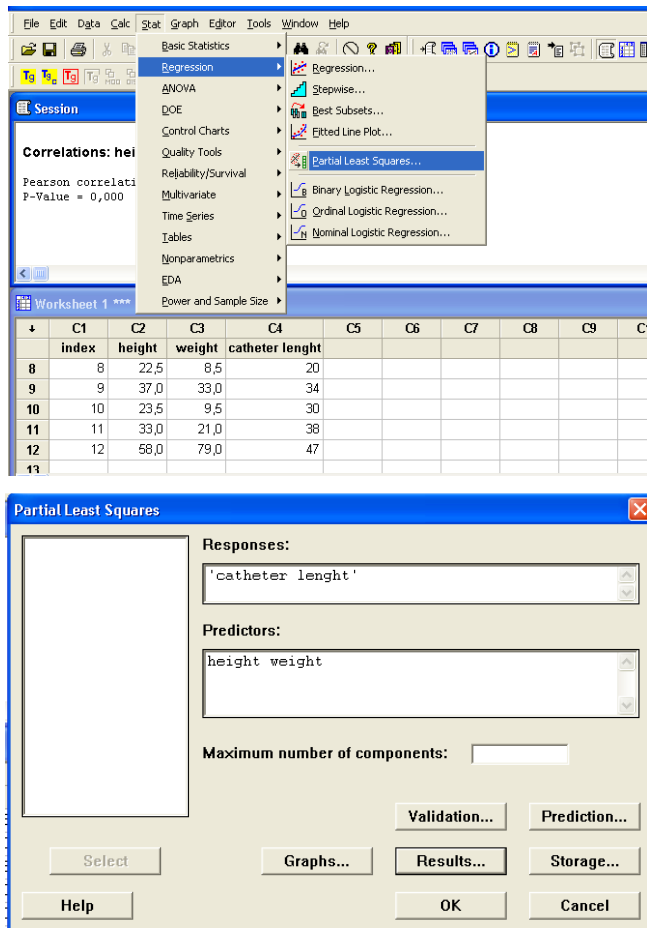
Dengan menggunakan metode OLS dan digunakan software MINITAB 14 diperoleh persamaan regresi dugaannya adalah

$$\hat{y} = 20,4 + 0,211 \text{ height} + 0,191 \text{ weight}$$

Dari uji F untuk model menyatakan bahwa secara bersama-sama  $\beta_1$  dan  $\beta_2$  signifikan mempunyai kontribusi terhadap variable y. Sedangkan saat dilihat koefisien determinasinya (adjusted  $R^2$ ) sebesar 78,7% menunjukkan bahwa sebenarnya dengan kedua variable di dalam model berkontribusi cukup besar menjelaskan hubungan antara variable predictor dan respon. Tetapi ketika diuji secara individual, masing-masing nilai t hitung berada pada area penerimaan  $H_0$ , atau jika dilihat p-value dari t hitungnya lebih dari 0,05 sebagai tingkat signifikansi (alpha) maka dapat dinyatakan masing-masing parameter koefisien regresi tidak signifikan berbeda dari nol. Hal ini menimbulkan dugaan adanya multikolinieritas variable bebasnya. Dapat dilihat pada nilai VIF ternyata lebih dari 10. Atau berikut output koefisien korelasi antara variable bebasnya melalui MINITAB

Pearson correlation of height and weight = 0,961  
P-Value = 0,000

Dengan koefisien korelasi sekitar 96,1% menunjukkan multikolinieritas yang sangat tinggi dalam model regresi sehingga solusi yang dapat diambil adalah memilih salah satu variable bebas tetap di dalam model atau menggunakan model regresi lain seperti regresi ridge, regresi PLS atau menggunakan metode lain. Dalam hal ini, penulis memilih menggunakan regresi PLS dengan bantuan MINITAB 14. Berikut ini langkah-langkah penggunaan dan hasil outputnya



Pada kotak dialog, masukkan variable catheter length sebagai *respons* dan heigt dan weight untuk *predictors*. Untuk **Maximum number of components**: ketikkan banyaknya komponen yang akan dihitung atau jika dibiarkan saja maka secara otomatis, default, oleh Minitab akan dihitung sejumlah banyaknya predictor atau cross validates 10 komponen. Selanjutnya Klik **Validation**, pilih **Leave-one-out**. klik **OK**. Kemudian untuk **Graphs**, pilih **Model selection plot**, **Response plot**, **Std Coefficient plot**, **Distance plot**, **Residual versus leverage plot**, and **Loading plot**. Jangan dipilih **Coefficient plot**. klik **OK**.

**Berikut ini output dari Minitab**

### PLS Regression: catheter length versus height; weight

```
Number of components selected by cross-validation: 1
Number of observations left out per group: 1
Number of components cross-validated: 2
```

```
Number of components set to: 2
```

Analysis of Variance for catheter lenght

Source	DF	SS	MS	F	P
Regression	2	607,188	303,594	21,27	0,000
Residual Error	9	128,479	14,275		
Total	11	735,667			

Model Selection and Validation for catheter lenght

Components	X	Variance	Error SS	R-Sq
	1	0,98055	129,727	0,823660
	2	1,00000	128,479	0,825357

Untuk menampilkan koefisien prediktornya pada pilihan **Storage**, klik **Coefficient** dan **Standardized coefficient**

Dan hasilnya adalah sebagai berikut :

Coef1	SCoef1
0,311937	0,455269
0,144757	0,461247

Dengan menggunakan nilai koefisien standardized, persamaan regresi dugaannya sekarang adalah

$$\hat{y} = 0,455269 z_1 + 0,461247 z_2$$

Dimana  $z_1$  dan  $z_2$  masing-masing adalah variable predictor yang sudah terstandarisasi. Dari uji F model regresi PLS menunjukkan model signifikan dan koefisien determinasi  $R^2$  meningkat menjadi sebesar 82,36%.

### Daftar Pustaka

- [1] Bastien, P., Vinzi, VE., Tenenhaus, M., 2004. *Partial Least Square Generalized Linear Regression*. Computational Statistics & Data Analysis 48 (2005) 17-46
- [2] Herve Abdi (2003). *Partial Least Square (PLS) Regression*. Encyclopedia of Social Sciences Research Methods
- [4] D.C. Montgomery and E.A. Peck (1982). *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- [5] Myers, R.H. (1996). *Classical and Modern Regression with Applications*. Boston : PWS-KENT Publishing Company

- [6] Neter, J., W. Wasserman, Kutner, MH. (1990). *Applied Linear Statistical Models Third Edition*, Richard D. Irwin, Inc., Homewood, Illinois
- [7] Retno, S. (2005) *Regresi Partial Least Square dalam Kasus Regresi Berganda*. Proceeding Seminar Nasional Matematika Dan Pendidikan Matematika Jurdik Matematika UNY, 2005.
- [9] Rosseuw, P.J and Leroy, A. (2003) *Robust Regression and Outlier Detection*. New Jersey: John Wiley& Sons Inc.