

# Analisa Regresi Berganda

Tjipto Juwono, Ph.D.

June 2018



SURYA  
UNIVERSITY

# Data

Home	Cost (\$)	Temp ( $^{\circ}F$ )	Ins (ln.)	Age (y)
1	250	35	3	6
2	360	29	4	10
3	165	36	7	3
4	43	60	6	9
5	92	65	5	6
6	200	30	5	5
7	355	10	6	7
8	290	7	10	10
9	230	21	9	11
10	120	55	2	5

Home	Cost (\$)	Temp ( $^{\circ}F$ )	Ins (ln.)	Age (y)
11	73	54	12	4
12	205	48	5	1
13	400	20	5	15
14	320	39	4	7
15	72	60	8	6
16	272	20	5	8
17	94	58	7	3
18	190	40	8	11
19	235	27	9	8
20	139	30	7	5

# Data

- $\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3$
- $\hat{Y}$ : Biaya pemanasan rumah di daerah dingin (dalam \$)
- $X_1$ : Suhu rata-rata di luar ( $^{\circ}F$ )
- $X_2$ : Tebal rata-rata dinding insulator (*inch*)
- $X_3$ : Usia pemanas (*Year*)

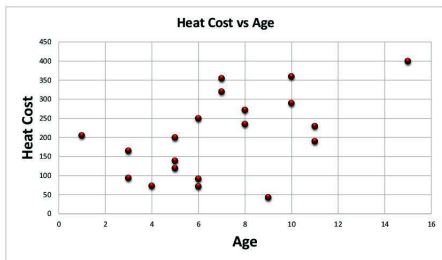
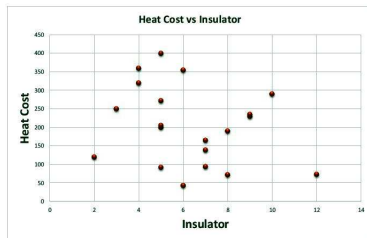
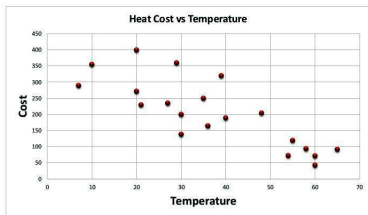
# Ringkasan Output (Excel)

<b>Regression Statistics</b>						
Multiple R	0.8967552991					
R Square	0.8041700665					
Adjusted R Square	0.767451954					
Standard Error	51.0485535828					
Observations	20					
<b>ANOVA</b>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	3	171220.472833637	57073.49	21.9011820324	6.561782477E-006	
Residual	16	41695.2771663631	2605.955			
Total	19	212915.75				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	427.1938033196	59.6014293122	7.167509	2.2376417E-006	300.8444174795	553.54318916
Temp	-4.5826626258	0.7723193534	-5.93364	2.1003502E-005	-6.2199065158	-2.9454187357
Insu	-14.8308626874	4.7544122805	-3.11939	0.0066059632	-24.9097664756	-4.7519588993
Age	6.1010320606	4.0121201661	1.52065	0.1478624843	-2.4042827407	14.6063468618

# Scatter Plots

- Scatter plot antara masing-masing variabel independen dan variabel dependen memberikan informasi awal yang berguna
- Dapat dilihat apakah ada korelasi negatif/positif
- Apakah ada korelasi yang cukup kuat

# Scatter Plots



# Menguji Asumsi dari Regresi Ganda

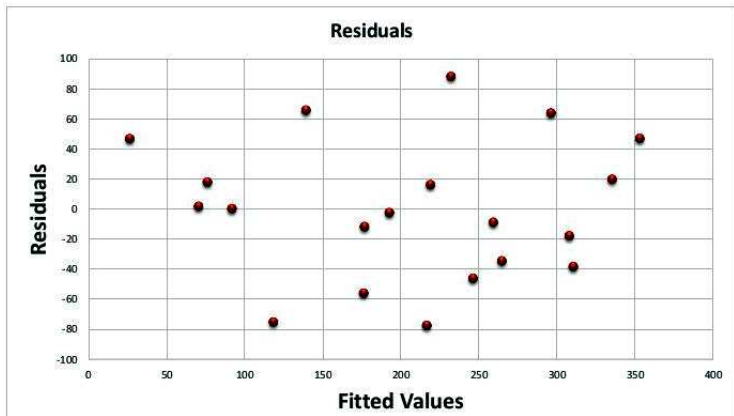
**Linearitas.** Ada hubungan dalam bentuk garis lurus antara variabel dependen dan variabel-variabel independen.

**Homoskedastisitas.** Variasi residual sama untuk semua  $Y$ .

**Normalitas.** Residu mengikuti distribusi normal.

**Multikolinearitas.** Variabel-variabel independen tidak berkorelasi satu sama lain.

# Residuals: Linearitas dan Homoskedastisitas





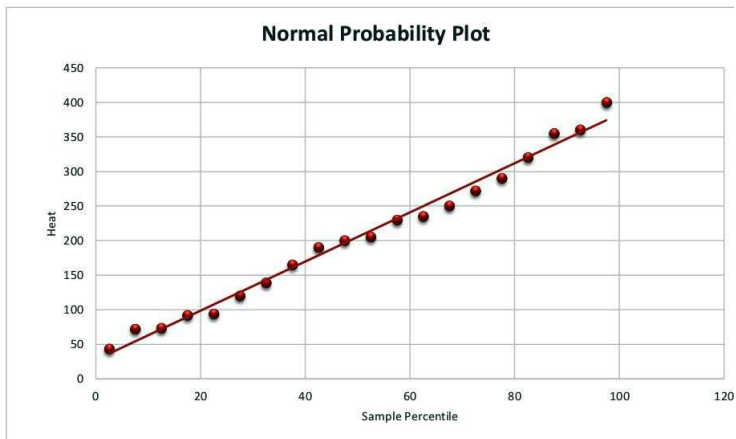
## Residuals: Linearitas

- Nilai residual terdistribusi secara random baik secara positif maupun negatif dalam distribusi yang kurang lebih sama.
- Tidak ada pola apapun dalam distribusi nilai residual

## Residuals: Homoskedastisitas

- Nilai residual terdistribusi secara random baik secara positif maupun negatif dalam distribusi yang kurang lebih sama.
- Distribusi residual untuk nilai  $Y$  yang kecil sama dengan untuk  $Y$  yang besar.

# Normality



# Multikolinearitas

	Temp	Ins	Age
Temp			
Ins	-0.103		
Age	-0.486	0.064	

- Multikolinearitas tidak mempengaruhi kemampuan persamaan regresi dalam memprediksi
- Namun multikolinearitas yang kuat membuat variabel independen menjadi redundant dan bertingkah laku tidak seperti yang diharapkan (umpamanya kita mengharapkan kenaikan positif satu variabel independen berpengaruh positif pada variabel dependen, tetapi itu ternyata tidak terjadi)
- Multikolinearitas menyebabkan kekeliruan kesimpulan dalam pengujian hipotesa

# Multikolinearitas

## Batas Multikolinearitas

Dianggap lemah jika pada interval  $-0.7 \leq r \leq 0.7$

Beberapa petunjuk adanya problem multikolinearitas:

- Suatu variabel independen yang penting ternyata menjadi tidak signifikan
- Suatu koefisien regresi yang diharapkan positif, ternyata negatif, atau sebaliknya
- Ketika suatu variabel independen ditambahkan, atau dihilangkan, terjadi perubahan drastis pada koefisien regresi sisanya.

# Multikolinearitas

Tes yang lebih tepat untuk multikolinearitas adalah *Variance Inflation Factor*

$$VIF = \frac{1}{1 - R_j^2} \quad (1)$$

$R_j^2$  dihitung dengan cara memilih salah satu variabel independen sebagai variabel dependen dan menjadikan variabel independen sisanya sebagai variabel independen. Jika diperoleh nilai  $VIF > 10$ , maka dianggap ada masalah kolinearitas dan variabel independennya harus dicoret dari analisa.

# Persamaan Regresi

## Persamaan Regresi

$$\hat{Y} = 427.194 - 4.583X_1 - 14.831X_2 + 6.101X_3 \quad (2)$$

	<i>Coefficients</i>
Intercept	427.1938033196
Temp	-4.5826626258
Insu	-14.8308626874
Age	6.1010320606

## Tabel Anova

Source	df	SS	MS	F
Regression	k	SSR	MSR=SSR/k	MSR/MSE
Residual	n-(k+1)	SSE	MSE=SSE/(n-(k+1))	
Total	n-1	SS Total		

$$SSR = \sum (\hat{Y} - \bar{Y})^2 \quad (3)$$

$$SSE = \sum (Y - \hat{Y})^2 \quad (4)$$

$$SS \text{ Total} = \sum (Y - \bar{Y})^2 \quad (5)$$



## Multiple Standard Error of Estimate

Misalkan diketahui:

- Biaya pemanasan 250\$
- Suhu di luar  $35^{\circ}F$
- Ketebalan insulator 3 inches
- Usia pemanas 6 tahun

Estimasi biaya pemanasan adalah:

$$\begin{aligned}\hat{Y} &= 427.194 - 4.583X_1 - 14.831X_2 + 6.101X_3 \\ &= 427.194 - 4.583(35) - 14.831(3) + 6.101(6) \\ &= 258.90\end{aligned}\tag{6}$$

## Multiple Standard Error of Estimate

- Residual  $(Y - \hat{Y}) = 250 - 258.90 = -8.90$
- Residual ini merupakan error random untuk salah satu data
- Residual Kuadrat  $(Y - \hat{Y})^2 = 79.21$
- Jika kita ulang perhitungan ini untuk 19 data lainnya, dan menjumlahkannya, maka kita memperoleh SSR pada tabel ANOVA.
- Dengan informasi ini kita dapat menghitung Multiple Standard Error of Estimate.

$$s_{yx} = \sqrt{MSE} \quad (7)$$

$$= \sqrt{\frac{SSR}{n - (k + 1)}} \quad (8)$$

$$= \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - (k + 1)}} \quad (9)$$

# Coefficient of Determination

## Coefficient of Multiple Determination

$$R^2 = \frac{SSR}{SS_{Total}} \quad (10)$$

## Adjusted Coefficient of Multiple Determination

$$R^2_{adj} = 1 - \frac{SSE/(n - (k + 1))}{SS_{Total}/(n - 1)} \quad (11)$$

# Ringkasan Output (Excel)

<b>Regression Statistics</b>						
Multiple R	0.8967552991					
R Square	0.8041700665					
Adjusted R Square	0.767451954					
Standard Error	51.0485535828					
Observations	20					
<b>ANOVA</b>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	3	171220.472833637	57073.49	21.9011820324	6.561782477E-006	
Residual	16	41695.2771663631	2605.955			
Total	19	212915.75				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	427.1938033196	59.6014293122	7.167509	2.2376417E-006	300.8444174795	553.54318916
Temp	-4.5826626258	0.7723193534	-5.93364	2.1003502E-005	-6.2199065158	-2.9454187357
Insu	-14.8308626874	4.7544122805	-3.11939	0.0066059632	-24.9097664756	-4.7519588993
Age	6.1010320606	4.0121201661	1.52065	0.1478624843	-2.4042827407	14.6063468618

## Hipotesa

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \textit{Tidak semua } \beta_i \textit{ sama dengan nol} \quad (12)$$

Dari tabel ANOVA:

$$\begin{aligned} F &= \frac{MSR}{MSE} \\ &= 21.9 \end{aligned} \quad (13)$$

Dari Tabel-F:

$$F(3, 16) = 3.24 \quad (14)$$

## Individual Test

Temperature	Insulator	Usia Pemanas
$H_0 : \beta_2 = 0$	$H_0 : \beta_3 = 0$	$H_0 : \beta_4 = 0$
$H_1 : \beta_2 \neq 0$	$H_1 : \beta_3 \neq 0$	$H_1 : \beta_4 \neq 0$

### Menguji Koefisien Regresi Secara Individu

$$t = \frac{b_i - 0}{s_{b_i}} \quad (15)$$

## Individual Test

- Untuk temperature, diperoleh  $t = -5.937$
- Untuk insulator, diperoleh  $t = -3.119$
- Untuk usia pemanas, diperoleh  $t = 1.521$
- Untuk  $\alpha = 0.05$ ,  $t_{\text{critical}} = -2.120$

Kesimpulan: Koefisien regresi untuk temperature dan insulator tidak sama dengan nol. tetapi untuk usia pemanas koefisien regresinya dapat nol. Lebih lanjut lagi,  $p - value$  untuk usia pemanas adalah lebih dari 0.05. Variabel usia pemanas bukanlah prediktor yang signifikan dan dapat dicoret dari analisa.