

Model Regresi Berganda

Hubungan linear (dlm parameter) antara 1 peubah tak bebas & 2 atau lebih peubah bebas

Intersep-Y
Populasi

Slope Populasi

Random
Error

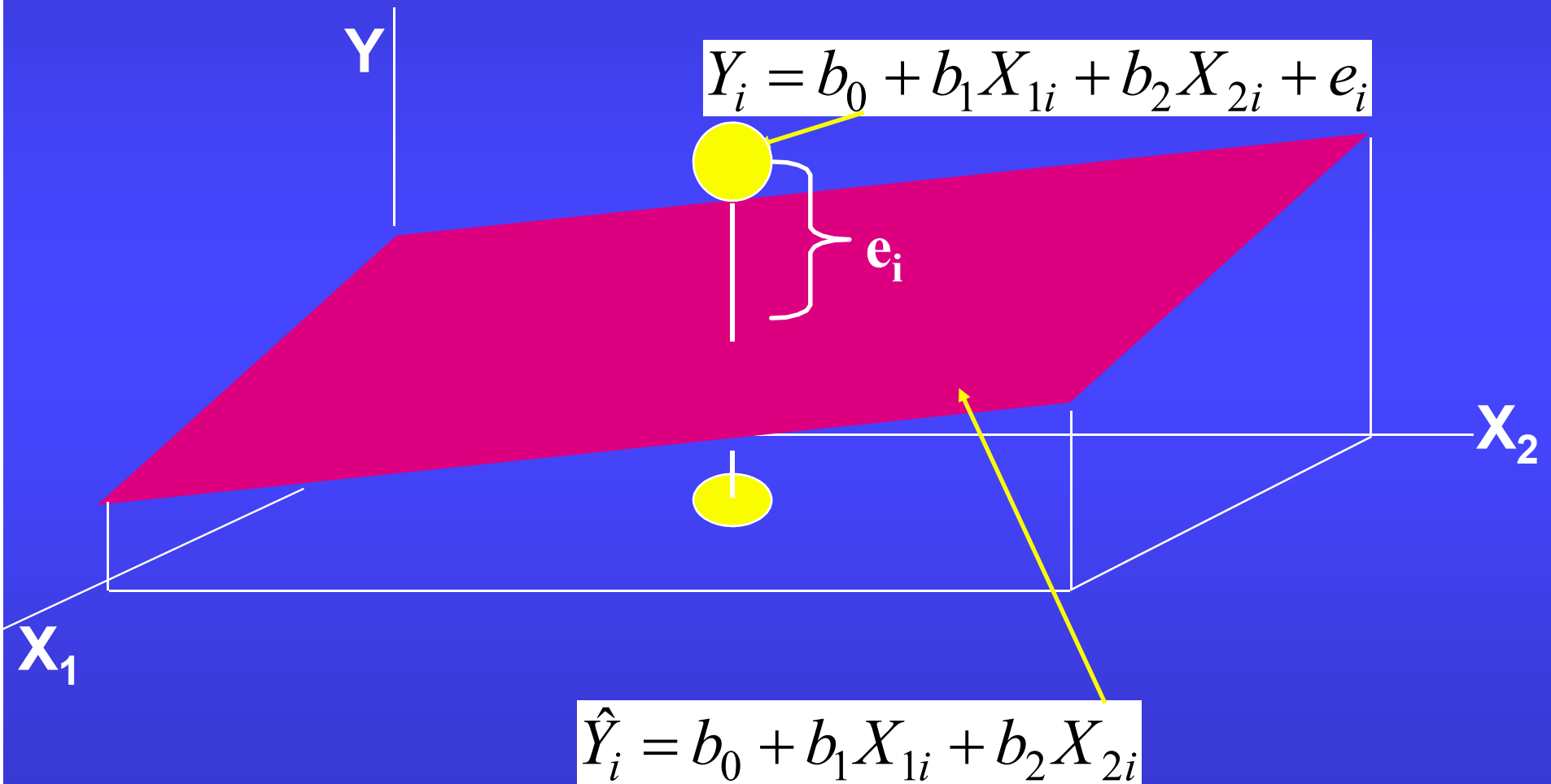
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_p X_{pi} + e_i$$

Peubah tak bebas
(Respons) utk sampel

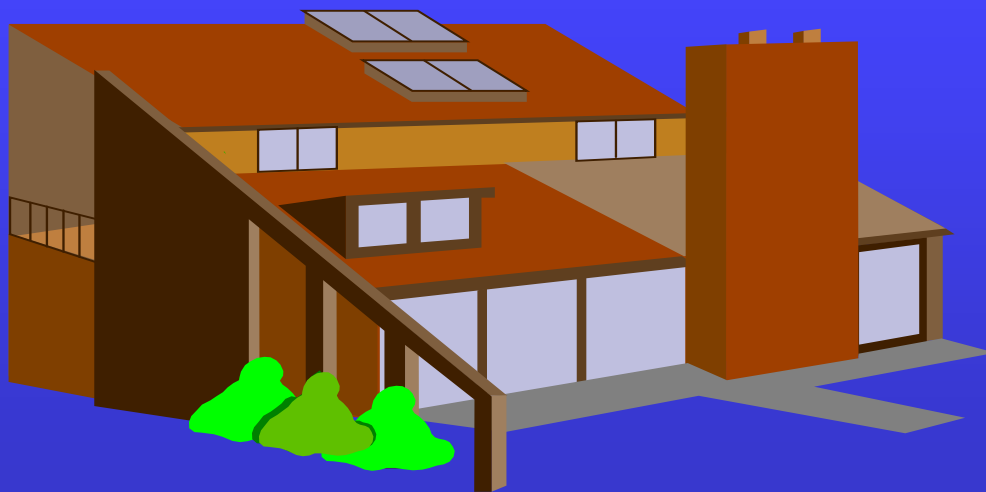
Peubah bebas (*Explanatory*)
utk model sampel

Model Regresi Berganda (Sampel) utk 2 Peubah Bebas



Model Regresi Berganda: Teladan

Tentukan suatu model utk memprediksi bahan bakar pemanas (Galon) yg digunakan sebuah rumah-satu-keluarga di bln Januari berdasarkan rata-rata temperatur ($^{\circ}\text{F}$) dan ketebalan isolasi (inchi).



Oil (Gal)	Temp($^{\circ}\text{F}$)	Insulation
275.30	40	3
363.80	27	3
164.30	40	10
40.80	73	6
94.30	64	6
230.90	34	6
366.70	9	6
300.60	8	10
237.80	23	10
121.40	63	3
31.40	65	10
203.50	41	6
441.10	21	3
323.00	38	3
52.50	58	10

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$$

$$\bar{Y} = \frac{\sum Y_i}{N}; y_i = (Y_i - \bar{Y})$$

$$\bar{X}_1 = \frac{\sum X_{1i}}{N}; x_{1i} = (X_{1i} - \bar{X}_1)$$

$$\bar{X}_2 = \frac{\sum X_{2i}}{N}; x_{2i} = (X_{2i} - \bar{X}_2)$$

$$b_0 = \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$$b_1 = \hat{\beta}_1 = \frac{(\sum x_{1i} y_i)(\sum x_{2i}^2) - (\sum x_{2i} y_i)(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2}$$

$$b_2 = \hat{\beta}_2 = \frac{(\sum x_{2i} y_i)(\sum x_{1i}^2) - (\sum x_{1i} y_i)(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2}$$

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$$

$$\text{Var}(e) = s^2 = \frac{\sum e_i^2}{N - p - 1} = KTS$$

Koef Variasi: $CV = \frac{s}{\bar{Y}}$

$$\text{Var}(b_1) = s_{b_1}^2 = \frac{s^2}{\sum x_{1i}^2 (1 - r^2)}$$

$$r_{x_1 x_2} = \frac{\sum x_{1i} x_{2i}}{\sqrt{\sum x_{1i}^2 \sum x_{2i}^2}}$$

$$\text{Var}(b_2) = s_{b_2}^2 = \frac{s^2}{\sum x_{2i}^2 (1 - r^2)}$$

$$\text{Cov}(b_1, b_2) = s_{b_1 b_2} = \frac{-s^2 r}{(1 - r^2) \sqrt{\sum x_{1i}^2 \sum x_{2i}^2}}$$

Dugaan Model Regresi Sampel

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

Output Excel

	Coefficients	<u>Satuan</u>
Intercept	562.1510092	Galon
X Variable 1	-5.436580588	Gal/°F
X Variable 2	-20.01232067	Gal/inc

$$\hat{Y}_i = 562.151 - 5.437 X_{1i} - 20.012 X_{2i}$$

Dgn isolasi yg tetap, utk tiap /
kenaikan 1 °F dlm temperatur, rata-
rata jml bahan pemanas yg di-
gunakan turun sebesar 5.437 galon.

Dgn temperatur tetap, utk tiap /
kenaikan 1 inchi isolasi, rata-
rata penggunaan bahan pemanas
turun sebesar 20.012 galon.

Penggunaan Model utk Prediksi

Dugalah rata-rata jml bahan bakar pemanas yg digunakan utk sebuah rumah jika rata-rata temperaturnya 30 °F dan isolasinya 6 inchi.

$$\begin{aligned}\hat{Y}_i &= 562.151 - 5.437X_{1i} - 20.012X_{2i} \\ &= 562.151 - 5.437 \cdot 30 - 20.012 \cdot 6 \\ &= 278.969\end{aligned}$$

Dugaan bahan bakar pemanas yg digunakan sebanyak 278.97 galon

Koefisien Determination

Output Excel

Regression Statistics	
Multiple R	0.982654757
R Square	0.965610371
Adjusted R Square	0.959878766
Standard Error	26.01378323
Observations	15

$$r_{Y,12}^2 = \frac{SSR}{SST}$$

r^2 terkoreksi

- Utk membandingkan model dgn jml peubah bebas berbeda
- Merefleksikan jml peubah bebas dan ukuran contoh
- lebih kecil dari r^2

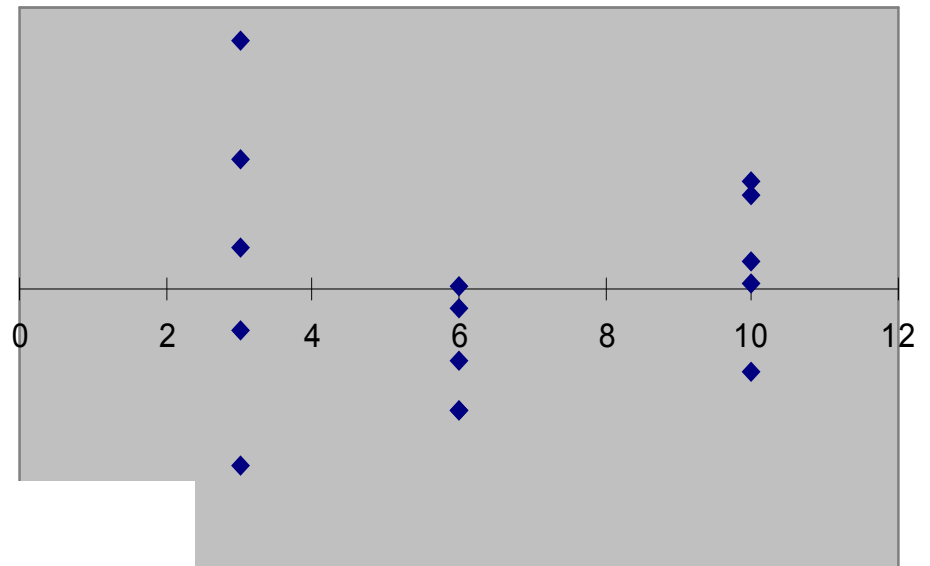
Beberapa Plot Sisaan (Dibakukan)

- Sisaan Vs \hat{Y}_i
 - Mungkin perlu transformasi peubah Y
- Sisaan Vs X_1
 - Mungkin perlu transformasi peubah X_1
- Sisaan Vs X_2
 - Mungkin perlu transformasi peubah X_2
- Sisaan Vs Time
 - Mungkin punya sifat autokorelasi

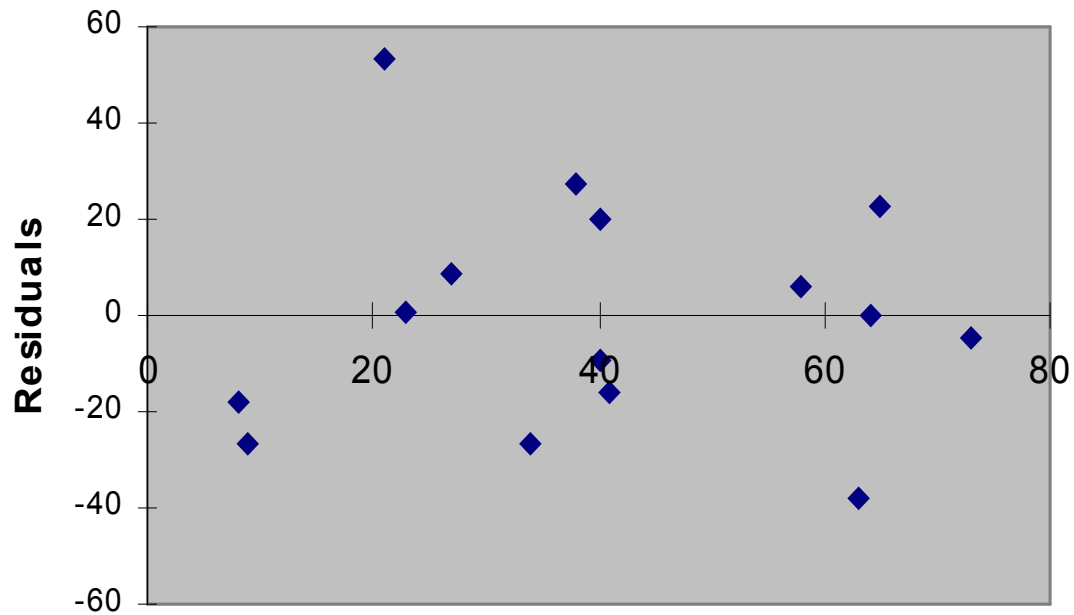
Plot Sisaan: Teladan

Output Excel

Insulation Residual Plot



Temperature Residual Plot



Tidak terlihat
suatu Pola

I. Uji Model Secara Keseluruhan

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_1 + \dots + \beta_p X_p + \varepsilon_i$$

•Apakah Model dapat menjelaskan keragaman Y

•Hipotesis Statistik:

H_0 : Model tdk dpt menjelaskan keragaman Y

$$\sigma^2_{\text{regresi}} = \sigma^2_{\varepsilon} \text{ atau } \sigma^2_{\text{regresi}} / \sigma^2_{\varepsilon} = 1$$

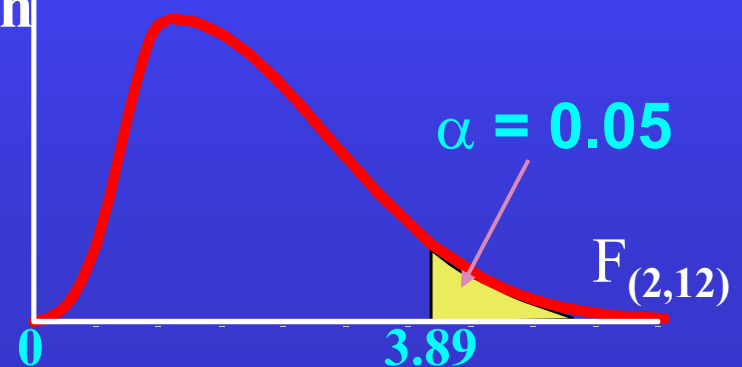
$$\beta_1 = \beta_2 = \dots = \beta_p = 0 \text{ (tidak dapat menjelaskan)}$$

H_1 : Model dpt menjelaskan keragaman Y

$$\sigma^2_{\text{regresi}} > \sigma^2_{\varepsilon} \text{ atau } \sigma^2_{\text{regresi}} / \sigma^2_{\varepsilon} > 1$$

Minimal ada $\beta_i \neq 0$ (ada peubah bebas yg mempengaruhi Y)

•Statistik uji-F = KTR/KTS $\sim F_{(p, n-1-p)}$



Uji Model Secara Keseluruhan: Analisis Ragam (ANOVA)

ANOVA	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	228014.6	114007.3	168.4712028	1.65411E-09
Residual	12	8120.603	676.7169		
Total	14	236135.2			

*p = 2, jumlah peubah
penjelas*

n - 1

p value

$$\text{Statistik Uji } F = \frac{KTR}{KTS}$$

Uji Model Secara Keseluruhan

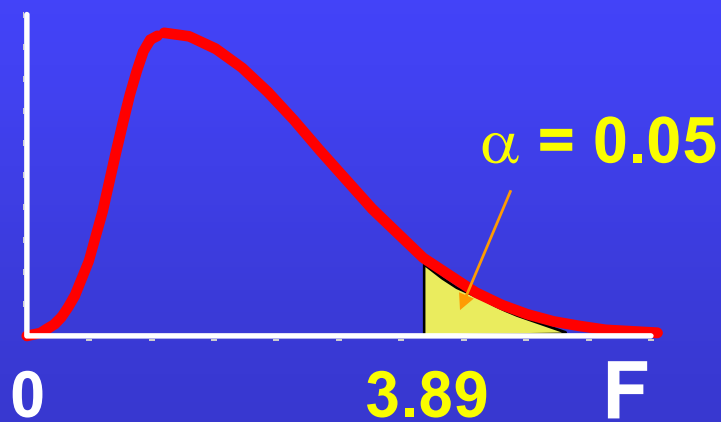
$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

H_1 : Minimal ada $\beta_i \neq 0$

$$\alpha = .05$$

db = 2 dan 12

Nilai Kritis:



Statistik Uji:

$$F = 168.47$$

(Output Excel)

Keputusan:

H_1 pada $\alpha = 0.05$

Kesimpulan:

- Model dpt menjelaskan keragaman Y
- Minimal ada satu peubah bebas yg mempengaruhi Y

II. Uji Signifikansi Masing² Peubah Bebas

- Apakah peubah bebas X_i mempengaruhi Y
- Hypotesis Statistik:

$H_0: \beta_i = 0$ (X_i tidak mempengaruhi Y)

$H_1: \beta_i \neq 0$ (X_i mempengaruhi Y)

Statistik Uji t utk X_1
(Temperatur)

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	562.151009	21.09310433	26.65094
X Variable	-5.4365806	0.336216167	-16.1699
X Variable	-20.012321	2.342505227	-8.54313

Statistik Uji t utk X_2
(Isolasi)

Teladan Uji t

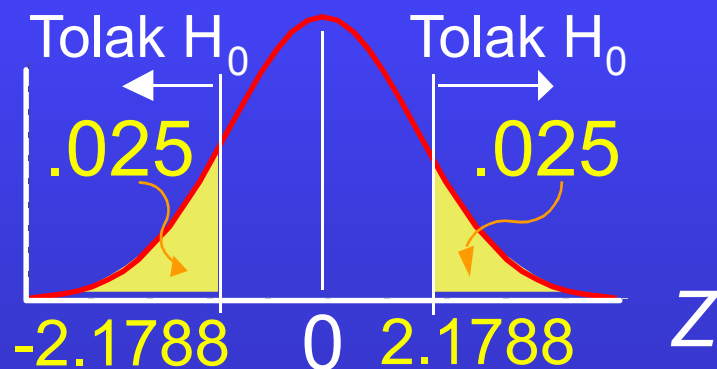
Apakah temperature berpengaruh nyata pada konsumsi bulanan bahan bakar pemanas? Ujilah pada $\alpha = 0.05$.

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$db = 12$$

Nilai Kritis:



Statistik Uji:

$$t = -16.1699$$

Keputusan:

Tolak H_0 at $\alpha = 0.05$

Kesimpulan:

Terbukti nyata, pengaruh temperatur pada konsumsi bahan bakar pemanas.

Dugaan Selang Kepercayaan utk Slope

Tentukan SK 95% bagi slope populasi β_1 (pengaruh temperatur thd Konsumsi Bahan Bakar Pemanas)

$$b_1 \pm t_{n-p-1} S_{b_1}$$

	Coefficients	Lower 95%	Upper 95%
Intercept	562.151009	516.1930837	608.108935
X Variable 1	-5.4365806	-6.169132673	-4.7040285
X Variable 2	-20.012321	-25.11620102	-14.90844

$$-6.169 \leq \beta_1 \leq -4.704$$

Rata-rata konsumsi BBP berkurang antara 4.7 galon sampai 6.17 galon tiap kenaikan suhu 1° F.

Pengujian Sebagian Model

- **Kontribusi keragaman peubah X_i thd Model (jika semua peubah lain sdh ada dlm model)**
 - **Dinotasikan dgn $JKR(X_i | \text{semua peubah selain } X_i)$**
 - **→ Koefisien determinasi parsial X_1 dgn Y jika X_2 konstan $r_{Y1.2}^2$**
- **Evaluasi Model secara Terpisah**
- **Berguna dlm Memilih Peubah Bebas**

Pengujian Sebagian Model: JKR

Kontribusi X_1 jika X_2 sdh tercakup dlm model:

$$\text{JKR}(X_1 | X_2) = \text{JKR}(X_1 \text{ dan } X_2) - \text{JKR}(X_2)$$

Dari bagian ANOVA
regresi utk model :

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

Dari bagian ANOVA
regresi utk model :

$$\hat{Y}_i = b_0 + b_2 X_{2i}$$

Uji F Parsial utk Kontribusi X_i

- **Hipotesis:**
 - H_0 : Peubah X_i tidak signifikan memperbaiki model setelah yg lainnya tercakup dlm model
 - H_1 : Peubah X_i signifikan memperbaiki model setelah yg lainnya tercakup dlm model
- **Statistik Uji:** Dgn db = 1 dan $(n - p - 1)$
 - $F = \frac{JKR(X_i | SemuaSelainnya)}{KTS}$

Koefisien Determinasi Parsial

$$r_{Y1.2}^2 = \frac{JKR(X_1|X_2)}{JKT - JKR(X_1 \text{ dan } X_2) + JKR(X_1|X_2)}$$

Dari bagian ANOVA
regresi utk model:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

Dari bagian ANOVA
regresi utk model:

$$\hat{Y}_i = b_0 + b_2 X_{2i}$$

Note: $JKR(X_1 | X_2) = JKR(X_1 \text{ dan } X_2) - JKR(X_2)$

Ujilah pada $\alpha = .05$ utk menentukan apakah peubah temperatur signifikan memperbaiki model setelah peubah isolasi ada dlm model

H_0 : X_1 tidak memperbaiki model (X_2 sdh ada)

$\alpha = .05$, db = 1 dan 12

Nilai Kritis = 4.75

H_1 : X_1 memperbaiki model

ANOVA	(For X_1 and X_2)	
	SS	MS
Regression	228014.6263	114007.313
Residual	8120.603016	676.716918
Total	236135.2293	

ANOVA	(For X_2)
	SS
Regression	51076.47
Residual	185058.8
Total	236135.2

$$F = \frac{JKR(X_1|X_2)}{KTS} = \frac{228,015 - 51,076}{676,717} = 261.47$$

Kesimpulan: Tolak H_0 . X_1 memperbaiki model

Akibat Multikolinieritas

- Koefisien Determinasi tinggi, tapi banyak koefisien yg tdk nyata
- Sulit memisahkan pengaruh dari masing-masing faktor.
- Tanda koefisien bukan hubungan yg sebenarnya
- Koefisien tdk nyata, bukan berarti peubah tsb tdk berpengaruh.

Contoh Output Minitab

The regression equation is

$$\text{Demand} = 850 - 5.03 P + 4.74 Pr + 0.277 \text{ Advertise} + 0.0107 \text{ Income} + 1.31 T$$

Predictor	Coef	SE Coef	T	P
Constant	850.0	134.8	6.31	0.000
P	-5.0326	0.4394	-11.45	0.000
Pr	4.7435	0.9696	4.89	0.000
Advertise	0.2774	0.1046	2.65	0.014
Income	0.010658	0.001316	8.10	0.000
T	1.3091	0.7534	1.74	0.095

S = 33.6398 R-Sq = 91.5% R-Sq(adj) = 89.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	291364	58273	51.49	0.000
Residual Error	24	27159	1132		
Total	29	318523			

