Metodologi Ekonometrika

Langkah-langkah metodologi ekonometrika:

1. Pernyataan Teori atau hipotesis (dugaan)
2. Spesifikasi model matematika dari teori atau hipotesis
3. Spesifikasi model statistika atau ekometrika
4. Pengumpulan data
5. Estimasi parameter dalam model ekonometrika
6. Pengujian Hipotesis
7. Peramalan atau prediksi
8. Menggunakan model untuk pengendalian atau tujuan kebijakan

Ilustrasi tahapan metode

1. Pernyataan atau hipotesis

   Teori Marginal Propensity to Consume (MPC), y.i. laju perubahan konsumsi untuk setiap penambahan 1 unit (rupiah) pendapatan, MPC berada antara 0 dan 1

2. Spesifikasi Model matematika konsumsi

   $Y = \beta_1 + \beta_2 X, \quad 0 < \beta_2 < 1$

   dimana

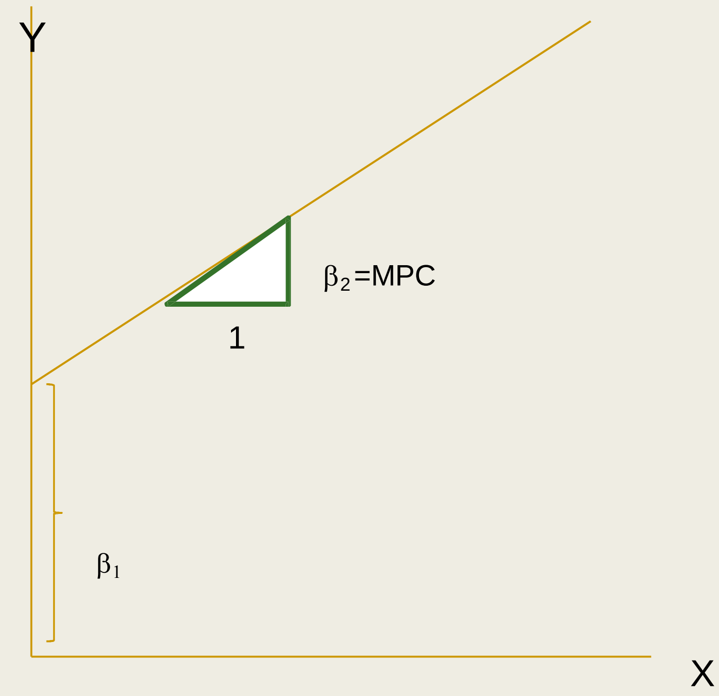
   Y = pengeluaran Konsumsi

   X = pendapatan

   β_1, β_2 = parameter persamaan (model)

   β_1= intersep

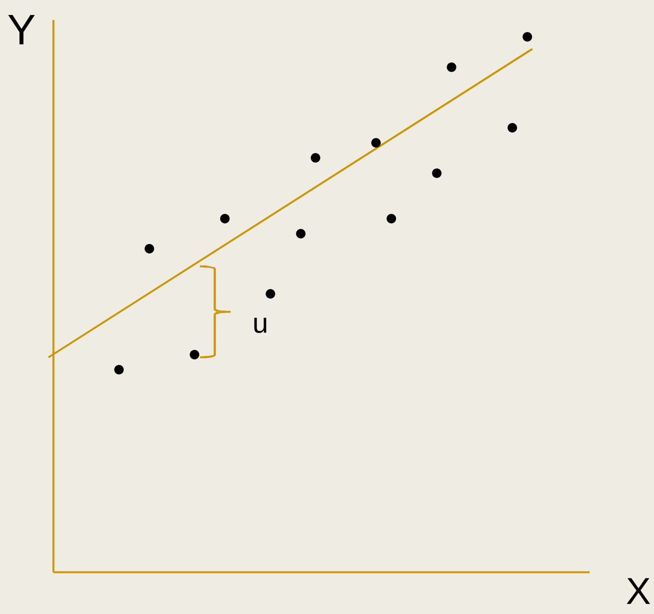
   β_2 = slope , dalam hal ini sbg MPC



3. Spesifikasi Model ekonometrika

$Y = \beta_1 + \beta_2 X + u$, $0 < \beta_2 < 1$ € bentuk persamaan: fungsi linier dimana

u = disturbance, error term yang bersifat random atau stokastik, yang menggambarkan perilaku probabilistik dari Y € model probabilistik (berlawanan dengan deterministik)

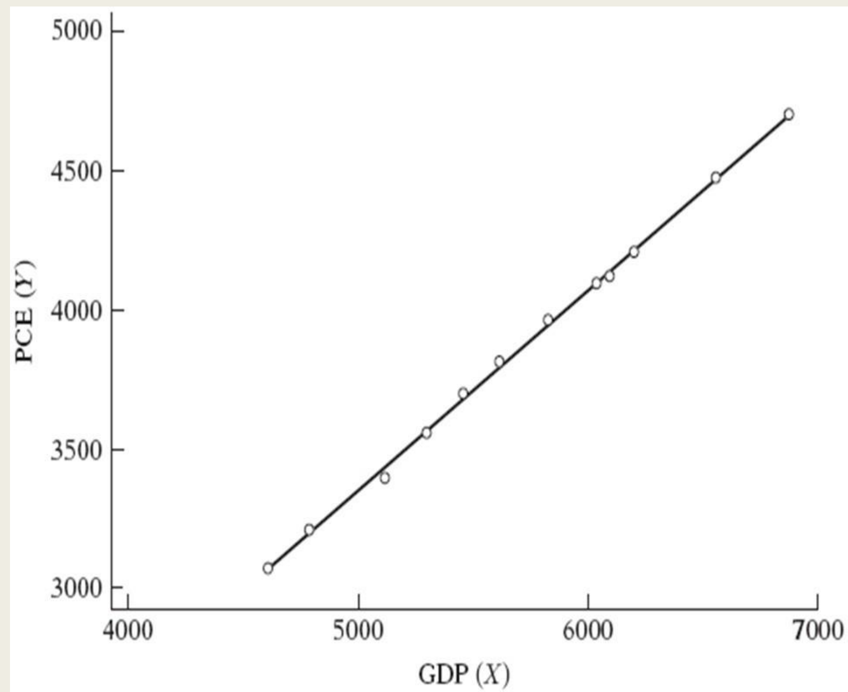


4. Pengumpulan data

Untuk memperoleh nilai numeric β*1 dan* β*2,* kita membutuhkan data. Lihat tabel berikut, yang menghubungkan antara *personal consumption expenditure* (PCE) and the *gross domestic product (GDP)*. Data dalam nilai riil.

Year	*Y*	*X*
1982	3081.5	4620.3
1983	3240.6	4803.7
1984	3407.6	5140.1
1985	3566.5	5323.5
1986	3708.7	5487.7
1987	3822.3	5649.5
1988	3972.7	5865.2
1989	4064.6	6062.0
1990	4132.2	6136.3
1991	4105.8	6079.4
1992	4219.8	6244.4
1993	4343.6	6389.6
1994	4486.0	6610.7
1995	4595.3	6742.1
1996	4714.1	6928.4

Plot Data



5. Estimasi dalam Model ekonometrika

- *Analisis regresi adalah alat utama yang digunkaan untuk memperoleh dugaan koefisien-koefisien*. Dengan metode regresi dan data pada tabel di atas diperoleh dugaan β_1 *and* β_2, *yaitu, −184.08 dan 0.7064*. dengan demikian fungsi konsumsi dugaan:
- $\hat{Y} = -184.08 + 0.7064X_i$

- Garis regresi lumayan bagus dengan slope coefficient (i.e., MPC) = 0.70, setiap peningkatan pendapatan riil 1 dollar akan meningkatkan sekitar 70 sen (secara rata- rata) dalam konsumsi riil.

6. Pengujian Hipotesis

Untuk mengetahui apakah koefisien dugaan yang diperoleh di atas sejalan dengan harapan teori maka perlu diuji. Keynes mengharapkan MPC bernilai positif tetapi kurang dari satu ($\beta_2 \leq 1$). Dalam kasus ini kita mendapati MPC = 0.70. Sebelum mengkonfirmasi teory konsumsi Keynes, kita harus meneliti apakah angka dugaan berada dibawah satu. Dengan kata lain secara statistika 0,70 kurang dari satu. Jika benar maka temuan ini mendukung Teori Keynes.

- Konfirmasi dengan cara seperti itu atau sanggahan terhadap teori ekonomi berbasis bukti sampel adalah dasar dari teori statistika yang dikenal dengan *statistical inference* (hypothesis testing).

7. Peramalan dan Prediksi

- Misalkan kita bermaksud meprediksi rata-rata belanja konsumsi untuk tahun 1997. GPD 1997 adalah $7269.8 milyar:

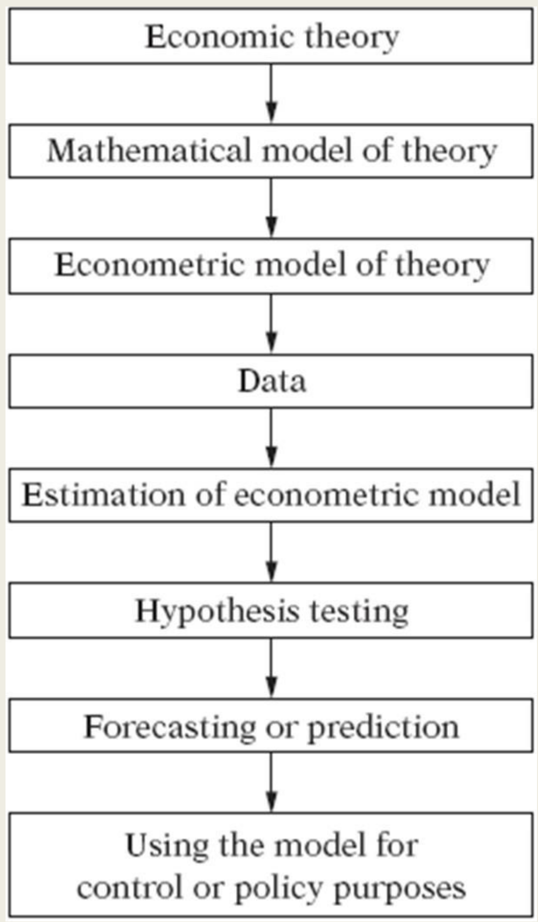
$$\hat{Y}_{1997} = -184.0779 + 0.7064\ (7269.8) = 4951.3$$

- Angka actual dari belanja konsumsi yang dilaorkan pada tahun 1997 adalah $4913.5 milyar, sementara prediksi model adalah $4951.3 milyar. Jadi kelebihan $37.82 milyar, atau kesalahan prediksi adalah $37.82 milyar atau sekitar 0.76% dari nilai actual GDP 1997

8. Menggunakan model untuk Pengendalian atau Penyusunan Kebijakan

- Dengan persamaaan terestimasi di atas, pemerintah percaya bahwa belanja konsumen sekitar $4900 milyar akan menahan laju pengangguran pada level 4.2%. Berapa level pendapatan yang akan menjamin tercapaiknya target belanja konsumsi tsb?
- Jika hasil regresi di atas dianggap benar, aritmatika sederhana akan menunjukkan :
- $4900 = -184.0779 + 0.7064X$
- Dengan menyelesaikan persamaan diperoleh *X = 7197 (kira-kira), Jadi dengan level pendapatan $7197 milyar, dan MPC =0,70,* akan melahirkan belanja konsumen sekitar $4900 milyar.

- Figure I.4 summarizes the anatomy of classical econometric modeling.



MULTIPLE REGRESSION ANALYSIS

Bentuk Umum Model

Bentuk model populasi:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + ui$$

dimana

Yi = nilai dependen variabel pada obeservasi ke-i

X_1, X_2, ..., X_k = variabel independen X1, X2,...., Xk

k = banyaknya variabel independent

β_0, β_1, β_2,...., β_k, = koefisien persamaan (sebagai parameter)

ui = gangguan acak atas nilai observasi di pengamatan ke-i.

β0 = intersep, pengaruh rata-rata untuk Y dari semua variabel yang tidak dimasukkan ke dalam model

β_1, β_2,...., β_k = partial regression coefficient

Asumsi Classical Linear Regression Model (CLRM)

- Zero mean of ui, atau
  - $E(ui|X_{1i}, X_{2i}, ...X_{ki}) = 0$ untuk setiap i
- No serial autocorrelation atau
  - Cov(ui, uj)= 0
- Homoscedasticity. Atau
  - Var(ui) = σ^2 , untuk semua i
- Zero covariance antara ui dan setiap variabel X, atau
  - Cov(ui., X1i) = cov(ui, X2i) = ...=cov(ui, Xki) = 0
- No spesification bias, atau
  - The model correctly specified
- No exact collinearity relationship between the X's variables (Explain!) atau
- Additional assumption:
  - Model is Linear in paramaters
  - Value of regressor are fixed (non stochastic)
  - There is sufficient variability in the values of regressor variables

Interpretation of Multiple Regression Equation

Given the assumption of CLRM, the expectation of the equation:

$$E(Y_i | X_{1i}, X_{2i}, ..., X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

That is the conditional mean or expected value of Y conditional upon the given or fixed value of X_1, X_2, ..., X_k

The meaning of Partial Regression Coefficient

Perhatikan kembali persamaan berikut:

$$E(Y_i \mid X_{1i}, X_{2i}, ..., X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

β_1, β_2, ..., β_k = partial slope coefficient atau partial regression coefficient

Makna koefisien:

β_1 = besarnya perubahan pada rata-rata Y atau E(Y) jika X1 berubah satu unit, dengan menjaga variabel-variabel lain tidak berubah (ceteris paribus). Atau effek langsung atau effect bersih dari perubahan satu unit pada X1 terhadap Y.

Demikian juga untuk koefisien yang lainnya. Explain

hubungan sederhana antara Y , X1 dan X2

Metode Estimasi OLS

Model regresi sampel

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + ... + \hat{\beta}_k X_{ki} + \hat{u}i$$

dimana

$\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ = estimator atau penduga bagi paramater populasi

Jika punya data sampel maka OLS kita dapat memperoleh satu set angka dugaan (estimate) bagi parameter-2 tersebut.

Bagaimana prinsip dasar metode OLS? Minimumkan SSE (atau RSS dalam Gujarati)

$$\min \sum u_i^2 = \sum (Yi - \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + ... + \hat{\beta}_k X_{ki})^2$$

Rumus penduga OLS dengan matriks

Rumus pendugaan parameter dengan solusi sistem persamaan sangat rumit.
Rumus dengan pendekatan matriks:

$$\hat{\beta} = \left(\mathbf{X'X}\right)^{-1}\left(\mathbf{X'Y}\right)$$

Kita tidak akan memperdalam proses pendugaan secara teknis. Sudah banyak software ekonometrika yang dapat mengerjakan tugas ini.

Standar Error Penduga Koefisien?

$$\begin{aligned}
\text{var-cov}(\hat{\boldsymbol\beta}) &= E[(\hat{\boldsymbol\beta}-\boldsymbol\beta)(\hat{\boldsymbol\beta}-\boldsymbol\beta)'] \\
&= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\} \qquad \textbf{(3)} \\
&= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]
\end{aligned}$$

where in the last step use is made of the fact that $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Noting that the *X*'s are nonstochastic, on taking expectation of (3) we obtain

$$\begin{aligned}
\text{var-cov}(\hat{\boldsymbol\beta}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

$E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}.$

Database Cross-Section Child Mortality

Fertility and other data for 64 Countries

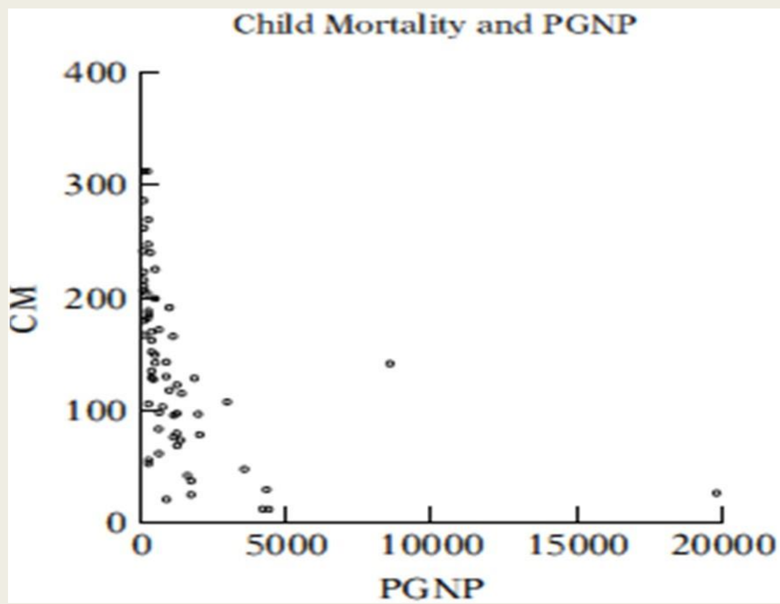
Obs	CM	FLRP	PGNP	TFR	Obs	CM	FLRP	PGNP	TFR
1	128	37	1870	6.66	33	142	50	8640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00
4	197	65	570	6.25	36	41	66	1620	3.91
5	96	76	2050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1180	3.93	43	191	31	1010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1730	3.50	45	37	88	1730	3.46
14	165	31	1150	7.41	46	103	35	780	5.66
15	94	77	1160	4.21	47	67	85	1300	4.82
16	96	80	1270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1340	7.17
24	12	81	4240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4370	2.86
27	107	87	3020	6.66	59	121	41	1310	4.88
28	72	63	1420	7.28	60	115	62	1470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19830	5.23	62	47	85	3630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

Note:

CM = Child mortality, the number of deaths of children under age 5 in a year per 1000 live births.

FLRP = Female literacy rate, percent. PGNP= per capita GNP in 1980.

TFR = total fertility rate, 1980–1985, the average number of children born to a woman, using agespecific fertility rates for a given year.



Analogikan dengan model konsumsi jasa komunikasi !

Partial Regression Coefficient

Dependent Variable: CM Method:
Least Squares Sample(adjusted): 1 64
Included observations: 64 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	168.3067	32.89165	5.117003	0.0000
PGNP	-0.005511	0.001878	-2.934275	0.0047
FLR	-1.768029	0.248017	-7.128663	0.0000
TFR	12.86864	4.190533	3.070883	0.0032
R-squared	0.747372	Mean dependent var		141.5000
Adjusted R-squared	0.734740	S.D. dependent var		75.97807
S.E. of regression	39.13127	Akaike info criterion		10.23218

Persamaaan regresi dugaan:

$CM_i = 168.307 - 0.0055PGNP - 1.768FLR + 12.869TFR + u_i$

Interpretasikan setiap koefisien!

Bagaimana menghitung Standar Error Penduga Koefisien?

Regresi dengan data distandardisasika ke variabel normal baku

For our child mortality example, the results are as follows:

$$\widehat{CM}^{*} = -\ 0.2026\ PGNP_i^{*} - 0.7639\ FLR_i^{*} \qquad \textbf{(7.6.3)}$$
$$se = \quad (0.0713) \qquad\qquad (0.0713) \qquad\qquad r^2 = 0.7077$$

Dimana:

$$CM*_i = [CM - mean(CM)]/\sigma_{CM}$$

$$FLR*_i = [FLR - mean(FLR)/\sigma_{FLR}$$

Persamaan tersebut tidak mengandung intersep krn rata-rata CM* dan FLR* adalah nol.

Makna koefisien:

- Jk PGNP naik sebesar satu standar deviasinya, maka CM akan berkurang sebesar -0.2026 std deviasinya dengan asumsi variabel FLR tidak berubah.
- Jika FLR naik 1 std deviasinya, maka CM berkurang sebesar 0.7639 std deviasinya dengan asumsi PGNP tidak berusah
- Variabel FLR memberikan pengaruh lebih kuat terhadap CM dibandingkan PGNP

Goodness of Fit (Koefisien determinasi)

Dalam kasus regresi linier sederhana, r^2, atau koefisien determinasi mengukur proporsi variasi dalam variabel Y yang dapat dijelaskan oleh variasi dalam variabel X sendiri.

Dalam model regresi berganda atau multiple regression, r^2, memiliki makna proporsi variasi dalam variabel Y yang dapat dijelaskan secara simultan oleh variasi semua variabel X atau variabel penjelas; dan diberi simbol R^2.

Sifat R^2:

a. Meningkat dengan bertambahnya variabel penjelas
b. Menurun dengan bertambahnya observasi

Untuk mengatasi sifat-sifat tersebut agar tidak keliru dalam menafsirkan maka diperkenalkan R^2 adjusted dengan rumus:

$$R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

k= banyak var bebas dalam model
n = banyak observasi

Perbandingan: $R^2adj < R^2$

R^2adj diigunakan untuk membandingkan kecocokan-suai atau kesesuaian suatu model peresamaan dengan model-2 lainnya.

- Apakah kita dapat membandingkan R^2 antara dua model berikut?

$$\ln Y_i = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u_i$$

$$Y_i = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + u_i$$

- Jawaban tidak bisa. Kenapa?
- Dua angka R^2 bisa dibandingan jika
  - Jumlah Observasi sama
  - Banyak variabel sama
  - Bentuk fungsional persamaan sama

Contoh Model Multiple Regression: Polynomial Regression Model

Bentuk Second-degree polynomial atau quadratic function

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

Bentuk k^{th}-degree polynomial:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i$$

Kasus Hubungan Output dan Biaya atau Fungsi Biaya, dengan bentuk fungsi third-degree polynomial, atau qubic function.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

TABLE 7.4 TOTAL COST (Y) AND OUTPUT (X)

Output	Total cost, \$
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

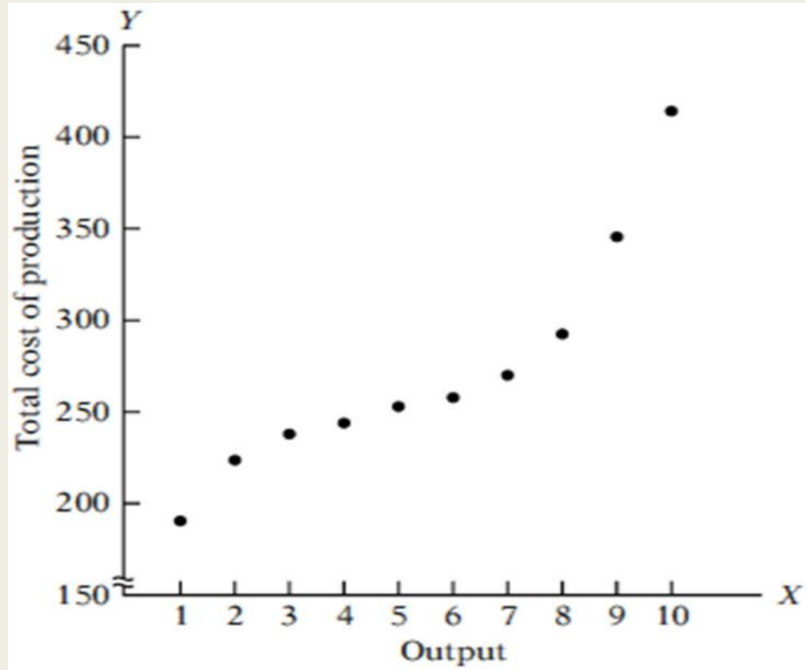


FIGURE 7.2 The total cost curve.

Hipotesis dari model, dengan mempertimbangkan scatter diagram:

1. β_0, β_1, β_3 >0
2. $\beta_2 < 0$
3. $\beta_2{}^2 < 3\ \beta_1\ \beta_3$

Hasil empiris:

$$\hat{Y}_i = \underset{(6.3753)}{141.7667} + \underset{(4.7786)}{63.4776X_i} - \underset{(0.9857)}{12.9615X_i^2} + \underset{(0.0591)}{0.9396X_i^3}$$

$$R^2 = 0.9983$$

Pengujian (hipotesis) Model

- Distribusi sampling parameter dugaan. Asumsi kenormalan dalam u:

We assumed that the ui follow the normal distribution with zero mean and constant variance σ^2. We continue to make the same assumption for multiple regression models. With the normality assumption we find that the OLS estimators of the partial regression coefficients, which are Identical with the maximum likelihood (ML) estimators, are best linear unbiased estimators (BLUE). the estimators β_2, β_3, and β_1 are normally distributed with means equal to true β_2, β_3, and β_1 and varians $\sigma^2 (X'X)^{-1}$. Oleh karena varians estimator diganti oleh penduga varians sampel maka estimators β_2, β_3, and β_1 mengikuti distribusi t-students dengan formula sebagai berikut:

$$t = \frac{\hat{\beta}_1 - \beta_1}{\mathrm{se}(\hat{\beta}_1)}$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{\mathrm{se}(\hat{\beta}_2)}$$

$$t = \frac{\hat{\beta}_3 - \beta_3}{\mathrm{se}(\hat{\beta}_3)}$$

Dengan derajat bebas: n – 3
Atau secara umum dengan derajat bebas: df=n – k – 1; dimana n=banyak observasi; dan k=banyak variabel penjelas.

Output Komputer: Contoh kasus Child Mortality

Dependent Variable: CM Method: Least Squares Date: 10/04/12 Time: 15:49
Sample(adjusted): 1 64
Included observations: 64 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	263.6416	11.59318	22.74109	0.0000
PGNP	-0.005647	0.002003	-2.818703	0.0065
FLR	-2.231586	0.209947	-10.62927	0.0000
R-squared	0.707665	Mean dependent var		141.5000
Adjusted R-squared	0.698081	S.D. dependent var		75.97807
S.E. of regression	41.74780	Akaike info criterion		10.34691
Sum squared resid	106315.6	Schwarz criterion		10.44811
Log likelihood	-328.1012	F-statistic		73.83254
Durbin-Watson stat	2.186159	Prob(F-statistic)		0.000000

P-value dari two- tails

- Berlaku untuk uji dua arah
- Jk ujinya satu arah maka P-Valu harus dibagi dua

Persamaan regresi dugaan:

CM = 263.6416 -0.00565 PGNP -2.231 FLR + u

Pengujian Koefisien regresi individu

- Disain Hipotesis:

➢ PGNP tidak berpengaruh secara linier terhadap CM, menggunakan level of significance atau taraf nyata 5% atau , $\alpha = 0.05$

H_0: $\beta_2 = 0$ vs H_1: $\beta_2 \neq 0$

Tolak Ho jk $t_{stat} > 2.000$ atau $t_{stat} < -2.000$, atau jika *p-value* < α

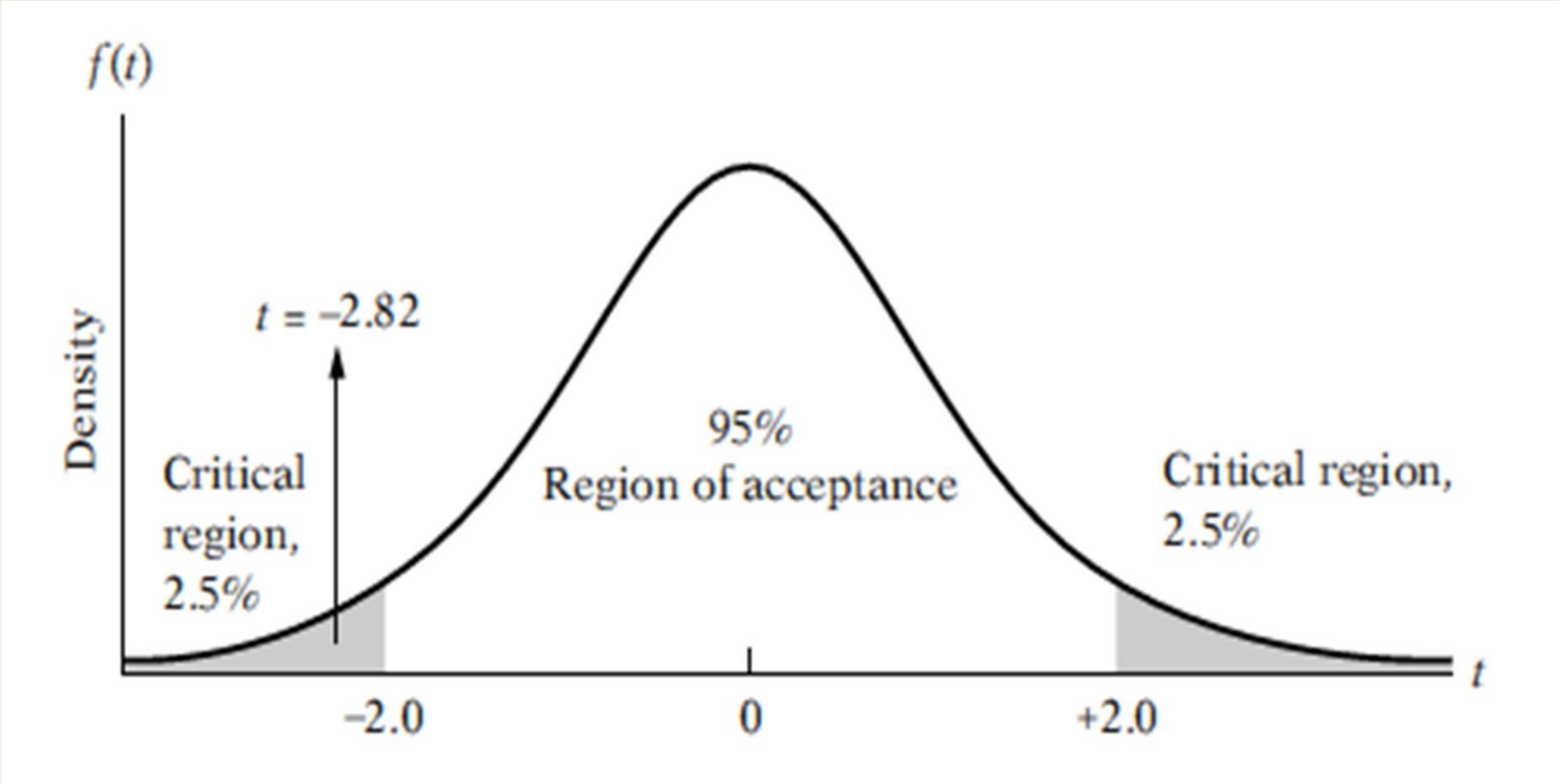
t-stast dari penduga β_2 (khusus untuk uji 2 arah) :

$$t = \frac{-0.0056}{0.0020} = -2.8187$$

Keputusan : totak Ho, karena tstat < -2.000,

Kesimpulan: koefisien $\beta 2$ yang sebenarnya berbeda dari nol , sehingga PGNP memiliki pengaruh linier signifikan terhadap CM.

- Posisi t_{stat} dalam peta daerah penerimaan dan penokan hipotesis, dalam rentang *Confidece Interval* 95% pada variabel t dengan df=60.



- Disain Hipotesis:

➢ PGNP berpengaruh negatif secara linier terhadap CM, menggunakan level of significance atau taraf nyata 5% atau , $\alpha = 0.05$

$H_0: \beta_2 >= 0$ vs $H_1: \beta_2 < 0$

Tolak Ho jk $t_{stat} < -1.671$ atau *p-value* < 2α (khusus untuk uji 1 arah). t-stast dari penduga β_2:

$$t = \frac{-0.0056}{0.0020} = -2.8187$$

Keputusan : totak Ho, karena tstat < -1.671,

Kesimpulan: koefisien β_2 yang sebenarnya negatif, sehingga PGNP memiliki pengaruh negatif signifikasi terhadap CM.

- Lakukan Pengujian Hipotesis untuk Paramater lainnya!