

BAB III

METODOLOGI PENELITIAN

A. Tujuan Penelitian

Berdasarkan masalah-masalah yang telah peneliti rumuskan, maka tujuan penelitian ini adalah untuk:

1. Mengetahui besarnya pengaruh tingkat pendidikan terhadap kemiskinan.
2. Mengetahui besarnya pengaruh pendapatan per kapita terhadap kemiskinan.
3. Mengetahui besarnya pengaruh tingkat pendidikan dan pendapatan per kapita terhadap kemiskinan.

B. Objek Penelitian

Penelitian ini dilaksanakan dengan mengambil data Tingkat Pendidikan Jawa Barat, Pendapatan Per kapita Jawa Barat dan Kemiskinan Jawa Barat di BPS Pusat (Badan Pusat Statistik) yang berada di DKI Jakarta.

Objek penelitian dalam penelitian ini adalah 26 Kabupaten/Kota yang ada di Provinsi Jawa Barat dan dilakukan penelitian hanya 25 Kabupaten/Kota karena tahun 2006 sampai 2008 Kabupaten Bandung Barat belum melakukan pemekaran. Lokasi ini dipilih karena jumlah penduduknya terbanyak pertama dan kemiskinan di Jawa Barat merupakan terbesar ketiga di Pulau Jawa setelah Provinsi Jawa Timur.

C. Metode Penelitian

Metode yang digunakan dalam penelitian adalah metode ekspos fakto dengan pendekatan korelasional. Metode ini dipilih karena merupakan metode yang sistematis dan empiris. Metode ekspos fakto adalah suatu penelitian yang dilakukan untuk meneliti peristiwa yang telah terjadi dan kemudian meruntut ke belakang untuk mengetahui faktor-faktor yang dapat menimbulkan kejadian tersebut.⁶¹ Sehingga dengan pendekatan korelasional ini, akan dapat dilihat pengaruh antara tiga variabel yaitu, variabel bebas (Tingkat Pendidikan dan Pendapatan Perkapita) yang mempengaruhi dan diberi simbol X1, X2, dan variabel terikat (Kemiskinan) yang dipengaruhi dan diberi simbol Y.

D. Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder yang bersifat kuantitatif yaitu data yang telah tersedia dalam bentuk angka. Sedangkan data yang digunakan dalam penelitian ini adalah data panel yang merupakan gabungan antara data *time series* (runtun waktu) dan data *cross section* (silang).⁶² Data *time series* sebanyak lima tahun dari tahun 2007-2011 dan data *cross section* sebanyak 25 Kabupaten/Kota di Jawa Barat. Data tersebut diperoleh dari studi kepustakaan yang bersumber dari publikasi yang diterbitkan oleh Badan Pusat Statistik (BPS) untuk memperoleh data tingkat pendidikan, pendapatan perkapita dan kemiskinan.

⁶¹Sugiyono. *Metode Penelitian Bisnis*. Jakarta: Alfabeta. 2004. p. 7

⁶²Nachrowi, *Pendekatan Populer dan Praktis Ekonometrika untuk Analisis Ekonomi dan Keuangan* (Jakarta: LPFE UI, 2006), p. 309

Data tingkat pendidikan berupa angka rata-rata lama sekolah di Jawa Barat yang diambil selama 5 tahun dan sebanyak 25 Kabupaten/Kota, data pendapatan per kapita pada PDRB Harga Konstan dibagi dengan jumlah penduduk Jawa Barat selama 5 tahun dan sebanyak 25 Kabupaten/Kota, dan data kemiskinan pada jumlah penduduk miskin Jawa Barat yang dilihat selama 5 tahun dan sebanyak 25 Kabupaten/Kota.

E. Teknik Operasional Variabel

1. Tingkat Pendidikan

a. Definisi Konseptual

Tingkat Pendidikan adalah urutan pendidikan (proses untuk mendapatkan pengetahuan dan pemahaman lebih tinggi) formal mulai dari pendidikan dasar sampai dengan pendidikan tinggi dengan memperhitungkan tingkat pendidikan yang sedang diduduki dan pendidikan tertinggi yang ditamatkan.

b. Definisi Operasional

Tingkat Pendidikan adalah kegiatan individu dalam pembelajaran yang diukur dari rata-rata lama sekolah yakni perbandingan banyaknya penduduk usia 15 tahun ke atas yang menjalani pendidikan di Jawa Barat sebanyak 25 Kabupaten/Kota dan diambil selama 5 tahun dari tahun 2006 sampai 2010 yang diambil dari Badan Pusat Statistik.

2. Pendapatan Per Kapita

1. Definisi Konseptual

Pendapatan perkapita adalah proses pertambahan pendapatan masyarakat yang terjadi di wilayah tersebut, yaitu kenaikan seluruh nilai tambah yang terjadi di wilayah (daerah) tersebut, pertambahan pendapatan ini diukur dalam nilai riil (dinyatakan dalam harga konstan). Pendapatan per kapita didapat dari hasil bagi pendapatan regional dengan jumlah penduduk tahun tersebut.

2. Definisi Operasional

Pendapatan perkapita adalah proses pertambahan pendapatan pada PDRB (harga konstan) yang dibagi dengan jumlah penduduk yang diambil dari 25 Kabupaten/Kota selama 5 tahun dari tahun 2006 sampai 2010 yang diambil dari Badan Pusat Statistik.

3. Kemiskinan

1. Definisi Konseptual

Kemiskinan adalah sebuah kondisi yang berada di bawah nilai standar kebutuhan minimum, baik untuk makanan dan non makanan yang disebut garis kemiskinan (*poverty line*) atau batas kemiskinan (*poverty threshold*) dari kebutuhan yang bersifat materil.

2. Definisi Operasional

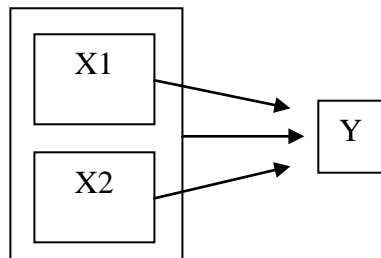
Kemiskinan adalah suatu kondisi seseorang atau kelompok masyarakat yang mengalami berbagai kekurangan (ketidakberdayaan) yang dilihat dari

jumlah penduduk miskin di Jawa Barat sebanyak 25 Kabupaten/Kota selama 5 tahun dari tahun 2006 sampai 2010 diambil dai Badan Pusat Statistik.

F. Konstelasi Pengaruh Antar Variabel

Variabel penelitian ini terdiri dari tiga variabel yaitu variabel bebas (tingkat pendidikan digambarkan dengan simbol X1 dan pendapatan perkapita digambarkan dengan simbol X2) dan variabel terikat (kemiskinan yang digambarkan dengan simbol Y).

Sesuai dengan hipotesis yang diajukan bahwa terdapat pengaruh variabel X1 dan X2 terhadap variabel Y, maka konstelasi pengaruh variabel X1 dan X2 terhadap variabel Y adalah:



Keterangan:

- (X1) : Tingkat Pendidikan
(X2) : Pendapatan Per Kapita
(Y) : Kemiskinan
→ : Arah Pengaruh

G. Teknik Analisis Data

1. Panel Data

Data panel adalah jenis data yang merupakan gabungan antara data runtun waktu (*time series*) dan lintas sektor (*cross section*). Oleh karenanya, data panel memiliki gabungan karakteristik kedua jenis data tadi, yaitu terdiri atas beberapa objek dan meliputi beberapa periode.⁶³ Estimasi panel data akan meningkatkan derajat kebebasan, mengurangi kolinearitas antara variabel penjelas dan memperbaiki efisiensi estimasi.

Keunggulan penggunaan data panel dibanding data runtun waktu dan data lintas sektor adalah :

- 1) Estimasi data panel dapat menunjukkan adanya heterogenitas dalam tiap unit.
- 2) Dengan data panel, data lebih informatif, mengurangi kolinieritas antara variabel, meningkatkan derajat kebebasan dan lebih efisien.
- 3) Data panel cocok digunakan untuk menggambarkan adanya dinamika perubahan.
- 4) Data panel dapat lebih mampu mendeteksi dan mengukur dampak.
- 5) Data panel bisa digunakan untuk studi dengan model yang lebih lengkap.
- 6) Data panel dapat meminimumkan bias yang mungkin dihasilkan dalam regresi.

Data panel merupakan gabungan dari data *cross section* dan data *time series*, dalam hal modelnya dapat dituliskan sebagai berikut:

⁶³ Wing Wahyu Winarno, *Analisis Ekonometrika dan Statistika dengan Eviews Edisi Ke-3* (Yogyakarta; STIM YKPN, 2011), p.1.2

$$Y = \alpha + \beta x_1 + \beta x_2 + \varepsilon_{it} \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \quad (3.1)$$

dimana :

Y	= Kemiskinan
N	= banyaknya observasi (provinsi)
T	= banyaknya waktu
N x T	= banyaknya data panel
TP	= Tingkat Pendidikan
PK	= Pendapatan Per Kapita

Untuk mengestimasi parameter model dengan data panel, terdapat beberapa teknik yang ditawarkan, yaitu:

a. Model *Common Effect*

Model *common effects* atau *pooled regression* merupakan model regresi data panel yang paling sederhana. Model ini pada dasarnya mengabaikan struktur panel dari data, sehingga diasumsikan bahwa perilaku antar individu sama dalam berbagai kurun waktu atau dengan kata lain pengaruh spesifik dari masing-masing individu diabaikan atau dianggap tidak ada. Dengan demikian, akan dihasilkan sebuah persamaan regresi yang sama untuk setiap unit *cross-section*. Sesuatu yang secara realistis tentunya kurang dapat diterima. Karena itu, model ini sangat jarang digunakan dalam analisis data panel. Persamaan regresi untuk model *common effect* dapat dituliskan sebagai berikut:

$$Y = \alpha + \beta x_1 + \beta x_2 + \varepsilon_{it} \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \quad (3.2)$$

dimana :

Y	= variabel dependen (Kemiskinan)
α	= koefisien regresi
x_1	= variable independen (Tingkat Pendidikan)
x_2	= variable independen (Pendapatan Per Kapita)
β	= estimasi parameter

ε_{it}	= <i>aerror term</i>
N	= jumlah (provinsi)
T	= jumlah periode waktu.

Berdasarkan asumsi struktur matriks varians-covarians residual, maka pada model *common effects*, terdapat 4 metode estimasi yang dapat digunakan, yaitu:

- (1) *Ordinary Least Square (OLS)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat homoskedastik dan tidak ada *cross sectional correlation*,
- (2) *Generalized Least Square (GLS) / Weighted Least Square (WLS): Cross Sectional Weight*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan tidak ada *cross sectional correlation*,
- (3) *Feasible Generalized Least Square (FGLS)/Seemingly Uncorrelated Regression (SUR)* atau *Maximum Likelihood Estimator (MLE)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan ada *cross sectional correlation*,
- (4) *Feasible Generalized Least Square (FGLS)* dengan proses *autoregressive (AR)* pada *error term*-nya, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan ada korelasi antar waktu pada residualnya.

b. Model Fixed Effect

Jika model *common effect* cenderung mengabaikan struktur panel dari data dan pengaruh spesifik masing-masing individu, maka model *fixed effect*

adalah sebaliknya. Pada model ini, terdapat efek spesifik individu α_i dan diasumsikan berkorelasi dengan variabel penjelas yang teramati X_{it} .

Berdasarkan asumsi struktur matriks varians-kovarians residual, maka pada model *fixed effect*, terdapat 3 metode estimasi yang dapat digunakan, yaitu:

- (1) *Ordinary Least Square (OLS/LSDV)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat homoskedastik dan tidak ada *cross sectional correlation*,
- (2) *Weighted Least Square (WLS)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan tidak ada *cross sectional correlation*,
- (3) *Seemingly Uncorrelated Regression (SUR)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan ada *cross sectional correlation*.

c. Model *Random Effect*

Keputusan untuk memasukkan peubah *dummy* dalam model *fixed effects* akan menimbulkan konsekuensi tersendiri yaitu dapat mengurangi banyaknya derajat kebebasan yang pada akhirnya akan mengurangi efisiensi dari parameter yang diestimasi. Untuk mengatasi masalah tersebut maka dapat digunakan model *random effects*.

Dalam model ini, parameter yang berbeda antar individu maupun antar waktu dimasukkan ke dalam *error*, karena hal inilah model ini sering juga

disebut sebagai *error component model*. Bentuk model *random effects* dapat dijelaskan dengan persamaan berikut:

$$Y = \alpha + \beta x_1 + \beta x_2 + \varepsilon_{it} \quad \varepsilon_{it} = u_i + v_t + w_{it} \quad (3.3)$$

dimana:

$$\begin{aligned} u_i &\sim N(0, \delta U^2) = \text{error component cross section} \\ v_t &\sim N(0, \delta V^2) = \text{error component time series} \\ w_{it} &\sim N(0, \delta W^2) = \text{error component combinations} \end{aligned}$$

Melihat persamaan di atas, maka dapat dinyatakan bahwa Model Efek Random menganggap efek rata-rata dari data *cross-section* dan *time-series* direpresentasikan dalam *intercept*. Sedangkan deviasi efek secara random untuk data *time-series* direpresentasikan dalam v_t dan deviasi untuk data *cross-section* dinyatakan dalam u_i .

Telah diketahui bahwa: $\varepsilon_{it} = u_i + v_t + w_{it}$. Dengan demikian varians dari *error* tersebut dapat dituliskan dengan:

$$\text{Var}(\varepsilon_{it}) = \delta U^2 + \delta V^2 + \delta W^2 \quad (3.4)$$

Hal ini tentunya berbeda dengan model OLS yang diterapkan pada panel data, sebagaimana telah dijelaskan di atas, yang mempunyai varian *error* sebesar:

$$\text{Var}(\varepsilon_{it}) = \delta W^2 \quad (3.5)$$

Dengan demikian, Model Efek Random bisa diestimasi dengan OLS bila $\delta U^2 = \delta V^2 = 0$. Kalau tidak demikian, Model Efek Random diestimasi dengan metode *Generalized Least Square* (GLS). Asumsi yang digunakan dalam Model Efek Random ini adalah *error* secara individual tidak saling berkorelasi, begitu pula dengan *error* kombinasinya. Penggunaan pendekatan

random effects dapat menghemat derajat kebebasan dan tidak mengurangi jumlahnya seperti pada pendekatan *fixed effects*. Hal ini berimplikasi pada parameter hasil estimasi akan menjadi efisien. Semakin efisien maka model akan semakin baik. Terkait dengan beberapa pilihan teknik untuk permodelan panel data, sebelum model diestimasi dengan model yang tepat, terlebih dahulu dilakukan uji spesifikasi apakah *Common Effect*, *Fixed Effect* dan atau *Random Effect* memberikan hasil yang sama. Penyeleksian model estimasi data panel antara lain:

1) Uji Chow

Uji Chow dapat digunakan untuk memilih teknik dengan metode pendekatan *Pooled Least Square* (PLS) atau metode *Fixed Effect* (FE).

Prosedur Uji Chow adalah sebagai berikut:

- a. Buat hipotesis dari Uji Chow

$H_0 = \text{model } common \text{ effect}$

$H_1 = \text{model } Fixed \text{ Effect}$

- b. Menentukan kriteria uji

H_0 Apabila nilai F statistik $>$ F tabel, maka hipotesis ditolak yang artinya kita harus memilih teknik FE.

H_1 Apabila nilai F statistik $<$ F tabel, maka hipotesis diterima yang artinya kita harus memilih teknik PLS.

2) Uji Hausman

Uji Hausman digunakan untuk memilih antara metode pendekatan *Fixed Effect* (FE) atau *Random Effect* (RE). Prosedur Uji Hausman adalah sebagai berikut:

- a. Buat hipotesis dari Uji Hausman: =*random effect* dan =*fixed effect*.
- b. Menentukan kriteria uji: apabila *Chi-square* statistik > *Chi-square* tabel dan *p-value* signifikan, maka hipotesis ditolak, sehingga metode FE lebih tepat untuk digunakan. Dan apabila *Chi-square* statistik < *Chi-square* tabel dan *p-value* signifikan, maka hipotesis diterima, sehingga metode RE lebih tepat untuk digunakan.

2. Uji Hipotesis

a. Uji F

Uji statistik F merupakan pengujian koefisien regresi secara keseluruhan. Pengujian ini menunjukkan apakah semua variabel bebas yang dimasukkan kedalam model mempunyai pengaruh secara bersama-sama terhadap variabel terikat.

Hipotesis pengujian yang digunakan adalah:

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

H_1 : paling sedikit salah satu nilai $\beta_i \neq 0$, dengan $i = 1, 2, \dots, k$.

Statistik uji F dihitung dengan formula sebagai berikut:

$$F_{(n-k-1, nT-nk)} = \frac{R^2 / (n + k - 1)}{(1 - R^2) / (nT - n - k)} \quad (3.10)$$

Hipotesis nol ditolak jika yang berarti bahwa minimal ada satu variabel bebas yang signifikan berpengaruh terhadap variabel tidak bebasnya. Keputusan ini dapat juga didasarkan pada perbandingan nilai *p-value* dengan tingkat signifikansinya. Hipotesis nol ditolak jika nilai *p-value* lebih kecil dari tingkat signifikansi (α).

b. Koefisien determinasi (R^2)

R^2 digunakan untuk mengukur kebaikan atau kesesuaian suatu model persamaan regresi. Besaran R^2 dihitung dengan rumus:

$$R^2 = \frac{\sum \left(\hat{Y}_i - \bar{Y} \right)^2}{\sum \left(Y_i - \bar{Y} \right)^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \quad (3.11)$$

Dan R^2 *adjusted* dihitung dengan rumus:

Keterangan :

ESS: Jumlah kuadrat yang dijelaskan

RSS: Jumlah kuadrat residual.

TSS: Jumlah kuadrat total.

n: Jumlah observasi (negara)

T: Jumlah periode waktu.

k: Banyaknya variabel bebas tanpa intersep.

Adjusted R² digunakan karena sudah menghilangkan pengaruh penambahan variabel bebas dalam model, karena nilai R^2 akan terus naik seiring dengan penambahan variabel bebas. Karena itu kita harus berhati-hati dalam menggunakan nilai R^2 ketika menilai kebaikan dan kesesuaian suatu model persamaan regresi. Penggunaan *adjusted R²* sudah memperhitungkan jumlah derajat bebas.

Dimana nilai R^2 terletak diantara 0 sampai dengan 1, nilai $0 \leq R^2 \leq 1$. Jika $R^2 = 0$, berarti variabel bebas tidak bisa menjelaskan perubahan variabel terikat, maka model dapat dikatakan buruk. Jika $R^2 = 1$, berarti variabel bebas mampu menjelaskan perubahan variabel terikat dengan sempurna. Kondisi

seperti dua hal tersebut hampir sulit diperoleh. Kecocokan model dapat dikatakan lebih baik kalau R^2 semakin dekat dengan 1.

c. Uji t

Uji t digunakan untuk mengetahui apakah variabel-variabel bebas secara parsial berpengaruh signifikan terhadap variabel tak bebasnya.

Hipotesis pengujian:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Statistik uji yang digunakan adalah statistik uji t -student. Adapun formulanya adalah sebagai berikut:

$$t_{\text{hitung}} = \frac{\hat{\beta}_i}{\text{se}(\hat{\beta}_i)} \quad (3.12)$$

adalah nilai penduga parameter ke- i , adalah simpangan baku dari nilai penduga parameter ke- i .

Hipotesis nol ditolak jika .Keputusan ini dapat juga didasarkan pada perbandingan nilai p -value dengan tingkat signifikansinya (α).Hipotesis nol ditolak jika nilai p -value lebih kecil dari (α) .Hal ini berarti secara parsial variabel bebas ke- i signifikan memengaruhi variabel tidak bebasnya dengan tingkat kepercayaan sebesar $(1-\alpha) \times 100$ persen.

3. Uji Asumsi klasik

Untuk membangun persamaan regresi panel yang terbaik dari kriteria ekonometrika, perlu dilakukan penyelidikan dan penanganan adanya

masalah-masalah yang berkaitan dengan pelanggaran asumsi dasar. Berikut ini adalah asumsi-asumsi yang diperlukan dalam analisis regresi:

a. Normalitas

Pengujian normalitas dilakukan untuk mengetahui apakah residual berdistribusi normal atau tidak. Berdasarkan dalil limit pusat (*central limit theorem*), ada kecenderungan *residual* yang terjadi sebenarnya menyebar secara normal. Jika residual ε , merupakan jumlah *residual* dari beberapa sumber maka apapun sebaran peluang masing-masing *residual* itu, akan mendekati sebaran normal bila komponen *residual*-nya semakin banyak. Pemeriksaan kenormalan terhadap residual dapat dilakukan menggunakan plot persentil-persentil (P-P Plot). Jika plot mengikuti garis lurus, maka residual mengikuti sebaran normal (Draper dalam Firmansyah, 2009).

Pengujian normalitas dilakukan untuk mengetahui apakah residual berdistribusi normal atau tidak. Berdasarkan dalil limit pusat (*central limit theorem*), ada kecenderungan residual yang terjadi sebenarnya menyebar secara normal. Jika residual merupakan jumlah residual dari beberapa sumber, maka apapun sebaran peluang masing-masing residual itu, akan mendekati sebaran normal bila komponen residual semakin banyak. Salah satu metode yang dapat digunakan untuk menguji Normalitas adalah *Jarque-Bera test*. Uji statistik ini dapat dihitung dengan rumus berikut:

$$JB = \frac{n}{6} \left[S^2 + \frac{(K-3)^2}{4} \right] \quad (3.13)$$

Di mana:

n = jumlah sampel

S = kemencengan

K = peruncingan

Jarque-Bera test mempunyai distribusi *chi square* dengan derajat bebas dua.

Jika hasil *Jarque-Bera test* lebih besar dari nilai *chi square* pada $\alpha=5$ persen,

maka tolak hipotesis nul yang berarti tidak berdistribusi normal. Jika hasil

Jarque- Bera test lebih kecil dari nilai *chi square* pada $\alpha=5$ persen, maka

terima hipotesis nul yang berarti *error term* berdistribusi normal.