

## **BAB III**

### **METODOLOGI PENELITIAN**

Pada bab ini akan dibahas mengenai ruang lingkup penelitian, metode pengumpulan data, dan model. Selain itu, akan dijelaskan pula deskripsi variabel, teknik pengolahan data, hingga pengujiannya. Metode yang digunakan didasarkan pada kajian teoritis dan empiris untuk menjawab permasalahan yang akan diteliti.

#### **3.1. Ruang Lingkup Penelitian**

Penulis membatasi ruang lingkup permasalahan yang akan di bahas untuk mendapatkan hasil dan pembahasan penelitian yang baik dan tercapainya tujuan dalam penelitian ini. Ruang lingkup permasalahan dalam penelitian ini mencakup 35 kabupaten/kota di Provinsi Jawa Tengah. Data dalam penelitian ini menggunakan data panel sekunder tahun 2010 hingga tahun 2015.

Adapun penelitian ini dilakukan untuk menguji dan menganalisis pengaruh dari Produk Domestik Regional Bruto (PDRB) per kapita, pendidikan yang diukur dengan rata-rata lama sekolah, kesehatan yang diukur dengan angka harapan hidup, tingkat pengangguran terbuka, serta upah minimum terhadap kemiskinan di 35 kabupaten/kota di Provinsi Jawa Tengah pada tahun 2010 hingga 2015.

#### **3.2. Metode Pengumpulan Data**

Data yang digunakan dalam penelitian ini adalah data sekunder berjenis data panel dengan jumlah deret waktu 6 tahun yaitu dari tahun 2010 hingga 2015 dan 35 unit *cross section* yang diperoleh dari *website*, buku, dan publikasi Badan Pusat

Statistik (BPS) Provinsi Jawa Tengah yang berhubungan dengan topik penelitian ini. Data yang dikumpulkan untuk penelitian ini adalah data tingkat kemiskinan, PDRB per kapita, rata-rata lama sekolah, angka harapan hidup, tingkat pengangguran terbuka, dan upah minimum kabupaten/kota di Provinsi Jawa Tengah pada tahun 2010 – 2015.

### 3.3. Model Penelitian

Model yang akan digunakan dalam penelitian ini merujuk pada model penelitian sebelumnya dengan menggunakan *Panel Data Regression Model*. Model yang digunakan merupakan pengembangan dari model yang dikembangkan oleh Olavarria-gambi (2003), Bakhtiari & Meisami (2010), Agrawal (2008), Stevans & Sessions (2001), Marinho, dkk. (2017), dan Miranti (2017) yang digunakan oleh penulis sebagai referensi dengan adanya modifikasi pada variabel-variabel yang digunakan serta mengkombinasikan variabel. Berikut adalah spesifikasi model untuk menganalisis bagaimana faktor-faktor yang diidentifikasi mempengaruhi tingkat kemiskinan

$$POV_{it} = \beta_1 + \beta_2 \ln GRDPCap_{it} + \beta_3 Educ_{it} + \beta_4 Health_{it} + \beta_5 Unem_{it} + \beta_6 \ln MinWage_{it} + U_{it}$$

Di mana,

*POV* : Tingkat kemiskinan per kabupaten/kota di Provinsi Jawa Tengah

*GRDPCap* : Produk Domestik Regional Bruto (PDRB) per kapita per kabupaten/kota di Provinsi Jawa Tengah

*Educ* : Pendidikan per kabupaten/kota di Provinsi Jawa Tengah

<i>Health</i>	: Kesehatan per kabupaten/kota di Provinsi Jawa Tengah
<i>Unem</i>	: Tingkat pengangguran terbuka per kabupaten/kota di Provinsi Jawa Tengah
<i>MinWage</i>	: Upah minimum per kabupaten/kota di Provinsi Jawa Tengah
<i>ln</i>	: Notasi dalam bentuk logaritma
$\beta_1$	: Konstanta
$\beta_2; \beta_3; \beta_4; \beta_5; \beta_6$	: Koefisien regresi variabel independen
<i>U</i>	: <i>Error term</i>
<i>i</i>	: Kabupaten/kota
<i>t</i>	: Data <i>time series</i> tahun 2010 – 2015

### 3.4. Operasionalisasi Variabel

Dalam penelitian ini, akan menggunakan tingkat kemiskinan sebagai variabel dependen. Variabel dependen adalah variabel terikat atau variabel yang dipengaruhi oleh variabel independen (Kothari, 2004). Variabel independen yang akan digunakan dalam penelitian ini yaitu PDRB per kapita, pendidikan yang diukur dengan rata-rata lama sekolah, kesehatan yang diukur dengan angka harapan hidup, tingkat pengangguran terbuka, dan upah minimum. Adapun definisi operasional dari berbagai variabel tersebut adalah sebagai berikut:

**Tabel 3.1. Operasionalisasi Variabel dalam Penelitian**

No.	Variabel	Keterangan	Definisi	Sumber Data	Satuan
1	<i>POV</i>	Tingkat Kemiskinan	Persentase jumlah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan. Tingkat kemiskinan yang digunakan didapat dengan konsep <i>Head Count Index</i> yang merupakan persentase penduduk yang berada di bawah garis kemiskinan	BPS Provinsi Jawa Tengah	%
2	<i>GRDPCap</i>	PDRB per kapita	PDRB harga konstan berdasarkan lapangan usaha yang dibagi dengan jumlah penduduk	BPS Provinsi Jawa Tengah	Rp
3	<i>Educ</i>	Pendidikan yang diukur dengan rata-rata lama sekolah	Rata-rata jumlah tahun belajar yang telah diselesaikan oleh penduduk yang berusia 25 tahun ke atas dan tidak termasuk tahun yang mengulang	BPS Provinsi Jawa Tengah	Tahun
4	<i>Health</i>	Kesehatan yang diukur dengan angka harapan hidup	Rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir.	BPS Provinsi Jawa Tengah	Tahun
5	<i>Unem</i>	Tingkat Pengangguran Terbuka	Persentase jumlah pengangguran terhadap jumlah angkatan kerja (penduduk usia 15 tahun ke atas). Pengangguran terbuka meliputi penduduk yang tidak punya pekerjaan dan sedang mencari pekerjaan, atau sedang mempersiapkan suatu usaha, atau merasa tidak mungkin	BPS Provinsi Jawa Tengah	%

			mendapatkan pekerjaan, atau sudah punya pekerjaan tetapi belum mulai bekerja		
6	<i>MinWage</i>	Upah Minimum	Upah minimum bulanan yang terdiri dari upah pokok termasuk tunjangan tetap yang berlaku di suatu daerah dalam penelitian ini adalah kabupaten/kota.	BPS Provinsi Jawa Tengah	Rp

Sumber: Definisi berdasarkan Badan Pusat Statistik

### 3.5. Teknik Pengolahan Data

Data dalam penelitian ini diolah menggunakan pendekatan kuantitatif dengan metode data panel. Metode ini menggabungkan dua jenis data, yaitu data berupa deret waktu (*time series*) dan data individu (*cross section*). Menurut Baltagi dalam Gujarati (2009), penggunaan data panel dalam suatu penelitian memiliki beberapa keunggulan, yaitu:

1. Estimasi data panel dapat menunjukkan heterogenitas dalam setiap individu.
2. Data panel akan memberikan informasi yang lebih banyak, variasi yang lebih banyak, kolinearitas antar variabel sedikit, dan *degree of freedom* yang lebih banyak sehingga akan lebih efisien.
3. Metode panel lebih baik untuk menentukan perubahan dinamis dibanding metode *cross section*.
4. Data panel merupakan data yang paling baik untuk mendeteksi dan mengukur dampak yang secara sederhana tidak dapat dilihat pada data *time series* dan *cross section*.
5. Data panel membantu untuk analisis studi yang lebih kompleks.

6. Data panel dapat meminimumkan bias yang bisa terjadi jika kita meregresi dalam jumlah yang besar.

Estimasi model regresi dengan menggunakan data panel menurut Gujarati (2009) dilakukan melalui tiga pendekatan, di mana pendekatan tersebut akan dianalisa untuk menentukan metode regresi data panel yang tepat yaitu:

- 1) Model *Pooled Least Square*

Model ini merupakan model yang sama dengan prinsip *Ordinary Least Square* yaitu dengan cara meminimalkan jumlah *error* kuadrat. Dalam hal ini, data yang digunakan adalah data panel yang diterapkan dalam bentuk *pooled* sehingga bukan hanya data yang bersifat *time series* saja atau data yang bersifat *cross section* saja. Estimasi model regresi data panel adalah yang paling sederhana dengan asumsi bahwa *intercept* dan koefisien *slope* atau parameter adalah konstan antar waktu dan *cross section* (*common effect*). Model estimasi ini dituliskan dalam bentuk persamaan berikut:

$$Y_{it} = \alpha_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_n X_{nit} + \mu_{it}$$

- 2) *Fixed Effect Model*

Model ini adalah estimasi model regresi pada data panel dengan asumsi bahwa koefisien *slope* adalah konstan sementara nilai *intercept* diasumsikan berbeda antar unit *cross-section* namun *intercept* diasumsikan bernilai tetap antar waktu (*fixed effect*). Terkait kendala asumsi *pooled least square* yang sulit dipenuhi maka agar dapat menghasilkan nilai koefisien *slope* atau parameter yang berbeda antar unit *cross section*, metode yang sering digunakan yaitu dengan

memberikan variabel *dummy* yang disebut *fixed effect model* atau *least square dummy variable* (LSDV) dengan bentuk persamaan sebagai berikut:

$$Y_{it} = \alpha_1 + \alpha_2 D_2 + \dots + \alpha_n D_n + \beta_2 X_{2it} + \dots + \beta_n X_{nit} + \mu_{it}$$

### 3) *Random Effect Model*

Model ini mengasumsikan bahwa nilai koefisien *slope* adalah konstan sementara nilai *intercept* berbeda antar individu dan antar waktu atau bersifat *random effect*. Dalam hal ini, asumsi dilakukan karena upaya memasukkan variabel *dummy* ke dalam *fixed effect model* dapat berakibat pada berkurangnya derajat kebebasan di mana pada akhirnya akan mengurangi efisiensi parameter yang diestimasi. Dalam model ini, parameter yang berbeda antar unit *cross section* maupun antar waktu dimasukkan ke dalam *error*. Bentuk model dituliskan dalam persamaan berikut:

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \dots + \beta_n X_{nit} + e_{it} + \mu_{it}$$

Di mana  $e_{it} + \mu_{it}$  merupakan faktor kesalahan yang acak. Faktor kesalahan tersebut terdiri dari komponen spesifik dari individu yang konstan sepanjang waktu ( $e_{it}$ ) dan komponen faktor kesalahan yang bervariasi antar individu dan waktu. Baik  $e_{it}$  maupun  $\mu_{it}$ , diasumsikan tidak memiliki hubungan dengan  $X_{it}$ .

Adapun data panel yang digunakan dalam penelitian ini adalah data sekunder yang didapatkan oleh penulis dan data tersebut akan digunakan untuk analisis deskriptif yang kemudian diolah menggunakan bantuan *software Microsoft Excel* dan *STATA 13*. Setelah dilakukan pengolahan data dengan metode panel maka dilakukan tiga tahap pengujian, yaitu:

## 1) Uji Chow

Pengujian ini dilakukan untuk mengetahui model manakah di antara model *pooled least square* atau *fixed effect model* yang lebih baik digunakan dalam penelitian ini. Hipotesis yang digunakan dalam Uji Chow adalah sebagai berikut:

$H_0$ : Model *pooled least square* lebih baik

$H_A$ : *Fixed effect model* lebih baik

Dengan kriteria sebagai berikut:

1. Apabila  $(Prob > F) < \alpha$  maka  $H_0$  ditolak yang artinya *fixed effect model* lebih baik digunakan dari pada model *pooled least square*.
2. Apabila  $(Prob > F) \geq \alpha$  maka  $H_0$  tidak dapat ditolak yang artinya model *pooled least square* lebih baik digunakan dari pada model *fixed effect model*.

## 2) Uji Hausman

Pengujian ini dilakukan untuk mengetahui model mana yang lebih baik digunakan di antara *fixed effect model* atau *random effect model*. Hipotesis yang digunakan dalam Uji Hausman adalah sebagai berikut:

$H_0$ : *Random effect model* lebih baik

$H_A$ : *Fixed effect model* lebih baik

Dengan kriteria sebagai berikut:

1. Apabila  $(Prob > \chi^2) < \alpha$  maka  $H_0$  ditolak yang artinya *fixed effect model* lebih baik digunakan dari pada *random effect model*.
2. Apabila  $(Prob > \chi^2) \geq \alpha$  maka  $H_0$  tidak dapat ditolak yang artinya *random effect model* lebih baik digunakan dari pada *fixed effect model*.



### 3) Uji Breusch – Pagan Langrangian Multiplier

Pengujian ini dilakukan untuk mengetahui model mana yang lebih baik digunakan di antara *random effect model* atau *pooled least square*. Hipotesis yang digunakan dalam Uji Breusch – Pagan Lagrange Multiplier adalah sebagai berikut:

$H_0$ : *Pooled least square* lebih baik

$H_A$ : *Random effect model* lebih baik

Dengan kriteria sebagai berikut:

1. Apabila  $(Prob > \bar{\chi}^2) < \alpha$  maka  $H_0$  ditolak yang artinya *random effect model* lebih baik digunakan dari pada *pooled least square*.
2. Apabila  $(Prob > \bar{\chi}^2) \geq \alpha$  maka  $H_0$  tidak dapat ditolak yang artinya *pooled least square* lebih baik digunakan dari pada *random effect model*.

## 3.6. Pengujian Masalah dalam Analisis Regresi Linear

Secara teoretis telah diungkapkan bahwa salah satu metode pendugaan parameter dalam model regresi linear adalah *Ordinary Least Square* (OLS). Metode OLS digunakan berlandaskan pada sejumlah asumsi tertentu. Ada beberapa asumsi yang harus dipenuhi, pada prinsipnya model regresi linear yang dibangun sebaiknya tidak boleh menyimpang dari asumsi BLUE (*Best, Linear, Unbiased Estimator*).

### 3.6.1. Uji Multikolinearitas

Multikolinearitas merupakan permasalahan pada suatu model regresi apabila terdapat hubungan linear yang sempurna antar satu variabel independen dengan variabel independen lainnya. Multikolinearitas seringkali disebabkan oleh

pengambilan data yang tidak baik, spesifikasi model dari penelitian, ataupun lebih banyaknya variabel dibandingkan jumlah observasi.

Gejala multikolinearitas ini dapat dideteksi dengan melihat nilai  $R^2$ . Apabila nilai  $R^2$  tinggi, namun variabel independen yang signifikan mempengaruhi variabel dependen hanya sedikit atau tidak ada sama sekali, maka bisa dikatakan terdapat masalah multikolinearitas dalam model. Multikolinearitas dapat dideteksi dengan melihat korelasi antar variabel independen. Apabila korelasi antar variabel lebih dari 0,8 maka dapat dikatakan terjadi masalah multikolinearitas dalam model, begitu juga sebaliknya (Gujarati & Porter, 2009).

### **3.6.2. Uji Heteroskedastisitas**

Heteroskedastisitas terjadi karena varian dari *error terms* tidak lagi konstan. Konsekuensi dari adanya heteroskedastisitas dalam sistem persamaan adalah bahwa penaksiran tidak lagi efisien karena mempunyai varians yang tidak lagi minimum. Namun, masalah heteroskedastisitas dapat diminimalisir dengan menggunakan *Robust variance estimate* (Wooldridge, 2012). Jika pada pemilihan model didapat *random effect model*, maka heteroskedastisitas dapat diminimalisir karena menurut Gujarati & Porter (2009) pada hasil regresi menggunakan *random effect model* merupakan estimasi yang dilakukan dengan *generalized least square* (GLS) yakni transformasi variabel sehingga memenuhi asumsi standar kuadrat terkecil, dimana hasil estimasi dari GLS yakni homoskedastis sehingga pada metode GLS tidak terdapat masalah heteroskedastisitas, di mana persebaran data menjadi konstan atau tidak terdapat outlier pada data. *Modified Wald test* dapat menentukan apakah suatu

model memiliki masalah heteroskedastisitas atau tidak (Baum, 2001). Hipotesis pada pengujian ini adalah:

$H_0$ : Model regresi tidak memiliki masalah heteroskedastisitas

$H_A$ : Model regresi memiliki masalah heteroskedastisitas

dengan kriteria:

1. Jika nilai probabilitas observasi lebih kecil dari tingkat signifikansi, maka model regresi tersebut memiliki masalah heteroskedastisitas karena hipotesis utama yang ditolak
2. Jika nilai probabilitas observasi lebih besar dari tingkat signifikansi, maka model regresi tersebut tidak memiliki masalah heteroskedastisitas karena hipotesis utama yang tidak dapat ditolak.

### 3.6.3. Uji Autokorelasi

Adanya korelasi antara anggota serangkaian observasi yang diurutkan menurut waktu (*time series*) atau ruang (*cross section*) merupakan istilah dari autokorelasi. Konsekuensi apabila terjadi autokorelasi dalam sistem persamaan adalah, pertama, penaksir tidak lagi efisien maka selang keyakinan menjadi lebar secara tak perlu sehingga pengujian arti (signifikasnsi) kurang kuat, kedua, pengujian signifikansi parsial tidak lagi sah, dan jika diterapkan akan memberikan kesimpulan yang menyesatkan secara serius mengenai arti statistik dari koefisien yang ditaksir. Jika pada pemilihan model didapat *random effect model*, maka autokorelasi dapat diminimalisir karena *random effect model* menggunakan estimasi model *generalized least square* sehingga masalah autokorelasi sudah teratasi dengan sendirinya (Wooldridge, 2002). Autokorelasi dapat dideteksi dengan

menggunakan uji Wooldridge (2002) pada data panel dengan hipotesis sebagai berikut:

$H_0$ : Model regresi tidak memiliki masalah autokorelasi

$H_A$ : Model regresi memiliki masalah autokorelasi

Dengan kriteria:

1. Jika nilai probabilitas observasi lebih kecil dari tingkat signifikansi, maka model regresi tersebut memiliki masalah autokorelasi karena hipotesis utama yang ditolak
2. Jika nilai probabilitas observasi lebih besar dari tingkat signifikansi, maka model regresi tersebut tidak memiliki masalah autokorelasi karena hipotesis utama yang tidak dapat ditolak.

### **3.7. Pengujian Statistik**

Pengujian statistik dilakukan untuk memperkuat hasil estimasi sebuah penelitian. Adapun pengujian statistik yang dilakukan meliputi koefisien determinasi ( $R^2$ ), uji signifikansi simultan, dan uji signifikansi parsial.

#### **3.7.1. Koefisien Determinasi ( $R^2$ )**

Ukuran ringkas yang menginformasikan kepada kita seberapa baik sebuah garis regresi sampel sesuai dengan datanya merupakan definisi dari koefisien determinasi ( $R^2$ ) (Gujarati & Porter, 2009). Terdapat dua sifat dari  $R^2$  yaitu:

1. Nilai  $R^2$  tidak pernah negatif
2. Besarnya nilainya berapa pada kisaran 0 hingga 1. Jika nilai  $R^2$  semakin mendekati angka 1 maka ada kesesuaian garisnya tepat atau antara regresi

dengan regresor memiliki hubungan dan sebaliknya ketika semakin menjauh dari angka 1 maka tidak ada hubungan antara regresi dan regresor.

### 3.7.2. Uji Signifikansi Simultan

Penggunaan uji F dilakukan untuk mengetahui tingkat signifikansi atau melihat adanya pengaruh antara variabel independen secara keseluruhan terhadap variabel dependen dalam model pada tingkat signifikansi tertentu, dengan hipotesis sebagai berikut:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$  (semua variabel independen dalam model secara bersama-sama tidak mempengaruhi variabel dependen)

$H_A$ : paling tidak, ada satu  $\beta \neq 0$  (semua variabel independen dalam model secara bersama-sama mempengaruhi variabel dependen)

Dengan statistik uji:

$$F = \frac{R^2/(k - 1)}{(1 - R^2)/(n - k)}$$

Di mana:

$R^2$ : koefisien determinasi

$k$ : jumlah variabel independen

$n$ : jumlah sampel

Dengan kriteria sebagai berikut:

1. Jika  $F_{stat} > F_{table}$  atau  $Prob F < \alpha$  maka  $H_0$  ditolak. Artinya bisa dikatakan bahwa semua variabel independen dalam model mempengaruhi variabel dependen.

2. Jika  $F_{stat} \leq F_{table}$  atau  $Prob F \geq \alpha$  maka  $H_0$  tidak dapat ditolak. Artinya bahwa semua variabel independen dalam model tidak mempengaruhi variabel dependen.

Namun, bila hasil estimasi regresi menggunakan *random effect model*, maka digunakan uji *Wald Chi-Square* untuk melihat apakah semua variabel independen mempengaruhi variabel dependen secara signifikan (StataCorp, 2013), dengan hipotesis yang sama serta kriteria sebagai berikut:

1. Jika  $Prob \chi^2 < \alpha$  maka  $H_0$  ditolak. Artinya bisa dikatakan bahwa variabel independen dalam model secara bersama-sama mempengaruhi variabel dependen
2. Jika  $Prob \chi^2 \geq \alpha$  maka  $H_0$  tidak dapat ditolak. Artinya bahwa semua variabel independen dalam model secara bersama-sama tidak mempengaruhi variabel dependen.

### 3.7.3. Uji Signifikansi Parsial

Penggunaan uji t dilakukan untuk mengetahui tingkat signifikansi atau melihat adanya pengaruh antara variabel independen dengan variabel dependen secara individu atau secara masing-masing variabel independen terhadap variabel dependen dalam model pada tingkat signifikan tertentu, dengan hipotesis sebagai berikut:

$H_0 : \beta_n = 0$  (variabel independen tidak mempengaruhi variabel dependen secara signifikan)

$H_A : \beta_n \neq 0$  (Variabel independen mempengaruhi variabel dependen secara signifikan)

Dengan statistik uji:

$$t = \frac{\hat{\beta}_n - \beta_n}{se(\hat{\beta}_n)}$$

Di mana:

$\hat{\beta}_n$ : koefisien variabel independen ke-  $n$

$\beta_n$ : nilai dari hipotesis nol

$se(\hat{\beta}_n)$ : simpangan baku dari variabel independen ke-  $n$

Dengan kriteria sebagai berikut:

1. Jika  $t_{stat} > t_{table}$  atau  $-t_{stat} < -t_{table}$  ( $Prob\ t < \alpha$ ) maka  $H_0$  ditolak. Artinya variabel independen mempengaruhi variabel dependen secara signifikan.
2. Jika  $-t_{table} \leq t_{stat} \leq t_{table}$  ( $Prob\ t \geq \alpha$ ) maka  $H_0$  tidak dapat ditolak. Artinya variabel independen tidak mempengaruhi variabel dependen secara signifikan.

Namun, bila hasil estimasi regresi menggunakan *Random Effect Model*, digunakan uji z untuk melihat pengaruh masing-masing variabel independen terhadap variabel dependen (StataCorp, 2013), dengan hipotesis yang sama serta kriteria sebagai berikut:

1. Jika  $Prob\ z < \alpha$  maka  $H_0$  ditolak. Artinya variabel independen mempengaruhi variabel dependen secara signifikan.
2. Jika  $Prob\ z \geq \alpha$  maka  $H_0$  tidak dapat ditolak. Artinya variabel independen tidak mempengaruhi variabel dependen secara signifikan.