

## BAB 3

### METODE PENELITIAN

#### 3.1 Pendekatan Penelitian

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan kuantitatif, dengan menggunakan metode regresi logit. Regresi logit merupakan salah satu pendekatan untuk membuat sebuah model probabilitas dari variabel biner (Gujarati, 2012:173). Variabel respon biner merupakan variabel dependen yang mempunyai dua nilai yaitu 0 dan 1 (Gujarati, 2012:172). Pendekatan tersebut bertujuan untuk menguji pengaruh karakteristik rumah tangga (independen) terhadap tingkat probabilitas partisipasi sekolah di Provinsi Jawa Timur (dependen). Karakteristik rumah tangga meliputi pendidikan kepala rumah tangga, jenis kelamin dan pekerjaan kepala rumah tangga. Penelitian ini menganalisis pengaruh karakteristik rumah tangga terhadap probabilitas partisipasi sekolah.

#### 3.2 Identifikasi Variabel

Dalam penelitian ini, variabel penelitian terdiri dari variabel terikat dan variabel bebas. Variabel terikat adalah probabilitas partisipasi sekolah di provinsi Jawa Timur, variabel bebas diantaranya *dummy* pendidikan kepala rumah tangga, *dummy* jenis kelamin, *dummy* pekerjaan.

### 3.3 Definisi Operasional Varibel

Definisi operasional bertujuan mendefinisikan atau memberi penjelasan mengenai variabel-variabel yang telah didefinisikan sehingga diketahui pemaparan yang lebih jelas yang juga disertai oleh skala pengukuran dan dapat dioperasionalkan. Variabel-variabel yang digunakan dalam model analisis ini meliputi:

- a. **Probabilitas partisipasi sekolah anak usia 16-18 tahun** : Probabilitas partisipasi sekolah anak usia 16-18 tahun untuk bersekolah dan tidak bersekolah. Bernilai 1 jika pada saat kelompok usia 16-18 tahun dalam status masih bersekolah dan bernilai 0 jika tidak bersekolah. Semakin nilainya mendekati 1, maka semakin tinggi probabilitas anak untuk bersekolah.
- b. **Karakteristik rumah tangga**, dalam penelitian ini terdiri dari beberapa variabel, yaitu : Pendidikan terakhir kepala rumah tangga, jenis kelamin, pekerjaan kepala rumah tangga, data tersebut diambil dari Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2012 yang berada dalam kolom individu.

#### 1. Pendidikan kepala rumah tangga

Yang dimaksudkan dengan pendidikan kepala rumah tangga dalam penelitian ini adalah tingkat pendidikan terakhir kepala rumah tangga berdasarkan ijazah/STTB tertinggi yang dimiliki. Jika pendidikan terakhir kepala rumah tangga apabila bersekolah hingga SMP, maka pendidikan

kepala rumah tangga yang ditamatkan termasuk kategori pendidikan rendah sedangkan apabila tamatan pendidikan terakhir kepala rumah tangga yang dimiliki SMA atau menengah ke atas hingga perguruan tinggi maka termasuk kategori pendidikan menengah ke atas. Bernilai 1 jika pendidikan kepala rumah tangga menengah ke atas bernilai 0 jika pendidikan kepala rumah tangga rendah.

## **2. Jenis kelamin**

Yang dimaksudkan dengan jenis kelamin anak dalam penelitian ini adalah laki-laki dan perempuan. Variabel ini berbentuk dummy yang bernilai 1 jika kepala rumah tangga berjenis kelamin laki-laki dan 0 jika berjenis kelamin perempuan.

## **3. Pekerjaan**

Dalam penelitian ini yang dimaksudkan dalam pekerjaan adalah pekerjaan kepala rumah tangga sebagai buruh atau bukan buruh. Bernilai 1 jika buruh dan bernilai 0 jika bukan buruh.

### **3.4 Jenis dan Sumber Data**

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, yaitu berupa data *cross section* tahun 2012. Data tersebut berasal dari data Survei Sosial Ekonomi Nasional (Susenas) yang diselenggarakan oleh Badan Pusat Statistika (BPS) Tahun 2012. Data yang digunakan dalam penelitian ini merupakan triwulan pertama yaitu Maret tahun 2012.

Susenas merupakan survei rumah tangga yang dirancang untuk mengumpulkan data sosial ekonomi kependudukan berupa data kor (pokok) dan data modul (rinci). Menurut observasinya susenas kor dibagi menjadi kor individu dan kor rumah tangga. Data yang dikumpulkan antara lain bidang kesehatan/gizi, pendidikan, perumahan/lingkungan, kegiatan sosial ekonomi, konsumsi dan pengeluaran rumah tangga.

Susenas didesain memiliki 3 modul (Modul Konsumsi/Pengeluaran Rumah Tangga, Modul Sosial Budaya dan Pendidikan dan Modul Perumahan dan Kesehatan) dan pengumpulan data setiap modul dilaksanakan setiap 3 tahun sekali. Pada tahun 2011 pengumpulan data kor dan modul konsumsi/pengeluaran rumah tangga dilaksanakan setiap triwulan (Maret, Juni, September dan Desember).

Jumlah sampel setiap triwulan sebanyak 75 ribu rumah tangga yang tersebar di seluruh Provinsi dan Kabupaten/Kota di Indonesia. Data setiap triwulan disajikan baik tingkat nasional maupun provinsi dan dibedakan menurut perkotaan dan pedesaan.

Data dalam penelitian ini diperoleh dari kuisisioner Susenas Maret 2012 modul Konsumsi/Pengeluaran Rumah tangga. Pengeluaran rumah tangga dibedakan menjadi konsumsi makanan dan bukan makanan. Konsumsi makanan terdiri dari jenis sayuran, buah-buahan, bahan minuman dan lainnya. Konsumsi bukan makanan terdiri dari beberapa kelompok yaitu kolom perumahan & fasilitas rumah tangga, kolom barang dan jasa, kolom pakaian, alas kaki dan tutup kepala,

kolom barang-barang tahan lama, kolom pajak dan asuransi, dan kolom keperluan pesta dan upacara.

Data dalam analisis regresi adalah data partisipasi sekolah yang diperoleh dari pertanyaan Blok IV.A nomor tiga (kor) tentang hubungan dengan kepala rumah tangga dan nomor lima (kor) tentang umur. Data jenis kelamin anak diperoleh dari kuisisioner pertanyaan blok IV.A nomor empat (kor). Data pekerjaan orang tua sebagai buruh di sektor industri diperoleh dari Blok V D nomor 24 (kor) tentang orang tua sedang bekerja dan nomor 30 dan 31 (kor) . Data jenis kelamin kepala rumah tangga diperoleh dari pertanyaan Blok IV A nomor 3 (kor) tentang hubungan dengan kepala rumah tangga dan nomor 4 (kor) tentang jenis kelamin. Data pendidikan terakhir kepala rumah tangga diperoleh dari kuisisioner pertanyaan Blok IV A nomor 3 (kor) tentang hubungan dengan kepala rumah tangga dan Blok V C nomor 17 tentang ijazah/STTB tertinggi yang dimiliki.

Sampel individu yang diobservasi adalah tingkat nasional pada tahun 2012 berasal dari susenas modul konsumsi dan susenas kor individu. Jumlah sampel susenas modul konsumsi Blok IV.2 tahun 2012 sebanyak 7.595.616 observasi dan susenas kor individu tahun 2012 sebanyak 1.114.445 observasi. Sampel tersebut kemudian dibatasi pada individu yang hubungan kepala rumah tangga adalah anak, masih bersekolah dan memiliki kepala rumah tangga. Sehingga observasi berkurang menjadi 12.846 observasi.

### **3.5 Prosedur Pengumpulan Data**

Prosedur pengumpulan data pada penelitian ini dilakukan dengan cara mengumpulkan data yang diperlukan dari sumber-sumber yang telah disebutkan.

Untuk analisis regresi, data sampel susenas kor dan modul disatukan (*merge*) dalam satu lembar kerja untuk mengumpulkan variabel yang dibutuhkan. Langkah berikutnya proses penyaringan (*filter*) individu, dimana sampel dengan hubungan dengan kepala rumah tangga adalah anak, usia 16-18 tahun, masih bersekolah atau tidak bersekolah dan anak tersebut dalam rumah tangga. Data jenis kelamin anak, pendidikan terakhir kepala rumah tangga, pekerjaan berada di susenas kor individu.

### **3.6 Teknik Analisis**

Analisis data dalam penelitian ini menggunakan model regresi Logit karena variabel terikat yang digunakan dalam penelitian ini bersifat binari. Model logit digunakan untuk mengestimasi probabilitas terjadinya partisipasi sekolah anak usia 16-18 tahun. Proses analisis atau perhitungan menggunakan *software* (perangkat lunak) komputer yaitu Stata 13.

Terdapat beberapa variabel yang akan di analisis yaitu variable dependen probabilitas partisipasi sekolah Provinsi Jawa Timur, variabel independen dari analisis data tersebut adalah karakteristik rumah tangga diantaranya pendidikan kepala rumah tangga, jenis kelamin, pekerjaan.

#### **3.6.1 Model Regresi Logit**

Metode regresi Logit digunakan untuk mengestimasi probabilitas terjadinya partisipasi sekolah terhadap anak yang memiliki keluarga yang bekerja sebagai buruh dengan alat analisis yang dipakai untuk mengolah data yaitu

program STATA 13. Model logit digunakan untuk penelitian ini dengan sistem matematis yang lebih sederhana.

Persamaan regresi model logit diperoleh dari penurunan persamaan probabilitas dari kategori-kategori yang akan diestimasi. Persamaan probabilitas tersebut adalah (Gujarati:595-596) :

$$P_i = E ( Y = 1 ) \mid X_i = \frac{1}{1+e^{-(\beta_1+\beta_2X_i)}} \dots\dots\dots(3.1)$$

Persamaan tersebut disederhanakan dengan mengasumsikan  $(\beta_1+\beta_2X_i)$  adalah  $Z_i$ , Sehingga menghasilkan persamaan berikut :

$$P_i = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} \dots\dots\dots(3.2)$$

Pada persamaan tersebut dapat terlihat bahwa  $Z_i$  berada dalam kisaran  $-\infty$  hingga  $+\infty$  dan  $P_i$  berada pada kisaran 0 hingga 1 dimana  $P_i$  tidak linier terhadap  $Z_i$ ,  $P_i$  tidak linier terhadap  $X_i$  dan juga terhadap  $\beta$ . Hal ini menimbulkan permasalahan estimasi sehingga prosedur regresi *ordinary least square (OLS)* tidak dapat dilakukan. Solusi dari permasalahan tersebut adalah dengan melinierkan persamaan (2.5) dengan menerapkan logaritma natural pada kategori dengan nilai 0, sehingga didapatkan persamaan berikut :

$$1 - P_i = \frac{1}{1+e^z} \dots\dots\dots(3.3)$$

Persamaan tersebut disubstitusikan dengan persamaan (3.3) menjadi :

$$\frac{P_i}{1-P_i} = \frac{1+e^z}{1+e^{-z}} \dots\dots\dots(3.4)$$

Persamaan ini disebut dengan rasio kecenderungan (*odds ratio*) terjadinya kategori dengan nilai 1. Selanjutnya dengan menerapkan logaritma natural terhadap odds ratio tersebut akan menghasilkan persamaan sebagai berikut :

$$L_i = \ln \left[ \frac{P_i}{1-P_i} \right] = z_i = \beta_1 + \beta_2 X_i \dots \dots \dots (3.5)$$

Dalam persamaan 2.4,  $L_i$  adalah log dari *odds ratio* atau disebut model logit. Model tersebut tersebut tidak hanya linier terhadap  $X$ , namun juga linier terhadap parameter  $\beta$ . Karena  $P_i$  terletak antara 0 dan 1,  $L_i$  antara  $-\infty$  dan  $+\infty$ , maka meskipun  $L_i$  linier dalam  $X$ , tetapi  $P_i$  tidak linier dalam  $X$ . Nilai  $\beta_1$  merupakan *intercept*, yang dapat diartikan sebagai partisipasi sekolah adalah sebesar  $\beta_1$  ketika variabel-variabel lain bernilai nol.  $\beta_2$  adalah perubahan dalam  $L_i$  apabila  $X$  berubah 1. Sehingga model logit untuk penelitian ini dapat ditulis sebagai berikut :

$$L_i = \ln(P_i|1 - P_i) = \beta_1 + \beta_2 \text{ dummy pendidikan kepala rumah tangga}_i + \beta_3 \text{ dummy jenis kelamin}_i + \beta_4 \text{ dummy pekerjaan} + e_i \quad (3.6)$$

Keterangan :

$P_i$  : Probabilitas Partisipasi sekolah

$\beta_i$  : *intercept*

$\beta_2, \beta_3, \beta_4$  : Parameter (koefisien) regresi

*dummy pendidikan kepala rumah tangga* : Pendidikan kepala rumah tangga rendah jika SD-SMP bernilai (0), SMA-Perguruan tinggi menengah keatas (1)

*dummy jenis kelamin* : Jenis kelamin anak perempuan (0), laki-laki (1)



*dummy* pekerjaan : Pekerjaan kepala rumah tangga jika buruh (1) , pekerjaan jika bukan buruh (0)

Pengujian statistik dilakukan untuk mengetahui apakah variabel independen atau variabel bebas mempengaruhi variabel dependen atau variabel terikat secara signifikan.

a. Uji secara bersama-sama model logit dengan Likelihood ratio (LR)

*Likelihood Ratio (LR)* mirip dengan pengujian *F-test* pada model regresi linier. Fungsi dari keduanya sama yaitu digunakan untuk menguji hipotesis nol bahwa semua *slope* koefisien secara bersama-sama dengan nol. *LR* mengikuti distribusi  $x^2$  dengan *df* sama dengan jumlah variabel penjelas (tidak termasuk intersept) (Gujarati, 2003:606). Hipotesisnya adalah :

$$H_0 = \beta_1 = \beta_2 \dots = \beta_n = 0$$

Berarti tidak ada pengaruh secara bersama-sama dari variabel bebas (independen) terhadap variabel terikat (dependen) partisipasi sekolah.  $H_1$ =kemungkinan tidak ada salah satu parameter yang tidak sama dengan nol. Berarti terdapat pengaruh secara bersama-sama dari variabel bebas (independen) terhadap variabel terikat (dependen) partisipasi sekolah.

Kriteria pengujian nilai *LR* adalah jika  $x^2_{hitung} > x^2_{tabel}$ , maka  $H_0$  ditolak. Selain itu untuk melakukan pengujian hipotesis dapat juga dilakukan dengan konsep *p-value*. Konsep ini membandingkan tingkat keyakinan ( $\alpha$ ) sebesar 5% dengan nilai *p-value*. Jika nilai *p-value* kurang dari  $\alpha$  maka  $H_0$  ditolak

(Gujarati,2003:137). Jika  $H_0$  ditolak maka ada pengaruh secara bersama-sama dari variabel independen terhadap variabel dependen, dan sebaliknya.

b. Uji Parsial signifikansi koefisien dengan *z-statistic*

Di dalam model logit akan menggunakan metode estimasi yang berbeda seperti yang digunakan dalam metode regresi linear. Metode estimasi yang digunakan dalam model ini adalah metode *maximum likelihood*, umumnya adalah metode yang digunakan oleh sampel besar, tujuannya untuk meminimumkan error.

Penggunaan *maximum likelihood* diharapkan dapat mendekati nilai variabel yang diestimasi dengan nilai variabel yang sebenarnya. Metode *maximum likelihood* membuat hasil estimasi dari nilai standart error menjadi asymptotic. Hal inilah yang menyebabkan proses signifikansi secara statistik pada koefisien tidak menggunakan *t-statistic*, melainkan *z-statistic* yang terdapat pada tabel distribusi normal dan penting untuk di ingat apabila jumlah sampel yang ada cukup besar, distribusi t menjadi konvergen ke distribusi normal, dan ketika terdistribusi normal apabila nilai rata-ratanya nol dan nilai variannya satu.

Hipotesis dari uji z adalah sebagai berikut :

$$H_0 : \beta_i = 0, \text{dimana } i = 0,1$$

Artinya, tidak ada pengaruh variabel independen terhadap variabel dependen secara parsial.

$$H_1 : \beta_i \neq 0 \text{ dimana } i = 0,1$$

Artinya, ada pengaruh variabel independen terhadap variabel dependen secara parsial

c. *Goodness of fit*

Uji *Goodness of fit* dilakukan dengan melihat R-square untuk melihat seberapa baik atau suatu model dapat menjelaskan hubungan antara variabel dependen dengan variabel independennya. Pada model regresi logistik, pengujian *Goodness of fit* dapat dilihat dengan nilai *Pseudo R<sup>2</sup>* yang memiliki arti sebagai *R-square*.

Nilai *Pseudo R<sup>2</sup>* berkisar antara 0-1 jika nilai mendekati 1 maka terjadi hubungan yang kuat antara variabel independen dengan variabel dependen. Ketika nilai mendekati 0 berarti tidak ada hubungan antara variabel independen dan variabel dependen (Gujarati,2003:84). Hal yang harus diperhatikan didalam suatu model adalah signifikansi antara variabel independen dan arah koefisien variabel tersebut. Sedangkan besaran *pseudo R<sup>2</sup>* tidak diutamakan. Oleh karena itu dalam penelitian ini menggunakan z-stat sebagai parameter utama kelayakan model, meskipun pengujian *pseudo R<sup>2</sup>* akan tetap dilakukan. Pilihan pengujian *goodness of fit* yang digunakan dalam penelitian ini adalah uji *goodness of fit* Hosmer-Lemeshow.

Pada pengujian *goodness of fit* Hosmer-Lemeshow menyajikan pearson chi-square *goodness of fit test* untuk mengetahui ketepatan model (*the fitted model*) terhadap nilai observasinya. Pengujian ini diukur dengan nilai chi-square. Adapun hipotesisnya sebagai berikut :

$H_0$  = tidak menolak model

$H_1$  = menolak model

Dalam pengujian Hosmer-Lemeshow, menentukan  $H_0$  ditolak atau  $H_0$  tidak ditolak, dapat dilakukan dengan membandingkan  $\text{prob} > \text{chi-square}$  dari nilai  $\text{pearson chi-square}$  dengan tingkat keyakinan ( $\alpha$ ) sebesar 5%. Jika nilai  $\text{prob} > \text{chi-square}$  kurang dari  $\alpha$  maka  $H_0$  ditolak dan jika nilai  $\text{prob} > \text{chi-square}$  lebih dari  $\alpha$  maka  $H_0$  diterima berarti tidak menolak model, sehingga model tersebut tepat atau cocok untuk digunakan (*the fitted model*), dimana model dapat memprediksi nilai observasinya.

d. Interpretasi Logit

Menginterpretasikan model logit perlu dilakukan. Apabila nilai logit suatu variabel bertanda positif, ketika nilai variabel meningkat, nilai log dari *odds* (kecenderungan) terjadinya partisipasi sekolah (masih sekolah) ( $Y=0$ ) meningkat; Sebaliknya ketika tanda dari nilai logit tersebut adalah negatif, maka nilai log dari *odds* (kecenderungan) terjadinya partisipasi sekolah (tidak bersekolah) ( $Y=1$ ) menurun seiring dengan peningkatan nilai  $X$  (Gujarati,2003:600).

e. Interpretasi Odds ratio

Untuk menginterpretasikan nilai dari koefisien logit, koefisien hasil estimasi model logit harus ditransformasikan ke dalam antilogaritma natural sehingga di dapatkan *odds ratio* yang merupakan rasio antara dua peluang(kecenderungan).

*Odds ratio* di interpretasikan sebagai nilai yang menunjukkan pengaruh variabel dependen (partisipasi sekolah) sebesar  $e^\beta$  dikali.