

BAB III

METODE PENELITIAN

A. Objek Penelitian

Pada penelitian ini penulis menggunakan data hasil survey yang diperoleh dari Indonesian Family Life Survey (IFLS). IFLS menyediakan informasi secara ekstensif mengenai bidang sosio ekonomi, kesehatan dan di bidang rumah tangga maupun individu. IFLS merupakan survei yang paling komprehensif yang pernah dilakukan di Indonesia (strauss, Witoelar & Sikoki, 2016). Survei ini diadakan atas kerja sama antara organisasi penelitian Amerika Serikat RAND, lembaga Demografi Universitas Indonesia, Pusat Studi Kependudukan dan Kebijakan Gadjah Mada.

Survei ini yang dilakukan di 24 provinsi di Indonesia berupa data *cross section* mencakup provinsi, Sumatera Barat, Jambi, Riau, Kepulauan Riau, Bangka Belitung, Sumatera Selatan, Sumatera Utara, Lampung, seluruh provinsi di Jawa, Bali, NTB, seluruh provinsi di Kalimantan, Sulawesi Selatan dan Sulawesi Barat. Subjek penelitian ini difokuskan pada individu dalam rumah tangga yang berusia 15 tahun atau lebih yang merupakan individu dalam penelitian *Indonesian Family Life Survey* (IFLS). Data IFLS yang digunakan pada penelitian ini adalah IFLS-5 (2014-2015).

B. Jenis Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari lembaga survei yaitu Indonesia Family Life Survey (IFLS). Data yang digunakan dalam penelitian ini adalah data *cross section* yang meliputi area seluas 24 provinsi di Indonesia. Dalam penelitian ini menggunakan data IFLS gelombang 5 pada tahun 2014-2015 di mana berjumlah 14.335 responden.

Kelebihan IFLS-5 dibanding survei sebelumnya yaitu IFLS-5 telah menggunakan sistem Computer-Assisted Personal Interview (CAPI) dan tidak lagi menggunakan kuisisioner kertas. Program CAPI telah dipersiapkan dan diuji coba selama kurang lebih 18 bulan. Selain itu, pengambilan data pada IFLS-5 juga telah menggunakan alat perekam suara sehingga kualitas data dapat terkontrol dengan baik (Strauss, 2004).

C. Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini menggunakan teknik dokumentasi yaitu penelitian menggunakan bahan yang didapat dari Indonesia Family Life Survey (IFLS) yang terkait dengan topik penelitian dengan melakukan pencatatan langsung berupa data *cross section* pada data IFLS-5 tahun 2014-2015.

D. Definisi Operasional Variabel Penelitian

Penelitian ini menggunakan upah (*wage*) tenaga kerja sebagai variabel dependent. Dan terdapat dua variabel independen meliputi capaian pendidikan dan pengalaman kerja. Serta tiga variabel kontrol yaitu usia, jam kerja, dan status pernikahan. Definisi dari masing-masing variabel menjelaskan sebagai berikut:

1. Variabel Dependen

Variabel dependen atau variabel terikat yaitu variabel yang dipengaruhi karena adanya variabel independe. Pada penelitian ini variabel dependen yang digunakan yaitu upah (*wage*) tenaga kerja dari Indonesian Family Life Survey tahun 2014-2015. Upah (*wage*) adalah penerimaan yang diperoleh tenaga kerja berupa uang dalam waktu satu bulan. Nilai upah tersebut dibuat menjadi logaritma.

2. Variabel Independen

Variabel independen atau variabel bebas yaitu variabel yang mempengaruhi atau yang menjadi sebab perubahan variabel dependen (Basuki dan Yuliadi, 2015). Dalam penelitian ini variabel independen merupakan hal yang mempengaruhi tingkat upah tenaga kerja. Adapun yang menjadi variabel independen dalam penelitian ini yaitu capaian pendidikan, dan pengalaman kerja.

- a. Capaian pendidikan yaitu tingkatan untuk menunjukkan jumlah tahun atau lamanya pendidikan yang ditempuh oleh individu dalam rumah tangga IFLS ke 5 tahun 2014-2015.

- b. Pengalaman kerja yaitu jumlah usia tenaga kerja yang dikurangi dengan lamanya tahun masa pendidikan yang di tempuh dikurangi usia pertama kali masuk sekolah dalam rumah tangga IFLS ke 5 tahun 2014-2015.

3. Variabel Kontrol

Variabel kontrol yaitu variabel yang mengatur hubungan antara variabel independen dan dependen, agar faktor-faktor dari luar yang tidak dibutuhkan tidak dapat mempengaruhi hubungan antara variabel independen dan variabel dependen atau agar tidak terjadi bias. Dalam penelitian ini variabel kontrol terdiri dari :

- a. Umur yaitu usia individu dalam rumah tangga IFLS ke 5 tahun 2014-2015, yang diukur menurut ulang tahun terakhir. Usia responden dalam penelitian ini minimal 15 tahun keatas.
- b. Jam kerja yaitu waktu yang digunakan oleh individu untuk bekerja selama seminggu dalam rumah tangga IFLS ke 5 tahun 2014-2015.
- c. Status perkawinan yaitu status perkawinan individu dalam rumah tangga IFLS ke 5 tahun 2014-2015 saat berlangsungnya survei. Adapun *dummy* variabel ditentukan dengan 1= menikah; 0=belum/tidak menikah.

E. Uji Kualitas Data

Menurut Basuki dan Yuliadi (2015) uji asumsi klasik yang digunakan dalam regresi linier dengan pendekatan Ordinary Least Squared (OLS) meliputi uji Linearitas, Autokorelasi, Heteroskedastisitas, Multikolenearitas dan Normalitas. Walaupun demikian, tidak semua uji asumsi klasik harus

dilakukan pada setiap model regresi linier dengan pendekatan OLS. Asumsi yang harus terpenuhi dalam analisis model regresi ini adalah normalitas, heterokedastisitas, dan tidak terdapat multikolinearitas.

1. Uji Normalitas

Uji normalitas yaitu uji yang berfungsi untuk menguji distribusi residual yang telah distandarisasi pada model regresi berdistribusi normal atau tidak. Residual dikatakan berdistribusi normal jika distribusi residual terstandarisasi tersebut sebagian besar mendekati nilai rata-ratanya. Dalam melihat suatu normalitas data dapat menggunakan Uji Shapiro Wilk. Untuk menentukan hasil uji normalitas dapat dilakukan dengan melihat nilai signifikansinya, jika $p\text{-value} > 0,05$ maka data residual pada penelitian tersebut terdistribusi secara normal. Sedangkan jika $p\text{-value} < 0,05$ maka data tidak terdistribusi secara normal. Apabila uji normalitas menunjukkan bahwa yang digunakan dalam penelitian ini cenderung tidak normal maka dapat digunakan asumsi *Central Limit Theorem* yaitu jika jumlah observasi cukup besar ($n > 30$), maka asumsi normalitas dapat diabaikan (Gujarati, 2009).

2. Uji Multikolinearitas

Uji multikolinieritas yaitu uji yang bertujuan untuk menguji keberadaan korelasi antara variabel independen dan model regresi. Suatu model regresi yang bagus seharusnya tidak terjadinya korelasi di antara variabel independennya (Ghozali, 2007). Cara mendeteksi adanya multikolinearitas adalah dengan mengamati nilai Variance Inflation Factor (VIF) dan *tolerance*. Batas VIF adalah 10 dan nilai dari *tolerance* adalah 0,1.

Jika nilai VIF lebih besar dari 10 dan nilai *tolerance* kurang dari 0,1 maka terjadi multikolinieritas. Bila ada variabel independen yang terkena multikolinieritas maka variabel tersebut harus dikeluarkan dari model penelitian.

Terdapat beberapa indikator yang bisa dipergunakan untuk melihat ada tidaknya multikolinieritas pada persamaan suatu regresi (Gujarati, 2007) yaitu:

- Nilai R^2 yang dihasilkan suatu estimasi model yang sangat tinggi, tetapi variable independent banyak yang tidak signifikan mempengaruhi variable dependen.
- Menganalisis matrik korelasi yang cukup tinggi (umumnya di atas 9,0) maka hal ini merupakan indikasi adanya multikolinieritas.
- Melihat nilai *tolerance* dan nilai *variance inflation factor* (VIF). Suatu regresi bebas dari masalah multikolinieritas apabila nilai *tolerance* kurang dari 0,1 dan nilai VIF lebih dari 1,0.

3. Uji Heterokedastisitas

Menurut Gujarati (2005), uji heteroskedastisitas yaitu uji yang dilakukan bertujuan untuk melihat apakah adanya peningkatan dari *variance* dalam variabel dependen. Jika *variance* dari residual satu pengamatan ke pengamatan lain tetap, maka disebut homoskedastisitas dan jika berbeda disebut heteroskedastisitas. Untuk menentukan apakah terdapat heteroskedastisitas dalam penelitian ini adalah dengan melihat grafik *scatter*

plot, apabila hasil data menyebar, yaitu di atas dan di bawah nilai nol maka model regresi layak pakai karena bebas heteroskedastisitas.

4. Teori Regresi *Robust*

Pada tahun 1972 Andrews memperkenalkan regresi *robust*. Regresi *robust* adalah suatu metode regresi yang digunakan apabila distribusi dari *error* menunjukkan tidak normal atau terdapat beberapa *outlier* yang berpengaruh terhadap model (Olive, 2005). Apabila data memiliki distribusi yang menunjukkan tidak normal dan mengandung *outlier* sehingga *outlier* dapat mempengaruhi hasil dari estimasi kuadrat terkecil. Regresi *robust* adalah alat untuk menganalisa data mengandung *outlier* dan menunjukkan hasil yang resisten dimana terdapat *outlier* (Turkan & dkk, 2012). Efisiensi dan *breakdown point* bertujuan untuk menjelaskan tingkat ke-*robust*-an (kekekaran) dari hasil estimasi *robust*. Efisiensi yang bertujuan untuk menunjukkan seberapa bagus hasil suatu estimasi *robust* sebanding dengan metode kuadrat terkecil tanpa *outlier*. Apabila makin tinggi efisiensi dan *breakdown point* dari hasil suatu estimator maka semakin resisten suatu model dalam menganalisa data yang terdapat *outlier*.

Metode regresi *robust* menurut (Gujarati, 2005) yaitu suatu metode untuk mengestimasi koefisien regresi yang tidak peka terhadap penyimpangan asumsi yang mendasarinya. *M-estimator* merupakan metode regresi *robust* yang sering digunakan. *M-estimator* dipandang dengan baik untuk mengestimasi parameter yang disebabkan oleh pencilan (*outlier*).

Beberapa masalah yang dapat di tangani menggunakan teknik regresi robust, menurut Chen (2002), yaitu :

1. Masalah *outlier* yang terdapat pada variabel dependen.
2. Masalah *outlier* yang terdapat pada variabel independen.
3. Masalah *outlier* yang terdapat pada keduanya, yaitu variabel dependen dan variabel independen.

5. Model Analisis Regresi Linier Berganda

Regresi linier berganda adalah analisis yang bertujuan untuk mengukur besarnya pengaruh variabel Independent terhadap variabel dependent dan memperkirakan variabel dependent menggunakan variabel independent. Gujarati (2009), mengatakan analisis regresi sebagai kajian terhadap hubungan satu variabel yang disebut sebagai variabel yang diterangkan (*the explained variable*) dengan satu atau dua variabel yang menerangkan (*the explanatory*). Jika variabel independent lebih dari satu, maka analisis regresi disebut regresi linier berganda dikarenakan pengaruh beberapa variabel independent akan dikenakan pada variabel dependent. Penelitian ini mengadopsi persamaan pendapatan Mincer yang digunakan dalam penelitian ini yaitu:

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 pengalaman_i + \beta_3 Jam\ Kerja_i + \beta_4 Status\ Pernikahan_i + \beta_5 Usia_i + U_i$$

Keterangan:

Wage_i = Upah individu i

Educ_i = Tahun sekolah individu i

Umur _i	= Umur individu i
Jam Kerja _i	= Jam kerja individu per minggu i
Pengalaman _i	= Pengalaman kerja individu i
Status Pernikahan _i	= Status pernikahan individu i
U _i	= Error term

Pada penelitian ini tahun sekolah diganti menggunakan capaian pendidikan, kemudian menggunakan variabel independen lain seperti pengalaman kerja, usia, jam kerja, dan status pernikahan sehingga model persamaannya sebagai berikut:

$$\logwage_i = \beta_0 + \beta_1 educ_i + \beta_2 pengalaman_i + \beta_3 Jam\ Kerja_i + \beta_4 Status\ Pernikahan_i + \beta_5 Usia_i + U_i$$

Keterangan:

logWage	= log upah
Educ	= Pendidikan dalam 6 jenjang pendidikan
Usia	= Usia dalam satuan tahun
Jam Kerja	= Jam kerja dalam satuan jam per minggu
Pengalaman	= Pengalaman kerja dalam satuan tahun
Status Pernikahan	= Status pernikahan (Menikah=1, Belum Menikah=0)
U _i	= error term
β ₁ , β ₂ , β ₃ , β ₄ , β ₅	= Koefisien regresi