

ANALISIS REGRESI LINIER BERGANDA PADA DATA SURVEY UNTUK PEMODELAN TOTAL PENGELUARAN DI JAWA TENGAH, INDONESIA

Alan Prahutama^{1*} dan Rita Rahmawati²

^{1,2} Departemen Statistika, Fakultas Sains dan Matematika Universitas Diponegoro

*e-mail: alanprahutama@lecturer.undip.ac.id

DOI: 10.14710/j.gauss.13.2.394-404

Article Info:

Received: 2024-10-18

Accepted: 2024-12-06

Available Online: 2024-10-07

Keywords:

Complex survey design data; regression analysis; total expenditure

Abstract: Survey data is collected using a sampling design to capture Population phenomena. One of the sampling techniques used is complex design. Complex design is a sampling technique other than Simple Random Sampling (SRS). Sampling weight is used for the estimate from the complex design to result in unbiased estimation. Regression analysis using survey data needs to consider complex design. This study models total expenditure in Central Java province using March 2021 Susenas data, with a sample of 29,870 households, 3,005 primary sampling units, and 65 strata. The best model produced from this study is 25.51%. Based on the regression model, the characteristics of households with higher expenditure on categorical independent variables include households living in urban areas, heads of households are male, married, graduates/postgraduates, working in the service sector, and households not receiving government assistance. Meanwhile, the positive coefficient of continuous variables include the age of the head of the household, hours worked during one week, and the number of households living in one house.

1. LATAR BELAKANG

Analisis regresi berganda merupakan salah satu metode statistika yang banyak digunakan untuk pemodelan variable dependent terhadap beberapa variable independent. Estimasi pemodelan dalam regresi menggunakan *ordinary least square* (OLS) yang mengasumsikan bahwa pengamatan merupakan variable random yang bersifat *iid* (independent dan identical distributed) (Draper & Smith, 1998). Pengamatan atau observasi data yang bersifat *iid* berasal dari pengamatan sampel acak atau simple random sampling (SRS). Teknik pengambilan sampling selain SRS (misal, clustering, strata, dan multistage) merupakan teknik pengambilan sample yang tidak memenuhi *iid*, dikarenakan terdapat intra-cross correlation dalam clustering, dan distribusi yang tidak identic (Chambers & Skinner, 2003). Teknik pengambilan sample selain SRS dalam survey dikenal dengan complex survey design. Teknik ini dilakukan karena keterbatasan sumber daya, dan umumnya diterapkan pada survey lingkup nasional atau cakupa wilayah yang luas. Untuk menghasilkan estimasi yang tak bias, estimasi pada complex design memperhatikan survey design dan survey weight (Hahs-Vaughn et al., 2011; Lumley, 2010). Estimasi variansi pada complex design menghasilkan nilai yang lebih besar dibandingkan SRS. Hal ini akan berdampak pada analisis data survey (misal, non-SRS) tanpa memperhatikan complex design akan menyebabkan bias (Heeringa et al., 2010). Pada pengujian hipotesis nilai variansi yang lebih besar memiliki kecenderungan untuk gagal tolak H_0 . Oleh karena itu pada pemodelan linear regression untuk data survey yang mengandung complex design perlu mempertimbangkan survey weight agar menghasilkan estimasi yang tak bias.

Salah satu survey yang menggunakan complex survey design adalah Susenas (Survey Sosial Ekonomi Nasional). Data Susenas merupakan survey yang dilaksanakan di Indonesia untuk menangkap fenomena sosial dan ekonomi di masyarakat yang dilaksanakan dua kali

dalam setahun. Data Susenas memberikan informasi bagi pemerintah terkait kemiskinan, kondisi rumah tangga, keluarga berencana, teknologi dan konsumsi atau pengeluaran. Total pengeluaran merupakan salah satu komponen yang ada dalam Susenas, dan dipergunakan untuk menghitung kemiskinan. Estimasi model total pengeluaran merupakan hal yang penting karena dapat diketahui karakteristik total pengeluaran dari berbagai kelompok atau faktor. Beberapa penelitian telah dilakukan untuk memodelkan pengeluaran diantaranya Sekhampu (2012) melakukan pemodelan pengeluaran makanan menggunakan beberapa faktor demografi kepala rumah tangga diantaranya umur, jenis kelamin, status pernikahan, and status pekerjaan. Sekhampu (2012) juga melakukan pemodelan menggunakan analisis regresi, tetapi tidak melibatkan survey design dan survey weight. Venn et al (2018) juga melakukan penelitian mengenai total pengeluaran makanan dengan menambahkan variabel tingkat pendidikan kepala rumah tangga sebagai sosial-ekonomi faktor. Smith et al (2013) melakukan penelitian tambahan uang dari pemerintah terhadap pengeluaran makanan. Penelitian-penelitian tersebut hanya berfokus pada pengeluaran makanan saja, tidak berfokus pada pengeluaran non-makanan. Sementara beberapa penelitian antara lain Jacobson et al.(2010); Utami & Ayu (2017); (Ali et al. (2018); Abubakar et al.(2021)telah melakukan pemodelan pengeluaran makanan dan non-makanan atau total pengeluaran menggunakan data survey, tetapi tidak melibatkan complex survey design. Oleh karena itu, pada penelitian ini memodelkan total pengeluaran, yang merupakan gabungan dari pengeluaran makanan dan non-makanan, menggunakan regresi linier berganda dengan complex design berdasarkan faktor-faktor sosial ekonomi rumah tangga.

2. LITERATURE REVIEW

Pemodelan regresi linier berganda dapat dinotasikan sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (1)$$

dimana y merupakan variable dependent, x merupakan variable independent, β merupakan parameter model regresi, dan ε adalah residual dari model. Estimasi dalam model regresi menggunakan *ordinary least square* (OLS), persamaan (1) jika ditulis dalam bentuk matrix menjadi $\mathbf{y} = \mathbf{BX} + \boldsymbol{\varepsilon}$, estimasi parameter model regresi adalah

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2)$$

Asumsi dalam model regresi meliputi (1) Estimasi model regresi merupakan linier dalam parameter; (2) Variansi dari residual adalah konstan $Var(\varepsilon_i | \mathbf{x}_i) = \sigma^2$ (Homokedastisitas); (3) Residual berdistribusi normal $\varepsilon_i \sim N(0, \sigma^2)$; (4) Residual bersifat independent $Cov(\varepsilon_i, \varepsilon_j | \mathbf{x}_i, \mathbf{x}_j) = 0$ for $i \neq j$.

Pemodelan regresi linier untuk complex survey design meliputi estimasi pada parameter dan standard error atau variansi dari parameter. Estimasi standard eror menjadi penting, karena melibatkan pengujian hipotesis didalam model regresi. Perbedaan teknik SRS dengan complex design salah satunya terhadap estimasi variansi, sehingga nilai standard error pada regresi linier dengan complex design, akan berbeda dengan standard regresi linier. Complex design melibatkan survey weight, sehingga estimasi parameter model regresi linier berganda untuk complex survey design menggunakan *weighted least square* (WLS), dimana pembobotan yang digunakan merupakan *sampling weight*. Estimasi

parameter model regresi linier dengan complex design dapat dituliskan sebagai berikut (Heeringa et al., 2010):

$$\hat{\beta}_{weight} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (3)$$

Persamaan (3) diperoleh dari estimasi WLS untuk estimasi model regresi dengan populasi bersifat terbatas, dimana estimasi parameter dilakukan untuk meminimumkan fungsi objektif sebagai berikut:

$$f(\beta) = \sum_{i=1}^N (y_i - \mathbf{x}_i \beta)^2 \quad (4)$$

Fungsi objektif $f(\beta)$ dianggap sebagai populasi terbatas dari *sum square error* (SSE), maka SSE yang bersifat weighted, dapat dianggap menggunakan sampling weight sebagai pembobotan. Oleh karena itu nilai SSE dari weighted untuk complex survey design dapat dituliskan sebagai berikut:

$$SSE_{weight} = \sum_{h=1}^H \sum_{\alpha}^{a_h} \sum_{i=1}^{n_{h\alpha}} w_{h\alpha i} (y_{h\alpha i} - \mathbf{x}_{h\alpha i} \hat{\beta})^2 \quad (5)$$

dimana, H adalah banyaknya, a_h adalah banyaknya cluster pada strata ke- h , and n adalah banyaknya sampel. Jadi nilai estimasi pada persamaan (3) merupakan hasil dari estimasi parameter yang mempertimbangkan survey design dan survey weight sebagai complex survey design.

Estimasi variansi atau standard error pada complex survey design menggunakan pendekatan *Taylor Series Linearization* (TSL). Estimasi parameter model $\hat{\beta}$ dapat dinotasikan sebagai perbandingan dua buah total yang dapat dituliskan sebagai berikut :

$$\hat{\beta} = \frac{\sum_h \sum_{\alpha} \sum_{i=1}^{n_{h\alpha}} w_{h\alpha i} y_{h\alpha i} x_{h\alpha i}}{\sum_h \sum_{\alpha} \sum_{i=1}^{n_{h\alpha}} w_{h\alpha i} x_{h\alpha i}^2} = \frac{t_{xy}}{t_{x^2}} \quad (6)$$

Menggunakan pendekatan metode TSL untuk melakukan estimasi variansi, sehingga estimasinya didapatkan sebagai berikut:

$$\text{var}(\hat{\beta})_{weight} \cong \frac{\text{var}(t_{xy}) + \beta^2 \text{var}(t_{x^2}) - 2\beta \text{cov}(t_{xy}, t_{x^2})}{(t_{x^2})^2} \quad (7)$$

Selain estimasi variansi, estimasi matriks *variance-covariance* juga bisa didapatkan sehingga dalam pengujian hipotesis untuk nilai *t-statistics* dan *wald chi-square* atau *wald F-statistics* dapat dihitung. Perhitungan nilai *F-wald statistics* untuk pengujian secara serentak, dan *t-statistics* untuk pengujian partial, menggunakan formula yang sama dengan standard regresi linier. Untuk nilai statistic uji serentak dituliskan pada persamaan (8), dan uji partial dituliskan pada persamaan (9).

$$t_{complex} = \frac{\hat{\beta}_{weight}}{\sqrt{\text{var}(\hat{\beta})_{weight}}} \quad (8)$$

$$F_{complex} = \frac{SSR_{weight} / k}{SSE_{weight} / (n - k - 1)} \quad (9)$$

SSR merupakan total kuadrat regresi dari estimasi model menggunakan complex survey design, dan k merupakan banyaknya variabel independent.

Model evaluasi dalam regresi linier dengan complex survey design, sama dengan standard regresi linier tetapi yang membedakan hanyalah menggunakan sampling weight dalam proses estimasi parameter dan variansi. Salah satu evaluasi model dalam regresi linier adalah menggunakan nilai R^2 . Nilai R^2 pada complex survey design, dapat dihitung menggunakan persamaan (10)

$$R_{weight}^2 = 1 - \frac{SSE_{weight}}{SST_{weight}} \quad (10)$$

Pengujian asumsi normalitas residual pada model regresi linier untuk complex survey design dapat menggunakan pengujian normalitas pada standard regresi linier. Heeringa et al. (2010), merekomendasikan untuk menggunakan standard residual untuk evaluasi pengujian normalitas. Sementara, untuk pengujian homogenitas bisa dilakukan dengan membuat scatter plot antara residual dengan *predicted value* \hat{y} . Jika sebaran plot tidak konstan, maka mengindikasikan terdapat heterogeneity of residual.

3. METODOLOGI PENELITIAN

Penelitian ini menggunakan data sekunder dari BPS, yaitu data Susenas Maret 2021 pada provinsi Jawa Tengah. Sampling design pada data Susenas menggunakan metode *stratified-two stage sampling*. Sampling unit pada Susenas adalah 40% dari sensus blok pada sensus penduduk 2020. Pada tahap pertama adalah memilih *primary sampling units* (PSU) adalah memilih sensus blok secara sistematik dengan *probability proportional to size* (PPS). Tahap kedua adalah memilih *secondary sampling units* (SSUs), yaitu memilih 10 rumah tangga dari sensus blok yang terpilih.

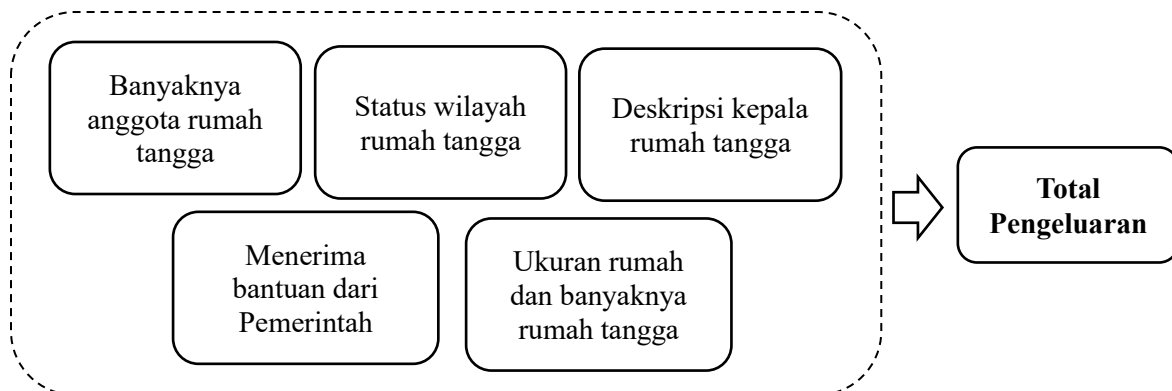
Sampling weight (W) dalam Susenas dihitung dari (BPS, 2021):

$$F = \frac{M_{khi} n_{kh}}{M_{kh}} \times \frac{n_{kh}}{n_{kh}} \times \frac{\bar{m}}{M_{khi}^{up}} = \frac{M_{khi} n_{kh} \bar{m}}{M_{kh} M_{khi}^{up}} \quad (11)$$

$$W = \frac{1}{F} \quad (12)$$

Dimana k merupakan kabupaten, h adalah strata, and i merupakan sensus blok. n adalah banyaknya sample dalam sensus blok. M adalah banyaknya rumah tangga, M^{up} adalah banyaknya rumah tangga yang sudah mengalami pembaruan selama proses survey, dan \bar{m} adalah 10 rumah tangga yang terpilih.

Pada penelitian ini, jumlah observasi sebanyak 29,870 rumah tangga, banyaknya PSU adalah 3,005 dan banyaknya strata adalah 65. Kerangka konseptual penelitian ini digambarkan pada Gambar 1. Analisis yang dilakukan melibatkan PSU, strata, dan sampling weight sebagai bentuk dari complex design yang digunakan untuk analisis.



Gambar 1. Kerangka Konseptual Penelitian

Gambar 1. menjelaskan bahwa total pengeluaran dipengaruhi oleh beberapa dimensi dan dapat diasumsikan sebagai berikut:

1. Semakin banyak anggota rumah tangga, maka total pengeluaran semakin besar. Selain itu perekonomian di pedesaan dan di perkotaan menunjukkan perbedaan (Sekhampu, 2012).
2. Kepala rumah tangga mempunyai peran penting di perekonomian rumah tangga. Salah satunya terkait pendapatan rumah tangga, dimana pendapatan mempengaruhi pengeluaran rumah tangga (García & Grande, 2010).
3. Rumah tangga yang menerima bantuan pemerintah, mengindikasikan bahwa rumah tangga tersebut miskin. Rumah tangga miskin merupakan rumah tangga dengan pengeluaran per kapita per bulan kurang dari garis kemiskinan (Smith et al., 2013).
4. Karakteristik rumah tangga diasumsikan mempengaruhi pengeluaran rumah tangga. Karakteristik rumah tangga dalam penelitian ini menggunakan ukuran rumah (Jacobson et al., 2010) dan jumlah keluarga yang tinggal dalam satu rumah.

Berdasarkan conceptual frame tersebut, pada penelitian ini dapat diidentifikasi beberapa variabel yang dapat memodelkan total pengeluaran rumah tangga. Variabel dependent adalah total pengeluaran rumah tangga per bulan (satuan Rupiah). Variabel independent antara lain status wilayah rumah tinggal X_1 (1= kota; 2=desa), Banyaknya anggota rumah tangga X_2 , status kepala rumah tangga X_3 (1=tidak menikah; 2=menikah; 3=ceraai hidup; 4=ceraai mati); jenis kelamin kepala rumah tangga X_4 (1=laki-laki, 2=perempuan); umur kepala rumah tangga X_5 , tingkat pendidikan kepala rumah tangga X_6 (0=tidak pernah bersekolah; 1=SD; 3=SMP; 4=SMA; 5=Diploma; 6=Sarjana; 7=Pasca Sarjana), lapangan pekerjaan kepala rumah tangga X_7 (0=tidak pernah bekerja; 1=pertanian; 2=industri; 3=jasa), lama jam bekerja dalam satu minggu X_8 , banyaknya rumah tangga yang tinggal dalam satu rumah X_9 , ukuran rumah X_{10} , Rumah tangga menerima bantuan Program Keluarga Harapan (PKH) X_{11} . Tahapan pemodelan didalam analisis regresi dengan complex survey design meliputi (1) Model specification; (2) Model Estimation; (3) Model Evaluation; (4) Model interpretation/ inference.

4. HASIL DAN PEMBAHASAN

Estimasi design-based merupakan estimasi parameter (misal rata-rata, proporsi, total, dan variansi) yang didasarkan pada design pengambil sampel yang telah ditentukan (Lohr, 2010). Dengan mempertimbangkan complex design dari teknik pengambilan sampel dalam Susenas, yaitu PSU dan strata, estimasi design-based untuk variable yang bersifat kontinue

disajikan dalam Tabel 1. Pada variabel kontinue dihitung estimasi rata-rata, standard error (SE), dan interval kepercayaan (CI) 95%.

Tabel 1. Estimasi Design-Based Untuk Variabel Kontinue

Variable	Mean	SE	CI 95%	
			Lower	Upper
Y	3,689,080	33,428.36	3,623,535	3,754,625
X ₂	3.518	0.011	3.496	3.540
X ₅	50.768	0.103	50.566	50.971
X ₈	40.726	0.187	40.360	41.093
X ₉	1.351	1.367	0.008	1.336
X ₁₀	94.731	0.589	93.577	95.886

Berdasarkan hasil estimasi design based, rata-rata total pengeluaran rumah tangga di Jawa Tengah sebesar Rp 3,689,080,-. Rata-rata banyaknya anggota rumah tangga adalah 3.5 (3 sampai 4 orang dalam 1 rumah tangga). Rata-rata umur kepala rumah tangga adalah 50 tahun. Rata-rata jumlah jam kerja kepala rumah tangga dalam satu minggu adalah 40-41 jam. Rata-rata ukuran rumah adalah 95m², dan dalam satu rumah hanya ada satu rumah tangga. Sedangkan estimasi design-based untuk variabel kategorik, dihitung nilai proporsi untuk setiap kategori. Hasil estimasi design-based untuk variabel kategorik disajikan dalam Tabel 2.

Tabel 2. Estimasi Proporsi Untuk Variabel Kategorik

Variabel	Kategori	Proporsi	SE	CI 95%	
				Batas Bawah	Batas Atas
X ₁	Kota	0.512	0.002	0.508	0.516
	Desa	0.488	0.002	0.484	0.492
X ₃	Tidak menikah	0.022	0.001	0.019	0.024
	Menikah	0.810	0.003	0.804	0.815
	Cerai hidup	0.032	0.001	0.029	0.034
	Cerai mati	0.137	0.002	0.132	0.141
X ₄	Laki-laki	0.845	0.003	0.840	0.850
	Perempuan	0.155	0.003	0.150	0.160
X ₇	Tidak bekerja	0.125	0.002	0.120	0.130
	Pertanian	0.341	0.004	0.333	0.349
	Industri	0.240	0.004	0.233	0.247
	Jasa	0.294	0.004	0.287	0.301
X ₈	Tidak pernah sekolah	0.049	0.002	0.045	0.052
	SD	0.494	0.004	0.487	0.501
	SMP	0.167	0.002	0.162	0.171
	SMA	0.217	0.003	0.212	0.222
	Diploma	0.014	0.001	0.013	0.016
	Sarjana	0.054	0.002	0.051	0.058
X ₁₁	Pasca Sarjana	0.006	0.001	0.005	0.007
	Ya	0.171	0.003	0.166	0.177
	Tidak	0.829	0.003	0.823	0.834

Pada pengambilan sampel ini dilakukan stratifikasi berdasarkan perkotaan dan pedesaan, sehingga proporsi antara perkotaan dan pedesaan adalah hampir sama. Status kepala rumah tangga 81% adalah menikah, dan 84.5% adalah laki-laki. Sebesar 34% kepala rumah tangga bekerja pada sektor pertanian, 29.4% pada sektor jasa dan 24% berada pada sektor industri. 49.4% kepala rumah tangga mempunyai pendidikan hanya sampai jenjang

SD dan 21.7% pada sektor SMA. Selain itu hanya 17.1% rumah tangga yang menerima program PKH dari pemerintah.

Pengujian multikolinieritas pada variabel kontinue menunjukkan bahwa nilai Variance inflation Factor (VIF) < 5, hal ini menunjukkan tidak ada multikolinieritas pada variabel kontinue. Untuk variabel yang bersifat kategorik tidak dilakukan pengujian multikolinieritas, karena dalam pemodelan menggunakan dummy variabel.

Tabel 3. Nilai VIF Pada Variabel Independent

Variabel	VIF
X_2	1.60
X_5	1.65
X_8	2.17
X_9	1.28
X_{10}	1.08

Pada pemodelan ini variabel kontinue tidak dilakukan standardisasi, karena pemodelan menggabungkan antara variabel independent bernilai kategorik dan kontinue. Hasil estimasi parameter model regresi linier berganda dengan complex survey design disajikan pada Tabel 4.

Tabel 4. Estimasi Parameter Regression Dalam Desain Survei Kompleks Untuk Semua Variabel Independen

Variabel	Koefisien	Standard Error	t-value	Prob	CI 95%	
					Batas Bawah	Batas Atas
X_1 _desa	-380848	51431.57	-7.4	< 0.001	-481693	-280002
X_2	551925.1	22534.17	24.49	< 0.001	507740.8	596109.5
X_3 _Menikah	477521.7	129478.3	3.69	< 0.001	223644.3	731399
X_3 _Cerai hidup	585797.4	166371.1	3.52	< 0.001	259581.7	912013.2
X_3 _Cerai Mati	567380.6	146961	3.86	< 0.001	279223.7	855537.4
X_4 _Perempuan	-192091	72326.06	-2.66	0.008	-333906	-50276
X_5	-2972.52	2008.161	-1.48	0.139	-6910.07	965.019
X_6 _SD	221166.1	62330.48	3.55	< 0.001	98950.3	343381.9
X_6 _SMP	399909.8	79254.2	5.05	< 0.001	244510.4	555309.1
X_6 _SMA	1153371	87403.78	13.2	< 0.001	981992.1	1324750
X_6 _Diploma	3460714	310943.2	11.13	< 0.001	2851026	4070403
X_6 _Sarjana	3469670	207451.3	16.73	< 0.001	3062906	3876435
X_6 _Pasca Sarjana	8778168	1070695	8.2	< 0.001	6678779	1.09E+07
X_7 _Pertanian	-599254	94339.26	-6.35	< 0.001	-784231	-414276
X_7 _Industri	-202292	113781.6	-1.78	0.076	-425392	20807.39
X_7 _Jasa	231079.4	112812.4	2.05	0.041	9880.028	452278.8
X_8	7179.031	1750.172	4.1	< 0.001	3747.344	10610.72
X_9	-130126	55319.18	-2.35	0.019	-238594	-21657.3
X_{10}	10819.33	873.4582	12.39	< 0.001	9106.678	12531.98
X_{11} _Tidak	609765.7	37884.35	16.1	< 0.001	535483.1	684048.2
Konstanta	-544196	185010.9	-2.94	0.003	-906960	-181432

Untuk pengujian secara serentak, hipotesisnya adalah sebagai berikut:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{11} = 0$$

H_1 : paling tidak ada satu $\beta_j \neq 0$ dengan j merupakan variabel independent

Hasil analisis didapatkan nilai F-test sebesar 157.96 dengan probability < 0.0001, hal ini menunjukkan bahwa secara keseluruhan model signifikan. Nilai R-square yang dihasilkan dari model tersebut sebesar 25.52%. Menurut Heeringa et al.,(2010), untuk penelitian sosial,

nilai R-square berkisar antara 20-40% sudah baik. Sedangkan untuk pengujian partial, hipotesisnya adalah sebagai berikut:

$H_0 : \beta_j = 0$ (variabel independen ke- j tidak signifikan dalam model)

$H_1 : \beta_j \neq 0$ (variabel independen ke- j signifikan dalam model)

Berdasarkan hasil pengujian partial model X_5 dan X_7 *Industri* tidak signifikan dalam model (dapat dilihat pada Tabel 4). Untuk mendapatkan model yang terbaik (signifikan disemua variabel independent), variabel X_5 perlu dikeluarkan didalam model, sementara untuk variabel X_7 *Industri* tidak perlu dikeluarkan dari model karena bernilai kategorik.

Hasil estimasi setelah variabel X_5 dikeluarkan didalam model dapat dilihat pada Tabel 5. Nilai F-statistics yang dihasilkan sebesar 165.62 dengan probability < 0.0001, ini menunjukkan bahwa model yang dihasilkan dari selected independent variables adalah significant. Nilai R square yang dihasilkan sebesar 25.51%, tidak jauh berbeda dengan menggunakan semua variabel independent.

Tabel 5. Estimasi model regresi dengan complex survey design untuk variabel independent yang terpilih.

Variabel	Koefisien	SE	t-value	Prob	CI 95%	
					Batas Bawah	Batas Atas
X_1 Desa	-373492	50645.84	-7.37	< 0.001	-472797	-274187
X_2	554448.2	22116.48	25.07	< 0.001	511082.9	597813.6
X_3 Menikah	436799.7	125192.9	3.49	< 0.001	191325.1	682274.3
X_3 Cerai hidup	543197	164315.6	3.31	0.001	221011.8	865382.2
X_3 Cerai Mati	498164.3	140307.5	3.55	< 0.001	223053.3	773275.2
X_4 Perempuan	-180817	71805.6	-2.52	0.012	-321612	-40022.9
X_6 SD	240523.9	60470.76	3.98	< 0.001	121954.6	359093.2
X_6 SMP	441463.8	70845.21	6.23	< 0.001	302552.6	580375
X_6 SMA	1192689	79858.49	14.94	< 0.001	1036105	1349273
X_6 Diploma	3495243	309675.9	11.29	< 0.001	2888039	4102446
X_6 Sarjana	3509181	211122.3	16.62	< 0.001	3095219	3923144
X_6 Pasca Sarjana	8815458	1068025	8.25	< 0.001	6721305	1.09E+07
X_7 Pertanian	-590384	95954.64	-6.15	< 0.001	-778529	-402239
X_7 Industri	-172188	117764.3	-1.46	0.144	-403097	58721.27
X_7 Jasa	256821.3	117060	2.19	0.028	27293.4	486349.2
X_8	7473.491	1687.918	4.43	< 0.001	4163.87	10783.11
X_9	-133139	54941.75	-2.42	0.015	-240867	-25410.5
X_{10}	10757.93	863.6936	12.46	< 0.001	9064.422	12451.43
X_{11} Tidak	608702.9	37962.34	16.03	< 0.001	534267.5	683138.4
_cons	-712418	157323.4	-4.53	< 0.001	-1020893	-403943

Interpretasi model regresi dapat dilakukan pada setiap variabel dan diasumsikan variabel lainnya bernilai tetap atau konstan. Interpretasi dilakukan dengan melihat nilai koefisien setiap parameter. Hasil interpretasi model regresi total pengeluaran di Jawa Tengah menunjukkan bahwa pengeluaran rumah tangga yang berada di desa lebih rendah daripada pengeluaran yang di kota. Pengeluaran rumah tangga dengan status kepala rumah tangga adalah tidak menikah lebih rendah daripada status kepala rumah tangga menikah, cerai hidup, dan cerai mati. Kepala rumah tangga dengan jensi kelamin wanita cenderung mempunyai total pengeluaran yang lebih rendah daripada laki-laki. Kepala rumah tangga yang pernah mengenyam pendidikan memiliki pengeluaran yang lebih besar dibandingkan yang tidak pernah bersekolah. Hal ini diasumsikan bahwa seseorang dengan pendidikan yang tinggi, mampu mendapatkan pekerjaan dan gaji yang lebih baik dibandingkan dengan yang tidak pernah mengenyam pendidikan atau memiliki tingkat pendidikan yang lebih

rendah. Sementara, kepala rumah tangga yang bekerja pada sektor jasa, cenderung memiliki pengeluaran yang lebih tinggi dibandingkan dengan sektor lainnya (pertanian dan industry). Berdasarkan bantuan dari pemerintah, rumah tangga yang tidak menerima bantuan program PKH dari pemerintah memiliki total pengeluaran yang lebih tinggi dibandingkan rumah tangga yang menerima bantuan program. Bantuan program PKH ditujukan untuk rumah tangga miskin, rumah tangga miskin didefinisikan sebagai rumah tangga yang memiliki total pengeluaran dibawah garis kemiskinan. Pada variabel independent yang bersifat kontinue, yaitu umur kepala rumah tangga, lama jam bekerja selama satu minggu dan ukuran rumah memiliki koefisien parameter bernilai positif, artinya semakin bertambah umur kepala rumah tangga, jam bekerja dan ukuran rumah maka akan bertambahnya total pengeluaran rumah tangga. Penelitian ini tidak melakukan standarisasi untuk variabel continue, sehingga tidak bisa dilihat perbandingan antar variabel yang bersifat kontinue.

Selain itu, hasil interpretasi model regresi ini, dapat ditindaklanjuti dengan melakukan estimasi tabulasi silang antara setiap variabel independent yang bernilai kategorik dengan variable dependent. Estimasi tabulasi silang dilakukan berdasarkan estimasi design based untuk rata-rata.

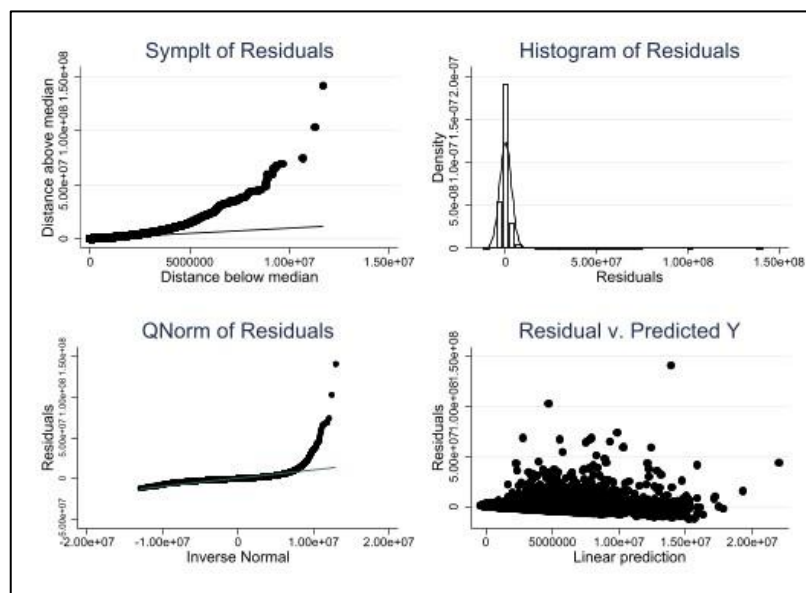
Table 6. Estimasi tabulasi silang antara variabel independent kategorik dengan variabel dependent

Variabel	Mean	SE	CI 95%	
			Batas Bawah	Batas Atas
Kota	4,208,938	58,102.8	4,095,012	4,322,864
Desa	3,143,050	31,471.2	3,081,342	3,204,758
Laki-Laki	3,874,526	36498.43	3,802,961	3,946,091
Perempuan	2,677,332	49977.44	2,579,338	2,775,326
Tidak Menikah	2,804,642	110,455.3	2,588,064	3,021,219
Menikah	3,915,566	37,734.9	3,841,576	3,989,555
Cerai Hidup	3,055,304	119,768.0	2,820,466	3,290,142
Cerai Mati	2,636,770	50,664.4	2,537,428	2,736,111
Tidak pernah sekolah	2,018,499	58,235.0	1,904,313	2,132,685
SD	2,965,688	26,764.9	2,913,208	3,018,168
SMP	3,507,220	40,942.7	3,426,941	3,587,500
SMA	4,533,291	59,725.6	4,416,183	4,650,400
Diploma	7,149,694	319,186.7	6,523,842	7,775,546
Sarjana	7,043,036	236,074.9	6,580,148	7,505,925
Pasca Sarjana	13,300,000	1,125,978.0	11,000,000	15,500,000
Tidak bekerja	3,211,125	72,110.6	3,069,732	3,352,517
Pertanian	2,876,720	32,658.9	2,812,683	2,940,756
Industri	3,753,694	54,965.1	3,645,920	3,861,467
Jasa	4,780,810	68,188.2	4,647,108	4,914,511
Ya	2,855,843	32,174.6	2,792,756	2,918,930
Tidak	3,861,458	38,339.34	3,786,284	3,936,633

Pada Tabel 6, untuk status wilayah rumah tinggal pada kategori pedesaan dan perkotaan nilai CI 95% tidak beririsan. Ini menunjukkan bahwa terdapat perbedaan rata-rata total pengeluaran rumah tangga di desa dan di kota. Pengeluaran rumah tangga di kota lebih tinggi daripada di desa. Pada variabel status kepala rumah tangga, nilai CI 95% pada kategori tidak menikah, cerai hidup, and cerai mati saling beririsan. Ini menunjukkan tidak ada perbedaan total pengeluaran pada kategori-kategori tersebut. Sementara, untuk variabel tingkat pendidikan kepala rumah tangga, pada kategori diploma dan sarjana terdapat irisan pada CI 95%, ini menunjukkan tidak ada perbedaan total pengeluaran antara tingkat pendidikan diploma dan sarjana. Sementara untuk kategori tingkat pendidikan lainnya (tidak pernah bersekolah, SD, SMP, SMA, and Pasca Sarjana), terdapat perbedaan total pengeluaran

diantara kategori-kategori tersebut. Untuk variabel pekerjaan kepala rumah tangga, terdapat perbedaan total pengeluaran diantara kategori pada variable lapangan pekerjaan. Kategori pengeluaran yang paling rendah adalah pada bidang pertanian, ini mengindikasikan bahwa kepala rumah tangga yang bekerja di sektor pertanian cenderung memiliki pengeluaran yang paling kecil diantara sektor lainnya. Sementara untuk yang berstatus tidak bekerja, memungkinkan bahwa kepala rumah tangga tersebut mendapatkan dana pensiun untuk memenuhi pengeluaran rumah tangga. Sektor yang paling tinggi total pengeluarannya adalah sektor jasa.

Pengujian diagnostik model regresi linier berganda disajikan pada Gambar 2. Pada Gambar 2 mengindikasikan bahwa model tidak mengikuti distribusi normal. Hal ini bisa dilihat dari keempat diagram bahwa “symplt of residuals” and “QNorm of Residuals” tidak mengikuti garis lurus atau linier. Pada plot tersebut terlihat adanya pencilan dalam model. Hal ini diperkuat juga pada “Histogram of residual”, ekor distribusi cenderung condong ke kiri. Sementara sebaran “residual v. predicted Y” cenderung beberapa titik tidak berkelompok, yang mengindikasikan sebaran tidak normal.



Gambar 2. Pengujian Diagnostik model total pengeluaran rumah tangga di Jawa Tengah

5. KESIMPULAN

Penelitian ini mampu memodelkan total pengeluaran rumah tangga di Jawa Tengah dengan beberapa dimensi antara lain banyaknya anggota rumah tangga, status wilayah, deskripsi kepala rumah tangga, menerima bantuan dari pemerintah, ukuran rumah serta banyaknya rumah tangga dalam satu atap. Model menghasilkan nilai R-square 25.51%, yang dikatakan sudah cukup baik untuk data survey. Model regresi yang dihasilkan tidak memenuhi asumsi normalitas, dikarenakan adanya pencilan. Jika pengamatan yang mengandung pencilan dibuang, maka survey design dan sampling weight akan berubah, sehingga tidak mencerminkan populasi. Penelitian ini tidak melakukan investigasi lebih lanjut terkait menggunakan regresi robust untuk data survey yang mengandung pencilan. Karakteristik total pengeluaran di provinsi Jawa Tengah yang lebih tinggi berdasarkan variabel kategorik antara lain tinggal di perkotaan, kepala keluarga berjenis kelamin laki-laki, dengan status sudah menikah, dengan pendidikan terakhir adalah sarjana atau pasca sarjana, bekerja pada

bidang jasa, dan tidak menerima program bantuan dari pemerintah. Sedangkan untuk variabel kontinue mempunyai koefisien yang positif, artinya jika semakin bertambah satu satuan unit pada variabel tersebut, akan meningkatkan total pengeluaran, antara lain pada variabel umur kepala rumah tangga, jumlah jam bekerja selama satu minggu, dan ukuran rumah tinggal. Pada penelitian ini tidak bisa membandingkan kontribusi terbesar variabel independent kontinue terhadap total pengeluaran karena tidak dilakukan standardisasi variabel.

DAFTAR PUSTAKA

- Abubakar, A. H., Idowu, A. A., & Ogiugo, G. O. (2021). Comparative Study of Income Expenditure on Household Consumption of Food and Non-Food Items in Rural and Urban Areas of Nigeria: Non-Parametric Approach. *RUJMASS*, 7(1), 83–90.
- Ali, A. F. M., Ibrahim, M. F., & Aziz, M. R. A. (2018). Food and Non-Food Expenditure Trends Among the Poor and Needy in Kelantan, Malaysia. *Journal of Entrepreneurship and Business*, 6(1), 59–78.
- BPS. (2021). *Buku Pedoman 1: Survey Sosial Ekonomi Marret 2021*. BPS.
- Chambers, R. ., & Skinner, C. . (2003). *Analysis of Survey Data*. John Wiley and Sons.
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. John Wiley & Sons, Inc.
- García, T., & Grande, I. (2010). Determinants of food expenditure patterns among older consumers. The Spanish case. *Appetite*, 54(1), 62–70.
- Hahs-Vaughn, D. L., McWayne, C. M., Bulotsky-Shearer, R. J., Wen, X., & Faria, A. M. (2011). Methodological considerations in using complex survey data: An applied example with the head start family and child experiences survey. *Evaluation Review*, 35(3), 269–303.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. Chapman & Hall Taylor and Francis.
- Jacobson, D., Mavrikiou, P. M., & Minas, C. (2010). Household size, income and expenditure on food: The case of Cyprus. *Journal of Socio-Economics*, 39(2), 319–328.
- Lohr, S. L. (2010). *Sampling: Design and Analysis* (2nd ed.). Brooks/Cole Cengage Learning.
- Lumley, T. (2010). *Complex Surveys: A Guide to Analysis using R*. John Wiley and Sons.
- Sekhampu, T. J. (2012). Socio-economic determinants of household food expenditure in a low income township in South Africa. *Mediterranean Journal of Social Sciences*, 3(3), 449–453.
- Smith, C., Parnell, W. R., Brown, R. C., & Gray, A. R. (2013). Providing additional money to food-insecure households and its effect on food expenditure: A randomized controlled trial. *Public Health Nutrition*, 16(8), 1507–1515.
- Utami, J. P., & Ayu, S. F. (2017). Food and Non-Food Consumption Expenditure In Medan City and Its Affecting Factors (Case Study of Java and Batak Tribes). *Proceedings of the 2nd International Conference on Social and Political Development (ICOSOP 2017)*, 488–501.
- Venn, D., Dixon, J., Banwell, C., & Strazdins, L. (2018). Social determinants of household food expenditure in Australia: The role of education, income, geography and time. *Public Health Nutrition*, 21(5), 902–911.