

Received: 25 Januari 2021

Revised: 10 Juni 2021

Accepted: 23 Juni 2021

Published: 30 Juni 2021

Analisis Angka Kelahiran pada Remaja Indonesia Usia 15-19 Menggunakan Regresi Binomial Negatif

Silvia^{1,a)}, Geraldine Immanuel Tangyong^{1,b)}, Yekti Widyaningsih^{1,c)}

¹*Departemen Matematika, FMIPA, Universitas Indonesia*

E-mail: ^{a)}silvia82@sci.ui.ac.id, ^{b)}geraldine.immanuel@sci.ui.ac.id, ^{c)}yekti@sci.ui.c.id (Penulis Korespondensi)

Abstract

According to the BKKBN, the birth rate for adolescent girls is quite high. This needs special attention, because the number of adolescents in Indonesia is very large and at the age of 15-19 a person should still be in the process of getting an education. This study aims to determine the factors affecting the birth rate of women aged 15-19 years in 2012 based on provinces in Indonesia. The factors in this study are the Human Development Index (IPM) by province, open unemployment rate (percentage) by province, number of poor people by province, literacy rate, and average length of schooling. This study uses Poisson regression analysis, especially negative binomial regression. The test results show that the data has an overdispersion. To overcome this condition, negative binomial regression is used. Factors that significantly influence the birth rate of women aged 15-19 years are literacy rates, average length of schooling, and number of poor people.

Keywords: Poisson Regression, Negative Binomial Regression, Generalized Linear Models, Overdispersion, Birth Rate of Women

Abstrak

Menurut BKKBN, angka kelahiran untuk perempuan usia remaja terbilang cukup tinggi. Hal ini perlu menjadi perhatian khusus, karena jumlah remaja di Indonesia sangat besar dan pada umur 15-19 tahun seseorang seharusnya masih dalam proses mengenyam pendidikan. Penelitian ini bertujuan untuk mengetahui faktor-faktor yang memengaruhi angka kelahiran perempuan umur 15-19 tahun pada tahun 2012 berdasarkan provinsi di Indonesia. Faktor-faktor dalam penelitian ini adalah Indeks Pembangunan Manusia (IPM) menurut provinsi, tingkat pengangguran terbuka (dalam persen) menurut provinsi, jumlah penduduk miskin menurut provinsi, angka melek huruf, dan rata-rata lama sekolah. Penelitian ini menggunakan analisis regresi Poisson, khususnya regresi binomial negatif. Hasil pengujian menunjukkan bahwa data mengalami overdispersi. Untuk mengatasi kondisi ini digunakan regresi Binomial Negatif. Faktor yang berpengaruh signifikan terhadap angka kelahiran pada perempuan berumur 15-19 tahun adalah angka melek huruf, rata-rata lama sekolah, dan jumlah penduduk miskin.

Kata-kata kunci: Regresi Poisson, Regresi Binomial Negatif, Generalized Linear Models, Overdispersi, Angka Kelahiran pada Perempuan Umur 15-19 Tahun

PENDAHULUAN

Angka kelahiran atau sering juga disebut natalitas merupakan frekuensi kelahiran hidup pada suatu populasi, dimana angka tersebut didapatkan dari jumlah kelahiran hidup per seribu penduduk setiap tahunnya. Di Indonesia, angka kelahiran dapat dikatakan cukup tinggi. Setiap tahun terdapat sekitar 4,8 juta anak lahir di Indonesia, maka Indonesia mengambil urutan keempat, setelah Cina, India, dan Amerika Serikat, dengan populasi yang hampir mencapai 270 juta jiwa.

Di Indonesia, angka kelahiran cukup tinggi diikuti dengan angka kelahiran pada remaja. Menurut data dari Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), angka kelahiran untuk perempuan pada rentang umur 15-19 tahun terbilang cukup tinggi, dimana angkanya mencapai 48 dari 1000 kelahiran. Di Provinsi Kalimantan Barat, angka kelahiran mencapai 104 per 1000 kelahiran (Puspitasari, 2015). Angka tersebut tentu saja lebih besar dari rata-rata nasional. Dengan jumlah remaja di Indonesia yang sangat besar (67 juta jiwa), maka angka kelahiran pada usia remaja ini harusnya menjadi perhatian khusus dari pemerintah, mengingat pada rentang umur 15-19 tahun, seseorang seharusnya masih dalam proses mengenyam pendidikan.

Studi oleh Zahab, Dharmawan, dan Winami (2017, pp.56-65) meneliti hubungan antara perilaku pacaran remaja usia 15-19 tahun di Indonesia. Hasil studi tersebut adalah tidak adanya hubungan antara cara berpacaran pada remaja terhadap tingginya angka kelahiran pada remaja. Studi oleh BKKBN dan Fakultas Psikologi UBAYA dalam laporan penelitian BKKBN tahun 2013 menganalisis hal yang serupa, yaitu faktor yang memengaruhi wanita melahirkan pada usia 15-19 tahun. Secara keseluruhan, faktor-faktor internal yang memengaruhi adalah kebutuhan material, kebutuhan seksual, dan kebutuhan psikologis. Merujuk pada faktor eksternal, pengaruh diberikan oleh faktor sosial budaya, faktor teknologi, dan faktor lingkungan. Kedua studi tersebut memandang fenomena kelahiran pada usia remaja tersebut melalui aspek sosial, budaya, psikologi, dan lingkungan. Bagaimanakah pengaruh dari aspek kesejahteraan masyarakat dan kependudukan?

Terdapat beberapa faktor yang memengaruhi tingkat kelahiran, antara lain tingkat aborsi, struktur usia-jenis kelamin yang ada, kepercayaan sosial dan religius, tingkat buta aksara (melek huruf) pada wanita, kemakmuran secara ekonomi, tingkat kemiskinan, angka kematian bayi, urbanisasi, usia pernikahan, tersedianya pensiun, dan berbagai faktor lainnya (Faqih, 2017).

Merujuk pada faktor-faktor tersebut, studi ini bertujuan untuk menganalisis angka kelahiran pada remaja Indonesia usia 15-19 berdasarkan pada Indeks Pembangunan Manusia (IPM) menurut provinsi, tingkat pengangguran terbuka (dalam persen) menurut provinsi, jumlah penduduk miskin menurut provinsi, angka melek huruf, dan rata-rata lama sekolah. Pada data yang digunakan, tingkat pendidikan atau melek huruf sangatlah dipertimbangkan, mengingat perempuan pada rentang umur 15-19 tahun merupakan rentang umur yang seharusnya masih dalam bangku pendidikan.

Data yang digunakan merupakan data Angka Kelahiran Pada Perempuan Umur 15-19 Tahun per Provinsi pada tahun 2012, disertai dengan informasi mengenai faktor-faktor yang akan dianalisis. Variabel respons angka kelahiran pada perempuan umur 15-19 tahun merupakan variabel diskrit. Kelahiran pada perempuan umur 15-19 tahun merupakan kejadian yang tidak sering terjadi. Analisis regresi linear tidak sesuai untuk variabel respons yang demikian. Oleh karena itu, digunakan model regresi Poisson dalam memodelkannya. Jika ekuidispersi tidak terpenuhi, maka alternatif lainnya adalah menggunakan regresi binomial negatif.

METODOLOGI

Data

Data mengandung 33 observasi dengan tujuh variabel. Variabel respon pada penelitian ini adalah angka kelahiran pada remaja usia 15-19 tahun. Sedangkan variabel bebasnya adalah X1 (persentase penduduk miskin), X2 (angka melek huruf), X3 (rata-rata lama sekolah dalam tahun), X4 (indeks pembangunan manusia), X5 (tingkat pengangguran terbuka), X6 (jumlah penduduk miskin dalam ribuan jiwa).

Analisis data pada penelitian ini menggunakan bantuan *software R Studio*. Adapun langkah-langkah analisis datanya adalah: (1) Memeriksa terjadinya overdispersi pada data yang dianalisis; (2) Memeriksa distribusi variabel respon (3) Mengajukan model regresi; (4) Menentukan model terbaik untuk data; (5) Melakukan pendugaan parameter; (6) Melakukan pengujian parsial parameter; (7) Melakukan uji kecocokan model; (8) Menginterpretasikan parameter; dan (9) *Cross Validation*.

Metode Penelitian

Analisis Regresi

Analisis regresi mempelajari tentang hubungan antar variabel, dimana hubungan yang diperoleh pada umumnya dinyatakan dalam bentuk persamaan matematik dimana terdapat hubungan fungsional antar variabelnya. Dalam analisis regresi terdapat dua jenis variabel, yaitu variabel bebas (prediktor) dan variabel terikat (respon). Pada umumnya, variabel bebas dinotasikan dengan X_1, X_2, \dots, X_k sedangkan variabel terikat dinotasikan dengan Y .

Analisis regresi memiliki banyak manfaat dan diterapkan dalam berbagai bidang, baik dalam bidang ekonomi, lingkungan, industri, dan sebagainya. Kegunaan analisis regresi sendiri adalah sebagai berikut;

1. Mengetahui dan memprediksi pengaruh variabel prediktor terhadap variabel respon.
2. Membuat estimasi rata-rata dan nilai variabel respon dilihat dari variabel prediktor.
3. Menguji hipotesis karakteristik dependensi.
4. Memprediksi nilai rata-rata variabel prediktor yang didasari pada nilai variabel prediktor diluar jangkauan sampel.

Generalized Linear Model (GLM)

Generalized Linear Model merupakan perluasan dari model regresi linier, dimana terdapat asumsi prediktor memiliki efek linier akan tetapi tidak mengasumsikan distribusi tertentu dari variabel respon. *Generalized Linear Model* digunakan ketika variabel responnya mengikuti distribusi keluarga eksponensial. Pada umumnya, GLM terdiri dari tiga komponen: komponen random, sistematik, dan fungsi penghubung yang menghubungkan dua komponen untuk menghasilkan prediktor linier (Lord & Persaud, 2000). Asumsi penting untuk GLM adalah *random error* hanya terjadi pada variabel terikat dan bahwa variabel-variabel bebas diketahui tanpa *error* (Maher & Summersgill, 1996). GLM dapat dinyatakan sebagai berikut:

- a. y_1, y_2, \dots, y_n merupakan n variabel bebas yang mengikuti distribusi peluang anggota keluarga eksponensial, diketahui $E[Y_i] = \mu_i = g^{-1}(\mathbf{X}\beta)$ sehingga $g(\mu_i) = \mathbf{X}\beta$. $g(\cdot)$ merupakan fungsi penghubung atau *link function*.
- b. $x_{i1}, x_{i2}, \dots, x_{ik}$ merupakan n variabel bebas (prediktor linear) yang dinotasikan dalam bentuk $\mathbf{x}'_i\beta$, dimana $\mathbf{x}'_i\beta = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}$, $i = 1, 2, 3, \dots, n$, $j = 1, 2, \dots, k$.
- c. $g(\cdot)$ Merupakan fungsi pemetaan yang menghubungkan prediktor linear dengan mean responnya adalah $\mathbf{x}'_i\beta = g(\mu)$. Beberapa fungsi penghubung yang umum adalah

- **Identity Link**

$$g(\mu_i) = \mu_i = \mathbf{X}\beta$$

- **Logit Link**

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \mathbf{X}\beta \leftrightarrow \mu_i = \frac{e^{\mathbf{X}\beta}}{1 + e^{\mathbf{X}\beta}}$$

- **Log Link**

$$g(\mu_i) = \log(\mu_i) = \mathbf{X}\beta \leftrightarrow \mu_i = e^{\mathbf{X}\beta}$$

- **Probit Link**

$$g(\mu_i) = \Phi^{-1}(\mu_i) = \mathbf{X}\beta \leftrightarrow \mu_i = \Phi(\mathbf{X}\beta)$$

- **Complementary Log-Log Link (Gompit Link)**

$$g(\mu_i) = \log(-\log(1 - \mu_i)) = \mathbf{X}\beta \leftrightarrow \mu_i = 1 - \exp\{-e^{\mathbf{X}\beta}\}$$

- **Power Link**

$$g(\mu_i) = \mu^\lambda = \mathbf{X}\beta \leftrightarrow \mu_i = (\mathbf{X}\beta)^{\frac{1}{\lambda}}$$

d. Variansi dari variabel random Y dapat ditulis sebagai $Var(Y) = V(\mu_i) = V(g^{-1}(\mathbf{X}\beta))$

Regresi Poisson

Regresi Poisson merupakan bagian dari *Generalized Linear Models* dan digunakan untuk data cacah (*count*) atau kejadian diskrit yang langka. Keuntungan penggunaan regresi Poisson yaitu distribusi ini dirancang untuk distribusi variabel terikat diskrit dan yang seringkali sangat menceng (*skewed*).

Distribusi Poisson memodelkan probabilitas diskrit dengan:

$$P(Y) = \frac{e^{-\mu} \mu^Y}{Y!}$$

dengan Y merupakan jumlah kejadian dan μ merupakan rata-rata kejadian dari periode tersebut (Pardoe, 2018). Model regresi Poisson dapat ditulis:

$$y_i = \mu_i + \varepsilon_i, i = 1, 2, \dots, n$$

dimana y_i merupakan jumlah kejadian dan μ_i merupakan rata-rata jumlah kejadian.

Pemetaan log untuk regresi Poisson adalah:

$$g(\mu_i) = \ln \mu_i$$

maka didapat hubungan antara mean respon dengan prediktor liniernya sebagai berikut:

$$\mu_i = g^{-1}(\mathbf{X}\beta) = e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}}$$

Dalam metode regresi Poisson, diasumsikan bahwa mean dan variansi dari respon adalah sama. Namun, terdapat suatu masalah utama, yaitu nilai variansi lebih besar dari mean (*overdispersi*) dan nilai variansi lebih kecil dari mean (*underdispersi*). Untuk mengatasi kejadian *overdispersi*, seringkali digunakan model Binomial Negatif sebagai ekstensi dari model regresi Poisson.

Regresi Binomial Negatif

Regresi Binomial Negatif merupakan ekstensi dari regresi Poisson dan digunakan untuk mengatasi masalah yang terjadi pada pemodelan regresi Poisson. Model binomial negatif dianggap sebagai distribusi yang lebih umum digunakan untuk data *count* dan membantu mengatasi masalah *overdispersi* (Bouk, 2016).

Distribusi binomial negatif memiliki banyak cara dalam pendekatannya, salah satunya sebagai barisan percobaan Bernoulli dan sebagai distribusi campuran Poisson-Gamma. Pendekatan klasik sebagai barisan percobaan Bernoulli memberikan probabilitas diskrit dengan:

$$P(Y = y) = \binom{y+k-1}{k-1} p^k (1-p)^y, k = 0, 1, 2, \dots$$

dengan Y merupakan jumlah kegagalan yang dialami sebelum terjadinya sukses ke-k dengan peluang suksesnya adalah (Widya, 2011).

Untuk mengakomodir *overdispersi* pada regresi Poisson, maka penurunan distribusi binomial negatif tidak berhubungan sebagai barisan dari percobaan Bernoulli, melainkan sebagai distribusi campuran Poisson-Gamma (Widya, 2011). Misalkan Y variabel acak berdistribusi Poisson dengan parameter λ dengan fungsi probabilitas bersyarat $f(y|\lambda)$. Saat terjadi *overdispersi*, λ merupakan nilai dari variabel

acak Λ yang berdistribusi Gamma dengan parameter $\frac{1}{\alpha}$ dan $\alpha\mu$ dengan fungsi probabilitas $h(\lambda)$. Fungsi probabilitas marginal dari Y adalah $f(y) = \int_{\lambda} f(y|\lambda)h(\lambda) d\lambda = f(y; \mu, \alpha)$

Distribusi binomial negatif dinyatakan sebagai berikut:

$$f(y; \mu, \alpha) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\mu}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1 + \alpha\mu}\right)^{y_i} \tag{1}$$

Persamaan (1) sesuai dengan distribusi binomial negatif yang didekatkan melalui percobaan Bernoulli, dengan $k = \frac{1}{\alpha}$ dan $p = \frac{1}{1 + \alpha\mu}$. Distribusi binomial negatif memiliki mean μ dan variansi $\mu + \alpha\mu^2$. Definisi variabel acak yang digunakan adalah jumlah kejadian dengan dispersi pada data sebesar $k = 1/\alpha$.

Hubungan antara mean respon dengan prediktor liniernya sebagai berikut:

$$g(\mu_i) = \ln \mu_i$$

$$\mu_i = g^{-1}(\mathbf{X}\beta) = e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}} = e^{(\mathbf{x}'\beta)}$$

dengan μ_i merupakan nilai ekspektasi dari y_i yang berdistribusi binomial negatif.

Pendugaan Parameter pada Regresi Binomial Negatif

Pendugaan parameter menggunakan metode *Maximum Likelihood Estimator (MLE)*. Fungsi *log likelihood* dari distribusi binomial negatif dinyatakan sebagai berikut:

$$\mathcal{L}(\mu; y, \alpha) = \sum_{i=1}^n y_i \ln \left(\frac{\alpha\mu}{1 + \alpha\mu}\right) - \frac{1}{\alpha} \ln 1 + \alpha\mu + \ln \Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln \Gamma(y_i + 1) - \ln \Gamma\left(\frac{1}{\alpha}\right)$$

Substitusi $\mu = \exp(\mathbf{x}'\beta)$

$$\mathcal{L}(\mu; y, \alpha) = \sum_{i=1}^n y_i \ln \left(\frac{\alpha \exp(\mathbf{x}'\beta)}{1 + \alpha \exp(\mathbf{x}'\beta)}\right) - \frac{1}{\alpha} \ln 1 + \alpha \exp(\mathbf{x}'\beta) + \ln \Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln \Gamma(y_i + 1) - \ln \Gamma\left(\frac{1}{\alpha}\right)$$

Pendugaan parameter ini menggunakan metode *iteratively re-weighted least squares (IRLS)* atau disebut juga metode Fisher *scoring*.

Uji Signifikansi Parameter Regresi Binomial Negatif

Uji yang digunakan adalah uji Wald, dengan hipotesis:

$H_0: \beta_j = 0$ (pengaruh variabel prediktor ke-j tidak signifikan)

$H_1: \beta_j \neq 0$ (pengaruh variabel prediktor ke-j signifikan)

Statistik uji Wald: $W = \left[\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right]^2$

H_0 ditolak jika $W > \chi_{\alpha,1}^2$ (Keswari, Sumarjaya, dan Suciptawati, 2014, p.111)

Saat membandingkan pengaruh dari variabel yang satu dengan yang lain, tidak bisa dikatakan bahwa *p-value* yang lebih kecil berkaitan dengan efek yang lebih besar. Akan tetapi, *p-value* yang lebih kecil mengindikasikan bahwa terdapat bukti yang lebih kuat bahwa pengaruh yang diberikan oleh variabel tersebut signifikan (Dahiru, 2008).

Interpretasi Parameter Regresi Binomial Negatif

Misalkan X_j adalah variabel prediktor ke- j . Untuk setiap kenaikan 1 unit nilai x_j dengan variabel prediktor lainnya tetap, akan diperoleh

$$\ln \hat{E}(Y|X_1 = x_1, \dots, X_j = x_j + 1, \dots, X_k = x_k) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j(x_j + 1) + \dots + \hat{\beta}_k x_k$$

Sehingga $\ln \hat{E}(Y|X_1 = x_1, \dots, X_j = x_j + 1, \dots, X_k = x_k) - \ln \hat{E}(Y|X_1 = x_1, \dots, X_j = x_j, \dots, X_k = x_k)$
 $= \hat{\beta}_j$

$$\frac{\hat{E}(Y|X_1 = x_1, \dots, X_j = x_j + 1, \dots, X_k = x_k)}{\hat{E}(Y|X_1 = x_1, \dots, X_j = x_j, \dots, X_k = x_k)} = e^{\hat{\beta}_j} \tag{2}$$

Persamaan (2) merupakan rasio dari rata-rata Y sesudah dan sebelum kenaikan 1 unit nilai x_j dengan menganggap variabel prediktor lainnya tetap. Maka, interpretasi yang tepat adalah untuk setiap kenaikan 1 unit dari nilai variabel penjelas kontinu X_j , rata-rata Y akan mengalami perubahan sebesar $\exp(\hat{\beta}_j)$ kali, dengan variabel lainnya tetap.

Uji Goodness of Fit

H_0 : keseluruhan model berguna

H_1 : keseluruhan model tidak berguna

Pengujian kecocokan model dengan hipotesis ini dapat dilakukan dengan menggunakan uji Pearson dan Deviance. Uji Pearson membandingkan nilai ekspektasi dan observasi dari tiap sel, sedangkan uji Deviance membandingkan *likelihood* model *saturated* (model yang memiliki tingkat kecocokan sempurna) dengan model yang didapat (Budyanra, 2017, p.40). Statistik uji Pearson untuk regresi binomial negatif dihitung dengan persamaan (3):

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \exp(\mathbf{X}_i \hat{\beta}))^2}{\exp(\mathbf{X}_i \hat{\beta})} \tag{3}$$

Statistik uji Deviance untuk regresi binomial negatif dihitung dengan persamaan (4):

$$D = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\exp\{\mathbf{X}_i \hat{\beta}\}} \right) - (y_i - \exp\{\mathbf{X}_i \hat{\beta}\}) \right] \tag{4}$$

Hipotesis nol ditolak jika nilai statistik uji Pearson dan Deviance lebih besar dari $\chi^2_{\alpha, n-k-1}$

Pseudo R-squared

Statistik *pseudo-R*² atau R_p^2 dihitung dengan persamaan (5):

$$R_p^2 = 1 - \frac{L_f}{L_l} \tag{5}$$

Dimana L_f adalah nilai fungsi *log-likelihood* dari model yang lengkap dan L_l adalah nilai fungsi *log-likelihood* dari model yang hanya mengandung *intercept*. Perlu diperhatikan bahwa interpretasi koefisien determinasi R^2 tidak dapat diterapkan kepada *pseudo-R*². Nilai yang cukup besar mengindikasikan model yang baik, dan nilai yang sangat kecil mengindikasikan model kurang baik.

Cross Validation

Cross validation dilakukan untuk memeriksa apakah model dapat memprediksi observasi yang baru dengan akurat. Ukuran statistik yang digunakan sebagai parameter untuk mengevaluasi keakuratan model regresi diantaranya:

1. *Root Mean Squared Error* (RMSE): ukuran RMSE memberikan rata-rata *error* dari prediksi yang diberikan oleh model. Nilai RMSE yang mengecil meningkatkan keakuratan model
2. *Mean Squared Error* (MAE): ukuran MAE memberikan rata-rata perbedaan nilai aktual dan prediksi secara absolut. Nilai MAE yang kecil menunjukkan bahwa model baik.
3. R^2 : ukuran R^2 menunjukkan seberapa besar persentase variasi dari variabel dependen dijelaskan oleh model secara keseluruhan, model yang baik memberikan nilai R^2 yang besar. (Mishra, 2020)

Metode *cross validation* yang akan digunakan adalah metode *repeated K-fold cross-Validation*. Data dikelompokkan menjadi k subset, setiap subset memperoleh giliran menjadi data *test*, selagi subset-subset yang lain menjadi data *training*. Proses kemudian diulang beberapa kali untuk menghitung *error* dari prediksi secara keseluruhan.

HASIL DAN PEMBAHASAN

Model regresi Poisson yang diajukan adalah

$$Y = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6) + \varepsilon$$

Y : angka kelahiran remaja usia 15-19 tahun

X_1 : persentase penduduk miskin

X_2 : angka melek huruf

X_3 : rata-rata lama sekolah dalam tahun

X_4 : indeks pembangunan manusia

X_5 : tingkat pengangguran terbuka

X_6 : jumlah penduduk miskin dalam ribuan jiwa

Untuk dapat mengajukan model regresi Poisson, asumsi ekuidispersi perlu dipenuhi. Akan diperiksa kemungkinan terjadinya overdispersi pada data.

$$\text{Statistik Uji: } \frac{\text{Pearson's Chi Square}}{n - p - 1} = \frac{\chi^2}{n - p - 1}$$

$$\frac{\chi^2}{n - p - 1} = \frac{112.3071}{33 - 6 - 1} = 4.319504$$

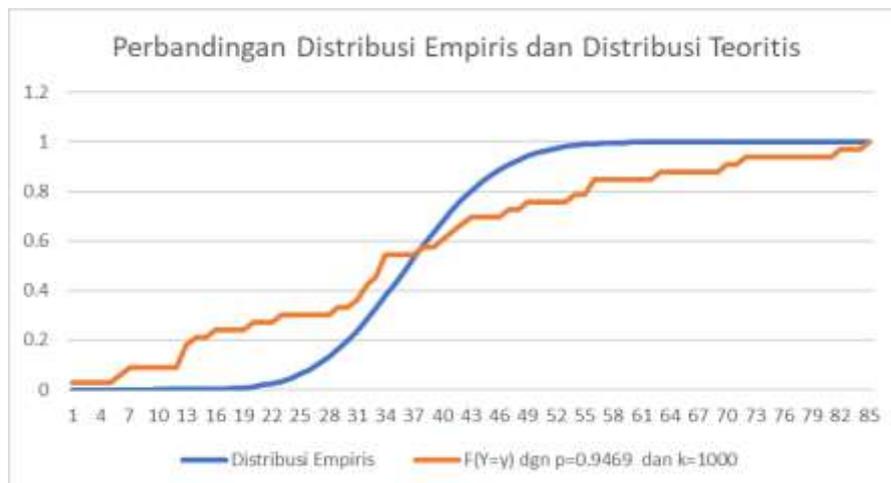
Dengan tingkat kepercayaan 95%, karena statistik uji $\frac{\chi^2}{n-p-1} = 4.319504$ lebih besar dari 1, H_0 ditolak atau data mengalami overdispersi.

Untuk mengatasi masalah overdispersi, dapat diajukan model regresi Binomial Negatif. Model regresi Binomial Negatif mengasumsikan bahwa variabel respons berdistribusi binomial negatif. Oleh karena itu, diasumsikan banyaknya angka kelahiran pada remaja usia 15-19 tahun berdistribusi binomial negatif.

Besar peluang terjadinya kelahiran akan diestimasi menggunakan *software* Rstudio. Diperoleh parameter p untuk distribusi binomial negatif bernilai 0.9469. Akan dibuktikan asumsi banyaknya angka kelahiran pada remaja usia 15-19 tahun berdistribusi binomial negatif dengan parameter $k = 1000$ dan $p = 0.9469$ menggunakan uji Kolmogorov-Smirnov.

$$\text{Statistik uji: } D = \max |F_0(x) - S_N(x)| = 0.268677483$$

Dengan tingkat kepercayaan 90%, karena $D = 0.268677483 < D_{0.99,33} = 0.277$, maka H_0 tidak ditolak atau Y berdistribusi binomial negatif dengan parameter $k = 1000$ dan $p = 0.9469$.



GAMBAR 1. Grafik perbandingan distribusi empiris dan teoritis

Model inisiasi regresi Binomial Negatif yang diajukan adalah

$$Y = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6) + \varepsilon$$

- Y : angka kelahiran remaja usia 15-19 tahun
- X₁ : persentase penduduk miskin
- X₂ : angka melek huruf
- X₃ : rata-rata lama sekolah dalam tahun
- X₄ : indeks pembangunan manusia
- X₅ : tingkat pengangguran terbuka
- X₆ : jumlah penduduk miskin dalam ribuan jiwa

Akan diperiksa kemungkinan terjadinya multikolinearitas pada variabel prediktor menggunakan bantuan *software* Rstudio.

TABEL 1. Keputusan hasil pengujian multikolinearitas

Variabel	VIF	Kriteria
X ₁	1.610093 (*)	10
X ₂	2.658844 (*)	10
X ₃	4.771168 (*)	10
X ₄	4.278078 (*)	10
X ₅	1.849951 (*)	10
X ₆	1.599745 (*)	10

(*) tidak terjadi multikolinearitas

Karena nilai VIF untuk seluruh variabel prediktor < 10 , maka tidak ada indikasi terjadinya multikolinearitas.

Dalam menentukan model terbaik, akan dilakukan *stepwise regression* yang prosesnya akan dilakukan secara *backward*. Kemudian akan dipilih model dengan nilai AIC yang paling kecil.

TABEL 2. Pemilihan model terbaik dengan kriteria AIC

<i>Initial Model</i>	$Y = \exp(\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6) + \varepsilon$
AIC	Model
279.0648	$Y = \exp(\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6) + \varepsilon$
277.1212	$Y = \exp(\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_5 + \beta_5X_6) + \varepsilon$
276.1045	$Y = \exp(\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_6) + \varepsilon$
275.5529	$Y = \exp(\beta_0 + \beta_1X_2 + \beta_2X_3 + \beta_3X_6) + \varepsilon$
<i>Final Model</i>	$Y = \exp(\beta_0 + \beta_1X_2 + \beta_2X_3 + \beta_3X_6) + \varepsilon$

Berdasarkan pada hasil tersebut, maka model regresi Binomial Negatif yang akan digunakan adalah

$$Y = \exp(\beta_0 + \beta_1X_2 + \beta_2X_3 + \beta_3X_6) + \varepsilon$$

Y : angka kelahiran remaja usia 15-19 tahun

X₂ : angka melek huruf

X₃ : rata-rata lama sekolah dalam tahun

X₆ : jumlah penduduk miskin dalam ribuan jiwa

Selanjutnya dengan bantuan *software* Rstudio diperoleh output berikut.

TABEL 3. Nilai dugaan parameter model regresi binomial negatif

Variabel	Parameter	Estimasi parameter	Standard error	Statistik uji Wald	p-value
(Intercept)	β_0	3.789	0.9510	15.87407	6.77×10^{-5}
X ₂	β_1	0.03845	0.01336	8.282849533	0.0042
X ₃	β_2	-0.4286	0.07334	34.15250174	5.0958×10^{-9}
X ₆	β_3	-0.0001158	0.00003833	9.127240207	0.0025

Berdasarkan output model regresi pada TABEL 3, kita peroleh persamaan regresi Binomial Negatif sebagai berikut:

$$\hat{y} = \exp(3.789 + 0.03845X_2 - 0.4286X_3 - 0.0001158X_6)$$

Setelah mendapatkan model regresi, dilakukan pengujian signifikansi parameter regresi. Dengan tingkat kepercayaan 95%, karena nilai statistik uji Walds untuk X₂, X₃, X₆ seluruhnya lebih besar dari nilai $\chi^2_{0.05,1} = 3.841$ dan p-value statistik uji Walds untuk X₂, X₃, X₆ lebih kecil dari $\alpha = 0.05$, maka H₀ ditolak atau pengaruh variabel prediktor masing-masing signifikan.

Berdasarkan kesimpulan tersebut, selain pengaruh dari faktor-faktor sosial, budaya, dan psikologis yang telah dianalisis oleh studi sebelumnya yang serupa, terdapat pula pengaruh yang signifikan dari variabel-variabel pada aspek kesejahteraan masyarakat dan kependudukan, yaitu angka melek huruf, rata-rata lama sekolah dalam tahun, dan jumlah penduduk miskin dalam ribuan jiwa. Untuk mengetahui dominasi pengaruh dari variabel prediktor tersebut, kita urutkan p-value dari yang terkecil hingga yang terbesar. P-value yang kecil mengindikasikan bukti yang lebih kuat bahwa pengaruh variabel prediktornya signifikan. Urutan p-value terkecil hingga terbesar diberikan oleh statistik uji Walds untuk X₃ (rata-rata lama sekolah dalam tahun), X₆ (jumlah penduduk miskin dalam ribuan jiwa), dan terakhir X₂ (angka melek huruf). Oleh karena itu, variabel yang memberikan bukti lebih kuat mengenai

signifikansi pengaruh variabel tersebut secara terurut adalah rata-rata lama sekolah dalam tahun, dilanjutkan jumlah penduduk miskin, dan efek terkecil diberikan oleh angka melek huruf.

Selanjutnya akan diuji kecocokan model secara keseluruhan menggunakan statistik uji Pearson dan Deviance.

TABEL 4. Statistik uji Pearson dan Deviance untuk uji *Goodness of Fit*

Pengujian	Statistik uji	<i>p-value</i>
Deviance	33.06536	0.2751107
Pearson	30.42088	0.3932201

H_0 : keseluruhan model berguna

H_1 : keseluruhan model tidak berguna

Uji hipotesis menggunakan $\alpha = 0.05$ dan hasil perhitungan ditunjukkan pada TABEL 4. Diperoleh *p-value* uji Deviance = 0.2751107 dan *p-value* uji Pearson = 0.3932201. Aturan keputusan dan wilayah kritis: H_0 ditolak jika *p-value* < $\alpha = 0.05$

Kesimpulan: dengan tingkat kepercayaan 95%, karena *p-value* uji Deviance = 0.2751107 dan *p-value* uji Pearson = 0.3932201 keduanya lebih besar dari $\alpha = 0.05$, maka H_0 tidak ditolak atau keseluruhan model berguna.

Diperoleh bahwa model berguna dan seluruh variabel prediktornya berpengaruh secara signifikan, oleh karena itu, parameter-parameter model regresi binomial negatif yang diperoleh selanjutnya dapat diinterpretasikan.

$$\hat{y} = \exp(3.789 + 0.03845X_2 - 0.4286X_3 - 0.0001158X_6)$$

Interpretasi parameter $\hat{\beta}_1 = 0.03845$

Untuk setiap kenaikan satu unit angka melek huruf, maka angka kelahiran remaja usia 15-19 tahun akan mengalami perubahan sebesar $\exp(0.03845) = 1.039198767$ kali dengan variabel lainnya konstan.

Interpretasi parameter $\hat{\beta}_2 = -0.4286$

Untuk setiap kenaikan satu tahun rata-rata lama sekolah, maka angka kelahiran remaja usia 15-19 tahun akan mengalami perubahan sebesar $\exp(-0.4286) = 0.6514204453$ kali dengan variabel lainnya konstan.

Interpretasi parameter $\hat{\beta}_3 = -0.0001158$

Untuk setiap kenaikan seribu jumlah penduduk miskin, maka angka kelahiran remaja usia 15-19 tahun akan mengalami perubahan sebesar $\exp(-0.0001158) = 0.9998842067$ kali dengan variabel lainnya konstan.

Pseudo R-Squared adalah nilai *R-Square* semu yang sama atau identik dengan *R-Square* pada *ordinary least squares* (OLS). Nilai *pseudo R-Squared* (R_p^2) terhitung sebesar 0.5461984 dimana nilainya cukup besar. Dapat dikatakan bahwa model regresi binomial negatif yang diperoleh cukup baik.

Untuk mengetahui apakah model dapat memprediksi observasi yang baru secara akurat, dilakukan *cross validation*. Dipilih metode *repeated k-fold cv* dengan jumlah replikasi sebanyak tiga kali dan jumlah subset sebanyak 10. *Output* yang diperoleh dengan bantuan *software* Rstudio ditunjukkan pada TABEL 5.

TABEL 5. Cross validation metode repeated k-fold

<i>Cross validation (10 fold, repeated 3 times)</i>			
<i>link</i>	RMSE	Rsquared	MAE
log	16.08581	0.7278169	13.44715

Perhatikan bahwa *link function* yang digunakan untuk model regresi binomial kita adalah *link log*. Angka RMSE yaitu 16.08581 cukup kecil, nilai *Rsquared* sebesar 72.78169% cukup besar, dan MAE sebesar 13.44715 cukup kecil. Hal ini mengindikasikan bahwa model regresi binomial negatif yang digunakan cukup akurat dalam memprediksi observasi yang baru.

KESIMPULAN DAN SARAN

Kesimpulan

Variabel yang berpengaruh secara signifikan terhadap angka kelahiran remaja usia 15-19 tahun pada 2012 adalah angka melek huruf, rata-rata lama sekolah, dan jumlah penduduk miskin. Model yang terbaik adalah model regresi binomial negatif dengan persamaan:

$$\hat{y} = \exp(3.789 + 0.03845X_2 - 0.4286X_3 - 0.0001158X_6)$$

Keseluruhan model berguna dan model dapat memprediksi dengan cukup baik observasi yang baru berdasarkan hasil *cross validation*.

Saran

Berdasarkan hasil analisis yang diperoleh, faktor-faktor terkait angka melek huruf, rata-rata lama sekolah, dan jumlah penduduk miskin nampaknya dapat dijadikan fokus untuk menanggulangi tingginya angka kelahiran remaja usia 15-19 tahun pada provinsi-provinsi tertentu yang angka kelahiran di usia dininya melewati rata-rata angka kelahiran Negara. Interpretasi parameter untuk variabel prediktor rata-rata lama sekolah menyatakan bahwa untuk setiap kenaikan 1 tahun rata-rata lama sekolah pada provinsi, maka angka kelahiran remaja usia 15-19 tahun pada provinsi tersebut berkurang lebih dari setengah kalinya dibandingkan tanpa kenaikan 1 tahun rata-rata lama sekolah (variabel prediktor lainnya tetap). Secara subjektif, ini menggambarkan keterkaitan literasi dengan angka kelahiran di usia dini. Pemerintah dan dinas terkait diharapkan dapat konsisten dalam mencanangkan pemerataan pendidikan ke seluruh provinsi, serta menjadikan pendidikan bagi para remaja sebagai perhatian khusus untuk ke depannya. Harapan penulis, ini dapat sekaligus mengurangi angka kelahiran remaja di usia yang masih sepatasnya berada di bangku pendidikan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada RPM FMIPA UI. Penelitian ini didanai oleh Hibah Riset FMIPA UI Tahun 2021, No.ND-871/UN2.F3.D/PPM.00/2021. Ucapan terima kasih juga ditujukan kepada semua pihak yang sudah membantu, yang tidak dapat penulis sebutkan satu persatu.

REFERENSI

- Badan Pusat Statistik. (2012). Angka Kelahiran pada Perempuan Umur 15-19 Tahun Menurut Provinsi 2012, Persentase Penduduk Miskin Menurut Provinsi (Persen), Angka Melek Huruf (Persen), Rata-rata Lama Sekolah (Tahun), Indeks Pembangunan Manusia, Tingkat Pengangguran Terbuka (Persen), Jumlah Penduduk Miskin (Ribuan Jiwa) [online]. Available at: <https://www.bps.go.id/> (Accessed: 24 December 2020)
- Bokuk, Maria Ansila. (2016). Pendugaan Model Regresi Binomial Negatif dengan Metode Kemungkinan Maksimum. Unpublished manuscript, Sanata Dharma University [online]. Available at: <http://repository.usd.ac.id/6699/>
- Budyandra, Budyandra & Azzahra, G. N. (2017) 'Penerapan Regresi Logistik Ordinal Proportional Odds Model pada Analisis Faktor-Faktor yang Mempengaruhi Kelengkapan Imunisasi Dasar Anak Balita di Provinsi Aceh Tahun 2015' *Media Statistika* 10(1), p37-47 [online]. Available at: http://ejournal.undip.ac.id/index.php/media_statistika (Accessed: 31 December 2020)

- Dahiru, T. (2008) 'P-Value, A True Test of Statistical Significance? A Cautionary Note' African Journals Online 6(1) p21-26 [online]. Available at: [10.4314/ajpm.v6i1.64038](https://www.ajol.info/handle/ajol/10.4314/ajpm.v6i1.64038) (Accessed: 20 May 2021)
- Faqih, Achmad. (2010) Kependudukan - Teori, Fakta, dan Masalah. 1st edn. Yogyakarta: Dee Publish.
- Keswari, N. M. R, Sumarjaya, I. W, & Suciptawati, N. L. P. (2014) 'Perbandingan Regresi Binomial Negatif dan Regresi Generalisasi Poisson dalam Mengatasi Overdispersi' E-Jurnal Matematika 3(3) p107-115 [online]. Available at: <https://ojs.unud.ac.id/index.php/mtk/article/view/12001/8301> (Accessed: 31 December 2020)
- Mishra, Rishu. (2020) Cross-Validation in R Programming [online]. Available at: <https://www.geeksforgeeks.org/cross-validation-in-r-programming/> (Accessed: 2 January 2021)
- Pardoe, Iain. (2018) Poisson Regression [online]. Available at: <https://online.stat.psu.edu/stat462/node/209/> (Accessed: 2 January 2021)
- Puspitasari, Wuryanti. (2015). 'Angka Kelahiran pada Remaja Masih Tinggi', ANTARA News, 6 January [online]. Available at: <https://www.antaranews.com/berita/472511/angka-kelahiran-pada-remaja-masih-tinggi> (Accessed: 31 December 2020)
- Setiasih, Setiasih and Srisiuni, Srisiuni and Wahyuningsih, Sri and Setyaningrum, Idfi and Pudjibudojo, Jatie Kusmiati Kusna and Karunia, Nurlita Endah. (2013). Faktor-Faktor yang Mempengaruhi Wanita Melahirkan pada Umur 15-19 Tahun. Available at: <http://repository.ubaya.ac.id/21415/> (Accessed: 20 May 2021)
- Wahyuni, Widya. (2011). Penaksiran Parameter Model Regresi Binomial Negatif pada Kasus Overdispersi. Unpublished manuscript, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia, Depok. [online]. Available at: <http://lib.ui.ac.id/file?file=digital/20228266-S209-Penaksiran%20parameter.pdf>
- Zahab, H., Dharmawan, Y. & Winarni, S. (2017) 'Hubungan Antara Perilaku Pacaran Remaja dan Pernikahan di Bawah Usia 20 Tahun Terhadap Angka Kelahiran Menurut Kelompok Umur 15-19 Tahun di Indonesia' Jurnal Kesehatan Masyarakat (Undip) 5(3) p56-65 [online]. Available at: <https://ejournal3.undip.ac.id/index.php/jkm/article/view/17167/16437> (Accessed: 20 May 2021)