

Workshop Analisis Regresi Logistik untuk Penelitian di bidang Ilmu Sosial dan Pendidikan

(Workshop on Logistic Regression Analysis for Research in Social Sciences and Education)

**Dhoriva Urwatul Wutsqa, Kismiantini, Rosita Kusumawati, Syarifah Inayati,
Sahid, Ezra Putranda Setiawan**

**Jurusan Pendidikan Matematika, Fakultas MIPA, Universitas Negeri Yogyakarta*

Abstrak

Penelitian-penelitian di bidang ilmu sosial dan pendidikan seringkali melibatkan variabel-variabel respon dengan tipe kategorik, yang memerlukan analisis menggunakan regresi logistik. Dibandingkan regresi linear, model regresi logistik lebih kompleks dalam hal pengolahan maupun interpretasinya. Oleh karena itulah, dilakukan kegiatan Pengabdian kepada Masyarakat dalam bentuk workshop analisis regresi logistik untuk penelitian di bidang ilmu sosial dan pendidikan. Kegiatan ini diikuti oleh praktisi lulusan S1 dan mahasiswa berbagai program studi pascasarjana di Indonesia. Workshop dilaksanakan secara daring selama dua hari dengan metode pemberian materi dan demo program R secara langsung. Pelatihan ini diawali dengan visualisasi data kategorik lalu dilanjutkan materi regresi logistik biner dengan prediktor kontinu pada hari pertama. Pada hari kedua, materi yang disampaikan adalah regresi logistik biner dengan prediktor kategorik dan regresi logistik dengan prediktor kategorik dan kontinu. Data yang digunakan sebagai contoh adalah data dalam penelitian ilmu sosial dan pendidikan. Berdasarkan hasil angket, pengamatan dan tanya jawab dengan peserta pelatihan, tampak bahwa peserta bersemangat mengikuti kegiatan pelatihan ini. Peserta dapat menggunakan perintah-perintah analisis regresi logistik untuk data penelitian ilmu sosial dan pendidikan serta dapat memberikan interpretasi dari *output* program R secara tepat.

Kata kunci: regresi logistik, data kategorik, riset sosial, riset pendidikan.

Abstract

Studies in social science and educational science frequently use categorical response variables, which can be analyzed using logistic regression. Compared to linear regression, logistic regression is more complex, especially on the procedure and interpretation. Based on this situation, we held a community service in a workshop on logistic regression analysis for research in social science and education. Participants of this workshop were bachelor graduates and students from many Graduate programs in Indonesia. This event was held online for two consecutive days in explanations and tutorials on using the R program. On the first day, the participants discuss two topics, i.e., visualization of categorical data and binary logistics regression with the continuous predictor. On the next day, the usage of the categorical predictor in the binary logistics regression was introduced. The last session was dedicated to binary logistic regression using both categorical and continuous predictors. From the observation, questionnaire, and discussion with the participants, it can be inferred that almost all participants are enthusiastic about joining this workshop. Participants can use binary logistics regression commands to analyze social and educational data and give a meaningful interpretation of the output of R software.

Key words: logistic regression, categorical data, social research, educational research.

PENDAHULUAN

Data kategorik atau *categorical data* diperoleh jika variabel diukur dengan hanya mengklasifikasikan responden ke dalam beberapa kelompok. Sebagai contoh data responden yang terdiri atas jenis kelamin, suku, tingkat pendidikan, serta motivasi belajar yang dikategorikan dalam rendah, sedang, dan tinggi. Penelitian-penelitian di bidang ilmu sosial, kependidikan seringkali melibatkan variabel-variabel dengan tipe kategorik. Sebagai contoh, penelitian untuk menyelidiki apakah mahasiswa akan lulus tepat waktu atau tidak tepat waktu, apakah siswa akan dapat menyelesaikan tugas dengan baik atau tidak. Hubungan antara anak dengan tingkat kecerdasan (tinggi, sedang, rendah) dengan kepercayaan diri (tinggi, sedang, rendah) juga merupakan bentuk analisis antar variabel kategorik.

Salah satu analisis data statistik yang bertujuan untuk melihat hubungan antara variabel independen dengan dependen adalah analisis regresi. Pada umumnya, analisis regresi yang telah dipelajari pada mata kuliah Statistika Dasar di berbagai program studi adalah analisis regresi linear sederhana dan ganda. Pada analisis regresi ini, variabel respon atau variabel dependennya harus berskala interval atau rasio. Metode ini tidak dapat dipergunakan bila variabel respon yang hendak dianalisis bertipe kategorik.

Analisis regresi logistik adalah analisis regresi dengan variabel dependen atau respon yang melibatkan data kategori. Sebagai contoh, dalam Subekti (2014) regresi logistik multinomial digunakan untuk menentukan pilihan sekolah lanjutan atas pada siswa/i Sekolah Menengah Pertama (SMP). Imaslikah *et al.* (2013) menggunakan regresi logistik ordinal guna menentukan faktor-faktor yang mempengaruhi predikat kelulusan mahasiswa S1 di ITS. Setyaningrum *et al.* (2014) meninjau faktor-faktor yang mempengaruhi perkembangan kognitif pada anak yang mengikuti kegiatan Pendidikan Anak Usia Dini (PAUD). Contoh-contoh tersebut juga menunjukkan adanya beberapa macam regresi logistik, yakni regresi logistik biner, regresi logistik multinomial, serta regresi logistik ordinal.

Secara umum, prosedur analisis regresi logistik secara umum hampir sama dengan analisis regresi linear pada umumnya. Proses dapat dimulai dari pengenalan model, inferensi statistik baik dengan uji hipotesis, interval kepercayaan, dan uji asumsi model. Akan tetapi karena model lebih kompleks, proses analisis regresi logistik juga lebih kompleks dibandingkan dengan analisis regresi linear (Mood, 2010). Cara menginterpretasikan hasil perangkat lunak merupakan kemampuan yang mendukung proses analisis regresi logistik, dan menginterpretasi model menjadi hal yang sangat penting. Pelaporan atau penyajian hasil analisis regresi logistik juga harus dilakukan secara cermat, mengingat adanya variasi pelaporan dalam jurnal-jurnal kependidikan (Peng *et al.*, 2002a; Peng *et al.*, 2002b).

Mengingat manfaat yang sangat besar dari analisis regresi logistik untuk penelitian di bidang ilmu sosial dan kependidikan, maka Program Studi Statistika Universitas Negeri Yogyakarta yang berada di bawah naungan universitas dengan basis kependidikan memiliki tanggung jawab untuk menyebarluaskan ilmu statistika yang bermanfaat untuk analisis data dalam penelitian bidang ilmu sosial dan kependidikan. Terlebih di Universitas Negeri Yogyakarta bahkan di Yogyakarta pada umumnya, tidak sedikit program studi yang berada dalam naungan ilmu pendidikan dan ilmu-ilmu sosial. Oleh karena itu, dalam kegiatan pengabdian kepada masyarakat ini diselenggarakan *workshop* analisis regresi logistik untuk penelitian di bidang ilmu sosial dan pendidikan.

SOLUSI/TEKNOLOGI

Analisis regresi logistik, lebih-lebih yang melibatkan beberapa variabel prediktor, cukup sulit dilakukan secara manual. Oleh karena itu, komputer dengan perangkat lunak statistika yang tepat sangat diperlukan agar prosedur analisis ini dapat dilakukan dengan cepat.

Dalam kegiatan *workshop* analisis regresi logistik untuk penelitian di bidang ilmu sosial dan pendidikan ini dipergunakan perangkat lunak R. Dibandingkan SPSS, perangkat lunak R mungkin kurang dikenal oleh para pengguna statistika di bidang ilmu sosial. Akan tetapi, perangkat lunak

R memiliki beberapa kelebihan, di antaranya dari segi ketersediaan pada berbagai sistem operasi, kebebasan untuk menggunakan sampai waktu yang tak terbatas, serta bersifat publik/GNU sehingga tidak memerlukan pembayaran lisensi (Rosadi, 2015).

Visualisasi Data

Visualisasi data adalah representasi grafis dari data untuk membantu pengguna untuk memahami konteks dan signifikansi dari suatu data. Tujuan utama dari visualisasi data adalah untuk mengkomunikasikan informasi secara jelas dan efisien kepada pengguna lewat grafik informasi yang dipilih, seperti tabel dan grafik. Pada pembahasan ini akan ditinjau beberapa cara membuat visualisasi data dasar, yang meliputi diagram pencar (*scatter plot*), histogram, diagram batang (*bar chart*), diagram kotak-garis (*box plot*), diagram densitas (*density plot*), serta diagram peta.

Perangkat lunak Statistika R menawarkan seperangkat fungsi dan pustaka (*library*) untuk membangun visualisasi dan menyajikan data. Beberapa program yang harus diinstalasi pada komputer atau laptop peserta antara lain: (1) Program R, dapat diunduh melalui <https://www.r-project.org/>, (2) program Rtools, diunduh pada situs <https://cran.r-project.org/bin/windows/Rtools/>, dan 3) *RStudio Desktop Open Source License* sebagai tampilan antarmuka yang diunduh pada <https://www.rstudio.com/products/rstudio/download/>. Selanjutnya diinstalasi pula sejumlah pustaka di R, yakni *corrplot*, *tidyverse*, *reshape2*, *purrr*, *readr*, *plotly*, *rwoldmap*, dan *ggplot2*. Fungsi dan penjelasan untuk masing-masing fungsi tersebut dapat dilihat pada situs CRAN.

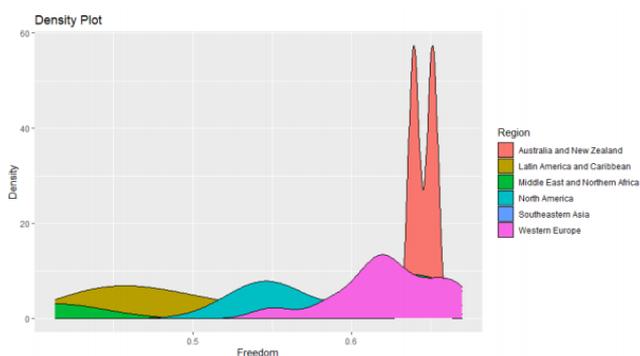
Untuk membuat visualisasi sederhana, dapat dipergunakan sejumlah pustaka pada R sebagaimana ditunjukkan pada Tabel 1. Selanjutnya, visualisasi-visualisasi data yang lebih canggih dapat dibuat menggunakan *ggplot2*, yakni pustaka pada R yang sangat berguna dan fleksibel untuk memproduksi grafik-grafik yang elegan. Paket ini diimplementasikan oleh Hadley Wickham.

Tabel 1. Macam-macam visualisasi data dalam R dan perintah untuk membuatnya

Jenis diagram	Perintah R
Diagram pencar (<i>scatter plot</i>)	plot()
Matriks diagram pencar (<i>scatter plot matrix</i>)	pairs()
Diagram kotak garis (<i>box plot</i>)	boxplot()
Diagram baris (<i>strip chart</i>)	stripchart()
Histogram	hist()
Diagram densitas (<i>density plot</i>)	density()
Diagram batang (<i>bar chart</i>)	barplot()
Diagram garis (<i>line plot</i>)	line() atau plot()
Diagram lingkaran (<i>pie chart</i>)	pie()
Diagram titik	dotchart()

Selanjutnya, visualisasi-visualisasi data yang lebih canggih dapat dibuat menggunakan *ggplot2*, yakni pustaka pada R yang sangat berguna dan fleksibel untuk memproduksi grafik-grafik yang elegan. Paket ini diimplementasikan oleh Hadley Wickham. Beberapa jenis visualisasi data yang dapat dibuat menggunakan pustaka ini adalah *box plot*, *violin plot*, *dot plot*, *stripchart*, *density plot*, *histogram plot*, *scatter plot*, *bar plot*, *line plot*, *pie chart*, dan *Q-Q plot*.

Konsep dari *ggplot2* adalah membagi plot dalam tiga bagian fundamental berbeda, yakni data, *aesthetic*, dan geometri. Data adalah suatu *data frame* yang akan divisualisasikan, sedangkan *aesthetic* digunakan untuk mengidentifikasi variabel independen (x) dan variabel dependen (y). Komponen ini juga dapat untuk mengontrol warna, ukuran, dan bentuk titik-titik, panjang *bar*, dan lain sebagainya. Bagian terakhir yakni *geometry* mendefinisikan tipe grafik (*histogram*, *box plot*, *line plot*, *density plot*, *dot plot*, dan sebagainya), titik (*geom_point*, digunakan untuk diagram pencar, dsb), garis (*geom_line*, digunakan pada data runtun waktu, regresi, dsb), serta *boxplot* (*geom_boxplot*, digunakan untuk *boxplot*). Pada Gambar 1 ditunjukkan visualisasi data yang dihasilkan dari fungsi *ggplot2*.



Gambar 1. Contoh penggunaan pustaka ggplot2 dalam visualisasi data berupa diagram densitas (*density plot*)

Regresi Logistik dengan Prediktor Kontinu

Model regresi logistik merupakan model regresi untuk respon atau variabel dependen biner. Biasanya respon biner ini dikategorikan sebagai sukses vs. gagal. Selanjutnya, peluang sukses dinotasikan sebagai $P(Y_i = 1)$, sedangkan peluang gagal dinotasikan sebagai $P(Y_i = 0)$.

Model regresi logistik merupakan logit dari peluang sukses, yakni logaritma dari *odds*, dan berbentuk linear.

$$\begin{aligned} \text{logit} [P(Y_i = 1)] &= \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} \\ &= \beta_0 + \beta \mathbf{X} \end{aligned}$$

atau

$$\begin{aligned} \text{logit}(\pi) &= \beta_0 + \beta \mathbf{X} \\ \ln\left(\frac{\pi}{1-\pi}\right) &= \beta_0 + \beta \mathbf{X} \end{aligned}$$

Dengan demikian diperoleh hubungan

$$\pi = P(Y_i = 1) = \frac{\text{odds}}{1 + \text{odds}} = \frac{\exp(\beta_0 + \beta \mathbf{X})}{1 + \exp(\beta_0 + \beta \mathbf{X})}$$

Pendugaan parameter pada model regresi logistik dilakukan dengan metode kemungkinan maksimum (*maximum likelihood*). Nilai dugaan parameter dapat diperoleh dengan pendekatan numerik. Selanjutnya, ketepatan model dapat diperiksa menggunakan uji Hosmer-Lemeshow. Inferensi atau uji signifikansi masing-masing variabel independen pada model regresi logistik dapat dilakukan dengan dua cara, yakni uji rasio Likelihood dan uji Wald.

Pada uji rasio likelihood, dipergunakan dua macam model yakni *non-restricted model* yang melibatkan semua variabel independen dan *restricted model* yang mempergunakan beberapa variabel independen saja. Hipotesis nol (H_0) yang menyatakan bahwa semua koefisien pada *non restricted model* yang tidak ada pada *restricted model* bernilai nol diuji dengan statistik

$$G^2 = -2 \log \left(\frac{L_0}{L_1} \right)$$

dengan L_0 dan L_1 berturut-turut menyatakan nilai *likelihood* pada *restricted* dan *non restricted model*. Diketahui bahwa statistik G^2 berdistribusi chi kuadrat.

Uji Wald sangat mirip dengan uji parsial t pada model regresi linear. Pada uji ini, hipotesis nol yang digunakan adalah nilai koefisien $\beta = 0$. Statistik uji yang dipergunakan adalah z yang berdistribusi normal baku. Bila ukuran sampel besar, dapat dihitung statistik uji z^2 yang berdistribusi chi kuadrat.

Interpretasi hasil estimasi koefisien model regresi logistik dengan prediktor kontinu dapat dilakukan sebagai berikut. Misalkan b adalah hasil estimasi koefisien model regresi logistik pada prediktor X . Pada nilai $Y = 1$, kenaikan satu unit pada nilai prediktor X akan menaikkan nilai *odds* sebesar e^b kali. Sebagai contoh, misal suatu model regresi logistik dibentuk untuk menilai keberhasilan siswa menempuh ujian Fisika Dasar ($Y = 1$ untuk lulus, $Y = 0$ untuk gagal), dengan prediktor X adalah nilai matematika. Bila didapat koefisien sebesar 0,65, dapat dikatakan bahwa untuk setiap kenaikan nilai matematika sebesar 1 poin, *odds* siswa tersebut untuk lulus ujian Fisika akan meningkat sebesar 65%.

Estimasi *intercept* atau konstanta pada regresi logistik juga harus dilakukan dengan hati-hati. Secara umum, konstanta dapat dipahami sebagai probabilitas suatu subjek dengan nilai variabel independen nol memperoleh hasil $Y = 1$. Dalam contoh model regresi logistik di atas, misal diperoleh estimasi konstanta sebesar $-1/2$, maka peluang bahwa seseorang yang memiliki nilai matematika sebesar 0 lulus ujian Fisika adalah sebesar

$$\pi = \frac{\exp(-0.5 + 0)}{1 + \exp(-0.5 + 0)} = 0.378.$$

Terlihat bahwa interpretasi konstanta maupun koefisien pada regresi logistik tidaklah sama dengan interpretasi pada regresi linear.

Regresi logistik juga dapat dipandang sebagai salah satu kejadian khusus dari model linear umum (*general linear model*). Oleh karena itulah prosedur regresi logistik pada perangkat lunak R dapat dilakukan dengan perintah **glm** pada pustaka **stats**, dengan menambahkan argumen **family = "binomial"**. Objek yang dihasilkan dari perintah tersebut memuat sejumlah elemen, yakni nilai-nilai hasil estimasi koefisien berikut standard error, dan statistik uji Wald serta nilai signifikansinya. Terlihat pula nilai residual untuk *deviance*.

Regresi Logistik dengan Prediktor Kategorik

Dalam suatu penelitian, hubungan antara satu atau dua variabel prediktor kategorik dengan variabel respon kategorik seringkali dianalisis dengan pembuatan tabel kontingensi yang diikuti oleh uji chi kuadrat. Analisis semacam ini juga dapat dilakukan dengan menggunakan metode regresi logistik. Kelebihannya, model regresi logistik dapat mengakomodasi adanya prediktor dengan beberapa kategori, serta mengakomodasi adanya lebih dari satu variabel prediktor. Adanya interaksi antar variabel prediktor juga dapat diuji dengan mudah pada regresi logistik ini. Seperti pada analisis dengan tabel kontingensi, nilai rasio *odds* beserta selang kepercayaannya juga dapat dihasilkan melalui regresi logistik.

Sebagai contoh, dimiliki data preferensi politik sejumlah sampel warga Amerika Serikat berupa dukungan terhadap calon presiden dari Partai Republik atau Partai Demokrat. Ingin diketahui apakah preferensi politik tersebut dipengaruhi oleh sejumlah faktor, di antaranya jenis kelamin (variabel kategorik dengan dua nilai yakni pria dan wanita) serta tingkat penghasilan (variabel kategorik dengan lima nilai dari sangat rendah hingga sangat tinggi). Perangkat lunak R dapat digunakan untuk menghitung nilai dugaan parameter, nilai dugaan rasio *odds* dan selang konfidensinya, serta statistik Hosmer-Lemeshow untuk uji hipotesis kecocokan model regresi logistik tersebut.

Regresi Logistik dengan Prediktor Kontinu dan Kategorik

Dalam praktis penelitian, regresi logistik yang digunakan umumnya memiliki sejumlah prediktor baik kategori maupun kontinu. Sebagai contoh, penelitian Wang (2015) di Taiwan menunjukkan bahwa orang-orang asli dari pegunungan, wanita, berusia lanjut, lebih sehat, lebih sejahtera, menempuh pendidikan tinggi, serta menghidupi nilai-nilai Barat cenderung lebih bahagia. Riset Lu *et al.* (2016) di Provinsi Guangdong, RR China, menunjukkan bahwa kepuasan kerja pada staf kesehatan di wilayah tersebut dipengaruhi oleh sejumlah variabel, di antaranya pekerjaan, tingkat pendidikan, status profesional, lama bekerja, penghasilan tahunan, serta frekuensi shift malam.

Salah satu data riset yang tersedia luas untuk dikaji dengan regresi logistik adalah dari *The Indonesian Family Life Survey* (IFLS). Pada data tersebut terdapat ukuran tingkat kebahagiaan (*happiness*), yang dapat direduksi menjadi dua pilihan yakni 1: bahagia, dan 0: tidak bahagia. Beberapa prediktor yang dapat dimasukkan pada model adalah jenis kelamin, usia, pendidikan tertinggi, status perkawinan, status pekerjaan, agama, daerah pedesaan/perkotaan, beserta status kesehatan yang diukur sendiri. Dengan adanya sejumlah prediktor kontinu, diperlukan sejumlah uji diagnostik pada model, di antaranya asumsi hubungan linear antara prediktor kontinu dan logit dari variabel respon, tidak adanya multi kolinearitas, serta tidak adanya *influential point* atau nilai ekstrem yang sangat berpengaruh. Ketiga asumsi tersebut berturut-turut dapat diuji dengan plot, nilai *variance inflation factors* (VIF), serta nilai *standardized residuals*.

HASIL DAN DISKUSI

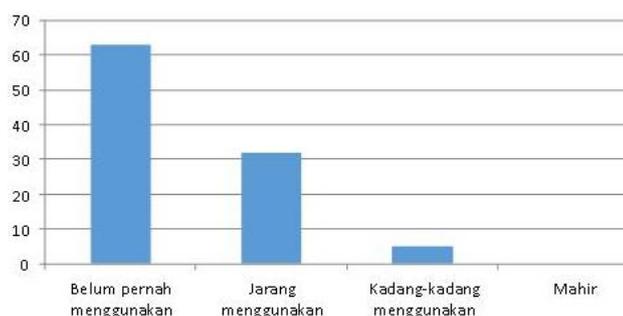
Rangkaian pelaksanaan Pengabdian kepada Masyarakat (PPM) ini dilakukan melalui beberapa tahap, di antaranya (1) koordinasi tanggal dan bentuk pelaksanaan kegiatan PPM (2) penyiapan materi dan demo program, (3) pendaftaran peserta, serta (4) pelaksanaan PPM. Berikut akan dipaparkan hasil kegiatan (1) hingga (3) dalam subbagian ‘pra pelaksanaan’, diikuti penjelasan lengkap tentang pelaksanaan kegiatan.

Pra Pelaksanaan

Pada tahap penyiapan materi dan demo program, dilakukan pembagian narasumber pada masing-masing sesi. Materi yang telah disusun dalam bentuk *handout* dan bahan demo program untuk masing-masing sesi kemudian didiskusikan bersama agar tercapai kesinambungan. Seluruh *handout* dan file data yang diperlukan untuk mencoba program diletakkan pada folder *google drive* agar dapat diakses langsung oleh peserta pada saat pelaksanaan *workshop*.

Pada tahap pendaftaran peserta, dilakukan publikasi kegiatan *workshop* dengan media poster yang dikirimkan ke sejumlah media sosial. Pada tahap ini diperoleh peminat *workshop* sebanyak 100 orang, yang terdiri dari 53 orang mahasiswa Pascasarjana (S2/S3), 27 mahasiswa program Sarjana/Diploma, 18 orang guru/dosen yang tidak sedang menempuh studi, serta 2 orang profesi lainnya. Menarik pula dicermati bahwa peminat kegiatan *workshop* berasal dari berbagai daerah di Indonesia, termasuk di antaranya masing-masing satu pendaftar dari Universitas PGRI NTT di Kupang, Institut Pendidikan Tapanuli Selatan di Padangsidimpuan, dan Universitas Mataram.

Tim juga melakukan survei sederhana sejauh mana peserta telah mengenal program R. Hasil survei ini ditunjukkan pada Gambar 2.



Gambar 2. Kemampuan awal peserta *Workshop* dalam menggunakan perangkat lunak R (sumber : olahan dari formulir pendaftaran)

Berdasarkan Gambar 2, terlihat bahwa lebih dari 50% peserta *workshop* belum pernah menggunakan perangkat lunak R. Oleh karena itu tim mengirimkan video terkait pengenalan dan instalasi program R dua hari sebelum pelaksanaan kegiatan *workshop*. Peserta dapat mempelajari video tersebut agar memiliki gambaran terkait

perangkat lunak R yang akan dipergunakan pada *workshop*. Tim memanfaatkan informasi ini untuk menyusun urutan materi *workshop* dan menguji kedalaman materi yang telah dipersiapkan.

Pelaksanaan

Kegiatan *Workshop* analisis regresi logistik untuk penelitian ilmu sosial dan pendidikan dilaksanakan secara daring melalui *zoom* pada hari Jumat dan Sabtu tanggal 6-7 November 2020. Rangkaian *workshop* dilakukan selama 16 jam yang terdiri atas kegiatan satu sesi pengenalan instalasi program R, 4 sesi materi utama, dan 2 sesi tugas mandiri. Sesi pertama (2 jam) diisi dengan pemaparan materi terkait visualisasi data kategorik menggunakan Program R. Sesi kedua (2 jam) diisi dengan pemaparan materi terkait regresi logistik biner dengan prediktor kontinu menggunakan Program R. Sesi ketiga (1,5 jam) diisi dengan pemaparan materi terkait dengan analisis regresi logistik biner dengan prediktor kategorik menggunakan Program R. Sesi keempat (2 jam), diisi dengan pemaparan materi regresi logistik biner dengan prediktor kategorik dan kontinu menggunakan Program R. Peserta *workshop* diberikan tugas mandiri untuk menyelesaikan masalah analisis regresi logistik pada data penelitian ilmu sosial dan pendidikan menggunakan program R, kemudian peserta diminta mengirimkan hasilnya.

Kegiatan *workshop* diikuti oleh para praktisi lulusan S1, mahasiswa pasca sarjana, dan beberapa dosen di lingkungan universitas di Yogyakarta dan sekitarnya. Pembicara dalam pelatihan ini adalah semua tim PPM. Mahasiswa sebagai anggota tim dilibatkan dalam teknis pelaksanaan kegiatan.

Metode yang digunakan dalam *workshop* adalah tutorial dan demo. Pada hari pertama, pelatihan ini diawali dengan pengertian data kategorik, visualisasi data dengan program R menggunakan grafik plot pencar, histogram, bar chart, boxplot, density plot, dan visualisasi peta. Paket-paket Program R yang diperlukan untuk visualisasi data kategorik meliputi *corrplot*, *tidyverse*, *reshape2*, *purrr*, *readr*, *plotly*, *ggplot2*, dan *rworldmap*. Setelah sesi pertama selesai, dilanjutkan sesi kedua yaitu pemaparan materi

terkait dengan regresi logistik biner dengan prediktor kontinu menggunakan Program R dan dilanjutkan demo kepada para peserta. Pada sesi kedua hari pertama disampaikan definisi model regresi logistik, estimasi parameter dengan metode maksimum likelihood, interpretasi parameter, inferensi parameter dengan uji hipotesis, model fit, dan contoh aplikasinya menggunakan program R. Selanjutnya pada hari kedua sesi pertama disampaikan materi regresi logistik biner dengan prediktor kategorik meliputi pengenalan model dengan satu prediktor kategorik, satu prediktor kategorik lebih dari dua level, dan dua prediktor kategorik dan interaksi serta contoh aplikasinya dengan program R. Pada sesi kedua hari kedua dilanjutkan dengan penyampaian materi regresi logistik biner dengan prediktor kategorik dan kontinu meliputi model regresi logistik biner, pengujian signifikansi parameter tunggal, pengaruh marjinal, aplikasi regresi logistik biner pada penelitian ilmu sosial dan pendidikan, dan studi kasus kebahagiaan dari

data survei Indonesian Family Life Survey gelombang 5 (IFLS5) menggunakan regresi logistik biner dengan program R. Pada setiap sesi dilakukan tanya jawab dengan para peserta.

Evaluasi

Selama kegiatan workshop, peserta yang hadir tampak bersemangat mengikuti kegiatan workshop. Peserta dapat menggunakan perintah-perintah analisis visualisasi data kategorik dan analisis regresi logistik biner untuk data penelitian ilmu sosial dan pendidikan serta dapat memberikan interpretasi dari output yang diperoleh dari program R. Peserta terlihat cukup baik dalam mengikuti workshop baik pada sesi teori maupun demo program R. Indikasi ini terlihat dari perhatian, pertanyaan yang mereka sampaikan pada setiap sesi workshop. Dengan demikian secara keseluruhan dapat disimpulkan bahwa acara workshop telah berjalan dengan lancar dan target acara telah tercapai.

Tabel 3. Kemampuan peserta setelah mengikuti kegiatan *workshop*

No	Kemampuan	Respon peserta (%)			
		SS	S	T	STS
1	Memahami cara menginstal program R dan Rstudio	42,9	57,1		
2	Memahami manfaat mempelajari program R	50	50		
3	Memahami cara kerja program R	28,6	71,4		
4	Mencari dan menambahkan <i>packages</i> yang dibutuhkan	28,6	71,4		
5	Memasukkan data ke dalam program R dan membaca file	35,7	64,3		
6	Mengenal tipe-tipe data dan menyusun variabel kategorik dalam R	28,6	71,4		
7	Membuat plot pencar dan melakukan penyesuaian argumen fungsi	28,6	71,4		
8	Membuat plot pencar dengan ggplot dan melakukan penyesuaian	28,6	71,4		
9	Membuat bar chart dengan barplot dan melakukan penyesuaian	35,7	64,3		
10	Membuat box plot dengan ggplot dan menggambar peta dengan package <i>rworldmap</i>	28,6	71,4		
11	Menyusun model regresi logistik biner dengan prediktor kontinu	7,1	92,9		
12	Menyusun model regresi logistik biner dengan prediktor kategorik	7,1	92,9		
13	Menyusun model regresi logistik biner dengan variabel prediktor kategorik serta campuran prediktor kontinu dan kategorik	7,1	92,9		
14	Melakukan uji asumsi regresi logistik biner dan menginterpretasi terhadap koefisien model regresi	7,1	92,9		
15	Menghitung nilai <i>odds ratio</i> dan menginterpretasi hasil yang diperoleh	7,1	92,9		
	Rata-rata persentase	24,8	75,2		

Menjelang penutupan kegiatan, dilakukan evaluasi menggunakan google form secara daring. Berdasarkan hasil evaluasi yang ditampilkan pada Tabel 3, terlihat bahwa sebagian besar peserta merasa memiliki kemampuan dalam melakukan hal-hal yang telah diberikan dalam workshop selama dua hari tersebut. Terkait dengan materi visualisasi data, lebih dari 25% peserta sangat setuju dengan pernyataan bahwa mereka telah memiliki kemampuan untuk melakukan berbagai bentuk visualisasi data. Untuk materi terkait model regresi logistik, odds ratio, dan cara interpretasinya, terlihat bahwa seluruh peserta telah setuju dengan pernyataan bahwa mereka mampu menyusun model-model regresi logistik biner tersebut, melakukan uji asumsi yang diperlukan, serta menginterpretasi hasilnya.

KESIMPULAN

Melalui kegiatan Pengabdian kepada Masyarakat dalam bentuk workshop ini, peserta mendapat pengetahuan dan keterampilan analisis regresi logistik untuk penelitian di bidang ilmu sosial dan pendidikan menggunakan program R. Seluruh materi ini mendukung peserta untuk mempelajari dan menggunakan analisis regresi logistik yang lebih kompleks, misalnya dengan lebih dari dua kelas variabel respon. Di samping itu, kegiatan pengabdian ini juga meningkatkan kesadaran penggunaan perangkat lunak tak berbayar untuk analisis data. Tim Pengabdian kepada Masyarakat (PPM) juga telah menyusun handout visualisasi data kategorik, regresi logistik biner dengan prediktor kontinu, regresi logistik biner dengan prediktor kategorik, serta regresi logistik biner dengan prediktor kategorik dan kontinu dengan menggunakan program R.

UCAPAN TERIMA KASIH

Terima kasih kami haturkan kepada Fakultas MIPA Universitas Negeri Yogyakarta yang mendanai kegiatan PPM ini, juga kepada UPT Pusat Komputer UNY yang telah memberikan fasilitas akses Zoom untuk kegiatan ini.

DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis*, 2nd edition. New York: John Wiley & Sons.
- Imaslihkah, S., Ratna, M., & Ratnasari, V. (2013). Analisis Regresi Logistik Ordinal terhadap Faktor-faktor yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya. *Jurnal Sains dan Seni ITS*, 2(2), 177-182.
- Lu, Y., Hu, X-M., Huang, X.L. et al. (2016). Job satisfaction and associated factors among healthcare staff : a cross-sectional study in Guangdong Provincen, China. *BMJ Open*, (6) : e011388.
- Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1), 67-82
- Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002a). An introduction to logistic regression analysis and reporting. *The journal of Educational Research*, 96(1), 3-14.
- Peng, C. Y. J., So, T. S. H., Stage, F. K., & John, E. P. S. (2002b). The use and interpretation of logistic regression in higher education journals: 1988–1999. *Research in Higher Education*, 43(3), 259-293.
- Rosadi, D. (2015). *Analisis Statistika dengan R*. Yogyakarta : Gadjah Mada University Press.
- Setyaningrum, S. R., Triyanti, T., & Indrawani, Y. M. (2014). Pembelajaran di Pendidikan Anak Usia Dini dengan perkembangan kognitif pada anak. *Kesmas: National Public Health Journal*, 243-249.
- Subekti, P. (2014). Model regresi logistik multinomial untuk menentukan pilihan sekolah lanjutan tingkat atas pada siswa SMP. *Cauchy*, 3(2), 91-98.
- Wang, J.H. (2015) Happiness and Social Exclusion of Indigeneous Peoples in Taiwan – A social sustainability perspective. *PLoS ONE*, 10(2), e0118305.