

LAPORAN DISKUSI
REGRESI DENGAN VARIABEL RESPON KUALITATIF



Disusun oleh :

Christi Novelina	(2106725021)
Nadia Sukesi Sianipar	(2106700776)
Rachel Thyffani Margaretha S	(2106726900)
Selvira Salsabila	(2106656232)

Ekonometri (A)

DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS INDONESIA
TAHUN AJARAN 2023/2024

MODEL LPM

Linear Probability Model (LPM) adalah sebuah metode statistik yang digunakan untuk menganalisis hubungan antara variabel independen (biasanya variabel-variabel kategorik) dan probabilitas terjadinya suatu peristiwa biner atau dua kategori yang saling eksklusif (misalnya, ya/tidak, sukses/gagal, 1/0).

LPM mengasumsikan bahwa probabilitas bersifat linear terhadap variabel penjelas, serta dalam hal variabel respon yang berupa kualitatif (kategori) dianggap sebagai variabel dummy.

Model Sederhana LPM :

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$(Y_i) = \beta_0 + \beta_1 X_i \quad ; \quad E(u_i) = 0$$

Keterangan :

X_i = nilai variabel independen

$Y = 1$ jika terjadi peristiwa positif (Cth : memiliki rumah)

$Y = 0$ jika terjadi peristiwa negatif (Cth : tidak memiliki rumah)

β_0 = Intercept (Konstanta)

β_1 = Koefisien yang mengukur pengaruh variabel independen

Misal $p_i = P(Y = 1)$, $1 - p_i = P(Y = 0)$. Sehingga $Y_i \sim \text{Bernoulli}(p_i)$

Model LPM memiliki karakteristik yang mirip dengan model regresi linear, sehingga metode OLS dapat digunakan pada model LPM ini. Linear Probability Model (LPM) adalah model yang sederhana. Namun model ini memiliki kelemahan, diantaranya adalah sebagai berikut :

1. Prediksi probabilitas di luar rentang $[0,1]$, yang tidak masuk akal.
2. Residual tidak berdistribusi normal, karena mengikuti distribusi binomial (distribusi Bernoulli). Sebenarnya kelemahan ini tidak begitu bermasalah, karena akan menghasilkan estimator yang BLUE. Apabila datanya semakin banyak, distribusinya juga akan mendekati distribusi normal.
3. Varian residual mudah bersifat heteroskedastis, karena e_i berdistribusi binomial. Apabila varian residual tersebut bersifat heteroskedastis, maka estimatornya tidak lagi bersifat BLUE. Untuk menghilangkan masalah ini, dapat diterapkan analisis regresi dengan metode WLS (*weighted least square*)

4. Nilai koefisien determinasi (R^2) tidak lagi mampu menjelaskan kesesuaian garis regresi dengan datanya. R^2 yang umumnya lebih rendah.

Solusi untuk mengatasi kelemahan model LPM adalah :

1. Menambah ukuran sampel
2. Menggunakan WLS, kuadrat terkecil, atau teknik pemrograman matematika.

Model LPM tidak bisa dilakukan pada kasus yang nonlinier maka perlu model lain :

1. Model Probit
2. Model Logit

MODEL LOGIT

- Model logit adalah model regresi non-linear yang menghasilkan sebuah persamaan dimana dimana variabel dependen bersifat kategorikal dengan menggunakan fungsi logistik kumulatif (F)
- Kategori paling dasar dari model tersebut menghasilkan *binary values* seperti angka 0 dan 1.
- Persamaan umum :

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

logit(p) adalah logaritma odds dari peristiwa sukses dengan p adalah probabilitas peristiwa sukses

- Angka yang dihasilkan mewakili suatu kategori tertentu yang dihasilkan dari perhitungan probabilitas terjadinya kategori tersebut.
- Bentuk dasar probabilitas

Y_i	Probabilitas
0	$1 - P_i$
1	P_i
Total	1

- Biasa digunakan dalam data klasifikasi (contoh : kategori kepemilikan rumah)
- Variabel independen dapat bersifat nominal, ordinal, interval, ratio
- Persamaan probabilitas untuk model logit :

$$P_i = E(Y = 1)|X_i = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}$$

- Probabilitas untuk kategori 0 (yang bersifat negatif),

$$1 - P_i = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}}$$

- Probabilitas untuk kategori 1 (yang bersifat positif),

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i}$$

- Odds Ratio (Rasio kecenderungan) terjadinya kategori dengan nilai 1.

$$\frac{P_i}{1 - P_i}$$

- Persamaan Logaritma natural terhadap *odds ratio*

$$L_i = \ln\left(\frac{p_i}{1-p_i}\right) = Z_i = \beta_1 + \beta_2 X_i$$

Dengan ketentuan :

L_i = log dari *odds ratio* yang linear terhadap X dan parameter β

- Kelebihan :
 - Cocok untuk variabel dependen biner (memiliki dua kategori seperti ya/tidak)
 - Mengatasi asimetri dalam data
 - Memprediksi probabilitas suatu kejadian
 - Dapat menangani variabel prediktor berskala interval atau rasio
 - Mengidentifikasi faktor penting
 - Memungkinkan analisis kausasi
 - Dapat menangani data yang tidak seimbang (data yang memiliki jumlah observasi berbeda untuk kelas berbeda)

- Kekurangan :
 - Adanya asumsi linearitas dan independensi pengamatan pada data
 - Sulit menangani variabel kategori dengan banyak kategori
 - *Overfitting*
 - Estimasi tidak stabil bila kekurangan informasi
 - Tidak dapat menangani data hilang secara langsung
 - Tidak dapat menangani outlier yang signifikan dan banyak *missing value*
 - Interpretasi koefisien sulit karena dikonversi menjadi *odds ratio*

MODEL PROBIT

- Model probit adalah tipe regresi dalam statistik jika hanya ada dua pilihan, misalnya ya atau tidak
- Model probit menggunakan fungsi CDF atau disebut juga model normit
- Level threshold adalah nilai yang digunakan untuk membedakan antara dua atau lebih kategori atau tingkat dalam suatu variabel atau kondisi.
- Contoh : keputusan keluarga ke-i untuk memiliki rumah atau tidak tergantung pada utilitas I_i yang tidak dapat diobservasi dilakukan oleh variabel penjelas

Persamaan : $I_i = \beta_1 + \beta_2 X_i$ dengan X_i = pendapatan dari keluarga ke-i

Semakin besar nilai indeks I_i , semakin besar probabilitas keluarga memiliki rumah

$Y = 1$ (memiliki rumah), $Y = 0$ (tidak memiliki rumah)

Level threshold dari index (I_i^*) :

$I_i > I_i^* \Rightarrow$ keluarga akan memiliki rumah

$I_i < I_i^* \Rightarrow$ keluarga tidak akan memiliki rumah

$I_i = I_i^* \Rightarrow$ tidak dapat diobservasi

Probabilitas dari $I_i^* \leq 1$, dapat dihitung dari standarisasi CDF normal :

$$P_i = P(Y = 1|X) = P(I_i^* \leq I_i) = P(Z_i \leq \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i)$$

Keterangan :

$P(Y = 1|X)$ = Probabilitas suatu peristiwa terjadi memberikan nilai variabel X

Z_i = Variabel standard normal

F = CDF normal standard

$$\begin{aligned} F(I_i) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_i} e^{-z^2/2} dz \end{aligned}$$

Dengan menggunakan invers dari rumus P_i dapat diperoleh informasi dari I_i , β_1 dan β_2

$$I_i = F^{-1}(P_i) \Rightarrow F^{-1}(P_i) \Rightarrow \beta_1 + \beta_2 X_i$$

- Kelebihan Model Probit :
 1. Model probit dapat menangani data biner dengan baik, sementara model regresi lainnya lebih cocok untuk data kontinu.
 2. Model probit dapat dengan mudah diinterpretasikan karena hasilnya dapat dinyatakan dalam probabilitas.
 3. Model probit dapat dengan mudah diestimasi dengan menggunakan metode maximum likelihood

4. Model probit dapat dengan mudah diuji dengan menggunakan uji statistik, seperti uji Wald dan uji Likelihood Ratio, yang merupakan uji yang umum digunakan dalam analisis regresi.
- Kekurangan Model Probit :
 1. Model probit tidak dapat menangani variabel kategori yang memiliki lebih dari 2 kategori.
 2. Model probit tidak dapat menangani data yang memiliki skewness (penyimpangan dari simetris). Model ini diasumsikan memiliki distribusi Normal, sehingga tidak cocok untuk data yang tidak terdistribusi secara Normal.
 3. Model probit tidak dapat menangani data yang memiliki hubungan yang tidak linear.

PEMODELAN

Data yang akan digunakan oleh kelompok kami adalah data dari **Badan Pusat Statistik (BPS) Kota Depok** mengenai **Meningkatnya Jumlah Fasilitas Kesehatan Berdasarkan Kecamatan di Kota Depok**. Tujuan pemodelan LPM (Linear Probability Model), Logit, dan Probit adalah untuk menganalisis hubungan antara satu atau lebih variabel independen dengan variabel dependen yang bersifat biner (0 dan 1). Kota Depok memiliki 11 kecamatan, yaitu:

1. Kecamatan Sawangan
2. Kecamatan Bojongsari
3. Kecamatan Pancoran Mas
4. Kecamatan Cipayung
5. Kecamatan Sukmajaya
6. Kecamatan Cilodong
7. Kecamatan Cimanggis
8. Kecamatan Tapos
9. Kecamatan Beji
10. Kecamatan Limo
11. Kecamatan Cinere

Berikut ini adalah variabel-variabel yang akan digunakan oleh kelompok kami untuk dimodelkan menggunakan model LPM, Logit, dan Probit.

1. Rumah Sakit Umum tahun 2021 (RSU21), yaitu jumlah rumah sakit umum di Kota Depok berdasarkan kecamatan pada tahun 2021.
2. Rumah Sakit Umum tahun 2022 (RSU22), yaitu jumlah rumah sakit umum di Kota Depok berdasarkan kecamatan pada tahun 2022.
3. Rumah Sakit Khusus tahun 2021 (RSK21) yaitu jumlah rumah sakit khusus di Kota Depok berdasarkan kecamatan pada tahun 2021.
4. Rumah Sakit Khusus tahun 2022 (RSK22) yaitu jumlah rumah sakit khusus di Kota Depok berdasarkan kecamatan pada tahun 2022.
5. Puskesmas Rawat Inap tahun 2021 (PRI21), yaitu jumlah puskesmas rawat inap di Kota Depok berdasarkan kecamatan pada tahun 2021.
6. Puskesmas Rawat Inap tahun 2022 (PRI22), yaitu jumlah puskesmas rawat inap di Kota Depok berdasarkan kecamatan pada tahun 2022.
7. Puskesmas Non Rawat Inap tahun 2021 (PNRI21), yaitu jumlah puskesmas non rawat inap di Kota Depok berdasarkan kecamatan pada tahun 2021.
8. Puskesmas Non Rawat Inap tahun 2022 (PNRI22), yaitu jumlah puskesmas non rawat inap di Kota Depok berdasarkan kecamatan pada tahun 2022.
9. Klinik Pratama tahun 2021 (KP21), yaitu jumlah klinik pratama di Kota Depok berdasarkan kecamatan pada tahun 2021.
10. Klinik Pratama tahun 2022 (KP22), yaitu jumlah klinik pratama di Kota Depok berdasarkan kecamatan pada tahun 2022.

11. Posyandu tahun 2021 (PRI21), yaitu jumlah posyandu di Kota Depok berdasarkan kecamatan pada tahun 2021.
12. Posyandu tahun 2022 (PRI22), yaitu jumlah posyandu di Kota Depok berdasarkan kecamatan pada tahun 2022.
13. Jumlah Penduduk tahun 2022 (JP), yaitu jumlah penduduk di Kota Depok berdasarkan kecamatan pada tahun 2022.
14. Dokter (*Doctor*), yaitu jumlah dokter di Kota Depok berdasarkan kecamatan pada tahun 2022.
15. Dokter Gigi (*Dentist*), yaitu jumlah dokter gigi di Kota Depok berdasarkan kecamatan pada tahun 2022.
16. Tenaga Keperawatan (*Nursing*), yaitu jumlah tenaga keperawatan di Kota Depok berdasarkan kecamatan pada tahun 2022.
17. Tenaga Kebidanan (*Midwifey*), yaitu jumlah tenaga kebidanan di Kota Depok berdasarkan kecamatan pada tahun 2022.
18. Tenaga Kefarmasian (*Pharmaceutical*), yaitu jumlah tenaga kefarmasian di Kota Depok berdasarkan kecamatan pada tahun 2022.
19. Tenaga Kesehatan Masyarakat (*Public Health*), yaitu jumlah tenaga kesehatan masyarakat di Kota Depok berdasarkan kecamatan pada tahun 2022.
20. Tenaga Kesehatan Lingkungan (*Environmental Health*), yaitu jumlah tenaga kesehatan lingkungan di Kota Depok berdasarkan kecamatan pada tahun 2022.
21. Tenaga Gizi (*Nutritionist*), yaitu jumlah tenaga gizi di Kota Depok berdasarkan kecamatan pada tahun 2022.
22. Ahli Teknologi Laboratorium Medik (*Medical Laboratory*), yaitu jumlah ahli teknologi laboratorium medik di Kota Depok berdasarkan kecamatan pada tahun 2022.

Hasil dan Pembahasan

Sebelum membangun model Logit, dan Probit, terdapat langkah-langkah yang perlu dilakukan sebagai berikut.

1. Import Dataset

Digunakan fungsi “`read_excel()`” dari *package* “`readxl`” untuk meng-*import* dan membaca dataset yang digunakan.

```
> #Pemanggilan Data
> library(readxl)
> data <- read_excel("C:/Users/M-SPORT/Downloads/Data Fasilitas Kesehatan.xlsx")

> head(data)
# A tibble: 6 x 23
  Kecamatan RSU21 RSU22 RSK21 RSK22 PRI21 PRI22 PNRI21 PNRI22 KP21 KP22 P21 P22 JP Doctor Dentist Nursing Midwifey Pharmaceutical
  <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Sawangan 2 3 0 0 2 0 4 5 10 12 95 97 191700 200 43 356 160 85
2 Bojongsari 1 1 0 0 0 0 2 4 10 13 97 98 113600 79 23 39 44 42
3 Pancoran Mas 5 4 1 1 1 0 3 4 35 35 124 125 231600 92 122 621 166 171
4 Cipayang 0 0 0 0 1 0 1 2 8 8 68 68 181140 46 16 22 58 36
5 Sukmajaya 2 3 0 0 1 0 3 4 28 34 130 131 255960 314 99 451 123 114
6 Cilodong 2 2 0 0 1 0 2 3 20 20 91 92 177640 149 27 163 88 82

# 14 more variables: 'Public Health' <dbl>, 'Environmental Health' <dbl>, 'Nutritionist' <dbl>, 'Medical Laboratory' <dbl>
```

2. Menghitung Jumlah Fasilitas Kesehatan Tahun 2021 dan 2022

Selanjutnya, akan dilakukan perhitungan jumlah fasilitas kesehatan pada tahun 2021 dan 2022. Fasilitas kesehatan meliputi, rumah sakit umum, rumah sakit khusus, puskesmas

rawat inap, puskesmas non rawat inap, klinik pratama, dan posyandu. Dengan bantuan *software* R dan menggunakan fungsi “rowSums()” diperoleh sebagai berikut.

```
> #Menghitung Jumlah Faskes 2021 dan 2022
> faskes21 <- rowSums(data[, c("RSU21", "RSK21", "PRI21", "PNRI21", "KP21", "P21")])
> faskes21
[1] 113 110 169 78 164 116 160 159 115 57 63
> faskes22 <- rowSums(data[, c("RSU22", "RSK22", "PRI22", "PNRI22", "KP22", "P22")])
> faskes22
[1] 117 114 169 78 172 117 168 160 115 59 65
```

Berdasarkan *output* di atas diperoleh informasi bahwa

Kecamatan	Fasilitas Kesehatan 2021	Fasilitas Kesehatan 2022
Sawangan	113	117
Bojongsari	110	114
Pancoran Mas	169	169
Cipayung	78	78
Sukmajaya	164	172
Cilodong	116	117
Cimanggis	160	168
Tapos	159	160
Beji	115	115
Limo	57	59
Cinere	63	65

3. Membuat Variabel *Dummy*

Selanjutnya, akan dibuat variabel *dummy* pada variabel respon, yaitu fasilitas kesehatan 2021 dan fasilitas kesehatan 2022 dengan kategori bahwa apabila terdapat peningkatan jumlah fasilitas kesehatan, maka akan diberi nilai 1 dan apabila tidak terjadi peningkatan jumlah fasilitas kesehatan, maka akan diberi nilai 0. Dengan bantuan *software* R dan menggunakan fungsi “ifelse()” diperoleh sebagai berikut.

```
> #Pembuatan Variabel Binary pada Variabel Respon (faskes21 dan faskes 22)
> faskes <- ifelse(faskes22 > faskes21, 1, 0)
> faskes
[1] 1 1 0 0 1 1 1 1 0 1 1
```

4. Membuat *Dataframe*

Selanjutnya, akan dibuat *dataframe* yang berisikan variabel faskes, JP, *Doctor*, *Dentist*, *Nursing*, *Midwifery*, *Pharmaceutical*, *Public Health*, *Environmental Health*, *Nutritionist*, dan *Medical Laboratory*. Dengan bantuan *software* R dan menggunakan fungsi “*dataframe()*” diperoleh sebagai berikut.

```
> #Pembuatan Dataframe
> df.faskes <- data.frame(faskes, data$JP, data$Doctor, data$Dentist, data$Nursing,
+ data$Midwifery, data$Pharmaceutical, data$Public Health,
+ data$Environmental Health, data$Nutritionist, data$Medical Laboratory)
> df.faskes
```

	faskes	data.JP	data.Doctor	data.Dentist	data.Nursing	data.Midwifery	data.Pharmaceutical	data..Public.Health.	data..Environmental.Health.
1	1	191700	200	43	356	160	85	16	11
2	1	143600	79	23	39	44	42	10	2
3	0	251600	92	122	621	166	171	43	12
4	0	181140	46	16	22	58	36	9	2
5	1	255960	314	99	451	123	114	14	7
6	1	177640	149	27	163	88	82	14	7
7	1	253330	350	126	478	189	137	21	7
8	1	272890	84	37	108	86	66	20	5
9	0	172410	353	106	474	98	138	15	8
10	1	121700	32	16	24	36	34	7	0
11	1	101390	238	54	191	34	70	5	3

```
data.Nutritionist data..Medical.Laboratory.
1 18 67
2 4 11
3 23 100
4 4 9
5 16 59
6 13 28
7 13 75
8 7 15
9 17 78
10 2 3
11 8 37
```

Langkah berikutnya adalah **membangun model**

1. Model Logit

Kita akan membangun model logit antara meningkatnya jumlah fasilitas kesehatan (faskes) dan jumlah penduduk (JP), jumlah dokter (*Doctor*), jumlah dokter gigi (*Dentist*), jumlah perawat (*Nursing*), jumlah bidan (*Midwifery*), jumlah tenaga kefarmasian (*Pharmaceutical*), jumlah tenaga kesehatan masyarakat (*Public Health*), jumlah tenaga kesehatan lingkungan (*Environmental Health*), jumlah tenaga gizi (*Nutritionist*), dan jumlah ahli teknologi laboratorium medik (*Medical Laboratory*), di mana faskes sebagai variabel dependen. Dengan bantuan *software* R diperoleh model logit dan estimasi parameternya.

```
> #Logit
> mod.logit <- glm(faskes ~ data.JP + data.Doctor + data.Dentist + data.Nursing + data.Midwifery +
+ data.Pharmaceutical + data..Public.Health. + data..Environmental.Health. + data.Nutritionist + dat
+ a..Medical.Laboratory.,
+ data = df.faskes, family=binomial(link="logit"))
> summary(mod.logit)
```

```
Call:
glm(formula = faskes ~ data.JP + data.Doctor + data.Dentist +
  data.Nursing + data.Midwifery + data.Pharmaceutical + data..Public.Health. +
  data..Environmental.Health. + data.Nutritionist + data..Medical.Laboratory.,
  family = binomial(link = "logit"), data = df.faskes)
```

Deviance Residuals:

```
[1] 0 0 0 0 0 0 0 0 0 0 0
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.919e+02	3.284e+06	0	1
data.JP	2.501e-04	1.398e+01	0	1
data.Doctor	3.181e+00	2.432e+04	0	1
data.Dentist	-1.323e+01	1.403e+05	0	1
data.Nursing	1.083e+00	5.880e+03	0	1
data.Midwifery	2.050e-01	3.517e+03	0	1
data.Pharmaceutical	-3.626e+00	1.384e+04	0	1
data..Public.Health.	4.495e+01	3.287e+05	0	1
data..Environmental.Health.	-1.379e+02	1.490e+06	0	1
data.Nutritionist	3.001e+01	3.154e+05	0	1
data..Medical.Laboratory.	3.709e+00	1.051e+05	0	1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1.2891e+01 on 10 degrees of freedom
Residual deviance: 4.7154e-10 on 0 degrees of freedom
AIC: 22
```

Number of Fisher Scoring iterations: 23

Berdasarkan *output* di atas diperoleh model matematika sebagai berikut.

$$\hat{Y} = -1.919e^{+02} + 2.501e^{-04}JP + 3.181e^{+00}Doctor - 1.323e^{+01}Dentist + 1.083e^{+00}Nursing + 2.050e^{-01}Midwifey - 3.626e^{+00}Pharmaceutical + 4.495e^{+01}Public Health - 1.379e^{+02}Environmental Health + 3.001e^{+01}Nutritionist + 3.709e^{+00}Medical Laboratory + u_i$$

\hat{Y} pada model ini adalah logg-odds (logit) probabilitas bahwa jumlah fasilitas kesehatan mengalami peningkatan. Berdasarkan model di atas, dapat kita **interpretasikan** sebagai berikut.

a. Intercept

Diperoleh *intercept* sebesar $-1.919e^{+02}$. karena bernilai negatif dan probabilitas tidak boleh negatif, maka diubah menjadi 0. Hal ini menunjukkan bahwa probabilitas meningkatnya jumlah fasilitas kesehatan adalah 0. **tidak masuk akal untuk menafsirkan *intercept* itu sendiri dalam model ini.**

b. Koefisien variabel JP

Koefisien variabel JP adalah $2.501e^{-04}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang penduduk, maka log-odds bertambahnya jumlah fasilitas kesehatan meningkat $2.501e^{-04}$** dengan mempertahankan variabel lainnya tetap.

c. Koefisien variabel *Doctor*

Koefisien variabel *Doctor* adalah $3.181e^{+00}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang dokter, maka log-odds bertambahnya jumlah fasilitas kesehatan meningkat $3.181e^{+00}$** , dengan mempertahankan variabel lainnya tetap.

d. Koefisien variabel *Dentist*

Koefisien variabel *Dentist* adalah $-1.323e^{+01}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang dokter gigi, maka log-odds bertambahnya jumlah fasilitas kesehatan menurun $1.323e^{+01}$** , dengan mempertahankan variabel lainnya tetap.

e. Koefisien variabel *Nursing*

Koefisien variabel *Nursing* adalah $1.083e^{+00}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang perawat, maka log-odds bertambahnya jumlah fasilitas kesehatan meningkat $1.083e^{+00}$** , dengan mempertahankan variabel lainnya tetap.

f. Koefisien variabel *Midwifey*

Koefisien variabel *Midwifey* adalah $2.050e^{-01}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang bidan, maka log-odds bertambahnya jumlah fasilitas**

kesehatan meningkat $2.050e^{-01}$, dengan mempertahankan variabel lainnya tetap.

g. Koefisien variabel *Pharmaceutical*

Koefisien variabel *Pharmaceutical* adalah $-3.626e^{+00}$. Hal ini menunjukkan bahwa setiap bertambahnya seorang tenaga kefarmasian, maka log-odds bertambahnya jumlah fasilitas kesehatan menurun $3.626e^{+00}$, dengan mempertahankan variabel lainnya tetap.

h. Koefisien variabel *Public Health*

Koefisien variabel *Public Health* adalah $4.495e^{+01}$. Hal ini menunjukkan bahwa setiap bertambahnya seorang tenaga kesehatan masyarakat, maka log-odds bertambahnya jumlah fasilitas kesehatan meningkat $4.495e^{+01}$, dengan mempertahankan variabel lainnya tetap.

i. Koefisien variabel *Environmental Health*

Koefisien variabel *Environmental Health* adalah $-1.379e^{+02}$. Hal ini menunjukkan bahwa setiap bertambahnya seorang tenaga kesehatan lingkungan, maka log-odds bertambahnya jumlah fasilitas kesehatan menurun $1.379e^{+02}$, dengan mempertahankan variabel lainnya tetap.

j. Koefisien variabel *Nutritionist*

Koefisien variabel *Nutritionist* adalah $3.709e^{+00}$. Hal ini menunjukkan bahwa setiap bertambahnya seorang tenaga gizi, maka log-odds bertambahnya jumlah fasilitas kesehatan meningkat $3.709e^{+00}$ dengan mempertahankan variabel lainnya tetap.

2. Model Probit

Kita akan membangun model probit antara meningkatnya jumlah fasilitas kesehatan (faskes) dan jumlah penduduk (JP), jumlah dokter (*Doctor*), jumlah dokter gigi (*Dentist*), jumlah perawat (*Nursing*), jumlah bidan (*Midwifey*), jumlah tenaga kefarmasian (*Pharmaceutical*), jumlah tenaga kesehatan masyarakat (*Public Health*), jumlah tenaga kesehatan lingkungan (*Environmental Health*), jumlah tenaga gizi (*Nutritionist*), dan jumlah ahli teknologi laboratorium medik (*Medical Laboratory*), di mana faskes sebagai variabel dependen. Dengan bantuan *software* R diperoleh model probit dan estimasi parameternya.

Berdasarkan *output* di atas diperoleh model matematika sebagai berikut.

$$\hat{Y} = -5.234e^{+01} + 6.823e^{-05}JP + 8.678e^{-01}Doctor - 3.609e^{+00}Dentist + 2.955e^{-01}Nursing + 5.592e^{-02}Midwifey - 9.891e^{-01}Pharmaceutical + 1.226e^{+01}Public Health - 3.762e^{+01}Environmental Health + 8.187e^{+00}$$

```

> #probit
> mod.probit <- glm(faskes ~ data.JP + data.Doctor + data.Dentist + data.Nursing + data.Midwifey +
+ data.Pharmaceutical + data..Public.Health. + data..Environmental.Health. + data.Nutritionist + data..Medical.Laboratory.,
+ data = df.faskes, family=binomial(link="probit"))
> summary(mod.probit)

Call:
glm(formula = faskes ~ data.JP + data.Doctor + data.Dentist +
data.Nursing + data.Midwifey + data.Pharmaceutical + data..Public.Health. +
data..Environmental.Health. + data.Nutritionist + data..Medical.Laboratory.,
family = binomial(link = "probit"), data = df.faskes)

Deviance Residuals:
[1] 0 0 0 0 0 0 0 0 0 0 0 0

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    -5.234e+01  7.042e+05      0      1
data.JP         6.823e-05  2.998e+00      0      1
data.Doctor     8.678e-01  5.214e+03      0      1
data.Dentist    -3.609e+00  3.009e+04      0      1
data.Nursing     2.955e-01  1.261e+03      0      1
data.Midwifey    5.592e-02  7.541e+02      0      1
data.Pharmaceutical -9.891e-01  2.967e+03      0      1
data..Public.Health. 1.226e+01  7.047e+04      0      1
data..Environmental.Health. -3.762e+01  3.194e+05      0      1
data.Nutritionist  8.187e+00  6.763e+04      0      1
data..Medical.Laboratory. 1.012e+00  2.253e+04      0      1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1.2891e+01  on 10  degrees of freedom
Residual deviance: 2.2608e-10  on  0  degrees of freedom
AIC: 22

Number of Fisher Scoring iterations: 23

```

$$\text{Nutritionist} + 1.012e^{+00} \text{Medical Laboratory} + u_i$$

\hat{Y} pada model ini adalah fungsi distribusi kumulatif dari distribusi normal standar (fungsi probit) dari probabilitas bahwa jumlah fasilitas kesehatan mengalami peningkatan. Berdasarkan model di atas, dapat kita **interpretasikan** sebagai berikut.

a. Intercept

Diperoleh *intercept* sebesar $-5.243e^{+01}$. karena bernilai negatif dan probabilitas tidak boleh negatif, maka diubah menjadi 0. Hal ini menunjukkan bahwa probabilitas meningkatnya jumlah fasilitas kesehatan adalah 0. **tidak masuk akal untuk menafsirkan *intercept* itu sendiri dalam model ini.**

b. Koefisien variabel JP

Koefisien variabel JP adalah $6.823e^{-05}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang penduduk, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan meningkat $6.823e^{-05}$** , dengan mempertahankan variabel lainnya tetap.

c. Koefisien variabel *Doctor*

Koefisien variabel *Doctor* adalah $8.678e^{-01}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang dokter, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan meningkat $8.678e^{-01}$** , dengan mempertahankan variabel lainnya tetap.

d. Koefisien variabel *Dentist*

Koefisien variabel *Dentist* adalah $-3.069e^{+00}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang dokter gigi, maka nilai fungsi distribusi probit**

bertambahnya jumlah fasilitas kesehatan menurun $3.069e^{+00}$, dengan mempertahankan variabel lainnya tetap.

e. Koefisien variabel *Nursing*

Koefisien variabel *Nursing* adalah $2.955e^{-01}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang perawat, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan meningkat $2.955e^{-01}$, dengan mempertahankan variabel lainnya tetap.**

f. Koefisien variabel *Midwifery*

Koefisien variabel *Midwifery* adalah $5.592e^{-02}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang bidan, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan meningkat $5.592e^{-02}$, dengan mempertahankan variabel lainnya tetap.**

g. Koefisien variabel *Pharmaceutical*

Koefisien variabel *Pharmaceutical* adalah $-9.891e^{-01}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang tenaga kefarmasian, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan menurun $9.891e^{-01}$, dengan mempertahankan variabel lainnya tetap.**

h. Koefisien variabel *Public Health*

Koefisien variabel *Public Health* adalah $1.226e^{+01}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang tenaga kesehatan masyarakat, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan meningkat $1.226e^{+01}$, dengan mempertahankan variabel lainnya tetap.**

i. Koefisien variabel *Environmental Health*

Koefisien variabel *Environmental Health* adalah $-3.762e^{+01}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang tenaga kesehatan lingkungan, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan menurun $3.762e^{+01}$, dengan mempertahankan variabel lainnya tetap.**

j. Koefisien variabel *Nutritionist*

Koefisien variabel *Nutritionist* adalah $8.187e^{+00}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang tenaga gizi, maka nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan meningkat $8.187e^{+00}$, dengan mempertahankan variabel lainnya tetap.**

k. Koefisien variabel *Medical Laboratory*

Koefisien variabel *Medical Laboratory* adalah $1.012e^{+00}$. Hal ini menunjukkan bahwa **setiap bertambahnya seorang ahli teknologi laboratorium medik, maka**

nilai fungsi distribusi probit bertambahnya jumlah fasilitas kesehatan meningkat $1.012e^{+00}$, dengan mempertahankan variabel lainnya tetap.

Perbandingan Model Logit dan Model Probit

```

> #Membandingkan Model Logit, dan Probit
> library(huxtable)
> library(magrittr)
> huxreg(Logit = mod.logit, Probit = mod.probit) %>%
+   set_caption("Models estimating probability
+     of denying loans")%>%
+   add_colnames()

```

names	Logit Logit	Probit Probit
(Intercept)	-191.865 (3284356.597)	-52.344 (704207.575)
data.JP	0.000 (13.983)	0.000 (2.998)
data.Doctor	3.181 (24319.204)	0.868 (5214.345)
data.Dentist	-13.230 (140349.186)	-3.609 (30092.640)
data.Nursing	1.083 (5879.914)	0.295 (1260.728)
data.Midwifey	0.205 (3517.063)	0.056 (754.103)
data.Pharmaceutical	-3.626 (13837.019)	-0.989 (2966.832)
data..Public.Health.	44.948 (328679.249)	12.262 (70472.984)
data..Environmental.Health.	-137.895 (1489713.129)	-37.620 (319413.320)
data.Nutritionist	30.007 (315407.542)	8.187 (67627.363)
data..Medical.Laboratory.	3.709 (105065.003)	1.012 (22527.264)
N	11	11
logLik	-0.000	-0.000
AIC	22.000	22.000

*** p < 0.001; ** p < 0.01; * p < 0.05.

Column names: names, Logit, Probit

Berdasarkan *output* di atas, model logit dan model probit memiliki nilai AIC yang sama, yaitu 22.

Kesimpulan :

Diketahui bahwa model logit dan model probit memiliki nilai AIC yang sama, kelompok kami tetap memilih **model logit adalah model terbaik (*best model*)** karena model logit lebih sederhana daripada model probit. Hal ini dikarenakan model logit tidak melibatkan kalkulasi fungsi distribusi kumulatif seperti yang terjadi dalam model probit.

LAMPIRAN

Data dan *Codes* : [Data dan Codes Diskusi 1 Kelompok 8](#)

REFERENSI

Khrisna. (2023). Model Probabilitas linear. Olah Data Statistik.

<https://datariset.com/olahdata/detail/model-probabilitas-linear>

Kota Depok Dalam Angka 2023. (2023). Badan Pusat Statistik (BPS) Kota Depok.

<https://depokkota.bps.go.id/publication/2023/02/28/55242ac062da048aa76aaffd/kota-depok-dalam-angka-2023.html>

Rohman, M. A. (2023). *Logit model: APA Kelebihan Dan Kekurangannya?*. Sekolah Statistics and Data Analytics.

<https://sekolahstata.com/logit-model-apa-kelebihan-dan-kekurangannya/>

Rohman, M. A. (2023). Model Probit: Kelebihan Dan Kekurangannya. Sekolah Statistics and Data Analytics. <https://sekolahstata.com/model-probit-kelebihan-dan-kekurangannya/>

Santoso, S. (2010). Model Logit (logistic regression) Sebagai Model Probabilitas Linier Alternatif, Dalam Contoh Aplikasi teknik Industri. *Jurnal Teknik Industri*, 2(1), 59.
<https://doi.org/10.22219/jtiumm.vol2.no1.59-64>