

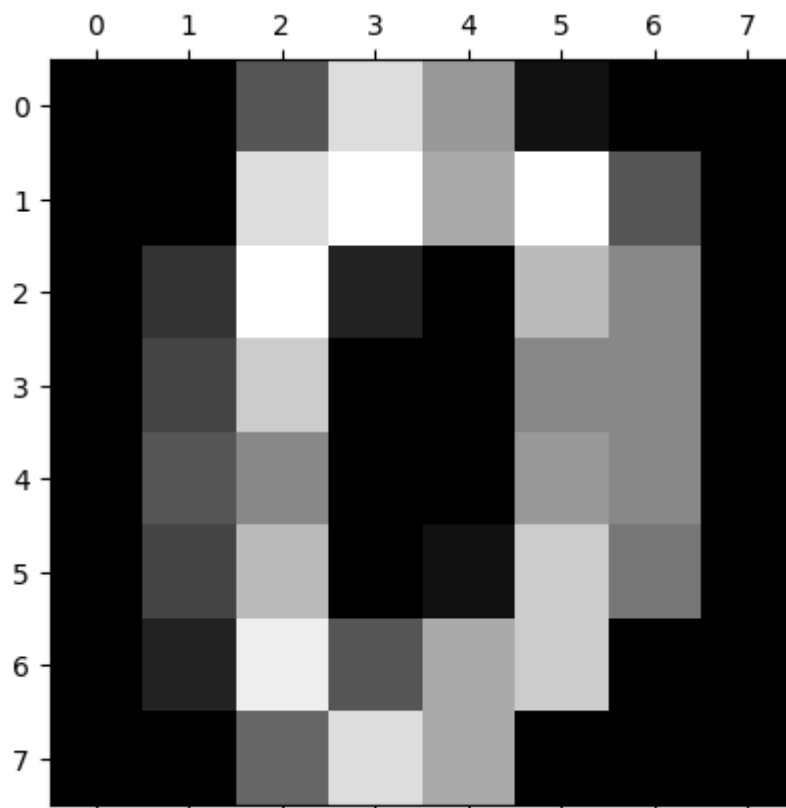
Percobaan 2

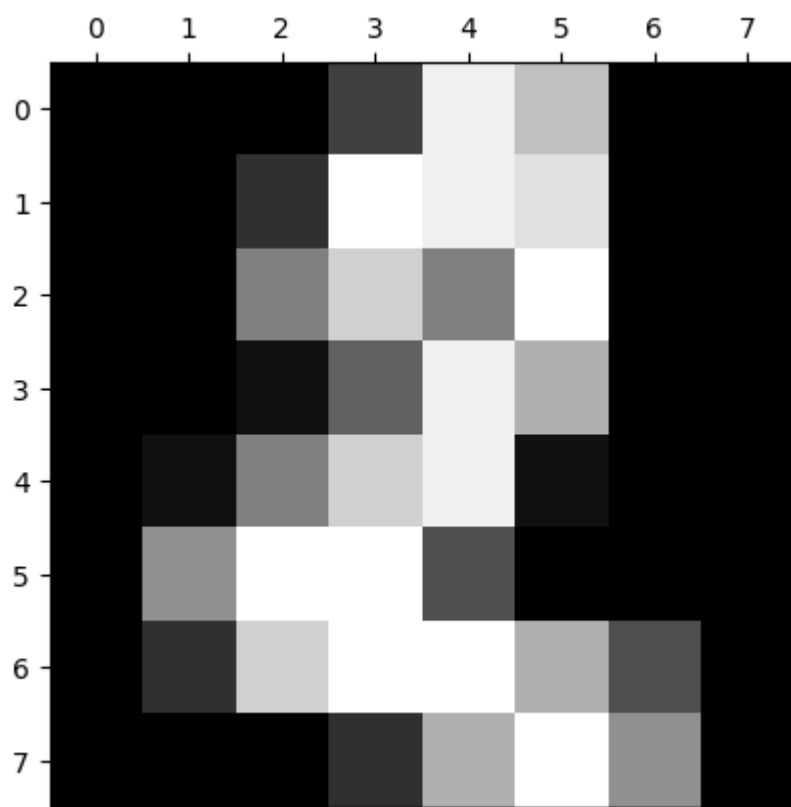
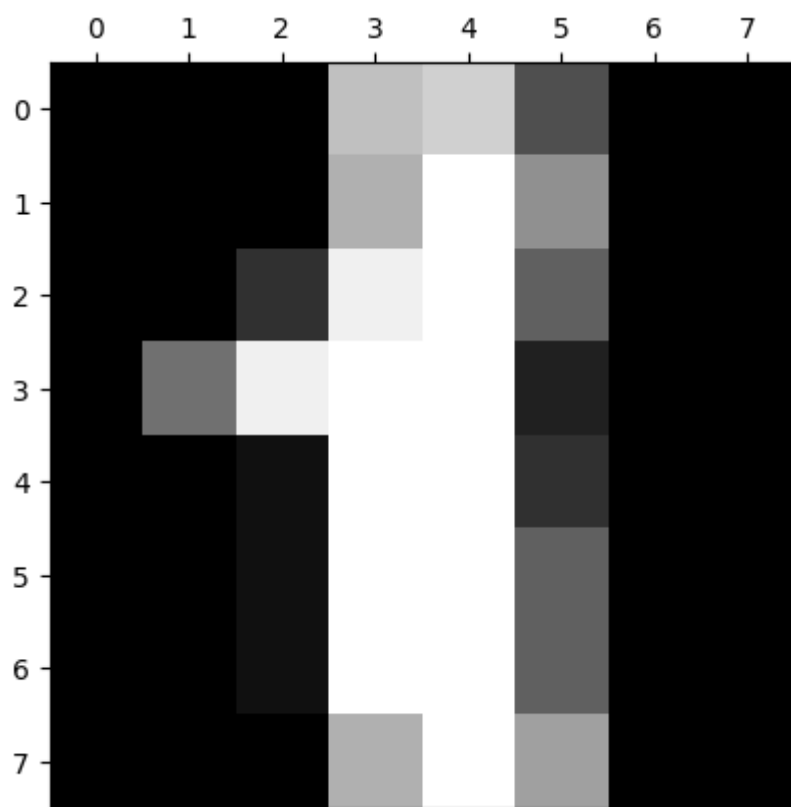
logistic regression untuk multiclass classification.

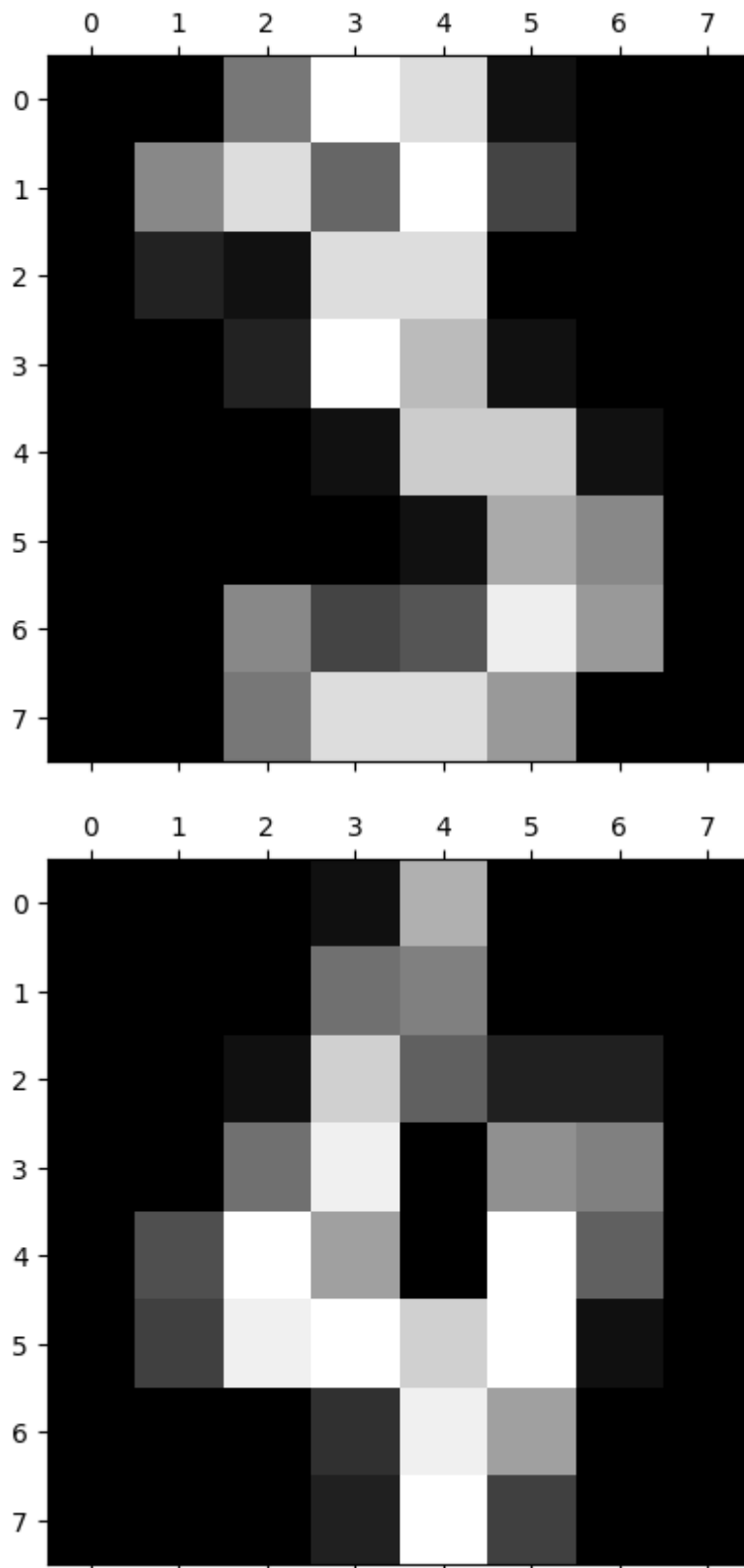
```
In [ ]: from sklearn.datasets import load_digits
%matplotlib inline
import matplotlib.pyplot as plt
digits = load_digits()


plt.gray()
for i in range(5):
    plt.matshow(digits.images[i])
```

<Figure size 640x480 with 0 Axes>







 **Analisis :** Menampilkan dataset mnist dari library sklearn mengatur mode warna menjadi grayscale, lalu menampilkannya dalam loop dari digit 0 hingga 4.

```
In [ ]: dir(digits)
```

```
Out[ ]: ['DESCR', 'data', 'feature_names', 'frame', 'images', 'target', 'target_names']
```

```
In [ ]: digits.data[0]
```

```
Out[ ]: array([ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.,  0.,  0., 13., 15., 10.,
          15.,  5.,  0.,  0.,  3., 15.,  2.,  0., 11.,  8.,  0.,  0.,  4.,
          12.,  0.,  0.,  8.,  8.,  0.,  0.,  5.,  8.,  0.,  0.,  9.,  8.,
           0.,  0.,  4., 11.,  0.,  1., 12.,  7.,  0.,  0.,  2., 14.,  5.,
          10., 12.,  0.,  0.,  0.,  0.,  6., 13., 10.,  0.,  0.,  0.]
```

Menerapkan logistic regression model

```
In [ ]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(digits.data, digits.target, test

from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
```

```
In [ ]: model.fit(X_train, y_train)
```

c:\Users\M S I\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\linear_model_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:


<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

n_iter_i = _check_optimize_result(

```
Out[ ]: LogisticRegression
LogisticRegression()
```

 **Analisis** : Membagi data menjadi train dan test, lalu menerapkan model logistic regresi


Mengukur akurasi model

```
In [ ]: model.score(X_test, y_test)
```

```
Out[ ]: 0.9611111111111111
```

```
In [ ]: model.predict(digits.data[0:5])
```

```
Out[ ]: array([0, 1, 2, 3, 4])
```

 **Analisis** : Kode `model.score(X_test, y_test)` digunakan untuk menghitung akurasi model regresi logistik pada data pengujian. Akurasi adalah rasio antara jumlah prediksi yang benar dan total jumlah prediksi. Dalam hal ini, model Anda menghasilkan akurasi sebesar 0.96111 atau 96.111% pada data pengujian.

Nilai ini cukup tinggi, yang menunjukkan bahwa model Anda melakukan pekerjaan yang baik dalam memprediksi label kelas yang benar untuk data pengujian. Dengan

kata lain, model Anda benar dalam sekitar 96.111% kasus pada data pengujian.

Classification_report

```
In [ ]: y_predicted = model.predict(X_test)
```

```
In [ ]: from sklearn.metrics import classification_report  
print(classification_report(y_test, y_predicted))
```

	precision	recall	f1-score	support
0	1.00	0.96	0.98	27
1	0.84	0.97	0.90	33
2	0.97	1.00	0.99	33
3	0.97	1.00	0.99	35
4	0.98	0.95	0.97	44
5	0.95	0.95	0.95	40
6	0.96	0.96	0.96	28
7	1.00	1.00	1.00	43
8	1.00	0.84	0.91	38
9	0.95	0.97	0.96	39
accuracy			0.96	360
macro avg	0.96	0.96	0.96	360
weighted avg	0.96	0.96	0.96	360

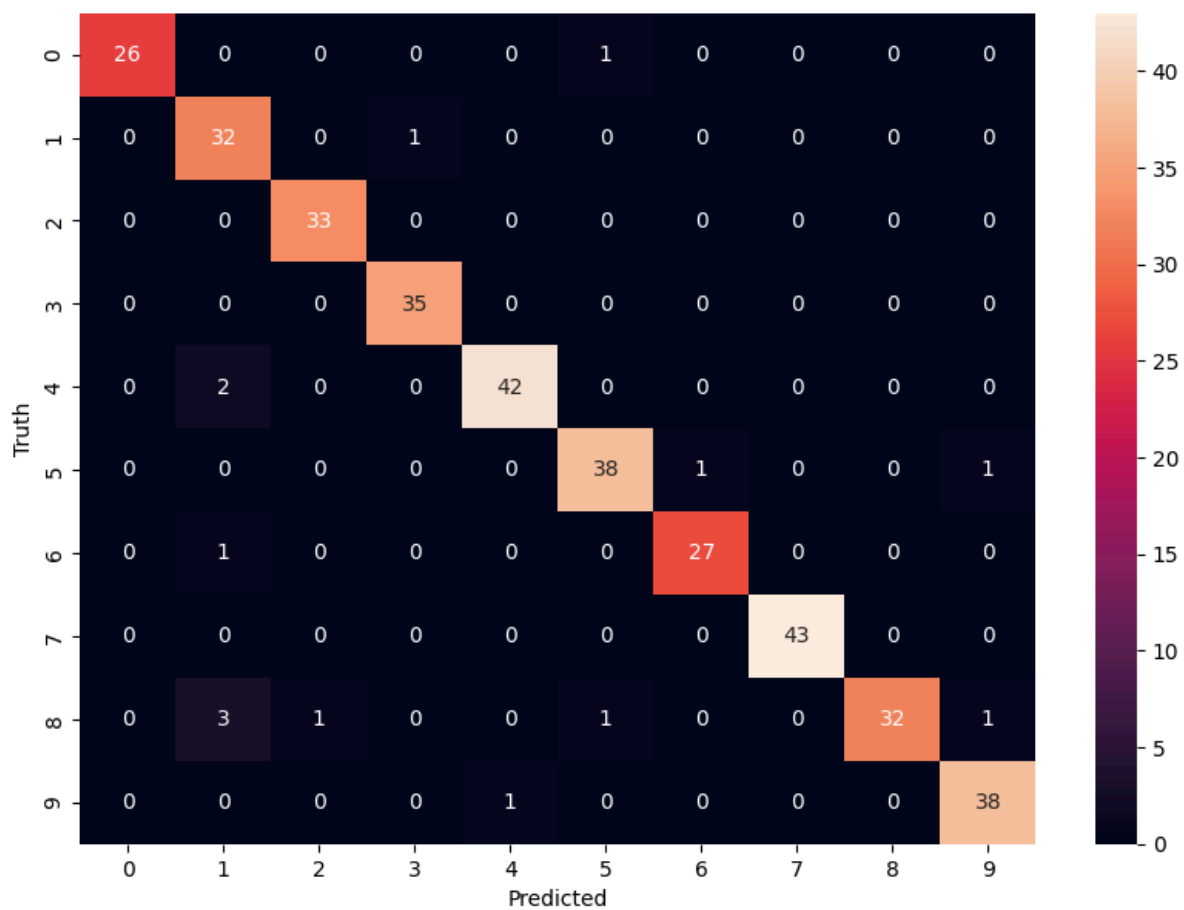
Confusion Matrix

```
In [ ]: from sklearn.metrics import confusion_matrix  
cm = confusion_matrix(y_test, y_predicted)  
cm
```

```
Out[ ]: array([[26,  0,  0,  0,  0,  1,  0,  0,  0,  0],  
               [ 0, 32,  0,  1,  0,  0,  0,  0,  0,  0],  
               [ 0,  0, 33,  0,  0,  0,  0,  0,  0,  0],  
               [ 0,  0,  0, 35,  0,  0,  0,  0,  0,  0],  
               [ 0,  2,  0,  0, 42,  0,  0,  0,  0,  0],  
               [ 0,  0,  0,  0,  0, 38,  1,  0,  0,  1],  
               [ 0,  1,  0,  0,  0,  0, 27,  0,  0,  0],  
               [ 0,  0,  0,  0,  0,  0,  0, 43,  0,  0],  
               [ 0,  3,  1,  0,  0,  1,  0,  0, 32,  1],  
               [ 0,  0,  0,  0,  1,  0,  0,  0,  0, 38]], dtype=int64)
```

```
In [ ]: import seaborn as sn  
plt.figure(figsize = (10,7))  
sn.heatmap(cm, annot=True)  
plt.xlabel('Predicted')  
plt.ylabel('Truth')
```

```
Out[ ]: Text(95.7222222222221, 0.5, 'Truth')
```



Praktikum Pemodelan Statistika Terapan

Dosen Pengampu Ronny Sutsetyoko



Wahyu Ikbal Maulana

3323600056

D3 SDT B

Politeknik Elektronika Negeri Surabaya

Percobaan 1

Studi dari World Values Surveys adalah untuk melihat faktor-faktor yang memengaruhi persepsi masyarakat tentang upaya pemerintah untuk mengurangi kemiskinan. Tujuannya adalah untuk memprediksi persepsi individu tentang upaya pemerintah untuk mengurangi kemiskinan berdasarkan faktor-faktor seperti kemiskinan, agama, negara, gelar, jenis kelamin, dan usia. Dalam studi kasus yang diberikan, persepsi individu dapat mengambil tiga nilai berikut - Terlalu Sedikit, Tentang Benar, Terlalu Banyak.

Untuk analisis ini digunakan data dari World Values Surveys untuk beberapa negara, yaitu Australia, Norwegia, Swedia, dan Amerika Serikat dari paket 'carData' di R.

```
In [ ]: library(carData)
library(MASS)
data(WVS)
head(WVS)
```

Warning message:
"package 'carData' was built under R version 4.3.2"

A data.frame: 6 × 6

	poverty	religion	degree	country	age	gender
	<ord>	<fct>	<fct>	<fct>	<int>	<fct>
1	Too Little	yes	no	USA	44	male
2	About Right	yes	no	USA	40	female
3	Too Little	yes	no	USA	36	female
4	Too Much	yes	yes	USA	25	female
5	Too Little	yes	yes	USA	39	male
6	About Right	yes	no	USA	80	female

```
In [ ]: summary(WVS)
```

```
      poverty      religion      degree      country      age
Too Little :2708    no : 786    no :4238  Australia:1874  Min.   :18.00
About Right:1862    yes:4595    yes:1143   Norway   :1127  1st Qu.:31.00
Too Much   : 811                                Sweden   :1003  Median :43.00
                                                USA      :1377  Mean   :45.04
                                                3rd Qu.:58.00
                                                Max.   :92.00

      gender
female:2725
male  :2656
```

Deskripsi data: Kemiskinan adalah variabel dependen terdiri dari 3 kategori: too little, about right, too much. Lima variabel independen berikut:

Agama : anggota suatu agama -tidak atau ya

Gelar : mempunyai gelar (tidak atau ya)

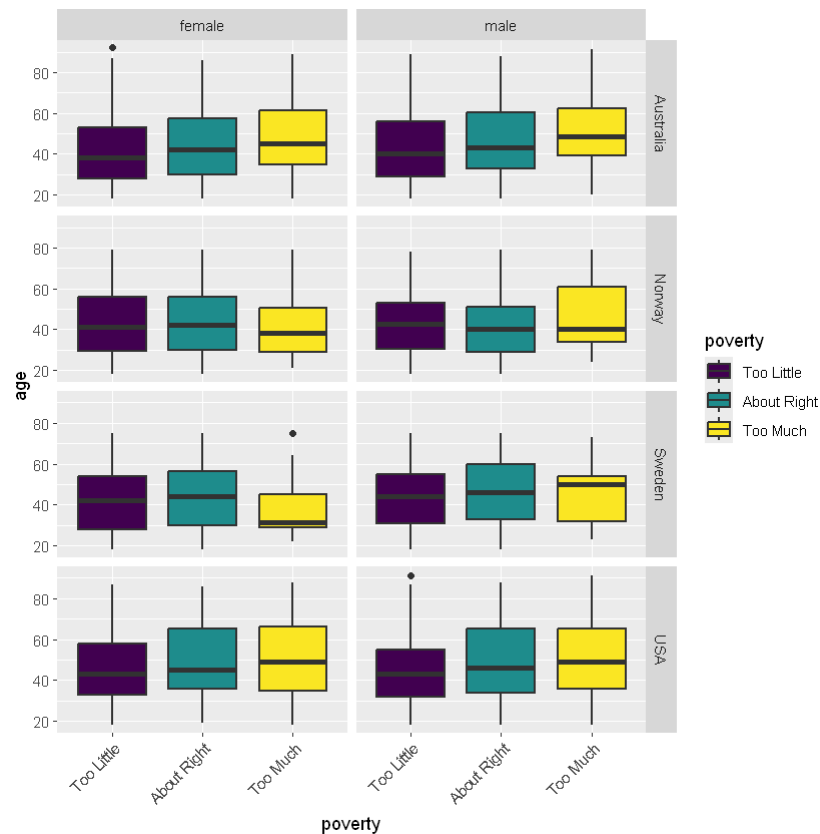
Negara : Australia, Norwegia, Swedia atau Amerika Serikat

Usia : usia (tahun)

Jenis Kelamin : laki-laki atau perempuan

Ringkasan Statistik

```
In [ ]: library(ggplot2)
ggplot(WVS, aes(x = poverty, y = age, fill = poverty)) + geom_boxplot(size = .75)
```



Alternatif menganalisis distribusi kemiskinan berdasarkan usia, jenis kelamin, dan negara

```
In [ ]: model_fit <- polr(poverty~religion+degree+country+age+gender, data = WVS, Hess = TRUE)
summary(model_fit)
```

Call:

```
polr(formula = poverty ~ religion + degree + country + age +  
      gender, data = WVS, Hess = TRUE)
```

Coefficients:

	Value	Std. Error	t value
religionyes	0.17973	0.077346	2.324
degreeyes	0.14092	0.066193	2.129
countryNorway	-0.32235	0.073766	-4.370
countrySweden	-0.60330	0.079494	-7.589
countryUSA	0.61777	0.070665	8.742
age	0.01114	0.001561	7.139
gendermale	0.17637	0.052972	3.329

Intercepts:

	Value	Std. Error	t value
Too Little About Right	0.7298	0.1041	7.0128
About Right Too Much	2.5325	0.1103	22.9496

Residual Deviance: 10402.59

AIC: 10420.59

Model regresi logistik Model Regresi Logistik Odds Proporsional menggunakan fungsi polr() dari paket MASS.

```
In [ ]: summary_table <- coef(summary(model_fit))  
pval <- pnorm(abs(summary_table[, "t value"]), lower.tail = FALSE) * 2  
summary_table <- cbind(summary_table, "p value" = round(pval, 3))  
summary_table
```

A matrix: 9 × 4 of type dbl

	Value	Std. Error	t value	p value
religionyes	0.17973194	0.077346042	2.323738	0.020
degreeyes	0.14091745	0.066193109	2.128884	0.033
countryNorway	-0.32235359	0.073766034	-4.369946	0.000
countrySweden	-0.60329785	0.079493909	-7.589234	0.000
countryUSA	0.61777260	0.070664761	8.742301	0.000
age	0.01114091	0.001560585	7.138933	0.000
gendermale	0.17636863	0.052972253	3.329453	0.001
Too Little About Right	0.72976353	0.104061643	7.012800	0.000
About Right Too Much	2.53247870	0.110349780	22.949558	0.000

Karena p-value untuk semua variabel < 0,05, maka secara statistik signifikan pada 95% CI.

Interpretasi Model Peluang Proporsional Untuk menginterpretasikan model ini, pertama-

tama kita perlu memahami cara kerja model peluang proporsional. Misalkan J adalah jumlah total kategori dari variabel terikat dan M adalah jumlah variabel bebas (Dalam kumpulan data yang diberikan, J=3 dan M = 5). Rumusan matematis Model Proporsional Odds

diberikan di bawah ini: Di sini, j adalah level dari kategori terurut dengan level J dan i

berhubungan dengan variabel independen Dalam kasus kami j = 1 mengacu pada too little j

= 2 mengacu pada about right j = 3 mengacu pada too much i = 1 mengacu pada 'agama' i

= 2 mengacu pada 'gelar' i = 3 mengacu pada 'negara' i = 4 mengacu pada 'usia' i = 5

mengacu pada 'gender' Menafsirkan model statistik: Variabel kategoris seperti gender dapat diartikan sebagai: individu laki-laki dibandingkan dengan individu perempuan, dikaitkan dengan kemungkinan yang lebih tinggi untuk memiliki persepsi positif tentang upaya pemerintah untuk mengurangi kemiskinan. Nilai t lebih besar dari 2 dan oleh karena itu signifikan secara statistik pada tingkat 5%.

Variabel kontinyu seperti usia dapat diartikan sbb: Dengan bertambahnya satu satuan usia log of odds memiliki persepsi positif tentang upaya pemerintah untuk mengurangi kemiskinan meningkat sebesar 0,011. Intersep: Secara matematis, intersep "too little | about right" sesuai dengan $\text{logit}[P(Y \leq 1)]$. Ini dapat diartikan sebagai log of odds percaya bahwa pemerintah melakukan 'Terlalu Sedikit' versus percaya bahwa pemerintah melakukan 'about right' atau 'too much'. Demikian pula, intersep 'about right | too much' sesuai dengan $\text{logit}[P(Y \leq 2)]$. Ini dapat diartikan sebagai log of odds percaya bahwa pemerintah melakukan too little atau about right versus percaya bahwa pemerintah melakukan too much Membuat prediksi pada data baru Katakanlah kita ingin memprediksi probabilitas yang sesuai dengan setiap persepsi untuk seorang individu — Test_Person dengan karakteristik berikut: Agama : ya Gelar : tidak Negara: Norwegia Umur : 30 jenis kelamin: laki-laki Komputasi Matematika Dengan menggunakan nilai intersep dan kemiringan dari Model Summary, kita dapat memperkirakan probabilitas yang diinginkan dengan cara berikut: Probabilitas yang sesuai dengan persepsi Terlalu Sedikit akan dihitung sebagai: $\text{logit}[P(Y \leq 1)] = 0.7298 - [(0.17973 \cdot 1) + (0.140920) + (-0.32235 \cdot 1) + (0.0111430) + (0.17637 \cdot 1)] \Rightarrow \text{logit}[P(Y \leq 1)] = 0.36185 \Rightarrow P(Y \leq 1) = \exp(0.36185) / (1 + \exp(0.36185)) = 0.589$ Dalam kasus kami, $P(Y \leq 1) = P(Y = 1) = 0.589$ Demikian pula, probabilitas yang sesuai dengan persepsi Tentang Benar akan dihitung sebagai: $\text{logit}[P(Y \leq 2)] = 2.5325 - [(0.17973 \cdot 1) + (0.140920) + (-0.32235 \cdot 1) + (0.0111430) + (0.17637 \cdot 1)] \Rightarrow \text{logit}[P(Y \leq 2)] = 2.16455 \Rightarrow P(Y \leq 2) = \exp(2.16455) / (1 + \exp(2.16455)) = 0.897$ Jadi, $P(Y = 2) = P(Y \leq 2) - P(Y \leq 1) = 0.897 - 0.589 \Rightarrow P(Y = 2) = 0.308$ Probabilitas yang sesuai dengan persepsi too much akan dihitung sebagai: Jadi, $P(Y = 3) = 1 - P(Y \leq 2) \Rightarrow P(Y = 3) = 0.103$ Dengan menggunakan fungsi prediksi di R

```
In [ ]: new_data <-
data.frame("religion"="yes", "degree"="no", "country"="Norway", "age"=30, "gender"="male")
round(predict(model_fit, new_data, type = "p"), 3)
```

Too Little: 0.589 **About Right:** 0.308 **Too Much:** 0.103