Andre Fransiscus Masalle

Tugas 3 Bootcamp RStudio

1. Model Regresi Logistik dari data titanic

```
> summary(m)

Call:
glm(formula = survived ~ pclass + sex + age, family = binomial(link = "logit"),
    data = titanic)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.58927    0.40572  11.311  < 2e-16 ***
pclass      -1.13324    0.11173 -10.143  < 2e-16 ***
sexmale     -2.49738    0.16612 -15.034  < 2e-16 ***
age         -0.03388    0.00628  -5.395 6.84e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1414.62  on 1045  degrees of freedom
Residual deviance:  983.02  on 1042  degrees of freedom
AIC: 991.02

Number of Fisher Scoring iterations: 4
```

Dapat dilihat bahwa variabel pclass, sex dan age mempengaruhi variabel survival secara signifikan. Namun akan di uji apakah memenuhi asumsi klasiknya dan goodness of fit.

- Cek Asumsi:
  - Uji autokorelasi dan outliers

```
> #cek asumsi
> check_outliers(m)
OK: No outliers detected.
- Based on the following method and threshold: cook (0.7).
- For variable: (Whole model)

> multicollinearity(m)
# Check for Multicollinearity

Low Correlation

   Term  VIF   VIF 95% CI Increased SE Tolerance Tolerance 95% CI
 pclass 1.39 [1.30, 1.52]         1.18      0.72     [0.66, 0.77]
    sex 1.05 [1.02, 1.19]         1.03      0.95     [0.84, 0.99]
    age 1.33 [1.25, 1.45]         1.16      0.75     [0.69, 0.80]
```

Setelah diuji asumsinya, tidak terdapat outliers maupun multikolinearitas.

- Uji Hosmer

```
> performance_hosmer(m, n_bins = 10)
# Hosmer-Lemeshow Goodness-of-Fit Tes

  Chi-squared: 39.882
           df:  8
      p-value:  0.000

Summary: model does not fit well.
```

Ternyata setelah di uji Hosmer, modelnya tidak fit.

- **Goodness of Fit**

```
> r2_nagelkerke(m)
i
Nagelkerke's R2
      0.4560168
```

Dapat disimpulkan bahwa *survivability* penumpang bisa dijelaskan oleh variabel-variabel yang kita teliti.

- **Odds Ratio Interpretation**

Odds ratio dapat diperoleh dari exponensial koefisiennya.

Dan didapatkan bahwa:

- Odds ratio pclass adalah 0,966622734
- Odds ratio sex adalah 0,321927794

2. Anlisis Diskriminan

- **Tes Normalitas**

```
> cek
$multivariateNormality
     Test        H      p value MVN
1 Royston 300.5813 2.657509e-65  NO

$univariateNormality
              Test  Variable Statistic   p value Normality
1 Anderson-Darling survived   196.2157  <0.001      NO
2 Anderson-Darling    sex     204.5210  <0.001      NO
3 Anderson-Darling    age       7.1298  <0.001      NO
4 Anderson-Darling RoundAge     7.1541  <0.001      NO
```

Namun dikarenakan datanya yang besar maka bisa dianggap bahwa datanya normal

- Berapa persen model dapat memodelkan data yang ada

```
Coefficients of linear discriminants:
                   LD1
RoundAge -0.02193064
pclass   -0.77001317
sex      -2.07040545
```

```
> sum(diag(prop.table(ct)))
[1] 0.7820268
```

Sebesar 78%

Andre Fransiscus Masalle