

Regresi Logistik ??

Respon → biner

Prediktor → kategorik



Kasus: 2 Prediktor Biner

Model untuk $P(Y = 1)$:

$$\text{logit}[P(Y = 1)] = \alpha + \beta_1 x + \beta_2 z$$

x dan z



Peubah
indikator/peubah
dummy

Kode Peubah *Dummy*

Table 4.3. Logits Implied by Indicator Variables in Model, $\text{logit}[P(Y = 1)] = \alpha + \beta_1x + \beta_2z$

x	z	Logit
0	0	α
1	0	$\alpha + \beta_1$
0	1	$\alpha + \beta_2$
1	1	$\alpha + \beta_1 + \beta_2$

Asumsi → tidak ada interaksi

Pengaruh satu faktor sama di setiap kategori pada faktor lain



Saat nilai $Z=z$ (tetap/fixed) → perubahan logit dari $x = 0$ ke $x = 1$ sebesar:

$$[\alpha + \beta_1(1) + \beta_2z] - [\alpha + \beta_1(0) + \beta_2z] = \beta_1$$

$\exp(\beta_1)$ equals the conditional odds ratio between X and Y

Saat nilai Z tetap, odd “sukses” untuk $x = 1$ sebesar $\exp(\beta_1)$ kali dibandingkan odd “sukses” untuk $x = 0$

The effects of AZT in slowing the development of AIDS symptoms (*New York Times, 15 February 1991*)

Table 4.4. Development of AIDS Symptoms by AZT Use and Race

Race	AZT Use	Symptoms	
		Yes	No
White	Yes	14	93
	No	32	81
Black	Yes	11	52
	No	12	43

X = AZT treatment,
 Z = race, and
 Y = whether AIDS symptoms developed
(1 = yes, 0 = no)

- Pengkodean
 $X=1$ untuk veteran yang segera diberi AZT,
 $X=0$ untuk selainnya
 $Z=1$ untuk white
 $Z=0$ untuk black

Output

Table 4.5. Computer Output for Logit Model with AIDS Symptoms Data

Log Likelihood -167.5756

Analysis of Maximum Likelihood Estimates

Parameter	Estimate	Std Error	Wald Chi-Square	Pr > ChiSq
Intercept	-1.0736	0.2629	16.6705	<.0001
azt	-0.7195	0.2790	6.6507	0.0099
race	0.0555	0.2886	0.0370	0.8476

LR Statistics

Source	DF	Chi-Square	Pr > ChiSq
azt	1	6.87	0.0088
race	1	0.04	0.8473

Obs	race	azt	y	n	pi_hat	lower	upper
1	1	1	14	107	0.14962	0.09897	0.21987
2	1	0	32	113	0.26540	0.19668	0.34774
3	0	1	11	63	0.14270	0.08704	0.22519
4	0	0	12	55	0.25472	0.16953	0.36396

Interpretasi Hasil

Dugaan Parameter

Dugaan pengaruh AZT sebesar $\hat{\beta}_1 = -0.720$.

Rasio Odds

Rasio odds antara “immediate AZT use” dan “development of AIDS symptoms” sebesar $\exp(-0.720) = 0.49$

Interpretasi Hasil (lanjutan)

Hipotesis untuk kebebasan antara “AZT treatment” dan “development of AIDS symptoms”, (asumsi ras yang sama) adalah $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

Statistik Uji (statistik Wald)

$$Z^2 = (\beta_1/SE)^2 = (-0.720/0.279)^2 = 6.65 \rightarrow (P = 0.010)$$

Kesimpulan

- Pengaruh penggunaan AZT **signifikan** terhadap “development of AIDS symptoms”
- Pengaruh ras **tidak signifikan** terhadap “development of AIDS symptoms” ($Z^2 = 0.0370$)

Uji Kemungkinan Maksimum (LR Test)

- $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$
- $(LR) = -2(L0 - L1) = 6,87$
- *comparing models (4.6)*

$$\text{logit}[P(Y = 1)] = \alpha + \beta_2 z \quad (4.6)$$

- *and (4.5)*

$$\text{logit}[P(Y = 1)] = \alpha + \beta_1 x + \beta_2 z \quad (4.5)$$

- A factor having **two categories** requires only a single indicator variable, taking value **1 or 0** to indicate whether an observation falls in the first or second category.
- A factor having ***l categories requires l – 1 indicator variables***
- An alternative **representation of factors in logistic regression** uses the way **ANOVA** models often express factors.

ANOVA-merepresentasikan model dari faktor

- Faktor yang memiliki i kategori, mengandung $i-1$ indikator peubah
- Model untuk merepresentasikan faktor

$$\text{logit}[P(Y = 1)] = \alpha + \beta_i^X + \beta_k^Z$$

- β_i^X → representasi dari pengaruh X .
Parameter ini mengukur pengaruh pada logit berdasarkan klasifikasi pada kategori i variabel X.
- β_k^z → representasi dari pengaruh z (z bukan pangkat hanya indeks)
- Jika X memiliki i level, maka akan ada parameter sebanyak i-1

- biasanya parameter kategori terakhir di sama dengankan 0
- Modelnya menjadi

$$\beta_1^x x_1 + \beta_2^x x_2 + \dots + \beta_{I-1}^x x_{I-1}$$

