$\boxed{\text{Model Selection}}$ $\dfrac{X_1 \cdots X_p}{\rightsquigarrow \quad \text{stepwise}}$ $+$ $\underline{AIC} / BIC.$ $\Rightarrow$ $\boxed{Y \sim X_1 \ X_2 \ X_3}$

$\boxed{\text{Model Fit}}$
- significance of predictors
- ✓ Interpretation of $\boxed{\beta}$
- Goodness-of-fit :
  Hosmer Lemeshow test
  H_0: model is adequate
  H_a: " inadequate

categorical x $(1/0)$
$$\left[\frac{\text{odds}(Y=1 | \text{company})}{\text{odds}(Y=1 | \text{ref})} = e^{\beta}\right]$$

numerical
$$\frac{\text{odds}(Y=1 | x+1)}{\text{odds}(Y=1 | x)} = e^{\beta}$$

$\boxed{\text{Model diagnostics}}$
- Model Assumption check : Residual plot.
  $Y \sim \text{Bernoulli}(\,p(Y=1 | x_{\cdots})\,)$
  $0 \leq \hat{p} \leq 1$
  $Y - \hat{p}$
  standardized R.P
- Influential points
  Cook's Distance

$\boxed{\text{prediction} \rightsquigarrow \text{classification} \atop \text{(new observation)}}$

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

$\overline{\log}$

odds

$\left[\begin{array}{l} Y: 0 / 1 \\ \underline{X}: \end{array}\right]$

$P(Y=1) \quad X$

$\boxed{P(Y=1)}$

New app at 2021 → predict then $\tilde{Y}$ ($\hat{P}$)
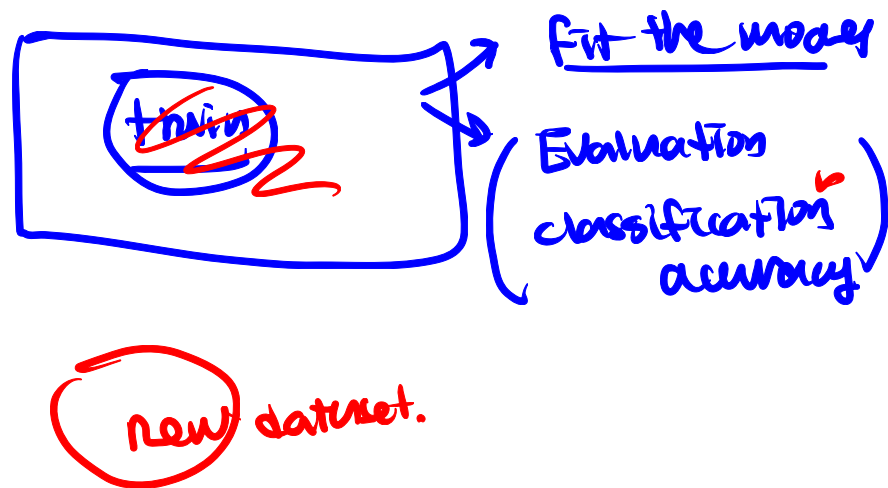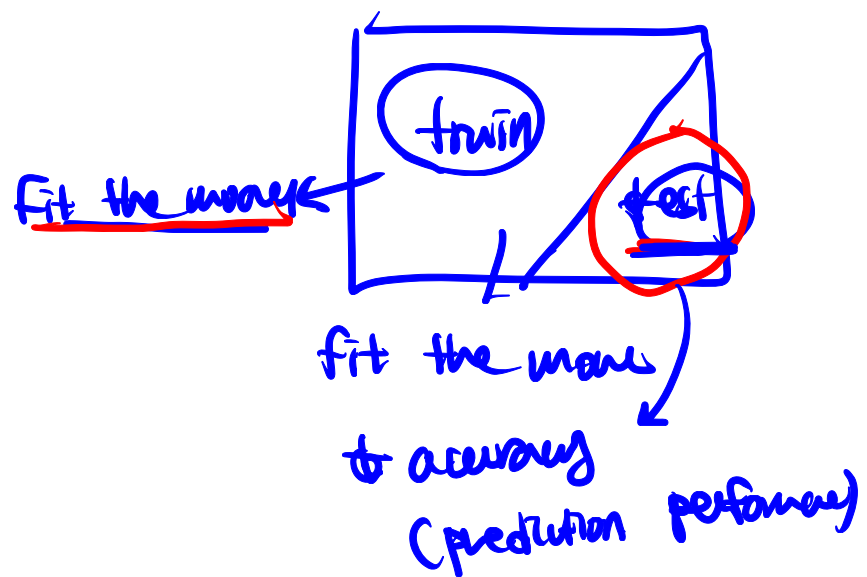
Goal

1
(graduate (fail)) → SAT

$Y \sim X$

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = 2.3 + 0.8 \cdot X \quad \text{SAT}$$

$$\hat{P}(Y=1) = \frac{\exp(2.3 + 0.8x)}{1 + \exp(2.3 + 0.8 \cdot x)}$$

cut-off — 0.5

# of obs w/ diff. $Y \sim \hat{Y}$

X	Y	$\hat{P}(Y=1)$	$\hat{Y}$ (cut: 0.5)	$\hat{Y}$ (cutoff) 0.35
1200	1	0.8	1	1
1500	1	0.7	1	1
900	1	0.5	0	1
560	0	0.3	0	0
400	0	0.2	0	0
700	1	0.4	0	1
1000	0	0.6	1	1

HW 4

Q 1, 2
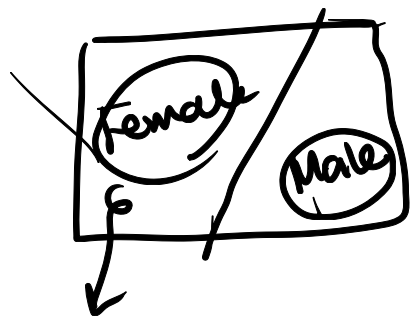
river breeder ⌐ Femal
                └ male

warning message
(0/1 occurs)

perfect separation

fitted prob. very close to 0 or 1.

Y

1

F (0/1)

X.

X

Q3. Q4

or
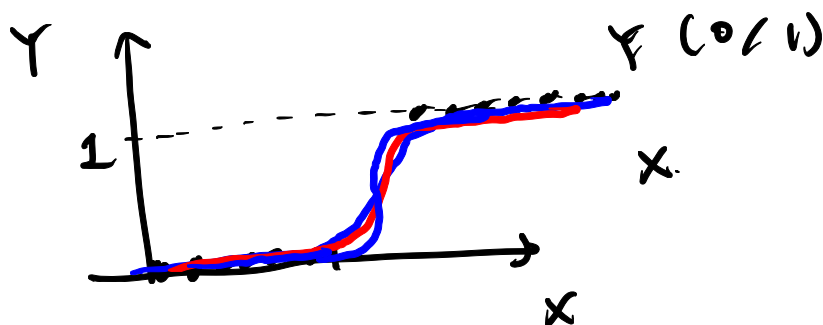
predation index
sleep exposure index

1 2 3 4 5
1 2 3 4 5

$4 \to 8$ / $4 \to 4$

$1 \to 2.$ / $1 \to 1$

[as factor (
  as.fact ( ]

1 2 3 4

Categorical X — nominal / ordinal

@ factor

Female / Male
w / B / A

kids / young adult / adult / senior

1 2 3 4 5

tobacco use

none / 1-3 / 4-10 / > 10
1 2 3 4

51 obs

$n = 51$

# of parameters

P

$p = 8$

degree of freedom $= n - p$