Dummy variables.

categorical data.

not continuous data

gender $\begin{cases} 0 & \text{male} \\ 1 & \text{female} \end{cases}$

typical:

survey results.

→ option $\begin{cases} \text{good} & ? \\ \text{OK} & ? \\ \text{bad.} & ? \end{cases}$

y quantitative result (response).

$\begin{cases} \text{good} \\ \text{OK} \\ \text{bad.} \end{cases}$ $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

$x = (0, 0, 1)^T$.

$(0, 1, 0)^T$.

$(1, 0, 0)^T$.

$y = x^T\beta + e$.

$\beta = (\beta_1, \beta_2, \beta_3)$.

measuring effect of feeling bad.

encoding → underlying $\boxed{\text{assumptions}}$.

$\begin{cases} \text{good} & 2 \\ \text{OK} & 1 \\ \text{bad} & 0 \end{cases}$ $\begin{matrix} 1 \\ 1. \end{matrix}$

$x \in \mathbb{R}$. $\beta \in \mathbb{R}$.

$y = x \cdot \beta + e$.

$\boxed{\mathbb{E}[Y]}$.

$x = \begin{cases} 2 \\ 1 \\ 0. \end{cases}$

good 2β

β.

0.

1. gives good / OK. OK / bad.
   their effects have same difference.

Encoding

NOT related to $\boxed{\hat{Y}}$. $E[Y|X]$.

$\begin{cases} 1 & \text{male} \\ 0 & \text{female} \end{cases}$ $\begin{cases} 1 & \text{female} \\ 0 & \text{male} \end{cases}$ $\hat{\beta}$.

2016 Spring final.

dummy, covariates

$$E[Y|U,X] = \beta_0 + \beta_1 U + \beta_2 X.$$

$$U = \begin{cases} 0 & \text{male} \\ 1 & \text{female} \end{cases} \qquad \hat\beta = (5, 2, 0).$$

$$\text{if } U = \begin{cases} \boxed{2} & \text{male} \\ \boxed{1} & \text{female} \end{cases} \qquad \boxed{2\beta_1} \quad \boxed{\beta_1}$$

$\beta_1 - 2\beta_1 = -\beta_1.$

what is $\hat\beta_{new}$?

β_1 $\boxed{-\text{ effect}}$ $\boxed{\text{male}}$.

$\boxed{\beta_1} + \boxed{\dots}$

$0 + \boxed{\dots}$ $\boxed{\beta_1 \leftrightarrow U.}$

measuring the effect.

$\boxed{\text{male}}$. 0.

linear model

linear transformation.

invertible.

equivalent model.

Solution.

$$U^{new} = aU^{old} + b.$$

$$\begin{cases} 2 = a\cdot 0 + b \\ 1 = a\cdot 1 + b \end{cases} \Rightarrow \begin{cases} a = -1. \\ b = 2 \end{cases}$$

New:
$$E[Y|U^{new}, X] = \beta_0' + \beta_1'(aU^{old} + b) + \beta_2' X.$$
$$= \beta_0' + 2\beta_1' - \beta_1' U^{old} + \beta_2' X.$$

original model:
$$E[Y|U^{old}, X] = \beta_0 + \beta_1 U^{old} + \beta_2 X.$$
$$= 5 + 2U^{old} + 0\cdot X.$$

$$\begin{cases} \beta_0' + 2\beta_1' = 5 \\ -\beta_1' = 2 \\ \beta_2' = 0. \end{cases} \Rightarrow \boxed{\begin{cases} \beta_0' = 9. \\ \beta_1' = -2 \\ \beta_2' = 0 \end{cases}}$$

$$\hat\beta_{new} = (9, -2, 0).$$
$$\hat\beta_{old} = (5, 2, 0).$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix} \quad \mu$$

$$\begin{pmatrix} y_n \\ y_{n+1} \\ \vdots \\ y_{2n} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix} + \vec{e}.$$

$\boxed{\begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix}}$ not identifiable. estimable.

$\rightarrow \boxed{\begin{pmatrix} \mu+\beta_1 \\ \mu+\beta_2 \end{pmatrix}}$ identifiable.

R programming. $lm(y \sim x)$.

$\mu \quad \beta_1 \quad \mu+\boxed{\beta_1}$.

$\uparrow$
0.

preassign it as 0.

$\boxed{\hat{\beta}_2}$. $\beta_2-\beta_1$.

$E Y_1 = \hat{\mu}+\hat{\beta}_1 \rightarrow \boxed{\mu+\beta_1}$.

$E Y_2 = \hat{\mu}+\hat{\beta}_2 \rightarrow \mu+\beta_1 + \beta_2-\beta_1$

$\boxed{\mu+\beta_2}$.

$\boxed{\text{contrast}}$ ① $\hat{\beta}_1 = 0$. $\quad \hat{\beta}_2 \rightarrow \beta_2-\beta_1$. $\quad lm(\cdot)$

② $\boxed{\hat{\beta}_1+\hat{\beta}_2 = 0}$.

$\mu+\beta_1 \longleftrightarrow EY_1$ $\quad 1^{st}$ drug

$\mu+\beta_2 \longleftrightarrow EY_2$ $\quad 2^{nd}$ drug.

$$\left\{\begin{array}{ll} \hat{\mu} & \mu+\beta_1+\frac{\beta_2-\beta_1}{2} = (1, \tfrac{1}{2}, \tfrac{1}{2}) \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix}. \\ \hat{\beta}_1 & -\frac{\beta_2-\beta_1}{2} \\ \hat{\beta}_2 & \frac{\beta_2-\beta_1}{2} \end{array}\right\} \quad \hat{\beta}_1+\hat{\beta}_2 = 0. \quad \begin{array}{l} (0 \;\; \tfrac{1}{2} \;\; -\tfrac{1}{2}) \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix} \\ (0 \;\; -\tfrac{1}{2} \;\; \tfrac{1}{2}) \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix}. \end{array}$$