Parametrization of Factor Effects

Binary Predictor

y_i = # of bike trips on day i
x_i = indicator of rain on day $i-1$, $x_i \in \{0,1\}$

Linear Model

$E[Y_i | X_i] = \beta_0 + \beta_1 X_i \Rightarrow E[Y_i | X_i = 0] = \beta_0$ = expected trips under no rain $= \mu_0$

$E[Y_i | X_i = 1] = \beta_0 + \beta_1$ = exp. trips under rain $= \mu_1$

$$\begin{aligned} \beta_0 &= \mu_0 \\ \beta_0 + \beta_1 &= \mu_1 \end{aligned} \Rightarrow \mu_1 - \mu_0 = \beta_0 + \beta_1 - \beta_0 = \underline{\beta_1}$$

β_1 = difference in mean = "contrast" of μ_1, μ_0.

OLS estimation: given $\underset{\sim}{y}, \underset{\sim}{x}$, obtain $\begin{cases} \hat{\mu}_0, \hat{\mu}_1 \\ \hat{\beta}_0, \hat{\beta}_1 \end{cases}$

Guess: $\hat{\mu}_0 = \text{mean}(y[x==0]) = \bar{y}_0$
$\hat{\mu}_1 = \text{mean}(y[x==1]) = \bar{y}_1$

$$\left\{ \begin{aligned} \beta_0 &= \mu_0 \\ \beta_1 &= \mu_1 - \mu_0 \end{aligned} \right\} \text{ suggests } \left[\begin{aligned} \hat{\beta}_0 &= \hat{\mu}_0 = \bar{y}_0 \\ \hat{\beta}_1 &= \hat{\mu}_1 - \hat{\mu}_0 = \bar{y}_1 - \bar{y}_0 \end{aligned} \right.$$

Are these the OLS estimates of $\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$?

<u>OLS Estimation:</u> $\hat{\beta} = \begin{pmatrix}\hat\beta_0 \\ \hat\beta_1\end{pmatrix} = (X^TX)^{-1}X^TY$

$$X = \begin{bmatrix}1 & 1\\ 1 & 0\\ \vdots & 0\\ \vdots & 1\\ \vdots & 0\\ 1 & 1\end{bmatrix} \qquad X^TY = \begin{bmatrix}1 & \cdots & \cdots & 1\\ 1 & 0 & 0 \cdots 0 & 1\end{bmatrix}\begin{bmatrix}Y_1\\ \vdots\\ \vdots\\ Y_n\end{bmatrix} = \begin{pmatrix}\sum Y_i\\ \sum Y_i X_i\end{pmatrix} \equiv \begin{bmatrix}S\\ S_1\end{bmatrix}$$

$S_1 = \text{SUM}(Y[X==1])$
$S = \text{SUM}(Y)$

$$(X^TX) = \begin{bmatrix}1 & \cdots & \cdots & 1\\ 1 & 0 & 0 \cdots 0 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ \vdots & 0\\ \vdots & 0\\ \vdots & 1\\ \vdots & 0\\ 1 & 1\end{bmatrix} = \begin{bmatrix}n & n_1\\ n_1 & n_1\end{bmatrix}$$

$$(X^TX)^{-1} = \frac{1}{nn_1 - n_1^2}\begin{bmatrix}n_1 & -n_1\\ -n_1 & n\end{bmatrix} = \frac{1}{\underbrace{n_1}_{\text{\# of 1's in X}}\underbrace{(n-n_1)}_{\text{\# of zeros}}}\begin{bmatrix}n_1 & -n_1\\ -n_1 & n\end{bmatrix}$$

$$\begin{pmatrix}\hat\beta_0\\ \hat\beta_1\end{pmatrix} = (X^TX)^{-1}X^TY = \frac{1}{n_1 n_0}\begin{bmatrix}n_1 S - n_1 S_1\\ n S_1 - n_1 S\end{bmatrix}$$

$$\Rightarrow \hat\beta_0 = \frac{1}{n_1 n_0}(n_1 S - n_1 S_1) = \frac{\cancel{n_1}}{\cancel{n_1} n_0}(S - S_1) = \frac{S_0}{n_0} = \underline{\bar{Y}_0} \quad \checkmark$$

$$\hat\beta_1 = \frac{1}{n_1 n_0}(nS_1 - n_1 S) = \frac{1}{n_1 n_0}\left((n-n_1)S_1 - n_1(S - S_1)\right)$$

$$= \frac{n_0}{n_1 \cancel{n_0}} S_1 - \frac{S - S_1}{n_0} = \underline{\bar{Y}_1 - \bar{Y}_0} \quad \checkmark$$

③

Alternative Parametrizations

Let $x_i = \begin{cases} +1/2 & \text{if rain} \\ -1/2 & \text{if no rain} \end{cases}$

$$E[Y \mid X] = \alpha_0 + \alpha_1 X = \alpha_0 + \alpha_1/2 \quad \text{under rain} = \mu_1$$
$$= \alpha_0 - \alpha_1/2 \quad \text{under no rain} = \mu_0$$

$$\Rightarrow \quad \frac{\mu_1 + \mu_0}{2} = \frac{\alpha_0 + \alpha_0}{2} = \underline{\underline{\alpha_0}}$$

$$\mu_1 - \mu_0 = \frac{\alpha_1}{2} + \frac{\alpha_1}{2} = \underline{\underline{\alpha_1}}$$

$$X = \begin{pmatrix} 1 & 1/2 \\ 1 & -1/2 \\ 1 & 1/2 \\ \vdots & \vdots \\ 1 & -1/2 \\ & 1/2 \end{pmatrix} \qquad X^T y = \begin{pmatrix} S \\ 1/2\,(S_1 - S_0) \end{pmatrix} \qquad \begin{pmatrix} \hat\alpha_0 \\ \hat\alpha_1 \end{pmatrix} = (X^T X)^{-1} X^T y$$

<u>Guess</u>: $\begin{pmatrix} \hat\alpha_0 \\ \tilde\alpha_1 \end{pmatrix} = \begin{pmatrix} (\bar y_1 + \bar y_0)/2 \\ (\bar y_1 - \bar y_0) \end{pmatrix}$ Check!

Let $x_1 = 1$ if rain, 0 else
$x_2 = 1$ if no rain, 0 else

$$E(y \mid x_1, x_2) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$$
$$= \gamma_0 + \gamma_1 \quad \text{under rain}$$
$$= \gamma_0 + \gamma_2 \quad \text{under no rain}$$

$$\Rightarrow \quad X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$\hat\gamma = (X^T X)^{-1} X^T y$$

$$(X^T X) = \begin{bmatrix} n & n_1 & n_0 \\ n_1 & n_1 & 0 \\ n_0 & 0 & n_0 \end{bmatrix} \begin{matrix} - r_1 \\ - r_2 \\ - r_3 \end{matrix} \qquad r_3 = r_1 - r_2$$

<u>Rank deficient design matrix!</u>

<u>Model is overparametrized!</u>

<u>not invertible!</u>

Try instead x_1 = rain indicator
x_0 = no rain indicator

$$E(Y \mid x_1, x_2) = \mu_0 x_0 + \mu_1 x_1 = \begin{cases} \mu_0 & x_0 = 1,\ x_1 = 0 \\ \mu_1 & x_1 = 1,\ x_0 = 0 \end{cases}$$

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & \vdots \\ \vdots & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad X^T X = \begin{pmatrix} n_0 & 0 \\ 0 & n_1 \end{pmatrix} \qquad (X^T X)^{-1} = \begin{pmatrix} 1/n_0 & 0 \\ 0 & 1/n_1 \end{pmatrix} \qquad X^T y = \begin{pmatrix} n_0 \bar{y}_0 \\ n_1 \bar{y}_1 \end{pmatrix}$$

$$(X^T X)^{-1} X^T y = \begin{pmatrix} \bar{y}_0 \\ \bar{y}_1 \end{pmatrix}$$

$$\left.\begin{array}{ll} E[y \mid x] = \beta_0 + \beta_1 x, & x \in \{0,1\} \\ E(y \mid x] = \alpha_0 + \alpha_1 x, & x \in \{-1/2, 1/2\} \\ E(y \mid x\} = \mu_0 x_0 + \mu_1 x_1, & x_0 \in \{0,1\},\ x_1 \in \{0,1\},\ x_1 x_0 = 0 \end{array}\right\} \text{Three } \underline{\text{parameterizations}} \text{ of the } \underline{\text{same}} \text{ mean model.}$$

All three models imply the same mean model for y:
Two unrelated means, under each of the two x-conditions.

"Set to zero" parameterization: $E(y \mid x) = \beta_0 + \beta_1 x$

$\beta_1 x$ = "effect" of x: Under one of the conditions (no rain), this is zero. (R-default)

"Sum to zero" parameterization: $E(Y \mid x) = \alpha_0 + \alpha_1 x$

Effect under rain $= \frac{\alpha_1}{2}$
no rain $= -\alpha_1/2$

rain effect + no rain effect = 0

"ANOVA" parameterization: $E(y \mid x) = \mu_x$