

# STAT 151A Lecture 17

Henry Liev

6 October 2025

## Remark 0.1 (Binary Variables)

$x \in \{0, 1\}$

We can relationship between  $Y$  and  $X$  by just calculating mean of  $Y$  when  $x = 0$  and mean of  $Y$  when  $x = 1$

$n_m = \bar{Y}_{male}, n_f = \bar{Y}_{female}$

$X = (\vec{M} \ \vec{F})$ ,  $\vec{M}$  binary vector of male or not,  $\vec{F}$  binary vector of female or not, project  $Y$  onto  $\text{Col}(X)$  build  $\mathbf{X}\hat{\beta}, \hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y$

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} n_m & 0 \\ 0 & n_f \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n_m} & 0 \\ 0 & \frac{1}{n_f} \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} \bar{Y}_m \\ \bar{Y}_f \end{pmatrix}$$

## Remark 0.2 (Alternative Design Matrix)

$(\vec{1} \ \vec{F})$  New  $\hat{\beta}$

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} n & n_f \\ n_f & n_f \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{nn_f - n_f^2} \begin{pmatrix} n_f & -n_f \\ -n_f & n \end{pmatrix}$$

$$\mathbf{X}^\top Y = \begin{pmatrix} \sum y_i \\ \sum_{i \in f} y_i \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y = \begin{pmatrix} \bar{y}_m \\ \bar{y}_f - \bar{y}_m \end{pmatrix}$$

## Remark 0.3

Baseline:  $y \sim 1$ ,  $y$  on intercept  $\mathbf{X} = [\vec{1}]$

Gender:  $y \sim \text{gender}$   $\mathbf{X} = [\vec{1} \ \vec{F}]$

Rate  $y \sim \text{rate}$   $\mathbf{X} = [\vec{1} \ \vec{r}]$

Parallel lines model (Gender + rate)  $y \sim \text{gender} + \text{rate}$   $\mathbf{X} = [\vec{1} \ \vec{F} \ \vec{r}]$

Non-parallel lines model (Gender  $\times$  rate)  $y \sim \text{gender} * \text{rate}$   $\mathbf{X} = [\vec{1} \ \vec{F} \ \vec{r} \ \vec{r} + \vec{F}]$

Principle of marginality: If you included an ??, include the main effects too