Example 2 (Berkeley Admission (Bickel et al., 1975)) *Potential students apply for admission to UC Berkeley. During application, students choose one of the six departments to which they apply, denoted by D. The admission decision is labelled as Y and the student's gender is labelled as X (x_0 male, x_1 female). A possible SCM $\mathcal{M}$ describing the situation might be*

$$\begin{aligned} X &\leftarrow \mathbb{1}(U_X > 0) \\ D &\leftarrow 1 + \lfloor U_D + 0.5 + \lambda X \rfloor \\ Y &\leftarrow \mathbb{1}(U_Y > \Phi^{-1}(0.1 + \alpha X + \beta D)), \end{aligned} \tag{4}$$

where $U_X \sim N(0,1)$, $U_D \sim Unif(0,5)$, $U_Y \sim N(0,1)$ and Φ is the cumulative distribution of a standard normal random variable. After simplification, the SCM can also be written as

$$\begin{aligned} X &\leftarrow Bernoulli(0.5) \\ D &\leftarrow Multionomial\big(1, (\frac{1}{10} + \lambda X, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{10} - \lambda X)\big) \\ Y &\leftarrow Bernoulli(0.1 + \alpha X + \beta D). \end{aligned} \tag{5}$$

Example 3 (COMPAS (Larson et al., 2016)) *The courts at Broward County, Florida, use machine learning tools to predict whether individuals released on parole are at high risk of re-offending within 2 years. The algorithm is based on the demographic information Z, race X (x_0 denoting White, x_1 Non-White), juvenile offense counts J, prior counts P and degree of charge D. After a period of using the algorithm, it is observed that Non-White individuals are 9% more likely to be classified as high-risk*

$$P(y \mid x_1) - P(y \mid x_0) = 9\%. \tag{6}$$

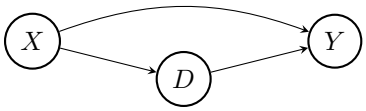
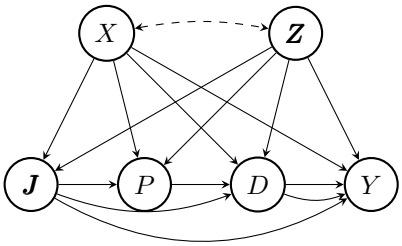
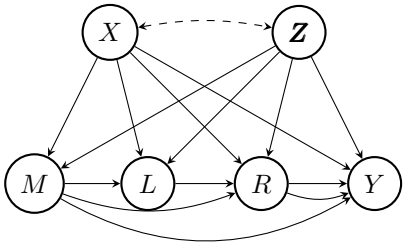
Does this mean that racial minorities are discriminated by the legal justice system?

Example 4 (US Government Census 2018) *The United States Census of 2018 collected information on the employees of the US Government. The information includes demographic information Z, gender X (x_0 male, x_1 female), marital and family status M, education information L and work-related information R. It is observed that male employees of the government on average earn 14 000 \$/year more than female employees, that is*

$$P(y \mid x_1) - P(y \mid x_0) = -14000\$. \tag{7}$$

Does this mean that the US Government discriminates against its female employees?

Table 1: Motivating examples: Berkeley admission dataset, COMPAS and Census 2018.

	Observed disparity	Causal graph	SCM
Berkeley	$P(Y \mid x_1) - P(Y \mid x_0) = -0.142$		$X \leftarrow \text{Bernoulli}(0.5)$ $D \leftarrow f_D(X, U_D)$ $Y \leftarrow f_Y(X, D, U_Y)$
COMPAS	$P(Y \mid x_1) - P(Y \mid x_0) = 0.086$		$X \leftarrow \text{Bernoulli}(0.5)$ $Z \leftarrow N(\mu, \sigma^2)$ $J \leftarrow f_J(X, Z, U_J)$ $P \leftarrow f_P(X, Z, J, U_P)$ $D \leftarrow f_D(X, Z, J, P, U_D)$ $Y \leftarrow f_Y(X, Z, J, P, D, U_Y)$
Census 2018	$\mathbb{E}[Y \mid x_1] - \mathbb{E}[Y \mid x_0] = -14297$		$X \leftarrow \text{Bernoulli}(0.5)$ $Z \leftarrow N(\mu, \sigma^2)$ $M \leftarrow f_M(X, Z, U_M)$ $L \leftarrow f_L(X, Z, M, U_L)$ $R \leftarrow f_R(X, Z, M, L, U_R)$ $Y \leftarrow f_Y(X, Z, M, L, R, U_Y)$