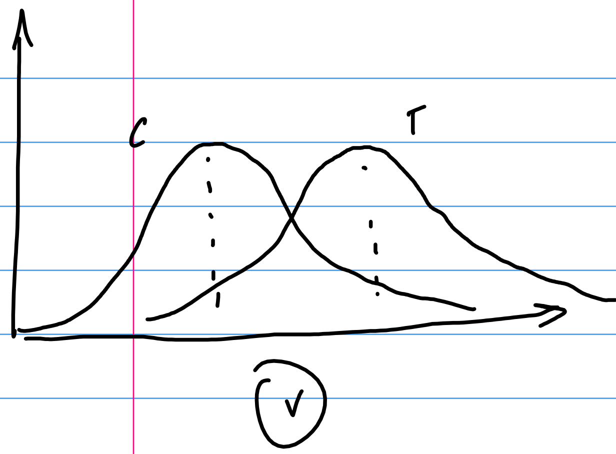
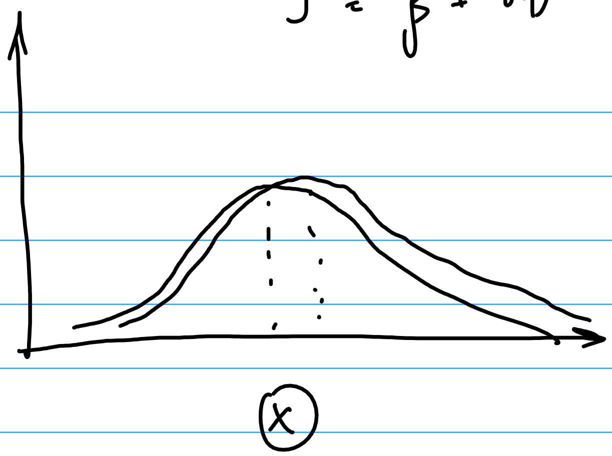
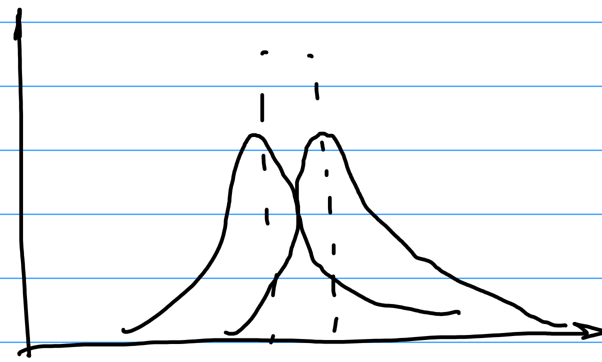
Variance: stratification

$y = \beta + \gamma \cdot D$







1. Diff - in - difference
2. CUPED (Controlled experiments using pre-experiment data)

Naive approach : $\Delta = \bar{y}_1 - \bar{y}_0$ $\qquad$ $y = \beta + \gamma \cdot D_T + \varepsilon$

Example : y - employment rate

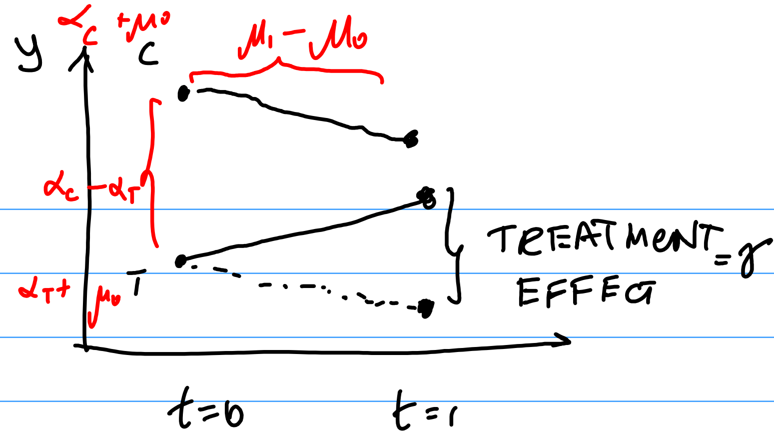
$t = 1$: policy - increasing min wage

1. $\Delta = \bar{y}_1 - \bar{y}_0$ $\qquad$ other factors changing over time aren't accounted

\+

2. $\Delta = \bar{y}_T - \bar{y}_C$ $\qquad$ selection Bias

"Diff - in - diff"



$$E(y \mid D=T, t=0) = \mu_0 + \alpha_T$$

$$E(y \mid D=T, t=1) = \mu_1 + \alpha_T + \delta$$

unknown TE

$$\Delta_T = \mu_1 - \mu_0 + \delta$$

$$E(y \mid D=C, t=0) = \mu_0 + \alpha_C$$

$$E(y \mid D=C, t=1) = \mu_1 + \alpha_C$$

$$\Delta_C = \mu_1 - \mu_0$$

$$\Delta_T - \Delta_C = \delta$$

$$\Delta y = \beta + \delta \cdot D_T + \varepsilon$$

2. CUPED

X - covariate

1) $y^{CUPED} = y - \theta \cdot X$

Frisch - Waugh

$$E(y) = E(\bar{y}) =$$

$y \mid D, X \quad \hat{\theta}$

$$E(\bar{y} - \theta \cdot X) + \theta E(X) =$$

1) $y \mid X \Rightarrow \tilde{y}$

2) $\tilde{y} \mid D \Rightarrow \hat{\theta}$

$$E(\bar{y} - \theta \cdot \bar{X}) + \theta E(X)$$

$$\tilde{y}^c = \bar{y} - \theta \cdot \bar{X} + \theta \cdot E(X) \rightarrow \min_{\theta}$$

$$\theta = \frac{cov(X, y)}{Var(X)} \qquad \hat{\theta} = \frac{\hat{cov}(X, y)}{\hat{var}(X)}$$

2) $$Var(\tilde{y}) = (1 - \rho^2) \cdot Var(y) \leq Var(y)$$

$$y = \alpha + \underbrace{\theta \cdot D}_{\text{var exp.}} + \underbrace{\beta \cdot X}_{\text{var ind}} + \underbrace{\varepsilon}_{\text{var unexplained}}$$

Generalized :

$$y = \alpha + \theta_1 \cdot D_{T_1} + \dots + \theta_k \cdot D_{T_k} +$$

$$+ \beta_1 \cdot X_1 + \dots + \beta_k \cdot X_k + \varepsilon$$

Stratification = CUPED with X- categorical

$$\mathrm{Var}(\bar{y}) = \underbrace{\sum w_k \cdot \frac{\sigma_k^2}{n}}_{\text{within}} + \underbrace{\sum w_k \cdot \frac{(\mu_k - \mu)^2}{n}}_{\text{between}} \geq$$

$$\sum w_k \cdot \frac{\sigma_k^2}{n} = \mathrm{Var}(\bar{y}_{strat})$$

$$\hat{y}_{strat} = w \cdot \bar{y}_1 + (1-w)\,\bar{y}_0 \qquad X = \begin{cases} 1, & w \\ 0, & 1-w \end{cases}$$

$$\hat{y}_{cup} = \bar{y} - \hat{\theta}\bar{X} + \hat{\theta} \cdot \underset{\substack{\| \\ w}}{E(X)}$$

$$\hat{\theta} = \frac{\hat{\mathrm{cov}}(y, x)}{\hat{\mathrm{Var}}(x)} = \underline{\bar{y}_1 - \bar{y}_0}$$

$$\theta = \frac{E(X \cdot y) - E(X) \cdot E(y)}{w \cdot (1-w)} =$$

$$= \frac{E(x^2 \cdot y_1) + E(x(1-x) \cdot y_1) - E(x) \cdot (E(x \cdot y_1) - E(1-x)y_0)}{w \cdot (1-w)} =$$

$$= \frac{\cancel{w^2 \cdot E(y_1)} + w \cdot (\cancel{1}-w) \cdot E(y_1) - \cancel{w^2 E(y_1)} + w(\cancel{1}-w) \cdot y_0}{w \cdot (\cancel{1}-w)}$$

$$= E(y_1) - E(y_0)$$

$$\bar{y} = X \cdot \bar{y}_1 + (1-X) \cdot \bar{y}_0$$

$$\hat{y}_{cv} = \bar{y} - (\bar{y}_1 - \bar{y}_0) \cdot \bar{x} + (\bar{y}_1 - \bar{y}_0) w =$$

$$\dots = w \cdot \bar{y}_1 + (1 - w) \cdot \bar{y}_0 = \hat{y}_{strat}$$