Reject: if p-value is small (α), CI: Collects all possible μ_0's that we can't reject in this way. Contains true μ $(1-\alpha)\%$ of the time

P-Value: "How likely it would be to observe estimator $\hat{\mu}$ as extreme than the one observed in the sample if null is true $H_0: \mu=\mu_0$

Estimator: Function of data for the sample: What we actually see (random). Estimand: Function of distribution of observable data (fixed number)

Let $D_i \begin{cases} 1, \text{treatment} \\ 0, \text{r-t.} \end{cases}$ $\forall i$, $Y_i(1)$ = outcome if treated, $Y_i(0)$ = outcome if r-t. $\therefore Y_i(1)-Y_i(0)$ individual i's treatment-effect $\Longrightarrow Y_i = D_i Y_i(1) + (1-D_i) Y_i(0) \therefore$ model (Y_i, D_i)

Target Parameter: $\underset{\text{(ATE)}}{E[Y_i(1)-Y_i(0)]}$. Sample $\{Y_i, D_i\}^N$. Estimator: $\frac{1}{N(1)}\sum_{i:D_i=1} Y_i - \frac{1}{N(0)}\sum_{i:D_i=0} Y_i$. Estimand: $E[Y_i|D_i=1]-E[Y_i|D_i=0]$ Compare to parameter $E[Y_i(1)]-E[Y_i(0)]$ $E[X]=\pi$, $Var(X)=\pi(1-\pi)$

CDF: $F(x)=P(X\le x)$. X discrete: $p(x)=P(X=x)$, $\sum_{x\in\mathcal{X}} p(x)=1$, continuous $f(x)=\frac{\partial}{\partial x}F(x)$, $f(x)\ge 0$, $\int_{x\in\mathcal{X}} f(x)dx=1$, Bernoulli: $X\in\{0,1\}$, $X\sim \text{Bernoulli}(\pi)$, $p(x)=\pi^x(1-\pi)^{1-x}$

$X\in\mathbb{R}$, $X\sim N(\mu,\sigma^2)$ $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}$ $\forall \alpha,\beta, (\alpha X+\beta)\sim N$. $X\in(a,b)$, $X\sim U(a,b)$, $f(x)=\frac{1}{b-a}$. Joint CDF: $F(x,y)=P(X\le x, Y\le y)$, pmf: $f(x)=P(X=x, Y=y)$

pdf: $f(x,y)=\frac{\partial^2}{\partial x\partial y}F(x,y)$. Marginal D. $p(x)=\sum_{y\in\mathcal{Y}} p(x,y)$ Conditional: $P(y|x)=\frac{P(Y=y, X=x)}{P(X=x)}=\frac{p(y,x)}{p(x)}$. Bayes: $p(y|x)=\frac{p(x|y)p(y)}{p(x)}$. $(X \perp Y)$ if $p(y|x)=p(y)\ \forall(y,x)$

Conditional I. $Y\perp X|W$ if $p(y|x,w)=p(y,w)\ \forall(y,x,w)$. Multivariate Normal: $x\in\mathbb{R}^k$, $\mu\in\mathbb{R}^k$, positive-definite $\Sigma\in\mathbb{R}^{k\times k}$, (X,Y), $X|Y$, $Y|X$

$\alpha X+\beta Y+c\sim N$. Mean: $E[X]=\sum_{x\in\mathcal{X}} x\,p(x)=\int_{x\in\mathcal{X}} x\cdot f(x)dx$, $E(a+bx)=a+bE(x)$, $Var(X)=E((X-E(X))^2)=E(X^2)-E(X)^2$, $Var(a+bX)=b^2Var(X)$, $Cov(X,Y)=E[(X-E(X))(Y-E(Y))]$

$Cov(X,Y)=E(XY)-E(X)E(Y)$, $Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$, $X\perp Y\Longrightarrow Cov(X,Y)=0$, $Var(aX+bY+c)=a^2Var(X)+2abCov(X,Y)+b^2Var(Y)$, $Cov(aX+c, bY+d)=abCov(X,Y)$

$Cov(X+Z,Y)=Cov(X,Y)+Cov(Z,Y)$. Conditional Exp: $E[Y|X=x]=\sum_{y\in\mathcal{Y}} y\,p(y|x)\equiv\int_{y\in\mathcal{Y}} y f(y|x)dy$, $E(f(X)+g(X)Y|X=x)=f(x)+g(x)E[Y|X=x]\ \forall f(\cdot),g(\cdot)$ Law of I.E.

(LIE): $E(Y)=E(E(Y|X))$, Mean Ind: Y m.i. of X: $E[Y|X=x]=E[Y]\ \forall x\in\mathcal{X}\Longrightarrow Corr(X,Y)=0$, $E\begin{bmatrix}X_{11}&\cdots&X_{1n}\\ \vdots&\ddots&\vdots\\ X_{n1}&\cdots&X_{nn}\end{bmatrix}=\begin{bmatrix}E(X_{11})&\cdots&E(X_{1n})\\ \vdots&&\vdots\\ E(X_{n1})&\cdots&E(X_{nn})\end{bmatrix}$, $Var\left(\begin{pmatrix}X_1\\X_2\end{pmatrix}\right)=$

$\begin{bmatrix}Var(X_1)&Cov(X_1,X_2)\\Cov(X_1,X_2)&Var(X_2)\end{bmatrix}$, $Cov\left(\begin{pmatrix}X_1\\X_2\end{pmatrix},\begin{pmatrix}Y_1\\Y_2\end{pmatrix}\right)=\begin{bmatrix}Cov(X_1,Y_1)&Cov(X_1,Y_2)\\Cov(X_2,Y_1)&Cov(X_2,Y_2)\end{bmatrix}$. Average Treatment Effect: ATE $=\overset{\tau}{E[Y_i(1)-Y_i(0)]}=\left(\text{If } D_i\perp(Y_i(1),Y_i(0))\right)=E[Y_i|D_i=1]-E[Y_i|D_i=0]$

Conditional Unconfoundedness $D_i\perp(Y_i(1),Y_i(0))|X_i$ $\therefore E[Y_i|D_i=1,X_i=x]=\underbrace{E[Y_i(1)|X_i=x]}_{\text{CATE(x)}}$ and $E[Y_i|D_i=0,X_i=x]=E[Y_i(0)|X_i=x]\Longrightarrow$ CATE(x)

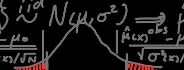
ATE(x) identified

$E[Y_i|D_i=1|X_i=x]-E[Y_i|D_i=0,X_i=x]=E[Y_i(1)-Y_i(0)|X_i=x]$ identified, $E(E[Y_i(1)-Y_i(0)|X_i])=E[Y_i(1)-Y_i(0)]$. Population Mean: $\{Y_i\}^N\overset{iid}{\sim}D$, $\mu=E[Y_i]$

Natural estimator: $\hat{\mu}=\frac{1}{N}\sum_i Y_i$ (random variable), $E[\hat{\mu}]=E[\frac{1}{N}\sum_i Y_i]=\mu$, $Var(\hat{\mu})=Var(\frac{1}{N}\sum_i Y_i)=\frac{\sigma^2}{N}$, Unbiased(i.e.) $E[\hat{\mu}]=\mu$. $\lim_{N\to\infty} Var(\hat{\mu})=0$

Null Hypothesis: H_0: $\mu=\mu_0$. P-value: reject H_0 if small (i.e.) Unlikely to observe $\hat{\mu}$ far from μ_0 if H_0:T. $\therefore$ p-value $=2\left(1-\Phi\left(\frac{|\hat{\mu}^{obs}-\mu_0|}{\sigma/\sqrt{N}}\right)\right)$ $\uparrow$

$\{Y_i\}^N\overset{iid}{\sim}N(\mu,\sigma^2)\Longrightarrow\hat{\mu}\sim N(\mu,\frac{\sigma^2}{N})$. $\hat{t}=\frac{\hat{\mu}-\mu_0}{\sigma/\sqrt{N}}\overset{\text{(under } H_0)}{\sim}N(0,1)\,|H_0$. $P(|\hat{t}|>t)=1-P(-t\le\hat{t}\le t)=1-(\Phi(t)-\Phi(-t))$. p-value $=1-(\Phi(|\hat{t}^{obs}|)-\Phi(-|\hat{t}^{obs}|))$



$p<0.05$ reject at 5%. 2.5% $\therefore$ H_0 if $|\hat{t}^{obs}|>1.96$, Don't reject if $|\hat{t}^{obs}|\le 1.96\Longrightarrow\frac{|\hat{\mu}^{obs}-\mu_0|}{\sigma/\sqrt{N}}\le 1.96\Longrightarrow\mu_0\in[\hat{\mu}^{obs}-1.96\sigma/\sqrt{N}, \hat{\mu}^{obs}+\dots]\triangleq$ CI: $P(\mu_0\in CI)=0.95$

Significance (size): P incorrectly rejecting H_0 when T. (Type I error). Power: P correctly rejecting H_0 when false (1 - Type II error) LLN: $N\to\infty$, $\hat{\mu}\xrightarrow{p}\mu$

CLT: $N\to\infty$, $\hat{\mu}\sim D\xrightarrow{d}N(\mu,\frac{\sigma^2}{N})$. CMT: $N\to\infty$, $g(\hat{\mu})$ continuous $\longrightarrow g(\mu)$. Fact: $E[(X_N-x)^2]\xrightarrow{}0\Longrightarrow X_N\xrightarrow{p}x$, Chebychev's: $P(|X_N-x|>\varepsilon)\le\frac{E[(X_N-x)^2]}{\varepsilon^2}$

X_N converges in P. to x, $X_N\xrightarrow{p}x$ if $\forall\varepsilon>0$ $P(|X_N-x|>\varepsilon)\to 0$. X_N consistent if $X_N\to\mu$. Proof: $E[(X_N-x)^2]=P(|X_N-x|>\varepsilon)E((X_N-x)^2||X_N-x|>\varepsilon)+$

$P(|X_N-x|\le\varepsilon)E[(X_N-x)^2||X_N-x|\le\varepsilon]\ge P(|X_N-x|>\varepsilon)\varepsilon^2$. LLN: $\{Y_i\}^N$ iid $\sim D$, $Var(Y_i)=\sigma^2$ $\therefore$ $\hat{\mu}_N=\frac{1}{N}\sum Y_i\xrightarrow{p}\mu=E[Y_i]$. $P(|\hat{\mu}_N-\mu|>\varepsilon)\to 0$, $P(|\hat{\mu}_N-\mu|\le\varepsilon)\to 1$

Proof: $E[\hat{\mu}_N]=\mu$, $Var[\hat{\mu}_N]=\frac{\sigma^2}{N}$ $\therefore$ $Var(\hat{\mu}_N)=E[(\hat{\mu}_N-\mu)^2]=\frac{\sigma^2}{N}\to 0\Longrightarrow\hat{\mu}_N\xrightarrow{p}\mu$. | X_N converges in distribution to a continuously distributed variable X,

i.e. $X_N\xrightarrow{d}X$ if $F_{X_N}(x)\xrightarrow{\text{pointwise}}F_X(x)\ \forall x$ $\therefore$ $P(X_N\le x)\to P(X\le x)$. CLT: $\{Y_i\}^N\overset{iid}{\sim}D$, $E[Y_i]=\mu$, $Var(Y_i)=\sigma^2<\infty$ $\therefore$ $\hat{\mu}=\frac{1}{N}\sum^N Y_i$, $X_N=\sqrt{N}(\hat{\mu}-\mu)\xrightarrow{d}N(0,\sigma^2)$

$\Longleftrightarrow\hat{\mu}\xrightarrow{d}N(\mu,\frac{\sigma^2}{N})$. $\vec{X}_N\in\mathbb{R}^k$, $\vec{X}_N\xrightarrow{d}\vec{X}$ if $F_{\vec{X}_N}(\vec{x})\to F_{\vec{X}}(\vec{x})$, $F_{\vec{X}_N}(\vec{x})=P(X_N^1\le x_1,\dots,X_N^k\le x_k)$, $x=\begin{pmatrix}x_1\\ \vdots\\ x_k\end{pmatrix}$, CMT: $g(\cdot)$ continuous, (1) $X_N\xrightarrow{p}X\Longrightarrow$

$g(X_N)\xrightarrow{p}g(X)$, (2) $X_N\xrightarrow{d}X\Longrightarrow g(X_N)\xrightarrow{d}g(X)$. Let $\hat{\sigma}^2=\frac{1}{N}\sum^N(Y_i-\hat{\mu})^2$. If $\{Y_i\}^N$, $\hat{\sigma}^2\xrightarrow{p}\sigma^2=Var(Y_i)$. Proof: $\hat{\sigma}^2=\frac{1}{N}\sum^N Y_i^2-\hat{\mu}^2$. $Z_i=Y_i^2$, $\frac{1}{N}\sum^N Z_i\xrightarrow{p}E[Z_i]$

$\therefore$ By LLN: $\frac{1}{N}\sum^N Y_i^2\xrightarrow{p}E[Y_i^2]$, $\hat{\mu}\xrightarrow{p}\mu=E[Y_i]$ $\therefore$ By CMT: $\hat{\mu}^2\xrightarrow{p}E[Y_i]^2$ and $\frac{1}{N}\sum^N Y_i^2-\hat{\mu}^2\xrightarrow{p}E[Y_i^2]-E[Y_i]^2=\sigma^2$. Slutky's: $X_N\xrightarrow{p}c\in\mathbb{R}$, $Y_N\xrightarrow{d}Y\Longrightarrow$

$X_N+Y_N\xrightarrow{d}c+Y$, $X_NY_N\xrightarrow{d}cY$, $Y_N/X_N\xrightarrow{d}\frac{Y}{c}$ ($c\ne 0$). By Slutky's, $\hat{t}=\frac{\hat{\mu}-\mu_0}{\hat{\sigma}/\sqrt{N}}\xrightarrow{d}N(0,1)$ $\therefore$ $P(\mu\in[\hat{\mu}\pm 1.96\frac{\hat{\sigma}}{\sqrt{N}}])=0.95$ ($N\to\infty$). Hypothesis T. Exp: $D_i\perp(Y_i(1),Y_i(0))$

$\tau=E[Y_i(1)-Y_i(0)]=E[Y_i|D_i=1]-E[Y_i|D_i=0]$. H_0: $\tau=\tau_0$. $\bar{Y}_1=\frac{1}{N(1)}\sum_{i:D_i=1}Y_i$, $\bar{Y}_0=\frac{1}{N(0)}\sum_{i:D_i=1}Y_i$ $\therefore$ $\begin{pmatrix}\sqrt{N_1}(\bar{Y}_1-E(Y_i(1)))\\ \sqrt{N_0}(\bar{Y}_0-E(Y_i(0)))\end{pmatrix}\xrightarrow{d}N(0,\Sigma)$, $\Sigma=\begin{pmatrix}Var(Y_i(1))&0\\0&Var(Y_i(0))\end{pmatrix}$, $N=N_1+N_2$

$\frac{N_1}{N}=\frac{1}{N}\sum D_i\xrightarrow{p}E[D_i]$, $\frac{N_0}{N}\xrightarrow{p}E[1-D_i]$ $\therefore$ $\sqrt{N}(\bar{Y}_1-\bar{Y}_0-E[Y(1)-Y(0)])\xrightarrow{d}N(0,\sigma^2)$, $\sigma^2=\frac{1}{E[D_i]}Var(Y_i(1))+\frac{1}{E[1-D_i]}Var(Y_i(0))$, $\hat{\sigma}^2=\frac{N}{N_1}\hat{\sigma}_1^2+\frac{N}{N_0}\hat{\sigma}_0^2$

$D_i\perp(Y_i(1),Y_i(0))|X_i$ $\therefore$ $\sqrt{N_x}(\bar{Y}_{1,x}-\bar{Y}_{0,x}-E[Y_i(1)-Y_i(0)|X_i=x])\xrightarrow{d}N(0,\sigma_x^2)$, $N_x=|i:X_i=x|$, $\sigma_x^2=\frac{1}{E[D_i|X_i=x]}Var(Y_i(1)|X_i=x)+\frac{1}{E[1-D_i|X_i=x]}Var(Y_i(0)|X_i=x)$

Linear Regression: Assume $\underset{\text{(CEF Linear)}}{E[Y_i|X_i=x]=\alpha+x\beta}$, $(\alpha,\beta)=\underset{\alpha,\beta}{\text{argmin}}\,E[(Y_i-(\alpha+\beta X_i))^2]$, $\min_u E[(Y_i-u)^2]\Longrightarrow E[-2(Y_i-u)]=0$ $\therefore$ $u=E[Y_i]$

$\min_{u(\cdot)} E[(Y_i-u(X_i))^2]=E[E[(Y_i-u(X_i))^2|X_i]]$ $\therefore$ $\forall x\in\mathcal{X}$ $u(x)\Longrightarrow\min E[(Y_i-u(x))^2|X_i=x]\Longrightarrow u(x)=E[Y_i|X_i=x]=\alpha+\beta x$ $\therefore$ $E[(Y_i-(\alpha+\beta X_i))^2]\le E[(Y_i-(a+bX_i))^2]\ \forall a,b$

$\frac{\partial}{\partial a}E[(Y_i-(a+bX_i))^2]=E[-2\underbrace{(Y_i-(a+bX_i))}_{\varepsilon_i}]=0$ (1), $\frac{\partial}{\partial b}E[(Y_i-(a+bX_i))^2]=E[-2X_i(Y_i-(a+bX_i))]=0$ (2), (1)$\Longrightarrow\alpha=E[Y_i]-\beta E[X_i]$, $\beta=\frac{E[X_i(Y_i-E(Y_i))]}{E[X_i(X_i-E(X_i))]}=\frac{E[(X_i-E(X_i))(Y_i-E(Y_i))]}{E[(X_i-E(X_i))^2]}=\frac{Cov(X_i,Y_i)}{Var(X_i)}$

$\beta=\frac{Cov(X_i,Y_i)}{Var(X_i)}$, $\alpha=E[Y_i]-E[X_i]\beta$ $\therefore$ $\hat{\beta}=\frac{\widehat{Cov}(X_i,Y_i)}{\widehat{Var}(X_i)}$, $\widehat{Cov}=\frac{1}{N}\sum^N(X_i-\bar{X})(Y_i-\bar{Y})$, $\widehat{Var}=\frac{1}{N}\sum^N(X-\bar{X})^2$, $\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X}$ OLS: $\min_{a,b}\frac{1}{N}\sum_i(Y_i-(a+bX_i))^2=(\hat{\alpha},\hat{\beta})$, OLS consistency:

$\hat{\beta}=\left(\frac{1}{N}\sum^N X_i^2-\bar{X}^2\right)^{-1}\left(\frac{1}{N}\sum^N X_iY_i-\bar{X}\bar{Y}\right)\xrightarrow{p}(E[X_i^2]-E[X_i]^2)^{-1}(E[X_iY_i]-E[X_i]E[Y_i])=Var(X_i)^{-1}Cov(X_i,Y_i)=\beta$, By LLN, Slutky's $\hat{\alpha}\xrightarrow{p}\alpha$. Regression Residual: ε_i

$\varepsilon_i=Y_i-(\alpha+X_i\beta)$, $E[\varepsilon_i|X_i]=0$, $Y_i=\alpha+X_i\beta+\varepsilon_i$, $E[\varepsilon_i]=0$, $E[\varepsilon_iX_i]=0$, $Cov(X_i,\varepsilon_i)=E[X_i\varepsilon_i]=E[X_i\cdot E[\varepsilon_i|X_i]]=0$, $\bar{Y}=\alpha+\bar{X}\beta+\bar{\varepsilon}$ $\therefore$ $Y_i-\bar{Y}=(X_i-\bar{X})\beta+(\varepsilon_i-\bar{\varepsilon})$

$\therefore$ $\hat{\beta}=\left(\frac{1}{N}\sum^N(X_i-\bar{X})^2\right)^{-1}\frac{1}{N}\sum^N(X-\bar{X})((X_i-\bar{X})\beta+(\varepsilon_i-\bar{\varepsilon}))=\beta+\left(\frac{1}{N}\sum^N(X_i-\bar{X})^2\right)^{-1}\frac{1}{N}\sum^N(X_i-\bar{X})(\varepsilon_i-\bar{\varepsilon})$ $\therefore$ $\sqrt{N}(\hat{\beta}-\beta)=\left(\frac{1}{N}\sum^N(X-\bar{X})^2\right)^{-1}\sqrt{N}\cdot\frac{1}{N}\sum^N(X_i-\bar{X})(\varepsilon_i-\bar{\varepsilon})$

$\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{d}N(0,\sigma^2)$, $\sigma^2=\frac{Var((X_i-E(X_i))\varepsilon_i)}{Var(X_i)^2}$, $\hat{\sigma}^2=\frac{\frac{1}{N}\sum_i((X_i-\bar{X})\hat{\varepsilon}_i)^2}{\left(\frac{1}{N}\sum_i(X_i-\bar{X})^2\right)^2}$, $\hat{\varepsilon}_i=Y_i-(\hat{\alpha}+X_i\hat{\beta})$, $\tau=E[Y_i(1)-Y_i(0)]$. $E[Y_i|D_i=d]=\alpha+\beta d$

ATE: $\hat{\beta}=\bar{Y}_1-\bar{Y}_0=\hat{\tau}$. Markov's: $P(X\ge c)\le\frac{E[X]}{c}$, Chebyshev's: $P(|X-E[X]|\ge\varepsilon)\le\frac{Var(X)}{\varepsilon^2}$, Jensen's: $E[g(X)]\ge g(E[X])\Longleftarrow$ convex

Randomization $\Longrightarrow D_i\perp(Y_i(1),Y_i(0))$ $\therefore$ $E[Y_i|D_i=1]=E[Y_i|D_i=1]=E[Y_i(1)|D_i=1]=E[Y_i(1)]$, $E[Y_i|D_i=0]=E[Y_i(0)]$, $E[f(X)+g(X)Y|X=x]=f(x)+g(x)E[Y|X=x]\ \forall f(\cdot),g(\cdot)$

Potential Outcomes $\begin{cases} Y_i(0): \text{Outcome that would be observed for an individual } i \text{ if they do not receive the treatment } (M_i=0) \\ Y_i(1): \ \dots \ (M_i=1) \end{cases}$

Multivariate Regression: $\mathbb{E}[Y_i|\vec{X}_i=x]$, $\underset{(k+1)\times 1}{\vec{X}_i}=(1, x_{i1},\dots,x_{iK})^T$, $\vec{\beta}=\begin{pmatrix}\alpha\\ \beta_1\\ \vdots \\ \beta_K\end{pmatrix}$ $\therefore \vec{X}_i^T\vec{\beta}=\alpha+\beta_1 x_{i1}+\dots+\beta_K x_{iK}$ | **Univariate**: $(\alpha,\beta)=\underset{a,b}{\operatorname{argmin}}\ \mathbb{E}[(Y_i-(a+bX_i))^2]$

Multivariate: $\vec{\beta}=\underset{\vec{b}}{\operatorname{argmin}}\ \mathbb{E}[(Y_i-\vec{X}_i^T\vec{b})^2]$. If **CEF linear**: $\mathbb{E}[Y|\vec{X}_i]=\vec{X}^T\vec{\beta}$, $CEF\neq linear$: $\vec{X}^T\vec{\beta}$ is **MSE-minimizer** approximator

$\vec{\beta}=\mathbb{E}[\vec{X}_i\vec{X}_i^T]^{-1}\mathbb{E}[\vec{X}_iY_i]$, $\hat{\beta}=(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^TY$, $\mathbb{X}\in\mathbb{R}^{n\times k}$, $Y\in\mathbb{R}^{n\times 1}$, $\hat{\beta}=\left(\frac{1}{N}\sum_1^N\vec{x}_i\vec{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_1^N\vec{x}_iY_i\right)$, $\hat{\vec{\beta}}\xrightarrow{p}\vec{\beta}$ (**Consistency**)

Asymptotic Normality: $\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{d}N(0,\Sigma)$, $\Sigma=\mathbb{E}[\vec{X}_i\vec{X}_i^T]^{-1}Var(\vec{X}_i\varepsilon_i)\mathbb{E}[\vec{X}_i\vec{X}_i^T]^{-1}$, $\varepsilon_i=Y_i-\vec{X}_i^T\vec{\beta}$ $\therefore$ $\hat{\Sigma}=\left(\frac{1}{N}\sum_1^N\vec{X}_i\vec{X}_i^T\right)^{-1}\left(\frac{1}{N}\sum_1^N\vec{X}_i\vec{X}_i\hat{\varepsilon}_i^2\right)\cdot\left(\frac{1}{N}\sum_1^N\vec{X}_i\vec{X}_i^T\right)^{-1}$, $\hat{\varepsilon}_i=Y_i-\vec{x}_i^T\hat{\beta}$. $SE(\hat{\beta}_j)=\sqrt{\hat{\Sigma}_{jj}}/\sqrt{N}$. **Frisch-Waugh-Lovell**: $Y_i=\beta_0+X_{i1}\beta_1+X_{i2}\beta_2+\varepsilon_i$. Obtain $\hat{\beta}_2$: (1) Reg $X_{i2}|X_{i1}$ $\therefore X_{i2}=\gamma_0+X_{i1}\gamma_1+u_i$; (2) $\hat{X}_{i2}=\hat{\gamma}_0+X_{i1}\hat{\gamma}_1$; $\hat{u}_i=X_{i2}-\hat{X}_{i2}$; 3) Reg $Y_i|\hat{u}_i$ $\therefore Y_i=\alpha+\underbrace{(X_{i2}-\hat{X}_{i2})}_{\hat{u}_i}\beta_2+v_i$. **Measures-Model-Fit**: $Y_i=\vec{X}_i^T\vec{\beta}+\varepsilon_i$, $R^2=\frac{Var(\vec{X}_i^T\vec{\beta})}{Var(Y_i)}$, $Var(Y_i)=Var(\vec{X}_i^T\vec{\beta})+Var(\varepsilon_i)\Rightarrow R^2=1-\frac{Var(\varepsilon_i)}{Var(Y_i)}$

$\hat{R}^2=\frac{\frac{1}{N}\sum_i(\vec{X}_i^T\hat{\vec{\beta}}-\bar{X}_i\hat{\beta})^2}{\frac{1}{N}\sum_i(Y_i-\bar{Y})^2}=1-\frac{1/N\sum_i\hat{\varepsilon}_i^2}{1/N\sum_i(Y_i-\bar{Y})^2}$. $(Y_{(1)},Y_{(0)})\perp D|X \Rightarrow$ **CATE(x)** $=E[Y_i|D_i=1,\vec{X}_i=x]-E[Y_i|D_i=0,\vec{X}_i=x]=E(Y_{(1)}-Y_{(0)}|\vec{X}=x]$; $\mathbb{E}[Y_i|D_i,\vec{X}_i]\approx D_i\beta+\vec{X}_i^T\vec{\gamma}$

$\Rightarrow$ **CATE(x) ≈ β, ATE ≈ β** $\Rightarrow Y_i=D_i\beta+\vec{X}_i^T\vec{\gamma}+\varepsilon_i$, $\hat{\beta}$ estimates ATE. **Fixed-Variables** $\{X_{i1},X_{i2},\dots\}$ $\begin{cases}X_{in}=1, i\in Case_n\\ X_{in}=0 \text{ if } \notin\end{cases}$ (Omit X_{i1}). Approximate **CEF**: $\mathbb{E}[Y_i|\vec{X}_i,D_i]=D_i\beta_D+X_{i1}\beta_1+X_{i2}\beta_2+\dots$

Omitted-Variable-Bias: Suppose $Y_i=\beta_0+\beta_DD_i+\beta_1X_{i1}+e_i$, $Y_i=\tilde{\beta}_0+\tilde{\beta}_DD_i+\varepsilon_i$, (Omitted) $\tilde{\beta}_D=\frac{Cov(Y_i,D_i)}{Var(D_i)}=\frac{Cov(\beta_0+\beta_DD_i+\beta_1X_{i1}+e_i,D_i)}{Var(D_i)}=\beta_D+\beta_1\frac{Cov(X_{i1},D_i)}{Var(D_i)}$, $Cov(e_i,D_i)=0$, $E[e_i|D_i,X_{i1}]=0$
$E[e_i]=0$
$E[D_ie_i]=E[D_iE[e_i|D_i,X_{i1}]]=0$, $\gamma_D=\frac{Cov(X_{i1},D_i)}{Var(D_i)}$ $\therefore \tilde{\beta}_D=\beta_D+\beta_1\gamma_D \Longrightarrow \hat{\beta}$ bias $\uparrow$ if X_{i1} highly correlated $\wedge$ D_i, If $\beta_1=0 \vee \gamma_D=0\Rightarrow \overset{0}{\underset{\|}{\text{OVB}}}$, $\hat{\tilde{\beta}}=\hat{\beta}_D+\hat{\beta}_1\frac{\widehat{Cov}(X_{i1},D_i)}{\widehat{Var}(D_i)}$, $Y_i=\beta_0+\beta_DD_i+\beta_1X_{i1}+\beta_2X_{i2}+e_i$
(Omit) X_{i2}: $\tilde{\beta}_D=\beta_D+\beta_2\gamma_D$, $X_{i2}=\gamma_0+\gamma_DD_i+\gamma_1X_{i1}+u_i$. **Heterogeneous T. Effects**: **CATE(x)** $=E[Y_i|D_i=1,\vec{x}_i=\vec{x}]-\mathbb{E}[Y_i|D_i=0,\vec{X}_i=\vec{x}]$, $E[Y_i|D_i=d,\vec{X}_i=\vec{x}]\approx\beta_Dd+\beta_{DX}(x_1d)+\vec{x}^T\vec{\gamma}$
CATE(x) $\approx\beta_D+\beta_{DX}x_1$ for someone with $X_{i1}=x_1$. For heterogeneity by X_{i1}: $Y_i=\beta_DD_i+\beta_{DX}(D_iX_{i1})+\vec{\gamma}^T\vec{X}_i+\varepsilon_i$. **Linear Combination**: $\sqrt{N}(\hat{\vec{\beta}}-\vec{\beta})\xrightarrow{d}N(0,\Sigma)$, $g(\vec{\beta})=\beta_1+\beta_2x$ continuous
CMT: $\therefore \sqrt{N}(\hat{\beta}_1+\hat{\beta}_2x-(\beta_1+\beta_2x))\xrightarrow{d}N(0,\sigma_x^2)$, $\sigma_x^2=\Sigma_{11}+x^2\Sigma_{22}+2x\Sigma_{12}$ $\therefore$ **C.I.** $(\beta_1+\beta_2x)$: $\hat{\beta}_1+\hat{\beta}_2x\pm1.96\hat{\sigma}_x/\sqrt{N}$, $\hat{\sigma}_x^2=\hat{\Sigma}_{11}+x^2\hat{\Sigma}_{22}+2x\hat{\Sigma}_{12}$. **Multiple Hypothesis T.**: Let $p<\alpha=5\%$, H_0: No T. effect
$\exists K$ iid groups. $N_{sig}:=\#$ significant results $\therefore \mathbb{P}(N_{sig}\geq1)=1-\mathbb{P}(N_{sig}=0)=1-\prod_i^K\mathbb{P}(G_j\,insig)=1-(0.95)^K$ $\therefore$ **Bonferroni**: Test H_{0_K} $\left|\alpha=\frac{5}{K}\%\right.$ $\therefore \mathbb{P}(N_{sig}>0)\leq\sum_K\mathbb{P}(K \text{ significant})=K\left(\frac{0.05}{K}\right)=0.05$

Difference-in-Differences: Assume $\exists$ periods $t=1,2$. Treated units $(D_i=1)$ treated in period 2; Never treated $(D_i=0)$
$Y_{it}=D_iY_{it}(1)+(1-D_i)Y_{it}(0)$ (Y_{it}: Observed Outcome $\Big|_{\text{period } t}^{\text{unit } i}$). **Assumptions**: (1) **No-anticipation** $Y_{i1}(0)=Y_{i1}(1)$ (Treatment$|_{t=2}$ does not influence outcomes$|_{t=1}$; (2) **Parallel Trends**: $\mathbb{E}[Y_{i2}(0)-Y_{i1}(0)|D_i=1]=\mathbb{E}[Y_{i2}(0)-Y_{i1}(0)|D_i=0]\Longleftrightarrow\mathbb{E}[Y_{i2}(0)|D_i=1]-\mathbb{E}[Y_{i2}(0)|D_i=0]=$
$\mathbb{E}[Y_{i1}(0)|D_i=1]-\mathbb{E}[Y_{i1}(0)|D_i=0]$ i.e. Δ in $Y(0)$ for treated $=\Delta$ in $Y(0)$ for control $\Longleftrightarrow$ Selection Bias$|_{t=2}$ = Selection Bias$|_{t=1}$; **D-i-D**: $\mathbb{E}[Y_{i2}-Y_{i1}|D_i=1]-\mathbb{E}[Y_{i2}-Y_{i1}|D_i=0]$
$=\mathbb{E}[Y_{i2}(1)-Y_{i1}(1)|D_i=1]-\mathbb{E}[Y_{i2}(0)-Y_{i1}(0)|D_i=0]=\mathbb{E}[Y_{i2}(1)-Y_{i1}(0)|D_i=1]-\mathbb{E}[Y_{i2}(0)-Y_{i1}(0)|D_i=0]=\mathbb{E}[Y_{i2}(1)-Y_{i2}(0)|D_i=1]+\mathbb{E}[Y_{i2}(0)-Y_{i1}(0)|D_i=1]-$
$\mathbb{E}[Y_{i2}(0)-Y_{i1}(0)|D_i=0]=\mathbb{E}[Y_{i2}(1)-Y_{i2}(0)|D_i=1]=\tau_{ATT}$ (**Average treatment effect on the treated**) $\therefore \tau_{ATT}=\underset{(\text{identified})}{\text{D-i-D}}$. Estimate $\hat{\tau}_{ATT}=$
$\bar{Y}_{12}-\bar{Y}_{11}-\bar{Y}_{02}-\bar{Y}_{01}$ ($\bar{Y}_{dt}$: sample mean for units $D_i=d|_t$). **D-i-D-Regression**: $Y_{it}=\beta_0+\beta_1Post_t+\beta_2D_i+\beta_3D_iPost_t+\varepsilon_{it}$ ($Post_t=\mathbb{1}[t=2]$). **CEF**:
$\mathbb{E}[Y_{it}|D_i=0,Post_t=0]=\beta_0$, $\mathbb{E}[Y_{it}|D_i=0,Post_t=1]=\beta_0+\beta_1$, $\mathbb{E}[Y_{it}|D_i=1|Post_t=0]=\beta_0+\beta_2$, $\mathbb{E}[Y_{it}|D_i=1,Post_t=1]=\beta_0+\beta_1+\beta_2+\beta_3$ $\therefore \beta_3=(\mathbb{E}[Y_{it}|D_i=1,Post_t=1]-\mathbb{E}[Y_{it}|D_i=1,Post_t=0])-(\mathbb{E}[Y_{it}|D_i=0,Post_t=1]-\mathbb{E}[Y_{it}|D_i=0,Post_t=0])$
$=\tau_{ATT}\Longrightarrow\hat{\beta}_3=(\bar{Y}_{12}-\bar{Y}_{11})-(\bar{Y}_{02}-\bar{Y}_{01})=\hat{\tau}_{ATT}$. **Multiple Periods**: $t=-\underline{T},\dots,\bar{T}$, s.t. treatment$|_{t=1}$, $\forall t=s\neq0$: $\hat{\beta}_s=(\bar{Y}_{1s}-\bar{Y}_{0s})-(\bar{Y}_{10}-\bar{Y}_{00})=\Delta_{t=s}-\Delta_{t=0}$
$Y_{it}=\phi_t+D_i\gamma+\sum_{s\neq0}D_i\times\mathbb{1}[t=s]\beta_s+\varepsilon_{it}$. **S.E. for Panel Regressions**: (OLS) $Y_i=\vec{X}_i^T\vec{\beta}+e_i$ for $(Y_i,\vec{X}_i)$ iid (independent). Clustered S.E.: Allow $(Y_{it},\vec{X}_{it})$ to be correlated if in same cluster i.e. $\{Y_{i1},Y_{i2},\dots\}\perp\!\!\!\perp\{Y_{j1},Y_{j2},\dots\}$ for $j\neq i$. **Instrumental Variables**: Let $Y_i=\alpha+\beta X_i+\varepsilon_i$ We do not assume $Cov(X,\varepsilon_i)=0$
Assume $\exists Z_i$ that affects Y_i only through its effect on X_i. Randomization $\Longrightarrow Cov(Z_i,\varepsilon_i)=0$, $Cov(Z_i,Y_i)=Cov(Z_i,\alpha+\beta X_i+\varepsilon_i)=\beta Cov(Z_i,X_i)$
$\therefore$ if $Cov(Z_i,X_i)\neq0\Rightarrow\beta=\frac{Cov(Z_i,Y_i)}{Cov(Z_i,X_i)}$ (identified) $\therefore Z_i$ instrument $\sim X_i$. $\beta=\frac{Cov(Z_i,Y_i)/Var(Z_i)}{Cov(Z_i,X_i)/Var(Z_i)}=\frac{\rho}{\pi}$, $Y_i=\kappa+\rho Z_i+v_i$ (**Reduced Form Reg**) and
$X_i=\mu+\pi Z_i+\eta_i$ (**First-Stage reg**) $Cov(Z_i,v_i)=Cov(Z_i,\eta_i)=0$ If **Z_i binary**: $\beta=\frac{\rho}{\pi}=\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[X_i|Z_i=1]-E[X_i|Z_i=0]}=\frac{\text{Reduced-form-effect}}{\text{first-form-effect}}$, **Validity-Condition**: $Cov(Z_i,\varepsilon_i)=0$
(i) **As-good-as-random assignment**: Individuals with higher/lower potential outcomes do not face systematically different values of Z_i. **Exclusion**:
the "assignment" of Z_i only affects Y_i through X_i: $Z\longrightarrow X\underset{\varepsilon}{\overset{\beta}{\rightleftarrows}}Y$. **Measurement-Error**: $Y_i=\alpha+\beta X_i^*+v_i$, $Cov(X_i^*,v_i)=0$ but only observe $X_i=X_i^*+\eta_i$
where $Cov(X_i^*,\eta_i)=Cov(X_i^*,v_i)=0$, $\frac{Cov(X_i,Y_i)}{Var(X_i)}=\frac{Cov(X_i^*+\eta_i,\alpha+\beta X_i^*+v_i)}{Var(X_i^*+\eta_i)}=\beta\frac{Var(X_i^*)}{Var(X_i^*)+Var(\eta_i)}\Longrightarrow$ **Attenuation-Bias**: $\frac{Cov(X_i,Y_i)}{Var(X_i)}/\beta\triangleq\lambda\in(0,1)$ $\therefore\lambda=\frac{Var(X_i^*)}{Var(X_i^*)+Var(\eta_i)}$ "**signal-to-noise**" ratio
If $Z_i=X_i^*+\xi_i$ with $Cov(X_i^*,\xi_i)=Cov(\eta_i,\xi_i)=Cov(v_i,\xi_i)=0\Longrightarrow\frac{Cov(Z_i,Y_i)}{Cov(Z_i,X_i)}=\frac{Cov(X_i^*+\xi_i,\alpha+\beta X_i^*+v_i)}{Cov(X_i^*+\xi_i,X_i^*+\eta_i)}=\beta\frac{Var(X_i^*)}{Var(X_i^*)}=\beta$. For **$Z_i$ binary**: $\hat{\beta}=\frac{\hat{Y}_{Z=1}-\hat{Y}_{Z=0}}{\bar{X}_{Z=1}-\bar{X}_{Z=0}}$, $\hat{\beta}=\frac{\hat{\rho}}{\hat{\pi}}$,
$Y_i=\hat{\kappa}+\hat{\rho}Z_i+\hat{v}_i$; $X_i=\hat{\mu}+\hat{\pi}Z_i+\hat{\eta}_i$, $Cov(Z_i,\varepsilon_i)\overset{0}{=}\wedge Cov(Z_i,X_i)\neq0\Longrightarrow$ identification: $\beta=\frac{Cov(Z_i,Y_i)}{Cov(Z_i,X_i)}$. For inference: $\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{d}N(0,\sigma^2)$. **I.V. | Covariates**: Assume $Cov(Z_i,\varepsilon_i|W_i)=0$ for some $(W_i)\Longrightarrow\frac{Cov(Z_i,Y_i|W_i)}{Cov(Z_i,X_i|W_i)}=\beta$. For binary Z_i: $\frac{E[Y_i|Z_i=1,W_i]-E[Y_i|Z_i=0|W_i]}{E[X_i|Z_i=1,W_i]-E[X_i|Z_i=0|W_i]}=\beta$. $\hat{\beta}=\frac{\hat{\rho}}{\hat{\pi}}$: $Y_i=\kappa+\rho Z_i+\vec{W}_i^T\vec{\gamma}+v_i$; $X_i=\mu+\pi Z_i+\vec{W}_i^T\vec{\psi}+\eta_i$, $\sqrt{N}\left(\begin{pmatrix}\hat{\rho}\\ \hat{\pi}\end{pmatrix}-\begin{pmatrix}\rho\\ \pi\end{pmatrix}\right)\xrightarrow{d}N(0,\Sigma)$.
Delta-Method: $g(\hat{\rho},\hat{\pi})=\frac{\hat{\rho}}{\hat{\pi}}$, For $(\hat{\rho},\hat{\pi})\approx(\rho,\pi)$ (Taylor): $g(\hat{\rho},\hat{\pi})\approx g(\rho,\pi)+\nabla g(\rho,\pi)(\hat{\rho}-\rho,\hat{\pi}-\pi)^T\underset{(ish)}{\Longrightarrow}\sqrt{N}\left(\frac{\hat{\rho}}{\hat{\pi}}-\frac{\rho}{\pi}\right)\approx\nabla g(\rho,\pi)\sqrt{N}\left(\begin{pmatrix}\hat{\rho}\\ \hat{\pi}\end{pmatrix}-\begin{pmatrix}\rho\\ \pi\end{pmatrix}\right)\longrightarrow N(0,\sigma)$
$=N(0,\nabla g(\rho,\pi)\Sigma\nabla g(\rho,\pi)')$. **Multiple Instruments**: $Z_{i1},\dots,Z_{iL}$ (Multiple Z_{il} can test-validity). (1) If Z_i a-g-a-r-a prior to the realization of some treatment X_i, w/ outcome Y_i, then Z_i valid for estimating effect X_i on Y_i: False: We also need exclusion restriction. (2) If the pre-trends of a treat. $\wedge$ control group are not statistically significantly diff. then a D-i-D identifies a causal effect: False: We also need trends stay parallel in post period. 3) If $f(x)=E[Y_i|X_i=x]$ linear, then Reg $Y_i|X_i$ identifies it: True. 4) Let D_i, Y_i, W_i (binary). $(Y_i(1),Y_i(0))\perp\!\!\!\perp D_i|W_i$. Let $Y_i=\alpha+\beta D_i+\gamma W_i+\delta D_iW_i+\varepsilon_i$
Show: the coefficients capture the CEF $E[Y_i|D_i=d|W_i=w]$: $E[Y_i|D_i=0,W_i=0]=\alpha$, $E[Y_i|D_i=1,W_i=0]=\alpha+\beta$, $E[Y_i|D_i=0,W_i=1]=\alpha+\gamma$, $E[Y_i|D_i=1,W_i=1]=\alpha+\beta+\gamma+\delta$

$\beta = E[Y_i | D_i=1, W_i=0] - E[Y_i | D_i=0, W_0]$, by $D_i | W_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) \Longrightarrow \beta = E[Y_i(1) - Y_i(0) | W_i = 0]$ (some CATE). 4) To estimate the effect of a state level policy Δ by D-i-D, There must be a panel dataset where we observe the same people over time: False: we only need average outcomes by status and time. So a "repeated cross-section" dataset which samples diff. people over time is enough. 5) If you double rows in a dataset OLS coefficients $\wedge$ robust standard errors do not change: FALSE: Coeff. dont change but SE $\downarrow$ by $\sqrt{2}$ unless you cluster by the original identifier.

6) Suppose $Y_i = \alpha + \beta X_i^2 + \varepsilon_i$, $E[\varepsilon_i | X_i] = 0$, $X_i \sim N(0,1)$. (If $Z_i \sim N \Longrightarrow E[(Z_i - E[Z_i])^3] = 0$) a) Show that Reg. of Y_i on a constant, X_i, X_i^2 identifies α, β: $E[Y_i | X_i, X_i^2] = E[\alpha + \beta X_i^2 + \varepsilon_i | X_i, X_i^2] = \alpha + \beta X_i^2 + E[\varepsilon_i | X_i] = \alpha + \beta X_i^2$ is linear $\therefore$ a regression identifies the CEF parameters (α, β), (b) Derive slope coeff. in a bivariate regression of Y_i on a constant and X_i: $\frac{Cov(Y_i, X_i)}{Var(X_i)} = \frac{Cov(\alpha + \beta X_i^2 + \varepsilon_i, X_i)}{Var(X_i)} = \beta \frac{Cov(X_i^2, X_i)}{Var(X_i)}$, $Cov(X_i^2, X_i) = E[(X_i^2 - E[X_i^2])(X_i - E[X_i])] = E((X_i - \overset{0}{\overset{\parallel}{E}}[X_i])^3)$ Show $\alpha + \beta$ identified by the constant of the bivariate regression in (b): $E[Y] - \frac{Cov(Y_i, X_i)}{Var(X_i)} E[X_i] = E[Y_i] = \alpha + \beta E[X_i^2] + E[\varepsilon_i] = \alpha + \beta$, since $E[X_i^2] = 1$, $E[\varepsilon_i] = E[E[\varepsilon_i | X_i]] = 0$. 7) Consider $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$; $X_{1i} = \gamma_0 + \gamma_1 X_{2i} + v_i$; $X_{2i} = \delta_0 + \delta_1 X_{1i} + w_i$, β coincides with slope parameter in the regression of Y_i onto $(1, W_i)$. 8) The OLS estimator for $\vec{\beta} = (\beta_1, \beta_2)^T$ in (7) is asymptotically normal as the sample size $n \to \infty$, $\sqrt{n}(\hat{\vec{\beta}} - \vec{\beta}) \xrightarrow{d} N(0, \Sigma)$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$, the asymptotic variance of $\sqrt{n}[(\hat{\beta}_1 - \hat{\beta}_2) - (\beta_1 - \beta_2)]$ is: $\sqrt{n}\left(\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} - \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}\right) \xrightarrow{d} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, $\sqrt{n}[(\hat{\beta}_1 - \hat{\beta}_2) - (\beta_1 - \beta_2)] \xrightarrow{d} Z_1 - Z_2$, $Var(Z_1 - Z_2) = Var(Z_1) - 2\,COV(Z_1, Z_2) + Var(Z_2) = \sigma_1^2 - 2\sigma_{12} + \sigma_2^2$

9) $\exists$ Panel Data (Y_{it}, D_i), $t = -4, \ldots, 4$, treatment $|_{t=1}$, D_i binary. (a) State assumption needed to identify ATT using Di-D: No anticipation: future treatment does not casually affect the current potential outcomes, Parallel Trends: The average time trend of the control outcomes are common between treated and control groups $(E[Y_{it}(0) - Y_{it-1}(0) | D=1] = E[Y_{it}(0) - Y_{it-1}(0) | D_i = 0]$

b) Why should we use the clustered SE in the construction of C.I $\forall t$: Panel Data (Y_{it}, D_i) are not iid over time $\therefore$ we should use CSE to take into account dependence across the observations. c) $\exists$ plot of treatment $\wedge$ control against time. i.e. log(monthly earnings) vs Months since programming class. Write a regression s.t. OLS estimates will yield the event-study coefficients: $Y_{it} = \alpha + \sum_{s \neq 0} \beta_s \cdot 1\{t=s\} + \gamma D_i + \sum_{s \neq 0} \delta_s D_i 1\{t=s\} + \varepsilon_{it}$

10) Let Y_i i's earnings as an adult, D_i indicator if grew up in a high-poverty neighborhood, Consider collecting a survey to ask earnings and if they grew up poor, (a) We collect $\vec{x}_i$ about familys background. What should $\vec{x}_i$ include to make conditional unconfoundedness assumption somewhat plausible? (Wealth of parents, education of parents, race, etc). (b) Suppose you are given a dataset with $(Y_i, D_i, \vec{x}_i)$ Write OLS specification that you might run to estimate the casual effect under cond. unconfoundedness, and a CATE: $E[Y(1) - Y(0) | X_i = x]$ linear in x: $Y_i = \alpha + \beta D_i + \vec{X}_i^T \gamma + D_i \cdot \vec{X}_i^T \delta + \varepsilon_i$. CATE $\approx \hat{\beta} + \vec{x}^T \hat{\delta}$

c) Z_i: voucher to move to better neighborhood, $Z_i = 1$ if lost lottery, $Z_i = 1$ if won, $D_i = 1$ poor, $D_i = 1^{\neg}$ poor.

	$Z=1$	$Z=0$
High-Poverty (mean)	0.5	0.4
Income mean	11,000	11,300

$\{0,1\}$

Calculate an I.V. instrument for growing up in a poor neighborhood effect in income. First-Stage: $0.5 - 0.4 \overset{0.1}{\parallel}$
Reduced-form: $11{,}000 - 11{,}300 = -300$ $\therefore$ I.V. estimator $= \frac{-300}{0.1} = -3000$. First S
First Stage: $D_i = \gamma_0 + \gamma_1 Z_i + v_i$, Reduced-form: $Y_i = \rho_0 + \rho_1 Z_i + u_i$, I.V. estimand is: $\beta = \frac{\rho_1}{\gamma_1}$

11) M_i: military service, Y_i: mental-health at 40's. $Y_i = \alpha + \beta M_i + \vec{A}_i^T \vec{\gamma} + \varepsilon_i$ $A_i \in \mathbb{R}^J$: characteristics relevant to mental health

Estimand: non-random, it is a function of the population distribution. When estimand and (target) parameter coincide $\Leftrightarrow$ identification. Ex: 1) $Y = a + X^3/b$, $b>0$, $X \sim N(0,1)$, $F_{a,b}(y) = \mathbb{P}(Y \le y)$, $Y \le y \Leftrightarrow a + X^3/b \le y \Leftrightarrow X \le \sqrt[3]{b(y-a)}$ $\therefore \mathbb{P}(Y \le y) = \Phi(\sqrt[3]{b(y-a)})$, $E[Y] = E(a + X^3/b) = a$ given $\mathbb{E}[X^3] = 0$ for $X \sim N(0,1)$ $\therefore Cov(Y, X) = E[YX]$ $E[aX + X^4/b]$, $\mathbb{E}[X^4] = 3$ for $X \sim N(0,1)$ $\therefore Cov(X,Y) = a \underbrace{bE[X]}_{zero} + \frac{3}{b} = \frac{3}{b}$; $\mathbb{P}(Y > 0) = 1 - \Phi(\sqrt[3]{b(0-a)}), = 1 - \Phi(0) = 0.5$

Let $W = a + X^3/b + Z$, Z mean-independent of X $\therefore E[W|X] = a + X^3/b + E[Z|W] = Y$ Need:

Ex 2: $f(x;p) = \begin{cases} p, & x=1 \\ 1-p, & x=0 \end{cases}$, $\{X_i\}^N \text{ iid}$ $\therefore \mathbb{E}[X_i] = \sum_{x=0}^{1} x f(x,p) = p$. Natural estimator of $\mathbb{E}[X_i]$: $\hat{p} = \frac{1}{N}\sum^N X_i$. Consistency: $Var(\hat{p}) \xrightarrow{p} 0$ $\therefore$ $Var(\hat{p}) = Var(\frac{1}{N}\sum^N X_i) = \frac{1}{N^2}\sum^N Var(X_i) = \frac{p(1-p)}{N} \xrightarrow{p} 0$. Unbiased as $\mathbb{E}[\hat{p}] = p$. Ex 3) $\{X_i\}^N$ iid $Exp(\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$, $E[X_i] = \frac{1}{\lambda}$ $\therefore \hat{X} = \frac{1}{N}\sum^N X_i$ $\therefore \mathbb{E}[\hat{X}] = \frac{1}{N}\sum^N \mathbb{E}[X_i] = \frac{1}{\lambda}$, $Var[X_i] = E[X_i^2] - E[X]^2 = \left(\int_0^\infty x^2 \lambda e^{-\lambda x} dx\right) - \left(\frac{1}{\lambda}\right)^2 = -x^2 e^{-\lambda x}\Big|_0^\infty + 2\int_0^\infty x e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$ $\therefore Var(\hat{X}) = \frac{1}{N^2}\sum Var(X_i) = \frac{1}{N\lambda^2}$, By CLT: $\sqrt{N}(\hat{X} - \frac{1}{\lambda}) \to N(0, \frac{1}{\lambda^2})$. Estimator for $\frac{1}{\lambda^2} = \bar{X}_n^2$, $\hat{\lambda} = \frac{1}{\bar{X}_n}$, By LLN: $\bar{X}_n \xrightarrow{p} E[X] = \frac{1}{\lambda}$ By CMT, since $g(x) = x^2$ continuous $\Rightarrow \frac{1}{\hat{\lambda}^2} \xrightarrow{p} \frac{1}{\lambda^2}$ (consistent). Ex 3) $Cov(aX, bY) = ab\,Cov(X,Y)$, $Std(aX) = a\,Std(X)$ $Std(bY) = b\,Std(Y) \Rightarrow Corr(aX, bY) = Corr(X, Y)$, Ex 4) $Y \perp\!\!\!\perp X \Rightarrow E[Y^3X] = E(Y^3)E(X)$, by LIE $E[Y^3X] = E[E[Y^3X]|X] = \mathbb{E}(X\,\mathbb{E}[Y^3|X]] = E[X\,\mathbb{E}(Y^3)]] = \mathbb{E}[X]\mathbb{E}[Y^3]$. Ex 5): P-value: "probability of seeing a test statistic as extreme as the observed under Ho T. Ex 6: $(\frac{1}{N}\sum_i X_i)^2$ consistent for $E[X_i]^2$: By LLN: $\frac{1}{N}\sum^N X_i \xrightarrow{p} E[X_i]$ $f(\cdot) = x^2$ continuous $\Rightarrow$ By CMT: $f(\frac{1}{N}\sum^N X_i) \xrightarrow{p} f(E[X_i])$. Ex 7: $D_i = \begin{cases} 1 \\ 0 \end{cases}$, outcomes $Y_i(1), Y_i(0)$, assume $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0))$ Prove that diff in Y_i identifies CATE: $\mathbb{E}[Y_i|D_i=1] - \mathbb{E}[Y_i|D_i=0] = \mathbb{E}[Y_i(1)|D_i=1] - \mathbb{E}[Y_i(0)|D_i=0] = \mathbb{E}(Y_i(1)] - \mathbb{E}[Y_i(0)] = \underbrace{E[Y_i(1) - Y_i(0)]}_{CATE}$, $Y_i = D_i Y_i(1) + (1-D_i)Y_i(0)$, Suppose $\tilde{Y}_i = Y_i + e_i$, $e_i \perp\!\!\!\perp D_i$: $\mathbb{E}[\tilde{Y}_i|D_i=1] - E[\tilde{Y}|D=0]$ $= \mathbb{E}[Y_i + e_i|D_i=1] - \mathbb{E}[Y_i + e_i|D_i=0] + E[e_i|D_i=1] - E[e_i|D_i=0] = \mathbb{E}[Y_i|D_i=1] - \mathbb{E}[Y_i|D_i=0$, Therefore: $\mathbb{E}[\tilde{Y}_i|D_i=1] - \mathbb{E}[\tilde{Y}_i|D_i=0]$ $= \mathbb{E}[Y_i|D_i=1] - \mathbb{E}[Y_i|D_i=0] = \mathbb{E}(Y_i(1) - Y_i(0)]$ $\therefore$ CATE Identified. Consider the sample means $\tilde{Y}_i$ is this consistent estimate of CATE? Yes, by LLN each sample mean is consistent $E[\tilde{Y}_i|D_i=d]$ $\therefore$ by CMT we have a consistent estimator for $\mathbb{E}[\tilde{Y}_i|D_i=1] - \mathbb{E}[\tilde{Y}_i|D_i=0] = \beta$ under identifying assumptions. Assume we observe a difference in sample means of $\tilde{Y}_i = 10{,}000$ with $SE = 4500$, what can you conclude about $E[Y_i|D_i=1] - E[Y_i|D_i=0]$? T-Stat: $\frac{10{,}000}{4{,}500} > 2$ $\therefore$ by our usual asymptotic approximation we can reject Ho with 95% confidence $\therefore$ Estimate is statistically significant at 5% level. Ex 8) Suppose $\{Y_i, X_i\}^N$, $Y_i = \beta X_i + \varepsilon_i$ for some β parameter. Suppose $E(\varepsilon_i|X_i) = 0$ $E(X_i) \ne 0$, show β id. by $E[Y_i]/E[X_i]$ $\therefore E[Y_i]/E(X_i) = E(\beta X_i + \varepsilon_i)/E(X_i) = \beta E(X_i)/E(X_i) + E(\varepsilon_i)/E(X_i)$ $= \beta + E(\varepsilon_i)/E(X_i)$. By LIE: $E(E[\varepsilon_i|X_i]) = E(\varepsilon_i) = 0 \Rightarrow E(Y_i)/E(X_i) = \beta$ (identification). Show $\hat{\beta} = \frac{\frac{1}{N}\sum Y_i}{\frac{1}{N}\sum X_i}$ consistent for β. By LLN: $\frac{1}{N}\sum Y_i \xrightarrow{p} E(Y_i)$, $\frac{1}{N}\sum X_i \xrightarrow{p} E(X_i)$, By CMT: $\frac{\frac{1}{N}\sum Y_i}{\frac{1}{N}\sum X_i} \xrightarrow{p} \frac{E(Y_i)}{E(X_i)} = \beta$. Show $\hat{\beta}$ unbiased estimate of β $E(\hat{\beta}) = E(E(\hat{\beta}|X_1, \dots, X_N)) = \mathbb{E}\left[E\left[\frac{1}{N}\sum Y_i)|X_1,\dots,X_N\right] / (\frac{1}{N}\sum X_i)\right] = \mathbb{E}\left[(\frac{1}{N}\sum E[Y_i|X_i])/(\frac{1}{N}\sum X_i)\right] = \mathbb{E}\left[(\frac{1}{N}\sum(\beta X_i + E[\varepsilon_i|X_i]))/(\frac{1}{N}\sum X_i)\right]$ $E[\beta(\frac{1}{N}\sum X_i)/(\frac{1}{N}\sum X_i)] = \beta$. Inference: "Process of learning about observable features of the full population from a random sample. Identification: "Process of learning about casual parameters from population estimands

A treatment D_i randomly assigned is independent of potential outcomes $Y_i(1) \wedge Y_i(0)$. But since $Y_i = Y_i(0) + (Y_i(1) - Y_i(0))D_i$ it is not independent of Y_i unless there are no treatment effects" i.e. $Y_i(1) - Y_i(0) = 0$ $\forall i$. $\forall i$

Ex 9: D_i, Y_i outcome of interest, W_i: another observed variable. Suppose CATE of D_i on Y_i constant $\therefore Y_i(1) - Y_i(0) = \beta$. Show $Y_i = \beta D_i + \varepsilon_i$ for some unobserved ε_i. $Y_i = Y_i(0)(1-D_i) + Y_i(1)D_i = (Y_i(1) - Y_i(0))D_i + Y_i(0) = \beta D_i + \varepsilon_i \Rightarrow \varepsilon_i = Y_i(0)$. Suppose we think D_i not randomly assigned $\therefore$ Show that: $E[Y_i|D_i=1] - E[Y_i|D_i=0] = \beta + b$, We have $E[Y_i|D_i=1] - E[Y_i|D_i=0] = E[\beta D_i + \varepsilon_i|D_i=1] - E[\beta D_i + \varepsilon_i|D_i=0]$ $= (\beta + E[\varepsilon_i|D_i=1]) - \mathbb{E}(\varepsilon_i|D_i=0] = \beta + (\mathbb{E}[\varepsilon_i|D_i=1] - E[\varepsilon_i](D_i=0) = \beta + b$, $b \ne 0$ whenever treated individuals have different untreated potential outcomes, on average, than untreated individuals. Suppose $\varepsilon_i = \mu + \lambda W_i$, What does $E[W_i|D_i=1] = E[W_i|D_i=0]$ mean for the id. of β: $b = E[\mu + \lambda W_i|D_i=1] - E[\mu + \lambda W_i|D_i=0] = \lambda(E[W_i|D_i=1] - E[W_i|D_i=1] = 0$ (i.e) β is identified.