Sem 6

$$D_i^T = \begin{cases} 1 & , T \\ 0 & , c \end{cases}$$

y^T V.o.I $y^{T-?}$ Unexp. Var

$$\underbrace{y_i}_{\text{Total var}} = \beta_1 + \underbrace{\beta_2 \cdot D_i}_{\text{Var Exper.}} + \underbrace{\gamma \cdot Z_i}_{\text{Exp. Var}} + \varepsilon_i$$

2 sample t-test $H_0: \mu_T = \mu_c$ $\qquad \hat{\beta}_2 = \frac{\hat{Cov}(D,y)}{\hat{Var}(D)} = \bar{y}_T - \bar{y}_c$

vs

t-test for sign. $H_0: \beta_2 = \mu_T - \mu_c = 0$ $\qquad \hat{\beta}_1 = \bar{y}_c$

$y = \alpha + \theta X + \varepsilon_i$

$y - \theta X = \alpha + \varepsilon_i$

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

2) $H_0: E(y_T^{CUPED}) = E(y_c^{CUPED})$

$y_{CUPED} | D_i \Rightarrow \beta_2$

① $y | X_1, X_2 \Rightarrow \hat{\beta}_1, \hat{\beta}_2$

② $y | X_2 \Rightarrow \hat{\varepsilon} | X_1 \Rightarrow \hat{\beta}_1$

CV → Control $y^{T-?}$

↳ CUPED

→ Categorical

↳ stratified

$\rho(X,Y) < 0$

↳ antithetic

Task 1

Consider an estimator of $E(Y)$:

$$\widehat{Y}_{cv} = \bar{Y} - \theta\bar{X} + \theta E(X),$$

(a) Find the variance of this estimator
(b) Minimise the variance by θ
(c) Derive the variance of $\widehat{Y}_{cv}$ under optimal θ

a) $E(\hat{y}_{cv}) = E(\bar{y} - \theta\bar{X} + \theta E(X)) = \mu_Y - \cancel{\theta\mu_x} + \cancel{\theta\mu_x}$

$Var(\hat{y}_{cv}) = Var(\bar{y} - \theta\bar{X} + \theta E(X)) =$

$= Var(\bar{y} - \theta\bar{X}) = Var(\bar{y}) + Var(\theta\bar{X}) - 2cov(\bar{y}, \theta\bar{X}) =$

$$= \frac{Var(y)}{n} + \theta^2 \frac{Var(X)}{\boxed{n}} - 2\theta \frac{Cov(y,X)}{n} \longrightarrow \min_\theta$$

b) $$\frac{\partial Var(\hat{y}_{cv})}{\partial \theta} = 2\theta \frac{Var(X)}{n} - 2\frac{Cov(y,X)}{n} = 0$$

$$\theta = \frac{Cov(y,x)}{Var(X)}$$

$y = \hat{\alpha} + \hat{\theta} x + \varepsilon$

c) $Var(\hat{y}_{cv}) = \frac{Var(y)}{n} + \frac{Cov^2(y,x)}{Var^{\cancel{2}}(x)} \cdot \frac{\cancel{Var}(x)}{n} - 2\frac{Cov(y,x) \cdot Cov(y,x)}{Var(x) \quad n}$

$$= \frac{Var(y)}{n}\left(1 - \frac{Cov^2(y,x)}{Var(x) \cdot Var(y)}\right) = Var(\bar{y}) \cdot (1 - \rho^2_{x,y})$$

$\frac{Var(y)}{n} = Var(\bar{y})$

$$\rho^2_{x,y} = \frac{\sigma^2_{xy}}{\sigma^2_x \cdot \sigma^2_y} \in (0,1)$$

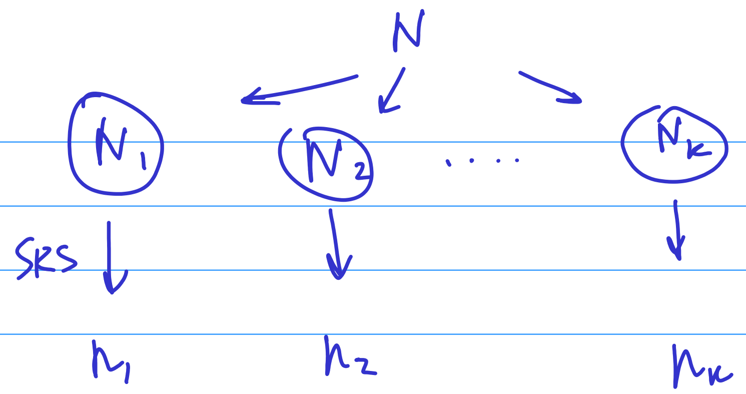
share of exp. var $\downarrow$

$$Var(\hat{y}_{cv}) = Var(\bar{y}) \cdot (\underbrace{1 - \rho^2_{x,y}}_{(0;1)}) = Var(\bar{y})(\underbrace{1 - R^2}_{\text{share of unexp. var}})$$

1) $Var(\hat{y}_{cv}) \leq Var(\bar{y})$

2) $\rho^2_{x,y} \uparrow \Rightarrow Var(\hat{y}_{cv}) \downarrow$

3) $\rho^2_{x,y} = R^2_{y|x}$



$$\Rightarrow \hat{y}_{strat} =$$
$$= \sum \omega_i \bar{y}_i$$

Task 2

Let X is binary with values 1 and 0. Let $p = E(X)$. Show that the following estimates are identical:

$$\widehat{Y}_{strat} = p\bar{Y}_1 + (1-p)\bar{Y}_0, \quad = \sum_{i=1}^{2} \omega_i \bar{y}_i = \frac{\sum y_i}{N} = \frac{N_1 \sum_{x=1} y_i}{N N_1} + \frac{\sum_{x=0} y_i N_0}{N N_0} = \frac{N_1}{N} \cdot \bar{y}_1 + \frac{N_0}{N_1} \cdot \bar{y}_0$$
$$\widehat{Y}_{cv} = \bar{Y} - \hat{\theta}\bar{X} + \hat{\theta}E(X)$$

$$\frac{n_1}{N} = \hat{p} \qquad \qquad p$$

$$X = \begin{cases} 1, & T \\ 0, & C \end{cases}$$

$$X_i \sim Be(p)$$
$$E(X_i) = p$$
$$Var(X_i) = pq$$
$$\hat{\theta} = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)} = \bar{y}_1 - \bar{y}_0$$

$$\hat{y}_{cv} = \bar{y} - (\bar{y}_1 - \bar{y}_0)\bar{X} + (\bar{y}_1 - \bar{y}_0)\, p =$$
$$= \frac{N_1}{\cancel{N}} \cdot \bar{y}_1 + \frac{N_0}{N} \bar{y}_0 -$$
$$- \cancel{\bar{y}_1} \cdot \frac{N_1}{N} + \bar{y}_0 \frac{N_1}{N} +$$
$$p\bar{y}_1 - p\bar{y}_0 =$$
$$= \bar{y}_0 \cdot \underbrace{\frac{N_0 + N_1}{N}}_{1} + p\bar{y}_1 - p\bar{y}_0 =$$

where $N_0 + N_1 = N$

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

Task 3

How to choose the sample sizes $n_1, n_2, ..., n_k$ so that the available resources are used in an effective way?

There are two aspects of choosing the sample sizes:

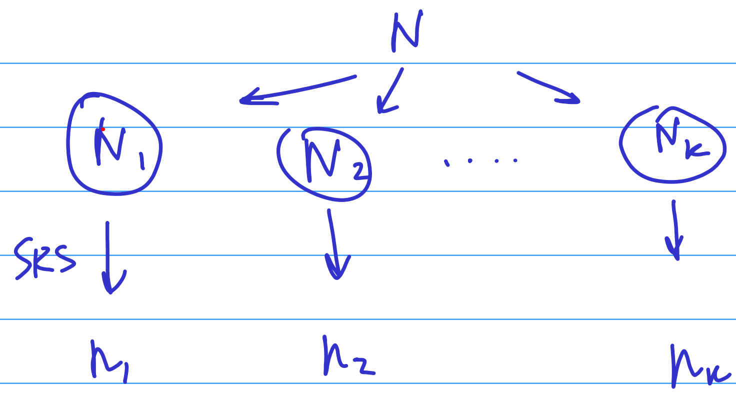
– Minimize the cost of survey for a specified precision.

– Maximize the precision for a given cost.

(a) Can both of these goals be achieved?

Goal $\nearrow$ max π / min c $\nearrow$ NPS $\quad$ $n_i \downarrow$

Goal $\searrow$ min Var, max precision $\quad$ $n_i \uparrow$ $\searrow$ max Power $\leftrightarrow$ max Info



(b) Consider the stratified sample variance estimator fith finite population correction:

$$\mathrm{Var}(\bar{y}_{st}) = \sum_{i=1}^{k} w_i^2 \underbrace{\mathrm{Var}_{fpc}(\bar{y}_i)}$$

$$S_j^2 = \frac{1}{n-1}\sum_{k=1}^{n_j}(y_k - \bar{y}_j)^2$$

where

$$\mathrm{Var}_{fpc}(\bar{y}_j) = \frac{N_j - n_j}{N_j n_j} S_j^2 = \frac{N_i - n_i}{N_i} \cdot \frac{S_j^2}{n_j} = \underbrace{\frac{N_j - n_j}{N_j}}_{fpc} \cdot \underbrace{\mathrm{Var}(\bar{y}_j)}_{CLT}$$

And let the cost function be

$$C = \underbrace{C_0}_{FC} + \underbrace{\sum_{i=1}^{k} C_i n_i}_{VC}$$

$$\frac{3-2}{3} = \frac{1}{3}$$

where

C : total cost

C_0 : overhead cost, e.g., setting up the office, training people etc

C_i : cost per unit in the i^{th} stratum

$\sum_{i=1}^{k} C_i n_i$: total cost within the sample.

Find minimal variance by n_i under this cost function. (Hint: consider the Lagrangian function).

M Б Г

[M Б] Г

1) $P(Г) = \frac{1}{3}$

2) $P(Г) = \frac{1}{3}$

[100 200] 300

[M, Б, Г]

200

1) $p(Г) = 1/3$

2) $P(Г) = 0$

$p(M) = 1/2$

$p(Б) = 1/2$

$$-\lambda = \psi^2$$

$$\mathcal{L} = \mathrm{Var}(\bar{y}_{st}) - \overbrace{\lambda}^{\text{shadow cost}}(C - C_0) =$$

$$= \sum w_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) s_i^2 + \psi^2 \sum c_i n_i =$$

$$= \sum w_i^2 \frac{1}{n_i} \cdot s_i^2 + \psi^2 \sum c_i n_i + C_1 =$$

$$= \sum \left(\sqrt{\frac{w_i^2 s_i^2}{n_i}} - \sqrt{\psi^2 \cdot c_i n_i} \right)^2 + \underbrace{2 \sum \frac{w_i s_i}{\sqrt{\cancel{n_i}}} \cdot \sqrt{c_i \cdot \cancel{n_i}} + C_1}_{C_2}$$

$$\frac{w_i s_i}{\sqrt{n_i}} = \psi \sqrt{c_i} \cdot \sqrt{n_i}$$

$$\Rightarrow n_i = \frac{1}{\psi} \cdot \frac{w_i \cdot s_i}{\sqrt{c_i}}$$

$w_i \uparrow \quad n_i \uparrow$

$s_i \uparrow \quad n_i \uparrow$

$c_i \uparrow \quad n_i \downarrow$

(c) $\sum c_i n_i = C_0^*$

$$\sum c_i \frac{1}{\psi} \cdot \frac{w_i \cdot S_i}{\sqrt{c_i}} = C_0^* \Rightarrow \psi$$

$n_i = \ldots \Rightarrow N = \sum n_i$