Kap 15: Stikprøver – tilfældig udvælgelse

23. maj 2002

Udvalgsbrøk: $f_i=n_i/N_i$			ALTERNATIV VARIATION (den variable kan antage 2 værdier: 0 eller 1)		GENEREL VARIATION (den variable kan antage alle værdier)
		Begreb	**Uden tilbagelægning**	**Med tilbagelægning**	**Uden tilbagelægning**
SIMPEL TILFÆLDIG UDVÆLGELSE	1		$Hyp(n,N,M)$	$Bin(n,\theta)$	$x_1,\ldots,x_n$ = værdi af var. målt på n enh i stikprøven
	2	Pop.gns. af obs. Variabel	$\theta=\bar{\mu}=\frac{1}{N}\sum_{V=1}^{N}\mu_V$		$\bar{\mu}=\frac{1}{N}\sum_{V=1}^{N}\mu_V$
	3	Skøn for 2	$\hat{\theta}=\frac{1}{n}\sum_{V=1}^{n}\mu_V$, μ_V = obs.værdier af x	$\theta\approx\hat{\theta}_m=\frac{x_m}{n}$	$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{X}$ for x_i=obs. afstok.var. X_i
	4	Pop.varians (anvendes til at udregne 6)	$\tau^2=\frac{1}{N-1}\sum_{v=1}^{N}(\mu_v-\theta)^2=\frac{N}{N-1}\theta(1-\theta)$		$\tau^2=\frac{1}{N-1}\sum_{v=1}^{N}(\mu_v-\bar{\mu})^2$
	5	Estimat for 4 (anvendes til at udregne 7)			$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$
	6	Usikkerhed på 3	$Var(\hat{\theta})=\frac{\tau^2}{n}(1-f)$	$Var(\hat{\theta})=\frac{\tau^2}{n}\cdot\frac{N-1}{N}=\frac{\theta(1-\theta)}{n}$	$Var\left[\bar{X}\right]=\frac{\tau^2}{n}\cdot(1-f)$
	7	Estimat for 6	$V\hat{a}r(\hat{\theta})=\frac{\hat{\theta}(1-\hat{\theta})}{n-1}(1-f)$	$Var(\hat{\theta})=\frac{\hat{\theta}(1-\hat{\theta})}{n}$	$V\hat{a}r\left[\bar{X}\right]=\frac{s^2}{n}\cdot(1-f)$
	8	Estimat for populationstotalen, $N\bar{\mu}$	$N\hat{\theta}$		$N\bar{x}$
	9	Estimat for usikkerhed på 7, $Var(N\bar{X})=\frac{N^2}{n}\cdot\tau^2\cdot(1-f)$	$V\hat{a}r(N\hat{\theta})=N^2\cdot V\hat{a}r(\hat{\theta})$		$V\hat{a}r(N\bar{X})=N^2\cdot V\hat{a}r(\bar{X})$
	10	$(Approx.)konf_{1-\alpha}(\hat{\theta})$	$\hat{\theta}\pm U_{1-\alpha/2}\cdot\sqrt{V\hat{a}r(\hat{\theta})}$		$\bar{x}\pm U_{1-\alpha/2}\cdot\sqrt{V\hat{a}r(\bar{X})}$
	11	$(Approx.)konf_{1-\alpha}(N\hat{\theta})$	$N\cdot\left[\hat{\theta}\pm U_{1-\alpha/2}\cdot\sqrt{V\hat{a}r(\hat{\theta})}\right]$		$N\cdot\left[\bar{x}\pm U_{1-\alpha/2}\cdot\sqrt{V\hat{a}r(\bar{X})}\right]$
	12	Min n, for $V\hat{a}r(\hat{\theta})=\sigma_0^2$ når θ er kendt - Rund altid OP til nærmeste hele tal	$n\geq\frac{\theta(1-\theta)}{\sigma_0^2+\theta(1-\theta)/N},\theta\approx\hat{\theta}$		$n\geq\frac{\tau^2}{\sigma_0^2+\frac{\tau^2}{N}},\tau^2\approx s^2$
	13	Min n, for $V\hat{a}r(\hat{\theta})=\sigma_0^2$ når θ er Ukendt - Rund altid OP til nærmeste hele tal	$n\geq\frac{1/4}{\sigma_0^2+1/4N}$, der svarer til worstcase: θ = 0,5		$n\geq\frac{1/4}{\sigma_0^2+\frac{1}{4N}}$, der svarer til worstcase: θ = 0,5
	14	Min n, for konf max L_0 bred, - Rund altid OP til nærmeste hele tal	$n\geq\frac{\theta(1-\theta)}{\left(\frac{L_0}{2\cdot U_{1-\alpha/2}}\right)^2+\frac{\theta(1-\theta)}{N}},\theta\approx\hat{\theta}.$ Er θ ukendt anvendes θ=0,5		$n\geq\frac{\theta(1-\theta)}{\left(\frac{L_0}{2\cdot U_{1-\alpha/2}}\right)^2+\frac{\theta(1-\theta)}{N}},\theta\approx\hat{\theta}.$ Er θ ukendt og $\hat{\theta}$ ej kan estimeres anvendes θ=0,5

Kap 15: Stikprøver – stratificeret udvælgelse

23. maj 2002

		Begreb	(A) TILFÆLDIG ALLOKERING	(B) PROPORTIONAL ALLOKERING	(C) OPTIMAL ALLOKERING
STRATIFICERET UDVÆLGELSE (generelt variation)			Kendetegn: Tilfældigt udvalgt stikprøve uden hensyn til variation indenfor strata (m uafhængige stikprøv.)	Kendetegn: $f_j = n/N$ for alle m strata	Kendetegn: n_j vælges så $var(X_S)$ minimeres (Kræver kendskab til var i strata inden opdeling!)
	0	Populationgennemsnittet	$\overline{\mu}$	$\overline{\mu}$	$\overline{\mu}$
	1	Stratum gennemsnittet (stor forskel $\Rightarrow$ fordel at gå fra (A) til (B))	$\overline{\mu}_j = \frac{1}{N_j}\sum_{v=1}^{N_j}\mu_{vj} = \theta_j$	$\overline{\mu}_j = \frac{1}{N_j}\sum_{v=1}^{N_j}\mu_{vj} = \theta_j$	$\overline{\mu}_j = \frac{1}{N_j}\sum_{v=1}^{N_j}\mu_{vj} = \theta_j$
	2	Stratumvægten	$W_j = {}^{N_j}\!/_{N}$	$W_j = {}^{n_j}\!/_{n}$, da $f_j = f \; \forall \; j$	$W_j = {}^{N_j}\!/_{N}$
	3	Stratum gennemsnittet (stor forskel $\Rightarrow$ fordel at gå fra (B) til (C))	$\tau_j^2 = \frac{N_j}{N_j - 1}\cdot\theta_j(1-\theta_j)$	$\tau_j^2 = \frac{N_j}{N_j - 1}\cdot\theta_j(1-\theta_j)$	$\tau_j^2 = \frac{N_j}{N_j - 1}\cdot\theta_j(1-\theta_j)$
	4	Stikprøvegennemsnittet	$\overline{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}$, når x_{ij} er obs. af stok.var, X_{ij}	$\overline{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}$, når x_{ij} er obs. af stok.var, X_{ij}	$\overline{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}$, når x_{ij} er obs. af stok.var, X_{ij}
	5	Stikprøvevariansen	$s_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2$	$s_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2$	$s_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2$
	6	(middelret) Skøn over populationsgennemsnittet, $\overline{\mu}$	$\overline{\mu} \approx \overline{X}_S = \sum_{j=1}^{m} W_j \cdot \overline{X}_j$, når x_{ij} = observeret værdi af stok.var.,X_{ij}	$\overline{\mu} \approx \overline{X}_P = \sum_{j=1}^{m} W_j \cdot \overline{X}_j = \sum_{j=1}^{m}\sum_{i=1}^{n_j} x_{ij}$ når x_{ij} = observeret værdi af stok.var.,X_{ij}	$\overline{\mu} \approx \overline{X}_S = \sum_{j=1}^{m} W_j \cdot \overline{X}_j$, når x_{ij} = observeret værdi af stok.var.,X_{ij}
	7	Usikkerhed på 6	$var\left[\overline{X}_S\right] = \sum_{j=1}^{m} W_j^2 \cdot \frac{\tau_j^2}{n_j}(1-f_j), f_j = n_j / N_j$	$\mathrm{var}\left[\overline{X}_P\right] = \frac{1-f}{n}\sum_{j=1}^{m} W_j \cdot \tau_j^2$	$\mathrm{var}\left[\overline{X}_O\right] = \frac{1}{n}\left(\sum_{j=1}^{m} W_j \cdot \tau_j\right)^2 - \frac{1}{N}\cdot\sum_{j=1}^{m} W_j \cdot \tau_j^2$
	8	Skøn over 7	$v\hat{a}r\left[\overline{X}_S\right] = \sum_{j=1}^{m} W_j^2 \cdot \frac{s_j^2}{n_j}(1-f_j), f_j = n_j / N_j$	$v\hat{a}r\left[\overline{X}_P\right] = \frac{1-f}{n}\sum_{j=1}^{m} W_j \cdot s_j^2$	$\mathrm{var}\left[\overline{X}_O\right] = \frac{1}{n}\left(\sum_{j=1}^{m} W_j \cdot s_j\right)^2 - \frac{1}{N}\cdot\sum_{j=1}^{m} W_j \cdot s_j^2$
	9	Populationstotalen, $N\overline{\mu}$ estimeres ved	$N\overline{X}_S$	$N\overline{X}_P$	$N\overline{X}_O$
	10	Usikkerheden på 9, $Var\left[N\overline{X}_S\right]$ estimeres ved	$v\hat{a}r\left[N\overline{X}_S\right] = N^2 \cdot v\hat{a}r\left[\overline{X}_S\right]$	$v\hat{a}r\left[N\overline{X}_P\right] = N^2 \cdot v\hat{a}r\left[\overline{X}_P\right]$	$v\hat{a}r\left[N\overline{X}_O\right] = N^2 \cdot v\hat{a}r\left[\overline{X}_O\right]$
	11	Approx. $Konf_{1-\alpha}$ =	$\overline{x}_S \pm U_{1-\alpha/2}\cdot\sqrt{V\hat{a}r\left[\overline{X}_S\right]}$	$\overline{X}_P \pm U_{1-\alpha/2}\cdot\sqrt{V\hat{a}r\left[\overline{X}_P\right]}$	$\overline{X}_O \pm U_{1-\alpha/2}\cdot\sqrt{V\hat{a}r\left[\overline{X}_O\right]}$
	12	Stikprøvestørrelse der giver variansen $\mathrm{var}(\overline{X}_s) = \sigma_0^2$ - Rund altid OP til nærmeste hele tal	Som for ustratificeret???	$n \geq \frac{\sum_{j=1}^{m} W_j \tau_j^2}{\sigma_0^2 + \frac{1}{N}\sum_{j=1}^{m} W_j \tau_j^2}, \tau_j^2 \approx s_j^2$	$n \geq \frac{\left(\sum_{j=1}^{m} W_j \tau_j\right)^2}{\sigma_0^2 + \frac{1}{N}\sum_{j=1}^{m} W_j \tau_j^2}, \tau_j^2 \approx s_j^2$
		Sammenligning af de tre varianser	Der fås ved lange udregninger: $\mathrm{var}\left[\overline{X}\right] \geq \mathrm{var}\left[\overline{X}_P\right] \geq \mathrm{var}\left[\overline{X}_O\right]$ jo større forskel der er på var i de m strata, jo større bliver gevinsten ved at skifte metode		