

Stikprøvetæori vejning/vægte/

Sommerskole

om vejning

1. Traditionel vejning ("lagkageregning"). Brug af "officielle" variable så som køn, alder, geografi.
2. Politisk vejning, en sofistikeret version af ovenstående her inddrages panel tankegangen

Stratifikation

<i>Strata</i>	<i>Antal</i>	<i>Univers</i>	<i>stikprøve</i>	<i>stikprøve</i>	<i>sum</i>	<i>gns</i>	<i>var</i>
		<i>vægte</i>	<i>antal</i>	<i>vægte</i>			
1	N_1	$W_1 = \frac{N_1}{N}$	n_1	$w_1 = \frac{n_1}{n}$	$Y_{1.}$	$\bar{Y}_{1.}$	S_1^2
2							
K	N_K	$W_K = \frac{N_K}{N}$	n_K	$w_K = \frac{n_K}{n}$	$Y_{K.}$	$\bar{Y}_{K.}$	S_K^2
<i>TOTAL</i>	N	1	n	1	$Y = Y_{..}$	—	—

Stratifikation

$$E(\bar{y}_{k.}) = \bar{Y}_k.$$

$$V(\bar{y}_{k.}) = \frac{(N_k - n_k)}{N_k} \frac{1}{n_k} S_k^2$$

$$\bar{y}_{strat} = \sum_{k=1}^K W_k \bar{y}_{k.}$$

$$E(\bar{y}_{strat}) = \bar{Y}$$

$$V(\bar{y}_{strat}) = \sum_{k=1}^K W_k^2 \frac{N_k - n_k}{N_k} \frac{1}{n_k} S_k^2$$

Vejning

$$\bar{y}_{strat} = \sum_{k=1}^K W_k \bar{y}_{k.} = \sum_{k=1}^K W_k \frac{1}{n_k} \sum_{m=1}^{n_k} y_{km} = \frac{1}{n} \sum_{k=1}^K W_k \frac{n}{n_k} \sum_{m=1}^{n_k} y_{km}$$

brug stikprøve vægte $w_k = \frac{n_k}{n}$

$$\frac{1}{n} \sum_{k=1}^K W_k \frac{1}{w_k} \sum_{m=1}^{n_k} y_{km} = \frac{1}{n} \sum_{k=1}^K \sum_{m=1}^{n_k} \frac{W_k}{w_k} y_{km}$$

$$\frac{1}{n} \sum_{k=1}^K \sum_{m=1}^{n_k} (vgt_k) y_{km}$$

$$\text{vægtene er } \frac{W_k}{w_k} = \frac{\text{univers_vægt}}{\text{stikprøve_vægt}}$$

Stratifikation

	alder	Univers	W_k	n_k	w_k	vgt	DF	DF	
							i stik	vægtet	
M	15-40	874.044	19,8	165	13,3	1,49	10	14,9	
M	40-69	1.057.619	24,0	368	29,6	0,81	30	24,3	
M	70+	233.798	5,3	77	6,2	0,86	12	10,3	
K	15-40	851.737	19,3	179	14,4	1,34	6	8,0	
K	40-69	1.053.420	23,9	359	28,9	0,83	35	29,1	
K	70+	340.962	7,7	96	7,7	1,00	4	4,0	
		4.411.580	100	1244	100	-	97	90,6	

Dansk Sociologi december 2009. Data er fra ESS år 2006 dvs. 3. runde

Meningsmåling

I 1997 udvalgte Gallup simpelt tilfældigt (så godt som muligt) 990 personer, som de stillede nogle spørgsmål.

Det sidste folketingsvalg var i 1994. Her fik liste A (socialdemokraterne) 34,6% af stemmerne og dermed fik de øvrige partier 65,4%

1. spørgsmål:

Hvilket part stemte De på ved FV i 1994?

(Stikprøve $n=990$)

parti	Stemte	procent		
A	278	28%		
Andet	712	72%		
TOTAL	990	100%		

Overvej hvilken fordeling du vil bruge til at beskrive ovenstående

Der stilles også spørgsmålet hvilket part ville De stemme på hvis der var FV i morgen?

Politisk indeks

De to spørgsmål er nu "krydses" mod hinanden.

vil stemme	stemte	stemte	total	
	A	Andet		
A	248	44	292	
Andet	30	668	698	
TOTAL	278	712	990	

$A = \{\text{vil stemme på A}\}$ $B = \{\text{stemte på A}\}$

$P(A)$ $P(A|B)$ og $P(A|B^c)$ B^c er B's komplementærhændelse

opskriv udtrykket for $P(A)$ ved brug af B og B^c

Er de to hændelser A og B uafhængige?

Politisk indeks

vil stemme	stemte	stemte	total	
	A	Andet		
A	248	44	292	
Andet	30	668	698	
TOTAL	278	712	990	

$$P(\text{vil stemme A} \mid \text{stemte på A}) = \frac{248}{278} = 0,8921$$

$$P(\text{vil stemme A} \mid \text{stemte på andet}) = \frac{44}{712} = 0,0618$$

$$P(\text{vil stemme på A}) = P(\text{vil A} \mid \text{stemte A})P(\text{stemte A}) + P(\text{vil A} \mid \text{stemte andet})P(\text{stemte andet})$$

$$= 0,8921 \cdot 0,346 + 0,0618 \cdot 0,654 = 34,9\% \quad (\text{mod si på } \frac{292}{990} = 29,5\%)$$

Politisk indeks

to strataer. Stemte på A, stemte på andet ved FV i 1994

strata		univers vgt			andel		
	N_k	W_k	n_k		\hat{p}_k	$W_k * \hat{p}_k$	
stemte A	1.453.200	0,346	278	248	0,8921		
stemte -A	2.746.800	0,654	712	44	0,0618		
TOTAL	4.200.000	1,000	990			0,349	

$$V(\hat{p}_{strat}) = \sum_{k=1}^2 W_k^2 \frac{N_k - n_k}{N_k} \frac{1}{n_k} S_k^2 \approx \sum_{k=1}^2 W_k^2 \frac{1}{n_k} S_k^2 =$$

$$\sum_{k=1}^2 W_k^2 \frac{1}{n_k} P_k (1 - P_k) = (0,87\%)^2$$

$$V(\hat{P}) = \frac{N-n}{N} \frac{1}{n} P(1 - P) = (1,45\%)^2$$

se også regneark

Ved valget den 11. marts 1998 fik A 35,9%

Universvægte: $W_K = \frac{N_K}{N}$ kan bruges til prop allokering
stikprøvevægte: $w_K = \frac{n_K}{n}$ er resultat af indsamlingen

"analyseinstitutsvægte" $vgt = \frac{W_K}{w_K} = \frac{\text{univers-vgt}}{\text{stikprøve-vgt}}$ "man vejer intuitivt på plads (s 151 i S. SAS)

Disse vægte er ofte dem der udleveres til når data er indsamlet af et institut (herunder også ESS ?)

Udvalgsvægte: $\frac{N_K}{n_K}$ den reciprokke udvalgsbrøk pr. stratum bruges når SAS skal beregne

variansen af en stratificeret stikprøve

$$DEFF_w = \frac{n \sum_{k=1}^n vgt_i^2}{(\sum_{k=1}^n vgt_i)^2} = \text{populært sagt, den andel man skal øge stikprøven}$$

med, for at opnå variansen

for en simpel tilfældig stikprøve. (s. 172 i S. SAS)