

Mandagsopgaven

Opgave 1

a

Vi ved variansen er givet ved:

$$V(\bar{y}_{si}) = \frac{(N-n)}{N} \frac{1}{n} S^2$$

Vi indsætter vores værdier:

$$V(\bar{y}_{si}) = \frac{(931-49)}{931} * \frac{1}{49} * 2248.6^2 \Leftrightarrow$$

$$V(\bar{y}) = 97757 \approx 313^2$$

Finder konfidensintervallet:

$$\bar{x} \pm 1.96 * \sqrt{V(\bar{y})} \Leftrightarrow$$

$$\bar{x} \pm 1.96 * \sqrt{312,7^2} \Leftrightarrow$$

$$\bar{x} \pm 612,89$$

b

$$\bar{y}_{si} = \frac{S^2}{\left(\frac{L_0}{2*1.96}\right)^2 + \frac{1}{N} S^2} \Leftrightarrow$$

$$\bar{y}_{si} = \frac{2248.6^2}{\left(\frac{126.61}{2*1.96}\right)^2 + \frac{1}{931} * 2248.6^2} \Leftrightarrow$$

$$\bar{y}_{si} = 925,5$$

c

Stratum	N _k	S _k	Vægt
HR	91	4710,6	0,0977
Sjælland	214	1732,7	0,2297
Fyn	105	940,3	0,1128
Jylland	521	1823,3	0,5596

Formlen er følgende:

$$V(\bar{y}_p) = \frac{N-n}{N} * \frac{1}{n} \sum_{k=1}^K W_k S_k^2$$

$$V(\bar{y}_p) = \frac{931-49}{931} * \frac{1}{49} \sum_{k=1}^K W_k S_k^2 \Leftrightarrow$$

$$V(\bar{y}_p) = \frac{931-49}{931} * \frac{1}{49} (4710,6^2 * 0,0977 + 1732,7^2 * 0,2297 + 940,3^2 * 0,1128 + 1823,3^2 * 0,5596) \Leftrightarrow$$

$$V(\bar{y}_p) = \frac{931-49}{931} * \frac{1}{49} * 4837431,12 \Leftrightarrow$$

$$V(\bar{y}_p) = 93526,30$$

$$Se(\bar{y}_p) = \sqrt{93526,30} = 305,82$$

Nu finder vi den optimale:

$$V(\bar{y}_{opt}) = \frac{1}{n} * \left(\sum_{k=1}^K W_k S_k\right)^2 - \frac{1}{N} \sum_{k=1}^K W_k S_k^2 \Leftrightarrow$$

$$V(\bar{y}_{opt}) = \frac{1}{49} * (\sum_{k=1}^K W_k S_k)^2 - \frac{1}{931} * 93526,30 \Leftrightarrow$$

$$V(\bar{y}_{opt}) = \frac{1}{49} * 1990,11^2 - \frac{1}{931} * 93526,30 \Leftrightarrow$$

$$V(\bar{y}_{opt}) = 80827,30 - 5195,95 \Leftrightarrow$$

$$V(\bar{y}_{opt}) = 75631,35 \Leftrightarrow$$

$$Se(\bar{y}_{opt}) = 75631,35^2 \approx 275,01$$

Hermed opnåes det, at

$$Se(\bar{y}_{opt}) < Se(\bar{y}_p) < Se(\bar{y}) \Leftrightarrow \\ 275,01 \leq 305,82 \leq 313$$

Opgave 2

Vi finder DEFF for propotional og optimal stratificering.

$$DEFF_p = \frac{Se(\bar{y}_p)}{Se(\bar{y})} \Leftrightarrow \frac{305,82}{313} = 0,977$$

$$DEFF_{opt} = \frac{Se(\bar{y}_{opt})}{Se(\bar{y})} \Leftrightarrow \frac{275,01}{313} = 0,879$$

Her ses det, at vi reducerer variansen meget mere ved at benytte optimal og propotional stratificering ift. en tilfældig stikprøve.

Opgave 3

$$\left(\frac{4200000 - 1032}{4200000} * \frac{1}{1032} * \frac{1032}{1032 - 1} * 0,264 * (1 - 0,264) \right)^{\frac{1}{2}} = 0,27$$

Dette må være større end F, da A har større tilslutning, hvilket påvirker positivt.

Opgave 4

$$1.96 * \left(\frac{4200000 - n}{4200000} * \frac{1}{n} * \frac{n}{n - 1} * 0,264 * (1 - 0,264) \right)^{\frac{1}{2}} = 0,0145 \Leftrightarrow \\ 1.96 * \left(\frac{4200000 - n}{4200000} \right)^{\frac{1}{2}} * \frac{1}{n^{\frac{1}{2}}} * \left(\frac{n}{n - 1} \right)^{\frac{1}{2}} * 0,264^{\frac{1}{2}} * (1 - 0,264^{\frac{1}{2}}) = 0,0145 \Leftrightarrow \\ 1.96 * \left(\frac{4200000 - n}{4200000} \right)^{\frac{1}{2}} * \left(\frac{n}{n - 1} \right)^{\frac{1}{2}} * 0,264^{\frac{1}{2}} * (1 - 0,264^{\frac{1}{2}}) = 0,0145 * n^{\frac{1}{2}} \Leftrightarrow \\ 1.96 * 0,264^{\frac{1}{2}} * (1 - 0,264^{\frac{1}{2}}) = 0,0145 * (n - 1)^{\frac{1}{2}} \Leftrightarrow \\ n = 3551$$

5

$$C + D + V \Leftrightarrow 12,9 + 5,3 + 17,7 = 35,9\%$$

Dette indsættes i.

$$1.96 * \left(\frac{4200000 - 1032}{4200000} * \frac{1}{1032} * \frac{1032}{1032 - 1} * 0,359 * (1 - 0,359) \right)^{\frac{1}{2}} = 0,0293 \Leftrightarrow$$

$$35,9 - 2,9 = 33$$

Vi ser, at den gamle stemmeprocent ikke er indeholdt i konfidensintervallet, hvorved vi forkaster hypotesen.

6

```

proc import datafile="/courses/d284cd65ba27fe300/Hurtigt/ESS9e03_rensset.sav"
out=ess9 replace;
proc contents data=ess9;
run;
proc format;
value STRAT
    1='unge m nd '
    2='mid. m nd '
    3=' ldre m nd '
    4='unge kvinder '
    5='mid. kvinder '
    6=' ldre kvinder '
;
data DK;
set ESS9;
if cntry='DK';
if agea < 40 then age1=1;
if 40<= agea< 70 then age1=2;
if agea =>70 then age1=3;
if age1=1 and gndr=1 then STRAT =1;
if age1=2 and gndr=1 then STRAT =2;
if age1=3 and gndr=1 then STRAT =3;
if age1=1 and gndr=2 then STRAT =4;
if age1=2 and gndr=2 then STRAT =5;
if age1=3 and gndr=2 then STRAT =6;
if trstep=. then trstep=5.5;
if trstun=. then trstun=5.5;
run;

proc means data=dk n maxdec=2 mean std;
class strat;
var trstep;
format strat strat.;
run;

*Opg 6*;
proc freq data=DK;
table strat;
format strat strat.;
run;

proc freq data=DK;
table trstep;
run;

```

Trust in the European Parliament				
trstep	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No trust at all	62	3.94	62	3.94
1	42	2.67	104	6.62
2	93	5.92	197	12.53
3	127	8.08	324	20.61
4	153	9.73	477	30.34
5	288	18.32	765	48.66
5.5	114	7.25	879	55.92
6	250	15.90	1129	71.82
7	223	14.19	1352	86.01
8	150	9.54	1502	95.55
9	51	3.24	1553	98.79
Complete trust	19	1.21	1572	100.00

Det kan udledes, at 114 personer har fået værdien 5,5. Dette er alle de respondenter, der ikke har svaret på spørgsmålet. Det vil nok være mere rigtigt at fjerne disse, da de kan trække middelværdien enten op eller ned, hvis svarene enten er meget høje eller meget lave.

7

Vi beregner:

$$\frac{1572 - 120}{1572} * \frac{1}{120} * 2,18^2 = 0.03658 = 0.03658 = 0,19^2$$

Hvorved bevist

8+9

```
*Opg 8*;
proc surveyselect data=dk seed=1 out=stik1 n=120;
run;

*Opg 9*;
proc surveymeans data=stik1;* N=480000;
var trstep;
*weight pspwght;
*weight anweight;
run;
```

Statistics						
Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
trstep	Trust in the European Parliament	120	5.462500	0.213199	5.04034443	5.88465557

Det kan udledes, der er en forskel på 0,02 i standardafvigelse fra den teoretiske med den reelle. Dette er tæt på den teoretiske, hvilket er positivt.

10

```
proc surveyselect data=dk seed=1 n=120 out=stik2 reps=5;
run;
proc surveymeans data=stik2; * N=480000;
var trstep;
*weight pspwght;
*weight anweight;
run;

proc means data=dk;
var trstep;
run;
```

Statistics						
Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
trstep	Trust in the European Parliament	600	5.281667	0.089811	5.10528479	5.45804854

Ovenstående er den reelle middelværdi og varians er $0,0898^2$. Den teoretiske beregnes vha. SAS til:

Analysis Variable : trstep Trust in the European Parliament				
N	Mean	Std Dev	Minimum	Maximum
1572	5.2150127	2.1789494	0	10.0000000

Her er der en meget højere varians, men finder et lavere gennemsnit.

11

STRAT	N	Mean	Std Dev (S)
unge mænd	279	5,65	2,17
mid. Mænd	420	4,85	2,15
ældre mænd	147	4,58	2,61
unge kvinder	230	6,13	1,65
mid. Kvinder	359	5,09	2,07
ældre kvinder	135	4,92	2,25
i alt	1.572	5,22	2,18

Dette er ikke en fremragende stratifikation. Det ses, at den samlede varians er lavere end nogle af de enkelte stratifikationer, hvorved stratifikationerne er irrelevante. Som Hans også sagde, er stratifikationer baseret på køn og alder ofte dårlige, hvilket dette er et eksempel på. Stratifikationer, der sænker variansen, er gode, da dette er målet ved at lave stratifikationer.

12

```
proc sort data=DK;
by strat;
run;
proc surveyselect data=DK n=20 out=udtrak1 seed=2008;
strata strat;
run;
```

I koden laver SAS stratifikationer, hvorved alle stratifikationer har samme andel i hver stratifikation. Ved at bruge kodelinjen,

```
proc surveymeans data=udtrak1;  
run;
```

Kan man finde middelværdien i stikprøven. Teoretisk er dette dog forkert, da andelen i stikprøven ikke stemmer overens med virkeligheden. (16% af den danske befolkning er ikke mænd over 70 år) Dette vil kunne rettes med at tilføje vægte mm., der tager højde for dette. Fx give mænd over 70 mindre betydning i stikprøven.

13

```
proc surveyselect data=dk seed=2008 out=udtrak2 rate=0.05;  
strata strat;  
run;  
proc means data=udtrak2;  
var trstep;  
run;
```

Programmet sørger for at trække de 5%, der er proportionalt allokeret, hvorfor vægtene ikke er nødvendige.