

Surveymetodik

Föreläsning 12

Måns Magnusson

Avd. Statistik, LiU

- 1 Grafik i surveyer
- 2 Analys i surveyer
- 3 Komplexa surveyer
- 4 Sammanfattning

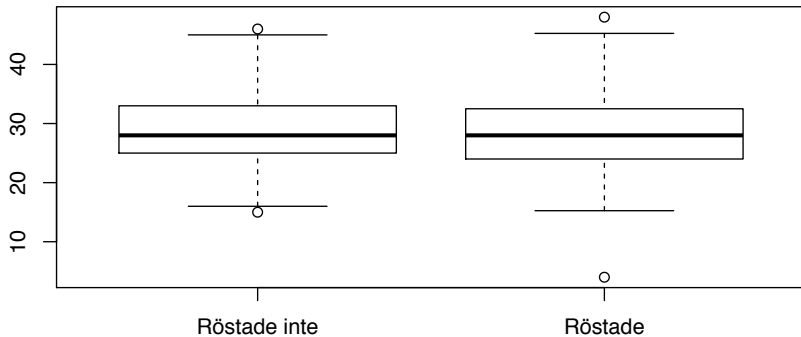
Section 1

Grafik i surveyer

- Varför tala om grafik i surveyer? Är det inte samma sak?
- Finns två huvudorsaker:
 - Surveydata är ofta stort (många observationer)
 - Vikterna behöver hanteras (bortfallskorrigerade grafer)
- Det finns tre huvudsakliga angreppssätt:
 - Basera graferna på skattningar
 - Explicit indikera vikterna i graferna
 - Dra ett urval från urvalet och visualisera
- Detta avnitt bygger huvudsakligen på Lumley (2010, kap 4) och på implementationer i surveypaketet

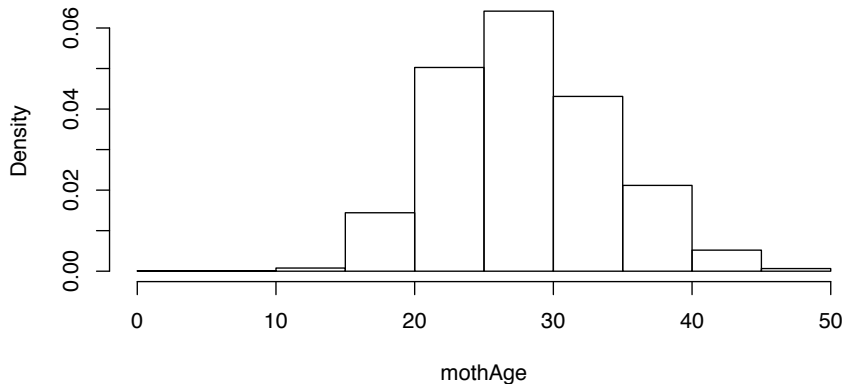
Boxplot

```
svyboxplot(mothAge ~ Valdeltagande, design = svy2010design)
```



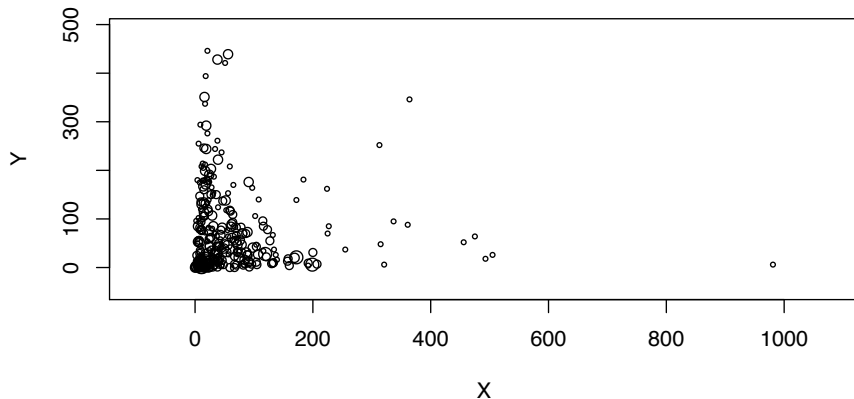
```
svyhist(formula = ~mothAge, design = svy2010design)
```

Histogram of mothAge



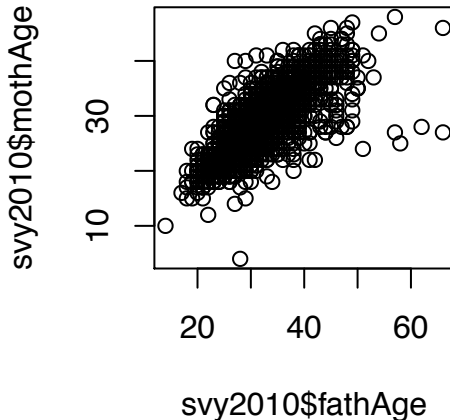
Scatterplot with weights

```
svyplot(LARGE92 ~ SMALL92, design = agSTRAT, style = "bubble")
```



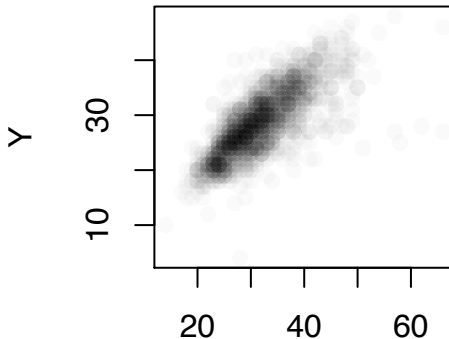
Problem med många observationer

```
plot(svy2010$fathAge, svy2010$mothAge)
```



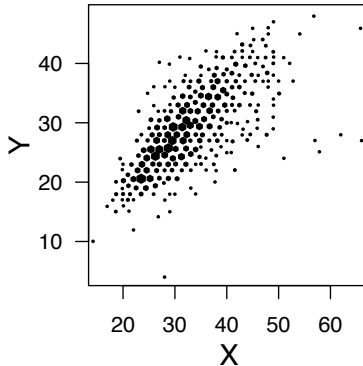
Problem med många observationer

```
svyplot(mothAge ~ fathAge, design = svy2010design, style = "transparent", pch = 19,  
        alpha = c(0, 0.02))
```



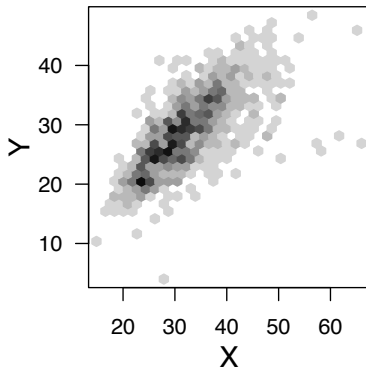
Problem med många observationer

```
svyplot(mothAge ~ fathAge, design = svy2010design, style = "hex")
```



Problem med många observationer

```
svyplot(mothAge ~ fathAge, design = svy2010design, style = "hex")
```



Section 2

Analys i surveyer

- Vid analys är den centrala frågan inferens. Vill vi dra slutsatser om...
 - Vår ändliga population?
 - En statistisk modell?
- Hur påverkas våra analyser av urvalsdesignen?
- Vid OSU med 'oänlig' population fungerar analyserna som vanligt

- Våra analyser påverkas på samma sätt som konfidensintervallen
 - $deff < 1$ vi blir säkrare än vid OSU
 - $deff > 1$ vi blir osäkrare än vid OSU
- På samma sätt påverkas χ^2 -test, t-test och regressionsmodeller
- Vi korregerar detta med ex. Wald-test eller Rao-Scott korrektion (se Lohr (2009, kap 10.3.2))

- Vi kan vara intresserade av att dra slutsatser om en ändlig population
- Observerar vi alla element saknar vi osäkerhet

Section 3

Komplexa surveyer

- Alla metoder vi beskrivit så här långt kan användas i en och samma undersökning
- Vi kan dra olika typer av urval i olika steg
- Samlingsnamnet för dessa metoder är **Komplexa surveyer**

Byggnaders energianvändning, tekniska status och inomhusmiljö (BETSI) Boverket (2009)

Boverket

- **Syfte:** Kartlägga det svenska byggnadsbeståndet
- **Målpopulation:** Byggnader med taxeringsvärde på minst 50 tkr och med minst 50 m² samt individer i småbostadshus eller lägenhet
- **Urval:**
 - Flerstegsurval
 - Steg I: Stratifierat klusterurval av kommuner (pps/ π ps)
 - Steg II: Stratifierat klusterurval av värderings/taxeringsenhet (OSU och pps/ π ps)
 - Steg III: Klusterurval av byggnad (OSU)
 - Steg IV: Lägenhet (OSU)
- **Bortfall:** 21-35 % (beroende på byggnad)
- **Datainsamlingsmetod:** Besiktningar och pappersenkäter
- **Periodicitet:** Ett tillfälle (?)

- Hur reder vi ut de olika nivåerna?
 - Börja på lägsta nivån och gå uppåt
- Exempel: Stratifierad population och trestegs klusterurval (*)

$$\bar{y}_{hij} = \frac{1}{m_{hij}} \sum_{k \in \mathcal{S}_{hij}} y_{hijk}$$

$$\hat{t}_{hij} = M_{hij} \cdot \bar{y}_{hij}$$

$$\hat{t}_{hi} = \sum_{j \in \mathcal{S}_{hi}} \frac{M_{hi}}{m_{hi}} \cdot \hat{t}_{hij}$$

$$\hat{t}_h = \sum_{i \in \mathcal{S}_h} \frac{M_h}{m_h} \hat{t}_{hi}$$

$$\hat{t} = \sum_{h=1}^H \hat{t}_h$$

- Varianserna kan utredas på samma sätt steg för steg (överkurs)
- I praktiken framställs ofta urvalsvikter på elementnivå
- Har vi tillgång till urvalsvikterna kan punktskattningar och medeltal beräknas enkel på följande sätt

$$\hat{t} = \sum_{h=1}^H \sum_{i \in \mathcal{S}_h} \sum_{j \in \mathcal{S}_{hi}} \sum_{k \in \mathcal{S}_{hij}} w_{hijk} \cdot y_{hijk}$$

eller

$$\hat{y}_{\mathcal{U}} = \frac{\sum_{h=1}^H \sum_{i \in \mathcal{S}_h} \sum_{j \in \mathcal{S}_{hi}} \sum_{k \in \mathcal{S}_{hij}} w_{hijk} \cdot y_{hijk}}{\sum_{h=1}^H \sum_{i \in \mathcal{S}_h} \sum_{j \in \mathcal{S}_{hi}} \sum_{k \in \mathcal{S}_{hij}} w_{hijk}}$$

- Detta räcker dock inte för att beräkna varianserna

Section 4

Sammanfattning

■ Obundet slumpmässigt urval

- Alla urvalsenheter/kluster har samma inklusionssannolikhet
- Fördelar
 - Enkel
- Nackdelar
 - Många gånger ineffektivt

■ Systematiskt urval

- Påminner om OSU, men är egentligen ett klusterurval
- Fördelar
 - Kan användas när det inte finns någon ram
- Nackdelar
 - Periodicitet i ramen

■ Stratifierade urval

- Målpopulationen (och rampopulationen) delas in i strata
- Varje strata kan ses som en egen undersökning - vanligt med OSU inom strata
- Vi vill stratifiera så det är stor skillnad mellan och liten skillnad inom strata
- Fördelar:
 - Effektivare urval
 - Garanterar respondenter i varje grupp
 - Kan kontrollera precision i redovisningsgrupper
 - Ofta hyfsat enkelt
- Nackdelar
 - Något mer komplicerat än OSU

■ Klusterurval

- Målpopulationen består av naturliga kluster/grupper
- Enstegs, tvåstegs eller flerstegs klusterurval
- Vi vill ha kluster med stor skillnad inom kluster och liten skillnad mellan kluster
- Fördelar
 - Kan vara mycket billigare än OSU
 - Kan vara enda möjligheten om ram saknas
- Nackdelar
 - Ineffektivare än OSU
 - Mer komplicerade beräkningar

■ PPS/ π PS-urval

- Olika inklusionssannolikheter för olika element/kluster
- En form av “kontinuerlig” stratifiering
- Fördelar
 - Effektiv design
 - Självvägande kluster
- Nackdelar
 - Komplicerat att dra urval
 - Mer komplicerade beräkningar

- Den vanliga estimatoren (Horwitz-Thopson)
 - Väntevärdesriktig
 - Enkel
- Kvotskattning
 - Bias vid mindre urval
 - En hjälpvariabel kan användas
 - Använder hjälpinformation med modellen $y = Bx$
 - Bra att använda vid klusterurval vid olika klusterstorlekar
- Regressionsskattning
 - Bias vid mindre urval
 - Flera hjälpvariabler kan användas
 - Använder hjälpinformation med modellen $y = \mathbf{B}x$
 - Används vid kalibrering

- Kvalitetsdimensionerna i en undersökning
 - Innehåll / relevans
 - Tillförlitlighet
 - Jämförbarhet och sam användbarhet
 - Aktualitet
 - Tillgänglighet och förståelighet
- Felen som kan uppstå i en undersökning
 - Urvalsfel
 - Specifikationsfel
 - Bortfallsfel
 - Mätfel
 - Täckningsfel
 - Bearbetningsfel

- De olika stegen i en undersökning
 - Faslå behov och kartlägga området
 - Design och planering
 - Förberedelser och tester
 - Datainsamling
 - Databearbetning
 - Analyser
 - Rapportering och kommunikation

■ Stegen vid enkätkonstruktion

- Utveckla enkäten
se Dillman et al. (2005); Dahmström (2011); SCB (2001); Japec et al. (1997)
- Expertgranska enkäten
se Dahmström (2011); SCB (2001); Japec et al. (1997)
- Kognitiva intervjuer/tester av enkäten
se Willis (2004) och SCB (2001, s. 49-60)
- Pilotstudie

■ Bortfallshantering

- Förebygga bortfall
- Genomföra en bortfallsanalys
- Kompensera för uppkommet bortfall
(vanligtvis med regressionsestimation / GREG)

- Boverket, 2009. Statistiska urval och metoder i boverkets projekt betsi. Tech. rep., Boverket.
- Dahmström, K., 2011. Från datainsamling till rapport: att göra en statistisk undersökning, 5th Edition. Studentlitteratur, Lund.
- Dillman, D., Gertseva, A., Mahon-Haft, T., 2005. Achieving usability in establishment surveys through the application of visual design principles. JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM- 21 (2), 183.
- Japac, L., Ahtiainen, A., Hörngren, J., Lindén, H., Lyberg, L., Nilsson, P., 1997. Minska bortfallet. URL http://www.pubkat.scb.se/statistik/_publikationer/0V9999_2000I02_BR_X97%C3%96P9701.pdf
- Lohr, S., 2009. Sampling: design and analysis, 2nd Edition. Thomson.
- Lumley, T., 2010. Complex surveys : a guide to analysis using R. Wiley-Blackwell, Oxford.
- SCB, 2001. Fråga rätt! : utveckla, testa, utvärdera och förbättra blanketter. Stockholm. URL http://www.scb.se/statistik/_publikationer/0V9999_0000T00_BR_X97%C3%96P9701.pdf