

# Survey metodik

## Föreläsning 9

Måns Magnusson

Avd. Statistik, LiU

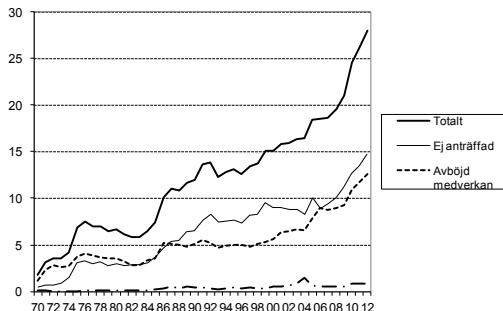
## 1 Bortfall

- Vad är bortfall?
- Bortfallsantaganden
- Bortfallets konsekvenser
- Bortfallets orsaker
- Hantera bortfall
- Redovisa bortfall

# Section 1

## Bortfall

- En del av (i princip) alla undersökningar med respondenter
- Har ökat kraftigt över tid



Figur : Bortfallsandel i AKU 1970-2013 (oviktat), SCB, 2012

- Kvalitetsgränser för bortfall idag: SCB 60% (?), Ekot 50%

- Skillnader i bortfall finns mellan olika datainsamlingsmetoder (\*)
  - högst i web och telefon,
  - lägre i (bra) postala enkäter,
  - enkäter i skolan fortfarande lågt bortfall
- Bortfallet introducerar **fel** i vår undersökning
  - Svarsprocenten är i sig inte nödvändigtvis en bra indikator på bortfallsfel (\*)
  - Det är vid en **bortfallsanalys** vi kan se om det finns risker för bortfallsfel
  - Vi behöver **hjälpinformation** för att både göra bortfallsanalyser och hantera uppkommet bortfall
- Använda bortfall som **indikator** på saker som inte fungerar

Högt bortfall = (troligtvis) dålig undersökning
- **Ramfel** kan betraktas på (ungefär) samma sätt som bortfall

*De element (objekt) i ramen som tillhör målpopulationen och man har planerat att undersöka, men som man ej fick något svar ifrån.  
- Dahmström (2011)*

- Två typer av bortfall finns
  - Objektbortfall
  - Partiellt bortfall
- Antaganden om bortfall
  - Missing completely at random (MCAR)
  - Missing at random (MAR)
  - Not missing at random (NMAR)
- Vi gör **ALLTID** något av ovanstående antaganden!  
Säger någon att de inte gör ett antagande om bortfallet - antar MCAR.

- En av de ansvariga statistikerna tappar/råkar förstöra en låda med enkäter.
- I en levnadsnivåundersökning delar inte personer som inte kan språket.
- I en skolundersökning är en del elever sjuka vid datainsamlingstillfället.
- Cryptosporidiumstudie: Personer som inte har kommunalt vatten tycker att de inte berörs av studien.
- Sexualvanestudie på gymnasiet: Personer som inte har haft sex tycker att studien inte rör dem och deltar inte.



- De direkta konsekvenserna av bortfall i undersökningar är
  - Bortfallet introducerar bortfallsfel i vår undersökning
  - Ett större slumpfel, p.g.a. mindre urval
  - Bias, som inte blir mindre då  $n \rightarrow \infty$
- Både slumpfelet och biasen behöver hanteras

- Vi kan betrakta bortfallet som två strata: svarsstrata ( $R$ ) och bortfallsstrata ( $M$ )
- Populationsmedelvärdet kan skrivas

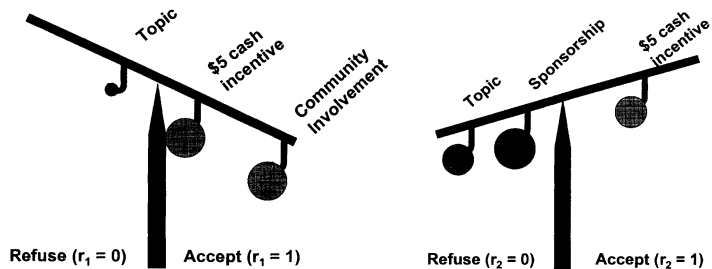
$$\bar{y}_U = \frac{N_R}{N} \cdot \bar{y}_{RU} + \frac{N_M}{N} \cdot \bar{y}_{MU}$$

- Om vi antar att  $E(\hat{y}) = \bar{y}_{RU}$  (ex. vid OSU) så kan vi räkna ut hur stor biasen blir (\*)

$$\text{Bias} = \frac{N_M}{N} (\bar{y}_{RU} - \bar{y}_{MU})$$

- Felets storlek beror på två faktorer
  - Bortfallsandelen  $\left(\frac{N_M}{N}\right)$
  - Skillnaden mellan svarande och icke-svarande  $(\bar{y}_{RU} - \bar{y}_{MU})$

- Vad orsakar bortfall?
- **Leverage-Salient-teorin** för deltagande i studier
  - Det finns **fördelar** för respondenter att delta i studier som...  
ex. Bidra, komma till tals, belöningar m.m.
  - Det finns **nackdelar** för respondenter att delta i studier som...  
ex. Uppgiftslämnarbörda, Oro för resultat/hantering, Misstro, Kränkande frågor
  - Det som påverkar är dels **hur stora** fördelarna/nackdelarna är och **hur viktiga** fördelarna/nackdelarna är **för respondenten**
  - Sammantaget avgör detta om folk deltar i studierna eller inte
- För mer information om denna teori se artikeln Groves et al. (2000).



Figur : Källa: Groves et al. (2000)

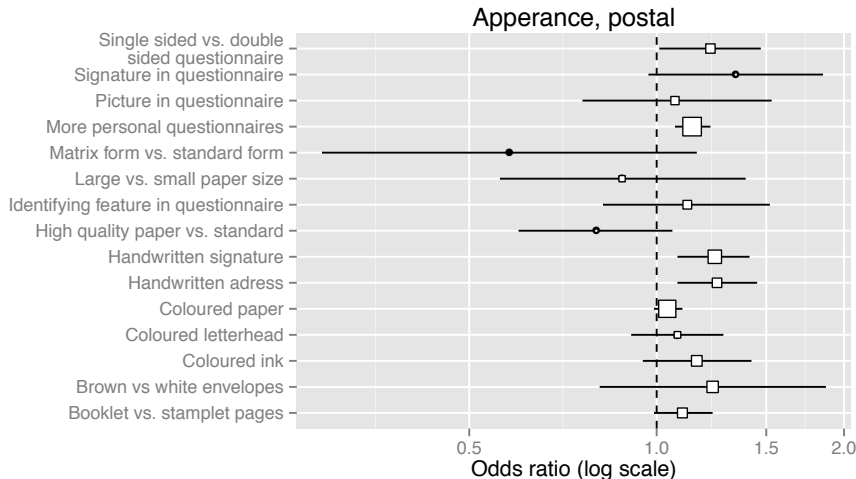
## Subsection 5

### Hantera bortfall

- 1 Förebygga (och förbereda för) bortfall
- 2 Genomföra en bortfallsanalys
- 3 Hantera uppkommet bortfall

- Det billigaste (och mest kotsnadseffektiva) är ofta att försöka **förebygga** (eller förekomma) bortfallsproblem
- Olika datainsamlingsmetoder ger olika bortfall
  - Att kombinera metoder kan minska bortfallet (ex. web + telefon)
- Inkludera variabler för bortfallsanalys och bortfallshantering
- Bra litteratur för att förebygga bortfall är Japac et al. (1997); Edwards et al. (2009)

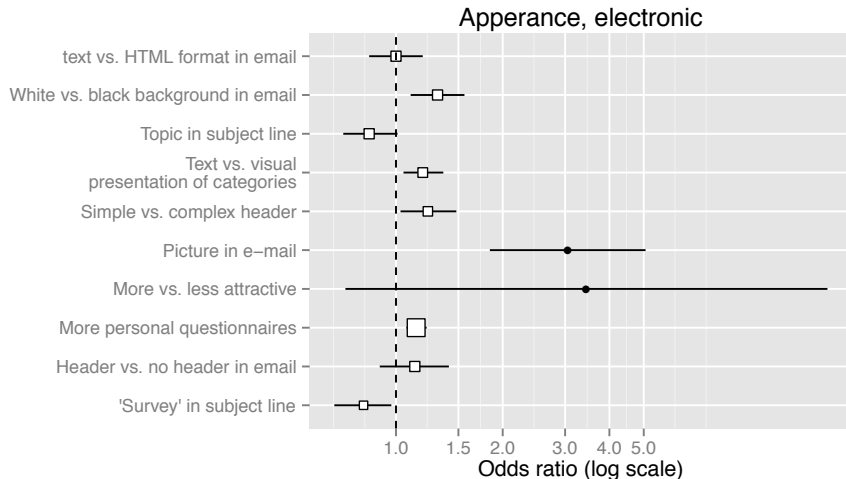
# Steg 1: Förebygga bortfall



Figur : Metastudie om bortfall av Edwards et al. (2009)

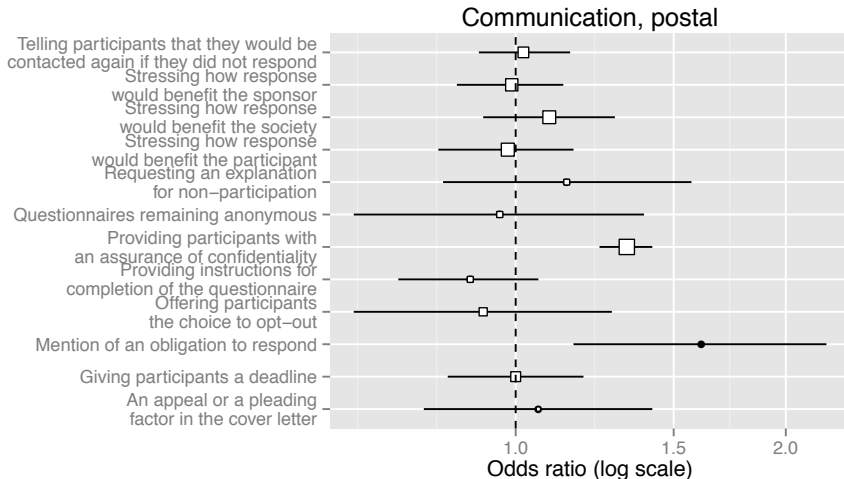


# Steg 1: Förebygga bortfall



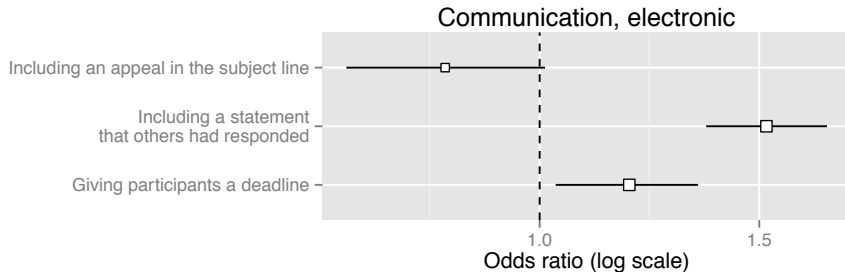
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygg bortfall



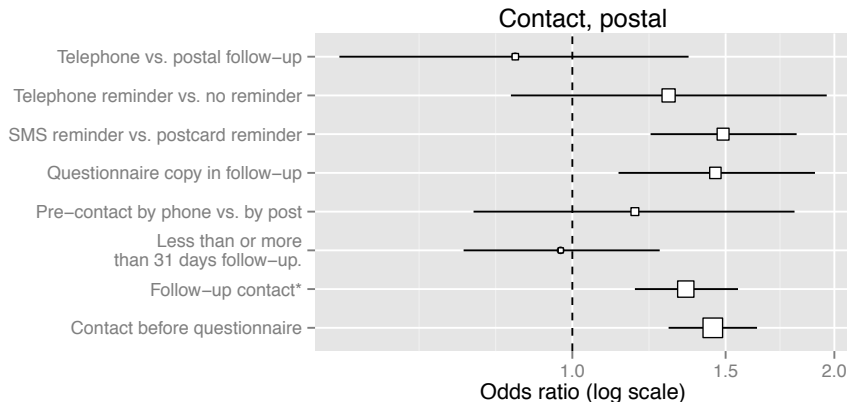
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



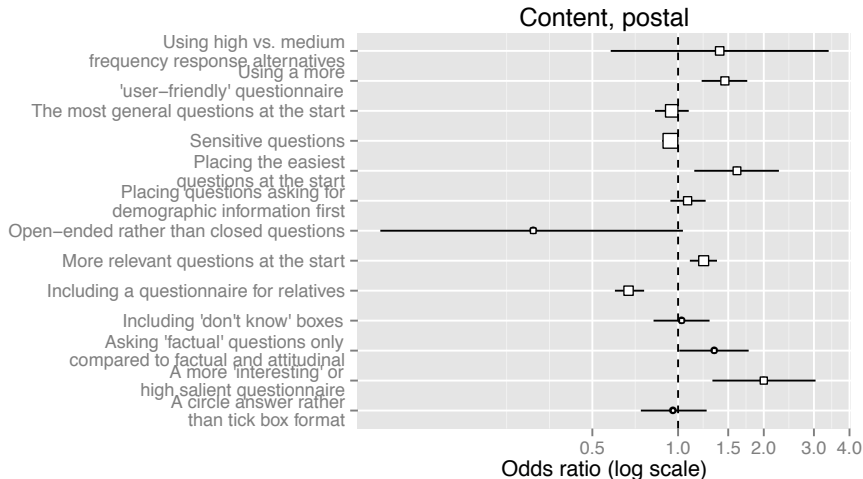
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



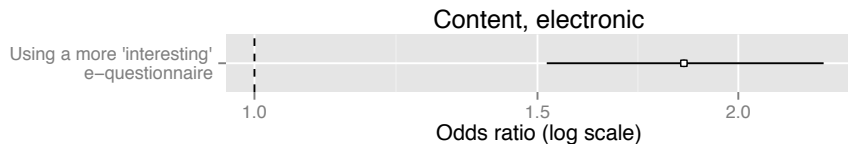
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



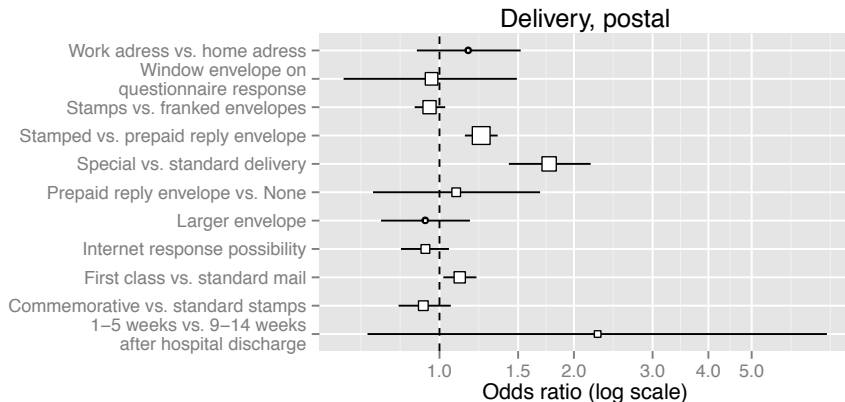
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



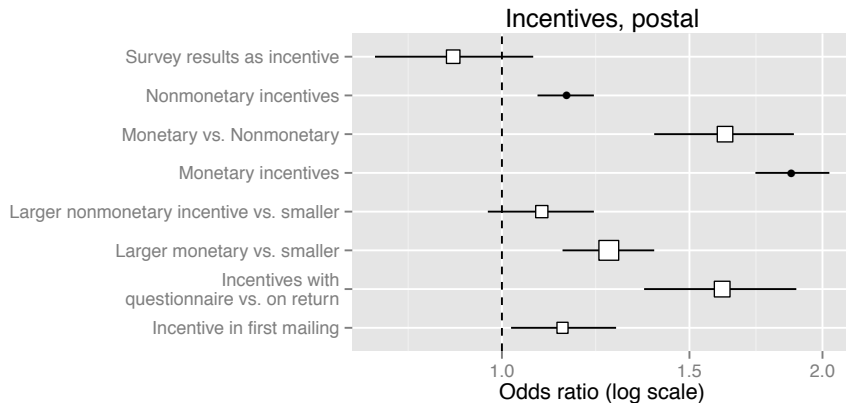
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



Figur : Metastudie om bortfall av Edwards et al. (2009)

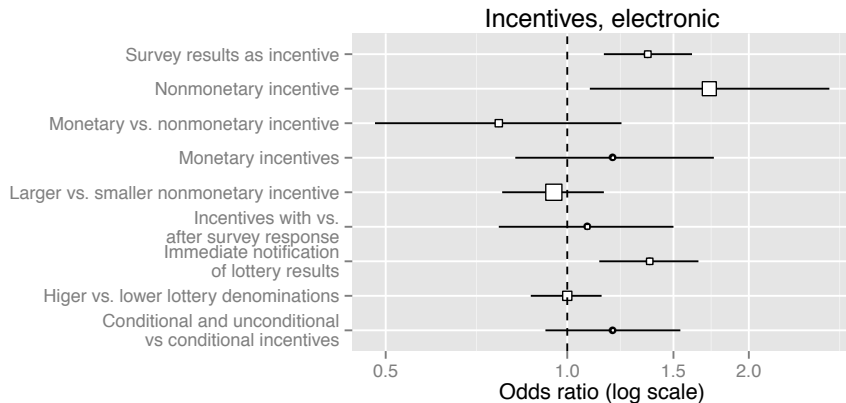
# Steg 1: Förebygg bortfall



Figur : Metastudie om bortfall av Edwards et al. (2009)

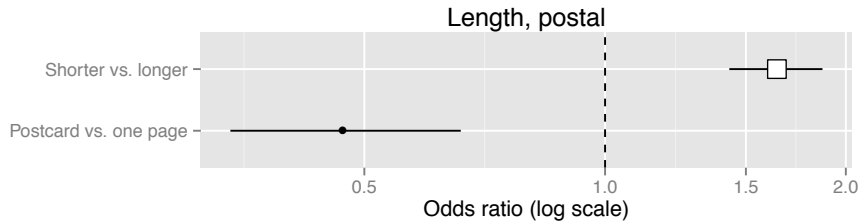


# Steg 1: Förebygga bortfall



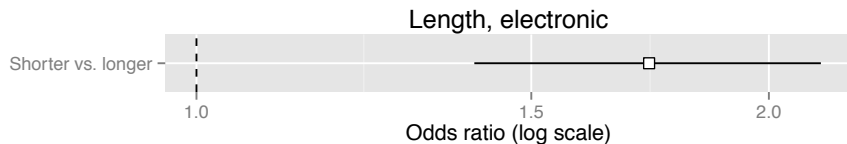
Figur : Metastudie om bortfall av Edwards et al. (2009)

## Steg 1: Förebygga bortfall



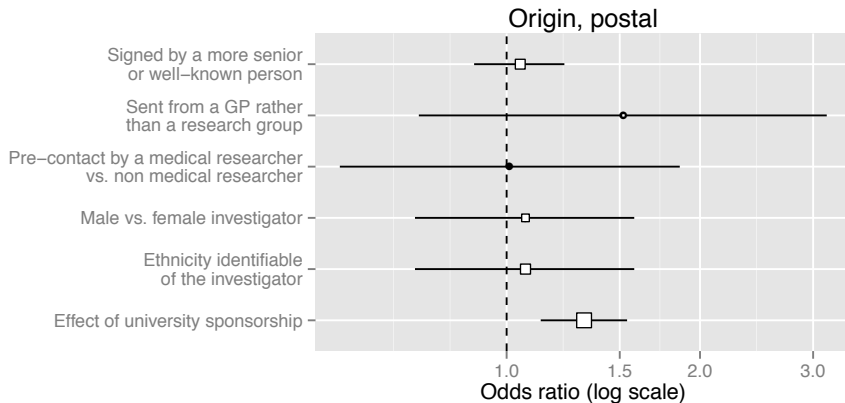
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



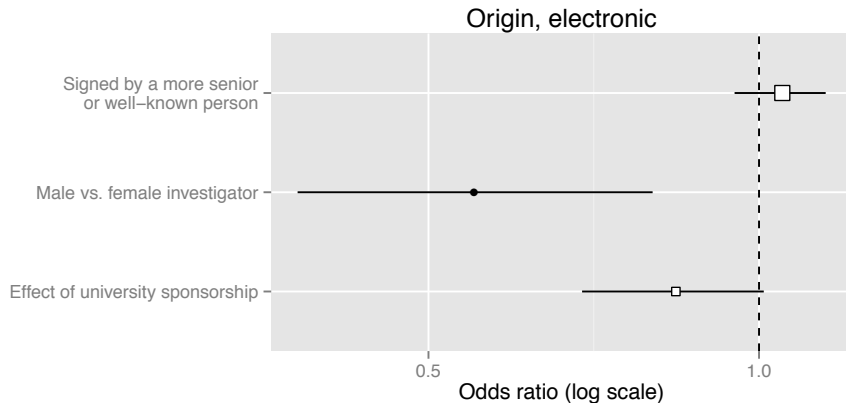
Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



Figur : Metastudie om bortfall av Edwards et al. (2009)

# Steg 1: Förebygga bortfall



Figur : Metastudie om bortfall av Edwards et al. (2009)

- För att förebygga problemet med att urvalet minskar på grund av bortfallet kan “översampla”
  - Ett större urval än det tänkta dras för att kompensera för det beräknade bortfallet.
- Det mest elementära sättet att göra detta är att bara räkna upp

$$n_* = \frac{n}{p_R}$$

där  $n_*$  är det urval man **drar**,  $n$  är det urval man **vill ha** och  $p_R$  är den **svarsandel** man uppskattar att man har.

- Identifiera vilka variabler som kan tänkas samvariera med bortfallet
  - Detta kräver att vi har hjälpvariabler.
  - Resonera, gärna tillsammans med ämnesexperter om detta
- Två situationer:
  - Vi har data för både svarande och icke-svarande: logistisk regression
  - Vi har bara data för svarande: tester mot populationsvärden
- För att reducera bortfallsfelet: identifiera variabler som både är korrelerade med  $y$  **och** med bortfalls sannolikheten

## Steg 3: Hantera uppkommet bortfall

- Hur vi hanterar bortfallet beror på vilket antagande vi kan göra om bortfallet
- Om vi antar **MCAR**
  - Går att bara lägga till nya respondenter (substitution)
  - Det är ofta bättre att göra detta innan studien drar igång
- Om vi antar **MAR** (den vanligaste situationen)
  - Omvägning/Kalibrering
  - (Multipel) imputering
  - Kräver en bortfallsmodell
- Om vi antar **NMAR**
  - Bortfallsuppföljning (Hansen-Hurwitz)
- GREG-estimatorn (regressionsestimatorn) så kalibrerar vi för bortfall och får en förbättrad precision samtidigt
  - Denna metod är standard idag i större undersökningar.



- En vanlig praktisk lösning är:
  - Partiellt bortfall - (Multipel imputering)
  - Objektbortfall - (Kalibrering)
- Olika typer av imputation
  - Medelvärdesimputering
  - Regressionsimputering (samma modell som vid kalibreringen)
  - Hot/Cold-deckimputering
- Problem med imputation - vi minskar vår osäkerhet
  - Lösningen: Multipel imputation

## Subsection 6

Redovisa bortfall

- För att beräkna bortfall korrekt (för jämförbarhet) används Svensk standard för bortfallsberäkning (se Surveysektionen, 2005) som kan laddas ned [\[här\]](#)
- Bortfall kan både beräknas oviktat och viktat.
  - Vid översampling för bortfall blir det viktade bortfallsmåttet ofta lägre än det oviktade
- Bortfallet bör redovisas efter kön, ålder m.m. samt efter
  - Ej anträffade
  - Vägrare
  - Övrigt (t.ex. språksvårigheter, sjukdom)

- För att beräkna oviktat bortfall används följande formel

$$BA_1(\text{oviktat}) = 1 - \frac{n_S}{n_S + n_B + n_O}$$

där  $S$  är svarande,  $B$  är bortfall och  $O$  är okänd målpopulation och  $\emptyset$  är övertäckning.

- För att beräkna viktat bortfall används

$$BA_1(\text{viktat}) = 1 - \frac{\sum_S d_k}{\sum_S d_k + \sum_B d_k + \sum_O d_k}$$

där  $d_k$  är designvikterna

- En undersökning har genomförts för att studera väljaropinion i Linköping. Vi behöver ett urval på  $n = 800$ .
- För att kompensera för bortfallet (och för att ha som redovisningsgrupper) har undersökningen stratifierats i två strata:
  - Yngre personer (15-29 år) och
  - Äldre personer (30 - 89 år)
- År 2012 bodde 35 664 yngre personer och 81 681 äldre personer i Linköping.
- Tidigare undersökningar har visat att personer under 30 år har ett bortfall på cirka 50 % medan den äldre gruppen har ett bortfall på cirka 20 %.
- Beräkna urvalet med oversampling i det två strata.

- Undersökningen genomförs med följande resultat

Resultat	Yngre medborgare	Äldre medborgare	Totalt
Svar (S)	241	509	740
Bortfall (B)	27	28	55
Okänd status (O)	159	52	211
Övertäckning ( $\emptyset$ )	59	8	67

- Beräkna studiens oviktade och viktade bortfall.

- Dahmström, K., 2011. Från datainsamling till rapport: att göra en statistisk undersökning, 5th Edition. Studentlitteratur, Lund.
- Edwards, P., Roberts, I., Clarke, M., Diguiseppi, C., Wentz, R., Kwan, I., Cooper, R., Felix, L., Prata, S., 2009. Methods to increase response to postal and electronic questionnaires (review). Cochrane Database of Systematic Reviews 3, 1–12.
- Groves, R., Singer, E., Corning, A., 2000. Leverage-saliency theory of survey participation: Description and an illustration. Public Opinion Quarterly, 299–308.
- Japac, L., Ahtainen, 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](http://www.pubkat.scb.se/statistik/_publikationer/0V9999_2000I02_BR_X97%C3%96P9701.pdf)
- SCB, 2012. Arbetskraftsundersökningarna 2011. Statistiska meddelanden AM12 SM1201, Statistiska centralbyrån, Stockholm.
- Surveysektionen, 2005. Standard för bortfallsberäkning. Stockholm.  
URL