

F9

Olika feltyper i en undersökning  
Kvalitetsredovisning

# Olika feltyper

- Feltyper kan klassificeras utifrån deras orsaker: felkällorna
  - Urvalsfel
  - Icke-urvalsfel
    - Täckningsfel
    - Bortfall
    - Mätfel
    - Bearbetningsfel

# Olika feltyper i en undersökning

- Urvalsfel
  - De fel som uppstår när vi drar slutsatser till populationen från ett (enda) urval i stället för att göra en totalundersökning.
    - *Den avvikelse vi får mellan värdet på skattningen i urvalet och värdet vi skulle ha fått i en totalundersökning*
  - Resultaten varierar mellan de olika (tänkta) urvalen. Om vi gör ett *slumpmässigt* urval, kan detta urvalsfel mätas.
  - Mått på urvalsfelet:
    - Felmarginalen
    - Konfidensintervallets längd

# Mått på urvalsfel

- Antag att vi har dragit ett slumpmässigt urval av  $n$  observationer och vill skatta populationsmedelvärdet  $\mu$
- Om vi drar ett tillräckligt stort stickprov gäller CGS  
 $\bar{X} \in \text{approx}N(\mu, V(\bar{X}))$
- Konfidensintervall  
 $\bar{x} \pm c\sqrt{V(\bar{X})}$   
där  $c$  beror av den valda konfidensgraden

# Mått på urvalsfel

- Exempel

$$\bar{x} = 5$$

$$s^2 = 1$$

$$n = 200$$

$$c = 1.96$$

$$5 \pm 1.96 \sqrt{\frac{1}{200}} \approx 5 \pm 0.14$$

- Felmarginal: 0.14
- Konfidensintervallets längd: 0.28
- Ju kortare intervall desto mindre urvalsfel
- Kan vi påverka urvalsfelet?

# Täckningsfel

- Fel som orsakas av att enheterna i målpopulationen och rampopulationen ej stämmer överrens
- Undertäckning:
  - Utgörs av de enheter som ingår i målpopulationen men ej i rampopulationen
  - Det allvarligaste täckningsfelet – kan leda till systematiska fel
- Övertäckning:
  - Utgörs av de enheter som ingår i rampopulationen men ej i målpopulationen
  - Ofta ej så allvarligt – kan dock minska stickprovsstorleken

# Bortfallsfel

- Fel som orsakas av att vi ej får användbara observationer på en eller flera variabler för enheter i ramen som tillhör målpopulationen och som blivit utvalda.
  - Det icke-urvalsfel som är lättast att mäta
  - Individbortfall
  - Partiellt bortfall
    - variabelbortfall

# Orsaker till och konsekvenser av bortfall

- Orsaker
  - Ej anträffbara
  - Vägrare
  - Övrigt
    - språksvårigheter
    - ålder, sjukdom
    - tekniska skäl
- Konsekvenser
  - Systematiskt fel – egenskaperna hos de svarande kan skilja sig åt från bortfallet
  - Hur stort kan det systematiska felet bli?

# Kan vi skatta konsekvenserna av bortfall?

- $W_S =$  andel svarande
- $W_B = 1 - W_S =$  andel bortfall
- $\bar{x}_S =$  observerat medelvärde bland svarande
- $\bar{x}_B =$  okänt medelvärde bland bortfall
- Skattning av medelvärdet i populationen:

$$\bar{x} = W_S \bar{x}_S + W_B \bar{x}_B$$

- Mått på snedvridning orsakad av bortfallet:

$$\begin{aligned} \bar{x}_S - \bar{x} &= \bar{x}_S - (W_S \bar{x}_S + W_B \bar{x}_B) = \bar{x}_S (1 - W_S) - W_B \bar{x}_B = \\ &= \bar{x}_S W_B - W_B \bar{x}_B = W_B (\bar{x}_S - \bar{x}_B) \end{aligned}$$

# Beräkning av den ”största” snedvridningen vid beräkning av proportion

- Sökt:  $P$  = andel i populationen
- $p_S$  = andel bland de svarande
- $p_B$  = andel bland bortfallet
- $p = W_S p_S + W_B p_B$  och kan variera mellan  $W_S p_S$  och  $W_S p_S + W_B$

# Exempel

- Vi vill undersöka hur stor genomsnittslön (i tusental) samt andelen rökare är i en population
- Vi har dragit ett urval av 1000 personer och fått svar från  $n_S = 600$ . Vi har erhållit följande:

$$\bar{x}_S = 25$$

$$p_S = 0.3$$

$$W_S = 0.6$$

$$W_B = 0.4$$

# Uppskattning av snedvridning

$$\bar{x}_S - \bar{x} = W_B (\bar{x}_S - \bar{x}_B) \quad \text{Om medelvärdet i bortfallet}$$
$$25 - \bar{x} = 0.4(25 - \bar{x}_B) \quad \text{exempelvis är 23 får vi:}$$

$$25 - \bar{x} = 0.4(25 - 23) = 0.8$$

Dvs vi överskattar medelvärdet med 800 kr jämfört med vad vi skulle ha fått om vi fått svar från alla.

Om vi antar att *ingen* i bortfallet är rökare:

$$p = W_S p_S = 0.6 \times 0.3 = 0.18$$

Om vi antar att *alla* i bortfallet är rökare:

$$p = W_S p_S + W_B = 0.6 \times 0.3 + 0.4 = 0.58$$

# Att få en uppfattning om möjlig snedvridning

- Förklaring till bortfall? Tror man att de avviker eller ej?
- Titta på olika svarsomgångar vid postenkäter exempelvis
- Redovisa bortfall separat för olika grupper
- Strukturjämförelser
  - Men varning!

# Förebyggande åtgärder mot bortfall

- Utbilda och informera
- Skapa en positiv inställning till statistik
- Testa frågeformuläret, gör provundersökningar
- Informationen i introduktionsbrevet – skyddet för den personliga integriteten
- Skicka påminnelsebrev/förnyade kontakter
- Byt kontaktform/insamlingsmetod
- Lova presenter
- Minska uppgiftslämnarbördan
- Använd anonymitetsskyddande åtgärder
  - CASI
  - Valurnemetoden

# Partiellt bortfall

- Variabelbortfall
  - Kan också leda till snedvridning
  - Statistiska beräkningar måste baseras på färre personer
  - Imputering

# Efterstratifiering

- Om vi har fått bortfall kan vi räkna med det i analyser genom att använda vikter, jmr medelvärdeskattning vid stratifierat urval.
- Säg att vi vet åldersfördelningen bland de som svarat respektive de som inte svarat.

Ålder	Antal svarande	Andel rökare	Medellön	Andel i populationen
18-34 år	100 av 300	0.60	20	0.30
35-64 år	500 av 700	0.24	26	0.70
18-64 år	600 av 1000	0.30	25	

# Efterstratifiering

- De ”vanliga” ojusterade värdena fick vi som

$$\bar{x}_s = \frac{100}{600} \times 20 + \frac{500}{600} \times 26 = 25$$

$$p_s = \frac{100}{600} \times 0.60 + \frac{500}{600} \times 0.24 = 0.3$$

- Efterstratifierar vi får vi istället

$$\bar{x}_{ps} = 0.3 \times 20 + 0.7 \times 26 = 24.2$$

$$p_{ps} = 0.3 \times 0.60 + 0.7 \times 0.24 = 0.348$$

Ålder	Antal svarande	Andel rökare	Medellön	Andel i populationen
18-34 år	100 av 300	0.60	20	0.30
35-64 år	500 av 700	0.24	26	0.70
18-64 år	600 av 1000	0.30	25	

# Standard för bortfallsberäkning/redovisning

- För att kunna jämföra kvaliteten i olika undersökningar behövs enhetliga mått vid bestämning av bortfallets storlek
- Elementen i urvalet kan klassificeras på följande sätt:
  - Svarande som tillhör målpopulationen,  $n_S$
  - Bortfall som tillhör målpopulationen,  $n_B$
  - Övertäckning (tillhör inte målpopulationen),  $n_{\bar{O}}$
  - Element som man ej vet om de är bortfall och alltså tillhör målpop eller utgör övertäckning (= enheter med okänd status),  $n_O$

# Bortfallsberäkning forts.

- Bortfallsandelen (BA) kan i enklaste fallet fås som

$$BA = 1 - \frac{n_S}{n_S + n_B + n_O}$$

- Man bör också redovisa
  - Antal svarande med användbara observationer
  - Bortfall i underkategorier (vägran, ej anträffad, övrigt)
  - Bortfall i intressanta redovisningsgrupper
  - Storlek på systematiska fel
  - Åtgärder för att minimera fel

# Mätfel

- Skillnaden mellan erhållet (uppgivet) och sant värde
- Orsaker:
  - Mätinstrumentet
    - Dåligt frågeformulär: ledande frågor, för omfattande, oklarheter
  - Mätmetoden
  - Intervjuaren
    - prestigebias, intervjuaren följer ej instruktioner, förhåller sig ej neutral till frågorna
  - Respondenten
    - svarar avsiktligt fel, vet ej svaret, gissar, förtränger vissa händelser, minnesfel, speciellt teleskopeffekten: man kan ej ange när en viss händelse har inträffat, förlägger den t ex senare i tiden.

# Åtgärder mot mätfel

- Testa frågeformuläret
- Utbilda och följ upp intervjuarnas arbete
- Använd anonymitetsskyddande åtgärder för känsliga frågor
- ”Minnesstärkande” åtgärder
  - be att almanackor o dyl tas fram
  - använd checklistor

# Bearbetningsfel

- Fel som uppstår vid:
  - Kodning – yrkeskodning svårt
  - Registreringen – lättare med automatisk dataregistrering med inbyggda kontroller
  - Datorbearbetning
    - Fel vid analysen – förutsättningar ej uppfyllda

# Totala felet

- Summan av urvalsfel och icke-urvalsfel
- Totala felet =  
urvalsfel + täckningsfel + bortfallsfel + mätfel +  
bearbetningsfel
- Kan vara mindre vid urvalsundersökningar än vid  
totalundersökningar tack vare större möjligheter  
att minska icke-urvalsfelen i det ”mindre  
formatet”.
- Viktigt med kvalitetsredovisning

# Exempel på olika fel i riktiga undersökningar

- Urvalsfel:
- Täckningsfel:
- Bortfallsfel:
- Mätfel:
- Bearbetningsfel:

# Kvalitetsredovisning

- Till en resultatredovisning bör höra en kvalitetsdeklaration
- Syftet:
  - Göra användarna uppmärksamma på fel som gör statistiken mindre användbar
  - Ge ett bättre beslutsunderlag
  - Användaren ska lättare kunna precisera kvalitetskrav inför nya undersökningar
- Vad ska kvalitetsdeklareras?
  - Innehåll
  - Tillförlitlighet
  - Aktualitet
  - Jämförbarhet och sammanvändbarhet
  - Tillgänglighet och förståelighet