

F6

Ytterligare urvalsmetoder och  
skattningsmetoder

# Flerstegsurval

- Anta att man vill göra ett urval som täcker ett stort geografiskt område vill använda besöksintervju som insamlingsmetod
  - Praktiskt omöjligt med OSU
- Tvåstegs gruppurval
  - Slumpmässigt urval av grupper (kluster)
    - Primära urvalsenheter (Pue)
    - Kan dras med exempelvis stratifierat urval/OSU
    - PPS = Probability proportional to size
  - Slumpmässigt urval av element inom grupper
    - Kan anpassa det så att alla element i populationen får samma inklusionssannolikhet
      - Vanliga formler för medelvärde och totalvärde kan användas
- Basurval
  - vanligt att "återanvända" samma Puer (samt intervjuare)

# Exempel PPS-urval av Pue (från KD s. 287)

- Inklusionssannolikheten ska vara proportionell mot något storleksmått
- Dra ett PPS-urval, från stratumet Kronobergs län, av 3 kommuner där sannolikheten är proportionell mot folkmängden i kommunen

Kommun	Folkmängd	Kumulativ folkmängd	Slumptal
Alvesta	19035	19035	000000-019035
Lessebo	8433	27468	019036-027468
Ljungby	27140	54608	027469-054608
Markaryd	9927	64535	054609-064535
Tingsryd	13484	78019	064536-078019
Uppvidinge	9912	87931	078020-087931
Växjö	73770	161701	087932-161701
Älmhult	15448	177149	161702-177149

# Panelundersökningar

- Syftet med en undersökning kan vara att bestämma förändringen mellan två tidpunkter av en parameter
- Oberoende tvärsnittsdata
  - Skattningar från två helt separata urval
- Paneldata
  - Skattningar från samma urval/delvis samma urval
  - Panelen består av de element som är desamma vid de två tidpunkterna
- Om en panelundersökning sträcker sig under en lång tid brukar den kallas longitudinell undersökning
  - Kohortundersökning: personer födda under en viss tid följs under en period
- Nackdel med panelundersökningar: bortfallet ökar vartefter tiden går (drop-outs)

# Panelundersökning

- Känd undersökning: AKU
  - Totala urvalet varje månad består av 8 paneler
  - Varje panel utfrågas med 3 månaders mellanrum under 2 år
  - Varje månad byts  $1/8$  av panelen ut mot nya personer
  - Roterande panelundersökning

# Panelundersökning

- Man kan få bättre skattningar av förändringar om man använder paneldata jämfört med tvärsnittsdata
- Varför?

# Urval från sällsynta populationer

- Anta att vi vill undersöka en population – sällsynt population - som utgör en liten del av en större population
- Ingen ram
- Screening
  - Vi kan dra ett stort urval av element från den större populationen med en billig metod och sedan avgöra om elementen tillhör den sällsynta populationen
    - Ex. telefonintervju
  - Urvalet kan ses som ett slumpmässigt urval från den sällsynta populationen

# Urval från sällsynta populationer

- Problem:
  - Kan behöva göra ett mycket stort urval
    - Man kan utnyttja hjälpinformation
  - Risk för felklassificering

# Efterstratifiering

- Vid stratifierat urval behöver vi en ram sorterad efter stratifieringsvariabeln
- Anta att vi tror att det finns ett samband mellan undersökningsvariabeln och en viss variabel som vi skulle vilja stratifiera efter för att få precisionsvinst
- Vi har inte tillgång till denna variabel förrän *efter* datainsamlingen
- Vi vet fördelningen för variabeln i populationen
- Vi kan efterstratifiera för att få högre precision på skattningarna

# Efterstratifiering

- Dra ett OSU av element från populationen
  - Klassificera varje element till ett visst stratum
  - En efterstratifierad skattning av populationsmedelvärdet fås som
- 
- Skattningen är approximativt VVR
  - Variansen för skattningen är approximativt densamma som vid PSU-urval
    - Förutsättning: Det ska finnas minst två observationer i varje stratum

# Exempel (från KD s 292)

- OSU av 230 småhus
- Skatta medelkostnaden i populationen

Husets ålder	Antal hus i området	Antal hus i urvalet	Medelkostnad kr
< 5 år	500	28	3500
5-10 år	1500	84	6700
> 10 år	2500	128	9200

# Uppgift

- Vi drog tidigare ett OSU på  $n = 4$  från följande population. Vi vet att i populationen finns det 7 kvinnor och 13 män.
- Dra ett nytt OSU på  $n = 8$ . Skatta medellängden i populationen med hjälp av efterstratifiering. Använd kön som stratifieringsvariabel

Ram	Längd	Slumptal
1. Karin	162	
2. Håkan	186	
3. Daniel	177	18236 87267
4. Pär	186	96895 84803
5. Hans	171	83031 27087
6. Gösta	169	56857 87431
7. Raul	170	93491 55950
8. Johan	188	72705 61329
9. Anita	168	29239 95179
10. Linda	169	00687 21151
11. Jessica	163	26135 75180
12. Gebre	175	11112 80844
13. Ellinor	159	43502 87558
14. Nicklas	175	51575 43361
15. Lars	168	58093 23652
16. Mikael	167	67709 64440
17. Peter	173	
18. Mattias	179	
19. Birgitta		157
20. Elisabet	158	

# Kvotskattningar

- Vi vill använda en hjälpvariabel vid skattning av undersökningsvariabeln
  - Vi känner till medelvärdet eller totalvärdet för hjälpvariabeln i populationen
  - Det finns en positiv korrelation mellan undersökningsvariabeln och hjälpvariabeln
  - Vi kan bilda en kvotskattning för att få en säkrare skattning av medelvärdet eller totalvärdet i populationen (högre precision)
  - Kvotskattningen är inte VVR, men om  $n > 20$  är felet försumbart





# Exempel (från KD s 294)

- Vi vill skatta den totala omsättningen i en population av  $N=100$  företag. Vi tror att det råder en positiv korrelation mellan  $Z$ =antalet anställda och  $X$ =omsättning. Det finns 805 anställda totalt. Genom OSU fås  $n=20$  företag som visar sig ha en medelomsättning på 2.5 milj kr. Antalet anställda var i genomsnitt 8.5.
- OSU-skattning:
- Kvot-skattning