

F3

Urvalsmetoder: Sannolikhetsurval resp.
icke-sannolikhetsurval,
OSU
(kap 9.1-9.4)

Urval

- Anta att vi ska göra en urvalsundersökning och samla in primärdata
 - Totalundersökning ofta inte möjlig
 - För dyrt
 - Tar för lång tid
 - Urvalsundersökning innehåller ofta mindre totalt fel
- Vi är intresserade av att skatta en parameter i populationen med hjälp av en skattning från urvalet
- Hur ska urvalet dras?

Slumpmässiga urval (sannolikhetsurval)

- Varje element i populationen skall ha en *känd* sannolikhet (chans) som är *större än 0* att bli utvald.
- Urvalsenheten kan vara det enskilda elementet eller *grupper* av element
- Vi ska titta på:
 - Obundet slumpmässigt urval (OSU)
 - Stratifierat urval
 - Systematiskt urval
 - Enstegs gruppurval (klusterurval)

Fördelar med slumpmässiga urval

- Vi kan använda inferensteori för att dra slutsatser från stickprovet (urvalet) till populationen
- Vi kan beräkna VVR skattningar av okända parametrar
- Vi kan mäta osäkerheten i skattningarna

Icke-slumpmässiga urval:

- Några element har 0 sannolikhet att väljas och/eller några element har en okänd sannolikhet att väljas
- Kan finnas tillfällen då de kan vara motiverade, men man får inte låtsas att urvalet är draget slumpmässigt och dra slutsatser till populationen som om så vore fallet.
 - Kvoturval ("quota sampling")
 - Intervjuaren väljer respondenter; val utifrån egna preferenser. Vissa "kvoter" skall fyllas så att urvalet med avseende på vissa bakgrundsvariabler blir som i populationen.

Icke-slumpmässiga urval forts.

- Användning av telefonkatalogen som ram
- Frivilliga ("voluntary") svarande/försökspersoner ("själval")
 - annonsering vid läkemedelsstudier
 - svar på webbenkäter
- Bekvämlighetsurval ("convenience sampling")
 - kvalitetskontroll då man tar de översta i ett parti varor.
 - val av vänner och släktingar som får svara
 - svar från hotellgäster/resenärer på en enkät som ligger på rummet

Icke-slumpmässiga urval forts.

- Typiska element ("purposive sampling", "judgement sampling")
 - Kan vara element valda subjektivt så att (man tror) att alla kategorier blir representerade för att t ex snabbt pejla opinionen; typiska element. Även vid provundersökningar.
- Snöbollsurval – då ingen ram finns över personerna.
 - Utgå från en känd grupp av personer med en speciell egenskap. Fråga om vilka de i sin tur känner som har denna egenskap osv; man fångar in allt större del av populationen.

Icke-slumpmässiga urval forts.

- Risker med icke-slumpmässiga urval
 - Systematiska fel i slutsatser
 - För bestämning av bortfallets storlek krävs en ram; här blir bortfallet dolt och okänt.

Repetition

- Vi vill välja ett "så bra" urvalsförfarande som möjligt
- Egenskaper hos en punktskattning
- Parameter:
- Skattning (t.ex. punktskattning):
 - Slumpvariabel (stokastisk variabel)
- Avvikelse:
- VVR skattning:

Repetition forts.

- Varians för skattningen:
- Standardavvikelsen för skattningen kan kallas medelfelet (standard error)

Repetition forts.

- Om en skattning har en liten standardavvikelse/liten osäkerhet har den hög *precision*
- Mått på osäkerheten i skattningen:

- Centrala gränsvärdessatsen

Repetition forts.

- Samplingfördelning

Urval

- Vi vill alltså göra en så "bra" urvalsundersökning som möjligt.
- Vi kan påverka:
 - Hur ramen ska konstrueras
 - Hur urvalet ska dras
 - Hur stort urvalet ska vara
 - Hur punktskattningen av den okända parameteren ska bestämmas
- Vi vill välja en urvalsdesign som är så "bra" som möjligt till en given kostnad
 - Bra: hög precision

OSU

- Anta att vi ska dra n element från en population med N st.
- OSU: Obundet Slumpmässigt Urval
 - Varje element har samma sannolikhet att bli valt och denna inklusionssannolikhet är n/N
 - Varje tänkbart stickprov om n element har samma sannolikhet
- Vi "styr" inte urvalet. Det är slumpen som styr vilka element som väljs
- Med återläggning
 - Dragningarna blir oberoende av varandra
- Utan återläggning

Skattning av medelvärdet i populationen

Parameter: μ

- Punktskattning:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- Skattade variansen för skattningen:

- OSU med återläggning

$$\hat{V}(\bar{X}) = \left(1 - \frac{1}{N}\right) \frac{s^2}{n}$$

- OSU utan återläggning

$$\hat{V}(\bar{X}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n}$$

Konfidensintervall för μ

- OSU utan återläggning:
- Om $n > 30$ fås, enligt CGS, ett 95%igt k.i. som:

Skattning av totalvärdet i populationen

- Parameter:
- Punktskattning:
- Skattade variansen för skattningen:
 - OSU med återläggning
 - OSU utan återläggning

Skattning av andel (proportion) i populationen

- Parameter:
- Punktskattning:
- Skattade variansen för skattningen:
 - OSU med återläggning
 - OSU utan återläggning

Exempel

- $X = 1$ om månadslön $> 30\ 000$, 0 annars
- $N = 500$, $n = 35$, $p = 0.4$
- Ett 95%igt k.i. för andelen (proportionen) P blir
 - Med återläggning
 - Utan återläggning

Skattning av totala antalet ettor i populationen

- Parameter:
- Punktskattning:
- Skattade variansen för skattningen:
 - OSU med återläggning
 - OSU utan återläggning

Bestämning av stickprovsstorlek n

- Val av stickprovsstorlek n beror på vilken precision på skattning vi vill ha.
 - Vi kan tex säga att standardavvikelsen för skattningen ej får överskrida ett visst tal D .
 - Exempel: Skattning av medelvärde:
-
- Problem?

Beräkning av stickprovsstorlek n

- Vid skattning av medelvärde μ , OSU utan återläggning, där vi vill att standardavvikelsen för skattningen ska vara högst D
- Vid skattning av proportion P , OSU utan återläggning, där vi vill att standardavvikelsen för skattningen ska vara högst D

Exempel

- Anta att vi vill skatta μ , medellön på ett viss företag, med $N = 500$ anställda, och tror att variansen på företaget ligger runt $\sigma^2 = 25$. Vi kommer dra ett OSU utan återläggning och vill att felmarginalen ska vara högst 2

OSU i praktiken

- Ram
- Identifiering av element
- Slumpmekanism
 - Slumptalstabell
 - Datoriserad slumptalsgenerator

Exempel: vi är intresserade av att skatta medellönen i en population

Ram	Slumptal
1. Fredrik	26135 75180 11112 80844
2. Johan	43502 87558 51575 43361
3. Ann-Britt	58093 23652 67709 64440
4. Gunnar	18236 87267 96895 84803
5. Mona	83031 27087 56857 87431
6. Peter	93491 55950 72705 61329
7. Daniel
8. Per	29239 95179 00687 21151
9. Katarina	
10. Carl	
...	
70. Anders	

Säg att vi vill dra ett OSU utan återläggning av $n = 30$ personer från $N = 70$ personer*. Vi väljer 2-siffriga slumptal:

*Ofta har vi en mycket större population än $N = 70$ och vi tar ofta ett större stickprov än $n = 30$.