

F4

Urvalsmetoder: Stratifierat urval  
(kap 9.5)

# Tidigare exempel

- Vi undersökte tidigare medellönen i ett företag med  $N = 500$  anställda. Vi fick ett konfidensintervall:
- Vi vet att några "förklaringsvariabler" till lön är kön, ålder, anställningsform, antal år i anställning osv.
- OSU kan pga slumpen ge ett urval med "för många" män exempelvis – medellönen kan då överskattas

# Stratifierat urval

- Populationen delas in i  $L$  st. strata men  $N_i$  element,  $i = 1, \dots, L$
- Från varje stratum drar man ett slumpmässigt urval av  $n_i$  element
- Alla strata blir representerade i stickprovet
- Inklusionssannolikhet:  $n_i/N_i$

# Några beteckningar

# Skattning av medelvärde i populationen

- Stratifierat urval med OSU i varje stratum:
- Parameter:
- Punktskattning:
- Skattade variansen för skattningen:

# Skattning av totalvärdet i populationen

- Stratifierat urval med OSU i varje stratum:
- Parameter:
- Punktskattning:
- Skattade variansen för skattningen:

# Exemplet forts.

- Vi vet att kvinnor och män har olika lön
  - Det finns ett samband mellan kön och lön
- Dela in populationen i två grupper, kvinnor och män, och dra ett OSU av personer från varje grupp (=från varje **stratum**)
  - Kön är då en **stratifieringsvariabel**

# Exempel

| Stratum | $N_i$   | $n_i$  |
|---------|---------|--------|
| Kvinnor | 200     | ?      |
| Män     | 300     | ?      |
| Totalt  | $N=500$ | $n=35$ |

- Vi har en population med 200 kvinnor och 300 män och vill dra ett urval av  $n = 35$  personer
- Hur ska vi fördela  $n=35$  mellan grupperna, dvs hur många kvinnor ska väljas och hur många män? Hur ska urvalet *allokeras*?
- Vi kan exempelvis göra ett *proportionellt stratifierat urval* (PSU)

# Proportionellt stratifierat urval (PSU)

- Fördela urvalet i samma proportioner som strata förhåller sig till hela populationen

$$n_1 = 35 \times \frac{200}{500} = 35 \times 0,4 = 14$$

$$n_2 = 35 \times \frac{300}{500} = 35 \times 0,6 = 21$$

- Vi kan välja 14 kvinnor och 21 män genom OSU från varje stratum
- Generellt väljer vi  $n_i$  element från stratum i enligt följande:

$$n_i = n \times \frac{N_i}{N}$$

# Skatta medellön, $\mu$

| Stratum | $N_i$ | $n_i$ | $\bar{x}_i$ | $s_i$ |
|---------|-------|-------|-------------|-------|
| Kvinnor | 200   | 14    | 23          | 4     |
| Män     | 300   | 21    | 26          | 3.5   |

- Väg ihop skattningarna från respektive stratum

$$\bar{x}_{str} = \frac{200 \times 23 + 300 \times 26}{500} = 24.8$$

- Variansen för skattningen blir:

# Skattning av andel (proportion) i populationen

- Stratifierat urval med OSU utan återläggning från varje stratum
- Parameter:
- Punktskattning:
- Skattade variansen för skattningen:

# Stratifierat urval

- Vi kan få högre precision (lägre varians) i skattningen vid stratifierat urval jämfört med OSU-urval
- När får vi det?
  - När vi delar in populationen i strata (grupper) som är homogena med avseende på undersökningsvariabeln
  - Om grupperna är lika (homogena) inom sig, så får vi lägre varians inom varje grupp (stratum)
  - Målet när vi delar in populationen i strata är att de ska bli så lika inom strata som möjligt och så olika mellan strata (heterogena) som möjligt med avseende på *undersökningsvariabeln*

# Stratifierat urval

- Att välja antal strata och stratumgränser
  - Hur ska vi välja antal strata samt stratumgränser så strata blir så homogena som möjligt?
  - Ju fler strata desto mindre precisionsvinst
  - Stratumgränser?
    - Om undersökningsvariabeln är lön och vi "vet" att det finns relativt tydliga "grupperingar" med höga, mellan och låga löner kan vi försöka bilda strata utifrån dessa grupper.  
Problem?

# Problem?

- Vi vill ha homogena strata med avseende på undersökningsvariabeln – ex lön – och vill dela in populationen så att vi får ett strata med låg lön och ett med hög lön alternativt ett med låg, ett med mellan och ett med hög lön.
  - Vi har inte tillgång till värden på undersökningsvariabeln lön!
  - Vi måste ta hjälp av någon *stratifieringsvariabel* (hjälpvariabel) som vi tror har ett SAMBAND med undersökningsvariabeln
  - Vi måste ha en *ram* sorterad efter stratifieringsvariabeln
  - Om vi tror att kön har ett samband med lön kan vi dela in populationen i två strata: män och kvinnor.
  - Om vi tror att ålder har ett samband med lön kan vi försöka hitta homogena strata med avseende på ålder (stratifieringsvariabeln) och om det finns ett samband mellan ålder och lön så hoppas vi att strata även blir homogena med avseende på lön (undersökningsvariabeln)

# Planering av ett stratifierat urval

- Vilken stratifieringsvariabel ska väljas?
- Hur många strata?
- Var ska stratumgränserna dras?
- Hur ska allokeringen göras?
  - Hur ska stickprovet fördelas?

# Allokering

- Vi drar ett OSU utan återläggning ur varje stratum – hur ska vi välja hur många element som ska dras från varje stratum?
- Proportionell allokering
  - Proportionellt stratifierat urval (PSU)
  - Välj proportionellt sett lika många element ur varje stratum som stratumet utgör av populationen
    - Ex om stratum 1 utgör 30% av populationen – välj 30% av stickprovsstorleken från stratum 1
  - Se till att alltid dra minst 2 element från varje stratum!
  - Om ett stratum är "litet" kan man istället för ett PSU välja att göra en totalundersökning i detta stratum

# Allokering

- Vi kan istället för PSU välja att dra lika många element från varje stratum
  - Exempelvis då vi är intresserade av att göra gruppjämförelser såsom att jämföra mäns och kvinnors medellöner
- Optimal allokering (Neyman allokering)
  - Variansen för skattningen minimeras
  - Dra fler element från stratum med större varians och färre från stratum med mindre varians
  - $n_i$  fås från följande: