

Kira Coder Gylling
PMU

DraUrval

1 Introduktion

I enlighet med målsättningen för den gemensamma produktionsmiljön, och i samband med avvecklandet av VB6-applikationerna, utvecklas en ny generell tjänst för urvalsdragning. Tjänsten är utvecklad i SAS i form av ett antal makron, i princip ett makro per urvalsmetod, som alla anropas genom ett antal wrappermakron.

De vanligaste urvalsmetoderna kan hanteras av urvalstjänsten; detaljer om vilka metoder som stöds och hur det görs beskrivs i avsnitt 2. I avsnitt 3 beskrivs lite kort urval från blandade designer och i avsnitt 4 till 6 nämns mer komplicerade designer. Vänligen kontakta Kira Gylling om du behöver hjälp med dessa.

I avsnitt 7 beskrivs mer ingående den huvudsakliga tillämpningen av urvalstjänsten, nämligen makrot FTGSAMU som är avsett för samordnade företagsurval inom ramen för det ekonomisk-statistiska systemet.

Slutligen, i avsnitt 8 beskrivs den nya modulen för skolorval som introducerades i version 1.1 av urvalstjänsten.

Urvalstjänsten kräver att användaren tillhandahåller ett SAS-dataset motsvarande urvalsramen, innehållandes information om bland annat stratumtillhörighet (vid stratifierat urval), önskad urvalsstorlek, storleksmått (vid π ps-urval) och i förekommande fall permanenta slumpetal.

För att komma åt betaversionen av urvalstjänsten körs följande två SAS-kommandon:

```
libname urv "T:\Gemensamma\Net  
apps\StatProd\DraUrval\x&sysaddrbits.";

options mstored sasmstore=urv;
```

2 En urvalsmetod

För fristående urval med en och samma urvalsmetod för varje stratum, eller för ej stratifierade urval, körs `EnkeltUrval`. Ett anrop ser ut som:

```
%EnkeltUrval(ram=, metod=, <metodspezifika  
parametrar>);
```

Input: Parametern `ram` anger den lokala ramen som urvalet ska dras från. Parametern `metod` anger vilken metod som ska användas och kan ta värdena `osu` (eller `srs`) för obundet slumpmässigt urval, `par` för Pareto π ps, `sekpo` för sekventiell Poisson, `parmix` för Pareto mixture, `pomix` för Poisson mixture, `sys` för systematiskt urval, `syspips` för systematiskt π ps, `ber` för Bernoulliurval samt `po` för Poissonurval.

Utöver `ram` och `metod` behövs metods specifika parametrar:

`stratum`: Stratifieringsvariabel, given som variabel på `ram`. Anges vid stratifierat urval.

`lillan`: Urvalsstorlek(ar), given som variabel på `ram` om stratifierat urval används.

`size`: Storleksmått, given som variabel på `ram`. Anges vid π ps-urval.

`lambda`: Urvalssannolikheter vid π ps-urval, om exempelvis totalundersökning av en delmängd av ramen önskas.

`B`: Blandningsgrad, given som ett tal mellan 0 och 1. Anges vid PoMix- och ParMixurval.

`sort`: Sorteringsvariabel för implicit stratifiering. Anges vid systematiskt urval om implicit stratifiering önskas.

`pi`: Urvalssannolikheter(er). Anges vid Bernoulli- och Poissonurval.

`seed`: Frö till slumpalsgeneratorn. Anges om reproducerbarhet önskas.

`slumptal`: Permanent slumpal. Anges i förekommande fall. Hämtas automatiskt om makrot körs inom inkapslingen `FTGSAMU`.

Output: Ett SAS-dataset kallat `Urvalsram`, som är en kopia av ramen tillsammans med en variabel som indikerar vilka objekt som är med i urvalet samt inklusionssannolikheter (`lambda`) om π ps-urval har använts.

3 Blandade urvalsmetoder

Om urvalsmetod varierar över stratumen körs MultiUrval.

```
%MultiUrval(ram=, metod=, lillan=, stratum=,  
<metods specifika parametrar>);
```

Ram är ett SAS-dataset och övriga parametrar (inklusive metod) är variabler på ramen. Variabeln metod specificeras per stratum och kan ha värdena osu, par, sekpo, pomix, parmix, sys, syspips, ber och po. Lillan, stratum samt metods specifika parametrar är desamma som för EnkeltUrval. Output från %MultiUrval är detsamma som det från %EnkeltUrval.

4 Klusterurval i två steg

Ett rudimentärt makro för klusterurval i två steg. Enbart osu stöds i första steget, men i andra steget stöds samtliga implementerade metoder.

```
%KlusterUrval(ram=, metod1=, metod2=, lillan1=,  
lillan2=, kluster=, <metods specifika parametrar>);
```

5 Tvåfasurval

Ett rudimentärt makro för tvåfasurval. Enbart osu stöds.

```
%TvafasUrval(ram=, idvar=, metod1=, lillan1=,  
stratum1=, metod2=, lillan2=, stratum2=);
```

6 Panelurval

Ett rudimentärt makro för panelurval. Alla implementerade metoder stöds. Vänligen kontakta Kira Gylling om du tänker använda detta makro.

```
%PanelUrval(ram=, metod=, npan=, idvar=, t=,  
<metods specifika parametrar>);
```

7 Tillämpning: FTGSAMU

FTGSAMU är en tillämpning av urvalstjänsten som används för att dra samordnade företagsurval inom ramen för det ekonomisk-statistiska systemet. Ett anrop är på formen:

```
%FTGSAMU(ram=, urvalsnr=, objtyp=, version=, metod=,
lillan=, stratum=, skarp=, <metodspecifika
parametrar>);
```

Input:

ram: Namn på det SAS-dataset som urvalet ska dras från.

stratum: Stratifieringsvariabel, angiven som variabel på ram.

lillan: Urvalsstorlekar, angivna som variabel på ram.

urvalsnr: Det nummer som undersökning har i det ekonomisk-statistiska systemet.

metod: Metod(er) som ska användas. Kan vara olika i olika stratum, i vilket fall de ska vara angivna som variabel på ramen. Metoder som stöds är samtliga implementerade metoder som:

- a) kan använda sig av permanenta slumpstal och
- b) har en på förhand fixerad urvalsstorlek.

Det vill säga i nuläget *osu*, *par*, *parmix*, *sekp* och *pomix*.

objtyp: Objekttyp som ska dras, exempelvis AE eller FE.

version: SAMU-version som ska användas.

skarp: Y om urvalet ska lagras i databas, behöver inte anges i annat fall.

Metodspecifika parametrar:

size: Storleksmått, angiven som variabel på ram. Anges endast vid π ps-urval.

lambda: Urvalssannolikheter vid π ps-urval, om exempelvis totalundersökning av en delmängd av ramen önskas.

B: Blandningsgrad, anges vid ParMix- eller PoMixurval.

Output:

Dataset Urvalsram som är en kopia på ramen tillsammans med en indikatorvariabel som anger vilka objekt som är utvalda (1 om utvalt, 0

annars) samt en variabel som anger urvalssannolikheter om π ps-urval används för undersökningen.

Urval lagrat i databas **SAMU_Urval** på server **q13**a om makrot är kallat med skarp=Y.

Exempel på användning:

```
%FTGSAMU(ram=myDataSet, stratum=stratum,  
lillan=lillan, metod=par, objtyp=AE,  
size=storleksmatt, urvalsnr=242, version=Aug2018,  
skarp=Y);
```

Exempel på hämtning av urval (proc sql):

```
libname samu_urv odbc schema=dbo  
  noprompt="driver=ODBC Driver 13 for SQL Server;  
           server=star.prod.sql;  
           database=SAMU_Urval;  
           Trusted_Connection=Yes";
```

proc sql;

```
  create table MittUrval as  
  select *  
  from samu_urv.AE  
  where Version = "Aug2018"  
  and UrvNr = 242  
  and Utvald = 1;
```

```
  create table StratumInfo as  
  select *  
  from samu_urv.Stratum  
  where Version = "Aug2018"  
  and UrvNr = 242;
```

```
  create table UrvalMedStratumInfo as  
  select a.*,  
         b.Lillan,  
         b.StoraN  
  from MittUrval a  
  join StratumInfo b  
  on a.Stratum = b.Stratum;
```

quit;

Exempel på hämtning av urval (datasteg):

```
libname samu_urv odbc schema=dbo  
  noprompt="driver=ODBC Driver 13 for SQL Server;  
           server=star.prod.sql;  
           database=SAMU_Urval;  
           Trusted_Connection=Yes";
```

data MittUrval;

```
    set samu_urv.AE;
    where Version = "Aug2018" and UrvNr = 242;
run;

data StratumInfo;
    set samu_urv.Stratum;
    where Version = "Aug2018" and UrvNr = 242;
run;

proc sort data=MittUrval;
    by Stratum;
run;

proc sort data=StratumInfo;
    by Stratum;
run;

data UrvalMedStratumInfo;
    merge MittUrval StratumInfo(keep=Stratum Lillan
StoraN);
    by Stratum;
run;
```

8 Skolurval

Inom uppdragsundersökningar av skolenheter behövs ett system för att sprida uppgiftslämnarbördan – dvs negativ samordning – men stratifieringar i olika undersökningar skiljer sig ofta från varandra och är vanligtvis inte kända på förhand, så ett helt annat system än det som används i %FTGSAMU behövs.

I DraUrval v1.1 är detta löst med s.k. Cotton & Hesses metod, som är en metod för urvalssamordning för (stratifierat) OSU. I likhet med SAMU används permanenta slumpstal, men istället för att transformera dem inför olika undersökningar så permuteras dem efter varje urvalsdragning. På så vis får de dragna enheterna de högsta slumpstalen och de ej dragna får de lägsta, så att i nästa undersökning dras i första hand de som inte drogs i den föregående.

OBS 1: För att kunna köra detta makro behöver du först be om behörighet till databasen **SKOLENH** på servern **skolenh.prod.sql** och skriva i mailet till behörig beställare att du behöver rollen **Skolenh_Enkatuppdrag**.

OBS 2: Makrot har en valfri flagga som heter ”skarp”. Används denna så förändras slumpstalen i databasen och de gamla sparas inte. Det är därför ytterst viktigt att bara köra skarpt om du är helt säker på att urvalet är det korrekta. Kör gärna en gång extra utan skarp-flaggan för att dubbelkolla att allt går rätt till innan du kör skarpt.

Makrot som körs heter %SkolUrval och ett anrop körs på formen:

```
%SkolUrval(ram=, stratum=, lillan=, idvar=, skarp=);
```

Input:

ram: Namn på det SAS-dataset som urvalet ska dras från.

stratum: Stratifieringsvariabel, angiven som variabel på ram.

lillan: Urvalsstorlekar, angivna som variabel på ram.

idvar: Identifierande ramvariabel, heter troligen skolenhetskod.

skarp: Y om urvalet ska användas i produktion, behöver inte anges i annat fall.

Output:

Dataset Urvalsram som är en kopia på ramen tillsammans med en indikatorvariabel som anger vilka objekt som är utvalda (1 om utvalt, 0

annars), och som även innehåller de permuterade permanenta slumpfallen. Slumfallen är även permuterade i tabellen **u_h_skolenhet** på ovan nämnda databas **SKOLENH** om makrot är kallat med `skarp=Y`.

Exempel på användning:

```
%SkolUrval(ram=myDataSet, stratum=stratum,  
lillan=lillan, idvar=skolenhetskod, skarp=Y);
```