

Seminar on Registers in Statistics - methodology and quality

21 - 23 May, 2007 Helsinki

Frame Populations and Register Populations

Anders Wallgren and Britt Wallgren, Statistics Sweden

Today:

National Accounts (NA) receive inconsistent macro data from different sources - these inconsistencies are handled by the NA

New project:

Microdata from these sources are integrated, inconsistencies are handled at the micro level – NA will then get consistent macro data

Three kinds of surveys:

Surveys based on own data collection:

1. Sample surveys
2. Censuses

Surveys based on administrative registers:

3. Register-based surveys

Three kinds of surveys:

Surveys based on own data collection:

1. Sample surveys
2. Censuses

Different ..

.. terms

.. methods

Surveys based on administrative registers:

3. Register-based surveys

Three kinds of surveys:

Surveys based on own data collection:

1. Sample surveys
2. Censuses

Different ..

.. terms

.. methods

Frame population

Surveys based on administrative registers:

3. Register-based surveys

Register population

Calendar year register

Registers and Time

When defining a register with regard to time, different register types must be distinguished.

1. **The current stock register** is updated with all available data on currently active/live objects. The current stock register is used as **frame population**.
2. **The register referring to a specific point in time**, such as the turn of the year, is updated to describe the object set at that point in time.

This kind of **register population** referring to the population on December 31 is used for almost all register-based surveys on persons at Statistics Sweden.

Chart 2. Integrated register based on five registers on persons

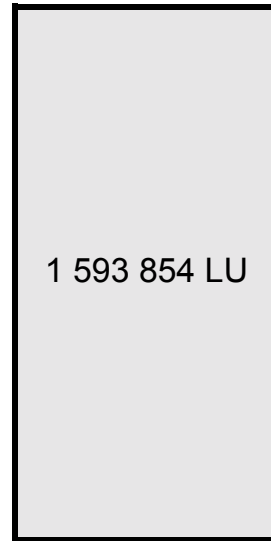
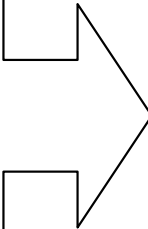
PIN	Classification variables from the Population Register					Statistical variables from the ...											
	Age	Sex	Region	var..	var..	Employment Reg			Education Reg			Income & Taxation			Occupation Reg		
						var..	var..	var..	var..	var..	var..	var..	var..	var..	var..	var..	var..
1																	
2																	
3																	
...																	
N																	

3. **The calendar year register** is the register version containing all objects that have existed at any point during a specific year. It should be used as **register population**.

November 2004 Frame

Source 1:
Population:
Continuous updates
based on changes in
the business register
kept by tax
authorities

Variables:
Activity code
Sector
NACE



Activity code November 2004

	Number of LU
'Not started'	335 003
'Active'	872 391
'Discontinued'	386 451

LU = Legal unit

↑
15-nov

November 2004 Frame



LU = Legal unit

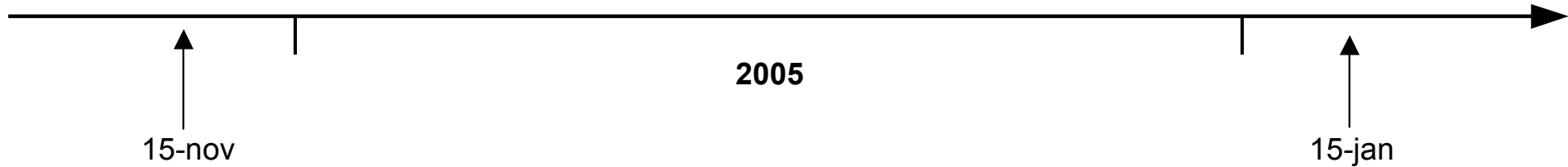
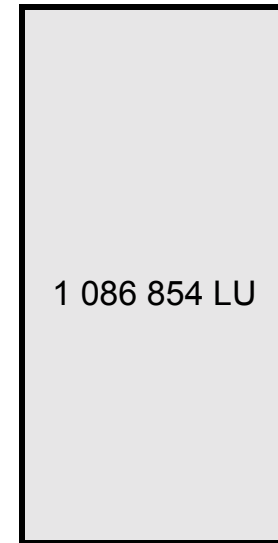
Activity code November 2004

	Number of LU
'Not started'	335 003
'Active'	872 391
'Discontinued'	386 451

Population:
Source 2, 3, 4, 5, 6
Tax declarations

Variables:
Source Statistics
Sweden Business
Register:
Sector
NACE

Calendar year register



November 2004 Frame

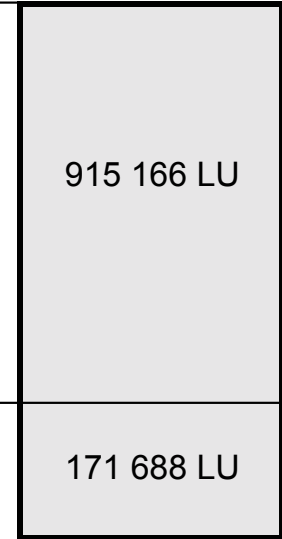
Calendar year register

Activity code November 2004

'Not started' 287 341
'Active' **93 114 overcoverage**
'Discontinued' 298 224



'Not started' 47 662 *undercoverage*
'Active' 779 277
'Discontinued' 88 227 *undercoverage*



LU = Legal unit

171 688 undercoverage



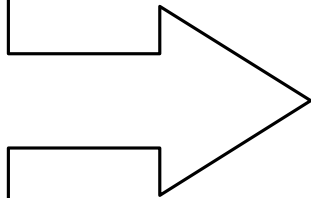
Chart 3. Six sources of a calendar year Business Register 2004

Register regarding 2004		Number of legal units, LU
Administrative source 1	Active in Business Register, November frame	872 391
Administrative source 2	Enterprises, yearly income declarations	980 305
Administrative source 3	VAT declarations	689 577
Administrative source 4	Yearly wage sums from statements of earnings	311 943
Administrative source 5	Monthly wage sums	310 628
Administrative source 6	Foreign trade	68 002

Chart 4. Calendar year population

Sources 2 – 6 are
Integrated => Calendar
year population

NACE, Sector, region
from source 1



	Classification variables from the Business Register		
BIN	NACE	Sector	Region
1			
2			
3			
...			
N			

1. This register will be used as **standardised population for all business surveys** that deliver data to the yearly National Accounts.
2. We also **measure coverage errors and inconsistencies** with this register.

2. We also **measure coverage errors and inconsistencies** with this register.

Chart 5. Undercoverage in the November frame of the Business Register 2004

Number of legal units	Adm source 2	Adm source 3	Adm source 4	Adm source 5	Adm source 6
Not in the Business Register	260 943	83 487	33 387	30 533	14 778
In the Business Register	719 362	606 090	278 556	280 095	53 224
Total	980 305	689 577	311 943	310 628	68 002

The Swedish Business Register will be redesigned

Errors due to **undercoverage in the November Frame of the Business Register 2004**

	Gross annual pay SEK Millions	Turnover SEK Millions	Number of LU
'Active' in the Business Register, gives old estimates	1 003 186	5 582 374	779 277
Missing completely in the Business Register	639	5 061	171 688
'Not started' in the Business Register	177	1 734	47 662
'Discontinued' in the Business Register	5 671	112 363	88 227
Total in Calendar Year Register, new estimates	1 009 673	5 701 532	1 086 854
Coverage error due to undercoverage, new - old estimate	6 487	119 158	307 577
Coverage error due to undercoverage, per cent	0,65	2,13	39,47

Coverage errors differ between industries!

Also **overcoverage errors**

Conclusions:

- Consistency:
All surveys are based on the same population
- Coverage:
Use *all* relevant administrative sources
- The Calendar Year Population is essential
- Frames can't be used as standardised populations