

# **Informe de l'Idescat sobre metodologia actual en l'estadística pública. La integració d'informació administrativa i estadística**

**Consell Català d'Estadística**

Reunió del 1 de desembre del 2011



Generalitat de Catalunya  
**Institut d'Estadística de Catalunya**





# **Informe de l'Idescat sobre metodologia actual en l'estadística pública**

## **La integració d'informació administrativa i estadística**

**Consell Català d'Estadística  
1 de desembre de 2011**



# **Informe Idescat: metodologia actual en l'estadística pública**

## **La integració d'informació administrativa i estadística**

1. El marc d'oportunitats. Externalitats
2. La identificació de fonts estadístiques i administratives
3. Algunes limitacions del potencial ús dels registres
4. Il·lustracions de pràctiques precedents, en curs i referents



# Informe Idescat: metodologia actual en l'estadística pública

## La integració d'informació administrativa i estadística (2)

### El marc d'oportunitats

- Informació per bastir directoris i marcs de llista o marcs d'àrea
- Producció estadística basada en informació administrativa de referència (register-based models)
- Producció estadística orientada a complementar la no-informació administrativa disponible
- Producció estadística combinada per a estimacions indirectes i/o validació
- Millores en l'actualitat i la granularitat de la informació (territorial)
- Explosió de dades generades a l'entorn públic i privat

### Externalitats

- Reducció de la càrrega que suporten les unitats informants
- Difícil sostenibilitat de les operacions estadístiques pesants



## **Informe Idescat: metodologia actual en l'estadística pública**

### **La integració d'informació administrativa i estadística (3)**

#### **La identificació i disponibilitat de dades estadístiques (oficials) i d'informació administrativa (individual)**

- El Registre de fitxers estadístics gestionat per l'Idescat (dades de persones físiques i jurídiques). Inclusió de dades/fitxers d'altres sistemes estadístics
- Intercanvis d'informació regular amb l'APDCAT (dades de persones físiques)
- Informes de l'Idescat sobre creació/modificació de registres administratius (dades de persones físiques i jurídiques)
- Protocols d'intercanvi de directoris generals i sectorials amb l'INE/SEE

#### **Tractament de la informació administrativa com una (altra) modalitat de recollida de dades primàries (cobertura, no-resposta, biaixos, etc.)**



# Informe Idescat: metodologia actual en l'estadística pública

## La integració d'informació administrativa i estadística (4)

### Algunes limitacions ...

- La informació administrativa no està (prou) estandarditzada ni estructurada
- Calen igualment operacions periòdiques de manteniment i actualització (treballs de camp)
- Canvis administratius o legals: trencament de les sèries estadístiques
- Capacitat de processament/tractament de grans volums d'informació
- Calibrar la confidencialitat de les dades individuals i la seva fiabilitat tècnica



# Informe Idescat: metodologia actual en l'estadística pública

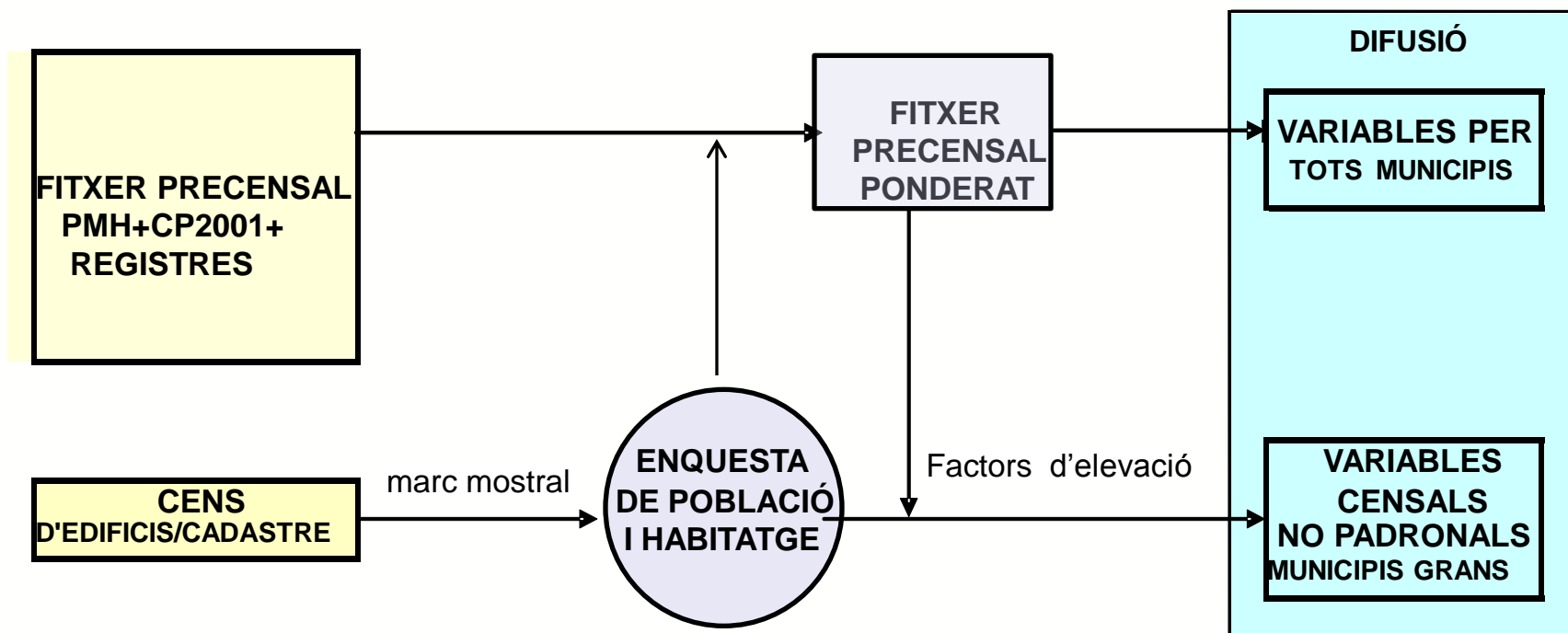
## La integració d'informació administrativa i estadística (i 5)

### II·lustracions de pràctiques precedents, en curs i referents

- **Complement i contrast d'informació administrativa**: cens de població 2011, cens agrari 2009
- **Fusió de registres**: estimació de les desigualtats de mortalitat i de factors de salut
- **Combinació de registres administratius**: índexs de privació material per seccions censals, mostra contínua de vides laborals
- Polivalència d'aprofitament: (micro)dades fiscals per a directoris, validació de dades, estimacions sectorials i estadística de síntesi
- Retroalimentació de fitxers/registres: ingressos i condicions de vida longitudinals d'individus (SCB)



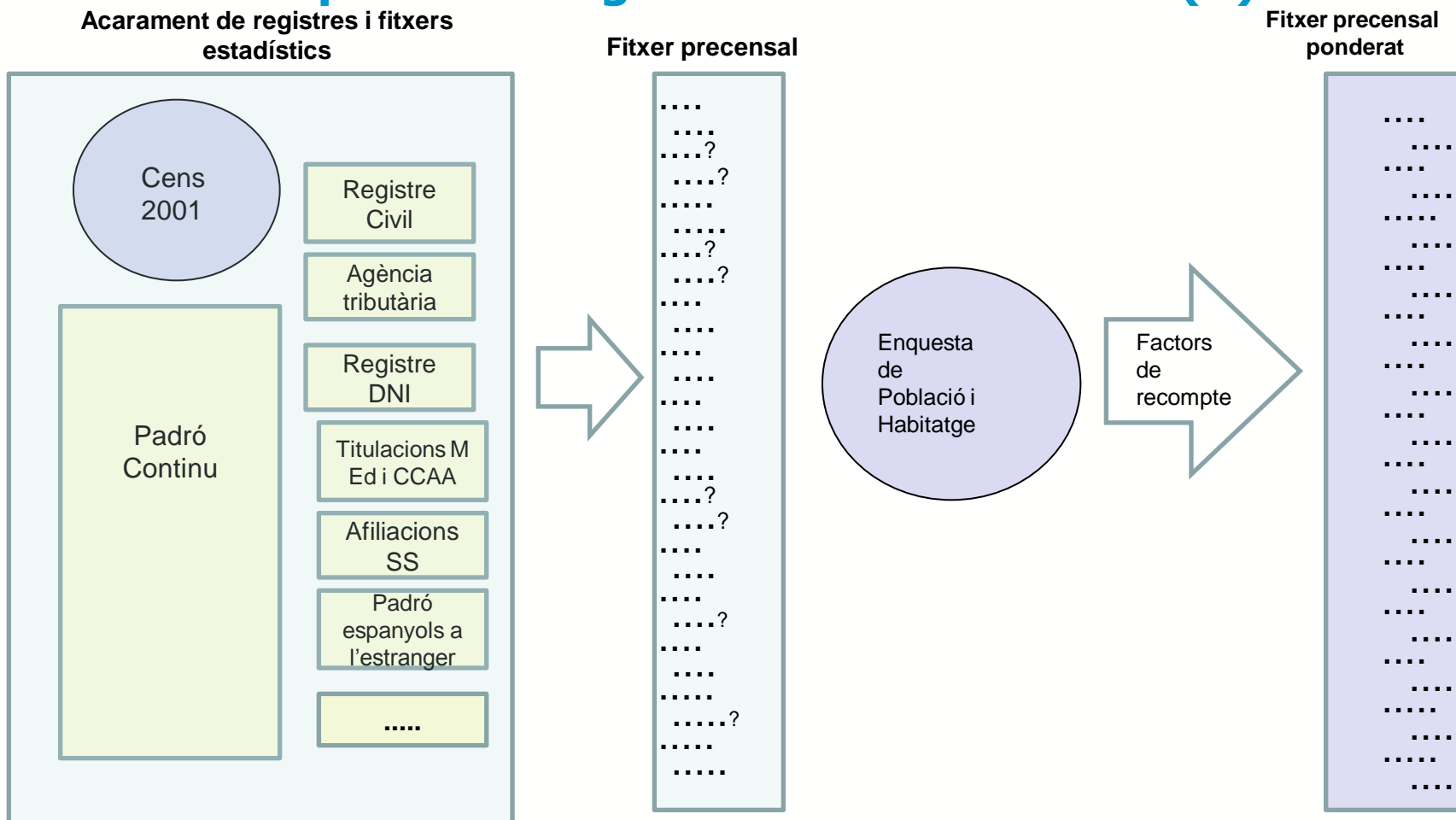
# 1. Els Censos de població 2011: una combinació d'enquestes i registres administratius (1)





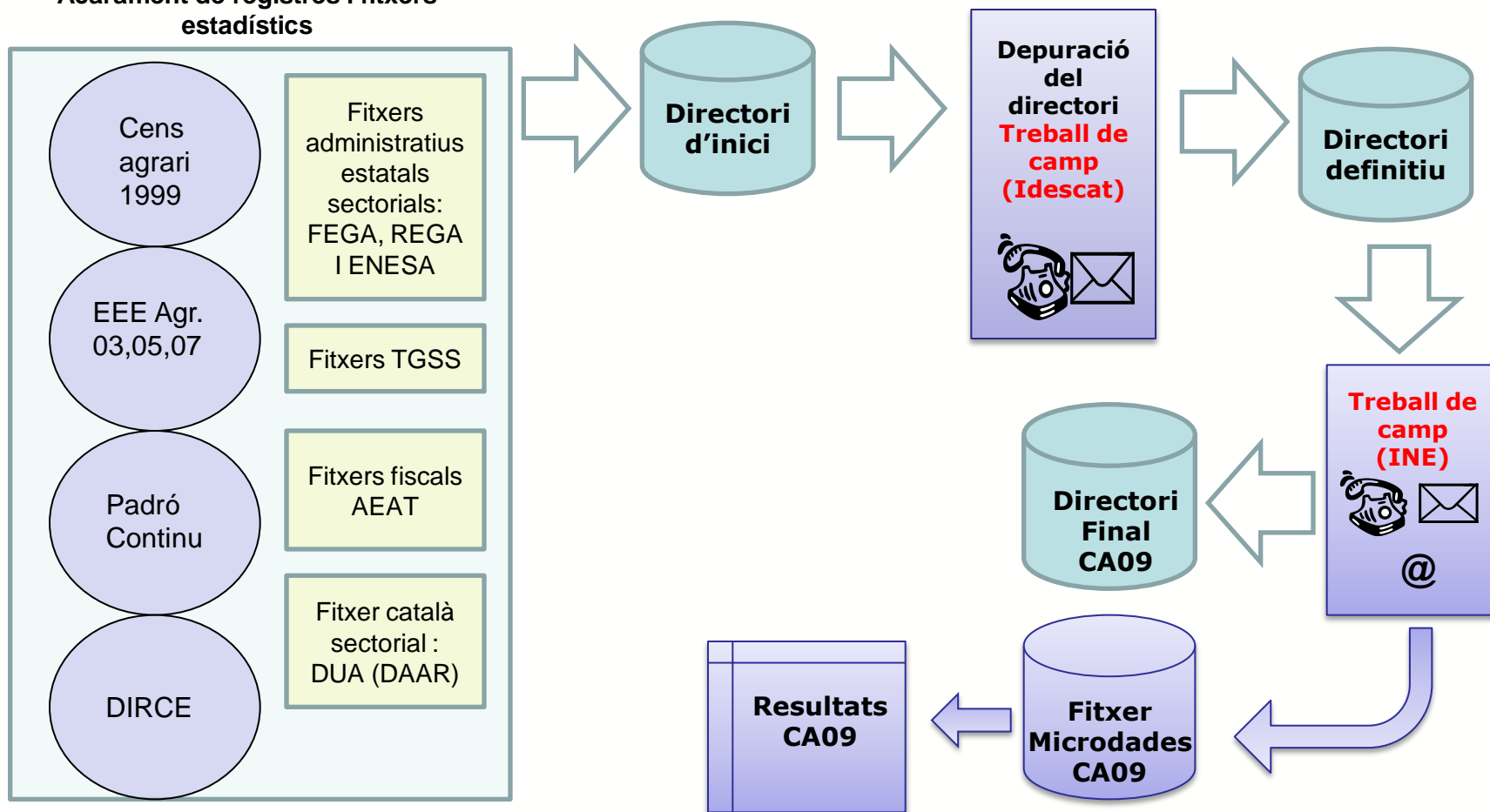


# 1. Els Censos de població 2011: una combinació d'enquestes i registres administratius (2)

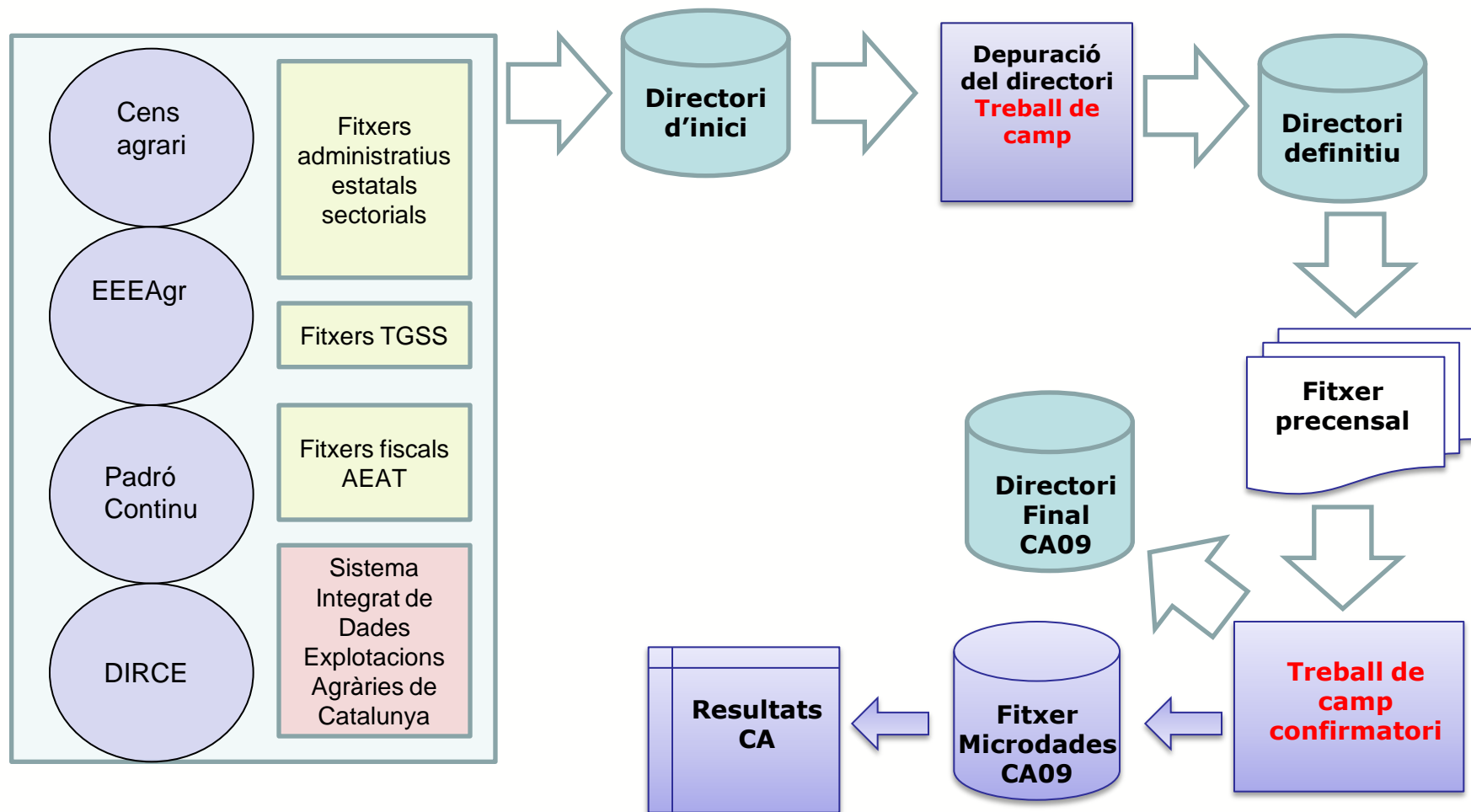


## 2. El Cens agrari 2009: registres administratius pels directoris de camp i complement d'informació

Acarament de registres i fitxers estadístics



## 2. El Cens agrari 2009: registres administratius pels directoris de camp i complement d'informació (2)

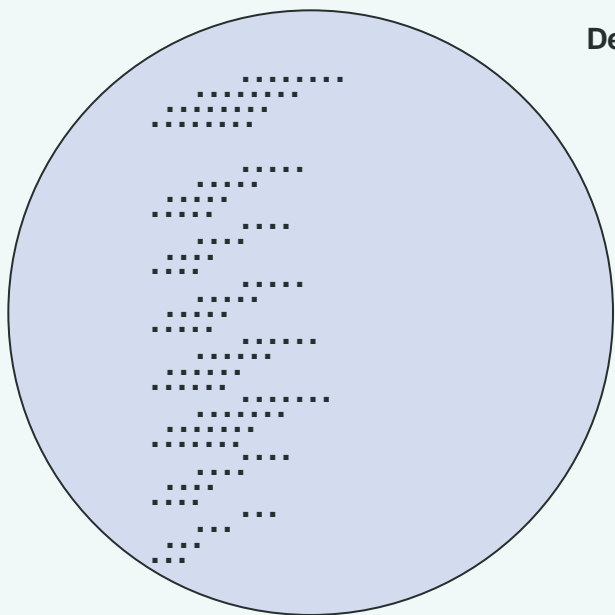




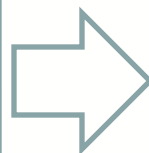
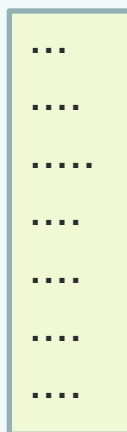
### 3. Fusió de registres: mortalitat segons nivell instrucció

Acarament de registres i fitxers estadístics

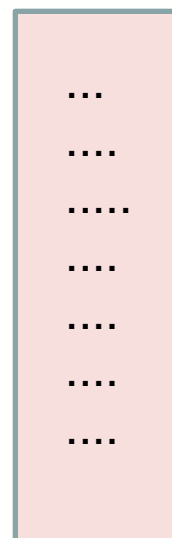
Cens de població 2001



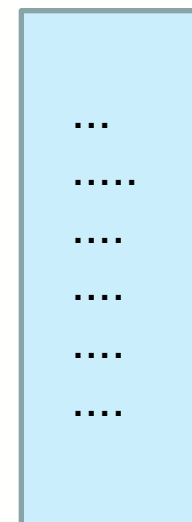
Defuncions



Defuncions segons  
nivell d'instrucció 2001



Difusió :  
Indicador de Mortalitat Estàndard (IME)  
per nivell d'instrucció

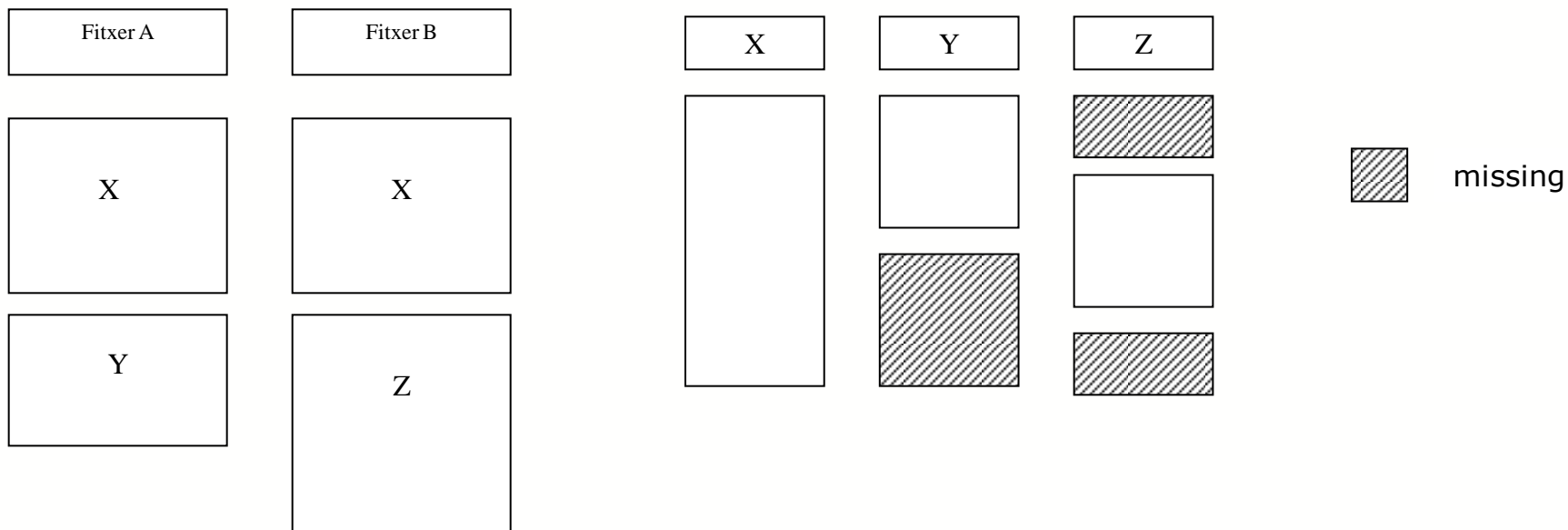




## 4. Fusió de registres: imputació de variables mancants

La fusió de dades consisteix a integrar la informació de dos fonts de dades independents sobre poblacions semblants. Implica la imputació d'un bloc complet de variables mancants. El vector X representa les variables comunes, mentre Y i Z són les variables específiques dels fitxers A i B, respectivament.

Una presentació alternativa:





## 4. Fusió de registres: imputació de variables mancants (1)

La fusió de dades pot aplicar-se des de diverses perspectives, per exemple predir en base a dos fonts de dades independents successives. Implica la imputació d'un bloc complet de variables mancants en base a valors anteriors

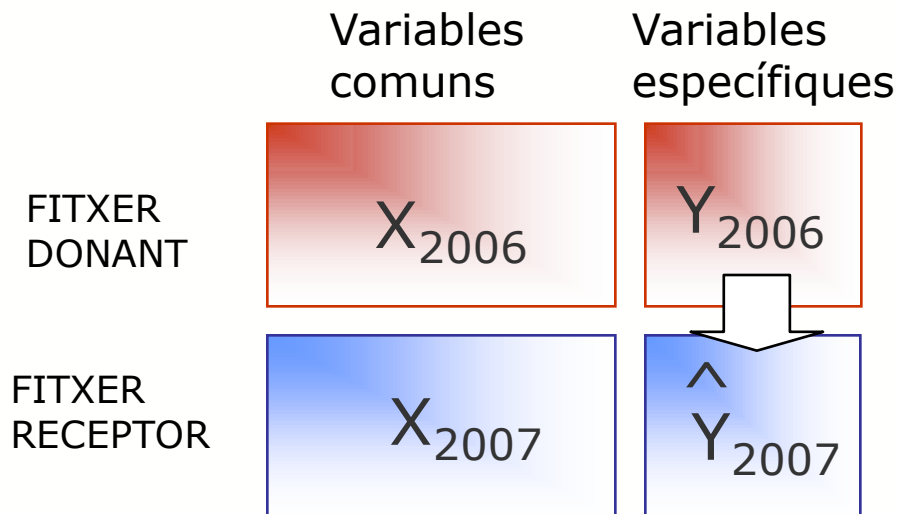
### Fusió unilateral

Fitxer donant  $(X_{2006}, Y_{2006})$

Fitxer receptor  $(X_{2007})$

Variables comunes  $x$

Variables específiques  $y$



### Estratègia

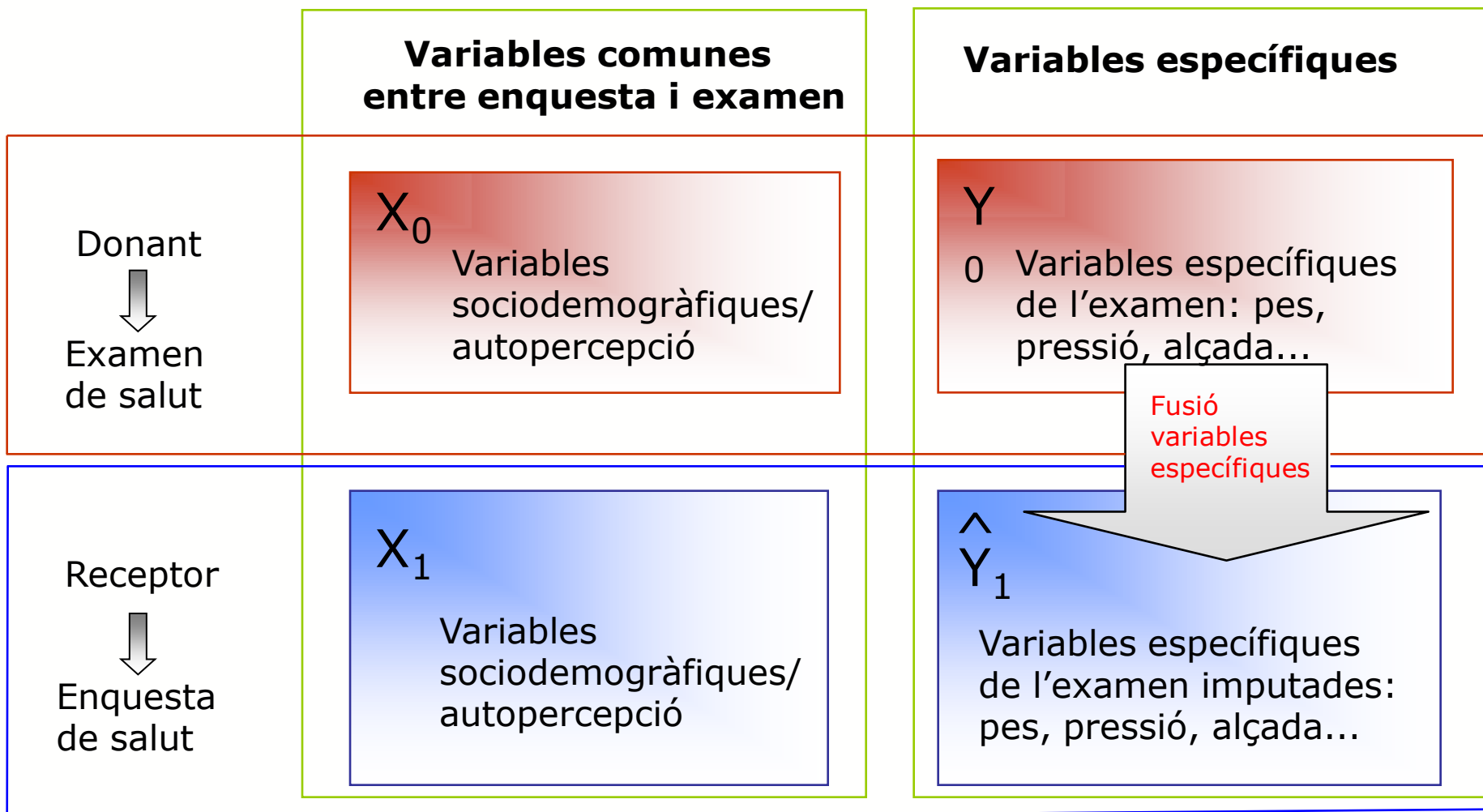
Selecció de las variables comunes amb més capacitat predictiva

Fusió de dades

Validació per remostreig



## 4. Fusió de registres: imputació de variables mancants (2)





## 4. Fusió de registres: imputació de variables mancants (3)

### Fusió de dades (2010)

#### Variables predictives

sexe; talla; regió sanitària; edat; índex de massa corporal; estudis; fumador; cintura; pressió arterial; activitat física; diabetis; alcohol; situació laboral

#### Variables específiques (fusió)

Talla mesurada  
 Pes mesurat  
 Perímetre de cintura  
 Perímetre de malucs  
 Pressió arterial sistòlica  
 Pressió arterial diastòlica

### Fusió de dades (2011)

#### Propensió a malalties del cor (Z)

sedentarisme; edat; sexe; fumador; hipertensió arterial; obesitat; diabetis; dislipèmia; colesterolèmia.

#### Propensió a diabetis (Z'):

alteracions de greixos a la sang; obesitat; sedentarisme; edat; hereditaris; hipertensió arterial; sexe

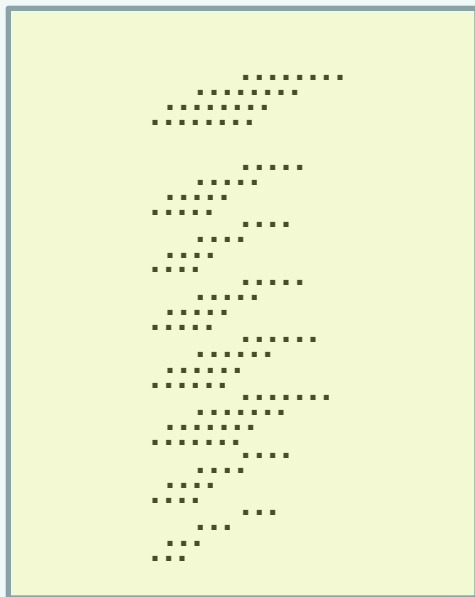




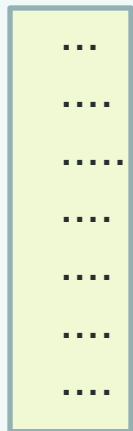
## 5. Combinació de registres: índexs de privació

Acarament de registres administratius de diferent orígens

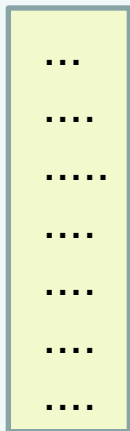
Registre de població  
de Catalunya



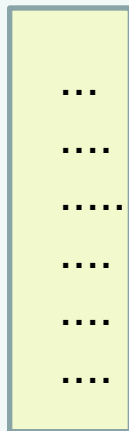
Beneficiaris  
cursos de  
formació  
SOC



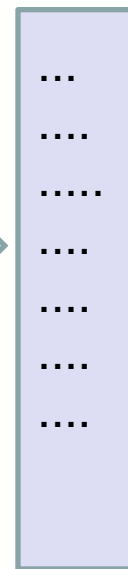
Beneficiaris  
RMI



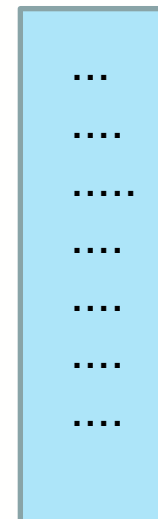
Altres  
registres



Georeferenciació:  
Assignació territorial  
adreces beneficiaris



Índex de privació  
(nivell secció censal)





## 6. Combinació de registres: MCVL

**Objectiu:** estudis longitudinals sobre les trajectòries laborals de la població a partir de la informació regular de la MCVL (Mostra Contínua de Vides Laborals)

**Mètode:** La MCVL és un fitxer de microdades construït a partir d'un mostreig aleatori simple del 4% de tota la població que ha tingut alguna relació econòmica amb la Seguretat social dins l'any de referència, ja sigui com a cotitzants o pensionistes. Es captura l'historial de les persones seleccionades remuntant-se fins allà on arribin els registres informatitzats.

Fitxers administratius combinats:

- Dades personals, història laboral, bases de cotització, etc.: fitxers generals d'afiliació de la TGSS
- Rendes i salaris: declaracions IRPF – AEAT
- Convivència: Registre de població (padrons municipals-INE)





## 7. Múltiples utilitats dels registres fiscals (1)

### Algunes fonts

- ❖ Impost sobre la renda de les persones físiques
- ❖ Impost sobre societats
- ❖ Registre de Duanes de l'AEAT
- ❖ Impost sobre Valor Afegit (especial interès TIOC i comptes)
- ❖ Impost sobre Activitats Econòmiques

### i algunes de les seves utilitats:

Dades agregades (dades tabulars)	{	Macromagnituds econòmiques (+ desagregació) Vendes i compres de les empreses Import/export. (comerç amb l'estranger)
-------------------------------------	---	--

Microdades (individuals)	{	Directoris i marc de mostres estadístiques Contrast/validació d'informació
-----------------------------	---	---

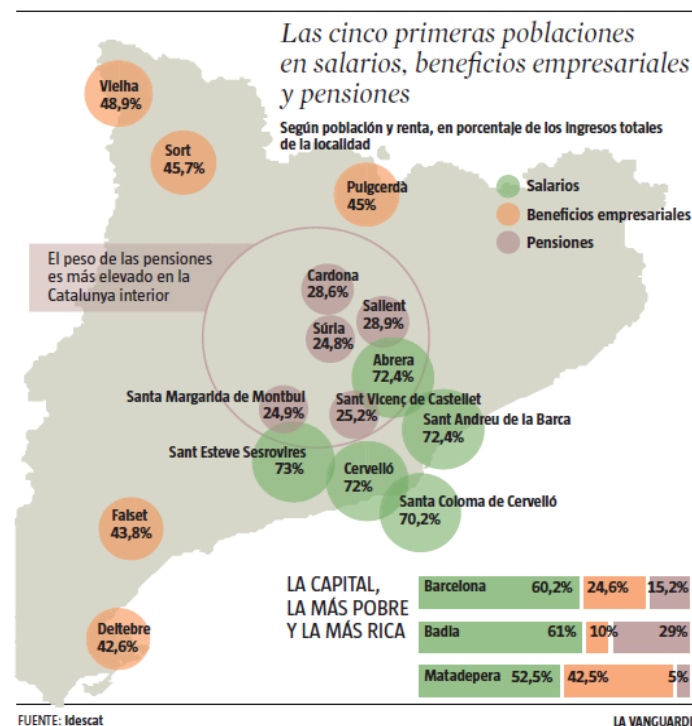
## 7. Múltiples utilitats dels registres fiscals (2)

### Exemples del seu ús a l'Idescat

Estadística de Comerç amb l'estranger contingut tecnològic



Macromagnituds territorials: RFD comarcal i municipal





## 8. Retroalimentació de registres: dades longitudinals (1)

### Household finances

#### General statistics

Survey	Responsible agency
› Household finances general statistics	SCB

### Household expenditures

Survey	Responsible agency
› Household budget survey (HBS)	SCB

### Income and income distribution

Survey	Responsible agency
› <u>Distributional analysis system for income and transfers</u>	SCB
› <u>Firm Register and Individual Data Bases (FRIDA)</u>	SCB
› <u>Household's finances</u>	SCB
› <u>Income and tax statistics</u>	SCB
› <u>Longitudinal Individual Data Base (LINDA)</u>	SCB
› <u>Wealth statistics</u>	SCB



Fitxers estadístics (enquesta)

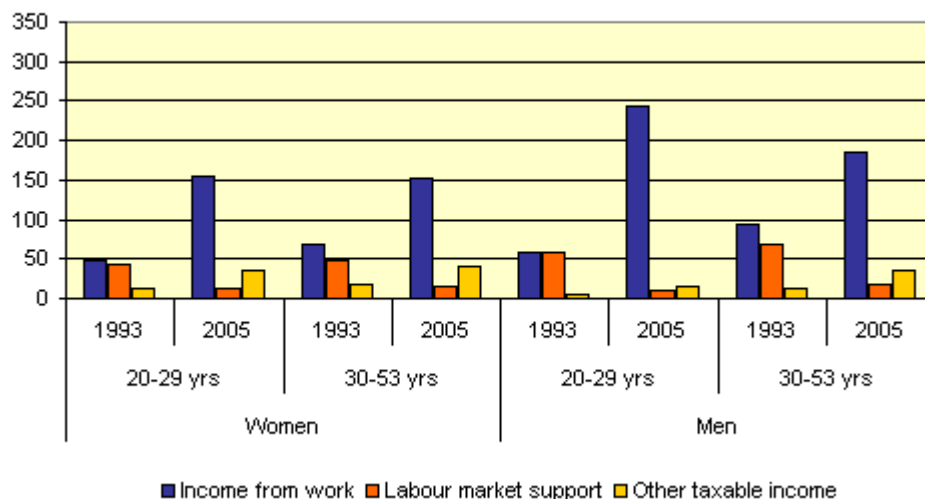
Fitxers administratius



## 8. Retroalimentació de registres: dades longitudinals (2)

### Longitudinal Individual Data Base (LINDA)

Taxable income for those seeking work in 1993 in 2005 prices, SEK thousands



- Fitxer longitudinal de dades individuals a partir d'una mostra panel de 300.000 individus (anual)

- La informació individual s'extreu dels fitxers fiscals i dels censos de població quinquennals (a partir dels 90s informació provinent de la gestió de l'atur).

- Informació representativa de la població des de 1960

Aplicabilitats:

- ✓ fusionable amb altres fitxers estadístics a demanda dels investigadors
- ✓ marc de mostra
- ✓ avaluació de polítiques públiques
- ✓ anàlisi longitudinal

[http://www.scb.se/Pages/Standard\\_38903.aspx](http://www.scb.se/Pages/Standard_38903.aspx)

## 8. Retroalimentació de registres: dades longitudinals (3)

### Household's finances

Overview Tables and Graphs Publications See also

Total amounts for all households[1], corrected 2011-11-11

To list of tables

	1991	1995	2001	2003	2004	2005	2006	2007	2008	2009	2010
<b>Primary income</b>											
Wage	851 650 792 912	1 059 241	1 072 724	1 091 575	1 132 509	1 171 425	1 239 652	1 255 774	1 262 417	1 289 471	
Business income	28 171	23 495	52 214	47 273	51 442	55 842	55 211	55 816	57 592	55 959	53 645
Interests/dividends	59 487	51 600	48 230	46 387	47 374	48 434	77 654	91 988	99 063	92 133	75 752
Capital income	50 582	16 112	61 762	52 974	70 327	111 958	156 851	205 631	103 116	101 002	123 941
Sum Primary income	989 890 884 118	1 221 448	1 219 360	1 260 717	1 348 745	1 461 142	1 593 086	1 515 545	1 511 510	1 542 808	
<b>Taxable transfers</b>											
Pensions	253 280 282 419	314 209	303 603	318 149	319 341	329 106	333 617	338 453	365 560	368 462	
Sickness and activity compensation <sup>[2]</sup>	0	0	0	52 982	58 026	58 813	59 732	61 714	55 055	56 126	48 566
Sickness compensation <sup>[3]</sup>	54 239	23 466	47 164	49 961	42 542	36 907	34 495	28 089	23 884	19 230	18 667
Parental allowance	23 060	21 951	18 632	21 386	21 475	24 243	24 910	24 381	25 863	27 479	28 541
Labour market assistance	24 536	61 724	36 147	38 951	47 017	45 616	41 533	26 426	20 882	30 892	32 919
Other	6 543	5 360	8 446	2 638	2 487	3 336	2 458	2 604	2 799	2 667	2 813
Sum Taxable transfers	361 661 394 922	424 599	469 520	489 697	488 256	492 235	476 832	466 937	501 953	499 967	
<b>Tax-free transfers</b>											
National child allowance	23 039	19 886	24 288	23 409	22 759	24 246	26 154	24 928	23 932	23 553	23 700
Study grants including study loans	13 368	16 179	22 515	25 655	25 224	24 262	24 355	23 593	21 866	24 808	25 569
Housing allowance	16 769	23 545	17 152	15 813	17 012	17 514	17 761	16 831	15 939	14 937	16 115
Social assistance <sup>[4]</sup>	6 570	12 904	9 695	8 099	7 921	10 100	9 327	10 021	9 886	10 424	11 282
Other	8 865	9 219	8 278	9 149	9 329	9 825	10 159	9 639	8 401	9 096	9 897
Sum Tax-free transfers	68 608 81 733	81 928	82 125	82 244	85 947	87 758	85 011	80 024	82 818	86 564	
<b>Taxes and public co-payments</b>	363 408 394 346	517 431	520 330	541 371	559 545	584 659	574 846	533 718	519 783	513 707	
Paid alimony and repaid study loans, etc.	19 273	19 558	28 638	27 795	28 654	28 829	29 726	30 084	26 495	27 487	27 405
<b>Disposable income</b>	1 037 477 946 869	1 181 906	1 222 880	1 262 634	1 334 573	1 426 749	1 550 000	1 502 293	1 549 011	1 588 227	

1) Also includes persons living in institutions, deceased, migrated to and emigrated from Sweden during the year.

2) Previously mentioned early retirement pension, then included in pensions.

3) Sickness benefit paid by the employer is included in the wage.

4) Formerly known as Socialbidrag.

### Household's finances survey

- Enquesta per mostreig anual
- Objectiu: obtenir la distribució de la renda entre diferents tipus de llars, nivell de despeses i nivells de vida en general
- Mostra extreta a partir de LINDA.
- Continguts sols disponibles a partir d'enquestes a la població