

Estadística pública, oficial o estadística a seques?

Frederic Udina

Institut d'Estadística de Catalunya

Assemblea
de la Societat Catalana d'Estadística
Barcelona, 7 de març de 2019



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



**SOCIETAT CATALANA
D'ESTADÍSTICA**

Outline

Què és l'estadística?

Què és l'estadística oficial?

Un cas paradigmàtic: la mesura de l'atur

Origen i evolució

A l'origen, l'**estadística** era la ciència de l'Estat.

Durant molts segles va seguir essent així: fins que apareix la probabilitat.

Llavors tenim el triangle DataAnalysis $\iff$ Probabilitat $\iff$ Inferència

I comencen a créixer les branques de l'arbre:

Biostat o biometria

Demografia

Epidemiologia

Psicometria

Tecnometria

Chemometrics

Sociometria

Bayesian stats (és una branca o tot un bosc nou?)

I ara apareix el dilema del DataScience de manetes del *big-data*!

La veritat és que la tentació és gran:

¿hauríem de dir que

L'estadística oficial és LA estadística

i els altres que vagin buscant nom?

En tot cas, gràcies



SOCIETAT CATALANA
D'ESTADÍSTICA

I endavant!



Generalitat de Catalunya
Institut d'Estadística
de Catalunya

Algunes definicions

L'**estadística pública** és la que s'interessa per l'activitat de la societat en general, per analitzar els problemes socials, econòmics, demogràfics, de salut, mediambientals, de l'ensenyament...

Com a subset: L'**estadística oficial** és l'estadística pública legalment regulada, duta a terme des dels governs com a bé públic i sotmesa a:

- ▶ Regulació legal
- ▶ Cobertura universal
- ▶ Garantia de qualitat i comparabilitat en el temps i en l'espai
- ▶ De resposta obligatòria per a persones i empreses
- ▶ Garantia de protecció de la confidencialitat i privacitat

Missió de l'Idescat

La missió de l'Institut d'Estadística de Catalunya (Idescat) és proveir informació estadística rellevant i d'alta qualitat, amb independència professional, i coordinar el Sistema estadístic de Catalunya, amb l'objectiu de contribuir a la presa de decisions, la recerca i la millora de les polítiques públiques.

Referents principals

Legislació principal

- ▶ Llei d'estadística de Catalunya
- ▶ Plans estadístics de Catalunya, actual 2017-2020
- ▶ Programes anuals d'actuació estadística, PAAE 2019

Codi de bones pràctiques de les estadístiques europees

<https://www.idescat.cat/metodes/bonespractiques/>

WEB DE L'ESTADÍSTICA OFICIAL DE CATALUNYA



Generalitat de Catalunya
Institut d'Estadística de Catalunya

idescat

Dades	Serveis	Mètodes	L'Institut	<i>cercar</i>
-------	---------	---------	------------	---------------

Codi de bones pràctiques de les estadístiques europees (edició 2017)

El Codi de bones pràctiques de les estadístiques europees és la pedra angular del marc comú de qualitat del Sistema estadístic europeu.

Es tracta d'un instrument autoregulator i es basa en setze principis que cobreixen l'entorn institucional, els processos estadístics i els resultats. Un conjunt d'indicadors de bones pràctiques i estàndards per a cadascun dels principis serveix d'orientació i referència per analitzar l'aplicació de bones pràctiques, i permet augmentar la transparència dins del Sistema estadístic europeu.



Generalitat de Catalunya
Institut d'Estadística
de Catalunya

Codi de bones pràctiques

El Codi es va adoptar totalment en el Pla estadístic de Catalunya 2017–2020.

La modificació de la llei d'estadística (Desembre 2016) va posar en primer pla

- ▶ l'ús de la informació d'origen administratiu
- ▶ L'ús de les dades per a la recerca i la millora de les polítiques públiques

Preàmbul

El Codi de bones pràctiques de les estadístiques europees

Entorn institucional

Principi 1: Independència professional

Principi 1Bis: Coordinació i cooperació

Principi 2: Mandat de recollida de dades i accés a les dades

Principi 3: Adequació dels recursos

Principi 4: Compromís de qualitat

Principi 5: Confidencialitat estadística i protecció de dades

Principi 6: Imparcialitat i objectivitat

Processos estadístics

Principi 7: Metodologia sòlida

Principi 8: Procediments estadístics adequats

Principi 9: Càrrega no excessiva per als enquestats

Principi 10: Relació cost-eficàcia

Productes estadístics

Principi 11: Rellevància

Principi 12: Precisió i fiabilitat

Principi 13: Oportunitat i puntualitat

Principi 14: Coherència i comparabilitat

Principi 15: Accessibilitat i claredat

Fonts d'informació estadística

Fonts primàries Les que es creen amb finalitat principal estadística, principalment enquestes o formularis específics.

Fonts secundàries Creades amb altra finalitat, s'aprofiten per a fer estadística. Origen administratiu, empreses públiques o privades, *big-data*,

L'aposta Idescat és la integració de fonts.

Dades	Serveis	Mètodes	L'Institut	<i>cercar</i>	gencat.cat
Demografia i societat Població Cultura · Llengua Educació Eleccions Justícia · Seguretat Salut Treball			Economia Macromagnituds Empreses · Finances Finances públiques Inversió i comerç exterior Preus Recerca · Tecnologia		Per àmbit geogràfic Províncies Àmbits del Pla territorial Comarques i Aran Municipis Districtes Seccions censals
Qualitat de vida Condicions de vida Habitatges i edificis Ingressos i consum de les llars Protecció social			Sectors econòmics Agricultura · Ramaderia · Pesca Comerç · Serveis Construcció Indústria · Energia Transport Turisme		Dades bàsiques Indicadors de conjuntura econòmica Indicadors anuals Indicadors de la Unió Europea Indicadors demogràfics i de territori Anuari estadístic de Catalunya El municipi en xifres
			Medi ambient · Territori Medi ambient Territori		
Relació d'estadístiques disponibles • Calendari de difusió de resultats • Rectificacions • Novetats					

TRANSFORMATIONAL CHANGE IN METHODOLOGIES (1)

From classical to contemporary statistical methods:

- From **designed** data to **found** data
- From **direct** measurements to **modelled** data
- From **single** source to **multiple** sources
- From **siloed** data sets to **integrated** data sets

(1) *Dr Siu-Ming Tam, Chief Methodologist, Australian Bureau of Statistics, NTTS 2017*

Un cas interessant

La mesura de l'atur, exemple de fonts diverses i de l'interès de combinar-les

- ▶ EPA: L'Enquesta de població activa, *Labour Force Survey*
- ▶ ATUREG: L'Atur registrat
- ▶ EPATERR: Una combinació per baixar l'EPA a àrees petites

Però abans cal definir curosament:

- ▶ Què és una persona activa?
- ▶ Què és un aturat?
- ▶ Taxa d'activitat, taxa d'atur...

Enquesta de població activa

EPA Enquesta de població activa

Període inicial i final

T3/1976 T4/2018

Freqüència

Trimestral

Desagregació territorial

Catalunya Províncies

Darrera actualització

12 de febre de 2019

Dades 2018

Propera actualització

Abril del 2019

Formació i condicions de treball
2018

L'Enquesta de població activa és la principal font d'informació estadística del mercat de treball. L'objectiu principal és conèixer la relació de la població amb l'activitat econòmica: població activa, població ocupada, població desocupada, taxes d'activitat, ocupació i atur.

L'Enquesta de població activa és una investigació per mostreig de caràcter continu duta a terme per l'INE a tot el territori espanyol. Trimestralment, l'Idescat porta a terme una explotació estadística detallada per a l'àmbit de Catalunya, de la qual obté l'ampliació de resultats següents:

- ☐ Resultats bàsics trimestrals i anuals de Catalunya i Espanya (inclou resultats comparatius per a les quatre províncies catalanes).
- ☐ Resultats detallats trimestrals i anuals de Catalunya.

Indicador destacat

Ocupats

Variació interanual

Període

T4/2018

Valor

2,3 %

Ocupats en la indústria

Variació interanual

T4/2018

-0,5 %

Ocupats en la construcció

Variació interanual

T4/2018

2,5 %

Ocupats en els serveis

Variació interanual

T4/2018

3 %

Ocupats en el sector cultural

Variació interanual

T4/2018

-2,4 %

Ocupats en el sector TIC

Variació interanual

T4/2018

-7,9 %

Ocupats en el sector turístic

Variació interanual

T4/2018

-3,5 %

T4/2018



Enquesta de població activa

Enquesta de població activa

Metodologia

- 1. Àmbit poblacional**
- 2. Àmbit temporal**
- 3. Disseny i grandària de la mostra**
- 4. Conceptes poblacionals**
- 5. Conceptes no poblacionals**
- 6. Variables de tabulació**
- 7. Últims canvis de metodologia**
 - 7.1. Canvis introduïts el 2014
 - 7.2. Canvis introduïts el 2005
 - 7.3. Canvis introduïts el 2002
- 8. Explotació de les variables de submostra**

L'estadística de força de treball té l'objectiu de conèixer amb una periodicitat d'alta freqüència la dimensió i la caracterització de la força de treball, fent una explotació específica de l'Enquesta de població activa (EPA, d'ara en endavant). Aquesta és una operació per mostreig de caràcter continu que realitza l'INE amb periodicitat trimestral per tot el territori espanyol. L'Idescat amplia trimestralment els resultats de l'EPA per a l'àmbit de Catalunya.

Enquesta de població activa

↓ Trieu una opció

T=Taula

Resultats bàsics trimestrals i anuals. Catalunya i Espanya

Codi de l'[índex](#) (ex. 5.2.01)

↳ Catalunya

↳ Espanya

Darrer període: 4t trimestre del 2018. Actualització: 29 de gener del 2019.

Resultats detallats trimestrals i anuals. Catalunya

↳ Població

↳ (+) Població total **T**

↳ (+) Població total amb nacionalitat espanyola **T**

↳ Població total en habitatges sense actius **T**

↳ Població total en habitatges sense perceptors d'ingressos **T**

↳ (+) Població de 16 anys i més **T**

↳ (+) Població de 16 anys i més, i altres **T**

↳ (+) Mitjana d'edat de la població de 16 anys i més **T**

↳ (+) Població de 16 anys i més que no està cursant estudis **T**

↳ (+) Mitjana d'anys de formació de la població de 16 anys i més **T**

↳ (+) Mitjana d'anys d'estudi de la població de 16 anys i més **T** (fins al 4t trimestre del 1999)

↳ (+) Població de 16 a 64 anys **T**

↳ Activitat

↳ Ocupació

↳ Ocupació assalariada

↳ Atur

↳ Inactivitat

↳ Població comptada a part (fins al 4t trimestre del 2001)

↳ Llars

Darrer període: 4t trimestre del 2018. Actualització: 29 de gener del 2019.



Enquesta de població activa

Població desocupada

Província

Catalunya. 1r trimestre del 2018

Altres trimestres ▾ ▶

	Valor
Barcelona	322,4
Girona	52,0
Lleida	21,6
Tarragona	62,7
Total	458,7

Unitats: Milers de persones.

Font: Enquesta de població activa, INE.

Nota: Els resultats d'aquesta taula des del 1r trimestre del 2001 en endavant estan afectats per l'aplicació de la nova definició d'atur establerta en el Reglament (CE) núm. 1897/2000 de la Comissió, de 7 de setembre del 2000. Per aquest motiu no són comparables amb els dels períodes anteriors.

Atur registrat

Metodologia

Definicions

Atur registrat: Nombre de persones de 16 anys i més que busquen ocupació, hagin treballat amb anterioritat o no, registrades a les Oficines de Treball de la Generalitat o del Servei Públic d'Ocupació Estatal. Actualment està regulat per l'Ordre ministerial d'11 de març de 1985 (BOE del 14 de març). D'acord amb els criteris formulats en aquesta disposició, en el concepte d'atur registrat s'hi inclouen les persones que tinguin demandes pendents de satisfer, a excepció dels col·lectius següents:

- **Treballadors ocupats:** els qui sol·liciten una feina compatible amb l'actual o per canviar-la per la que tenen, les persones que estan en suspensió o reducció de jornada per un expedient de regulació de treball i els qui estan ocupats en feines de col·laboració social.
- **Treballadors sense disponibilitat immediata per al treball o en situació incompatible amb l'ocupació:** jubilats, pensionistes d'invalidesa absoluta o gran invalidesa, persones de més de 65 anys, treballadors en situació d'incapacitat laboral transitòria o baixa mèdica, estudiants d'ensenyament oficial reglat sempre que tinguin menys de 25 anys o que siguin demandants de la primera ocupació i alumnes de formació professional ocupacional quan les seves hores lectives superin les vint setmanals, tinguin una beca de manutenció i siguin demandants de la primera feina.
- **Treballadors que demanin exclusivament una feina de característiques específiques:** els qui sol·liciten una ocupació a domicili per un període inferior a tres setmanes o una jornada setmanal inferior a vint hores, els qui s'inscriuen a les oficines d'ocupació com a requisit previ per participar en un procés de selecció per a un lloc de treball determinat i els qui sol·liciten una feina exclusivament per a l'estranger.
- **Treballadors eventuais agraris beneficiaris del subsidi especial de desocupació:** els qui l'estan percebent o que, tot i haver-lo exhaurit, no hagi transcorregut un període superior a un any des del dia del naixement del dret.

També se n'exclouen els demandants inscrits com a desocupats que es neguin injustificadament a participar en accions d'inserció o formació professional

Atur registrat

Atur registrat (a 31 de març). Recòmptes 

2018 

Descarregar

Atur registrat (a 31 de març). Recòmptes
Províncies. 2018

☐ Codi	Atur
☒ Litoral	
Barcelona	300.790
Girona	38.820
Lleida	22.245
Tarragona	49.606
Catalunya	411.461

Unitats: Persones.

Font: Departament de Treball, Afers Socials i Famílies.

Atur registrat

Atur registrat (a 31 de març). Recomptes 

2018 

Descarregar

Atur registrat (a 31 de març). Recomptes
Municipis. 2018

☐ Codi ☒ Literal	Atur
Abella de la Conca	9
Abrera	578
Àger	11
Agramunt	249
Aguilar de Segarra	11
Agullana	35
Aiguafreda	126
Barberà de la Conca	13
Barberà del Vallès	1.767
Barcelona	74.239
Baronia de Rialb, la	3
Bàscara	35
Bassella	3

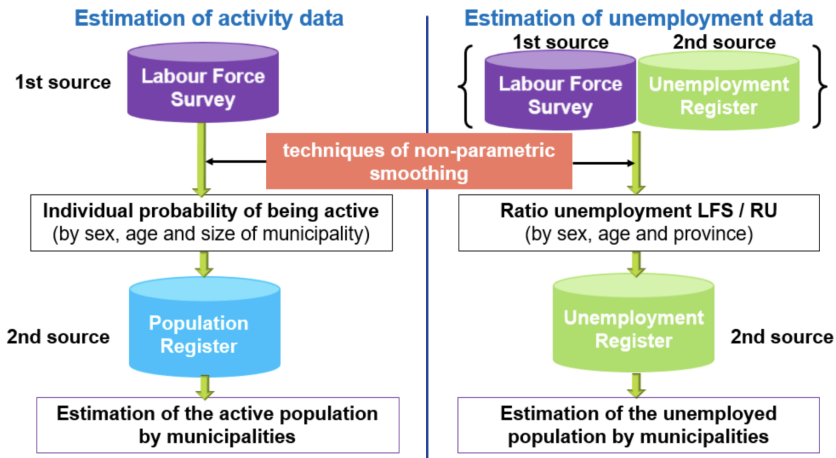
En resum (i simplificant)

Tenim una situació típica de confluència/conflicte de fonts:

- ▶ L'Enquesta de població activa, *Labour Force Survey*
 - ▶ Recull exactament el que volem
 - ▶ Costos alts, grandària mostral moderada, sempre insuficient
 - ▶ Error de mostreig controlable
 - ▶ Errors no-de-mostreig menys controlables
 - ▶ Manca de detall territorial i de subpoblacions
 - ▶ Podem resumir: biaix petit, variabilitat alta
- ▶ L'atur registrat
 - ▶ No mesura el que volem mesurar.
 - ▶ Costos molt baixos, reaprofitament de dades, cobertura censal
 - ▶ No error de mostreig
 - ▶ Error no-de-mostreig petit
 - ▶ Molt detall territorial i de subpoblacions
 - ▶ En resum: biaix gran, variabilitat petita, quasi nul·la
- ▶ EPATERR: Una combinació per baixar l'EPA a àrees petites

Idescat's Register-based Statistics Business and Labour Statistics

TERRITORIAL INDICATORS OF ECONOMICALLY ACTIVE POPULATION



EPATERR: Model activitat

GAM (Model additiu generalitzat) amb resposta binària.

$$\begin{aligned} P(\text{Actiu} = 1 \mid x = \text{Edat}, s = \text{Sex}, e, m, p, t) \\ = \frac{\exp(\alpha_{se} + f_{se}(x) + \beta_p + \delta_m)}{1 + \exp(\alpha_{se} + f_{se}(x) + \beta_p + \delta_m)} \end{aligned}$$

on

$f_{se}(x)$ és un p -spline que recull la relació no lineal de l'edat amb la resposta,

p : província

m : municipi

e : estrat del municipi segons grandària

t : taxa d'atur registrat

$\delta_m \sim N(0, \sigma_m^2)$

Utilitzem un model similar per estimar l'ocupació en cada municipi a partir de les ocupacions EPA a nivell més alt.

Model atur

Model p -spline amb resposta Gaussiana que estima el logaritme de la raó entre el nombre d'aturats EPA i el nombre d'aturats registrats.

Per cada sexe s , edat agregada x , província p ,

$$\log\left(\frac{\text{Atur}_{\text{EPA}}}{\text{Atur}_{\text{REG}}}\right) = \alpha_{sp} + f_{sp}(x) + \epsilon$$

on $\epsilon \sim N(0, \sigma^2)$

Ajust dels models en R

Model activitat:

```
model_act = gam(bb ~ s(EDAT, bs="ps", k=20, by=XX)
  + XX + PROV + s(cmun2,bs="re") + tasa_paro,
  weights=w, data=df2, method="REML",
  family=quasibinomial)
```

Model Ocupació:

```
model_ocupa = gam( oo ~ s(EDAT, bs="ps", k=20, by=XX)
  + XX + PROV + s(cmun2,bs="re"),
  weights=w, data=df2, method="REML",
  family=quasibinomial)
```

Model atur:

```
model_atur = gam(log((epa$Freq)/(regis$Freq))
  ~ s(age2, bs="ps", by=X_atur, k=9)
  + X_atur, type="REML", data=regis)
```

Moltes gràcies

fudina@idescat.cat
frederic.udina@upf.edu

Quines dades?

