

ESTIMACIÓ DE LA VARIÀNCIA MOSTRAL A L'ENQUESTA DE POBLACIÓ ACTIVA

MONTSERRAT GUILLÉN

XAVIER MARTÍN

Universitat de Barcelona i
Institut d'Estadística de Catalunya

L'Enquesta de Població Activa constitueix una de les principals fonts d'informació sobre el mercat de treball. Aquest treball tracta de com incorporar el disseny mostral complex de l'Enquesta en l'estimació dels errors mostrals de resultats referits a Catalunya. En primer lloc, es descriu el mètode d'estimació emprat per l'Instituto Nacional de Estadística. Seguidament, s'introdueixen els dos enfocaments fonamentals per a l'estimació dels errors en la submostra de Catalunya. Per acabar, es comenten alguns resultats sobre el nombre total d'ocupats.

Estimation of sampling variance of the Spanish Labour Force Survey

Keywords: Mostreig de disseny complex, Mercat de treball, Estimació de la variància.

* Aquest article ha estat elaborat dins el projecte **Implementació de programes en l'entorn SAS de fiabilitat i millores en l'estudi metodològic**. Hem d'agrair la col·laboració de Luis Antonio del Barrio Martín, cap de l'Àrea de Estadística de la Actividad, el Empleo y el Paro, i de Florentina Alvarez Alvarez, Subdirectora de Estadísticas Demográficas, de l'INE.

- Article rebut el desembre de 1995.

- Acceptat l'abril de 1996.

1. INTRODUCCIÓ

La principal font d'informació sobre el mercat de treball d'un país són les seves estadístiques de la població activa. Així ho reconeix, per exemple, l'Organització Internacional del Treball (OIT) quan recomana incloure les estadístiques de la població actualment activa i dels seus components en els programes d'estadístiques contínues [1]. En el cas de la Unió Europea, a més a més, aquesta recomanació de l'OIT es converteix en una obligació: els països membres han de dur a terme un cop l'any una enquesta sobre les forces de treball seguint estrictament les directrius metodològiques establertes per l'Eurostat. Cal tenir present que una de les missions que compleix l'enquesta comunitària sobre les forces de treball és la de proporcionar indicadors comparables que permetin repartir entre els països membres els recursos dels fons estructurals (Fons Social Europeu, Fons Europeu de Desenvolupament Regional) [2,3].

A Espanya, l'Enquesta de Població Activa (EPA) és una investigació per mostreig adreçada a les famílies que realitza l'Instituto Nacional de Estadística (INE) al conjunt de l'Estat des de l'any 1964. En l'actualitat, l'Enquesta de Població Activa proporciona resultats trimestrals i és l'EPA corresponent al segon trimestre la que constitueix l'enquesta comunitària sobre les forces de treball. Entre aquests resultats s'inclouen estimacions de les principals magnituds per comunitats autònomes però amb un nivell de detall molt inferior al de les estimacions del conjunt d'Espanya.

L'Institut d'Estadística de Catalunya és l'organisme estadístic oficial de Catalunya. L'interès de disposar d'informació per a l'anàlisi del mercat de treball a Catalunya va obligar l'Institut d'Estadística de Catalunya a escollir entre dues possibilitats: la primera, portar a terme la seva pròpia enquesta sobre les forces de treball; la segona, aprofitar els resultats de l'Enquesta de Població Activa de l'INE. Es va optar per la segona possibilitat atès que, entre d'altres, són criteris de decisió de l'Institut, per un costat, que la metodologia de les operacions estadístiques permeti la comparació dels seus resultats amb altres estadístiques similars, i, per un altre, que no resulti una duplicitat amb altres estadístiques ja existents [4]. A més, el pes específic de Catalunya a la mostra de l'INE abonava la idea que era possible una tabulació més rica de l'Enquesta de Població Activa que la disponible fins a aquell moment [5]. Per emprendre la tasca d'explotació de les cintes de l'EPA, però, calia primer dotar-se d'un instrument de càlcul dels errors de mostreig de les estimacions. Aquest instrument era necessari per tal que el disseny de la tabulació estigués definit, a més de per criteris d'interès temàtic, també per consideracions sobre la fiabilitat de la informació. Fruit d'aquesta necessitat es va desenvolupar la investigació que aquí es presenta.

2. DISSENY MOSTRAL DE L'ENQUESTA

L'Institut Nacional de Estadística utilitza en l'Enquesta de Població Activa un mostreig bietàpic amb estratificació de les unitats de primera etapa.

Les **seccions censals** (que constitueixen les unitats de primera etapa) s'estratifiquen primer segons un criteri geogràfic que té en compte la província i el tipus de municipi al què pertanyen. Després, les seccions censals de cada estrat s'agrupen per formar subestrats segons la seva categoria socio-econòmica.

Els **habitatges familiars** principals i els allotjaments fixos constitueixen les unitats de segona etapa. La selecció d'habitatges dins de cada secció es porta a terme per un procediment sistemàtic amb arrencada aleatòria. A cada habitatge es recull informació de totes les persones que hi tinguin la seva residència habitual.

La distribució de les seccions entre les províncies, els estrats i els subestrats es realitza de la forma següent:

- ① per a les províncies s'utilitza una afixació de compromís entre la proporcional i la uniforme (per a garantir una mínima representació de totes les províncies i per tal que el nombre de seccions sigui un múltiple de 12 —nombre de seccions que pot visitar un enquestador en un trimestre—);
- ② per als estrats d'una província s'utilitza una afixació proporcional a la població lleugerament corregida (s'incrementa el nombre de seccions de determinats estrats en els què s'espera que hi haurà major dispersió en les característiques estudiades);
- ③ entre els subestrats d'un estrat s'utilitza una afixació proporcional a la població.

3. CÀLCUL DELS ERRORS DE MOSTREIG DE L'INE

L'Institut Nacional de Estadística utilitza el mètode de les semimostres reiterades per a calcular els errors de mostreig de l'Enquesta de Població Activa [6]. Aquest mètode estima la variància d'un determinat resultat (indicador de l'error de mostreig) de la forma següent: calcula de forma repetida l'estimació que interessa utilitzant diferents parts (meitats) de la mostra per mesurar, després, la variabilitat d'aquestes estimacions.

L'INE fa 40 reiteracions utilitzant el següent procediment:

- a) Es divideix la mostra original en dues semimostres complementàries fent parelles de seccions a dins de cada estrat.
- b) S'assigna després la primera secció de cada parella a 20 reiteracions aleatòriament i la segona a les 20 restants. D'aquesta forma es garanteix que en totes les reiteracions hi participa la meitat de la mostra total i que cada secció apareix en la meitat de les reiteracions.

4. INFORMACIÓ DISPONIBLE A L'ENQUESTA

L'Institut d'Estadística de Catalunya rep regularment els fitxers trimestrals de l'Enquesta de Població Activa de Catalunya. Aquests fitxers, però, no contenen tota la informació necessària per a poder aplicar el procediment utilitzat per l'INE en l'estimació dels errors de mostreig¹.

En efecte, com s'ha vist en l'apartat anterior, el procediment de l'INE exigeix conèixer la secció i estrat a què pertany cada individu. L'estrat es pot deduir a partir de la província i el factor d'elevació associat a cada registre. La secció, malauradament, no és identificable de la informació continguda en el fitxer, per la qual cosa és impossible definir semimostres basades en parelles de seccions.

Per tal de poder construir un sistema similar a l'utilitzat per l'INE, es va establir provisionalment un criteri alternatiu que consisteix en realitzar parelles de famílies o habitatges.

5. ESTIMACIÓ DE TOTALS I DE LA SEVA VARIÀNCIA

Com a primera aproximació a l'estimació dels errors mostrals (variància i error estàndard) suposarem que l'EPA té un disseny de mostra aleatòria simple. Al següent apartat es presenta la manera de dur a terme aquesta aproximació i, després, la forma d'implementar un mètode de remostreig.

¹En el moment de publicació d'aquest article, aquesta mancança s'haurà probablement solventat i, per tant, es podrà millorar el procediment que es descriu en els propers apartats.

5.1. Sota el supòsit de mostra aleatòria simple

En primer lloc, cal esmentar que la forma de mesurar la fiabilitat dels resultats obtinguts es farà mitjançant l'estimació de la variància del resultat desitjat.

El principal interès de l'explotació de l'Enquesta de Població Activa resideix en proporcionar l'estimació de xifres totals de persones en una determinada situació (el nombre d'ocupats, per exemple). També és habitual voler aquest total desagregat segons les diferents categories d'una variable classificatòria (per exemple, el nombre d'ocupats al sector agrari). Haurem de ser capaços, per tant, de calcular un interaval de confiança o un marge que informi sobre la fiabilitat d'un estimador d'aquestes característiques.

En general, interessa acompanyar la taula d'estimacions de totals de persones en una determinada situació a Catalunya (per exemple, ocupades) classificades segons les categories de dues variables (sexes i sector, per exemple) d'una altra taula de les mateixes dimensions amb les estimacions de l'error associat a cada una de les cel·les de la taula original.

Cada individu de l'enquesta té associat un factor d'elevació que denotarem per w , i la variable estudiada que anomenarem X prendrà el valor 1 si l'individu té les condicions que es desitgen totalitzar o bé 0 si no és així.

En aquesta situació, l'estimació del total és:

$$\hat{X} = \sum_{i=1}^n x_i w_i.$$

El pes w_i és inversament proporcional a la probabilitat d'ésser inclòs a la mostra. L'elevació de cada individu indica el nombre d'individus de la població que representa.

Quan suposem que el mostreig realitzat és aleatori simple, l'expressió utilitzada per calcular la variància del total és:

$$\text{Var}(\hat{X}) = \sum_{i=1}^n \frac{(x_i w_i n - \hat{X})^2}{n(n-1)}.$$

Substituint l'expressió de \hat{X} , la variància es pot escriure com:

$$\text{Var}(\hat{X}) = \sum_{i=1}^n \frac{(x_i w_i n - \sum_{j=1}^n x_j w_j)^2}{n(n-1)},$$

que es pot corregir, finalment, per a tenir en compte la fracció mostrejada. En aquest cas, cal recordar que una bona aproximació per a conèixer el volum total de població

estudiada que intervé en el càlcul és:

$$\hat{N} = \sum_{i=1}^n w_i.$$

Per tant, la fracció de la mostra es pot aproximar per $\frac{n}{\hat{N}}$, i, finalment, l'expressió de la variància esdevé:

$$\widehat{\text{Var}}(\hat{X}) = \left(1 - \frac{n}{\hat{N}}\right) \sum_{i=1}^n \frac{(x_i w_i n - \sum_{j=1}^n x_j w_j)^2}{n(n-1)}.$$

Com que el que es desitja estimar és l'error de mostreig d'un total d'individus que compleixen una condició determinada (per exemple, ser ocupats) la variable d'interès és de tipus dicotòmic (1 - ser ocupat, 0 - no ser-ho) i, llavors, l'expressió es pot reduir a:

$$(1) \quad \widehat{\text{Var}}(\hat{X}) = \left(1 - \frac{n}{\hat{N}}\right) \frac{\sum_{i=1}^{n_1} (x_i w_i n - \sum_{j=1}^{n_1} x_j w_j)^2 + (n - n_1) (\sum_{i=1}^{n_1} x_i w_i)^2}{n(n-1)},$$

on n_1 és el nombre d'individus pels quals X pren el valor 1, i no s'utilitzen els individus pels quals X pren el valor 0.

Per obtenir una referència completa, es pot consultar [7].

5.2. Remostreig per a l'estimació de la variància

L'aproximació al càlcul de les variàncies basada en el supòsit de mostra aleatòria simple és incorrecte, donat que el disseny mostral de l'Enquesta de Població Activa s'allunya d'aquesta situació: es tracta d'un disseny complex. Existeixen dues components bàsiques que així ho determinen; en primer lloc, l'estratificació i, en segon lloc, la selecció de les diferents unitats mostrals en les diverses etapes del mostreig (seccions i famílies). L'entrevista a tots els membres d'una mateixa família és especialment rellevant, ja que cal esperar que les característiques socio-econòmiques dels membres seran iguals i que el comportament individual estarà relacionat, trencant així la independència entre les observacions. És conegut que els efectes d'aquests trets de la mostra són de caire oposat; si bé l'existència d'estrats adequats redueix la variància, l'existència de conglomerats l'augmenta, i és difícil *a priori* determinar la magnitud d'ambdues distorsions.

Wolter [8] presenta diferents mètodes que es poden emprar per a estimar la variància en situacions com la descrita. Els mètodes de remostreig són adequats quan el

nivell de complexitat del disseny (o dels estadístics d'interès) dificulta una aproximació analítica. Entre les diferents alternatives hi ha el mètode jackknife, el bootstrap i les semimostres equilibrades, a part de les expansions en sèrie de Taylor. Els tres primers mètodes es basen en obtenir estimacions a partir de repetir un mostreig de la mostra completa i en analitzar la variabilitat dels seus resultats. El procediment d'expansió en sèrie s'utilitza habitualment quan la principal dificultat es troba en l'estadístic que es vol calcular, el qual s'aproxima linealment, per a calcular la variància de la seva linealització.

En aquest treball ens proposem utilitzar un mètode de remostreig, que reproduueixi el disseny i permeti una aproximació de la variància. A tal efecte, i a la vista de la informació continguda en les dades, podem establir un mètode d'estimació de la variància per semimostres, semblant a l'aplicat per l'INE, on els parells no es formin a nivell de seccions sinó de famílies.

En general, per a poder aplicar el mètode d'estimació de la variància per semimostres, cal dividir la mostra original en dues parts complementàries. Per realitzar l'estimació de la variabilitat de l'estadístic es pren la meitat dels individus de la primera part i l'altra meitat de la segona (l'INE pren la meitat de les seccions de la primera part i l'altra meitat de la segona) i, multiplicant per dos el resultat obtingut, s'aconsegueix una estimació del total que es desitja. Si es reitera aquest procés un cert nombre de vegades en les quals els individus (seccions) se seleccionen aleatòriament, es poden aconseguir diverses estimacions del total i calcular la seva variabilitat.

L'estructura general del disseny mostral de la mostra completa ha de ser incorporada en la formació de les semimostres utilitzades en cada reiteració. Així, es garanteix que la semimostra tingui unes característiques similars a les de la mostra principal.

Per a poder formar semimostres en el cas del segment de dades per Catalunya i tenint en compte la impossibilitat de poder identificar les seccions, s'ha optat per construir les semimostres de la manera més fidel al disseny de la mostra original. Seguint les recomanacions de Wolter [8], i deprés d'analitzar alguns exemples reals de formació de les semimostres, hem resolt dividir les famílies en dos grups diferents dins de cada estrat (és a dir, fer parells de famílies). Per tant, actuem com si es tractés d'un mostreig estratificat de famílies i no de seccions. Com que no disposem de cap criteri de proximitat entre les famílies d'un mateix estrat, establim que els parells es formen de forma seqüencial en el fitxer d'observacions. Els dos grups de la mostra queden determinats quan diem que el primer grup el formen les famílies parells i que el segon grup les senars².

²S'ha previst provar un procediment aleatori de formació de parells de famílies, per tal d'evitar que l'ordre de l'arxiu provoqués biaix en els resultats.

* Per a poder construir les reiteracions, en lloc d'utilitzar una assignació aleatòria dels grups dels estrats a cada reiteració, utilitzem les matrius Hadamard que asseguren que finalment a cada semimuestra es té en compte un dels grups de l'estrat i que cada grup no apareix en més semimostres que el seu complementari (veure [8]). Les matrius Hadamard determinen quin dels dos grups de famílies de l'estrat intervé en la reiteració concreta. La seva utilització recomana prendre tantes reiteracions com el menor múltiple de quatre estrictament superior al nombre d'estrats. Com que el nombre d'estrats en les dades és 26, realitzarem 28 reiteracions.

A fi d'aplicar la metodologia d'estimació de la variància en la situació anterior s'ha elaborat una macro-funció en llenguatge SAS [9,10] suficientment general per ser aprofitada en explotacions similars.

6. RESULTATS

A continuació presentem els resultats obtinguts en l'explotació de les dades de l'Enquesta de Població Activa del primer trimestre de 1990. Es desitja obtenir una taula on aparegui el total estimat d'ocupats a Catalunya per sector d'activitat i sexe. Seguidament, cal donar una estimació del seu error.

Taula 1

*Distribució de freqüències mostrals de la població ocupada per sectors.
1r trimestre de 1990.*

Sector d'activitat	Freqüència	Percentatge
Agricultura	562	7.2
Energia i Aigua	99	1.3
Indústria bàsica	327	4.2
Trans. Metalls	703	9.0
Prod. Aliment.	341	4.4
Tèxtil, Cuiro	500	6.4
Fusta, Paper	476	6.1
Construcció	783	10.0
Comerç, Hosteleria	1617	20.7
Transports, Comunic.	459	5.9
Fin., Asseg. Llog.	462	5.9
A.P. Enseny. i San.	1478	18.9
Total	7807	100.0

Les Taules 1 i 2 mostren les distribucions de freqüències mostrals de la població ocupada per sectors i per sexe.

Taula 2

*Distribució de freqüències mostrals de la població ocupada per sexes.
1r trimestre de 1990.*

Sexe	Freqüència	Percentatge
Homes	5170	66.2
Dones	2637	33.8
Total	7807	100.0

6.1. Estimació del total i de la seva variància segons l'expressió de mostra aleatòria simple

A la Taula 3 es presenten els resultats de l'estimació de totals d'ocupats a Catalunya per sectors i sexe.

Taula 3

*Estimació de total d'ocupats per sectors i sexe.
1r trimestre de 1990.*

Sector d'activitat	Homes	Dones	Total
Agricultura	70832	19221	90053
Energia i Aigua	22394	2957	25351
Indústria bàsica	84430	23383	107813
Trans. Metalls	217623	32837	250460
Prod. Aliment.	67334	20501	87836
Tèxtil, Cuiro	79125	85959	165085
Fusta, Paper	116826	26163	142988
Construcció	180011	6791	186802
Comerç, Hosteleria	241506	175781	417288
Transports, Comunic.	122240	19337	141578
Fin., Asseg. Llog.	85958	435509	129468
A.P. Enseny. i San.	161420	243548	404969
Total	1449701	699989	2149690

Taula 4*Error estàndard relatiu del total d'ocupats. 1r trimestre de 1990.*

Sector d'activitat	Homes	Dones	Total
Agricultura	5.52	10.09	4.83
Energia i Aigua	12.55	35.76	11.84
Indústria bàsica	7.02	13.78	6.24
Trans. Metalls	4.43	11.15	4.10
Prod. Aliment.	7.48	13.24	6.50
Tèxtil, Cuiro	7.47	6.81	5.01
Fusta, Paper	5.87	12.23	5.28
Construcció	4.42	21.55	4.33
Comerç, Hosteleria	3.90	4.45	2.88
Transports, Comunic.	5.75	14.18	5.32
Fin., Asseg. Llog.	6.64	9.48	5.42
A.P. Enseny. i San.	4.81	3.89	2.98
Total	1.45	2.19	1.09

L'error estàndard relatiu es calcula com $E.S.R. = 100 \frac{S}{\hat{X}}$, on S és l'error estàndard estimat i \hat{X} el total estimat.

Taula 5

*Semiamplicitud de l'interval de confiança (95%) del nombre d'ocupats.
1r trimestre de 1990.*

Sector d'activitat	Homes	Dones	Total
Agricultura	7665	3801	8519
Energia i Aigua	5508	2073	5883
Indústria bàsica	11620	6316	13191
Trans. Metalls	18913	7177	20147
Prod. Aliment.	9872	5321	11187
Tèxtil, Cuiro	11587	11479	16213
Fusta, Paper	13450	6273	14793
Construcció	15604	2869	15848
Comerç, Hosteleria	18449	15339	23580
Transports, Comunic.	13783	5374	14756
Fin., Asseg. Llog.	11193	8082	13744
A.P. Enseny. i San.	15217	18590	23643
Total	41125	30058	46108

Taula 6

*Coefficients de variació del total d'ocupats.
1r trimestre de 1990. (Mètode de semimostres.)*

Sector d'activitat	Hòmes	Dones	Total
Agricultura	6.74	9.16	5.52
Energia i Aigua	12.34	18.48	9.91
Indústria bàsica	5.52	12.25	5.85
Trans. Metalls	6.21	17.47	4.70
Prod. Aliment.	8.82	12.11	8.66
Tèxtil, Cuiro	4.41	5.64	3.69
Fusta, Paper	4.12	7.13	3.27
Construcció	4.38	13.90	3.97
Comerç, Hosteleria	5.89	2.98	3.73
Transports, Comunic.	4.73	11.98	4.86
Fin., Asseg. Llog.	3.87	98.47	3.63
A.P. Enseny. i San.	5.39	3.50	3.96
Total	1.12	1.50	0.91

L'error estàndard relatiu es calcula com $E.S.R. = 100 \frac{S}{\bar{X}}$, on S és l'error estàndard estimat i \bar{X} el total estimat.

Taula 7

*Semiamplicitud l'interval de confiança (95%) del nombre d'ocupats.
1r trimestre de 1990. (Mètode de semimostres.)*

Sector d'activitat	Homes	Dones	Total
Agricultura	9360	3451	9748
Energia i Aigua	5416	1071	4926
Indústria bàsica	9130	5614	12369
Trans. Metalls	26493	11245	23091
Prod. Aliment.	11640	4867	14905
Tèxtil, Cuiro	6834	9499	11935
Fusta, Paper	9431	3654	9164
Construcció	15462	1850	14526
Comerç, Hosteleria	27869	10255	30499
Transports, Comunic.	11322	4542	13474
Fin., Asseg. Llog.	6526	7220	9218
A.P. Enseny. i San.	17062	16709	31446
Total	31801	20631	38181

Quan es realitza el tractament com si la mostra fos aleatòria simple, es consideren tots els individus en el mateix estrat.

La Taula 4 mostra els errors estàndard relatius (segons la fórmula clàssica (1)) corresponents als totals estimats de la Taula 3.

Mitjançant la Taula 4 és senzill identificar aquelles caselles on l'error mostral és més gran en relació al nombre d'ocupats estimat. Alternativament es presenta la Taula 5, on es mostra la meitat de l'amplada de l'interval de confiança calculat al 95%.

Sumant i restant la semiamplitud de la Taula 5 al total d'ocupats, obtindríem els límits superior i inferior de l'interval de confiança.

6.2. Estimació de la variància segons el mètode de semimostres reiterades

Per brevetat, reproduïrem aquí els resultats obtinguts corregint la fracció mostrejada³, els corresponents errors estàndard relatius en percentatge i la semiamplitud de l'interval de confiança al 95%.

En general, els errors de mostreig són inferiors quan s'utilitza el mètode de les semimostres, però aquest efecte es deu essencialment a la inclusió del factor de correcció per fracció mostrejada (per més detalls veure [9]).

Amb l'aproximació de les variàncies per remostreig, que té en compte l'estratificació i també l'agrupació dels individus en famílies (encara que no en seccions), el corresponent efecte de correlació (donada la manca d'independència entre els individus d'una mateixa família) provoca estimacions dels errors superiors als derivats de l'expressió de mostreig estratificat. Malgrat que l'expressió de mostra aleatòria simple, que ignora la totalitat del disseny d'Enquesta de Població Activa, proporciona una estimació dels errors conservadora, en alguns casos no és prou acurada. Per exemple, en la casella d'agricultura-homes, el mètode de les semimostres, que incorpora bona part del disseny mostral, proporciona una estimació de l'error mostral superior.

Convé basar les conclusions de fiabilitat sobre els resultats del total d'ocupats en aquesta metodologia d'estimació d'errors per remostreig, donat que incorpora més informació sobre el procés d'obtenció de la mostra. Cal esperar que el procediment que s'ha presentat es realitzi més acuradament quan es pugui tenir en compte la informació sobre seccions censals i estratificació.

³Aquesta correcció per fracció mostrejada redueix lleugerament la variància estimada, i en conseqüència els errors estàndard i els errors estàndard relatius

REFERÈNCIES

- [1] **Mehran, F. i Hussmanns, R.** (1990). *Surveys of economically active population, employment, unemployment and underemployment*. An ILO manual on concepts and methods, Oficina Internacional del Treball, Ginebra.
- [2] **Eurostat** (1992). *Enquête sur les forces de travail. Méthodes et définitions*. Oficina Estadística de les Comunitats Europees, Luxemburg.
- [3] **Comissió de les Comunitats Europees** (1988). *L'enquête sur les forces de travail comme instrument de la politique de l'emploi*, CECA-CEE-CEEA, Luxemburg.
- [4] **Llei 30/1991, de 13 de desembre, Pla estadístic de Catalunya 1992-1995**. Diari Oficial de la Generalitat de Catalunya núm. 1539 de 10 de gener de 1992.
- [5] **Institut d'Estadística de Catalunya** (diversos anys). *Mercat de treball. Ampliació de resultats anuals de l'Enquesta de Població Activa*, Barcelona.
- [6] **INE** (1996). *Encuesta de Población Activa. Informe Técnico*. Area de Diseño de Muestras y Evaluación de Encuestas de Población y Hogares. Madrid.
- [7] **Cochran, W.** (1977). *Sampling Techniques*. John Wiley.
- [8] **Wolter, K.M.** (1986). *Introduction to variance estimation*. Springer-Verlag.
- [9] **Martín, X. i Guillén, M.** (1992). *Estimación de totales y su error de muestreo en la Encuesta de Población Activa. Document de Treball*. Institut d'Estadística de Catalunya.
- [10] **SAS Institute** (1990). *SAS/IML Software. Usage and Reference Version 6. First edition*. SAS Institute. Cary, NC.

ENGLISH SUMMARY:

ESTIMATION OF SAMPLING VARIANCE OF THE SPANISH LABOUR FORCE SURVEY

Montserrat Guillén and Xavier Martín

The Enquesta de Població Activa is one of the main sources of information about the labour market. This article discusses how to incorporate the complex sample design of this survey in the estimation of sampling errors for results referred to Catalonia.

The first section introduces the reader to the context of active population surveys. The role of the *Institut d'Estadística de Catalunya* in producing reliable statistical results within this framework is emphasized.

In the second section, a description of the sample design used to obtain the *Enquesta de Població Activa* is detailed. It is a two-step design with stratification. Furthermore, once a household is selected in the sample, all the individuals in the household are interviewed.

Section 3 explains how the *Instituto Nacional de Estadística* estimates the variance for the results obtained from the latter survey. A half-sample method is used.

Section 4 summarizes the type of information that is received in the files corresponding to the *Enquesta de Població Activa de Catalunya*, which do not allow to use the variance estimation method that is reproduced in section 3.

In section 5, the general theoretical framework for the estimation of totals and their variance is introduced. A resampling approach is also described, and the article explains how to apply the half-sampling method to the subsample which is restricted to Catalonia.

Finally, section 6 shows some of the results obtained using the methodology introduced in the previous section and some comments about the estimation of the total working population are presented. The discussion includes a comparison of the variance estimates under two different hypothesis; namely, the simple random sample approach and the half-samples method that takes into account the sample design.

A list of related references is also included at the end.