



Generalitat de Catalunya
Institut d'Estadística de Catalunya

idescat

VI Jornada MET-MEIO 2013

Avaluació de models de risc de revelació estadística en microdades

Institució: Institut d'Estadística de Catalunya (Idescat)
Tutor Idescat: Mònica Gasulla
Professor: Enric Ripoll
Dades contacte: Via Laietana, 58
08003 Barcelona
Tel. 93 557 30 06
mgasulla@idescat.cat

El risc de revelació estadística és una mesura de la probabilitat d'identificació de les dades confidencials individuals per part d'un tercer, mitjançant algun fitxer o taula que qualsevol organisme ha difós amb finalitats estadístiques (o d'altres). En la literatura actual i d'acord amb les millors pràctiques que apliquen les agències estadístiques, s'han proposat diverses mesures tant del risc individual com del risc global, així com diversos mètodes per estimar-les.

Objectiu:

Avaluar i comparar les mesures que segueixen a continuació, aplicant un ventall molt ampli de tècniques estadístiques en un nou camp que pren cada dia més rellevància.

- ▶ Mètodes basats en la població:
 - Models d'inferència freqüentista i bayesiana basats en les distribucions Poisson, Binomial i Binomial Negativa.
 - Models Log-lineals.
 - Mètodes heurístics basats en el submostreig de les dades i detecció d'outliers especials.
- ▶ Mètodes basats en la mostra:
 - Regla del valor frontera o k-anonimat per a variables categòriques.
 - Estratègies de detecció d'outliers, tècniques 'clúster' aglomeratives i de densitat per a variables contínues.
- ▶ Mètodes derivats d'experiments de "record linkage" (enllaç de registres):
 - Basat en distàncies.
 - Probabilístic.

Requisits:

Coneixements d'estadística oficial, mostreig, inferència estadística, inferència bayesiana, anàlisi multivariant i ús de paquets estadístics (R).

Avaluació de l'error mostral de l'Enquesta de població activa (EPA)

Institució: Institut d'Estadística de Catalunya (Idescat)
Tutora Idescat: Núria Bové
Professor: Enric Ripoll
Dades contacte: Via Laietana, 58
08003 Barcelona
Tel. 93 557 30 06
nbove@idescat.cat

L'Enquesta de població activa (EPA) és la principal font d'informació estadística del mercat de treball i té per objectiu conèixer la dimensió i la caracterització de la força de treball. L'EPA és una operació estadística per mostreig dirigida a les llars, de periodicitat trimestral i de caràcter continu. L'Idescat realitza una explotació estadística trimestral detallada per a l'àmbit de Catalunya.

Objectiu:

Estudiar les tècniques de remostreig aplicables a l'EPA i escollir-ne la millor per tal de calcular els errors de mostreig associats a les principals magnituds del mercat de treball que publica l'Idescat. Alguns exemples de resultats que difon l'Idescat de l'Enquesta de població activa i que es tractaria de calcular els errors de mostreig són:

- ▶ Taxes d'activitat, ocupació i atur segons principals variables sociodemogràfiques
- ▶ Població activa, ocupada i desocupada segons sexe, edat, nivell de formació assolit i nacionalitat, entre altres
- ▶ Població activa i ocupada segons sectors i branques d'activitat
- ▶ Població ocupada i assalariada segons característiques de l'ocupació

Requisits:

Coneixements d'estadística oficial, mostreig, inferència estadística i nivell alt de programació en paquets estadístics (SAS o R).