

Informe tècnic: Comparació entre estimadors ponderats i no ponderats

Enquesta d'Inserció Laboral 2023. AQU Catalunya

Pedro Delicado (Universitat Politècnica de Catalunya)

AQU Catalunya té interès en avaluar la possibilitat d'eliminar les ponderacions de les unitats mostrals a les estimacions derivades de l'Enquesta d'Inserció Laboral (EIL). Entre d'altres motivacions, destaca l'informe tècnic [1], que l'any 2021 la *Higher Education Statistics Agency* (HESA, UK) va encarregar a l'*Institute for Social and Economic Research*, (ISER, University of Essex) per a avaluar la possibilitat de no ponderar les dades del *Graduate Outcome Survey* (GOS).

1 Informe ISER: ponderacions al GOS de l'HESA

El GOS de la HESA és, pel seu disseny, un **estudi censal** de la població d'interès. A la pràctica, però, hi ha una proporció d'enquestats que no responen el qüestionari. Aquestes no-respostes podrien implicar que la mostra finalment observada no representés bé la població en estudi. Per tal de corregir aquest *biaix per no-resposta*, l'HESA venia fent servir un sistema de ponderacions per tal d'assignar a cada unitat mostral el nombre total d'unitats poblacionals a les quals representa.

Les ponderacions venen donades per l'invers de l'estimació de les probabilitats de respondre a l'enquesta. Si el fet de respondre o no a l'enquesta és totalment independent de la variable d'interès de l'enquesta (*missing completely at random*, *MCAR*) aleshores, en una població de mida N , la probabilitat de respondre a una enquesta de mida n és n/N per a totes les unitats poblacionals. Com a conseqüència, el pes de totes les unitats mostrals serà el mateix i valdrà N/n . És a dir, suposant *MCAR* no cal ponderar els individus que responen a un qüestionari censal.

Quan el fet de respondre a l'enquesta pot dependre d'algunes característiques conegudes administrativament per a totes les unitats poblacionals, és possible ajustar models estadístics que estimin les probabilitats de respondre a l'enquesta. Es fan servir diferents models de regressió amb resposta binària (l'individu respon a l'enquesta o no ho fa), ajustats a tota la població (per exemple, regressió logística). Si, a més a més, es pot suposar que la resposta a la variable d'interès és independent del mecanisme de no-resposta quan es condiciona per les variables administratives conegudes a tota la població (*missing at random*, *MAR*), aleshores la ponderació que s'ha de donar a cadascuna de les unitats mostrals és l'invers de la probabilitat de respondre, estimada per algun d'aquests models estadístics. És això precisament el que l'HESA venia fent amb el GOS.

Els experts de l'ISER, al seu informe [1], van analitzar el GOS referit als graduats el curs 2018-2019 i van comparar els resultats quan no es feia servir cap ponderació amb els obtinguts quan es ponderava d'acord a les probabilitats de resposta estimades per 4 models diferents. Per a l'estimació de la mitjana poblacional de dues variables d'interès, es va calcular l'error quadràtic mitjà (EQM) dels cinc estimadors resultants, i la magnitud de la diferència entre l'estimador no-ponderat i els 4 estimadors ponderats (el que podria entendre's com a biaix per la no-resposta de l'estimador no ponderat). Es van considerar estimacions per a diferents sub-poblacions (diferents sexes, diferents universitats, diferents ensenyaments, etc.). Es va emprar el següent criteri per a jutjar un sistema de ponderacions: aquest es va considerar beneficiós quan implicava una reducció

d'EQM en més del 50% de les estimacions fetes (combinant alguna de les variables d'interès amb alguna de les diferents subpoblacions).

La principal conclusió de l'informe de l'ISER [1] va ser que les ponderacions aconseguien reduir l'EQM només en una minoria de les estimacions, i que la magnitud de les reduccions eren molt petites. Per tant, en aquella edició del GOS no havia necessitat de ponderar perquè no es van trobar diferències substancials entre la precisió dels estimadors ponderats i no ponderats.

2 Càlculs amb les dades de l'EIL-2023

L'EIL de l'AQU té una diferència fonamental respecte del GOS britànic: **l'EIL no és un estudi censal**, sinó que es tracta d'un mostreig estratificat (els estrats es defineixen creuant *centres d'ensenyament* i *tipus d'ensenyament*).

Al mostreig estratificat les unitats mostrals poden tenir un pes diferent quan les fraccions mostrals són diferents en diferents estrats. Si, per disseny, a l'estrat A s'entrevista a 1 de cada 4 individus, mentre que a l'estrat B s'entrevista a 1 de cada 2, és lògic que el pes de les unitats mostrals de l'estrat A sigui el doble del pes de les unitats mostrals de l'estrat B (mireu, per exemple, el primer paràgraf de la secció 1.3 de l'informe de l'ISER per a l'HESA). Si, a més de mostreig estratificat, podem trobar no-respostes, les ponderacions de les unitats mostrals poden ajudar també a corregir el biaix per no-resposta.

Al mostreig estratificat hi ha un cas on no és necessari ponderar les unitats mostrals:

- Fraccions mostrals constants, és a dir, mida mostral a cada estrat proporcional a la mida poblacional de l'estrat.
- Procés de no-resposta independent de l'estrat.
- Procés de no-resposta completament aleatori (MCAR).

Cal esmentar, però, que aquest no és l'únic escenari on els estimadors ponderats i no-ponderats podrien donar lloc a resultats estadísticament semblants.

Les ponderacions que l'AQU fa servir per a analitzar els resultats de l'EIL no tenen en compte cap model de regressió amb resposta binària (resposta/no-resposta) per estimar la probabilitat de no-resposta a cada unitat poblacional. Per contra, les ponderacions de l'AQU són simplement els inversos de les proporcions de respostes obtingudes a cada estrat. Aquestes ponderacions són correctes pels següent escenari de mostreig estratificat:

- Fraccions mostrals no necessàriament constants, és a dir, mida mostral a cada estrat no necessàriament proporcional a la mida poblacional de l'estrat.
- Procés de no-resposta que pot dependre de l'estrat.
- Procés de no-resposta completament aleatori (MCAR).

Per avaluar si a l'EIL de 2023 la ponderació de les unitats mostrals té influència en els resultats, hem replicat l'anàlisi fet per l'ISER al seu informe [1] sobre el GOS per a l'HESA. En particular, hem seguit aquest informe per a **calcular els EQMs d'estimadors de mitjanes i proporcions** i hem fet servir el **criteri de reducció de l'EQM en al menys el 50% de les estimacions** considerades per tal de qualificar com a beneficis un sistema de ponderacions.

Per tal d'avaluar si les diferències entre els valors estimats amb i sense ponderació són rellevants o no, calculem **l'interval de confiança 95% pel valor esperat de la**

diferència entre els dos estimadors. Quan les dues formes d'estimar (amb i sense ponderació) són semblants (i per tant no cal ponderar) s'espera que aquest interval contengui el valor 0.

Com a maneres addicionals, menys formals, d'avaluar si les diferències entre els valors estimats amb i sense ponderació són rellevants o no, hem definit els següents indicadors:

Diferència relativa entre els dos estimadors: Diferència entre els dos estimadors (en valor absolut) dividida entre la longitud de la intersecció dels intervals de confiança 95% construïts a partir dels dos estimadors.

Proporció de solapament entre intervals de confiança: Longitud de la intersecció dels intervals de confiança 95% construïts a partir dels dos estimadors, dividida entre la longitud de la unió dels dos intervals.

Quan les dues formes d'estimar (amb i sense ponderació) són semblants (i per tant no cal ponderar) s'obté un valor de la diferència relativa entre els dos estimadors proper a 0, i un valor de la proporció de solapament entre intervals de confiança proper a 1.

Aquesta ha estat la nostra manera de dur a terme l'estudi:

- Hem considerat 3 paràmetres d'interès:
 - Proporció d'ocupats.
 - Proporció de titulats que repetiria els mateixos estudis al mateix centre.
 - Mitjana de la valoració dels continguts teòrics rebuts (en una escala de 0 a 10).
- Hem considerat per separat els estudis de grau i de màster.
- Hem considerat les següents divisions de les poblacions de grau i de màster:
 - Població total.
 - Població dividida per sexes (2 subpoblacions).
 - Divisió per centres d'ensenyament (65 subpoblacions pels graus i 56 subpoblacions pels màsters).
 - Divisió per tipus d'ensenyament (115 ensenyaments pels graus i 48 subàmbits detallats pels màsters)

En total hi ha $3 \times ((1 + 2 + 65 + 115) + (1 + 2 + 56 + 48)) = 870$ estimacions on comparem els resultats dels estimadors amb i sense ponderacions.

Aquests són els fets més rellevants:

1. El mostreig estratificat és pràcticament proporcional als graus però no als màsters. Això podria tenir implicacions en el paper de les ponderacions en un i altre tipus d'estudis.
2. Per a la població total, aquests són els resultat numèrics dels tres indicadors de la magnitud de l'efecte de la ponderació:

	Graus			Màsters		
	Prop. d'ocupats	Prop rep. estudis	Mitjana sat. teor.	Prop. d'ocupats	Prop rep. estudis	Mitjana sat. teor.
$EQM_w - EQM_{uw}$	$1.16 \cdot 10^{-7}$	$-6.11 \cdot 10^{-8}$	-0.000207	$4.29 \cdot 10^{-7}$	$1.61 \cdot 10^{-6}$	-0.001360
$0 \in IC_{0.95}(\text{dif.estims})?$	SÍ	SÍ	NO	SÍ	SÍ	NO
Dif. relat. estims.	0.0158	0.0531	0.4325	0.0746	0.0739	1.4527
Prop. solap. ICs	0.9694	0.9041	0.5362	0.8702	0.8712	0.2561

Es segueix el següent:

- (a) L'efecte de la ponderació sobre l'EQM és petit i no sempre en favor de la ponderació.
- (b) L'efecte de la ponderació és menor per a graus que per a màsters.
- (c) L'efecte de la ponderació és menor quan s'estimen proporcions que quan s'estimen mitjanes.
- (d) La diferència dels estimadors amb i sense ponderació és significativament diferent de 0 per a les mitjanes de satisfacció (tant per a graus com per a màsters) però no pas per a les proporcions d'ocupació o de la potencial repetició dels estudis.
- (e) Altres indicadors menys formals:
 - La diferència relativa dels estimadors amb i sense ponderació és petita (excepte en el cas de la mitjana de satisfacció pels màsters).
 - La proporció de solapament dels intervals de confiança 95% és gran (excepte en el cas de la mitjana de satisfacció pels màsters).

3. De la comparació de les 870 estimacions es conclou el següent:

- (a) **L'EQM de l'estimador ponderat és menor que el de l'estimador no ponderat en el 39.54% dels casos**, són iguals en el 20.12%, i és major en el 40.34%.
- (b) **En el 77.93% dels casos, les estimacions amb i sense ponderacions no són significativament diferents, si treballem amb confiança 95%**. Aquest percentatge puja al 79.78% pel grau i baixa al 74.77% pels màsters.
- (c) La diferència relativa dels estimadors amb i sense ponderacions, definida abans, és menor o igual que 0.0458 en el 75% de les estimacions.
- (d) La proporció de solapament entre els intervals de confiança 95% construïts a partir dels estimadors amb i sense ponderacions, és major o igual que 0.9144 en el 75% de les estimacions.
- (e) Aquí també s'observa que l'efecte de la ponderació és menor per a graus que per a màsters.

3 Conclusions i recomanacions

Les conclusions de l'anàlisi fet amb les dades de l'EIL de 2023 sobre l'efecte de les ponderacions en els resultats finals són coincidents amb les que l'ISER reporta pel cas del GOS britànic:

- Les ponderacions redueixen l'EQM en una minoria dels casos considerats (en un 39.54%).
- La magnitud de les diferències entre els estimadors amb i sense ponderacions són petites en la majoria de les situacions. Concretament, en el 77.93% dels casos les estimacions amb i sense ponderacions no són significativament diferents.
- Addicionalment, en el 75% de les estimacions, els intervals de confiança 95% construïts a partir dels estimadors amb i sense ponderacions tenen un grau de solapament de com a mínim el 0.9144.

Es conclou doncs que a l'EIL de l'AQU no es perd qualitat si els resultats es reporten sense fer servir ponderacions.

Es recomana que es tinguin en compte les següents observacions i recomanacions:

- L'estudi s'ha fet amb unes dades concretes (EIL de 2023) i només tres paràmetres d'interès. Aleshores les conclusions de l'estudi podrien canviar si s'analitzessin altres onades de l'EIL o s'estimessin altres paràmetres. Es recomana repetir l'anàlisi feta a aquest informe per a futures edicions de l'EIL.
- S'ha constatat que l'efecte de la ponderació és més acusat per a màsters que per a graus. Una possible causa d'aquesta diferència podria ser la falta de proporcionalitat en l'assignació d'unitats mostrals als diferents estrats al mostreig en el cas dels màsters. Es recomana que s'augmenti el grau de proporcionalitat en el mostreig estratificat dels màsters per a futures edicions de l'EIL.

Referències

- [1] Lynn, P. and C. Xena (2021) Exploration of statistical weighting in the Graduate Outcomes Survey: technical report. *Institute for Social and Economic Research, University of Essex*.