

## Cómo citar:

Hernández Coscollà, Carles y Hernández Dobón, Francesc J. (2024): "La correlació de la desigualtat social i la desigualtat educativa en les comunitats i ciutats autònomes". *Arxius de Ciències Socials* 49, pp. 59-86 <https://doi.org/10.7203/acs.49.27567>

# LA CORRELACIÓ DE LA DESIGUALTAT SOCIAL I LA DESIGUALTAT EDUCATIVA EN LES COMUNITATS I CIUTATS AUTÒNOMES

CARLES HERNÁNDEZ<sup>1</sup> I FRANCESC J. HERNÁNDEZ<sup>2</sup>

## RESUM

AQUEST ARTICLE FORMULA UN INDICADOR DE DESIGUALTAT EDUCATIVA ESTANDARDITZAT A LES MESURES HABITUALS, EL CALCULA PER A LES COMUNITATS I CIUTATS AUTÒNOMES ESPANYOLES I MOSTRA LES SIGNIFICATIVES CORRELACIONS QUE S'HI PODEN ESTABLIR. TAMBÉ APROFUNDEIX EN L'ÚS DE CORBES LOGARÍTMQUES EN L'EXPRESSIÓ DELS RENDIMENTS DE L'ENSENYAMENT AMB LES DADES DE LES PROVES DIAGNÒSTIQUES INTERNACIONALS. FA SERVIR UN PROGRAMA DE MINERIA DE DADES (IA) PER A ELABORAR UN MODEL.

## PARAULES CLAU

DESIGUALTAT EDUCATIVA, DESIGUALTAT SOCIAL, ÍNDEX DE GINI, FÓRMULA DE BOLTZMANN, INTEL·LIGÈNCIA ARTIFICIAL

<sup>1</sup> Aprentell. <https://orcid.org/0009-0005-7753-6238>

<sup>2</sup> Professor jubilat UV. <https://orcid.org/0000-0001-5229-2998>

## A B S T R A C T

THIS ARTICLE FORMULATES AN INDICATOR OF EDUCATIONAL INEQUALITY STANDARDIZED TO THE USUAL MEASURES, CALCULATES IT FOR SPANISH AUTONOMOUS COMMUNITIES AND CITIES AND SHOWS THE SIGNIFICANT CORRELATIONS THAT CAN BE ESTABLISHED. IT ALSO DELVES INTO THE USE OF LOGARITHMIC CURVES IN THE EXPRESSION OF TEACHING PERFORMANCE WITH DATA FROM INTERNATIONAL DIAGNOSTIC TESTS. IT USES A DATA MINING (AI) PROGRAM TO DEVELOP A MODEL.

## KEYWORDS

EDUCATIONAL INEQUALITY, SOCIAL INEQUALITY, GINI INDEX, BOLTZMANN FORMULA, ARTIFICIAL INTELLIGENCE

## 1. Introducció

La desigualtat social és un assumpte clàssic de la sociologia. Podríem dir, fins i tot, que aquesta disciplina es formulà a partir de la sospita que els ideals proclamats per la Revolució Francesa (llibertat, igualtat i fraternitat) no s'havien acomplert. Precisament perquè s'havia enunciat aquest ideal en l'autorepresentació del moviment revolucionari, la noció de desigualtat heretà un principi normatiu, que permet distingir-la de la mera constatació descriptiva de diferències. Aquesta càrrega normativa és la que està subjacent a nocions com ara classe (a diferència d'estrat), gènere (per contraposició a sexe) o ètnia (a diferència de l'acientífica raça).

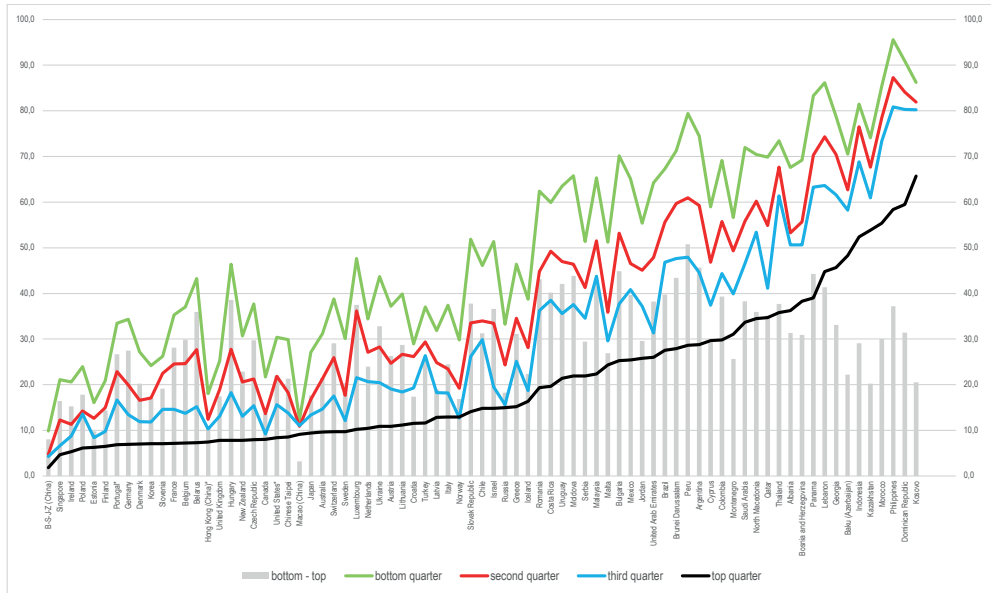
Les desigualtats educatives són un tipus de desigualtats socials. Hi ha una àmplia bibliografia que relaciona les diferències en el rendiment educatiu amb les altres desigualtats. De fet, aquest és el tema de la sociologia de l'educació, tal i com fou reformulada a partir del anys 60/70 del segle passat. En són fites destacades els estudis alemanys sobre la mobilitat social i la seua relació amb l'educació, vegeu, per exemple, es seminaris d'Adorno dels anys 60 (Adorno 2021, vol. II: 229-282), l'Informe Coleman sobre rendiments diferenciats per classe social i ètnia als EUA (una avaluació cinquanta anys després en: RASE 2016) o el llibre de Bourdieu i Passeron sobre la reproducció, amb dades de França (una consideració del seu llegat en: RASE 2022). Val a dir, Coleman o Bourdieu establiren que els *inputs* socials correlacionaven amb els *outputs* educatius. A partir d'aquestes recerques se n'han fet moltíssimes més, en les quals es relacionen classes, gèneres o ètnies amb rendiments diferents de l'ensenyament. Per exemple, la bibliografia produïda pel Grup de Recerca 13 de la Federació Espanyola de Sociologia es tan àmplia que no podríem citar-la ací.

A més, a hores d'ara el programa PISA aporta moltíssimes dades. A tall d'exemple, es poden veure els percentatges de baix rendiment segons els països en la prova de competència lectora de 2018, desagregats pels quartils d'índex socioeconòmic, que es presenten en el gràfic 1.

Els països estan ordenats pel baix rendiment del quartil superior. Com s'hi pot veure el baix rendiment del quartil superior en índex socioeconòmic (altrament dit, els fills i les filles de la capa social més rica) *sempre*

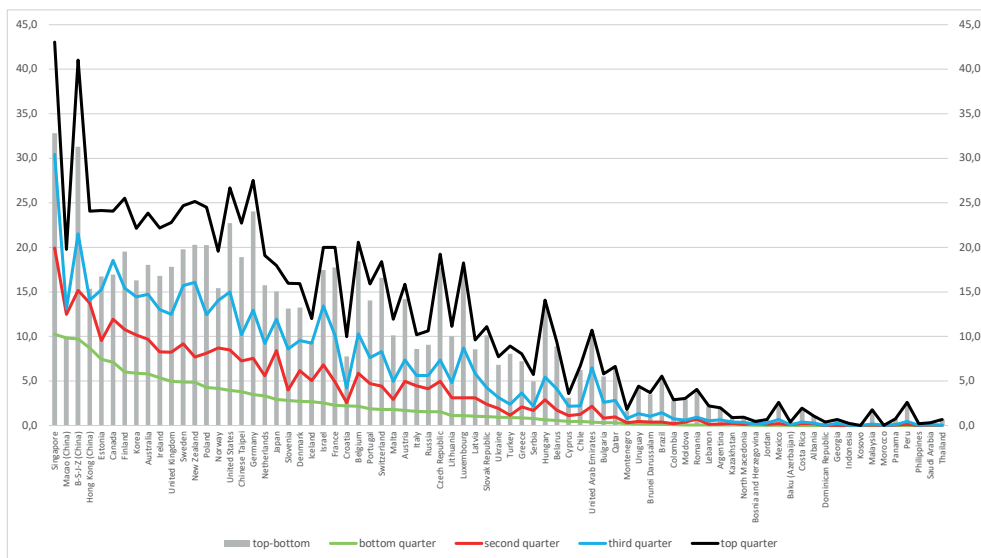
és inferior al baix rendiment del tercer quartil, que, al mateix temps, és inferior al del segon, i aquest al del quartil inferior en índex socioeconòmic (altrament dit, de la capa social més pobre). Les barres de color gris representen la diferència de baix rendiment entre el quartil inferior i el superior. La variable contrària, l'alt rendiment, presenta, com es podria suposar, un comportament invers, com acredita el gràfic 2.

GRÀFIC 1. PERCENTATGE DE BAIX RENDIMENT PISA 2018 EN COMPETÈNCIA LECTORA, SEGONS ÍNDEX SOCIOECONÒMIC.



Font: Elaboració pròpia PISA data base. Espanya fou descartada per problemes de mostreig.

GRÀFIC 2. PERCENTATGE D'ALT RENDIMENT PISA 2018 EN COMPETÈNCIA LECTORA, SEGONS ÍNDEX SOCIOECONÒMIC.



Font: Elaboració pròpia PISA data base. En *reading*, Espanya fou descartada per problemes de mostreig.

Ara bé, al nostre parer la correlació entre desigualtat de rendiment i factors socials, a partir del càlcul estadístic presenta dues limitacions. La primera és que realment no tracta de la desigualtat educativa *en ella mateixa*, sinó que estableix correlacions amb mesures de rendiment i, per tant, només pot arribar a identificar *factors* (com ara, Calero i Waisgrais, 2009, també amb dades de PISA). Val a dir, no disposa d'una *mesura* de la desigualtat educativa *per se*, com la que s'ha establert, per exemple, pel que fa a la desigualtat social. La conseqüència és que tampoc pot establir un principi normatiu. Val a dir, il·lustrar la diferència, però no establir la desigualtat (de manera anàloga a com un repertori de diferències salarials, etc., per països no és una mesura de la desigualtat).

Per tal de poder fonamentar que es tracta de *desigualtats*, el que farem en aquest article és precisament *adaptar* la manera habitual de mesurar les desigualtats socioeconòmiques als rendiments educatius, a continuació, aplicar-lo a les dades de PISA, i, posteriorment, relacionar aquesta mesura adaptada amb la mesura habitual de desigualtat. Després, avançarem en una modelització, fent servir la Intel·ligència Artificial. Amb aquest procediment no només argumentarem el que, com ja hem indicat és conegut (que la pobresa és un factor de baix rendiment), sinó que avançarem en la determinació de la desigualtat educativa *mateixa*. D'aquesta manera podem formular les hipòtesis de l'article.

## 2. Hipòtesi

- 1a) És possible adaptar l'Índex de Gini com a mesura de la desigualtat (social) educativa.
- 2a) La desigualtat educativa correlacionarà positivament amb la desigualtat social.

En aquestes hipòtesis cal considerar que la mesura del rendiment de l'ensenyament es pot realitzar pel percentatge de baix rendiment (o d'alt rendiment) o per la puntuació general.

### 3. Adaptació de l'Índex de Gini a la desigualtat (social) educativa.

Com és conegut, la desigualtat social es mesura per mitjà de dos procediments: la comparació de quantils i l'índex de Gini.

La comparació de quantils es pot realitzar de dues maneres. La primera és amb la comparació de les mitjanes dels quantils. Així, per exemple, la Unió Europea mesura la desigualtat com la proporció dels ingressos mitjans del quantil superior i l'inferior, allò que s'anomena indicador S80/S20 (Eurostat 2023).

Resulta trivial que és equivalent aquesta comparació si fem servir la mitjana d'ingressos o el conjunt dels ingressos de cada quantil. Si ens referim a les mitjanes, que podem representar precisament com a S20 (mitjana d'ingressos del primer quantil, els que estan a l'esquerra del 20% de la mostra ordenada si la representàrem en un eix) i com a S80 (mitjana d'ingressos del cinquè quantil, val a dir, la submostra més enllà del punt corresponent al 80%), la desigualtat  $\delta$  serà:

$$\delta = \frac{S80}{S20}$$

Si per a una població de  $N$  individus, ens referim al total dels ingressos per a cada quantil i anomenem T20 al total del quantil inferior i T80 al total del quantil superior, aleshores la proporció (serà  $\delta'$ ):

$$\delta' = \frac{T80}{T20}$$

Si dividim el numerador i el denominador per la població de cada quantil, val a dir  $\frac{N}{5}$ , aleshores:

$$\delta' = \frac{\frac{T80}{\frac{N}{5}}}{\frac{T20}{\frac{N}{5}}} = \frac{S80}{S20}$$

Per tant, resulta trivial que  $\delta = \delta'$ . Per això, parlarem de mitjanes d'ingressos encara que també podríem considerar els ingressos totals.

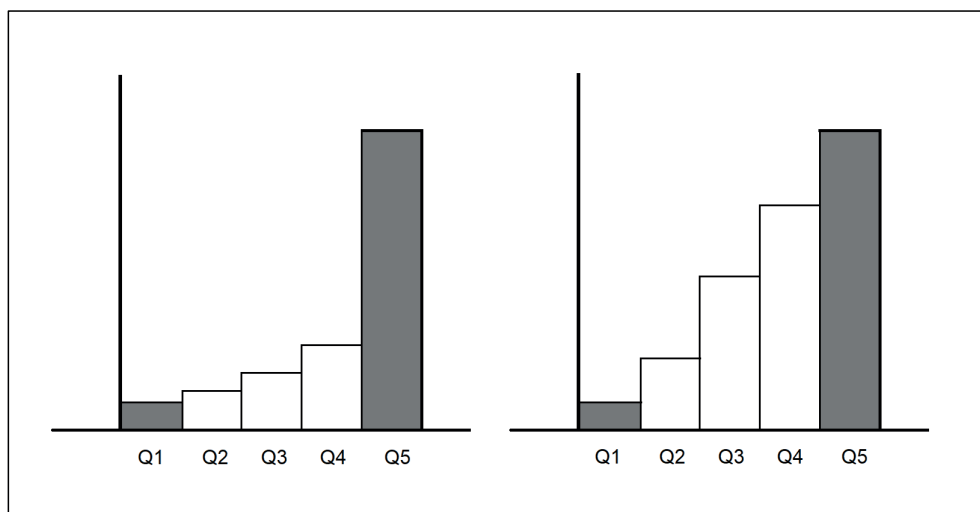
Altres indicadors de desigualtat, com ara els emprats per l'OCDE (OECD 2023) fan servir no només la comparació dels ingressos o les mitjanes d'un quantil, sinó el punt de tall entre un decil i altre, allò que podríem dir els interdecils. Així s'han formulat la relació entre els interdecils del decil superior i l'inferior, el que es podria formular com: P90/P10 (és clar que el punt de tall del decil superior és el percentil 90 i així successivament). la relació entre l'interdecil superior i el mitjà, val a dir, el valor de la mitjana, que seria: P90/P50, i el de la mitjana amb l'interdecil inferior, és a dir: P50/P10. OCDE també calcula S90/S40,

altrament dit la proporció entre els ingressos del quintil superior i els dels dos quintils inferiors (o les seues mitjanes, tant s'hi val, com hem explicat adés), el que s'anomena índex de Palma).

La comparació de quantils, bé es considere  $S$  o bé es considere  $P$ , o fins i tot les altres possibilitats mostrades, es susceptible de rebre dos crítiques, que ens semblen molt importants.

En primer lloc, els procediment bandeja altres quantils (per exemple,  $S_{80}/S_{20}$  desconsidera el 60% de la població, és a dir, els quintils 2n, 3r i 4t). Per això, un mateix valor de  $\delta$  pot correspondre a situacions ben diferents, com les que es representen en el gràfic 3.

GRÀFIC 3. SOCIETATS AMB LA MATEIXA DESIGUALTAT ( $S_{80}/S_{20}$ ), PERÒ DE COMPOSICIÓ BEN DIFERENT

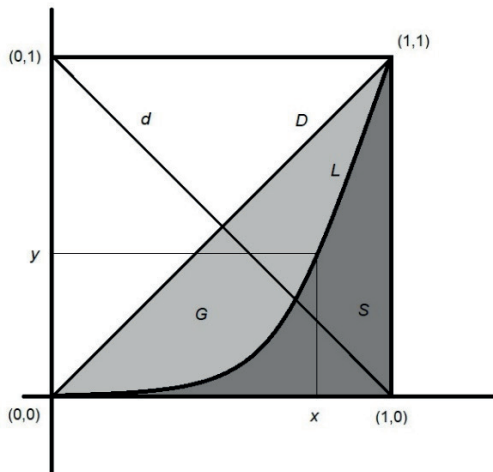


En segon lloc, la comparació de quantils ignora el fenomen dels mega-rics (Piketty 2013; també Piketty 2023). Aquestes dues crítiques no són aplicables a l'Índex de Gini.

La premissa bàsica de l'Índex de Gini (que pren el seu nom de l'economista italià Corado Gini) és que, en una societat plenament igualitària, la representació en els eixos cartesianes del percentatge acumulat de la població (en l'eix de les abscisses) i del percentatge acumulat d'un recurs econòmic —com ara els ingressos— (en l'eix de les ordenades) [el que per tant esdevé en el que anomenarem un quadrat unitat perquè és el quadrat que té una unitat de costat, ja que està definit pels punts: (0, 0), (1, 0), (1, 1) i (0, 1)], se superposaria a la diagonal creixent, la que uneix els punts (0, 0) i (1, 1). És fàcil d'entendre-ho: en una societat igualitària el 10% dels individus tindrien el 10% dels recursos; el 20% dels individus, el 20% dels recursos, etc. En una societat no igualitària, la representació dels punts no és la diagonal [la línia que uneix els punts (0, 0) i (1, 1)], sinó una corba (l'anomenada corba de Lorenz) per baix d'ella, però que coincideix amb la diagonal necessàriament en els punts (0, 0), (perquè a 0% de població acumulada els recursos són necessàriament el 0%) i pel punt (1, 1) (perquè el 100% dels recursos acumulats correspon al 100% de la població). Com menys igualitària (o més desigualitària) siga la societat, més s'allunya la línia de la diagonal creixent, i a l'inrevés. La proporció entre la superfície tancada per la diagonal i la corba per baix de la diagonal (que en el gràfic posterior anomenem  $\delta$ ) i tota la superfície per baix de la diagonal (que mesura  $\frac{1}{2}$  unitats gràfiques), propor-

ciona un coeficient, que multiplicat per 100 dona l'Índex de Gini. Ho podem expressar matemàticament de la manera següent. Si partim del gràfic 4,

GRÀFIC 4. REPRESENTACIÓ DEL CÀLCUL DE L'ÍNDEX DE GINI



Definirem l'Índex de Gini ( $IG$ ), a partir del Coeficient de Gini ( $CG$ ), de manera que:

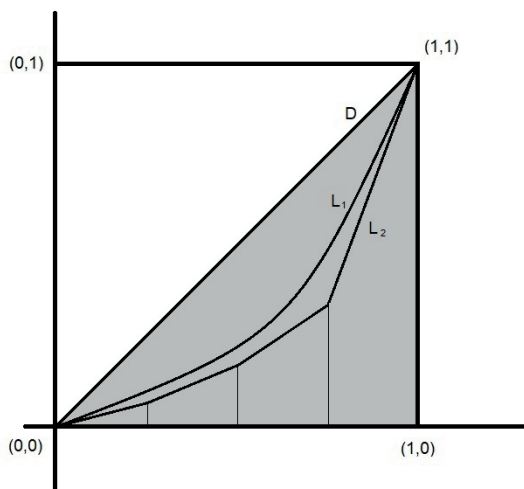
$$IG = 100 CG$$

$$CG = \frac{G}{G + S} = \frac{G}{\frac{1}{2}} = 2G$$

$$G = \frac{1}{2} - S$$

L'última fórmula és necessària perquè el més fàcil és calcular  $S$  i, a partir del seu valor, establir el de  $G$ . En el cas de disposar del valor de quantils (quintils, decils, percentils), el càlcul ha de contemplar la superfície que determinen aquests, com mostra el gràfic 5.

GRÀFIC 5. REPRESENTACIÓ DEL CÀLCUL DE L'ÍNDEX DE GINI AMB QUANTILS



En aquest cas, el càlcul de  $S$  és més senzill, i per tant, també el de  $G$ .

#### 4. Càlcul de la desigualtat educativa amb les dades PISA

Disposem de les puntuacions de PISA per a les comunitats i ciutats autònomes espanyoles de la sèrie del 2018, en les matèries de matemàtiques i ciències (la comprensió lectora es descartà per problemes de mostreig en les proves espanyoles), segons determinats percentils, més concretament els valors dels interdecils: P5, P10, P25, P50, P75, P90, P95 (OECD 2019, 257 i 259).

Per a poder aplicar la metodologia de l'Índex de Gini ( $IG$ ), que mesura la desigualtat social, a aquestes dades sobre rendiments de l'ensenyament, necessitem, en primer lloc fer una extrapolació de les dades corresponents a P0 i P100, la qual cosa és relativament senzilla, fent servir una línia de tendència amb els valors dels percentils disponibles.

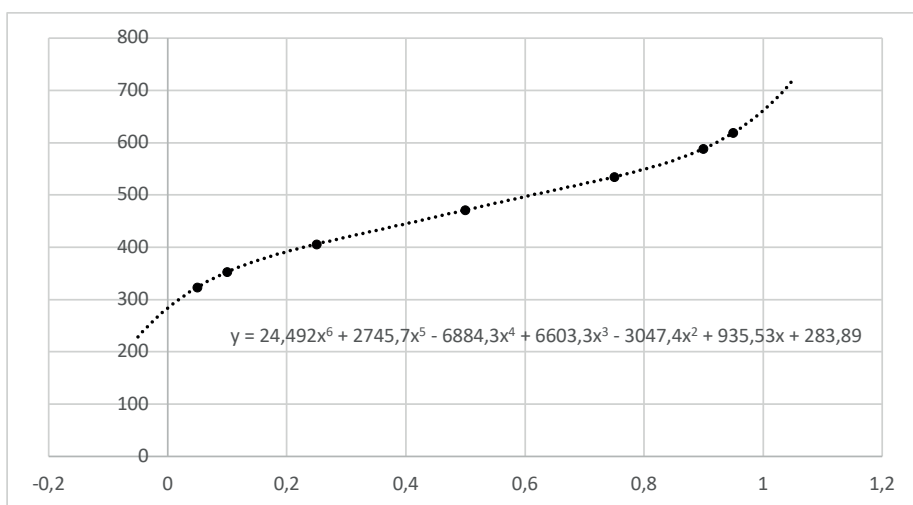
De les diverses línies de tendència disponibles, hem emprat la polinòmica d'ordre 6, que presenta la forma  $y=ax^6+bx^5+cx^4+dx^3+ex^2+fx+g$ , de manera que resulta trivial que en el cas de P0  $y=g$  (ja que si  $x=0$ , aleshores s'anul·len tots els sumands on hi ha una  $x$  de la fórmula:  $ax^6+bx^5+cx^4+dx^3+ex^2+fx+g$ ) i  $P100=a+b+c+d+e+f+g$  (ja que si  $x=1$ , aleshores  $x$  elevat a qualsevol valor és igualment 1 i la fórmula  $ax^6+bx^5+cx^4+dx^3+ex^2+fx+g$  se simplifica com:  $a+b+c+d+e+f+g$ ). Per exemple, en el cas de la primera fila de la taula posterior (la que serà la taula 1), tenim:

	P5	P10	P25	P50	P75	P90	P95
Andalusia	323,83	352,91	406,28	471,13	534,83	588,82	619,15

Es representen aquestes dades en un gràfic (vegeu gràfic 6).



GRÀFIC 6. REPRESENTACIÓ DELS PUNTS AMB LÍNIA DE TENDÈNCIA POLINÒMICA



El que proporciona el valor de P100 ( $24,492 + 2745,7 - 6884,3 + 6603,3 - 3047,4 + 935,53 + 283,89 = 661,21$ ) i el valor de P0 (283,89).

Amb les dades PISA i l'extrapolació esmentada, arribem a les taules següents<sup>3</sup>, per a ciències (taula 1) i matemàtiques (taula 2):

TAULA 1. PUNTUACIONS CIÈNCIES (PISA 2018 SEGONS ALGUNS PERCENTILS)

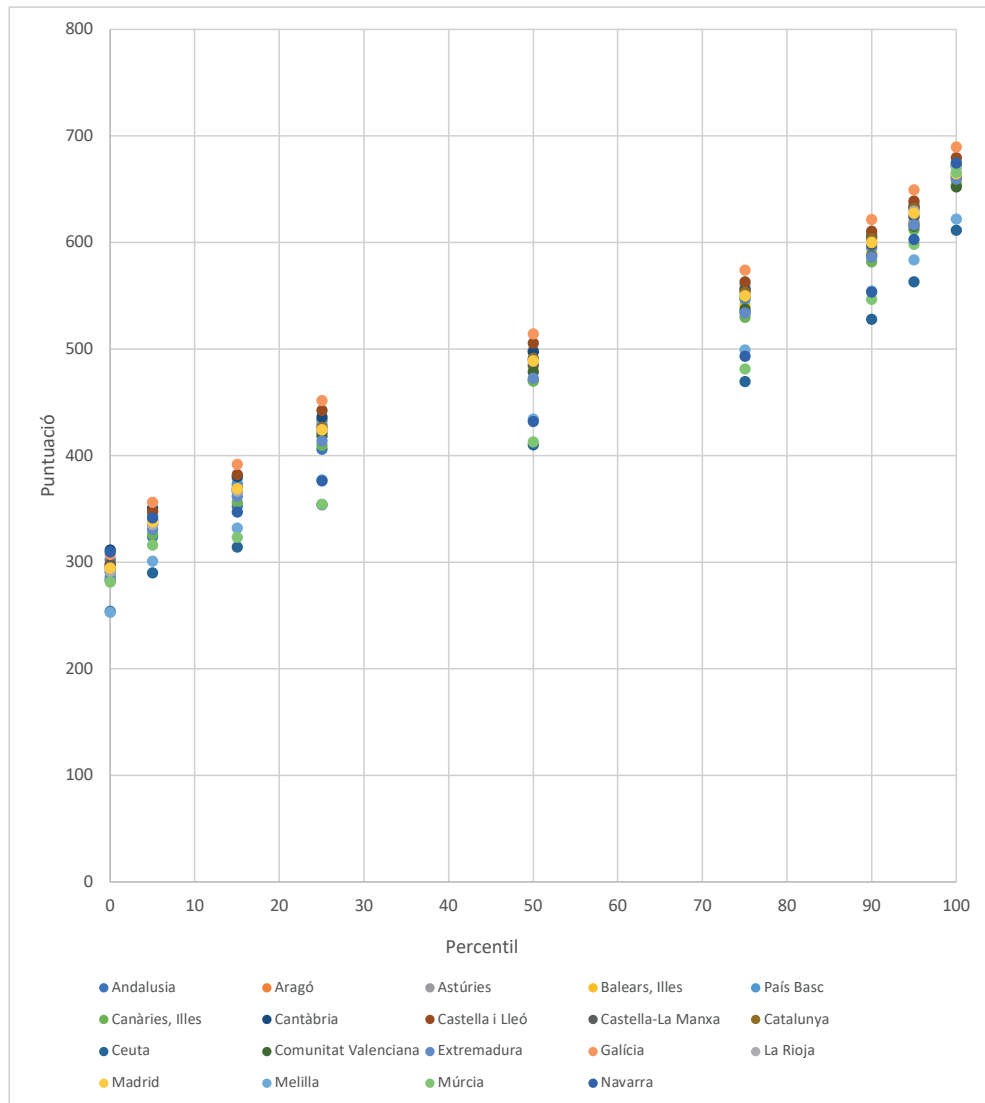
	P0	P5	P10	P25	P50	P75	P90	P95	P100
Andalusia	283,89	323,83	352,91	406,28	471,13	534,83	588,82	619,15	661,21
Aragó	297,10	340,15	372,61	433,44	497,95	556,52	606,46	634,57	673,68
Astúries	303,04	347,68	378,83	432,91	499,57	560,53	608,68	636,41	675,66
Balears, Illes	295,04	338,87	370,65	425,24	482,76	542,07	590,57	618,28	657,86
País Basc	302,54	343,52	373,79	429,08	490,29	547,02	595,26	623,59	664,04
Canàries, Illes	283,17	326,13	356,66	409,66	470,37	530	582,26	612,28	654,15
Cantàbria	311,66	351,19	380,81	436,15	497,65	556,11	605,53	633,26	671,95
Castella i Lleó	298,70	347,62	382,7	442,88	505,59	563,24	610,99	639,27	679,79
Castella-La Manxa	297,57	339,42	369,2	421,6	485,44	548,98	598,14	625,02	662,10
Catalunya	294,16	336,71	368,45	426,94	491,74	553,62	603,66	631,95	672,03
Ceuta	254,04	290,32	314,36	354,03	410,32	469,58	528,34	563,41	611,75
Comunitat Valenciana	290,47	335,56	366,61	418,45	478,95	537,81	587,69	615,25	652,57
Extremadura	286,79	330,98	362,02	414,04	473,02	534	586,56	616,92	659,78
Galícia	307,68	356,67	391,86	451,86	514,53	574,15	621,58	649,52	689,91
La Rioja	291,71	335,23	367,12	424,63	489,22	550,13	600,14	629,32	671,02
Madrid	294,80	338,3	369,39	424,68	489,06	550,12	600,39	627,72	664,84
Melilla	253,18	301,26	332,15	377,51	434,63	499,41	554,75	584,03	622,28
Múrcia	281,70	316,24	323,56	354,7	413,05	481,75	547,05	598,26	666,16
Navarra	309,80	341,87	347,45	376,55	432,08	493,6	553,91	603,23	674,88

Font: OECD PISA 2018 Taula I.B2.11 i elaboració pròpia (P0 i P100)

<sup>3</sup> Els càlculs estan realitzats amb més decimals que els reproduïts a les taules.

Representarem aquestes dades en el gràfic 7.

GRÀFIC 7. REPRESENTACIÓ DE LES COMUNITATS I CIUTATS AUTÒNOMES



TAULA 2. PUNTUACIONS MATEMÀTIQUES (PISA 2018 SEGONS ALGUNS PERCENTILS)

	P0	P5	P10	P25	P50	P75	P90	P95	P100
<b>Andalusia</b>	264,18	317,66	352,93	407,57	469,64	528,80	580,46	608,59	645,58
<b>Aragó</b>	297,63	343,17	376,75	437,17	501,37	560,06	606,97	634,05	672,61
<b>Astúries</b>	287,50	339,58	374,81	430,73	493,59	554,07	602,15	629,32	666,90
<b>Balears, Illes</b>	295,32	339,95	371,40	425,68	486,67	541,44	587,44	613,91	650,42
<b>País Basc</b>	300,88	350,28	384,98	443,33	505,00	559,66	604,37	630,47	667,52
<b>Canàries, Illes</b>	283,03	321,13	348,80	399,79	461,16	520,02	571,21	599,59	638,31
<b>Cantàbria</b>	305,12	353,60	387,38	443,84	503,29	557,55	604,76	630,94	666,62
<b>Castella i Lleó</b>	296,50	348,79	385,09	444,81	507,84	565,90	611,73	637,10	671,40
<b>Castella-La Manxa</b>	289,68	331,86	362,06	416,32	482,43	543,89	590,91	617,12	653,63
<b>Catalunya</b>	290,23	339,99	374,04	429,54	492,48	551,88	601,26	629,91	670,23
<b>Ceuta</b>	243,69	281,77	307,78	351,64	407,30	467,36	524,98	556,66	598,90
<b>Comunitat Valenciana</b>	283,99	332,74	365,49	417,21	474,73	532,07	581,48	607,98	642,49
<b>Extremadura</b>	273,98	322,92	356,73	412,05	472,11	529,99	578,15	605,21	642,52
<b>Galícia</b>	290,53	344,00	381,18	442,33	504,43	559,88	605,60	631,09	665,73
<b>La Rioja</b>	290,56	338,46	374,00	437,39	501,94	561,90	612,65	643,38	688,38
<b>Madrid</b>	285,52	333,53	367,37	425,29	489,78	549,08	597,89	625,84	664,71
<b>Melilla</b>	268,80	299,03	322,64	370,82	427,59	491,90	548,06	572,06	599,43
<b>Múrcia</b>	267,29	318,78	353,11	409,53	477,38	539,13	591,63	618,14	652,09
<b>Navarra</b>	310,97	354,90	386,95	444,24	506,08	563,74	612,37	639,95	678,49

Font: OECD PISA 2018 Taula I.B2.10 i elaboració pròpia (P0 i P100)

Explicarem el procediment per a calcular els corresponents Índex de Desigualtat Educativa, que abreviarem, per tractar-se d'un Índex de Gini adaptat:  $IG^*$ , per a cada matèria de PISA. Mostrarem el càlcul en el cas de ciències i estalviarem el de matemàtiques, que és idèntic.

En primer lloc, calculem el valor de cada rectangle, que lògicament és la marca de classe (la diferència entre els valors dels percentils dividida per 2) multiplicada per la base (0,05, 0,05, 0,15, 0,25, etc.). Els resultats s'arreglegen en la taula 3. Hi afegim la suma total.

TAULA 3. RECTANGLES CIÈNCIES

	P0:P05	P5:P10	P10:P25	P25:P50	P50:P75	P75:P90	P90:P95	P95:P100	Suma
Andalusia	15,19	16,92	56,94	109,68	125,75	84,27	30,20	32,01	470,95
Aragó	15,93	17,82	60,45	116,42	131,81	87,22	31,03	32,71	493,39
Astúries	16,27	18,16	60,88	116,56	132,51	87,69	31,13	32,80	496,00
Balears, Illes	15,85	17,74	59,69	113,50	128,10	84,95	30,22	31,90	481,95
País Basc	16,15	17,93	60,22	114,92	129,66	85,67	30,47	32,19	487,22
Canàries, Illes	15,23	17,07	57,47	110,00	125,05	83,42	29,86	31,66	469,77
Cantàbria	16,57	18,30	61,27	116,73	131,72	87,12	30,97	32,63	495,31
Castella i Lleó	16,16	18,26	61,92	118,56	133,60	88,07	31,26	32,98	500,80
Castella-La Manxa	15,92	17,72	59,31	113,38	129,30	86,03	30,58	32,18	484,42
Catalunya	15,77	17,63	59,65	114,84	130,67	86,80	30,89	32,60	488,85
Ceuta	13,61	15,12	50,13	95,54	109,99	74,84	27,29	29,38	415,90
Comunitat Valenciana	15,65	17,55	58,88	112,18	127,10	84,41	30,07	31,70	477,54
Extremadura	15,44	17,33	58,20	110,88	125,88	84,04	30,09	31,92	473,78
Galícia	16,61	18,71	63,28	120,80	136,09	89,68	31,78	33,49	510,43
La Rioja	15,67	17,56	59,38	114,23	129,92	86,27	30,74	32,51	486,28
Madrid	15,83	17,69	59,56	114,22	129,90	86,29	30,70	32,31	486,50
Melilla	13,86	15,84	53,22	101,52	116,76	79,06	28,47	30,16	438,88
Múrcia	14,95	16,00	50,87	95,97	111,85	77,16	28,63	31,61	427,04
Navarra	16,29	17,23	54,30	101,08	115,71	78,56	28,93	31,95	444,06

Font: Elaboració pròpia

A continuació es calcula, el valor acumulat de cada quartil i la proporció que representen respecte del total, tal i com apareix en la taula 4.

TAULA 4. PERCENTATGES ACUMULATS

	Q1	Q2	Q3	Q4	Ac Q1	Ac Q1:Q2	Ac Q1:Q3	Ac Q1:Q4
<b>Andalusia</b>	89,05	198,73	324,47	470,95	18,9%	42,2%	68,9%	100,0%
<b>Aragó</b>	94,20	210,63	342,44	493,39	19,1%	42,7%	69,4%	100,0%
<b>Astúries</b>	95,31	211,87	344,38	496,00	19,2%	42,7%	69,4%	100,0%
<b>Balears, Illes</b>	93,28	206,78	334,88	481,95	19,4%	42,9%	69,5%	100,0%
<b>País Basc</b>	94,30	209,22	338,88	487,22	19,4%	42,9%	69,6%	100,0%
<b>Canàries, Illes</b>	89,78	199,78	324,83	469,77	19,1%	42,5%	69,1%	100,0%
<b>Cantàbria</b>	96,14	212,87	344,59	495,31	19,4%	43,0%	69,6%	100,0%
<b>Castella i Lleó</b>	96,33	214,89	348,50	500,80	19,2%	42,9%	69,6%	100,0%
<b>Castella-La Manxa</b>	92,95	206,33	335,63	484,42	19,2%	42,6%	69,3%	100,0%
<b>Catalunya</b>	93,06	207,89	338,56	488,85	19,0%	42,5%	69,3%	100,0%
<b>Ceuta</b>	78,86	174,40	284,39	415,90	19,0%	41,9%	68,4%	100,0%
<b>Comunitat Valenciana</b>	92,08	204,26	331,35	477,54	19,3%	42,8%	69,4%	100,0%
<b>Extremadura</b>	90,97	201,86	327,73	473,78	19,2%	42,6%	69,2%	100,0%
<b>Galícia</b>	98,60	219,40	355,48	510,43	19,3%	43,0%	69,6%	100,0%
<b>La Rioja</b>	92,61	206,84	336,76	486,28	19,0%	42,5%	69,3%	100,0%
<b>Madrid</b>	93,08	207,29	337,19	486,50	19,1%	42,6%	69,3%	100,0%
<b>Melilla</b>	82,92	184,44	301,19	438,88	18,9%	42,0%	68,6%	100,0%
<b>Múrcia</b>	81,81	177,78	289,63	427,04	19,2%	41,6%	67,8%	100,0%
<b>Navarra</b>	87,82	188,90	304,61	444,06	19,8%	42,5%	68,6%	100,0%

Font: Elaboració pròpia

A continuació es calculen les àrees dels rectangles corresponents a les proporcions acumulades (fent servir novament les marques de classe i considerant que cada rectangle té com a base 0,25). Es fa la suma corresponent i es calcula el Coeficient i finalment l'Índex. Aquest és el contingut de la taula 5.

TAULA 5. ÀREES, COEFICIENT I ÍNDEX

	Àrea Q1	Àrea Q2	Àrea Q3	Àrea Q4	Suma	CG*	IG*
<b>Andalusia</b>	0,024	0,076	0,139	0,211	0,450	0,100	9,999
<b>Aragó</b>	0,024	0,077	0,140	0,212	0,453	0,094	9,406
<b>Astúries</b>	0,024	0,077	0,140	0,212	0,453	0,093	9,318
<b>Balears, Illes</b>	0,024	0,078	0,140	0,212	0,454	0,091	9,129
<b>País Basc</b>	0,024	0,078	0,141	0,212	0,455	0,091	9,074
<b>Canàries, Illes</b>	0,024	0,077	0,140	0,211	0,452	0,096	9,608
<b>Cantàbria</b>	0,024	0,078	0,141	0,212	0,455	0,090	9,021
<b>Castella i Lleó</b>	0,024	0,078	0,141	0,212	0,454	0,091	9,133
<b>Castella-La Manxa</b>	0,024	0,077	0,140	0,212	0,453	0,095	9,467
<b>Catalunya</b>	0,024	0,077	0,140	0,212	0,452	0,096	9,590
<b>Ceuta</b>	0,024	0,076	0,138	0,210	0,448	0,104	10,365
<b>Comunitat Valenciana</b>	0,024	0,078	0,140	0,212	0,454	0,093	9,277
<b>Extremadura</b>	0,024	0,077	0,140	0,211	0,452	0,095	9,509
<b>Galícia</b>	0,024	0,078	0,141	0,212	0,455	0,090	9,027
<b>La Rioja</b>	0,024	0,077	0,140	0,212	0,452	0,096	9,583
<b>Madrid</b>	0,024	0,077	0,140	0,212	0,453	0,095	9,474
<b>Melilla</b>	0,024	0,076	0,138	0,211	0,449	0,102	10,227
<b>Múrcia</b>	0,024	0,076	0,137	0,210	0,447	0,107	10,693
<b>Navarra</b>	0,025	0,078	0,139	0,211	0,452	0,095	9,542

Font: Elaboració pròpia. En les cel·les de la dreta de la primera fila de dades pot comprovar-se que hem fet els càlculs amb totes els decimals disponibles i no només amb les dades arrodonides de les taules.

Després de fer el càlcul idèntic per a les dades de matemàtiques de la taula 2, es pot arribar a la següent taula de síntesi (taula 6), on s'han afegit les dades de l'Índex de Gini que proporciona l'INE (2023), per a les comunitats i ciutats autònomes en l'última sèrie disponible (2020, per a l'any anterior) (l'INE no proporciona l'Índex de Gini de Navarra).

Hem d'advertir que els valors de  $IG^*$  són notablement *més reduïts* que els de la desigualtat econòmica. La raó és ben senzilla: té a veure amb l'escala convencional de PISA. Així, per exemple, mentre que la desigualtat de l' $IG$  de les comunitats i ciutats autònomes es mou en la forqueta 29,7-42,4, això correspon a una comparació de quintils (S80/S20) que està en 5,6 per al conjunt del país (última dada proporcionada per INE), però la comparació entre decils en el cas dels resultats PISA de ciències (allò que diríem S90/S10), es mou *només* entre 1,82 i 1,97, per la qual cosa el seu  $IG^*$  és menor (penseu que, per un mer efecte estadístic, si aplicàrem S90/S10 a les dades econòmiques, el valor hauria de ser encara superior al 5,6. Ara bé, aquesta és una qüestió menor, que no afecta a les correlacions, que és el moll de l'os de la nostra argumentació, perquè el coeficient de correlació de Pearson no està afectat per l'escala de les variables.

Els resultats generals s'arreglen en la taula 6.

TAULA 6. RESULTATS GENERALS

	IG	IG* Ciències	Puntuació general ciències	IG* Matemàtiques	Puntuació general matemàtiques
Andalusia	33,0	9,999	471	30,436	467
Aragó	29,7	9,406	493	28,822	497
Astúries	30,3	9,318	496	29,111	491
Balears, Illes	32,8	9,129	482	27,920	483
País Basc	31,3	9,074	487	27,405	499
Canàries, Illes	33,8	9,608	470	30,019	460
Cantàbria	30,1	9,021	495	27,017	499
Castella i Lleó	30,2	9,133	501	28,140	502
Castella-La Manxa	31,2	9,467	484	29,964	479
Catalunya	32,3	9,590	489	29,089	490
Ceuta	42,3	10,365	415	32,770	411
C. Valenciana	32,6	9,277	478	28,429	473
Extremadura	31,1	9,509	473	29,434	470
Galícia	29,7	9,027	510	27,993	498
La Rioja	30,7	9,583	487	29,728	497
Madrid	35,9	9,474	487	29,617	486
Melilla	42,4	10,227	439	31,811	432
Múrcia	31,7	10,693	479	31,417	474
Navarra	n.d.	9,542	492	27,812	503

Font: INE 2023 i elaboració pròpia.

La taula anterior proporciona la següent matriu de correlacions, fent servir el coeficient de correlació de Pearson ( $R$ ).

TAULA 7. MATRIU DE CORRELACIONS

	IG	IG* Ciències	Puntuació general ciències	IG* Matemàtiques	Puntuació general matemàtiques
IG	1,000	0,609	-0,902	0,732	-0,891
IG* Ciències		1,000	-0,703	0,902	-0,703
Puntuació general ciències			1,000	-0,808	0,964
IG* Matemàtiques				1,000	-0,842
Puntuació general matemàtiques					1,000

Font: Elaboració pròpia de les taules anteriors. Com es conegut, les matrius són simètriques, per la qual cosa no cal omplir totes les cel·les.

## 5. Conclusions parcials

Podem donar per verificades les dues hipòtesis (epígraf 2).

1a) Hem adaptat l'Índex de Gini com a mesura de la desigualtat (social) educativa, el que permet formular un Índex de Desigualtat Educativa amb resultats pertinents.

2a) La desigualtat educativa correlacionarà positivament amb la desigualtat social, amb valors del coeficient de correlació de Pearson per a les comunitats i ciutats autònomes, les matèries i els anys de referència de  $R=0,732$  (matemàtiques) i  $R=0,609$  (ciències). Això representa coeficients de determinitat de  $R^2=0,535$  (matemàtiques) i  $R^2=0,370$  (ciències).

A més, hem obtingut dues conclusions més. Una que podríem considerar esperable: hi ha una alta correlació entre la puntuació general en les dues matèries (matemàtiques i ciències), amb  $R=0,964$ . Fins i tot es podria considerar esperable l'alta correlació entre desigualtat en una disciplina i en l'altra, amb  $R=0,902$ . S'hi pot veure una correlació inversa entre rendiment (puntuació general) i els dos indicadors de desigualtat, el social i l'educatiu. De manera que la puntuació general de matemàtiques correlaciona de manera elevada i inversa amb la desigualtat que presenten les puntuacions (segons el quantils indicats), amb  $R=-0,842$  en matemàtiques i  $R=-0,703$  en ciències.

Verificades les hipòtesis avançarem en l'enunciat de dos corol·laris.

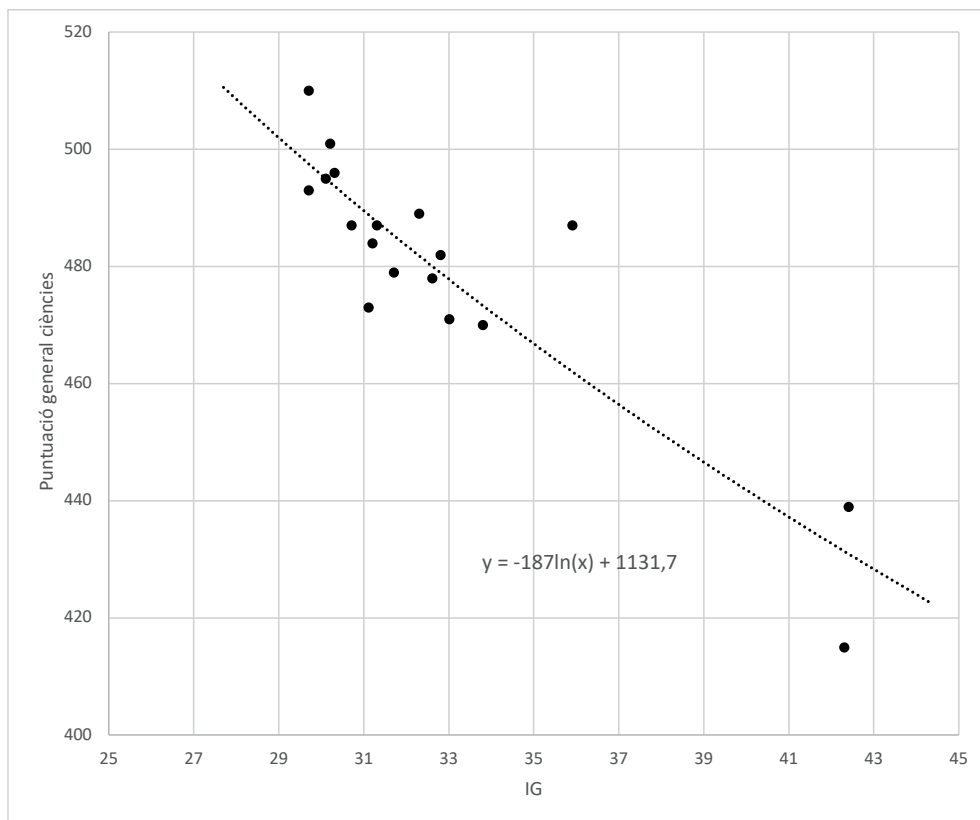
## 6. Aproximació a la modelització de la mesura del rendiment de l'ensenyament amb corbes logarítmiques (corol·lari 1)

A partir de les conclusions anteriors, podríem pensar que la *desigualtat social es podria considerar un predictor de la puntuació general de PISA i també de la desigualtat educativa*.

Partim de la taula 6 i representem els resultats de desigualtat social i puntuació general de ciències en el gràfic 8 i de matemàtiques en el gràfic 9. Farem servir línies de tendència logarítmiques i posarem les corresponents equacions en els gràfics.

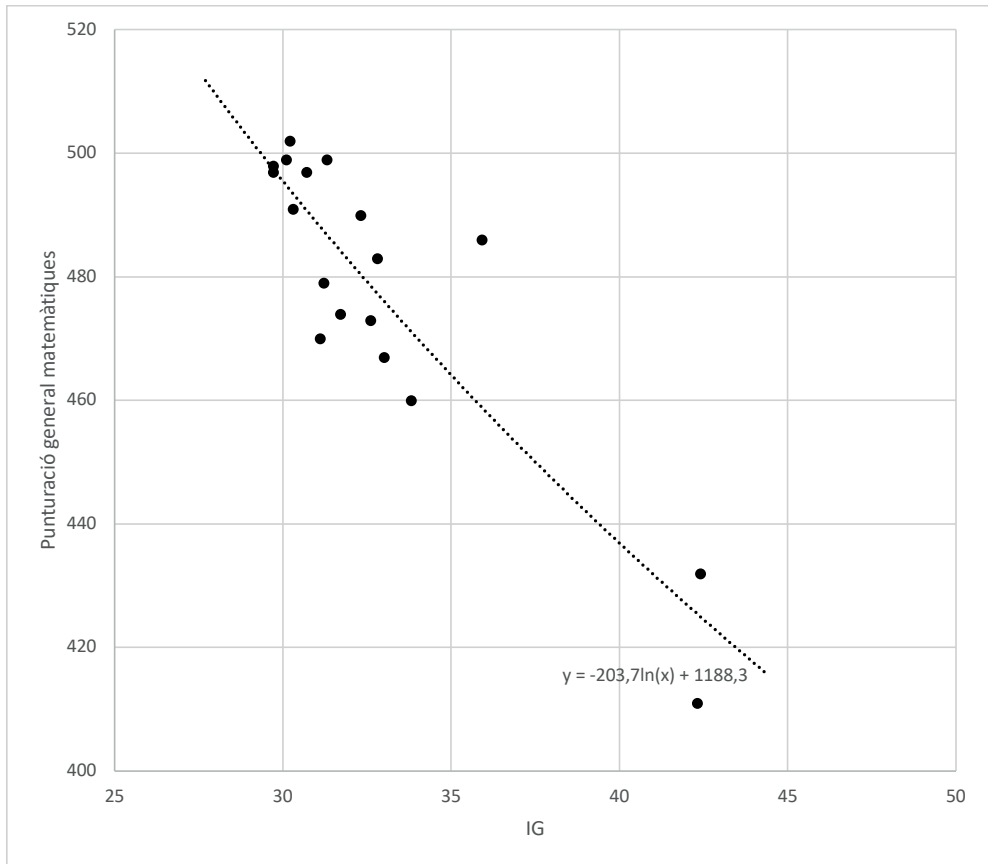


GRÀFIC 8. DESIGUALTAT SOCIAL I PUNTUACIÓ GENERAL DE CIÈNCIES



Font: Elaboració pròpia de la taula 6.

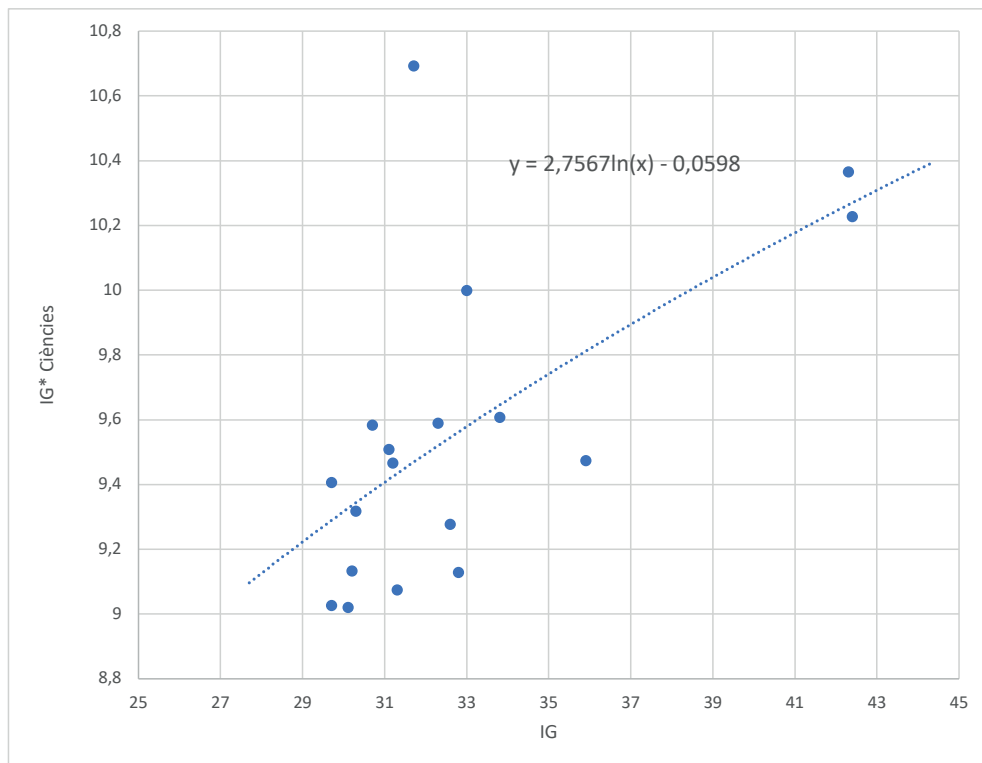
GRÀFIC 9. DESIGUALTAT SOCIAL I PUNTUACIÓ GENERAL DE MATEMÀTIQUES



Font: Elaboració pròpia de la taula 6.

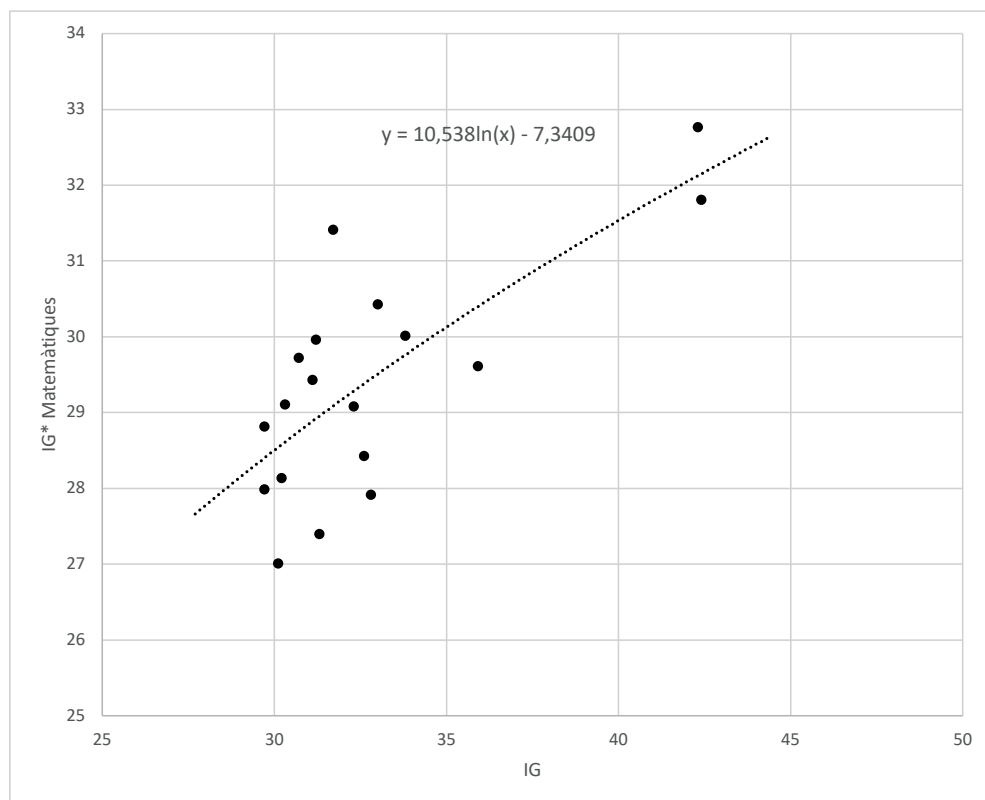
També podem representar la relació entre desigualtat social i desigualtat educativa, en les dues disciplines, ciències i matemàtiques, que representen els gràfics 10 i 11.

GRÀFIC 10. DESIGUALTAT SOCIAL I DESIGUALTAT EDUCATIVA EN CIÈNCIES



Font: Elaboració pròpia de la taula 6.

GRÀFIC 11. DESIGUALTAT SOCIAL I DESIGUALTAT EDUCATIVA EN MATEMÀTIQUES



Font: Elaboració pròpia de la taula 6.

Altrament dit: com a corol·lari, podem avançar models en els quals la variable independent siga la desigualtat social i les variables dependents siguin els rendiment, però també (i ací rau la novetat de la nostra aportació) la desigualtat educativa *mateixa*, diguem-ne, interna al rendiment. No la relació de la desigualtat educativa amb altres desigualtats socials, sinó la composició interna de la mateixa desigualtat.

Una altra aproximació a la mateixa qüestió es pot fer amb un programa d'Intel·ligència Artificial de *mineria de dades*, tot aplicant l'algorisme Random Forest. El resultat s'arreplega en el gràfic 12, amb la línia de tendència logarítmica. L'ús de la Intel·ligència Artificial ens exonera d'haver de fer càlculs estadístics de bondats dels models (com ara, el mètode Akaike), perquè el que fem és ensinistrar un model perquè ens proporcione el millor model possible.

Random Forest és un algorisme d'aprenentatge supervisat, que es presenta com un enfortiment dels *Tree Predictors*. Els *Tree Predictors* són mètodes d'aprenentatge automàtic, que prenen com a base els *Decisions Tree*. Aquests són procediments que divideixen el conjunt de dades en subconjunts més xicotets i, per dir-ho així, aprenen de les característiques dels subconjunts. Amb això poden prendre decisions basades en

seqüències *if-then-else*. Per això es fan servir en problemes tant de classificació<sup>4</sup> com de regressió. Als models *Tree Predictors* s'utilitza un únic arbre de decisió. Per això s'han formulat models més perfeccionats que fan servir uns *Tree Ensembles*, com és el cas de Random Forest, Gradient Boosting i AdaBoost. Aquests *Ensemble Methods* utilitzen tècniques que combinen les prediccions de diversos models base per a millorar la generalització i la precisió<sup>5</sup>.

Random Forest va ser proposat i desenvolupat per Leo Breiman, un estadístic i professor de la Universitat de Califòrnia a Berkeley. Breiman va introduir Random Forest en un article titulat «Random Forests» (Breiman 2001). aquest article que acumula a hores d'ara seixanta tres mil citacions.

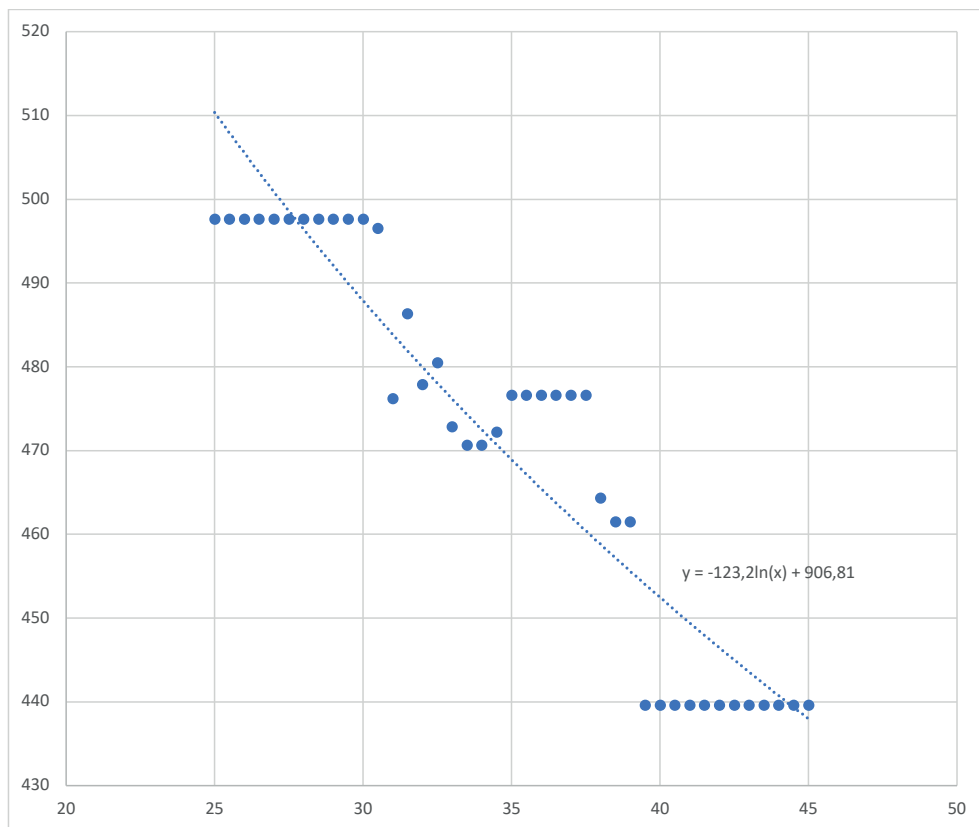
La idea darrere de Random Forest era abordar problemes de sobreajustament comuns als arbres de decisió individuals i millorar la precisió i estabilitat del model combinant múltiples arbres entrenats de forma independent. En síntesi, doncs, Random Forest és un algorisme d'aprenentatge supervisat utilitzat per a problemes de classificació i regressió. Pertany a la categoria de mètodes de conjunt, que combina prediccions de múltiples models base per millorar la precisió i la robustesa generals.

Actualment, Random Forest està implementat a diverses biblioteques i plataformes de programació. Aquestes biblioteques i plataformes permeten als desenvolupadors utilitzar l'algorisme en els projectes i anàlisi de dades. Alguns dels entorns i biblioteques més populars amb implementacions de Random Forest són: Scikit-learn (Python), R (R Language), Tensor Flow (Python), H2O.ai, Apache Spark MLlib; Weka (Waikato Environment for Knowledge Analysis). El programa Orange Data Mining 3.36.1, utilitzat per nosaltres, incorpora entre els seus algorismes de modelització Random Forest, així com d'altres esmentats: Tree, AdaBoost, Gradient Boosting, etc.

<sup>4</sup> Un popular exemple d'arbre de decisió amb objectius classificatoris és allò que s'anomena precisament «triatge» en les entrades d'urgències dels hospitals. Mitjançant una sèrie de preguntes i mesures (temperatura corporal, pressió sanguínia, saturació d'oxigen en sang, etc.) es decideix no només l'especialitat que ha d'atendre el malalt, sinó també el seu codi d'urgència.

<sup>5</sup> Podríem imaginar una persona que en accedir a les urgències d'un hospital passe per diversos serveis de «triatge», cadascú dels quals hi aplique, a més, una metodologia diferent.

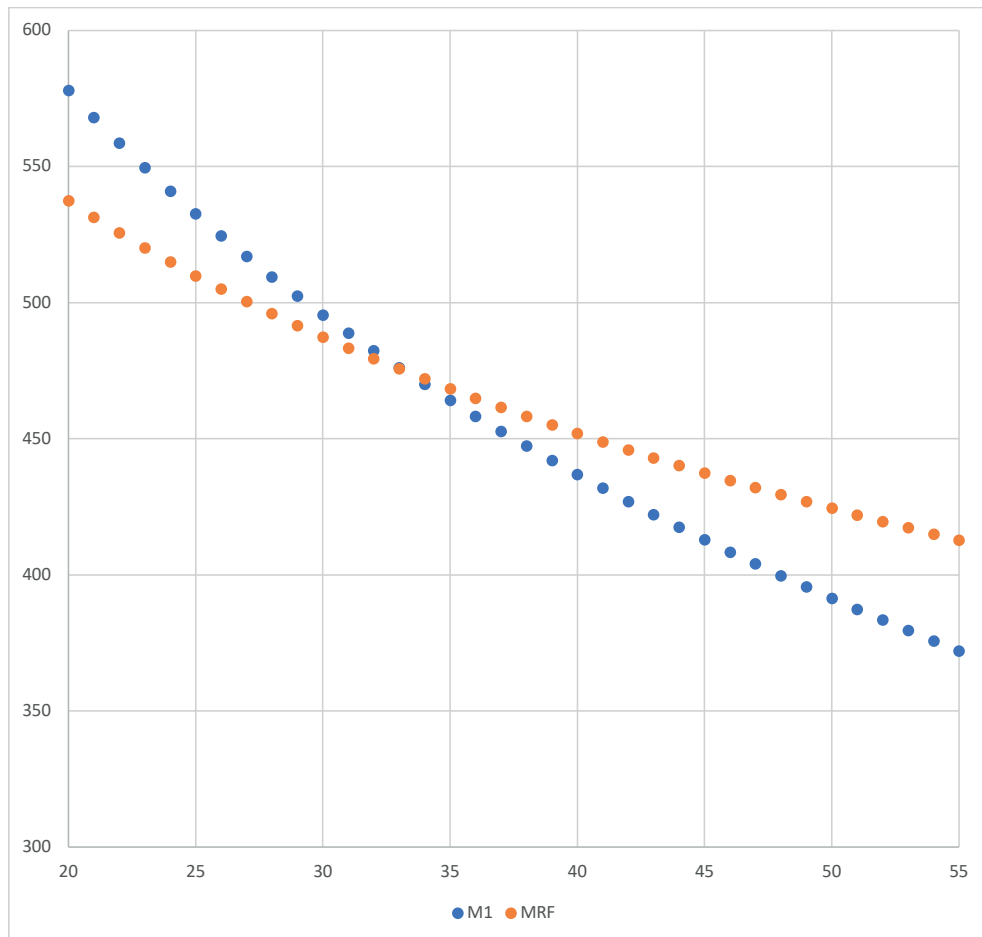
GRÀFIC 12. DESIGUALTAT SOCIAL I PREVISIÓ DE PUNTUACIÓ MATEMÀTICA AMB RANDOM FOREST



Font: Elaboració pròpia dels resultats de l'algorisme Random Forest i línia de tendència.

El gràfic 13 permet visualitzar la diferència entre el model original (M1) i l'elaborat amb Random Forest.

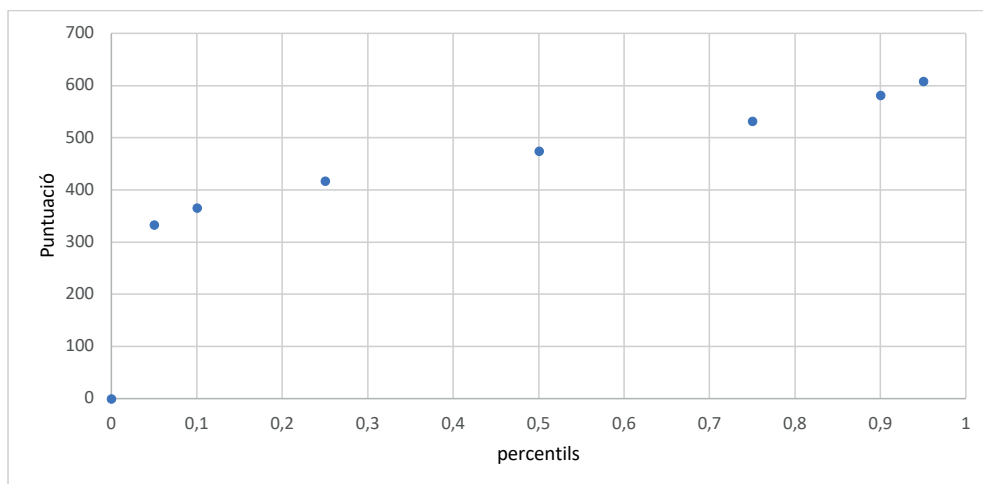
GRÀFIC 13. COMPARACIÓ DE MODELS: LÍNIA DE TENDÈNCIA LOGARÍTMICA I RANDOM FOREST



## 7. Aproximació a la modelització de la corba de la desigualtat educativa. Aplicació de l'equació de Boltzmann (corol·lari 2)

Agafem, a tall d'exemple, les puntuacions de PISA 2018 de la C. Valenciana en matemàtiques, segons la desagregació de percentils proporcionades per l'OCDE (val a dir, les que es troben en la fila corresponent de la taula 2 pel que fa als percentils 5, 10, 25, 50, 75, 90 i 95. La seua representació gràfica es correspon amb els punts del gràfic 14:

GRÀFIC 14. PUNTUACIÓ DE LA C. VALENCIANA EN MATEMÀTIQUES, SEGONS PERCENTILS



La disposició dels punts és semblant a la que origina una fórmula com la següent, que no és més que una adaptació de l'anomenada Equació de Boltzmann:

$$y = \mathbb{LN} \frac{(a+b)!}{a! b!} + \theta$$

En la qual:

- $\mathbb{LN}$  representa una sèrie de logaritmes naturals niuats:  $\ln$ ,  $\ln(\ln)$ ,  $\ln(\ln(\ln))$ ,  $\ln(\ln(\ln(\ln)))$ , etc.
- $\theta$  s'introdueix a fi de trencar l'eventual simetria vertical del gràfic d'aquesta fórmula i presenta els valors següents:

a) Si  $a > b$ , aleshores  $\theta = 0$ , i per tant:

$$y = \mathbb{LN} \frac{(a+b)!}{a! b!}$$



b) Si  $a=b$  (el que correspon al valor màxim de  $y$ ), aleshores podem introduir  $a=b=c$ :

$$y = \mathbb{LN} \frac{(a+b)!}{a!b!} = \mathbb{LN} \frac{(2c)!}{2(c)!}$$

c) Si  $a < b$

$$y = \mathbb{LN} \frac{(2c)!}{2(c)!} + \left[ \mathbb{LN} \frac{(2c)!}{2(c)!} - \mathbb{LN} \frac{(a+b)!}{a!b!} \right]$$

És clar que cada valor de  $\mathbb{LN}$  [val a dir  $\ln$ ,  $\ln(\ln)$ ,  $\ln(\ln(\ln))$ ,  $\ln(\ln(\ln(\ln)))$ , etc.] produirà un model diferent. En la taula 9 s'analitzen les dades de quatre models (com veurem, no cal fer-ne més) i en la taula 10 les proporcions respecte del valor central (és a dir: P50 d'una banda i  $a=b=c$  de l'altra). Les desviacions es presenten en la taula 11, amb la suma de desviacions en valors absoluts.

TAULA 9. MODELS LOGARÍTMICS

Percentil	$\ln$	$\ln(\ln)$	$\ln(\ln(\ln))$	$\ln(\ln(\ln(\ln)))$	PISA
05	18,136	2,897	1,064	0,062	332,74
10	30,482	3,417	1,228	0,206	365,49
25	53,845	3,986	1,382	0,324	417,21
50	66,783	4,201	1,435	0,361	474,73
75	79,722	4,416	1,488	0,398	532,07
90	103,085	4,985	1,642	0,516	581,48
95	115,430	5,504	1,806	0,660	607,98

Font: Elaboració pròpia

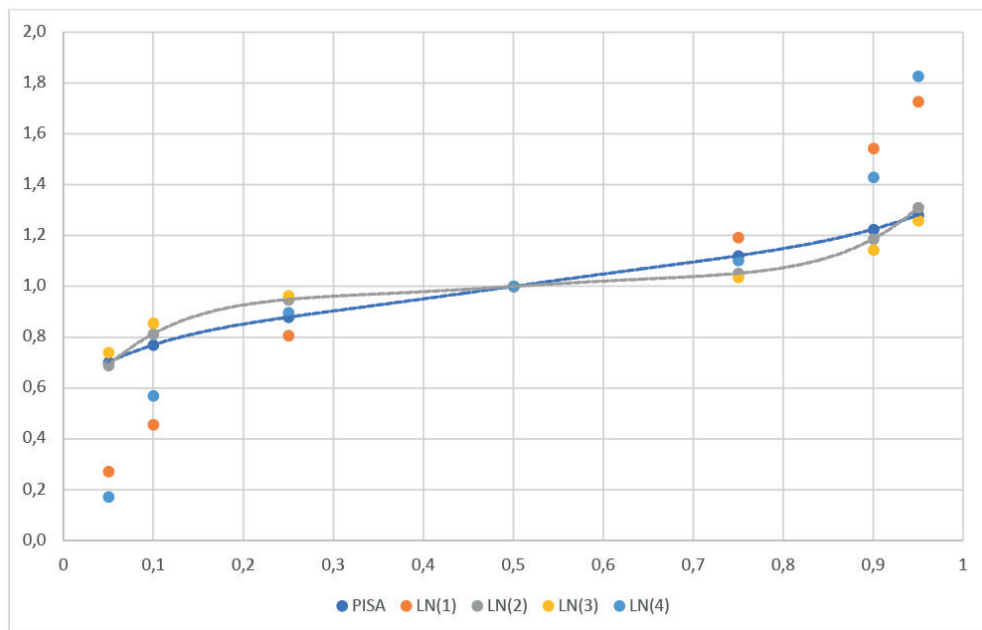
TAULA 10. PROPORCIONS RESPECTE DEL VALOR P50

Percentil	$\ln$	$\ln(\ln)$	$\ln(\ln(\ln))$	$\ln(\ln(\ln(\ln)))$	PISA
05	27,158%	68,975%	74,124%	17,163%	70,090%
10	45,643%	81,332%	85,605%	57,002%	76,989%
25	80,626%	94,875%	96,335%	89,669%	87,884%
50	100,000%	100,000%	100,000%	100,000%	100,000%
75	119,374%	105,125%	103,665%	110,331%	112,078%
90	154,357%	118,668%	114,395%	142,998%	122,486%
95	172,842%	131,025%	125,876%	182,837%	128,069%
<b>Suma dif. abs.</b>	1,655	0,475	0,655	0,885	

Font: Elaboració pròpia de la taula 9.

Com s'hi pot veure, les menors desviacions es registren en la columna corresponent a  $\ln(\ln)$ . Ara podem representar aquests valors en el gràfic 15, on s'han afegit les línies de tendència polinòmiques en els resultats PISA i en  $\ln(\ln)$ , que obté desviacions menors, segons la taula anterior.

GRÀFIC 15



La qüestió fonamental és: per què expressar les diferències de rendiments educatius amb una fórmula tan complicada com la fórmula anterior quan podem fer-ho amb escales més senzilles, encara que arbitràries (PISA no deixa de ser una escala tan convencional com les habituals 0 a 10)? La raó té a veure amb què la fórmula anterior és equivalent a una altra fórmula històrica: l'equació de Boltzmann de l'entropia ( $\mathcal{S}$ ), que se sol presentar així:

$$\mathcal{S} = k \ln \frac{(a + b)!}{a! b!}$$

On el logaritme es refereix a la combinació d'estats possibles de  $a + b$  elements que prenem de  $a$  en  $a$ . Val a dir, en expressió combinatòria:

$$\mathcal{S} = k \ln \binom{a + b}{a}$$

Com hem comprovat en altre treball, aquesta fórmula produeix altes correlacions tant amb les mesures absolutes com amb les mesures relatives –percentatges– (vegeu. Hernández 2023). Val a dir, no només permet expressar la proporció d'una mostra que apleix determinat requisit, sinó també integra la grandària de la mateixa mostra.

## 8. Discussió

Aquest article aconsegueix l'objectiu de formular una noció de desigualtat educativa, val a dir d'adaptar la mesura habitual de desigualtat social als rendiments de l'ensenyament, i ho fa de manera que acredita correlacions importants entre les dues variables amb diversos continguts d'ensenyament. A més, mostra la virtualitat hermenèutica de quantificar els rendiments d'ensenyament a partir d'equacions de corbes logarítmiques.

Encara que ací s'han presentat els resultats de PISA 2018, hem trobat resultats concordants en les edicions anteriors de PISA, com ara 2012 i 2015, que ja hem considerat en altres publicacions, tot i considerant dades dels Estats participants. En escriure aquestes ratlles s'anuncien els resultats de PISA 2021, encara que potser les microdades, com són els resultats per percentils, hauran d'esperar a la publicació dels Informes, com esdevingué en les sèries anteriors.

Pel que fa al contingut de la nostra aportació, ulteriors recerques hauran de verificar les conclusions, en particular la cridanera matriu de correlacions que hem presentat. L'interès d'aquesta verificació és allunyar la possibilitat que hi hagen factors espuris, relatius al procediment de càlcul. Encara que hem fet servir sempre les línies de tendència de major grau, també hem d'estar atents per si les relacions de logaritmes naturals niuats no produeix un cert efecte d'aplanament de les corbes.

Pel que fa als factors espuris, potser el lector haja pensat en la possibilitat que la mesura de la desigualtat plantejada  $IG^*$  estiga relacionada amb la mateixa desviació estàndard de les puntuacions. Certament, ha d'haver una correlació entre l'índex calculat i la desviació, però menor a les correlacions mostrades en la matriu anterior, amb valors de  $R=0,199$  en ciències i  $R=0,495$  en matemàtiques. Pel que fa a les puntuacions generals, la correlació amb les desviacions és de  $R=-0,280$  en ciències i  $R=-0,538$  en matemàtiques.

El debat sobre l'Índex de Gini ha cobrat importància recentment per la formulació de l'anomenada Corba del Gran Gatsby, que vincula desigualtat amb variables relatives a la transmissió de la riquesa. Encara que aquest vincle ha sigut objecte de polèmica (per a un repàs exhaustiu, vegeu Carabaña 2023), el fons del debat és la relació entre dues variables, una de les quals, la desigualtat educativa, s'ha precisat en el nostre article. Resta pendent poder formular la transmissió en termes operatius. Mentre no disposem d'aquesta variable de manera operativa, el debat s'ha d'ajornar.

Per últim, volem destacar que el canvi paradigmàtic del focus educatiu des de l'ensenyament a l'aprenentatge, hauria de portar a un qüestionament de l'acumulació de dades sobre rendiment d'ensenyament. Amb tot, a hores d'ara, no disposem de cap altra font de dades de la importància de les microdades de PISA per poder establir correlacions sòlides.

## 9. Referències

- Adorno, T. W. (2021): Die Frankfurter Seminare Theodor W. Adornos. Ed. D. Braustein. Berlín/Boston: Walter de Gruyter, 4 vols.
- Breiman, L. (2001): Random Forests. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Calero, J.; Waisgrais, S. (2009): Factores de desigualdad en la educación española. Una aproximación a través de las evaluaciones de PISA. *Papeles de Economía Española*, 119, 86-98.
- Carabaña, J. (2023): Desigualdad y movilidad de renta: la 'curva del Gran Gatsby' en 46 provincias españolas, *RIS* vol. 81, núm. 3. <https://doi.org/10.3989/ris.2023.81.3.22.00560>
- EUROSTAT (2023): Income quintile share ratio S80/S20 for disposable income by sex and age group - EU-SILC survey [ilc\_di11] (accés 19 octubre 2023).
- Hernández, F. (2023): Formulació d'un indicador unitari de competència i d'ús de la llengua per a l'avaluació de polítiques lingüístiques. *Treballs de Sociolingüística Catalana*, núm. 33, p. 141-157 doi: 10.2436/20.2504.01.207
- INE (2023): Atlas de distribución de renta de los hogares. Resultados por comunidades autónomas, provincias e islas. Índice de Gini y Distribución de la renta P80/P20. <https://www.ine.es/up/zEkXkJIZ> (accés 19 octubre 2023).
- OECD (2019): PISA 2018 Results (Volume I): What Students Know and Can Do. <http://dx.doi.org/10.1787/888934029109> (accés 19 octubre 2023).
- OECD (2023): OECD (2023), Income inequality (indicator). doi: 10.1787/459aa7f1-en (accés 19 octubre 2023).
- Piketty, Thomas (2013): *Le Capital au XXIe siècle*. París: Éditions du Seuil.
- Piketty, Thomas (2023): *Nature, culture et inégalités*. Nanterre: Société d'ethnologie.
- RASE (2016): A vueltas con la desigualdad, en el cincuentenario del Informe Coleman, *Revista de Sociología de la Educación*, vol. 9, núm. 1.
- RASE (2022): La sociología de la educación después de Bourdieu. *Revista de Sociología de la Educación*, vol. 15, núm. 2.