Intersecția Educației, Sărăciei și Șomajului: O Analiză Comparativă în Europa, 2022

Popa Alexandru  
Pricop Măriuca  
Vorniceanu Vlad

Cuprins

[**Introducere 3**](#_jemw0h34fgju)

[**Aplicația 1 4**](#_wqpat4khncpw)

[**Capitol teoretic 4**](#_agrehym4vd0z)

[Literature review 4](#_g6opgv110krz)

[Metodologia cercetarii 5](#_ze7uxqmr47xj)

[Capitolul aplicativ 6](#_52ake4jiyqjl)

[Date utilizate 6](#_3pcfzdgqn6af)

[Scurtă analiză descriptivă a datelor 6](#_oh4xxvsnr9ip)

[Rezultatele empirice ale cercetării 9](#_lbdgwuv6fciy)

[**Aplicația 2 13**](#_qnl495py6ii2)

[Capitol teoretic 13](#_grb9kdsy8n2h)

[Literature review 13](#_6auuwhk8hrmj)

[Metodologia cercetării 14](#_5nmqtus4jvjg)

[Capitol aplicativ 15](#_hoynmfrcsh0h)

[Date utilizate 15](#_lqeacy2jubff)

[Scurtă analiză descriptivă a datelor 15](#_t45piz6f33zi)

[Rezultatele empirice ale cercetării 17](#_4wlipsywmobt)

[Interpretare parametri 18](#_vfda7bw08osh)

[Concluzii 18](#_2e19bp3x103v)

[**Aplicația 3 27**](#_tbyr3qc070gr)

[Capitol teoretic 27](#_30mg4b9a73m5)

[Literature review 27](#_eumuy3466i9i)

[Metodologia cercetării 28](#_9z4ftg86yzx9)

[Capitol aplicativ 29](#_j4u84phxnhwo)

[Date utilizate 29](#_5mkfn24ir2md)

[Rezultatele empirice ale cercetării 31](#_tuzp49yq8849)

[Metoda celor mai mici pătrate grupată (Pooled OLS) 31](#_szg1oe8w0a66)

[Estimare Modelul cu efecte fixe 34](#_fy2kovpcmwg)

[Estimare Modelul cu efecte aleatoare 35](#_w9b9g0xi7uv7)

[**Bibliografie 42**](#_2u4y71hrgkpc)

# Introducere

În contextul unui peisaj european marcat de fluctuații economice și schimbări sociale profunde în 2022, acest studiu își propune să investigheze dinamica interconectată dintre rata sărăciei, șomajului și nivelul de educație în diferite țări europene. Acest demers analitic este crucial pentru a înțelege cum se reflectă politica economică și socială în viața cetățenilor și cum factori precum nivelul de educație și ocuparea forței de muncă contribuie la configurarea profilului socio-economic al continentului.

Analizăm datele pentru a detecta modele și anomalii, să identificăm zonele de rezistență și vulnerabilitate și să explorăm cum variațiile în educație și șomaj influențează riscul de sărăcie. Acest studiu se bazează pe o abordare metodică, începând cu selecția tematică relevantă, urmată de o colectare minuțioasă de date și analiza acestora. Articolele academice de specialitate au fost revizuite pentru a oferi o bază teoretică, iar aplicarea practică a fost realizată prin metode statistice și analitice.

De asemenea, studiul nostru își propune să dezvăluie cum starea de sărăcie poate impune restricții asupra accesului la educație de calitate și la oportunități echitabile de angajare, creând un ciclu vicios care perpetuează inegalitățile sociale. Prin această abordare holistică, sperăm să oferim o perspectivă clară asupra provocărilor și oportunităților prezente în țările europene, oferind astfel o bază pentru politici mai informate și mai eficiente în domeniul social și economic.

În cadrul studiului, ne-am concentrat asupra colectării și analizei datelor, selectând informații relevante care reflectă starea socio-economică a țărilor europene. Această colectare de date a fost urmată de o clasificare riguroasă și de o evaluare a informațiilor, cu scopul de a identifica corelații și tendințe semnificative. Am utilizat diverse metode statistice pentru a asigura o analiză obiectivă și cuprinzătoare, permițându-ne astfel să obținem o imagine clară a dinamicii dintre educație, sărăcie și șomaj.

Pe parcursul cercetării, am acordat o atenție specială impactului pe care diferitele niveluri de educație le au asupra ratei sărăciei și șomajului. Studiul explorează modul în care educația poate servește drept un catalizator pentru reducerea sărăciei și pentru îmbunătățirea oportunităților de angajare.

# Aplicația 1

## Capitol teoretic

### Literature review

În anul 2022, Europa s-a confruntat cu provocări semnificative în ceea ce privește rata sărăciei și șomajului, fenomene care reflectă complexitatea și diversitatea situațiilor socio-economice în cadrul Uniunii Europene. Prin analiza datelor și rapoartelor de specialitate, devine evident că aceste probleme sunt profund interconectate și influențate de o varietate de factori economici, sociali și politici.

Un aspect alarmant, evidențiat într-un articol de la DW, este rata înaltă a sărăciei în rândul copiilor din România, care a atins un nivel de 41,5% în 2022, cel mai ridicat din întreaga Uniune Europeană. Această situație gravă a fost exacerbată de efectele pandemiei COVID-19, creșterea prețurilor și impactul conflictului din Ucraina, care au afectat în special copiii din zonele rurale și pe cei din familii monoparentale sau numeroase(DW, 2023)​​.

Datele Eurostat relevă o altă fațetă îngrijorătoare: aproape jumătate dintre șomerii din UE-27 erau expuși riscului de sărăcie în 2018, cu cele mai înalte rate în Germania. Această statistică subliniază vulnerabilitatea persoanelor fără locuri de muncă și legătura strânsă dintre șomaj și sărăcie(EuroStat, 2020, 2021)​​​​.

Un rol esențial în abordarea acestor probleme îl are Fondul Social European Plus (FSE+), așa cum este discutat într-un articol de pe site-ul Parlamentului European. Cu un buget impresionant de 88 de miliarde EUR pentru perioada 2021-2027, FSE+ se concentrează pe combaterea sărăciei, în special în rândul copiilor și tinerilor, prin facilitarea accesului la educație, hrană și locuințe decente, precum și prin sprijinirea formării profesionale pentru tinerii aflați în căutarea unui loc de muncă(Actualitate Parlamentul European, 2021)​​.

Impactul inflației asupra sărăciei este un alt factor crucial. Inflația afectează disproporționat persoanele cu venituri mici, cum ar fi cazul în România, unde cei mai săraci 10% dintre cetățeni cheltuiesc aproape jumătate din veniturile lor pe alimente și băuturi nealcoolice. Acest fapt subliniază modul în care creșterile de prețuri pot agrava sărăcia și disparitățile sociale(Normedia, 2023)​​.

Analizând aceste date, devine clar că sărăcia și șomajul sunt probleme complexe care necesită o abordare integrată și politici bine coordonate. Educația, formarea profesională și susținerea economică sunt esențiale pentru a contracara aceste provocări și pentru a promova o societate mai echitabilă și inclusivă în Europa. Astfel, anul 2022 a fost un an de reflecție și acțiune, subliniind necesitatea urgentă de a aborda aceste provocări în mod eficient și sustenabil.

### Metodologia cercetarii

Procesul nostru de cercetare începe prin selectarea și colectarea datelor relevante de pe site-ul Eurostat, asigurându-ne că alegem un an recent pentru a dispune de informații actualizate. Apoi, vom analiza împreună datele pentru a identifica tendințe, modele și corelații importante. Ne vom sprijini unul pe celălalt în căutarea informațiilor și în luarea deciziilor privind conținutul analizei noastre. Pentru a efectua analize economice riguroase, vom utiliza aplicația EViews, care ne va permite să dezvoltăm modele și să obținem rezultate relevante și precise. Această abordare colaborativă și folosirea unei aplicații specializate ne vor ajuta să obținem rezultate de calitate în studiul nostru de caz.

## Capitolul aplicativ

### Date utilizate

Rata de sărăcie

Această variabilă măsoară proporția populației care se află sub pragul de sărăcie, adică cei care au un venit sub nivelul necesar pentru a asigura nevoile de bază.

Unitatea de măsură este procentul (%), și reflectă cât de mare este acest grup în totalul populației.

Datele pentru această variabilă au fost obținute de pe site-ul Eurostat, de la indicatorul ‘At-risk-of-poverty rate’ selectând anul 2022.

Rata șomajului

Această variabilă indică procentajul de persoane care sunt fără loc de muncă în raport cu forța de muncă totală, adică cei care sunt capabili și dispuși să lucreze.

Unitatea de măsură este procentul (%) și reflectă gradul de ocupare a forței de muncă.

Datele pentru această variabilă au fost obținute de pe site-ul Eurostat, de la indicatorul ‘Total unemployment rate’ selectând anul 2022.

### Scurtă analiză descriptivă a datelor

Pentru a face o analiză descriptivă a datelor colectate am folosit funcții din librăriile numpy, pandas și seaborn din python. Utilizând numpy și pandas, am calculat statistici descriptive precum media și deviația standard pentru fiecare indicator. Media oferă o înțelegere a valorii centrale a datelor, reflectând tendința generală, în timp ce deviația standard măsoară cât de larg sunt răspândite valorile în jurul mediei, oferind o perspectivă asupra variabilității sau uniformității setului de date. Folosind Seaborn, am creat grafice de tip boxplot pentru a vizualiza distribuția datelor. Boxplot-urile sunt eficiente în prezentarea medianei, a intervalelor intercartilice și a valorilor extreme. Aceste grafice sunt utile pentru a identifica variațiile și anomalii în date, permițând o interpretare mai clară a dispersiei și a tendințelor centrale.

Rata sărăciei

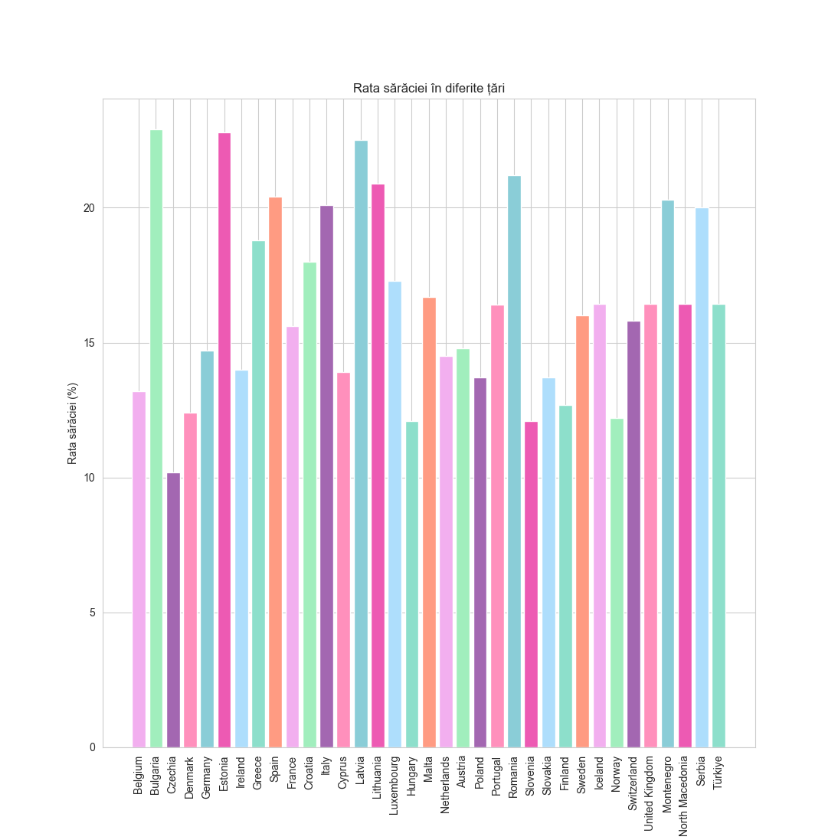
**Media**: 16.45%

**Deviația standard**: 3.43%

**Minim**: 10.20%

**Maxim**: 22.90%

Distribuția ratelor sărăciei este relativ simetrică, cu o ușoară concentrare în jurul mediei.



Rata șomajului

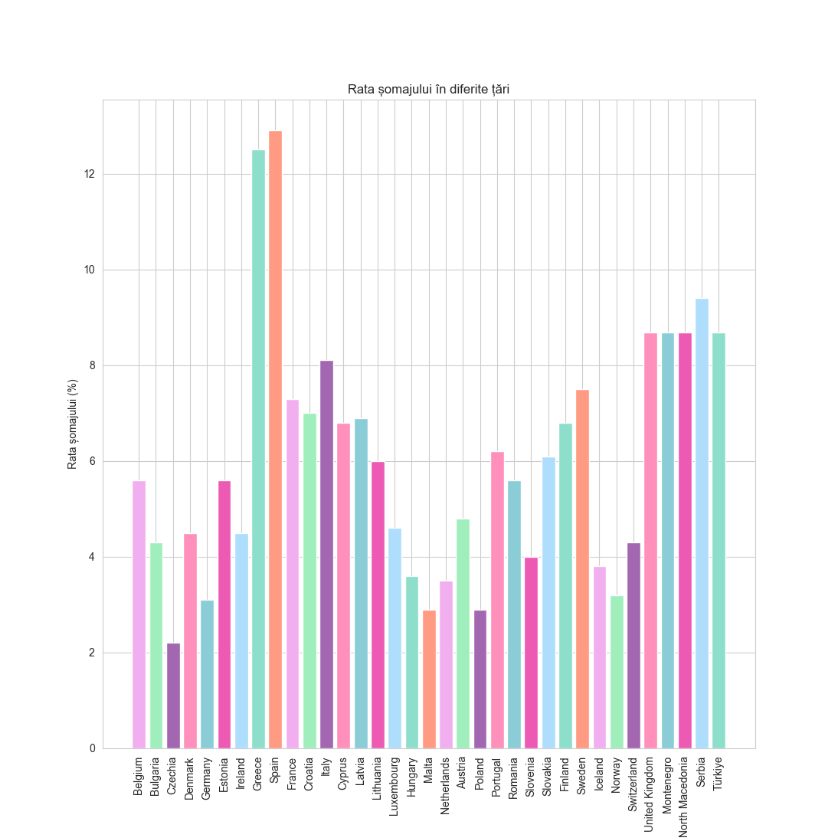
**Media:** 6.04%

**Deviația standard:** 2.58%

**Minim:** 2.20%

**Maxim:** 12.90%

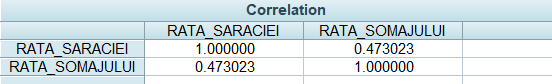
Distribuția ratelor șomajului prezintă o anumită variabilitate, cu valori concentrate în jurul mediei și câteva valori mai mari, indicând posibile excepții.



### Rezultatele empirice ale cercetării

Pentru cerințele de la aplicațiile 1 și 2 am folosit primele 2 variabile: rata sărăciei ca variabilă dependentă și rata șomajului ca variabilă independentă.

Pentru a măsura intensitatea dintre cele două variabile am folosit coeficientul de corelație Pearson.



Coeficientul de 0.473 indică o corelație pozitivă moderată între rata sărăciei și rata șomajului. Acest lucru înseamnă că, în setul de date analizat, pe măsură ce rata șomajului crește, există o tendință de creștere și a ratei sărăciei.

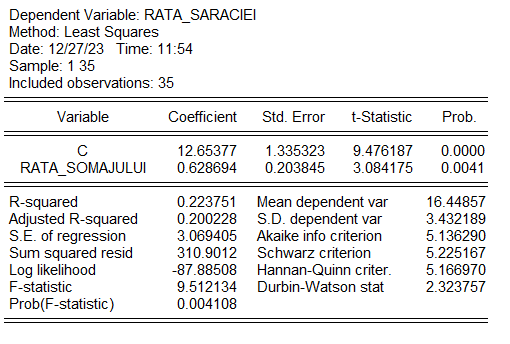


**Distribuția Datelor**: Punctele reprezentate pe grafic par să arate o tendință ascendentă, sugerând că există o relație pozitivă între rata șomajului și rata sărăciei. Cu alte cuvinte, țările cu rate mai mari ale șomajului tind să aibă, de asemenea, rate mai mari ale sărăciei.

**Intensitatea Relației**: Relația nu pare să fie perfect liniară, dar indicația este că, în general, pe măsură ce rata șomajului crește, la fel face și rata sărăciei. Coeficientul de corelație Pearson calculat anterior (aproximativ 0.47) confirmă această corelație moderată.

**Dispersia Datelor**: Există o variație considerabilă în rata sărăciei chiar și pentru țările cu rate similare ale șomajului, ceea ce sugerează că alți factori, în afara șomajului, pot influența nivelurile sărăciei într-o țară.

Estimarea parametrilor



Interpretări econometrice

**Coeficientul Interceptului** (C): Interceptul este 12.65377. Acesta reprezintă valoarea estimată a ratei sărăciei atunci când rata șomajului este zero. Coeficientul este statistic semnificativ, având un p-value de 0.0000.

**Coeficientul pentru rata șomajului**: Coeficientul este 0.628694, ceea ce sugerează că pentru fiecare creștere de 1% în rata șomajului, rata sărăciei crește cu aproximativ 0.629%. Acest coeficient este, de asemenea, statistic semnificativ, având un p-value de 0.0041.

**R-squared**: Valoarea R-squared este 0.223751, indicând că aproximativ 22.38% din variația ratei sărăciei este explicată de modelul nostru care include doar rata șomajului ca variabilă explicativă.

**Adjusted R-squared**: Valoarea ajustată R-squared este 0.200228, care este ajustată pentru numărul de predictorii în model și este o măsură mai precisă a bunătății ajustării pentru modelele cu mai multe variabile.

**F-statistic**: Valoarea F-statistic este 9.512134, sugerând că modelul este statistic semnificativ la nivel global, cu un p-value de 0.004108, ceea ce indică faptul că există o relație liniară puternică între variabile.

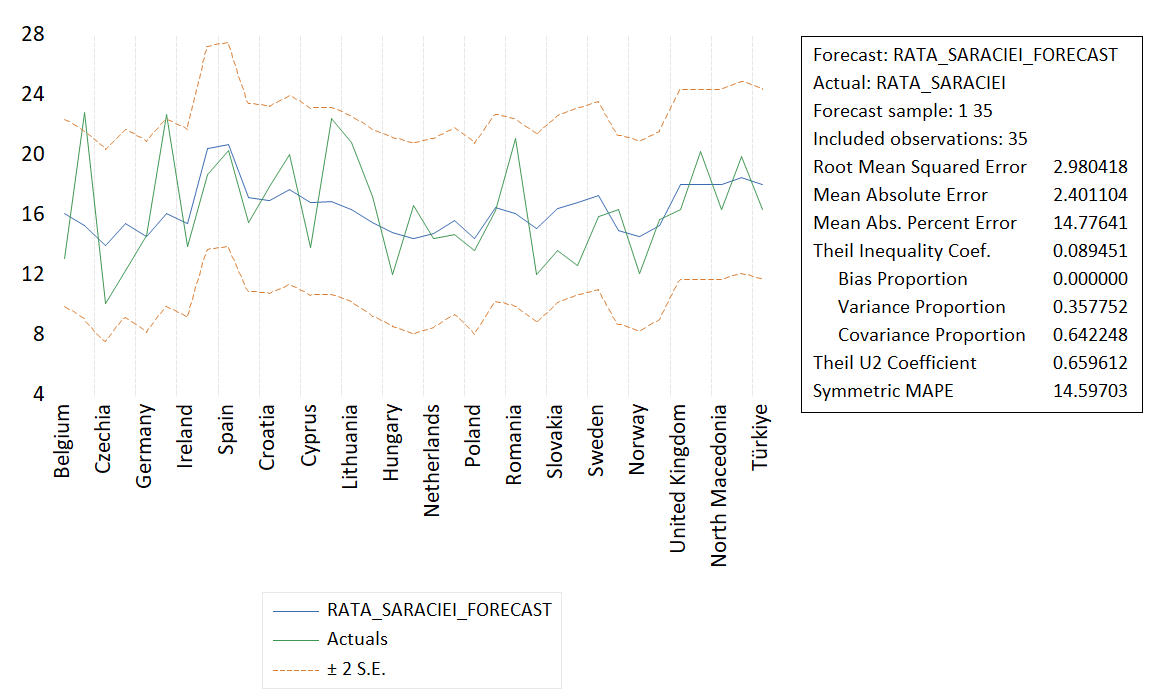
**Durbin-Watson statistic**: Valoarea statisticii Durbin-Watson este 2.323757, care este destul de aproape de 2, sugerând că nu există o autocorelație semnificativă a reziduurilor modelului.

Interpretări economice

Rezultatele indică faptul că există o relație pozitivă semnificativă între rata șomajului și rata sărăciei. Acest lucru poate sugera că politicile care vizează reducerea șomajului ar putea avea efecte benefice asupra reducerii sărăciei.

Deși coeficientul pentru rata șomajului este semnificativ, R-squared nu este foarte mare, indicând că și alte variabile, care nu sunt incluse în model, ar putea juca un rol important în explicarea ratei sărăciei.

Prognoză pentru rata sărăciei



Modelul de regresie estimează că există o corelație moderată între rata șomajului și rata sărăciei, cu un coeficient de corelație Pearson de aproximativ 0.47. Prognoza realizată pentru rata sărăciei arată erori de predicție relativ mari, cu un RMSE de 2.98 și un MAE de 2.40, indicând o acuratețe moderată a modelului. Intervalul de încredere larg subliniază incertitudinea semnificativă în predicțiile modelului de regresie simplă.

# Aplicația 2

## Capitol teoretic

### Literature review

În anul 2022, Europa a prezentat un tablou complex, cu o creștere ușoară a ratei sărăciei severe și deprivării materiale, de la 6.3% la 6.7%. Acest lucru reflectă scăderea puterii de cumpărare și o inversare a tendinței descrescătoare a anilor anteriori. Numărul persoanelor afectate de sărăcie severă a scăzut la 28.7 milioane, dar proporția gospodăriilor fără locuri de muncă semnificative a scăzut ușor, ilustrând impactul pandemiei COVID-19 asupra pieței muncii (European Commission, 2023, p. 36).

Mai mult, rata AROPE, care măsoară riscul de sărăcie și excluziune socială, a variat considerabil în funcție de nivelul de educație, accentuând importanța accesului la educație și formare profesională. Cu o rată AROPE de 34.5% pentru cei cu un nivel scăzut de educație, comparativ cu doar 10.5% pentru cei cu un nivel înalt de educație, diferențele sunt semnificative și reflectă un risc crescut de sărăcie și excluziune socială pentru persoanele cu niveluri educaționale inferioare (European Commission, 2023, p. 20).

Acest tablou este confirmat de tendințele generale, care arată că persoanele cu niveluri mai scăzute de educație se confruntă cu riscuri mai mari de sărăcie în muncă, subliniind rolul esențial al educației și dezvoltării competențelor. În ciuda unei rate a șomajului în scădere, sărăcia în muncă rămâne o problemă persistentă pentru cei cu niveluri mai scăzute de educație, la 18.4% în 2022, comparativ cu 8.7% și 4.1% pentru niveluri medii și superioare de educație, respectiv (European Commission, 2023, p. 14).

Rata șomajului în UE a ajuns la un nivel scăzut record în 2022 și nu este de așteptat să crească în 2023 și 2024, cu toate că există variații între statele membre. De exemplu, România a rămas stabilă, în timp ce Grecia și Spania au experimentat scăderi semnificative. Rata șomajului în rândul tineretului a scăzut și ea, cu cele mai mari scăderi înregistrate în Italia și Bulgaria (European Commission, 2023, p. 27).

Pe de altă parte, sărăcia subiectivă rămâne cea mai ridicată în rândul persoanelor cu niveluri scăzute de educație. Aproape o treime din populația UE cu un nivel scăzut de educație a fost considerată săracă în 2022, cu rate mai mari raportate în rândul populației cu niveluri scăzute de educație comparativ cu cele medii și înalte (Eurostat, 2023).

În completarea analizei actuale privind sărăcia, șomajul și educația în Europa, se observă că în 2022, 95.3 milioane de persoane din UE (22% din populație) au fost expuse riscului de sărăcie sau excluziune socială, arătând o stabilitate față de 2021 (Eurostat, 2023).

Rata șomajului în UE a scăzut, dar peste 27.5 milioane de persoane fie căutau activ un loc de muncă, fie erau subocupate. În mod semnificativ, aproape o treime din populația UE cu educație scăzută era considerată săracă, un contrast evident cu persoanele având niveluri educaționale mai înalte (Euronews, 2023).

În concluzie, datele din 2022 indică că educația și dezvoltarea abilităților rămân esențiale în promovarea angajării și abordarea sărăciei și excluziunii sociale, atât în interiorul, cât și în afara pieței muncii. Acestea subliniază necesitatea unor intervenții politice focalizate pe îmbunătățirea accesului la educație de calitate și pe creșterea oportunităților de angajare, în special pentru grupurile vulnerabile conform unui studiu realizat de Harms, M.B., Garrett-Ruffin, S.D (2023).

### Metodologia cercetării

Pentru această aplicație am folosit o metodologie asemănătoare celei de la prima aplicație: am identificat o nouă variabilă și am preluat datele pentru aceasta de pe site-ul EuroStat. În continuare am analizat cerințele folosind aplicația EViews, iar după am cercetat ce semnifică rezultatele obținute și am întocmit descrierile economice și econometrice.

## Capitol aplicativ

### Date utilizate

Pe lângă variabila dependentă rata sărăciei și variabila independentă rata șomajului am ales încă o variabilă independentă: nivelul mediu de educație. Această variabilă măsoară procentajul populației care a atins un nivel de educație superior, cum ar fi studii universitare sau echivalente. Unitatea de măsură este procentul (%) și reflectă gradul de educație al populației. Datele pentru această variabilă au fost obținute de pe site-ul Eurostat, de la indicatorul ‘Population by educational attainment level’ selectând anul 2022.

### Scurtă analiză descriptivă a datelor

Asemănător aplicației anterioare am folosit librării din python pentru a analiza datele rapid și pentru a desena grafice precise și ușor de înțeles.

Nivelul educației

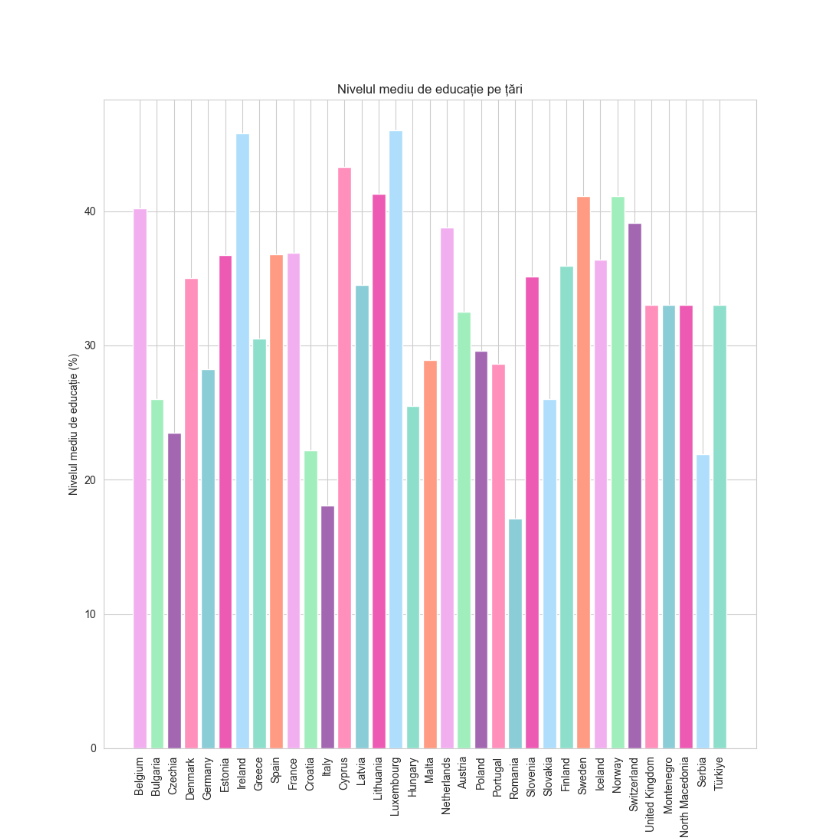
**Media**: 32.99%

**Deviația standard**: 7.42%

**Minim**: 17.10%

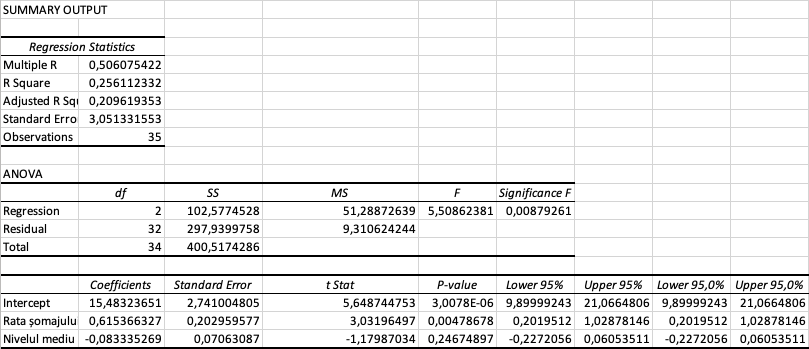
**Maxim**: 46.00%

Distribuția nivelului mediu de educație arată o varietate mai mare, cu valori dispersate pe un interval larg.

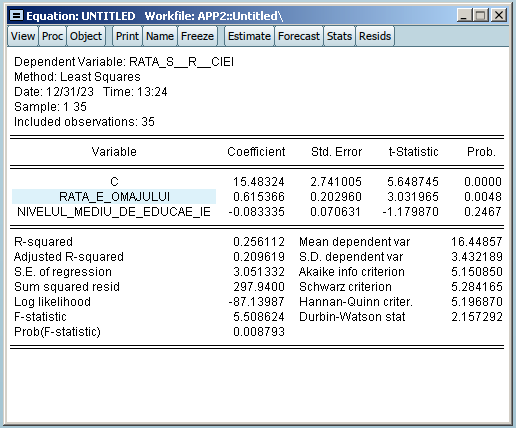


### Rezultatele empirice ale cercetării

2) Estimarea parametrilor modelului de regresie liniară multiplă

Tabel output Excel:  


Tabel output EViews



#### Interpretare parametri

* **Multiple R și R Square**:
  + **Multiple R**: Este coeficientul de corelație între variabila dependentă și variabilele independente. În cazul tău, este 0,506, ceea ce sugerează o corelație moderată între variabilele tale.
  + **R Square**: Acesta indică proporția variabilității în variabila dependentă care poate fi explicată de variabilele independente. Pentru tine, este 0,256, ceea ce înseamnă că aproximativ 25,6% din variabilitatea ratei sărăciei poate fi explicată de ratele șomajului și nivelul mediu de educație.
* **Adjusted R Square**:
  + **Adjusted R Square**: Acesta este R Square ajustat în funcție de numărul de variabile independente. Este util atunci când comparăm modele cu un număr diferit de variabile independente. Pentru tine, este 0,209.
* **Testul ANOVA (Testul F)**:
  + **Significance F**: Valoarea mică indică faptul că cel puțin una dintre variabilele independente are o contribuție semnificativă la model. Valoarea mică (0,00879) sugerează că modelul este semnificativ.
* **Coeficienții de Regresie**:
  + **Intercept**: Este valoarea estimate a ratei sărăciei când toate variabilele independente sunt zero. În cazul tău, este 15,48.
  + **Rata șomajului**: Pentru fiecare unitate de creștere a ratei șomajului, rata sărăciei crește cu 0,615.
  + Nivelul mediu de educație: Pentru fiecare unitate de creștere a nivelului mediu de educație, rata sărăciei scade cu 0,083.
* **P-value**:
  + O valoare mică pentru p-value (0,00479 pentru Rata șomajului) sugerează că această variabilă independentă este semnificativă în model.

#### Concluzii

* Modelul pare să fie semnificativ în predicția ratei sărăciei.
* Adjusted R Square ajustează R Square pentru a lua în considerare numărul de variabile independente. Cu cât acesta este mai aproape de R Square, cu atât este mai bine. În cazul tău, este 0,209, ceea ce indică că modelul tău este destul de simplu, dar încă furnizează o explicație rezonabilă pentru variația ratei sărăciei.
* Valorile p-value mici pentru coeficienții variabilelor independente sugerează că ambele variabile sunt semnificative în model.

a) Comparați Valorile pentru Adjusted R-squared. Concluzia?

Comparând valorile pentru Adjusted R-squared în Excel și EViews și având în vedere că acestea sunt aceleași, putem trage concluzia că rezultatele obținute din cele două platforme sunt consistente și converg către aceeași măsură de ajustare a coeficientului de determinare. Prin urmare, concluzia este că modelul de regresie liniară simplă furnizează o explicație relativ bună a variabilității în rata sărăciei, având în vedere variabilele independente implicate în analiză.

b) Ecuația de regresie estimată. Interpretarea parametrilor pantă estimaţi.

Rata sărăciei = 15,4832 + 0,6154×Rata șomajului − 0,0833 × Nivelul mediu de educație

Aceasta este ecuația care estimează rata sărăciei în funcție de rata șomajului și nivelul mediu de educație. Interpretarea coeficienților este următoarea:

* **Coeficientul pentru Intercept** (15,4832) reprezintă valoarea estimată a variabilei dependente (rata sărăciei) atunci când toate variabilele independente sunt zero.
* **Coeficientul pentru Rata șomajului** (0,6154) indică cât de mult se schimbă variabila dependentă pentru o unitate de schimbare a rata șomajului, menținând celelalte variabile constante.
* **Coeficientul pentru Nivelul mediu de educație** (-0,0833) indică cât de mult se schimbă variabila dependentă pentru o unitate de schimbare a nivelului mediu de educație, menținând celelalte variabile constante.

c) Testarea validităţii modelului de regresie (testul F)

Pentru testarea validității modelului de regresie, vom folosi testul F. Ipotezele pentru acesta sunt:

H0: modelul nu este valid statistic (MSR=MSE)

H1: modelul este valid statistic (MSR>MSE)

Deoarece Fcalc = 5,5086 iar Significance F (pragul de semnificație calculat, nu impus, al testului) este F-statistic = 0,0088 (valoare mai mică decât 0,05=alfa, nivelul de semnificație considerat sau impus al testului), atunci respingem H0 si acceptam H1, anume că modelul de regresie construit este unul valid statistic, pentru o probabilitate maximă de (100 - 0,0088\*100)% = 99,12% > 95%, și poate fi utilizat astfel pentru analiza dependenței dintre variabilele precizate.

d) Testarea semnificaţiei unui parametru pantă

În acest punct al lucrării vom testa semnificația parametrului pantă asociat variabilei “Nivelul mediu de educație”. Pentru asta, vom folosi testul T pentru tratarea coeficientului. Conform rezultatelor obținute prin intermediul EViews și Excel, avem următoarele date pentru variabila “Nivel mediu de educație”:

* Coeficient = -0,0833
* Std. Error = 0,0706
* T. Stat = -1,1799

Vom impune următoarele ipoteze:

* Ipoteza nulă (H0): Coeficientul asociat variabilei independente este egal cu zero (nu are impact semnificativ).
* Ipoteza alternativă (H1): Coeficientul asociat variabilei independente este diferit de zero (are impact semnificativ).

Formula testului T este:

Deoarece pragul de semnificație calculat (nu impus) al testului, P-value, este 0,2467 > 0,05= α, nu există suficiente dovezi statistice pentru a respinge ipoteza nulă că coeficientul asociat variabilei 'Nivelul mediu de educație' este zero. Astfel, nu putem considera acest coeficient ca fiind semnificativ diferit de zero în modelul de regresie. Intervalul de încredere [-0,2272; 0,0605] acoperă valoarea posibilă a coeficientului la un nivel de încredere de 95%, iar faptul că acest interval conține valoarea zero susține ideea că variabila **“Nivelul mediu de educație” nu are un impact semnificativ și stabil asupra variabilei dependente**.

3) Verificarea ipotezelor clasice formulate asupra modelului de regresie liniară

• **Verificare ipoteza de neautocorelare a erorilor (Durbin-Watson; Breusch-Godfrey)**

Testul Durbin-Watson:

Pentru un nivel de semnificație de 5%, formulăm ipotezele:

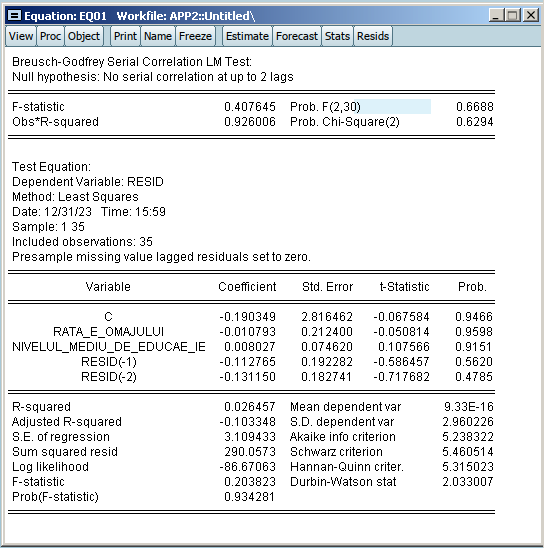
H0: Autocorelația erorilor aleatoare de ordin 1 este absentă (𝜌 = 0).

H1: Există autocorelație de ordin 1 a erorilor aleatoare (𝜌 ≠ 0).



Din calculele efectuate, obținem că DW=2,157292 => nu există autocorelație semnificativă a erorilor în modelul nostru de regresie. => acceptam H0

Testul Breusch-Godfrey



Conform rezultatelor obținute în urma testului Breusch-Godfrey, putem observa că probabilitatea F.Stat = 0,6688 care este mai mare decât nivelul de semnificație de 5% (0,05), prin urmare nu avem suficiente dovezi pentru a respinge ipoteza nulă de non autocorelație a ordinului superior al erorilor. De asemenea, probabilitatea Chi-Square(2) = 0,6264 care și ea depășește valoarea nivelului de semnificație, ceea ce susține ideea că nu există autocorelație semnificativă a ordinului superior.

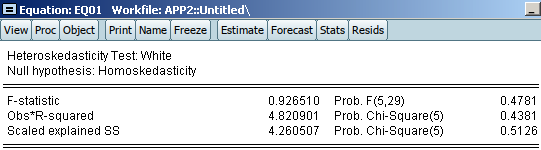
Prin urmare, pe baza acestor rezultate, nu există suficiente dovezi pentru a respinge ipoteza nulă de non autocorelație a erorilor în modelul de regresie.

• **Verificare ipoteza de homoscedasticitate a erorilor (doar Testul White)**

Avem ipotezele:

H0: erorile aleatoare sunt homoscedastice

H1: erorile aleatoare sunt heteroscedastice



F-statistic = 0,926510   
Obs\*R-squared = 4,820901

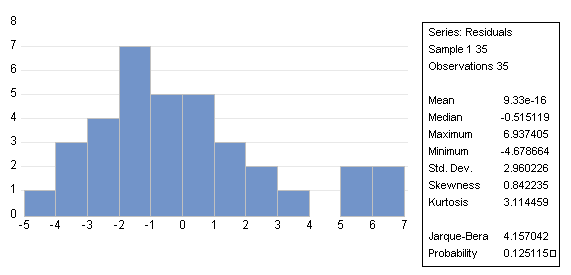
Observăm că Prob. (F-statistic) =0,4781 iar Prob. (Chi-Square ) = 0,4381.

Deoarece probabilitățile P-value < 0,05 respingem H0 și acceptăm ipoteza H1: erorile aleatorii sunt heteroscedastice.

• **Verificare ipoteza de normalitate a erorilor (Jarque-Bera)**

H0: Reziduurile au distribuţie normală ( S=0 şi K=3 )

H1: Reziduurile nu au distribuţie normală



Interpretarea rezultatelor  
**Jarque-Bera** = 4,157042

**Probability** = 0,125115

Deoarece Probabilitatea asociată statisticii JB este 0,125 > 0,05 vom accepta H0. Așadar, reziduurile au o distribuție normală.

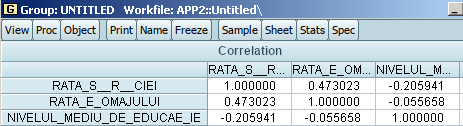
• **Verificare ipoteza privind multicoliniaritatea („matricea corelațiilor”; Klein)**

Avem ipotezele:

**Ipoteza Nulă():** Nu există multicoliniaritate semnificativă între variabilele independente

**Ipoteza Alternativă():** Există multicoliniaritate semnificativă între cele două variabile independente

1. **Metoda matricei corelațiilor**



Dacă respingem și acceptăm (există multicoliniaritate).

În cazul nostru, se poate observa că

: nu există multicoliniaritate semnificativă între variabilele independente .

1. **Criteriul lui Klein**

Regula lui Klein: Multicoliniaritatea poate fi o problemă dacă obținut dintr-o regresie auxiliară (a unei variabile exogene în raport cu celelalte variabile exogene din model) este mai mare decât , obținut din regresia lui Y în raport cu toate variabilele exogene din model.

Vom trata aceleași ipoteze de mai sus:  
**Ipoteza Nulă():** Nu există multicoliniaritate semnificativă între variabilele independente

**Ipoteza Alternativă():** Există multicoliniaritate semnificativă între cele două variabile independente

Variabilele sunt coliniare dacă:

Pasul 1: Din estimarea modelului complet se reține R-Squared, notat .

Pasul 2: Calculăm matricea de corelații liniare ale variabilelor explicative: (

Pasul 3: Dacă , variabilele sunt puternic corelate.

, iar

Întrucât > Variabilele sunt sunt corelate   
 Acceptăm ipoteza nulă **:** Nu există multicoliniaritate semnificativă între variabilele independente.

4) Previzionarea valorilor variabilei dependente Y pentru valori cunoscute ale

variabilelor explicative  
Pentru modelul de calcul utilizat, ecuația de regresie arată astfel:

Unde:

Y = variabila dependentă (Rata sărăciei)

= parametrul de interceptare

= coeficienții de regresie parțiali pentru variabilele independente ,

, = variabilele independente explicative (rata șomajului, nivelul mediu de educație)

= termenul de eroare

# Aplicația 3

## Capitol teoretic

### Literature review

În anii 2020-2022, Europa a fost martoră a unui tablou socio-economic complex, ilustrat de tendințe variate în privința sărăciei, șomajului și nivelului de educație. Datele preluate de pe site-ul Eurostat au furnizat o imagine detaliată asupra evoluției acestor fenomene într-un eșantion reprezentativ format din 35 de țări europene.

În timp ce Europa a experimentat o ușoară creștere a ratei sărăciei severe și deprivării materiale, atingând 6.7%, față de 6.3% în anii precedenți, analiza reflectă impactul pandemiei COVID-19 asupra puterii de cumpărare și a pieței muncii (European Commission, 2023, p. 36). În ciuda scăderii numerice a persoanelor afectate, la 28.7 milioane, procentajul reflectă o inversare a tendințelor anterioare.

Rata AROPE, care măsoară riscul de sărăcie și excluziune socială, a evidențiat diferențe semnificative în funcție de nivelul de educație. Persoanele cu educație scăzută au prezentat un risc de sărăcie și excluziune socială mai ridicat (34.5%), în comparație cu cele cu educație înaltă (10.5%) (European Commission, 2023, p. 20). Aceste constatări subliniază importanța accesului la educație și formare profesională în combaterea sărăciei.

Sărăcia în muncă a rămas o problemă persistentă pentru cei cu niveluri mai scăzute de educație, cu o rată de 18.4% în 2022, comparativ cu 8.7% și 4.1% pentru niveluri medii și superioare de educație (European Commission, 2023, p. 14). Chiar dacă rata șomajului a atins un minim record, sărăcia în muncă relevă provocări continue pentru aceste grupuri.

Rata șomajului în UE a atins un nivel scăzut record în 2022, cu variații între statele membre. În timp ce România a înregistrat stabilitate, Grecia și Spania au avut scăderi semnificative. De asemenea, rata șomajului în rândul tineretului a scăzut, evidențiind variații regionale (European Commission, 2023, p. 27).

Sărăcia subiectivă a rămas ridicată în rândul persoanelor cu educație scăzută, afectând aproape o treime din această populație în 2022 (Eurostat, 2023). Datele subliniază nevoia de intervenții specifice pentru îmbunătățirea condițiilor de viață pentru aceste grupuri.

Datele din 2022 indică că educația și dezvoltarea abilităților rămân esențiale în combaterea sărăciei și excluziunii sociale. Necesitatea unor politici focalizate pe îmbunătățirea accesului la educație de calitate și creșterea oportunităților de angajare devine evidentă (Harms, M.B., Garrett-Ruffin, S.D., 2023). Aceste concluzii au implicații semnificative pentru elaborarea și implementarea viitoarelor politici sociale și economice în Europa.

### Metodologia cercetării

Am adoptat aceeași metodologie pentru această aplicație precum în cea de-a doua aplicație. Totuși, datele utilizate au acoperit un interval de 3 ani (2020 - 2022) și au fost extrase de pe site-ul EuroStat. Am efectuat analiza cerințelor utilizând aplicația EViews, iar ulterior am investigat semnificația rezultatelor obținute.

## Capitol aplicativ

### Date utilizate

În cadrul analizei noastre, am adoptat o structură specifică pentru datele de tip panel, având dimensiuni cross-sectionale N=35 și dimensiuni temporale T=3, cu scopul de a organiza și gestiona eficient informațiile colectate în perioada 2020-2022. Această structură este prezentată detaliat în cadrul acestei secțiuni și a fost consolidată într-un singur tabel într-un sheet unic în documentul Excel.

Am introdus, de asemenea, o coloană suplimentară denumită "An" pentru a identifica anul corespunzător fiecărei înregistrări. În acest tabel unic, fiecare rând reprezintă o unitate de observație distinctă, "Țara", iar fiecare coloană corespunde unei variabile relevante, precum "Rata de sărăcie", "Rata șomajului" și "Nivelul mediu de educație".

Această structură simplificată facilitează analizele longitudinale, permițând evaluarea evoluției variabilelor pe parcursul anilor într-un singur loc. Informațiile colectate în anii specifici (2020, 2021 și 2022) sunt distribuite în cadrul aceleiași arhitecturi unice, oferind o abordare coerentă și sistematică. Aceasta facilitează analizele comparative între diferitele perioade și oferă o perspectivă clară asupra schimbărilor și tendințelor înregistrate în cadrul fiecărui an.

Această structură eficientă cu date de tip panel se dovedește esențială pentru analizele ulterioare, permițând investigarea detaliată a evoluției variabilelor în timp și compararea performanțelor între unitățile de observație, toate reunite într-un singur loc convenabil.

Structura datelor expusă anterioară se încadrează în cadrul unui model de regresie specific, și anume regresia cu date de tip panel. Această metodologie de analiză statistică constituie o abordare interdisciplinară, integrând elemente atât din regresia cu date în secțiune transversală (cross-section regression), cât și din regresia cu date serii de timp (time series regression). Este esențial de subliniat că regresia cu date de tip panel oferă avantajele analizei longitudinale, capturând atât variațiile între unitățile de observație (secțiune transversală), cât și variațiile în timp ale acelorași unități de observație (serii temporale).

Formula generală a unui model de regresie cu date de tip panel include termeni specifici pentru a controla variațiile specifice unităților de observație și pentru a furniza o abordare robustă în analiza datelor longitudinale. În cazul unui model de regresie cu efecte fixe, formula poate fi exprimată astfel:

unde:

- reprezintă variabila dependentă pentru unitatea de observație i la momentul t.

- , … sunt variabilele independente corespunzătoare unității de observației și momentului t.

- este parametrul de interceptare.

-… reprezintă coeficienții asociați variabilelor …

- reprezintă eroarea aleatoare.

Această formulă integrează atât caracteristicile transversale cât și cele temporale, contribuind la o analiză comprehensivă a datelor de tip panel.

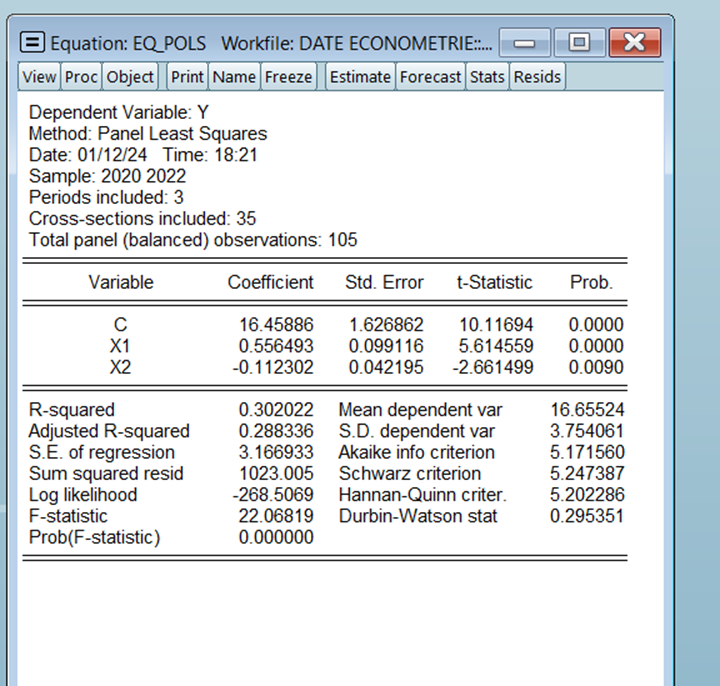
### Rezultatele empirice ale cercetării

#### Metoda celor mai mici pătrate grupată (Pooled OLS)

În cadrul analizei noastre empirice, am aplicat Metoda celor Mai Mici Pătrate Grupată (Pooled OLS) pentru evaluarea relațiilor dintre variabilele definitorii, respectiv Rata de Sărăcie, Rata Șomajului, și Nivelul Mediu de Educație, pe parcursul perioadei 2020-2022, pentru cele 35 de țări incluse în studiu. **Metoda celor Mai Mici Pătrate Grupată** reprezintă o tehnică de regresie utilizată în analiza datelor de tip panel, având drept scop estimarea parametrilor într-un model liniar. Această metodă consideră datele ca fiind "grupate" într-un singur set, ignorând variațiile specifice fiecărei unități transversale sau perioade temporale. Astfel, scopul acestei metode este să ofere o estimare globală a relațiilor dintre variabilele cercetate, fără a lua în considerare diferențele individuale între unitățile sau perioadele de timp.

În lumina ecuației de regresie estimate,

, rezultatele analizei noastre, prezentate în continuare, se bazează pe tabelul EViews furnizat mai jos. Am formulat și testat mai multe ipoteze pentru a înțelege relațiile dintre variabilele incluse în model.



**Ipotezele testate sunt:**

1. **Ipoteza de semnificație globală (F-test):**

* H0: Toți coeficienții de regresie sunt zero (niciun efect semnificativ în model).
* H1: el puțin un coeficient de regresie este diferit de zero.

1. **Ipoteza de semnificație individuală (t-test):**

* H0: Coeficientul asociat cu fiecare variabilă independentă este zero.
* H1: Coeficientul asociat cu fiecare variabilă independentă este diferit de zero.

1. **Testul Durbin-Watson:**

* H0: Nu există autocorelare în reziduurile modelului.
* H1: Există autocorelare în reziduurile modelului.

**Rezultatele regresiei Pooled OLS sunt:**

**Semnificație globală (F-test):**

* Statistica F este de 22,0681, iar probabilitatea asociată este 0.0000 (< 0.05), respingând ipoteza nulă. Acest rezultat indică existența cel puțin unei variabile explicative semnificative în model.
* Acceptăm H1

**Semnificație individuală (t-test):**

* Ambele variabile independente sunt semnificative statistic la un nivel de semnificație de 0,05. Probabilitățile asociate testelor pentru X1=0,0000 și X2=0,0090 (mai mici decât 0,05).
* Acceptăm H1

**Testul Durbin-Watson (DW):**

* Statistica DW este de 0,2953, o valoare foarte mică, sugerând posibila existență a autocorelării pozitive în reziduurile modelului.
* Acceptăm H1

**Calitatea ajustării modelului**

* R2=0,2883, ceea ce înseamnă 28,83% din variabilitatea lui Y este explicată de variabilitatea lui X1 și X2

**Direcția efectelor (Coeficienți):**

* Coeficientul asociat cu X1 ​ este pozitiv, indicând un impact pozitiv asupra lui Y.
* Coeficientul asociat cu X2 este negativ, sugerând un impact negativ asupra lui Y.

#### Estimare Modelul cu efecte fixe

Modelul cu efecte fixe este o metodă statistică utilizată în analiza datelor de tip panel, unde observațiile sunt colectate pentru aceleași unități individuale de-a lungul unui interval de timp. În acest context, "efectele fixe" reprezintă caracteristici constante individuale care sunt specifice fiecărei unități în parte și care nu se schimbă pe parcursul perioadei de analiză. Scopul acestui model este de a captura și controla variația neobservată la nivel individual, oferind astfel o perspectivă mai detaliată asupra influențelor constante care pot afecta variabilele analizate. Prin includerea acestor efecte fixe în ecuația regresiei, modelul cu efecte fixe permite estimarea și interpretarea mai precisă a relațiilor dintre variabilele dependente și independente în cadrul datelor de tip panel.

**Interpretarea datelor pentru acest model**:

1. **Termenul Constant** (Valoarea medie a efectelor fixe):

Valoarea medie a efectelor fixe (𝜶̅i ) este de 26.3355. Acesta reprezintă valoarea constantă la care se stabilizează variabilele individuale în model.

1. **Efectele Fixe Raportate** ():

Efectele fixe raportate sunt abateri de la valoarea medie ( pentru fiecare categorie. Cu alte cuvinte, ele arată cât de mult se abate fiecare categorie individuală față de media generală a efectelor fixe.

1. **Coeficienții Panta**:

Coeficienții de regresie (panta) au un grad de semnificație mai mare. Aceasta indică faptul că relațiile dintre variabilele independente și variabila dependentă sunt mai semnificative, sugerând că modelează mai precis variabilitatea observată.

1. **Coeficientul de Determinare R2**:

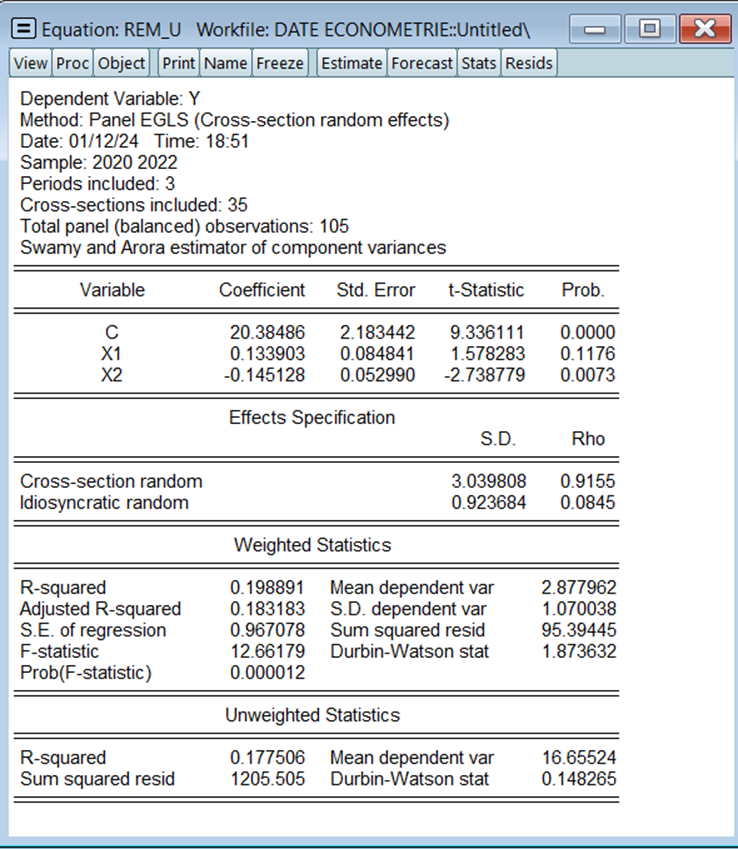
R2 =0.9604 sugerează că 96.04% din variabilitatea variabilei dependente este explicată de variabilele independente și efectele fixe. Aceasta indică o foarte bună ajustare a modelului la datele observate.

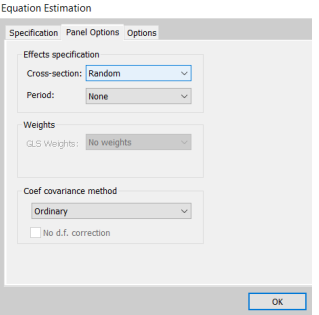
1. **Testul Durbin-Watson (DW)**:

Statistica DW este de 2.9887. O valoare aproape de 2 sugerează absența autocorelării în reziduurile modelului. Cu toate acestea, această valoare specifică (2.9887) este mai mare decât 2, indicând o posibilă absență a autocorelării pozitive.

#### Estimare Modelul cu efecte aleatoare

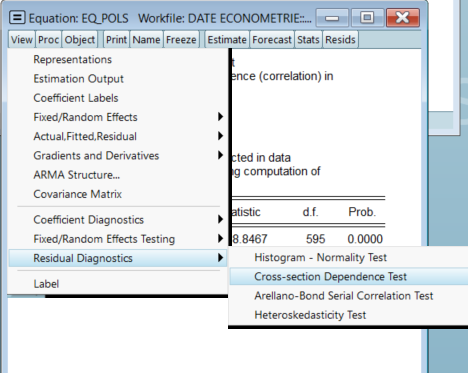
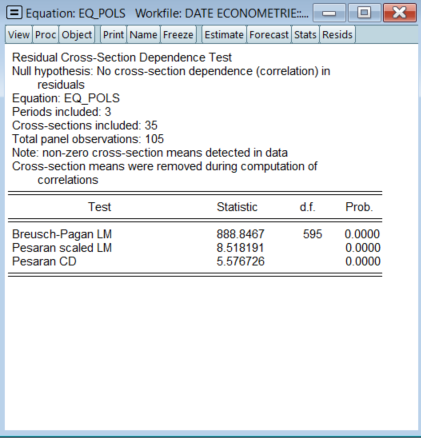
Modelul cu efecte aleatoare reprezintă o abordare robustă în analiza datelor de tip panel, în care observațiile sunt colectate pentru aceleași unități individuale de-a lungul unui interval de timp. În contrast cu modelele cu efecte fixe, unde se presupune că fiecare unitate are o contribuție constantă la variabilitatea dependenței, modelul cu efecte aleatoare permite variabilitatea acestor contribuții să fie distribuită aleator în întreaga populație. Efectele aleatoare reflectă astfel diferențele nesistematice între unități și oferă o perspectivă mai flexibilă asupra caracteristicilor individuale. Scopul acestui model este de a captura variabilitatea nedeterministică a efectelor individuale și de a estima impactul acestora asupra variabilelor dependente, permițând în același timp explorarea diversității între unități. Această abordare se dovedește utilă în situațiile în care există variabilitate nesistematică semnificativă între unități și permite estimarea eficientă a relațiilor dintre variabilele independente și variabila dependentă în contextul datelor longitudinale.





**Teste de dependență cross-secțională a panelului - testul Breusch-Pagan Lagrange Multiplier (LM)**

* H0: Metoda Pooled OLS este metoda potrivită
* H1: Este mai potrivită metoda FE sau RE

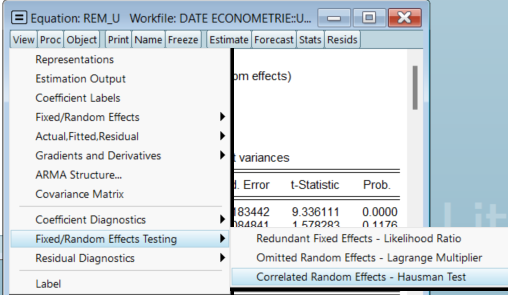


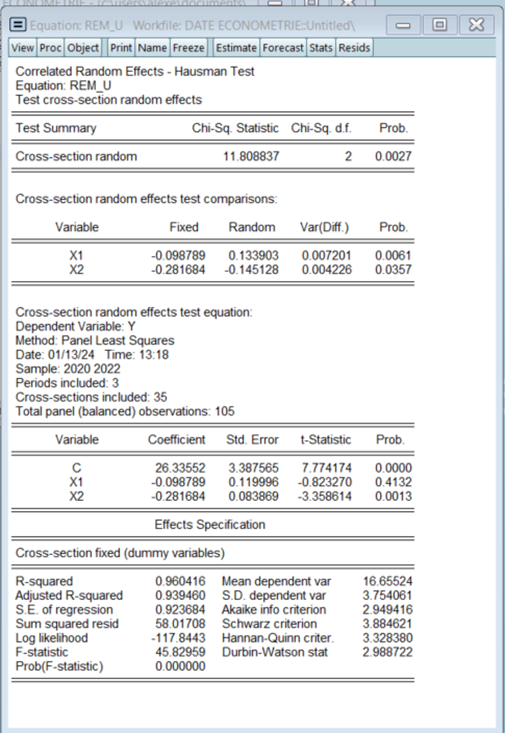
**Decizie**: Deoarece Prob(BP)=0,0000 < 0,05 => Este mai potrivită Metoda FE sau RE => Acceptăm H1.

**Testul Hausman (pentru efecte aleatoare corelate)**

* H0: => modelul REM este potrivit
* H1: => modelul FEM este potrivit

Testul Hausman pentru efecte aleatoare corelate reprezintă o metodă importantă în analiza datelor de tip panel, utilizată pentru a decide între utilizarea unui model cu efecte fixe și unul cu efecte aleatoare. În contextul efectelor aleatoare corelate, se presupune că există o corelație între efectele individuale și variabilele explicative, permițând variabilitatea acestora să fie distribuită aleator în întreaga populație. Testul Hausman are ca scop evaluarea dacă există sau nu o corelație sistematică între efectele aleatoare și variabilele explicative, cu alte cuvinte, dacă excluderea corelației din model este sau nu justificată. Prin compararea estimărilor dintre modelul cu efecte aleatoare și cel cu efecte fixe, testul Hausman furnizează o cale de a decide care model este mai potrivit în ceea ce privește explicarea variației observate în datele de panel. Astfel, acest test reprezintă o unealtă esențială în procesul de selecție și validare a modelului potrivit pentru analiza specifică a datelor longitudinale.





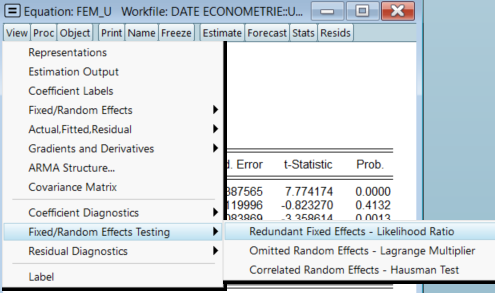
Decizie: Statistica testului are o distribuție Chi-Square cu df=2 (număr de variabile explicative din model). P-value = 0,0027 < 0,05 => acceptăm H1 => FEM este mai potrivit

**Interpretarea rezultatelor regresiei cu FEM:**

1. **Testul F - Regresia ESTE semnificativă**:
   1. H0: Toți coeficienții de regresie sunt zero (niciun efect semnificativ în model).
   2. H1: el puțin un coeficient de regresie este diferit de zero.
      1. **F-statistic (45.8295):** Această statistică măsoară semnificația generală a regresiei, adică dacă cel puțin una dintre variabilele independente contribuie semnificativ la explicarea variației în variabila dependentă.
      2. **Prob(F-statistic) = 0.0000 < 0.05:** Probabilitatea asociată F-statisticii este foarte mică, indicând că există suficiente dovezi pentru a respinge ipoteza nulă că toți coeficienții de regresie sunt zero. Acest rezultat susține semnificația generală a regresiei. => acceptăm H1.
2. **R2 = 0,9604**
   1. 96,04% din variația lui Y este explicată prin variația lui și
3. **Testul T - variabilele sunt semnificative**:
   1. H0: Coeficientul asociat cu variabila independentă este zero (adică variabila independentă nu are niciun efect semnificativ asupra variabilei dependente).
   2. H1: Coeficientul asociat cu variabila independentă nu este zero (adică variabila independentă are un efect semnificativ asupra variabilei dependente).
   3. (=> nu respingem H0
   4. => respingem H0 => acceptăm H1
4. **Testul Durbin-Watson:**
   1. H0: Nu există autocorelare în reziduurile modelului.
   2. H1: Există autocorelare în reziduurile modelului.
   3. DW = 2,9887 => Valoarea statistică Durbin-Watson (DW) de 2,9887 indică că există o ușoară indicație a autocorelației în reziduurile modelului de regresie. Această valoare DW este aproape de 4, ceea ce sugerează o posibilă autocorelație negativă în reziduuri. Autocorelația negativă ar putea indica faptul că valorile reziduurilor din model sunt corelate în sens opus între perioadele succesive.

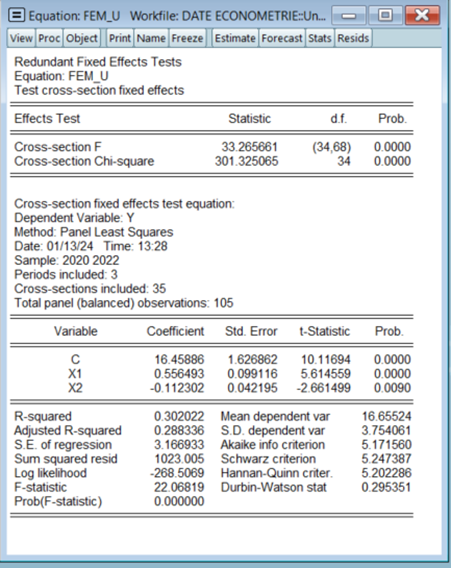
Testarea pentru efecte fixe

Testarea pentru efecte fixe reprezintă o metodă esențială în analiza datelor de tip panel, unde aceleași unități individuale sunt urmărite pe parcursul unui interval de timp. Modelul cu efecte fixe are ca principală caracteristică includerea unor termeni fixi specifici fiecărei unități individuale în regresie, capturând astfel variația constantă și specifică a acestora. Scopul acestui model este de a controla și cuantifica influența factorilor individuali constanți asupra variabilei dependente, permițând în același timp analiza variației între unități. Testarea pentru efecte fixe se concentrează asupra evaluării semnificației acestor efecte individuale, stabilind dacă există variații semnificative între unități și contribuind astfel la înțelegerea mai profundă a determinanților specifici care influențează variabila de interes în contextul datelor longitudinale.



* H0: nu există FE (FE sunt redundante, inutile) Este potrivită metoda OLS.
* H1: există FE (FE nu sunt redundante, sunt necesare) Este potrivit FEM.

Decizie: Probabilitatea asociată a celor 2 teste (Testele Cross-section F și Cross-section Chi-square) este 0,0000 <0,05 => acceptăm H1.



# Bibliografie

DW (2023) ‘Copiii din România, cei mai expuși sărăciei’, Societate, Euroopa, <https://www.dw.com/ro/copiii-din-rom%C3%A2nia-cei-mai-expu%C8%99i-s%C4%83r%C4%83ciei-dintre-copiii-europeni/a-64908444>

EuroStat (2020) ‘Statistici privind sărăcia veniturilor’, <https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=469045>

EuroStat (2021) ‘Statistici privind sărăcia veniturilor’, <https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=526266>

Actualitate Parlamentul European (2021) ‘Fondul social european: combaterea sărăciei și a șomajului’, <https://www.europarl.europa.eu/news/ro/headlines/society/20181129STO20519/fondul-social-european-combaterea-saraciei-si-a-somajului>

Nordmedia (2023) ‘STUDIU. În 2023, numărul românilor expuși riscului de sărăcie va crește. Persoanele cu venituri mai mici resimt mai tare creșterea prețurilor’, Economie, <https://www.normedia.ro/2023/01/28/studiu-ramani-saracie-2023-f34r/>

Eurostat (2023) ‘People at risk of poverty or social exclusion in 2022’, News Articles, <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/DDN-20230614-1>

Directorate-General for Employment, Social Affairs and Inclusion (European Commission) (2023). ‘Employment and social developments in Europe 2023’, Social policy, <https://op.europa.eu/en/publication-detail/-/publication/680d6391-2142-11ee-94cb-01aa75ed71a1>

Eurostat (2023) ‘Subjective poverty highest among lower educated’, News Articles, <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20231120-1>

Sudesh Baniya (2023) ‘Europe's labour problem: More than 27 million unemployed or underemployed in the EU in 2022’, Work, <https://www.euronews.com/next/2023/05/24/europes-labour-problem-more-than-27-million-unemployed-or-underemployed-in-the-eu-in-2022>

Harms, M.B., Garrett-Ruffin, S.D. (2023) ‘Disrupting links between poverty, chronic stress, and educational inequality’, npj Sci. Learn, <https://www.nature.com/articles/s41539-023-00199-2#citeas>

Eurostat ‘At-risk-of-poverty rate’, <https://ec.europa.eu/eurostat/databrowser/view/tespm010__custom_9304785/default/table?lang=en>

Eurostat ‘Population by educational attainment level, sex and age (%)’, <https://ec.europa.eu/eurostat/databrowser/view/edat_lfs_9903__custom_9305300/default/table?lang=en>

Eurostat ‘Total unemployment rate’, <https://ec.europa.eu/eurostat/databrowser/view/tps00203__custom_9304562/default/table?lang=en>