**ACADEMIA DE STUDII ECONOMICE BUCUREȘTI**

**FACULTATEA DE CIBERNETICĂ, STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ**

**Proiect pentru**

**Econometrie**

**Studierea fenomenului de mortalitate infantilă**

**Profesor coordonator: LECT. UNIV. DR. VASILESCU MARIA DENISA**

**Student: IFTIMIE ALEXANDRA-MARIA**

**Grupa: 1039**

**2020-2021**

**CUPRINS**

**1. Prezentarea problemei........................................................................................................................2**

**• Descrierea fenomenului analizat și a factorilor explicativi considerați. Explicarea legăturii din punct de vedere economic.......................................................................................................................2**

**• Tabelul cu datele, sursa datelor............................................................................................................3**

**2. Definirea modelului de regresie..........................................................................................................8**

**• Forma, variabilele și parametri modelului de regresie. Aproximarea grafică a modelului legăturii dintre variabile.........................................................................................................................8**

**3. Analiza preliminară a datelor.............................................................................................................9**

**• Verificarea preliminară a datelor........................................................................................................9**

**• Factorii independenți (ipoteza de multicoliniaritate).......................................................................10**

**4. Estimarea modelului de regresie......................................................................................................10**

**• Estimarea punctuală a parametrilor și interpretarea rezultatelor.................................................10**

**• Estimarea parametrilor prin interval de încredere..........................................................................11**

**• Testarea parametrilor modelului de regresie....................................................................................12**

**• Testarea validității modelului............................................................................................................13**

**• Interpretarea coeficientului de determinație....................................................................................13**

**5. Testarea ipotezelor clasice asupra modelului de regresie...............................................................13**

**• Testarea normalității erorilor............................................................................................................13**

**• Testarea ipotezei de homoscedasticitate............................................................................................14**

**• Testarea ipotezei de autocorelare a erorilor......................................................................................15**

**• Corectarea eventualelor probleme identificate.................................................................................16**

**6. Previziunea valorii variabilei Y în ipoteza modificării variabilelor factoriale..............................18**

**7. Concluzii............................................................................................................................................18**

**Bibliografie............................................................................................................................................19**

**1. Prezentarea problemei**

• *Descrierea fenomenului analizat și a factorilor explicativi considerați. Explicarea*

*legăturii din punct de vedere economic.*

Mortalitatea infantilă este de obicei definită ca numărul de decese ale copiilor mai mici de un an într-un anumit an la o mie de copii din această grupă de vârstă.

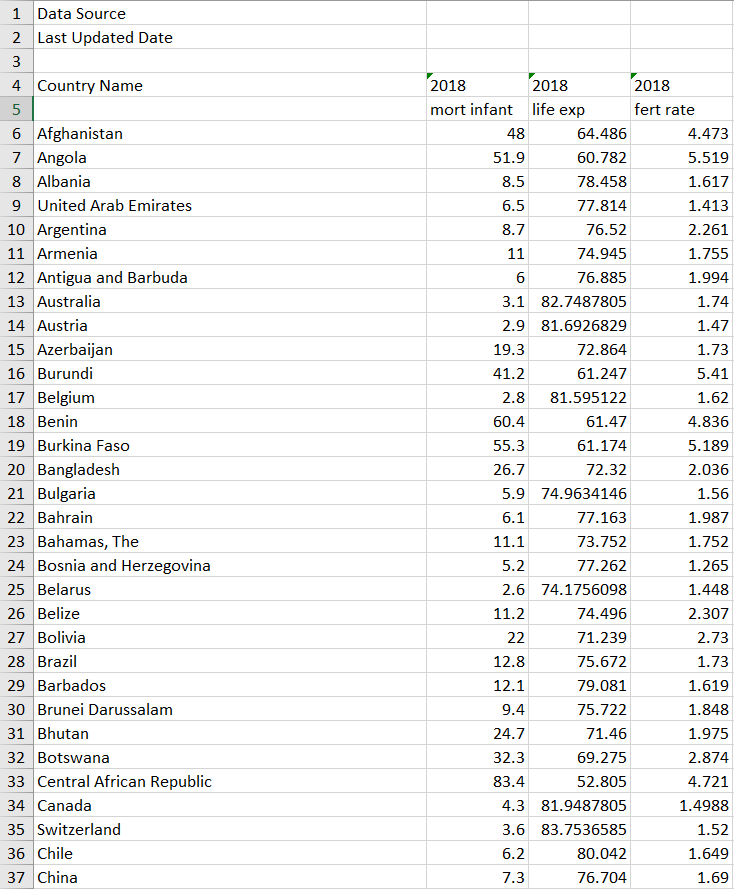
Utilitatea mortalității infantile ca indicator de sănătate depinde de contextul populației în care este utilizată. De exemplu, într-o țară dezvoltată, în care principalele cauze globale ale mortalității infantile nu sunt la fel de răspândite, ratele cauzelor specifice ale mortalității infantile ar putea fi utilizate în prioritizarea programării de prevenire (de exemplu, orientată către siguranța pasagerilor autovehiculelor). Cu toate acestea, într-o țară în curs de dezvoltare, ratele mortalității infantile pot juca un rol mai important ca indicator al sănătății la nivel general și al problemelor sociale și de mediu, cum ar fi malnutriția, salubritatea apei, sărăcia și accesul la sistemele de sănătate.

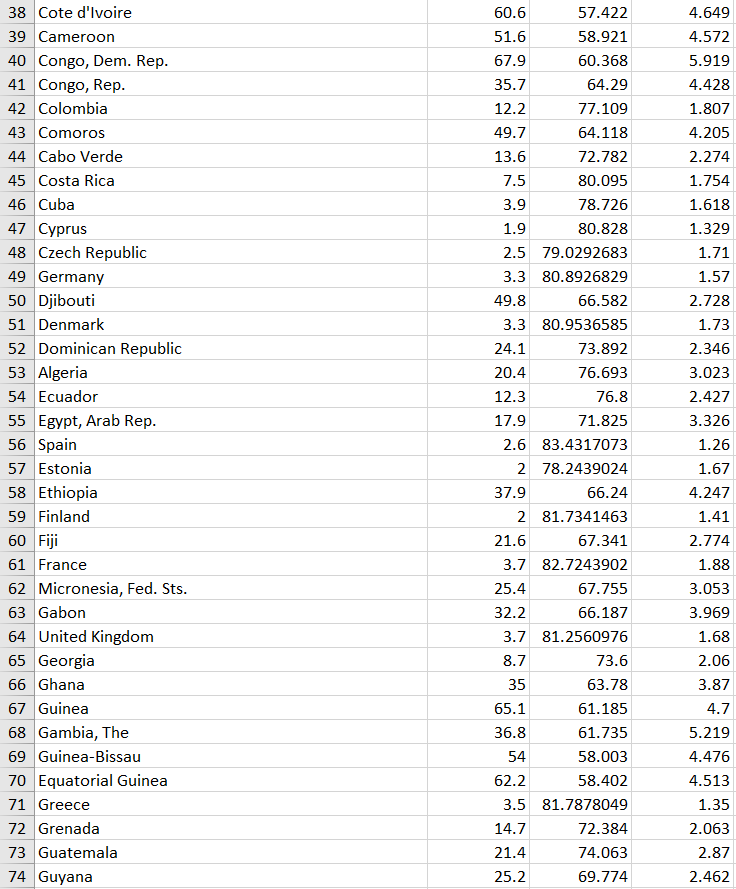
Lumea a făcut progrese remarcabile în supraviețuirea copiilor în ultimele trei decenii, iar milioane de copii au șanse de supraviețuire mai bune decât în 1990 — 1 din 27 de copii a murit înainte de a ajunge la vârsta de cinci ani în 2019, comparativ cu 1 din 11 în 1990. În plus, progresele înregistrate în reducerea ratei mortalității infantile au fost accelerate în perioada 2000-2019 comparativ cu anii 1990, rata anuală de reducere a ratei mortalității la nivel mondial sub cinci ani crescând de la 1,9 % în 1990-1999 la 3,7 % în 2000-2019. În ciuda progreselor globale în ceea ce privește reducerea ratei mortalității infantile în ultimele decenii, se estimează că 5,2 milioane de copii cu vârsta sub cinci ani au murit în 2019 — mai mult de jumătate dintre aceste decese au avut loc în Africa Subsahariană.

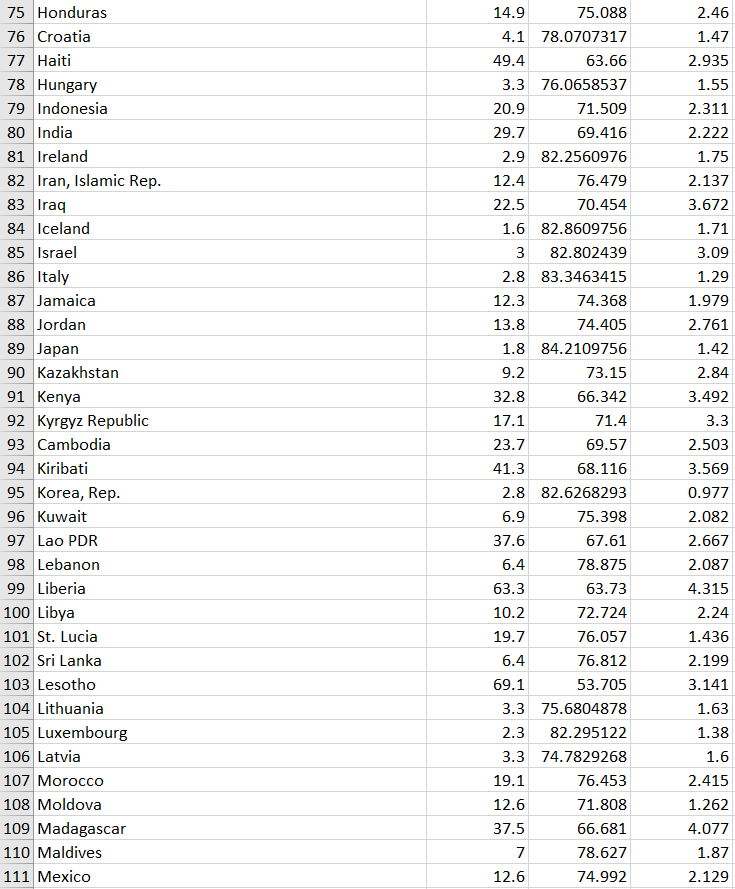
Pentru a analiza acest fenomen, mortalitatea infantilă, voi propune ca factori independenți rata fertilității și speranța de viață la naștere. Conform The World Bank, de unde voi procura datele studiate, *rata mortalității infantile* reprezintă numărul de sugari care mor înainte de a ajunge la vârsta de un an, la 1.000 de născuți vii într-un anumit an, în timp ce *rata totală a fertilității* reprezintă numărul de copii pe care i-ar naște o femeie dacă ar trăi până la sfârșitul anilor fertili și ar avea copii în conformitate cu ratele de fertilitate specifice vârstei din anul specificat, iar *speranța de viață* la naștere indică numărul de ani pe care un nou-născut i-ar trăi dacă modelele predominante de mortalitate la momentul nașterii sale ar rămâne aceleași pe tot parcursul vieții sale (link-urile se află în secțiunea de *Bibliografie*). Consider că rata fertilității influențează rata mortalității infantile prin adaos, posibilitatea să moară mai mulți copii este cu atât mai mare cu cât este mai mare șansa să se nască mai mulți copii, în timp ce o speranță de viață mică evidențază unele condiții de trai nu atât de favorabile (precum sărăcia, lipsa apei potabile, malnutriția, lipsa îngrijirilor medicale, războaie, etc...), ceea ce ar putea amenința viața nou-născutului.

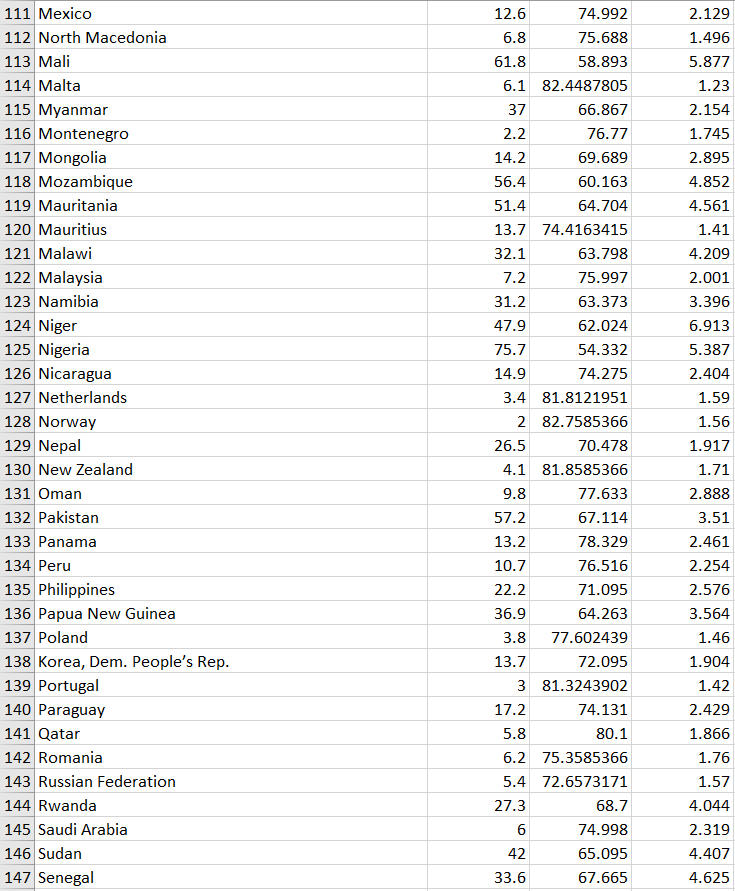
*• Tabelul cu datele*

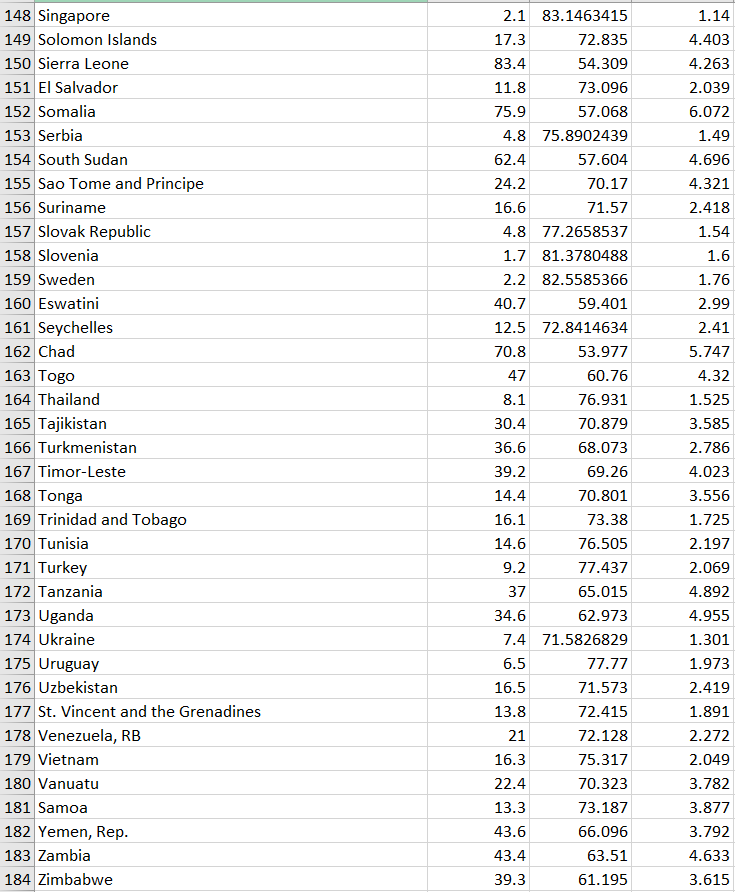
Datele pe care îmi propun să le analizez sunt la nivel mondial, din anul 2018.

Figura 1a, Tabelul cu datele analizate, după eliminarea valorilor cumulative și a entităților care nu au furnizat date.

 Figura 1b, Tabelul cu datele analizate, după eliminarea valorilor cumulative și a entităților care nu au furnizat date.

 Figura 1c, Tabelul cu datele analizate, după eliminarea valorilor cumulative și a entităților care nu au furnizat date.

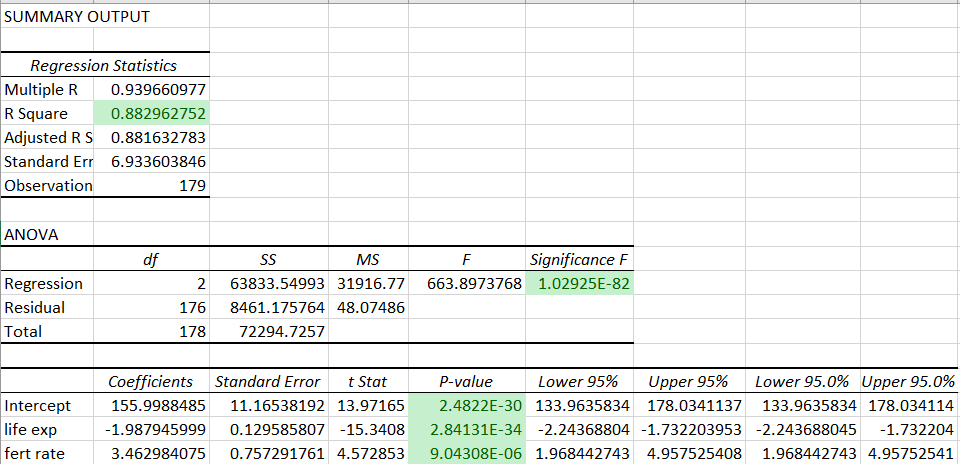
 Figura 1d, Tabelul cu datele analizate, după eliminarea valorilor cumulative și a entităților care nu au furnizat date.

 Figura 1e, Tabelul cu datele analizate, după eliminarea valorilor cumulative și a entităților care nu au furnizat date.

**2. Definirea modelului de regresie**

*• Forma, variabilele și parametrii modelului de regresie. Aproximarea grafică a modelului legăturii dintre variabile*

Pentru datele enumerate anterior, tabelul de regresie este următorul:

Figura 2, Tabelul de regresie al datelor analizate.

Analizând Figura 2, se poate observa faptul că atât modelul, cât și parametri sunt valizi, deoarece *Significance F,* respectiv *P-value* au valori mai mici decât nivelul de semnificație ales, 0,05. De asemenea, modelul de regresie ales explică în proporție de 88,29% fenomenul studiat.

Forma modelului de regresie:

**Y=155,9988485-1,987945999\*X1+3,462984075\*X2**

Interpretarea coeficienților modelului de regresie:

* Termenul liber are valoarea 155.9988485 și semnifică valoarea mortalității infantile atunci când valorile speranței de viață și a ratei fertilității sunt egale cu 0. Cum acest lucru nu este posibil din punct de vedere economic, putem afirma faptul că termenul liber nu are interpretare economică;
* Coeficientul speranței de viață este egal cu -1.987945999 și se explică prin faptul că la o creștere cu 1 an a speranței de viață, rata mortalității infantile va scădea cu aproximativ 2 puncte procentuale (ceteris paribus). Siguranța economică, îmbunatățiri în domeniul medical si ușurința accesului la ele sunt doar câteva dintre cauzele ce pot influența mortalitatea să scadă și speranța de viață să crească.

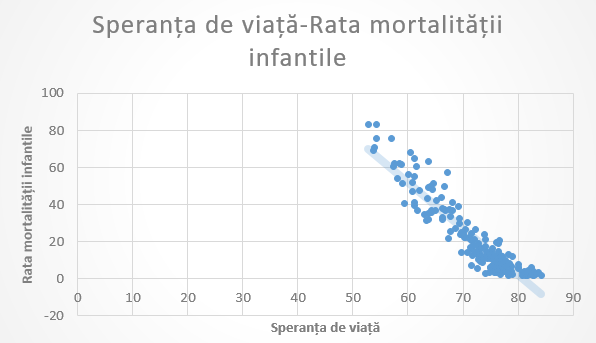


Figura 3, Evidențierea relației indirecte între speranța de viață și mortalitatea infantilă.

* Coeficientul ratei fertilității este egal cu 3.462984075 și se explică prin faptul că la o creștere cu un procent a ratei fertilității, rata mortalității infantile va crește cu aproximativ 3,46 puncte procentuale (ceteris paribus). Acest aspect se poate explica prin faptul că posibilitatea să moară mai mulți copii este cu atât mai mare cu cât este mai mare șansa să se nască mai mulți copii.

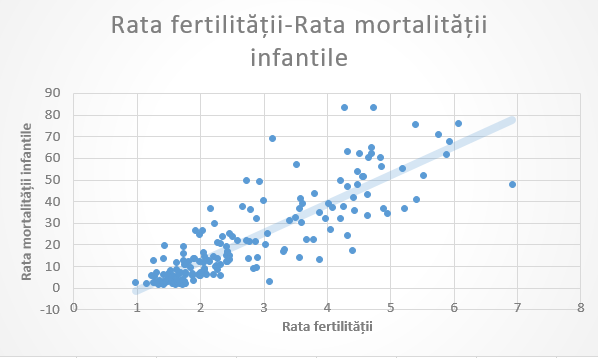


Figura 4, Evidențierea relației directe între rata fertilității și mortalitatea infantilă.

**3. Analiza preliminară a datelor**

*• Verificarea preliminară a datelor*

Înainte de începerea analizei modelului de regresie, am analizat datele studiate pentru a observa posibilele valori outlier, iar în acest scop am aplicat formula *medie±3\*abatare standard*  statisticilor descriptive ale seriilor de date analizate.

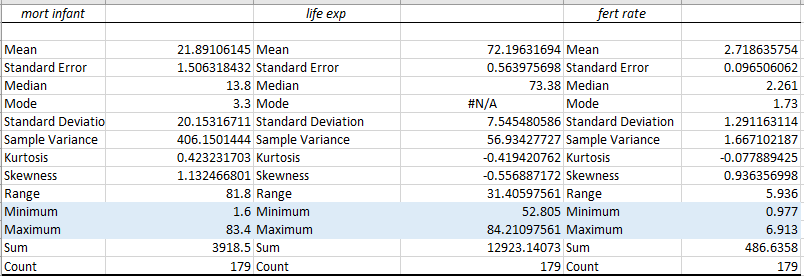


Figura 5, Statisticile descriptive ale celor 3 serii de date analizate, mortalitatea infantilă, speranța de viață și rata fertilității(de la stanga la dreapta)

Aplicând formula mai sus menționată, au rezultat urmatoarele intervale:



Figura 6, Intervale pentru mortalitatea infantilă (albastru), speranța de viață (galben) și rata fertilității (verde)

Corelând Figura 1 cu Figura 5 și datele inițiale, putem afirma faptul că pentru variabila dependentă mortalitate infantilă avem două valori extreme, corespunzătoare Republicii Centrafricană și Sierra Leone cu o rată a mortalității de 83,4%, pentru variabila independentă rata fertilității există un singur outlier, corespunzător Nigeriei, cu o rată a fertilității de 6,913%, în timp ce seria de date ce descrie speranța de viață nu prezintă valori extreme. Voi continua analiza modelului de regresie utilizând și valorile outlier, dar voi ține cont de acestea la posibilitatea corectării eventualelor probleme identificate la analiza erorilor modelului.

Analizând Figura 5, putem afirma următoarele:

* Pentru setul de date studiat, valoarea medie a mortalității infantile a fost de 21,89%, variind de la medie cu o abatere de 20,15 puncte procentuale;
* Valoarea medie a speranței de viață a fost 72 de ani, fluctuând de la medie cu valori cuprinse între 52 și 84 de ani;
* Media ratei fertilității a fost de 2,71%, iar diferența dintre cea mai mare și cea mai mica valoare a fost de 5,93 puncte procentuale.

*• Factorii independenți (ipoteza de multicoliniaritate)*

Pentru a determina dacă există multicoliniaritate între variabilele independente studiate, vom aplica criteriul lui Klein:

**Pasul 1.** Se estimează modelul de regresie și se pastrează valoarea coeficientului de determinație. Valoarea coeficientului este egala cu 0.88 (vezi Figura 2).

**Pasul 2.** Se determina matricea de corelație pentru variabilele explicative.

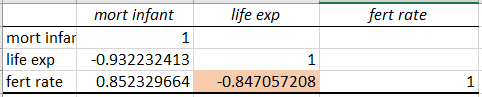


Figura 7, Matricea de corelație a variabilelor studiate.

**Pasul 3**. Daca |coef de corel| > R2, înseamnă că există multicoliniaritate.

**Decizia:** Cum coeficientul de corelație între variabilele explicative are o valoare mai mică decât coeficientul de determinație, putem afirma faptul că nu există multicoliniaritate între variabilele explicative studiate.

**4. Estimarea modelului de regresie**

*• Estimarea punctuală a parametrilor și interpretarea rezultatelor*

Pentru estimarea punctuală a parametrilor, vom contrui matricea B ce are drept componente valorile parametrilor:

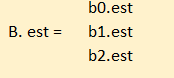


Figura 8, Matricea B, ce are în compenența sa valorile estimate ale parametrilor modelului de regresie

Știm de asemenea faptul că matricea B este egală cu



Figura 9, Formula de calcul a matricei B

În urma calculelor, a rezultat următoarea matrice:

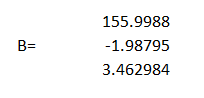


Figura 10, Matricea B, ce are în compenența sa valorile estimate ale parametrilor modelului de regresie

După cum se poate observa, în urma estimării parametrilor au rezultat aceleași valori ca cele identificate în tabelul de output al regresiei, ceea ce poate fi considerat o verificare a modului de calcul (a se vedea Figura 2).

Interpretarile parametrilor sunt cele afirmate mai sus. Pentru continuitatea lecturii, le voi reaminti:

* Termenul liber are valoarea 155.9988485 și semnifică valoarea mortalității infantile atunci când valorile speranței de viață și a ratei fertilității sunt egale cu 0. Cum acest lucru nu este posibil din punct de vedere economic, putem afirma faptul că termenul liber nu are interpretare economică;
* Coeficientul speranței de viață este egal cu -1.987945999 și se explică prin faptul că la o creștere cu 1 an a speranței de viață, rata mortalității infantile va scădea cu aproximativ 2 puncte procentuale (ceteris paribus).
* Coeficientul ratei fertilității este egal cu 3.462984075 și se explică prin faptul că la o creștere cu un procent a ratei fertilității, rata mortalității infantile va crește cu aproximativ 3,46 puncte procentuale (ceteris paribus).

*• Estimarea parametrilor prin interval de încredere*

Vom utiliza formula pentru estimarea pe interval de încredere a parametrilor:

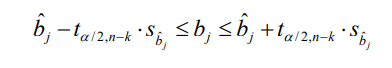


Figura 11, Formula de calcul pentru estimarea pe bază de interval de încredere a parametrilor modelului de regresie

În urma calculelor necesare, au rezultat urmatoarele, pentru un nivel de semnificație de 0.05:

* Termenul liber este cuprins între 133.9635834 și 178.0341137.
* Coeficientul variabilei independente speranța de viață este cuprins între -2.243688045 și -1.73220395.
* Coeficientul variabilei independente rata fertilității este cuprins între 1.968442743 și 4.957525408.

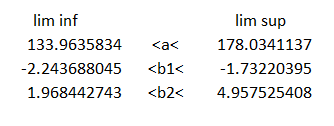


Figura 12, Estimarea pe bază de interval de încredere a parametrilor modelului de regresie pentru un nivel de semnificație de 0.05

De asemenea, aceste valori se regăsesc și în tabelul output al modelului de regresie (vezi Figura2).

*• Testarea parametrilor modelului de regresie*

Pentru a testa parametri modelului de regresie, vom utiliza testul t:

**Pasul 1**: Ipotezele: H0: 𝑏𝑗 = 0 (parametrul nu este semnificativ diferit de 0)

H1: 𝑏𝑗 ≠ 0, cu j = 0, 1 sau 2 (parametrul este semnificativ diferit de 0)

**Pasul 2**: Calcularea valorii statisticii t, tcalc, dupa formula:

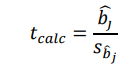


Figura 13, Formula de calcul a testului statistic

Pentru a calcula abaterile fiecărui parametru, vom calcula matricea Ω. În urma calculelor, au rezulat următoarele:

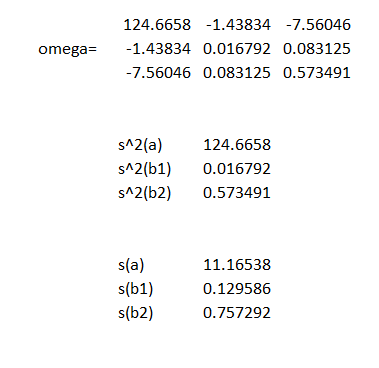


Figura 14, Matricea Ω, dispersiile și abaterile parametrilor modelului de regresie

Astfel, tcalc pentru fiecare dintre parametri este:

* Pentru parametrul variabilei dependente Y: 13.97165;
* Pentru parametrul variabilei speranța de viață: -15.3408;
* Pentru parametrul variabilei rata fertilității: 4.572853.

De asemenea, aceste valori se regăsesc și în tabelul output al modelului de regresie (vezi Figura 2).

**Pasul 3**: Valoarea critica tcritic­=1.97.

**Pasul 4**: Decizia: Întrucât toare valorile absolute ale testului statistic calculate sunt mai mari decât valoarea critică, vom respinge H0 și vom accepta H1. De asemenea, un alt mod de a testa parametri modelului de regresie este verificarea valorii p-value a fiecăruia dintre ei, prezentă în tabelul output al modelului de regresie (vezi Figura 2). Întrucât toate valorile p-value pentru toți parametri sunt mai mici decat 0,05, putem afirma faptul că parametri sunt semnificativi din punct de vedere statistic.

*• Testarea validitatii modelului*

Pentru a testa validitatea modelului, vom aplica testul F:

**Pasul 1:** Ipotezele: H0: Modelul nu este valid.

H1: Modelul este valid.

**Pasul 2**: Testul statistic:

Calculăm valoarea testului: Fcalc===663.8974.

**Pasul 3**: Valoarea critică:

Fcritic depinde de α și de (k-1,n-k) grade de libertate

Fcritic =3.047307

**Pasul 4:** Decizia statistica: Dacă valoarea Fcalc este mai mare decât valoarea Fcritic , atunci respingem H0 și acceptăm H1.

Concluzie: Acceptăm H1, modelul este valid. De asemenea, un alt mod de a testa validitatea modelului este de a cerceta valoarea lui Significance F (vezi Figura 2). Întrucât aceasta este mai mică decât valoarea nivelului de semnificație ales, 0,05, putem afirma faptul că modelul este unul valid.

*• Interpretarea coeficientului de determinație*

Coeficientul de determinație reprezintă R Square din Figura 2 și este egal cu 0,882963, ceea ce semnifică faptul că rata fertilității și speranța de viață explică fenomenul mortalității infantile în proporție de 88,29%.

**5. Testarea ipotezelor clasice asupra modelului de regresie**

*• Testarea normalității erorilor*

Pentru testarea normalității erorilor vom utiliza testul Jarque-Bera:

**Pasul 1**: Ipotezele: H0 – erorile sunt normal distribuite.

H1 – erorile nu sunt normal distribuite.

**Pasul 2**: Valoarea statisticii JB se calculează cu formula:

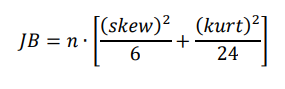


Figura 15, Formula de calcul pentru testul Jarque-Bera

Vom calcula statisticile descriptive ale erorilor pentru a găsi valoarea coeficienților de boltire și asimetrie.

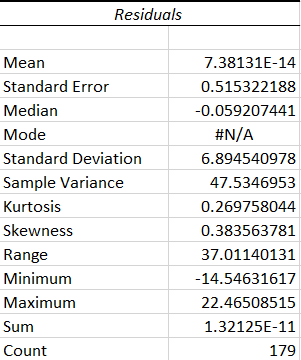


Figura 16, Statisticile descriptive ale erorilor

Valoarea testului Jarque-Bera=JB=4.931853493.

**Pasul 3**: Valoarea calculată a testului JB urmează o repartiţie χ2 cu 2 grade de libertate.

JBcritic=5.991464547

**Pasul 4**. Decizia: Dacă valoarea testului JB calculată este inferioară valorii tabelare χ2α;2 vom accepta ipoteza H0 (erorile sunt normal distribuite).

**Concluzie**: întrucât valoarea calculată a testului JB este mai mică decât valoarea critică, vom accepta ipotea H0, conform căreia erorile sunt normal distribuite.

*• Testarea ipotezei de homoscedasticitate*

Pentru verificarea ipotezei de homoscedasticitate vom utiliza testul White. Din considerente de spațiu, voi ilustra doar tabelul regresiei auxiliare, fără valorile variabilelor.

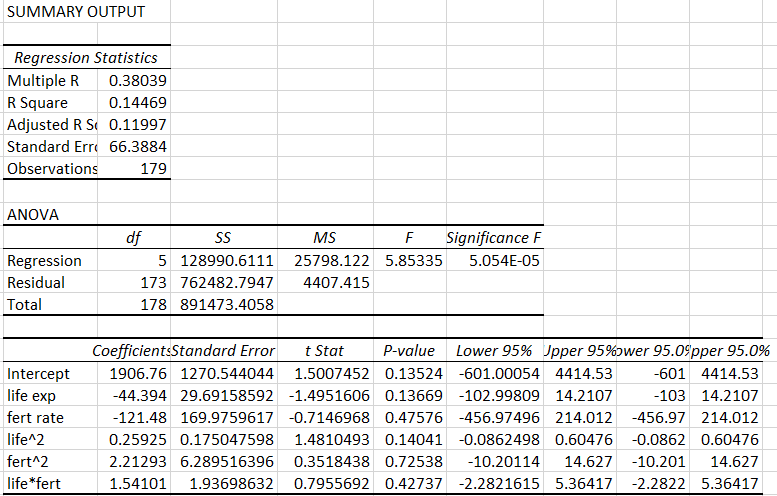


Figura 17, Tabelul output al regresiei auxiliare

Ipotezele testului White:

|  |
| --- |
| H0: Erorile sunt homoscedastice. |
| H1: Erorile sunt heteroscedastice.  Calculăm valoarea testului White, utilizând formula Whitecalc=n\*R2a, unde R2a  reprezintă coeficientul de determinație al modelului auxiliar de regresie.  Whitecalc=25.90018  Valoarea testului White urmează o repartiţie χ2 cu df grade de libertate. În cazul modelului de regresie auxiliar, df=5.  Valoarea critică=11.0705  Concluzie: întrucât valoarea calculată a testului White este mai mare decât valoarea critică, vom accepta ipotea H1, conform căreia erorile sunt heteroscedastice. |

*• Testarea ipotezei de autocorelare a erorilor*

Vom utiliza testul Durbin-Watson pentru a depista dacă erorile sunt autocorelate sau nu.

Ipotezele:

H0: Erorile nu sunt autocorelate.

H1: Erorile sunt autocorelate.

Statistica testului:

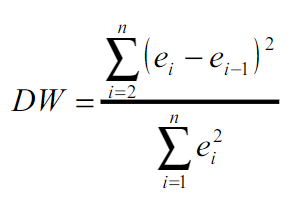


Figura 18, Statistica testului Durbin-Watson

Valoarea testului Durbin-Watson este 2.179585.

Calculam coeficientul de autocorelatie=1-= -0.08979.

Vom considera valorile critice d1 1.706 si d2 1.760.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | autocorelare |  | indecizie |  | independență |  | independență |  | indecizie |  | autocorelare |  |
| 0 | + | d1 | <- | d2 |  | 2 |  | 4- d2 | -> | 4-d1 | - | 4 |
| 0 |  | 1,706 |  | 1,760 |  | 2 | DWcalc | 2,24 |  | 2,294 |  | 4 |

Figura 19, Tabelul cu intervalele testului Durbin-Watson

Analizând tabelul cu intervalele testului Durbin-Watson, putem afirma că erorile nu sunt autocorelate.

*• Corectarea eventualelor probleme identificate*

Am observat faptul că erorile modelului de regresie sunt heteroscedastice. Vom încerca să remediem acest lucru, prin eliminarea valorilor extreme, apoi prin logaritmarea variabilelor.

Avem 3 valori extreme, Nigeria, Sierra Leone și Republica Centrafricană.

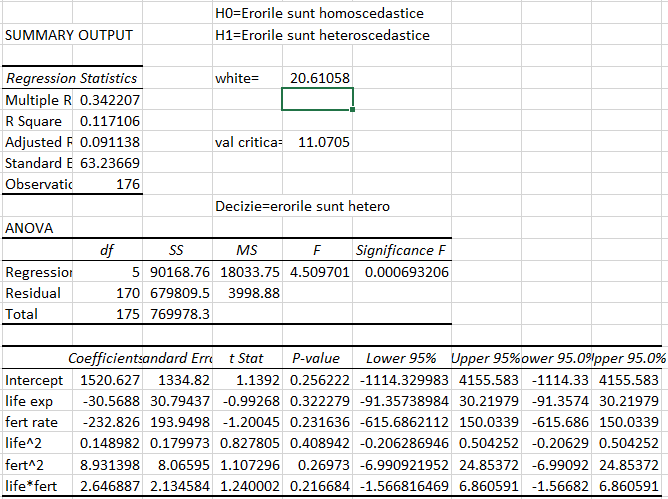


Figura 20, Tabelul output al regresiei auxiliare, împreună cu valoarea noului test White, după eliminarea valorilor outliers.

După eliminarea valorilor extreme, noua valoare a testului White este 20,61058. Întrucât este mai mare decât valoarea critică, vom accepta ipoteza H1­, conform căreia erorile sunt heteroscedastice.

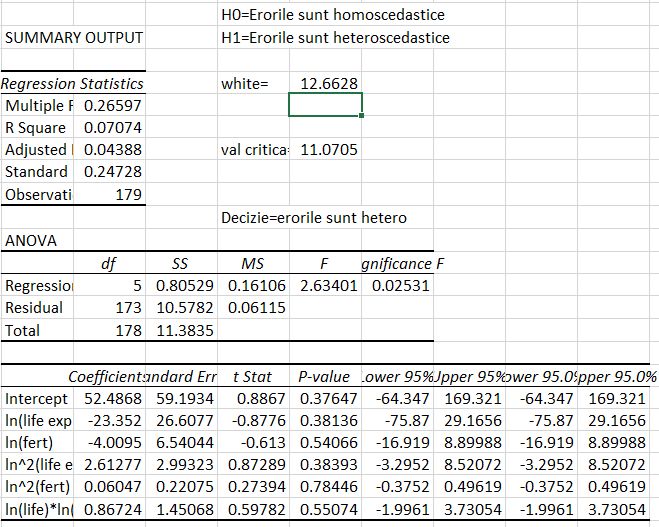


Figura 21, Tabelul output al regresiei auxiliare, împreună cu valoarea noului test White, după logaritmarea variabilelor.

După logaritmarea variabilelor, noua valoare a testului White este 12,6628. Întrucât este mai mare decât valoarea critică, vom accepta ipoteza H1­, conform căreia erorile sunt heteroscedastice.

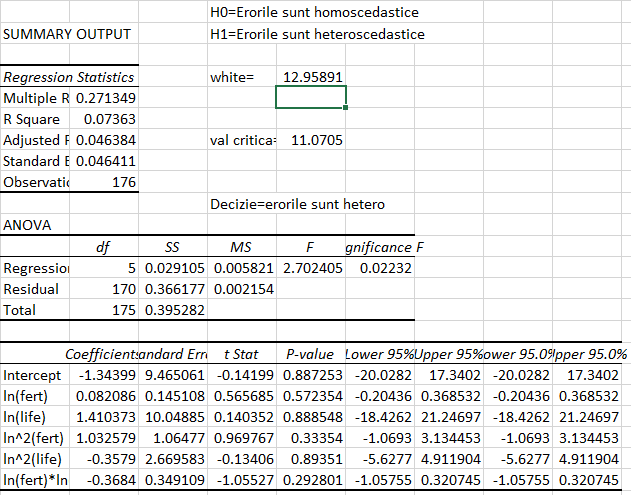


Figura 22, Tabelul output al regresiei auxiliare, împreună cu valoarea noului test White, după eliminarea valorilor outliers și logaritmarea variabilelor.

După eliminarea valorilor extreme și logaritmarea variabilelor, noua valoare a testului White este 12,95891. Întrucât este mai mare decât valoarea critică, vom accepta ipoteza H1­, conform căreia erorile sunt heteroscedastice.

**6. Previziunea valorii variabilei Y în ipoteza modificării variabilelor factoriale**

Pentru a face o previziune punctuală a valorii variabilei Y, voi lua în ipoteză pentru variabila rata fertilității valoarea 5% iar pentru speranța de viață valoarea 85 de ani.

Aplicând ecuația de regresie pe valorile propuse, opținem o valoare de 4.338359 pentru variabila Y, ceea ce semnifică faptul că într-o țară cu o rată a fertilității egală cu 5 și o speranță de viață de 85 de ani, mortalitatea infantilă va avea o valoare previzionată de 4,338359.

**7. Concluzii**

Prin utilizarea regresiei liniare multifactoriale, am descoperit existenţa unei corelaţii puternice între rata fertilității și speranța de viață, ceea ce explică fenomenul de mortalitate infantilă în proporție de 88%.

Ca o concluzie generală afirmăm că modelul econometric de regresie liniară este eficient şi rezultatul analizei oferă suficientă încredere. Cu toate acestea, un dezavantaj al modelului îl reprezintă heteroscedasticitatea erorilor, ce nu a putut fi corectată.

**Bibliografie**

* <https://www.encyclopedia.com/education/encyclopedias-almanacs-transcripts-and-maps/child-mortality>
* <https://data.unicef.org/topic/child-survival/under-five-mortality/>

Sursa datelor:

* <https://data.worldbank.org/indicator/SH.DYN.NMRT?view=chart>
* <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?view=chart>
* <https://data.worldbank.org/indicator/SP.DYN.LE00.IN?view=chart>