# **ACADEMIA DE STUDII ECONOMICE**

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

# **SPECIALIZAREA INFORMATICĂ ECONOMICĂ**

# 

# 

# 

# **PROIECT ECONOMETRIE**

# **CALITATEA VINULUI ȘI FERICIREA**

# 

# 

# 

# **Profesor coordonator: Studenți:**

# **Prof. Univ. Dr. Alexandru Adriana Ana Maria**

**Catina Ada Maria**

# 

# **Bucuresti, 2022**

# **Cuprins**

[**Introducere** 5](#_Toc125208758)

[**APLICAŢIA 1** 7](#_Toc125208759)

[***Model de regresie simplă*** *7*](#_Toc125208760)

[*1.1.1. Literature review 7*](#_Toc125208761)

[*1.2.1. Date utilizate 9*](#_Toc125208762)

[*1.2.2. Rezultatele empirice ale cercetării 10*](#_Toc125208763)

[*1.3. Concluzii 19*](#_Toc125208764)

[**APLICAŢIA 2** 21](#_Toc125208765)

[***Model de regresie multiplă*** *21*](#_Toc125208766)

[*2.1.1. Literature review 21*](#_Toc125208767)

[*2.1.2. Metodologia cercetării 22*](#_Toc125208768)

[*2.2.1.2. Rezultatele empirice ale cercetării 23*](#_Toc125208769)

[*2.2.2.1. Metodologia cercetării 31*](#_Toc125208770)

[*2.2.2.2. Rezultatele empirice ale cercetării 32*](#_Toc125208771)

[*2.3. Concluzii 40*](#_Toc125208772)

[**APLICAŢIA 3** 41](#_Toc125208773)

[***Model de date de tip panel*** *41*](#_Toc125208774)

[*3.1.1. Literature review 41*](#_Toc125208775)

[*3.1.2. Metodologia cercetării 42*](#_Toc125208776)

[*3.2.2. Rezultatele empirice ale cercetării 43*](#_Toc125208777)

[*3.3. Concluzii 50*](#_Toc125208778)

[***Bibliografie*** *51*](#_Toc125208779)

***Lista Tabele***

[Table 1.Principalele statistici descriptive ale indicilor 10](#_Toc125209488)

[Table 2.Modelul de regresie liniara 11](#_Toc125209489)

[Table 3. Testul Breusch-Pagan 14](#_Toc125209490)

[Table 4. Testul Breusch-Pagan corectat 14](#_Toc125209491)

[Table 5. Model Test White 15](#_Toc125209492)

[Table 6. Testul White 15](#_Toc125209493)

[Table 7. Testul White corectat 16](#_Toc125209494)

[Table 8. Testul Durbin-Watson 17](#_Toc125209495)

[Table 9. Model Test Durbin-Watson 17](#_Toc125209496)

[Table 10. Testul Durbin-Watson corectat 17](#_Toc125209497)

[Table 11.Testul Breusch-Godfrey 18](#_Toc125209498)

[Table 12. Testul Breusch-Godfrey corectat 18](#_Toc125209499)

[Table 13. Testul Jarque-Bera 18](#_Toc125209500)

[Table 14. Testul Jarque-Bera corectat 19](#_Toc125209501)

[Table 15.Modelul de regresie simpla final 20](#_Toc125209502)

[Table 16.Principalele statistici descriptive - regresie multipla 23](#_Toc125209503)

[Table 17.Regresie multifactorială între alcool, sulfați și acid volatil 24](#_Toc125209504)

[Table 18.Verificare nr obs > nr variabile independente 26](#_Toc125209505)

[Table 19.Variabilitatea in x 26](#_Toc125209506)

[Table 20.Media rezidurilor 27](#_Toc125209507)

[Table 21.VIF 27](file:///C:\Users\Ada\Desktop\1081_Carp_Catina.docx#_Toc125209508)

[Table 22.Testul Breusch-Pagan multiplu 27](#_Toc125209509)

[Table 23.Testul White multiplu 28](#_Toc125209510)

[Table 24.Testul Durbin-Watson multiplu 28](#_Toc125209511)

[Table 25.Testul Durbin-Watson multiplu corectat 29](#_Toc125209512)

[Table 26.Testul Breusch-Godfrey multiplu 29](#_Toc125209513)

[Table 27.Testul Breusch-Godfrey multiplu de ordin I și II corectat 29](#_Toc125209514)

[Table 28. Testul Jarque-Bera multiplu 30](#_Toc125209515)

[Table 29.Testul Jarque-Bera multiplu corectat 31](#_Toc125209516)

[Table 30. Tabelul modelului de interacțiune a regresiei multiple 33](#_Toc125209517)

[Table 31.Prefictia calitatii pentru 3 observatii 35](#_Toc125209518)

[Table 32.Modelul de regresie pentru calitatea vinului 36](#_Toc125209519)

[Table 33.Testul Reset 36](#_Toc125209520)

[Table 34. Testul CHOW 36](#_Toc125209521)

[Table 35.Algoritmul Boruta 40](#_Toc125209522)

[Table 36. Model OLS 44](#_Toc125209523)

[Table 37. Model FE 45](#_Toc125209524)

[Table 38. Test FE 46](#_Toc125209525)

[Table 39.Model RE 46](#_Toc125209526)

[Table 40. Test Hausman 46](#_Toc125209527)

[Table 41.Testare efecte fixe în timp - Ftest 47](#_Toc125209528)

[Table 42.Testare efecte fixe în timp - LM 47](#_Toc125209529)

[Table 43.Testare efecte aleatorii 48](#_Toc125209530)

[Table 44.Testare dependență transfersală 48](#_Toc125209531)

[Table 45. Testare dependență transversală - Pesaran 48](#_Toc125209532)

[Table 46.Testare autocorelare 49](#_Toc125209533)

[Table 47.Testare heteroschedasticitate 49](#_Toc125209534)

[Table 48. Testare efecte random - F test 49](#_Toc125209535)

[Table 49.Testare efecte random - Breusch-Pagan LM 50](#_Toc125209536)

[Table 50.Corecție autocorelare 50](#_Toc125209537)

[Table 51. Corectare heteroschedasticitate 50](#_Toc125209538)

# ***Lista Figuri***

[Figure 1. Graficul observatiilor cu dreapta estimata 10](#_Toc125209061)

[Figure 2. Graficul ACF 16](file:///C:\Users\Ada\Desktop\1081_Carp_Catina.docx#_Toc125209062)

[Figure 3. Histogram - Normality test 19](#_Toc125209063)

[Figure 4. Histograma rezidurilor model multiplu 31](#_Toc125209064)

[Figure 5.Modelul de regresie liniară 37](#_Toc125209065)

[Figure 6.Valoarea lui lambda pt care avem MSE minimizat 38](#_Toc125209066)

[Figure 7.Diagrama Trace 38](#_Toc125209067)

[Figure 8.Diagrama Lasso 39](#_Toc125209068)

[Figure 9.Testarea valorii lamda LASSO 39](#_Toc125209069)

[Figure 10.Explorarea heterogeneitatii in sectiunea transversala 43](#_Toc125209070)

[Figure 11.Explorarea heterogeneitatii in sectiunea temporala 44](#_Toc125209071)

[Figure 12.Grafic OLS 45](#_Toc125209072)

# 

# 

# **Introducere**

Modelul econometric reprezinta o imagine simplificata a relatiilor dintre variabilele economice, care priveste atat reprezentarea anatomica a proceselor economice (definirea variabilelor) cat si descrierea fiziologica (relatii, conditionari, mecanisme de functionare).

Analiza de regresie se ocupă cu previzionarea valorii medie a unei variabile dependente, cunoscându-se valorile fixate ale variabilelor independente. Regresia se folosește pentru a determina și a testa o relație de cauzalitate, pentru a previziona o variabilă dependentă în funcție de una (sau mai multe) variabile independente sau pentru a explica efectul în funcție de cauze.

Modelul clasic de regresie liniară reprezintă una dintre tehnicile statistice cele mai versatile şi mai des utilizate în analiza economică. Prin acest proiect ne propunem să realizăm analiza modelelor liniare de regresie și, în vederea atingerii acestui scop, vor fi avute în vedere două modele de regresie diferite.

Consumul de vin este asociat în mod normal cu mâncarea bună, companie bună și, în general, timpuri bune. Consumatorii de vin încep să dezvolte un interes din ce în ce mai mare pentru vin, care depășește simplul act de a-l bea. În primul rând, știind mai multe despre caracteristicile vinului, istoria vinului și tipurile de vin au devenit la modă. Acum, o adevărată pasiune pentru vin s-a extins și oamenii vor să afle mai multe despre ce se află în spatele etichetei.

Piața vinurilor s-a extins la nivel mondial în toate direcțiile posibile. Noi zone producătoare de vin au apărut în principal în Lumea Nouă, aducând noi denumiri și stiluri. Peste 36 de miliarde de sticle sunt acum produse în peste 1 milion de etichete distincte de aproximativ 150.000 de producători. Piața vinului este, de asemenea, mai competitivă în fiecare an.

Echipa noastră a ales această temă deoarece înglobează două topicuri care deja se regăsesc în discuțiile din viețile noastre, și anume calitatea vinului. În mod invariabil, oamenii din întreaga lume mențin o relație mai strânsă cu vinul, care depășește cu mult consumul unei băuturi alcoolice sau ceva care să se asorteze cu o masă bună la un restaurant bun sau la o reuniune de familie. Din ce în ce mai mult, mai ales în lumea nouă, consumatorii de vin doresc să aibă o experiență de degustare mai deplină. Ei caută să înțeleagă mai bine ce definește acest teritoriu.

Prima aplicație își propune să studieze prin intermediul unui model unifactorial de regresie influența alcolului, asupra calității vinului.

A doua aplicație reprezintă o continuare a aplicației unu, deoarece se dorește a se studia și influența altor factori asupra calității vinului, aplicația analizând prin intermediul unui model multifactorial de regresie influența atât a alcoului, cât și a altor produși chimici, respectiv acidul volatil și sulfații. Astfel, cu ajutorul acestui model se dorește a se evidenția semnificația factorilor care influențează calitatea vinului alb sau roșu.

A treia aplicație este un model de date de tip panel în care analizăm rangul de fericire al țărilor din Europa ,luat ca proxy pentru această cercetare. Astfel, observăm influența pe care o are rangul de fericire asupra mai multor factori, precum: PIB-ul pe cap de locuitor, speranța la o viață sănătoasă, generozitatea, percepția corupției și libertatea. Perioada de timp analizată este 2015 și 2022.

În continuare, sunt descrise şi analizate tehnicile de regresie puse la dispoziţie de către programul econometric R Studio: specificarea şi estimarea unui model de regresie, efectuarea de analize simple de diagnostic şi utilizarea rezultatelor estimării în analize suplimentare, cum ar fi efectuarea de prognoze.

# **APLICAŢIA 1**

## **Model de regresie simplă**

### ***1.1.1. Literature review***

În ultimii câțiva ani, interacțiunea dintre consumatorii mai conștienți și actorii lanțului de aprovizionare cu vin, au influențat fenomenul complex, și oarecum la modă, al consumului de vin. În Europa, numărul tot mai mare de cluburi și asociații de vinuri acordată atenție vinului pe fiecare canal media și multitudinea. Exista târguri de vinuri născute peste tot care arată importanța culturală consumului de vin.

Cercetările de marketing au investigat percepția calității în cumpărarea vinului și comportamentul de consum prin recenzii, analize calitative și cantitative (Combris et al., 1997; Angulo şi colab., 2000; Rocchi, 2000; Orth, Kruska, 2002; Lockshin, 200 3, Verdú Jover et al, 2004; Charters, Pettigrew, 2007). Calitatea a fost adesea recunoscută si este un factor important pentru a înțelege determinanții de cumpărare. Calitatea percepută se bazează atât pe atribute intrinseci, cât și pe cele extrinseci și au fost analizate o multitudine de indicii care influențează consumatorii.

De exemplu, conform Zeithaml (1988) prețul devine un semnal în mod specific atunci când: sunt disponibile puține informații, produsul nu poate fi evaluat, riscul de a face o alegere greșită este mare.

Chiar dacă intuitiv decizia de a cumpăra vinuri ar putea fi ponderată pe mai multe date, cercetările empirice arată numărul limitat de informații pe care consumatorii il folosesc in aceasta alegere iar, prețurile împreună cu mărcile fac alegerea mai puțin riscantă (Lockshin și colab., 2000).

După cum au subliniat Stefani (1997) și Rocchi (2000) pe piața vinului, diferențierea calitativă este însoţită de o asimetrie informaţională între producător şi consumator. După cum demonstrează Landon și Smith (1997) pentru a reduce decalajul de informații, consumatorul utilizează semnale ca firmă, reputație bazandu-se pe experienta anterioara.

Lockshin sugerează trei dimensiuni de „implicare” legate de produse, mărci și achiziții; consumatorul mai mult sau mai puțin implicat reacționează în moduri diferit cu privire la atribute precum prețuri, denumiri de origine, soiuri etc. În general , cu atât consumatorii mai implicați vor fi cei cu disponibilitate mai mare de a analiza și de a utiliza informațiile pentru achizițiile lor, în schimb, cu atât mai puțin implicate vor fi cei care folosesc semnale precum prețurile sau mărcile ca strategii de reducere a riscului.

Charters e Pettigrew (2007) a studiat prin metoda focus-grupurilor cu sesiuni de degustare, calitatea experimentată a 60 de consumatori australieni cu un nivel diferit de implicare. Au găsit o ierarhie a dimensiunilor calității în care natura multidimensională a calității este bine exprimată. Ele introduc distincția dintre dimensiunile catalitice și terminale. Plăcerea este rezultatul unei experiențe senzoriale (dimensiune catalitică ) de degustare. Valoarea “plăcere” a fost descoperita ca valoare terminală in multe alte studii (Fotopoulos et al., 2003) și Alexaki și colab., 1999)

Dimensiunile calității au importanță diferită în funcție de nivelul de implicare: mai multi consumatori implicați favorizează dimensiunile cognitive/terminale, iar cei mai puțin implicați identifică caracteristici senzoriale chiar dacă nu sunt întotdeauna capabili să le definească perfect.

***1.1.2. Metodologia cercetării***

Aplicația studiază calitatea vinului portughez „Vinho Verde” variantele rosii si albe. Studiul pune accentul pe influenta compusilor chimici asupra calitatii acestuia.

Pentru a determina relația dintre calitate și alcool, am utilizat modelul de regresie liniară simplă.

Regresia liniară reprezintă cea mai utilizată metodă de modelare. Modelul de regresie liniară simplă este cel mai simplu model econometric şi conţine două variabile între care există o legătură de dependenţă.

Forma generală a unui model de regresie liniară simplă este:

Y = α + β \* X

unde:

Y – variabila dependentă, calitatea vinului

X – variabila independentă, alcoolul

α , β - parametrii

Parametrii ecuației de regresie:

* αa - reprezintă ordonata la origine și arată valoarea variabilei Y când X = 0;
* βb - este panta dreptei (numit și coeficient de regresie).

În ecuaţia de regresie parametrii αa şi βb sunt necunoscuţi. Semnul parametrului de regresie b indică direcţia legăturii între variabile şi anume:

* βb > 0 legătură directă (pozitivă), adică dacă variabila X creşte cu o unitate, atunci şi variabila Y creşte în medie cu β ;
* βb = 0 nu există legătură de tip liniar;
* βb < 0 legătură inversă (negativă), adică dacă variabila X scade cu o unitate, atunci şi variabila Y scade în medie cu β unităţi.

Prin intermediul soft-ului R Studio, am estima modelul prin metoda celor mai mici pătrate, iar pentru testarea validității acestuia am calculat indicatorii de bonitate, aceștia fiind coeficientul de determinație, coeficientul de nedeterminație, coeficientul de determinație ajustat. Totodată, am testat ipotezele modelului de regresie liniară și semnificația parametrilor din model.

În cadrul aplicației sunt utilizate următoarele teste:

● Testul t;

● Testul Breusch-Pagan;

● Testul White;

● Testul Durbin-Watson;

● Testul Breusch-Godfrey

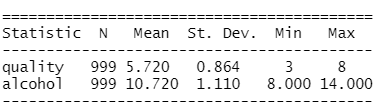
● Testul Jarque-Bera;

1.2.1. Date utilizate

Variabilele utilizate in modelul de regresie liniara au fost calitatea(variabila senzoriala) si alcoolul (variabila fizico-chimica). Aceste date au fost selectate de pe site-ul Kaggle.

Valorile nu au necesitat transformări.

Table 1.Principalele statistici descriptive ale indicilor



În ceea ce priveşte evoluţia calității vinului, se observă ca acesta are o evoluţie crescătoare pe parcursul cantității alcoolului (Figura 1).

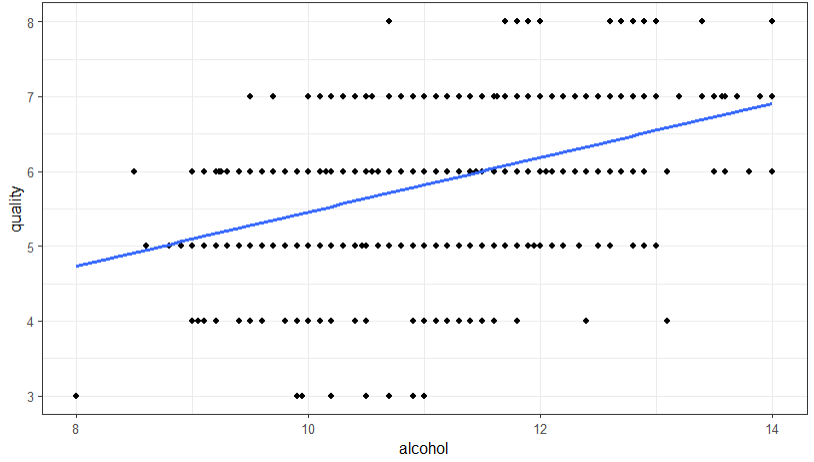


Figure 1. Graficul observatiilor cu dreapta estimata

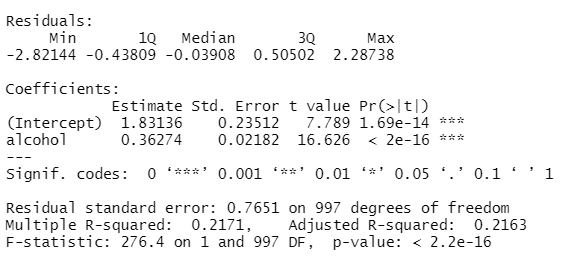
1.2.2. Rezultatele empirice ale cercetării

Pentru a determina intensitatea relaţiei dintre variabile (cât de bine este reprezentat setul de date cu ajutorul funcţiei de regresie) se stabileşte nivelul corelaţiei dintre acestea. Corelaţia indică intensitatea legăturii dintre variabile prin măsurarea gradului de împrăştierea a datelor înregistrate în jurul dreptei de regresie. Raportul de corelaţie va avea semnul parametrului β din ecuaţia de regresie. Vom calcula acum coeficientul de corelaţie Pearson:

Coeficientul indică o legătură directă (R > 0), moderata, întrucât se afla intre 0.3 si 0.7.

Vom continua cu estimarea parametrilor modelului. Analiza seriilor de date considerate, precum şi estimarea parametrilor modelului de regresie a fost realizată cu ajutorul pachetului de programe infromatice specializate R Studio, prin aplicarea funcției summary, ducând la obţinerea următoarelor rezultate:

Table 2.Modelul de regresie liniara

****

Dreapta de regresie estimată este:

Y = 1.83136 – 0.36274 \* X

α = 1.83136 – arată nivelul calitatii atunci când alcoolul tinde către 0

β = 0.36274 – este panta dreptei de regresie și înseamnă că atunci când alcoolul crește cu o unitate (1%), calitatea va creste, în medie, cu 0.36 unitati.

*Indicatorii de bonitate*

Pentru aflarea calității ajustării, calculăm indicatorii de bonitate:

**a. Coeficientul de determinație**

R2 = R-squared = 0.2170814 = > aproximativ 21.71% din componenta alcoolului explică variația calitatii.

**b. Coeficientul de nedeterminație**

K2 = 1- R2 = 0.7829186 => acțiunea altor factori neincluși în model și cuprinși în eroare explică aproximativ 78.29% din calitate.

**c. Coeficientul de determinație ajustat**

Adjusted R-squared= 0.2163. Arată legătura dintre alcool, calitate. Rezultă că aproximativ 21.63% din variația alcoolului este explicată de variația calitatii.

**d. Raportul de corelație**

Indică faptul că între cele două variabile există o legătură moderata și directă.

*Testarea semnificației parametrilor*

***Testul t metoda 1 - compararea testului t calculat cu cel critic***

În continuarea vom verifica semnificaţia parametrilor cu ajutorul testului t:

Ipoteza:

H0: α=0, β=0 ↔ rxy =0 (parametrii nu sunt semnificativi/ rxy nu este semnificativ, modelul nu este valid)

H1: α ≠0,β≠0 ↔ rxy≠0 (parametrii sunt semnificativi/ rxy este semnificativ din punct de vedere statistic, modelul este valid statistic)

– la nivel de semnificatie de 5% .

Deoarece |t\_stat | > |t\_critic | acceptăm ipoteza alternativă conform căreia coeficientul de corelație liniară Pearson este semnificativ din punct de vedere statistic.

***Testul t metoda 2: calcularea intervalelor de incredere***

Valoarea minima a intervalului pentru 95% nivel de incredere

coefficient - 1.96 \* se => 0.3199743

Valoarea maxima a intervalului pentru 95% nivel de incredere

coefficient + 1.96 \* se => 0.4054957

Intervalul : [0.3199743, 0.4054957] => intervalul nu include 0 => ipoteza nula e respinda iar coeficientul e semnificativ diferit de 0 pentru 95% nivel de incredere

Valoarea critica pentru nivel de semnificatie de 10%

Valoarea minima a intervalului pentru 90% nivel de incredere

coefficient - 1.65 \* se => 0.3267375

Valoarea maxima a intervalului pentru 90% nivel de incredere

coefficient + 1.65 \* se => 0.3987325

Intervalul : [0.32,0.39] intervalul nu include 0 => ipoteza nula e respinda iar coeficientul e semnificativ diferit de 0 pentru 90% nivel de incredere

Cele doua metode: de comparare a testului t cu valorile critice, si intervalele de incredere duc la aceleasi rezultate si concluzii.

*Testarea validității modelului*

Pentru testarea validității modelului de regresie, ipotezele sunt:

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

H1: modelul este valid statistic (MSR>MSE)

Putem afirma cu siguranță că modelul este semnificativ statistic în urma testului F (F-statistic = 276.4) > F critic ( Ftabelat=Fcritic=Fα;1;n-2=F0,05;1;997=3.85), deci se respinge ipoteza nulă H0 si se accepta ipoteza alternativa H1, modelul fiind valid statistic.

*Verificarea îndeplinirii ipotezelor modelului de regresie liniară*

1. Forma este liniară: Y = 1.83136 – 0.36274 \* X
2. Nr de observatii > nr variabile independente: Adevarat
3. Modelul de regresie este corect specificat: ecuatia modelului este specificata in mod corespunzator
4. Variabilitatea in x este pozitiva: 1 > 0
5. Media reziduurilor este 0: medie=1.02 aproape de 0
6. Rezidurile nu sunt corelate cu variabilele independente: p-value = 1 > 0.1 => nu sunt corelate
7. Testare heteroschedasticitate: var() = constant

Pentru a observa homoscedasticitatea erorilor aleatoare se vor aplica următoarele teste:

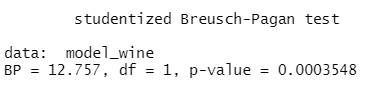
**TESTUL BREUSCH-PAGAN**

Testul se aplică pentru următoarele ipoteze:

H0: erorile sunt homoscedastice (rezidurile sunt distribuite cu varianta egala)

H1: erorile sunt heteroscedastice (rezidurile nu sunt distribuite cu varianta egala)

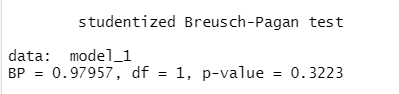
Table 3. Testul Breusch-Pagan

**

Precum se poate observa în tabelul de mai sus, Prob(F-statistic) = 0.0003548 și este mai mic de 5%, deci probabilitatea de a greși în respingerea ipotezei nule este mică. Astfel, acceptăm H1 și erorile aleatoare sunt heteroschedastice.

În consecință, vom aplica modelul de regresie log quality (log-lin) pentru corectarea heteroschedasticității.

Table 4. Testul Breusch-Pagan corectat



Obținem că p-value pentru statisticile calculate este mai mare de 5%, respectiv 0.3223, astfel există o probabilitate foarte mare de a greși în respingerea lui H0, deci acceptăm H0, conform căreia erorile aleatoare sunt homoscedasticitate.

**TESTUL WHITE**

Testul se aplică pentru următoarele ipoteze:

H0: există homoscedasticitate

H1: există heteroscedasticitate

Table 5. Model Test White

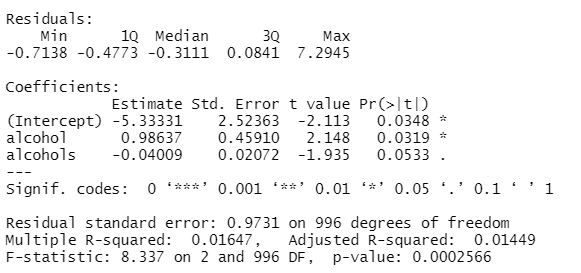
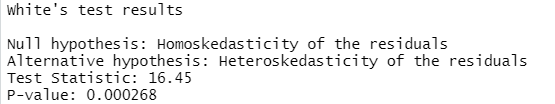
****

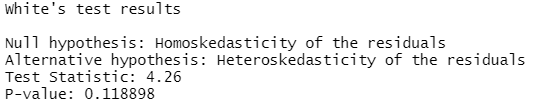
Table 6. Testul White

****

Precum se poate observa în tabelul de mai sus, p-value = 0.000268 și este mai mic de 5%, deci probabilitatea de a greși în respingerea ipotezei alternative este mare. Astfel, acceptăm H1 și erorile aleatoare sunt heteroscedastice.

În consecință, vom aplica modelul de regresie log quality (log-lin) pentru corectarea heteroschedasticității.

Table 7. Testul White corectat



Obținem că p-value pentru statisticile calculate este mai mare de 5%, respectiv 0.118898, astfel există o probabilitate foarte mică de a greși în respingerea lui H1, deci acceptăm H0, conform căreia erorile aleatoare sunt homoscedasticitate.

1. Neautocorelarea erorilor aleatoare

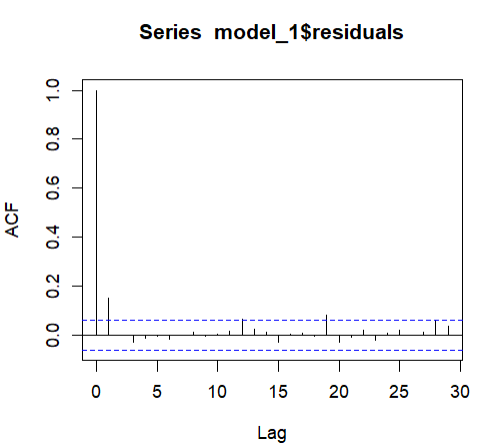
Inspectarea autocorelarii cu ajutorul graficului ACF (autocorelare) (Figura 2).

Figure 2. Graficul ACF

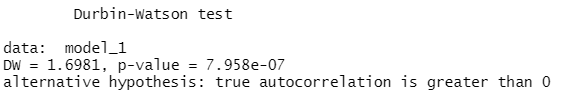
**TESTUL DURBIN-WATSON**

Folosim statistica Durbin-Watson cu ipotezele:

H0: ρ =0 (nu există autocorelarea erorilor aleatoare de ordinul I)

H1: ρ ≠0 (există autocorelarea erorilor aleatoare de ordinul I)

Table 8. Testul Durbin-Watson



Se obține DW = 1,6981 pentru modelul analizat. Valorarea critică ale statisticii DW este Prob(F-statistic) = 7.958e-07, care este mai mic de 1%. Astfel rezultă că există autocorelarea erorilor.

În consecință, vom crea un nou set de date pentru cel corectat, implementarea modelului cu o nouă variabilă adaugată în model, pentru corectarea autocorelării.

Table 9. Model Test Durbin-Watson

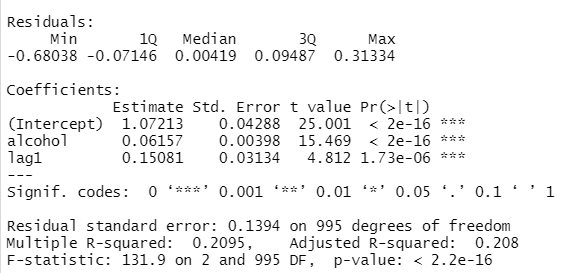
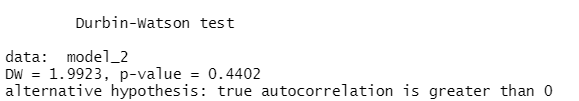


Table 10. Testul Durbin-Watson corectat

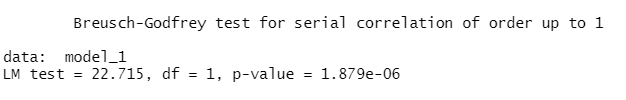


După corectarea autocorelării, noua valoare a indicatorului Durbin-Watson este DW = 1,9923 pentru modelul analizat. Noua valorare critică ale statisticii DW este Prob(F-statistic) = 0.4402, care este mai mare de 1% , deci nu mai avem autocorelare a erorilor aleatoare.

**TESTUL BREUSCH-GODFREY**

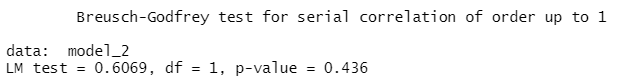
Se obține LM = 22.715 pentru modelul analizat. Valorarea critică ale statisticii LM este Prob(F-statistic) = 1.879e-06, care este mai mic de 1%. Astfel rezultă că există autocorelarea erorilor.

Table 11.Testul Breusch-Godfrey

****

În consecință, vom crea un nou set de date pentru cel corectat, implementarea modelului cu o nouă variabilă adaugată în model, pentru corectarea autocorelării.

Table 12. Testul Breusch-Godfrey corectat



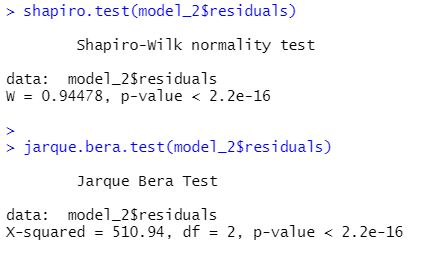
După corectarea autocorelării, noua valoare a indicatorului Breusch-Godfrey este LW = 0.6069, iar pentru ordin II este LM = 0.95772. Noua valorare critică ale statisticii DW este Prob(F-statistic) = 0.436, iar pentru ordin II este Prob(F-statistic) = 0.6195 care este mai mare de 1% , deci nu mai avem autocorelare a erorilor aleatoare.

1. Normalitatea distribuției erorilor aleatoare

Pentru testarea ipotezei de normalitate a erorilor aleatoare se va folosi testul Jarque-Bera, cu ipotezele:

H0: erorile aleatoare au o distribuție normală

H1: erorile aleatoare nu au o distribuție normală

Table 13. Testul Jarque-Bera

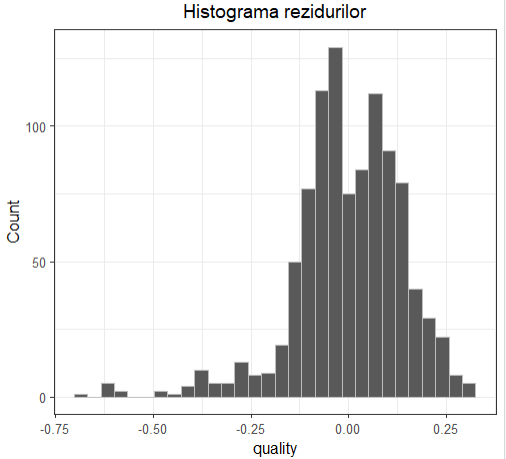
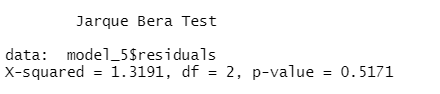


Figure 3. Histogram - Normality test

Probabilitatea testului Jarque-Bera este de 2.2e-16, care este mai mică decât 5%, deci se respinge ipoteza nulă și se acceptă ipoteza alternativă conform căreia erorile nu au o distribuție normală. Astfel, vom corecta modelul folosind distanta Cook si eliminand toate observatiile influente care ne afecteaza modelul.

Table 14. Testul Jarque-Bera corectat



Probabilitatea testului Jarque-Bera este de 0.5171, care este mai mare decât 5%, deci se acceptă ipoteza nulă conform căreia erorile au o distribuție normală.

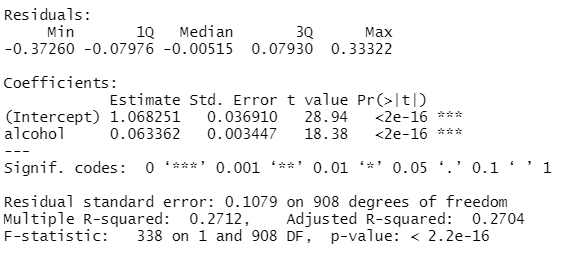
1.3. Concluzii

În urma prelucrărilor efectuate asupra datelor s-a obținut un model econometric de regresie liniară cu o bonitate medie de 27% în care alcoolul influențează calitatea vinului.

Modelul unifactorial rezultat în urma estimării este următorul:

*CALITATEA = 1.068251 + 0.063362 \* ALCOOL*

Table 15.Modelul de regresie simpla final

****

Cu ajutorul soft-ului Rstudio am estimat modelul anterior și am obținut următoarele concluzii:

● Utilizând statistica t se poate observa că valorile coeficientilor sunt semnificative, indiferent de metoda utilizata, iar între variabilele cercetate există o legătură semnificativă, cu o probabilitate de 95%, modelul de analiză este corect specificat.

● Coeficientul de determinare confirmă faptul că alcoolul este un factor ce influențează calitatea, fiind o influență de 27,12% a alcoolului asupra calitatii.

● Între alcool și calitatea vinului există o relație semnificativă, astfel se poate afirma că o creștere cu o unitate (1%) a alcoolului determina o crestere cu 36% a calitatii.

# **APLICAŢIA 2**

## **Model de regresie multiplă**

### ***2.1.1. Literature review***

Calitatea vinului, așa cum a spus Maynard Amerine, este mai ușor de detectat decât de definit. Acest lucru se datorează parțial faptului că calitatea este în primul rând subiectivă și puternic influențată de factori extrinseci. În mod corespunzător, definirea calității vinului în ceea ce privește chimia sa nu va fi niciodată mai mult decât parțial de succes. Cu toate acestea, cei mai serioși cunoscători de vin tind să cadă de acord cu privire la ceea ce constituie calitatea vinului, adică ceea ce au ajuns să le placă în mod subiectiv prin degustări extinse. Deși dezamăgitor de nebulos, a fost totuși suficient pentru a ghida cultivatorii de struguri și vinificatorii în alegerea procedurilor pe care le folosesc. Ceea ce nu este sigur este dacă acesta este un exemplu complex de orb care conduce orb. Acest gând tulburător este parțial susținut de acceptarea neîndoielnică a anumitor tradiții privind producția de struguri și vin, în special ideile despre boabe mici, viță de vie vechi, randament limitat, evitarea irigațiilor etc., ca fiind esențiale pentru calitatea vinului. Toate acestea s-au dovedit experimental a fi în esență false.

Atributele de calitate ale produselor vitivinicole: un studiu explicativ al motivației de cumpărare a consumatorilor printr-o abordare a lanțurilor mijloace-final Alessio Cavicchi – Departamentul de Studii privind Dezvoltarea Economică , Universitatea din Macerata. Annamaria Mazzeschi – Institutul de Educație, Universitatea din Londra Silvia Guerri – Departamentul de Economie Agricolă și Resurselor, Universitatea din Florența.

Din punctul de vedere al cercetătorului, complexitatea acestei probleme este binecunoscută: există dovezi empirice considerabile că consumatorii, atunci când cumpără un anumit produs alimentar sau vitivinicol, exprimați o preferință pentru anumite atribute (culoare , mărime, regiune de origine...) care sunt luate în considerare indicii de calitate (semnale) ale întregului produs. Din acest motiv, Steenkamp a clarificat conceptul în 1989 de calitate percepută pe pieţele alimentare ca mediere între caracteristicile produsului şi preferințele consumatorului. Aceasta înseamnă că percepția calității depinde de o evaluare individuală care este strict legat de mediu și de situația specifică de consum și care poate fi bazată pe informații incomplete (Holm, Kild evang, 1996). Descoperirea subiectivului și dimensiunea obiectivă și modul în care acestea se relaționează între ele pot fi o problemă importantă pentru a determina rentabilitatea firmelor pe o perioadă lungă. După cum a afirmat Lockshin (2003), calitatea vinului se poate baza pe percepții, cum ar fi prețul, recomandările prietenilor sau experților sau eticheta. Dar nivelurile de implicarea şi cunoaşterea pot fi determinanţi ai informaţiilor căutate şi în consecinţă ai percepția calității și comportamentul de cumpărare . Astfel, înțelegerea calității diferiților consumatori, percepția este un curent principal atât pentru cercetători , cât și pentru managerii care se ocupă de piața vinului.

### ***2.1.2. Metodologia cercetării***

Forma modelului aplicație este următoarea:

*Y =* α *+* β 1*\** X1 *+* β2 *\** X2 + β3 *\** X3

unde:

Y – variabila dependentă, calitatea vinului;

X1 – variabila independentă, alcoolul;

X2 – variabila independentă, acid volatil;

X3 – variabila independentă, sulfați.

Astfel, cu ajutorul soft-ului R Studio, am estimat modelul prin metoda celor mai mici pătrate și am testat validitatea acestuia prin calcularea indicatorilor de bonitate, aceștia fiind coeficientul de determinație, coeficientul de nedeterminație, coeficientul de determinație ajustat. Totodată, am testat ipotezele modelului de regresie liniară și semnificația parametrilor din model.

În cadrul aplicației sunt utilizate următoarele teste și criterii statistice:

● Testul t;

● Testul Breusch-Pagan;

● Testul White;

● Testul Durbin-Watson;

● Testul Breusch-Godfrey

● Testul Jarque-Bera;

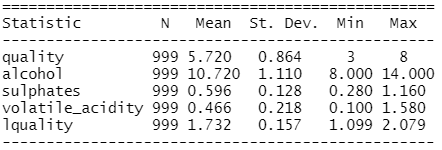
***2.2.1.1. Date utilizate***

Variabilele utilizate in modelul de regresie multipla au fost calitatea (variabila senzoriala), alcoolul , acidul volatil si sulfatii (variabile fizico-chimice).

Volatilitatea este tendința unei substanțe de a se vaporiza. Motivul pentru volatilitatea ridicată este că există o presiune ridicată a vaporilor la temperatura camerei normale.

Sulfatul este, în [chimia anorganică](https://ro.wikipedia.org/wiki/Chimie_anorganic%C4%83), un [anion](https://ro.wikipedia.org/wiki/Anion). Sulfatii (compușii care conțin acest ion) sunt [săruri](https://ro.wikipedia.org/wiki/Sare) ale [acidului sulfuric](https://ro.wikipedia.org/wiki/Acid_sulfuric).

Table 16.Principalele statistici descriptive - regresie multipla



2.2.1.2. Rezultatele empirice ale cercetării

Dreapta de regresie estimată este:

Y = 1.12422 + 0. 05621\* X1  +0.18599\* X2- 0.22637 *\** X3

α = 1.12422 – arată nivelul logaritmat al calității alcolului atunci când acidul volatil și și sulfații tind către 0;

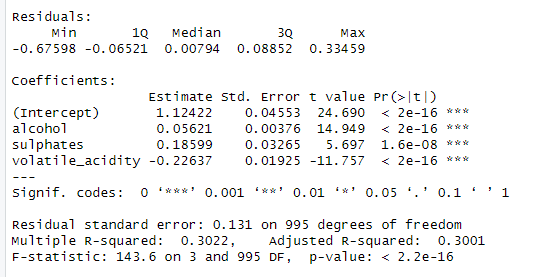
β 1 = 0. 05621 – arată că atunci când alcoolul creste cu o unitate (1%),logaritmul calității vinului va creste, în medie, cu 0.05 unități;

β2 = 0.18599 – arată că atunci când sulfații cresc cu o unitate (1%),logaritmul calității vinului va creste, în medie, cu 0.18 unități.

β3= - 0.22637 –arată că atunci când acidul volatile crește cu o unitate (1%), logaritmul calității vinului va scadea, în medie, cu 0.22 de unități.

În continuare se va analiza fiecare variabilă inclusă în model, pentru a evidenția evoluția sa economică, pentru care s-au selectat datele.

Table 17.Regresie multifactorială între alcool, sulfați și acid volatil



*Indicatorii de bonitate*

Pentru aflarea calității ajustării ,calculăm indicatorii de bonitate:

**a. Coeficientul de determinație**

R2 = R-squared = 0,3022= > arată că 30,22% din variația variabilei dependente este explicată de variația simultană a alcoolului, acidului volatil și sulfaților, adică o legătură puternică între variabila endogenă și cele trei variabile exogene

**b. Coeficientul de nedeterminație**

K2 = 1- R2 = = 0,6978 => acțiunea altor factori neincluși în model și cuprinși în eroare explică aproximativ 69.78% din emisiile de carbon.

**c. Coeficientul de determinație ajustat**

Adjusted R-squared= 0,3001. Rezultă că aproximativ 30.01% din variația logaritmării calității vinului este explicată de variația acidului volatil și sulfaților.

*Testarea semnificației parametrilor*

***TESTUL T***

Avem ipotezele:

H0: α =0 ( Parametrul α nu este semnificativ statistic)

β1 =0 ( Parametrul β1 nu este semnificativ statistic )

β2=0 ( Parametrul β2 nu este semnificativ statistic )

β3=0 ( Parametrul β2 nu este semnificativ statistic )

H1: α ≠0 ( Paramentrul α este semnificativ statistic )

β1≠0 ( Parametrul β1 este semnificativ statistic )

β2≠0 ( Parametrul β2 este semnificativ statistic )

β3≠0 ( Parametrul β2 este semnificativ statistic )

α = 1.12422

β 1 = 0. 05621

β2 = 0.18599

β3= - 0.22637

t1calculat= 14.9; t2calculat= 5.69; t3calculat= -11.

tcritic= t0.025;995 =1.96.

Cum | t1calculat | > tcritic si | t2calculat | > tcritic, | t3calculat | > tcritic. Pentru un nivel de semnificatie de 5%, toate sunt semnificative statistic. Pentru fiecare dintre cei 3 parametri respingem ipoteza nula acceptăm ipoteza alternativa ,toți parametrii sunt semnificativi statistic la pragul de semnificaţie de 5%. Acest lucru este întărit și de valorile foarte mici ale lui Prob. pentru fiecare parametru al modelului.

*Testarea validității modelului*

Pentru testarea validității modelului de regresie, ipotezele sunt:

H0: modelul nu este valid statistic

H1: modelul este valid statistic

Putem afirma cu siguranță că modelul este semnificativ statistic în urma testului F ,F-statistic = 138.2309 > F critic = 3.850822, deci se respinge ipoteza nulă H0, modelul fiiind valid pentru un nivel de semnificație Prob. (F-statistic)=5.626784e-30, mai mic față de 5%.

*Verificarea îndeplinirii ipotezelor modelului de regresie liniară*

1. Forma este liniară: Y = 1.12422 + 0. 05621\* X1  +0.18599\* X2- 0.22637 *\** X3

2. Nr de observatii > nr variabile independente:

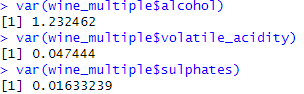
Table 18.Verificare nr obs > nr variabile independente



3. Modelul de regresie este corect specificat: ecuatia modelului este specificata in mod corespunzator;

4. Variabilitatea in x este pozitiva:

Table 19.Variabilitatea in x

**

Se aplică pentru următoarele ipoteze:

H0: este pozitivă

H1: este negativă

După cum se poate observa, valorile 1.232462, 0.047444 și 0.01633239 sunt mai mari ca 0, deci acceptăm ipoteza nulă unde Variabilitatea in x este pozitivă.

5. Media reziduurilor este 0:

Se aplică pentru următoarele ipoteze:

H0: aproape de 0

H1: depărtată de 0

Table 20.Media rezidurilor



Având în vedere că valoarea -2.226739e-18 este apropiată de 0, se acceptă H0.

6. Testare multicoliniaritate: nu avem valori pt VIF > 10

Se aplică pentru următoarele ipoteze:

H0: VIF < 10

H1: VIF > 10

Table 21.VIF

Se poate observa că valorile pentru factorul de inflație al variație sunti 1.013539, 1.012536, 1.023032 sunt mai mici ca 10, deci acceptăm H0

7. Rezidurile nu sunt corelate cu variabilele independente: p-value = 1 > 0.1 => nu sunt corelate

8. Testare heteroschedasticitate: var() = constant

Pentru a observa homoscedasticitatea erorilor aleatoare se vor aplica următoarele teste:

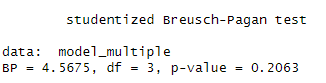
**TESTUL BREUSCH-PAGAN**

Testul se aplică pentru următoarele ipoteze:

H0: există homoscedasticitate

H1: există heteroscedasticitate

Table 22.Testul Breusch-Pagan multiplu



Precum se poate observa în tabelul de mai sus, Prob(F-statistic) = 0.2063 și este mai mare de 5%, deci probabilitatea de a greși în respingerea ipotezei nule este mare. Astfel, acceptăm H0 și erorile aleatoare sunt homoschedastice.

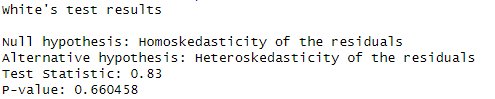
**TESTUL WHITE**

Testul se aplică pentru următoarele ipoteze:

H0: există homoscedasticitate

H1: există heteroscedasticitate

Table 23.Testul White multiplu



Obținem că Prob. F pentru statisticile calculate este mai mare de 5%, respectiv 0.660458, astfel există o probabilitate foarte mică de a greși în respingerea lui H1, deci acceptăm H0, conform căreia erorile aleatoare sunt homoscedasticitate.

8. Neautocorelarea erorilor aleatoare

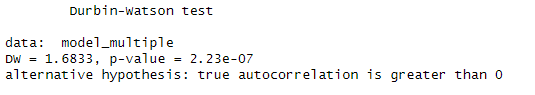
**TESTUL DURBIN-WATSON**

Folosim statistica Durbin-Watson cu ipotezele:

H0: ρ =0 (nu există autocorelarea erorilor aleatoare de ordinul I)

H1: ρ ≠0 (există autocorelarea erorilor aleatoare de ordinul I)

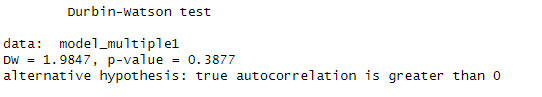
Table 24.Testul Durbin-Watson multiplu



Se obține DW = 1.6833 pentru modelul analizat. Valorarea critică ale statisticii DW este p-value = 2.23e-07, care este mai mic de 1%. Astfel rezultă că există autocorelarea erorilor.

În consecință, vom crea un nou set de date pentru cel corectat, cu implementarea modelului cu o nouă variabilă adaugată în model, pentru corectarea autocorelării.

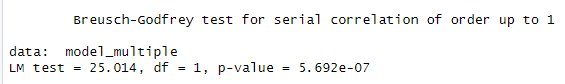
Table 25.Testul Durbin-Watson multiplu corectat



După corectarea autocorelării, noua valoare a indicatorului Durbin-Watson este DW = 1.9847 pentru modelul analizat. Noua valorare critică ale statisticii DW este p-value = 0.3877, care este mai mare de 1% , deci nu mai avem autocorelare a erorilor aleatoare.

**TESTUL BREUSCH-GODFREY**

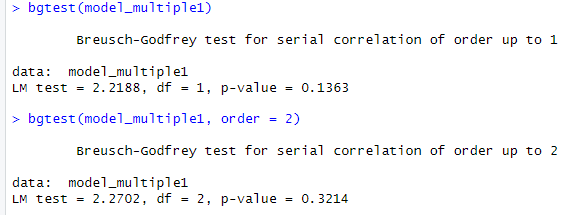
Table 26.Testul Breusch-Godfrey multiplu



Se obține LM = 25.014 pentru modelul analizat. Valorarea critică ale statisticii LM este Prob(F-statistic) = 5.692e-07, care este mai mic de 1%. Astfel rezultă că există autocorelarea erorilor.

În consecință, vom crea un nou set de date pentru cel corectat, cu eeimplementarea modelului cu o nouă variabilă adaugată în model, pentru corectarea autocorelării.

Table 27.Testul Breusch-Godfrey multiplu de ordin I și II corectat



După corectarea autocorelării, noua valoare a indicatorului Breusch-Godfrey este LW = 2.2188, iar pentru ordin II este LM = 2.2702. Noua valorare critică ale statisticii DW este Prob(F-statistic) = 0.1363, iar pentru ordin II este Prob(F-statistic) = 0.3214 care este mai mare de 1% , deci nu mai avem autocorelare a erorilor aleatoare.

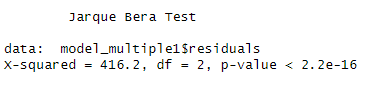
1. Normalitatea distribuției erorilor aleatoare

Pentru testarea ipotezei de normalitate a erorilor aleatoare se va folosi testul Jarque-Bera, cu ipotezele:

H0: erorile aleatoare au o distribuție normală

H1: erorile aleatoare nu au o distribuție normală

Table 28. Testul Jarque-Bera multiplu



Valoarea probabilitatății testului Jarque-Bera este de 2.2e-16, cu o probabilitate asociată de 0.531562, care este mai mică decât 5%, deci se respinge ipoteza nulă și se acceptă ipoteza alternativă conform căreia erorile nu au o distribuție normală.

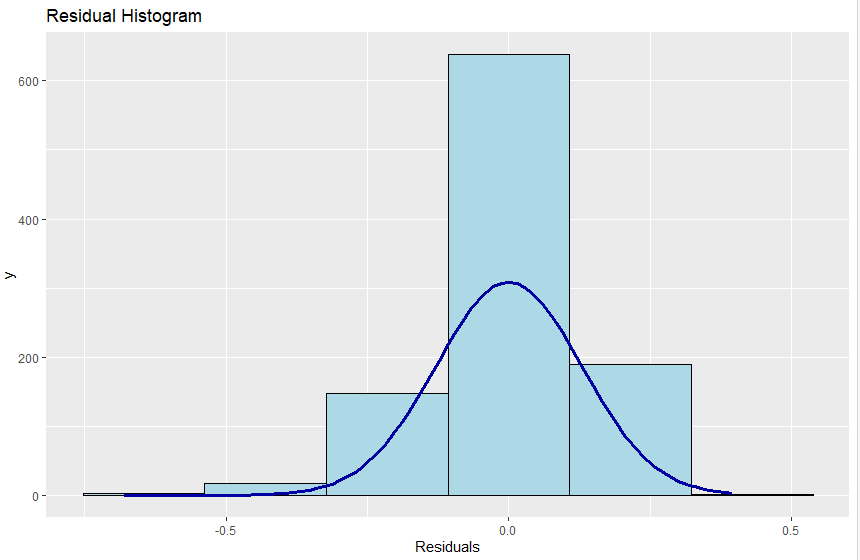
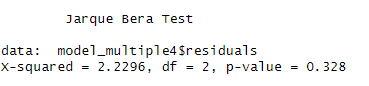


Figure 4. Histograma rezidurilor model multiplu

Table 29.Testul Jarque-Bera multiplu corectat



Probabilitatea testului Jarque-Bera este de 0.328, care este mai mare decât 5%, deci se acceptă ipoteza nulă conform căreia erorile au o distribuție normală.

2.2.2.1. Metodologia cercetării

Am dezvoltat modelul de regresie prin adoptarea altei forme functionale , adaugarea unei variabile de interacțiune între variabila alcool și cea nou introdusă, variabila dummy:

Y = α + β1\* X1 + β2 \* X2 + β3 \* X3 + β4\* D + X1\*D

unde:

Y – variabila dependentă, logaritm din calitatea vinului

X1 - logaritm din alcool

X2 - logaritm din sulfați

X3 - logaritm din acid volatil

D - vin alb

D = 1, vin alb

D = 0 , vin rosu

Astfel, cu ajutorul soft-ului R Studio, am estimat modelul prin metoda celor mai mici pătrate și am testat validitatea acestuia prin calcularea indicatorilor de bonitate, aceștia fiind coeficientul de determinație, coeficientul de nedeterminație, coeficientul de determinație ajustat. Totodată, am testat ipotezele modelului de regresie liniară și semnificația parametrilor din model.

2.2.2.2. Rezultatele empirice ale cercetării

Dreapta de regresie estimată este:

Y = 0.59951 + 0.46963\* X1 + 0.12156\* X2 – 0.0979 \* X3 – 0.97907\* D + 0.41033\*X1\*D

α = 0.59951 – arată nivelul logaritmat al calității alcolului atunci când acidul volatil și și sulfații tind către 0

β1= 0.46963 – arată că atunci când alcoolul creste cu o unitate (1%), calitatea vinului va creste;

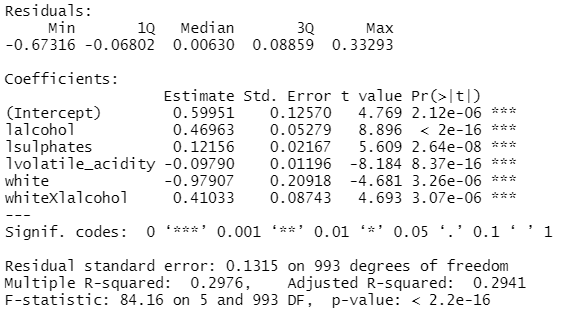
β2 = 0.12156 – arată că atunci când sulfații cresc cu o unitate (1%), calitatea vinului va creste;

β3 = – 0.0979 – arată că atunci când acidul volatil crește cu o unitate (1%), calitatea vinului va scadea;

β4 = – 0.97907 – este negativ, deci efectul vinului alb asupra calitatii va scadea;

β5 = 0.41033 – variabila de interacțiune între variabila alcool si tipul vinului este pozitiva.

Table 30. Tabelul modelului de interacțiune a regresiei multiple



*Indicatorii de bonitate*

Pentru aflarea calității ajustării ,calculăm indicatorii de bonitate:

**a. Coeficientul de determinație**

R2 = R-squared = 0,2976= > arată că 29,76% din variația variabilei dependente este explicată de variația simultană a vinului alb și a logaritmului dintre alcool, acidul volatil și sulfaților, adică o legătură puternică între variabila endogenă și cele trei variabile exogene.

**b. Coeficientul de nedeterminație**

K2 = 1- R2 = 0,7024 => acțiunea altor factori neincluși în model și cuprinși în eroare explică aproximativ 70.24% din logaritmul calității vinului.

**c. Coeficientul de determinație ajustat**

Adjusted R-squared= = 0,2941.Rezultă că aproximativ 29.41% din variația logaritmării calității vinului este explicată de variația vinului alb și variația logaritmului acidului volatil și sulfaților.

*Testarea semnificației parametrilor*

***TESTUL T***

Avem ipotezele:

H0: α =0 ( Parametrul  α nu este semnificativ statistic)

β1 =0 ( Parametrul β1 nu este semnificativ statistic )

  β2=0 ( Parametrul β2 nu este semnificativ statistic )

β3=0 ( Parametrul β3 nu este semnificativ statistic )

β4=0 ( Parametrul β4 nu este semnificativ statistic )

β5=0 ( Parametrul β4 nu este semnificativ statistic )

H1: α ≠0  ( Paramentrul α este semnificativ statistic )

      β1≠0  ( Parametrul β1 este semnificativ statistic )

        β2≠0  ( Parametrul β2 este semnificativ statistic )

β3=0 ( Parametrul β3  este semnificativ statistic )

β4=0 ( Parametrul β4 este semnificativ statistic )

β5=0 ( Parametrul β4 este semnificativ statistic )

tacalculat= 8.89 ; tscalculat= 15.6; tvcalculat= -8.1; twcalculat= -4.6; twacalculat= 4.69

tcritic=1.96.

* |tcalculat| > |tcritic | . Pentru un nivel de semnificatie de 5%, parametrii sunt semnificativi statistic.

*Testarea validității modelului*

Pentru testarea validității modelului de regresie, ipotezele sunt:

H0: modelul nu este valid statistic

H1: modelul este valid statistic

Putem afirma cu siguranță că modelul este semnificativ statistic în urma testului F ,F-statistic = 84.16> F critic = 2.380894, deci se respinge ipoteza nulă H0, modelul fiind valid.

Testul Chow

Pentru testarea validității modelului de regresie, ipotezele sunt:

H0: modelul nu este valid statistic

H1: modelul este valid statistic

Putem afirma cu siguranță că modelul este semnificativ statistic în urma testului F (F-statistic = 10.13654> F critic = 2.380912, deci se respinge ipoteza nulă H0, modelul fiiind valid pentru un nivel de semnificație Prob. (F-statistic)=4.809635e-08, mai mic față de 5%.

*Predicții*

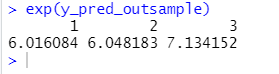
RMSE - Root Mean Squared Error => acest indicator de acuratete arata cât de departe se incadrează predicțiile față de valorile reale măsurate folosind distanța euclidiană. RMSE=0.1 < 1 => predicție bună.

MAE - Mean Absolute Error => similar cu RMSE, doar ca MAE este un scor liniar, ceea ce înseamnă că toate diferențele individuale sunt ponderate in mod egal. MAE=0.09 < 1 => predicție bună.

MSE - Mean Squared Error => măsoară în unități care este pătratul variabilei țintă și penalizează mai mult și mai sever erorile mai mari. MSE < 0 =>predicție bună.

MAPE - Mean Absolute Percentage Error => măsură ce indică media dispersiei dintre valorile previzionate si cele reale . MAPE = 0.05 < 1 => predicție bună.

În această secțiune vom previziona calitatea folosind modelul de regresie multipla. In tabelul de mai jos (Table 31. Prefictia calitatii pentru 3 observatii) am realizat o predictie pentru trei tipuri de valori a compusilor chimici care s-au dovedit a fi relevanti.

Table 31.Prefictia calitatii pentru 3 observatii

*Regresia neliniara*

**RAMSEY RESET**

Testul se aplică pentru următoarele ipoteze:

H0: modelul este bine specificat

H1: modelul nu este bine specificat

Table 32.Modelul de regresie pentru calitatea vinului

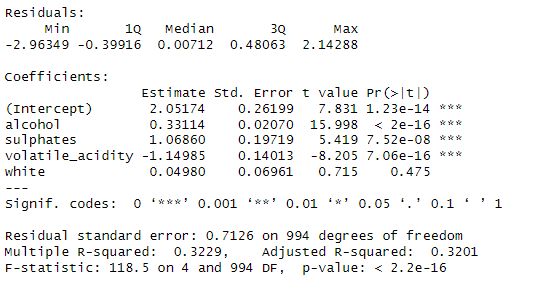
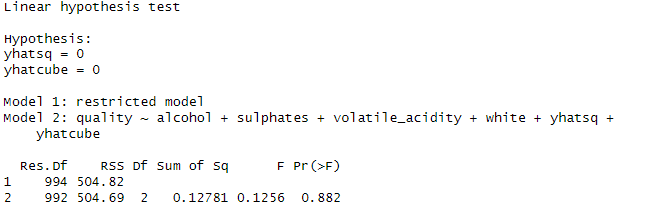


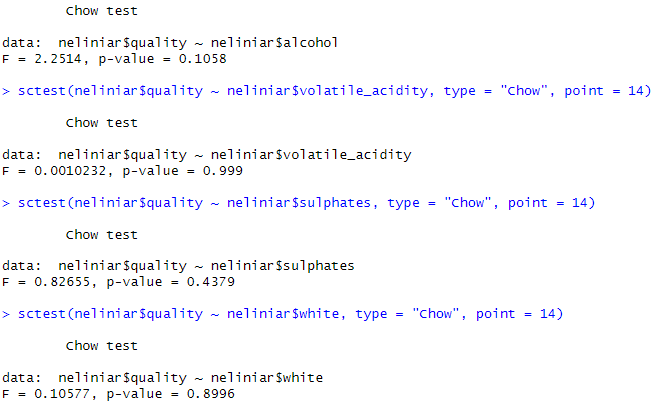
Table 33.Testul Reset

**

După cum se poate observa, valoarea probabilității pentru statisticile F este de 0.882 care este mai mare de 1%, deci se acceptă ipoteza nulă.

**TESTUL CHOW**

Table 34. Testul CHOW

**

Probabilitățile 0.1058, 0.999, 0.4379 și 0.8996 sunt mai mari ca 1%, astfel nu există rupturi în serie si este suficient un singur model de regresie.

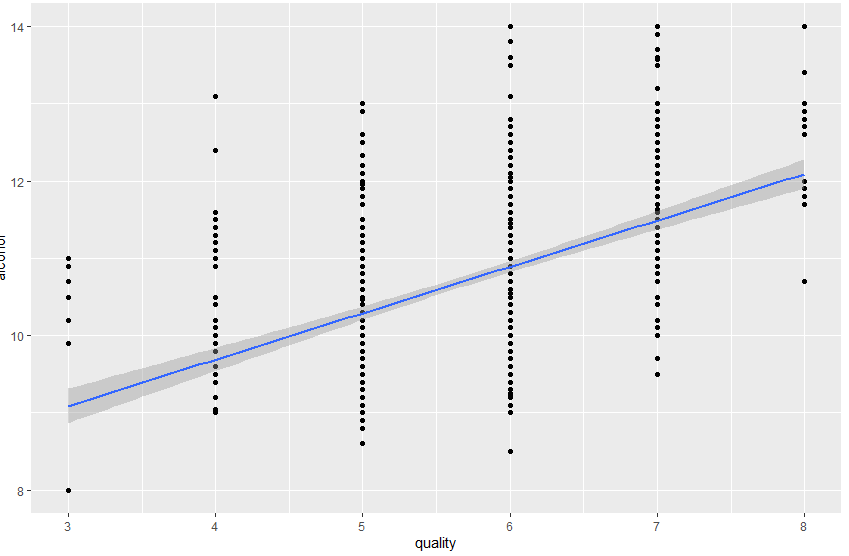


Figure 5.Modelul de regresie liniară

Comparam acuratetea si bonitatea modelelor

RMSE: LINIAR = 0.73

POLINOMIAL = 0.745

SPLINE = 0.74

GAM = 0.736

POLINOMIAL> SPLINE> GAM> LINIAR

R2: LINIAR = 19.74%

POLINOMIAL = 17.43%

SPLINE = 17.77%

GAM = 19.72%

LINIAR> GAM> SPLINE> POLINOMIAL

*Metode de penalitate*

**REGRESIA RIDGE**

În această parte, dorim estimarea unei valori lambda utilizând validarea încrucișată, astfel încât MSE să fie minimizat. Valoarea obținută generează un model Ridge valid.

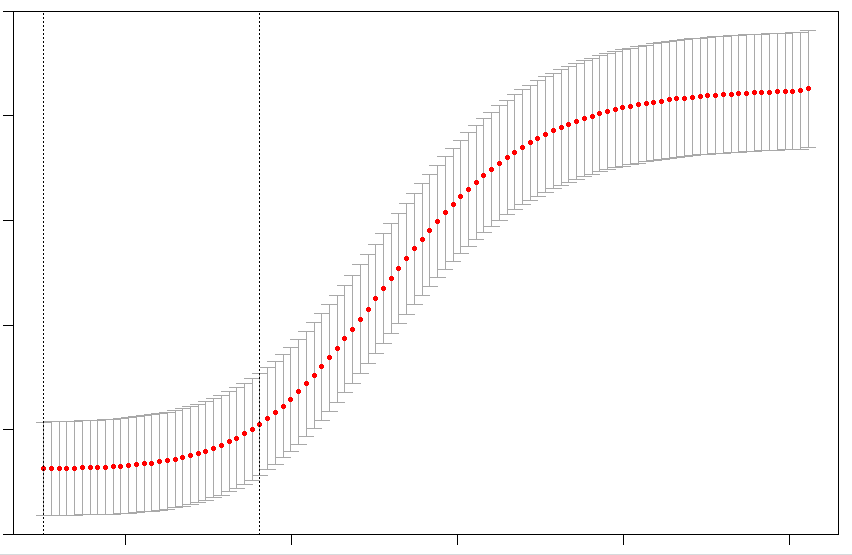


Figure 6.Valoarea lui lambda pt care avem MSE minimizat

Diagrama Trace(Figure 7.Diagrama Trace) pentru a vizualiza modul in care estimarile coeficientulilor s-au modificat ca urmare a cresterii valorii lui lambda.

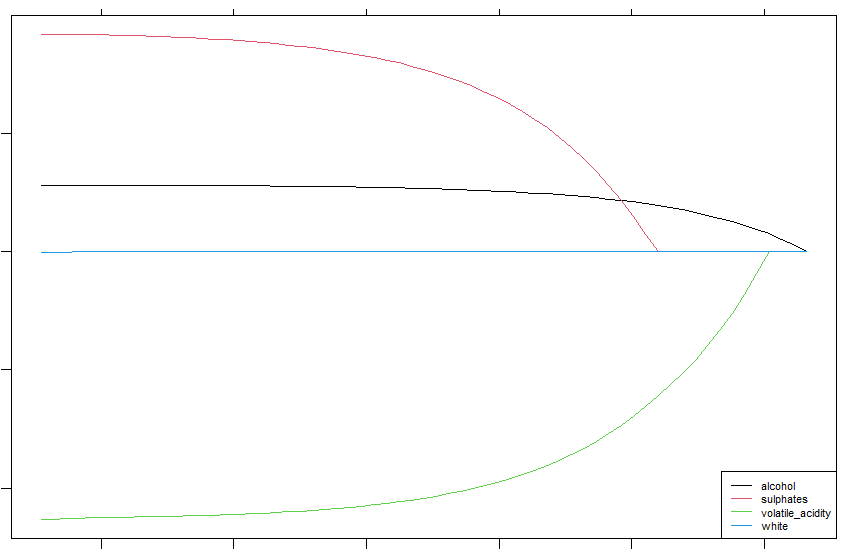


Figure 7.Diagrama Trace

O să folosim R2 pentru a putea compara modelele de penalitate, în cazul modelului Ridge, acesta fiind 30,16%.

**REGRESIA LASSO**

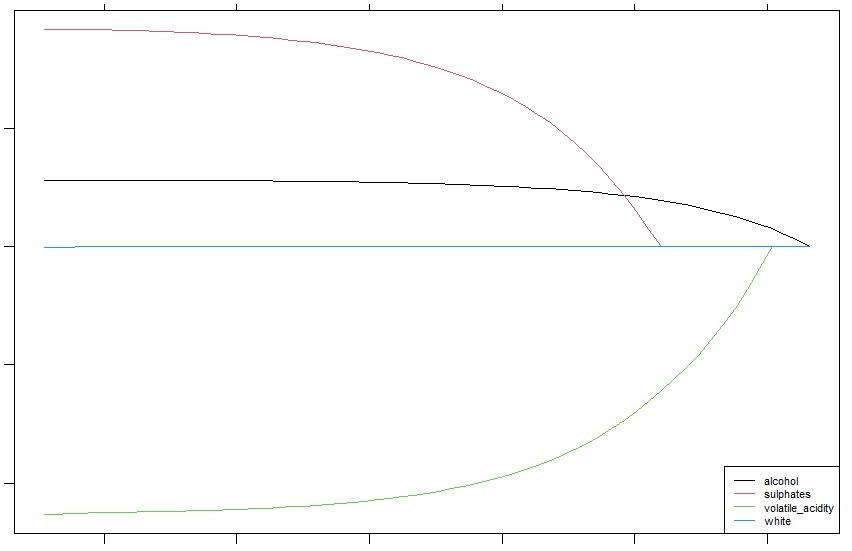


Figure 8.Diagrama Lasso

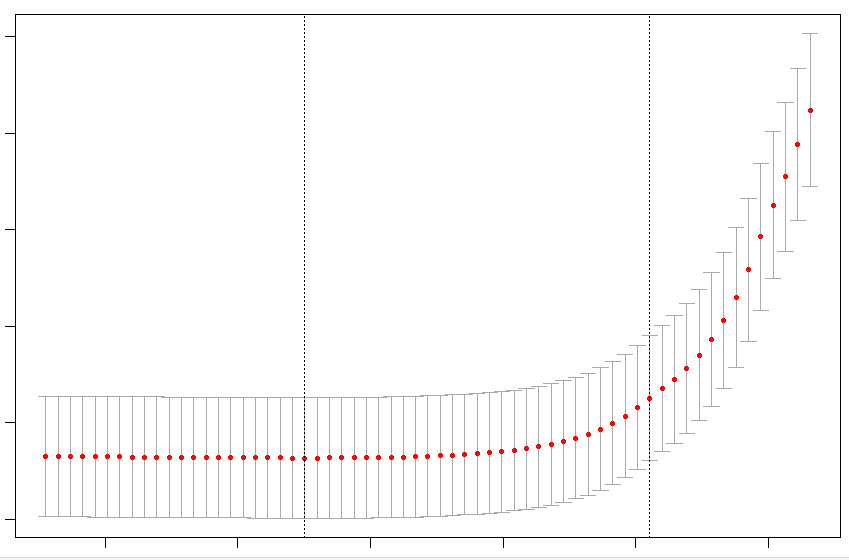


Figure 9.Testarea valorii lamda LASSO

Similar modelului Ridge, modelul Lasso folosește aceleași cinci variabile prezentate anterior, iar R2 rezultat din modelul Ridge este 30.19%.

**MODELUL ELASTIC NET**

Modelul Elastic Net reprezintă o combinație între Ridge și Lasso, dar pentru acest caz oferă o bonitate mai slabă, de 30.18%.

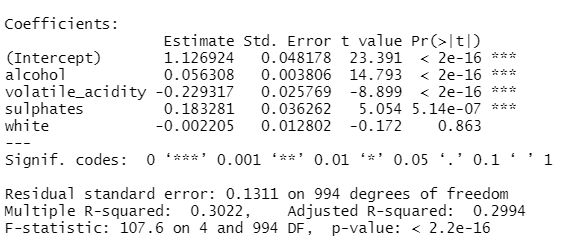
Comparând R2 pentru cele 3 cazuri, modelul cel mai bun este cel Lasso.

**ALGORITMUL BORUTA**

Încearcă să surprindă toate caracteristicile importante pe care le putea avea setul de date cu privire la o variabila rezultat.

Modelul reimplementat de algoritm este modelul de regresie multipla avand in componenta sa si variabila dummy. Astfel , a gasit aproape toate variabilele ca fiind importante in afara de dummy , insa bonitatea acestuia nu a fost modificata.

Table 35.Algoritmul Boruta

****

2.3. Concluzii

În urma analizelor efectuate asupra datelor s-au obținut două modele econometrice de regresie multifactorială. Cel de-al doilea, prin care s-a dorit îmbunătățirea modelului de regresie multiplă a rezultat drept valid statistic, în urma testării parametrilor prin Testul T, a reieșit faptul că aceștia sunt semnificativi statistici. In urma analizei indicilor de bonitate am comparat valorile lui R2 , iar modelul nou construit scadea valoarea lui R2 . Drept urmare, am hotărât să concluzionăm că modelul de regresie multifactorială în forma initiala este de fapt, cea mai bună versiune a modelului estimat.

Astfel, s-a obținut un model econometric de regresie liniară multifactorială cu o bonitate ridicată care surprinde modul în care alcoolul, sulfatii si acidul volatil influenteaza calitatea vinului.

Modelul multifactorial rezultat în urma estimării este următorul:

*CALITATEA = 1.12422 + 0. 05621\* ALCOOL +0.18599\* SULFATI- 0.22637 \* ACID-VOLATIL*

Limitarile modelului: Setul de date a fost limitat, rezumandu-se doar la compusi chimici. Nu există date despre tipurile de struguri, marca vinului, prețul de vânzare a vinului , acestia fiind principalii indici in vederea cumpararii vinului si acordandu-i o nota.

# **APLICAŢIA 3**

## **Model de date de tip panel**

### ***Literature review***

Fericirea este un obiectiv universal esențial al omului în viața lui, care poate îmbunătăți calitatea vieții. De la introducerea psihologiei pozitive, principala considerație a fost subliniată de studiul rolurilor anumitor factori în prezicerea fericirii, în special progresul tehnologiei.

Studiile internaționale privind bunăstarea subiectivă (SWB) arată diferențe consistente de nivel mediu între națiuni. Inglehart și Klingemann ( 2000 ) compară mai multe valuri de SWB la nivel de țară din World Values ​​Survey (WVS) și descoperă diferențe substanțiale între țări care persistă și sunt stabile.

În sondajul din 1998, 65% dintre danezi au fost foarte mulțumiți de viața lor, în timp ce doar 5 % dintre portughezi au declarat că sunt foarte mulțumiți. În mai multe sondaje de dinainte, proporția danezilor care erau foarte mulțumiți de viață a fost, de asemenea, de aproximativ 12 ori mai mare decât cea a portughezilor. Kenny ( 1999) constată, de asemenea, că variația indicelui de fericire al unei singure țări în perioade diferite este mult mai mică decât varianța indicelui fericirii din diferite țări în aceleași perioade.

În literatura de specialitate, evaluările bunăstării folosesc uneori întrebări care se referă la „satisfacția vieții”, uneori la „fericire” și uneori la bunăstarea subiectivă generală. Cu alte cuvinte, emoțiile sunt întotdeauna situate și încorporate în contexte culturale specifice. Cultura poate afecta fericirea în ceea ce privește cantitatea, gradul de fericire. De exemplu, o cultură a individualismului predomină în țările occidentale din Europa și America. Oamenii pun accent pe libertatea individuală, realizările individuale și urmărirea sentimentelor pozitive individuale.

Conform teoriei lui Maslow, după satisfacerea nevoilor de nivel inferior și supuse nivelurilor de abilități, generozitatea și ajutorul celorlalți sunt nevoi de nivel superior care pot spori nivelul de fericire . A fi amabil cu ceilalți îi face pe indivizi să se simtă bine.

Teoria comparației sugerează că satisfacerea nevoilor materiale are contribuții limitate la creșterea fericirii. Chiar dacă nivelul veniturilor crește, efectele asupra fericirii pot să nu fie semnificative. Conform teoriei nevoilor, nevoile de bază precum nevoile materiale pot fi satisfăcute cu ușurință, dar efectele satisfacerii acestor nevoi asupra fericirii sunt limitate, cu o utilitate marginală în scădere. Pe de altă parte, satisfacția la nivelurile spirituale superioare este nelimitată.

### ***3.1.2. Metodologia cercetării***

Forma modelului aplicației este următoarea:

*Y =* α *+* β 1*\** X1 *+* β2 *\** X2 + β3 *\** X3 + β4 *\** X4

Dreapta de regresie estimată este:

Y = 206 - 81 \* X1  - 41 \* X2 - 68 *\** X3 - 63*\** X4

unde:

Y – variabila dependentă, rang fericire

X1 – variabila independentă, libertate

X2 – variabila independentă, PIB pe cap de locuitor

X3 – variabila independentă, speranța sănătății

X4 – variabila independentă, generozitate

α = 206 – arată nivelul rangului de fericire;

β 1 = -81 – arată că atunci când libertatea crește cu o unitate (1%), rangul fericirii va scădea, în medie, cu 81 de unități;

β2 = -41 – arată că atunci când PIB-ul pe cap de locuitor crește cu o unitate (1%), rangul fericirii va scădea, în medie, cu 41 de unități;

β3= - 68 – arată că atunci când speranța sănătății crește cu o unitate (1%), rangul fericirii va scădea, în medie, cu 68 de unități;

β4= - 63 – arată că atunci când generozitatea crește cu o unitate (1%), rangul fericirii va scădea, în medie, cu 63 de unități;

Astfel, cu ajutorul soft-ului R Studio am estimat modelul ,am testat validitatea acestuia prin calcularea indicatorilor de bonitate, aceștia fiind coeficientul de determinație, coeficientul de nedeterminație, coeficientul de determinație ajustat. Totodată, am testat efectele fixe, ipotezele modelului de regresie liniară.În cadrul aplicației sunt utilizate următoarele teste și criterii statistice.

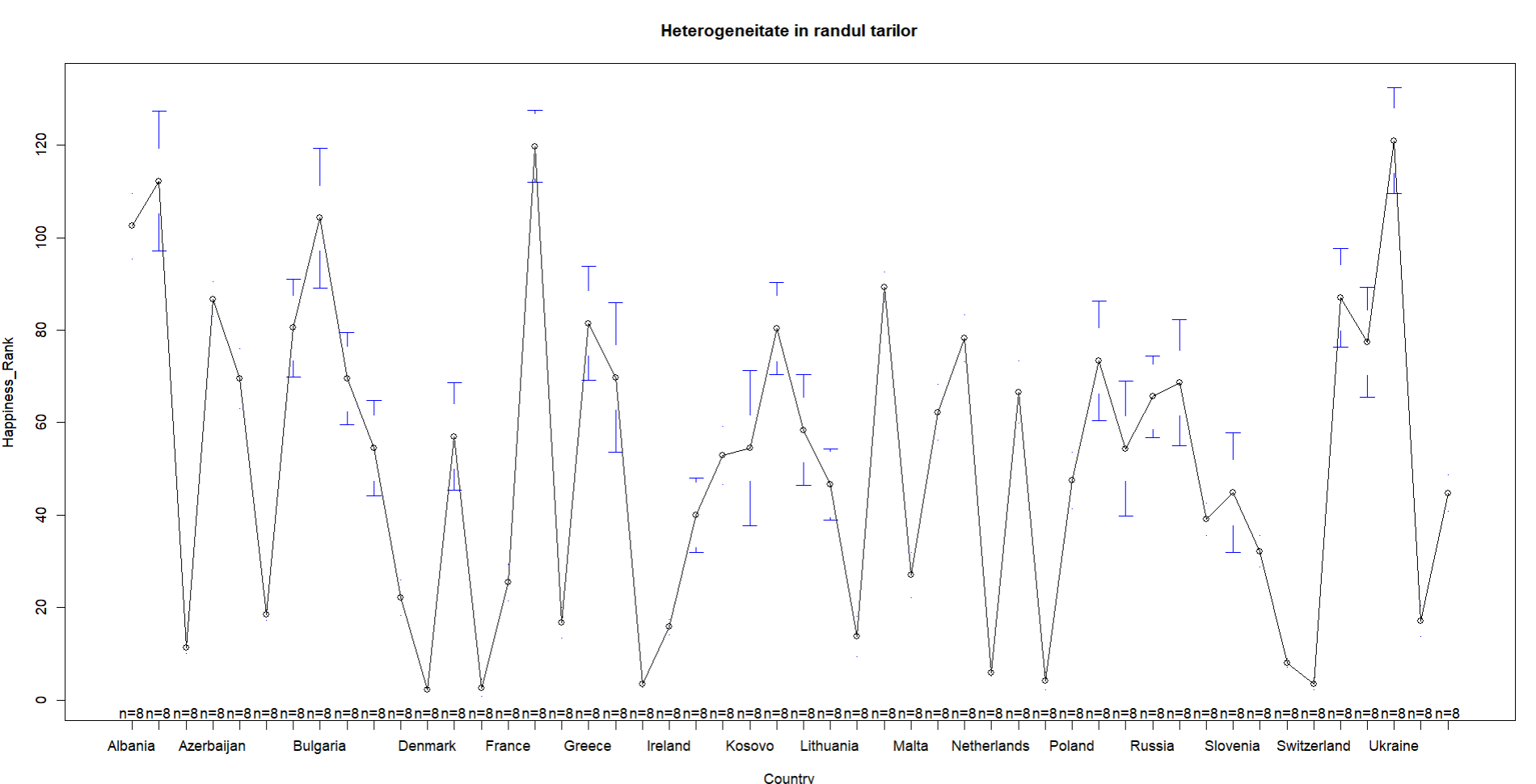
***3.2.1. Date utilizate***

Proiectul investighează ce este nevoie pentru a construi o țară fericită analizând relația dintre clasamentul fericirii țărilor și factorii lor la nivel macroeconomic. Arată modul în care cinci variabile cheie contribuie la explicarea eșantionul complet de scoruri medii anuale naționale pe toată perioada 2015 - 2022, au fost utilizate date pentru țările din Europa. Aceste date au fost selectate de pe site-ul Kaggle.

Datele utilizate au fost: rang fericire, libertate, PIB pe cap de locuitor,speranța de viață a sănătății și generozitate.

3.2.2. Rezultatele empirice ale cercetării

* Explorarea heterogeneitatii in sectiunea transversala

Figure 10.Explorarea heterogeneitatii in sectiunea transversala

* Explorarea heterogeneității în secțiunea temporală

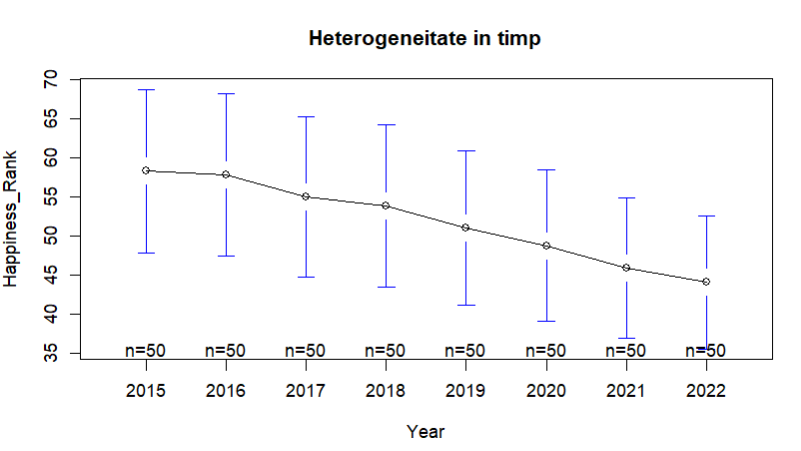
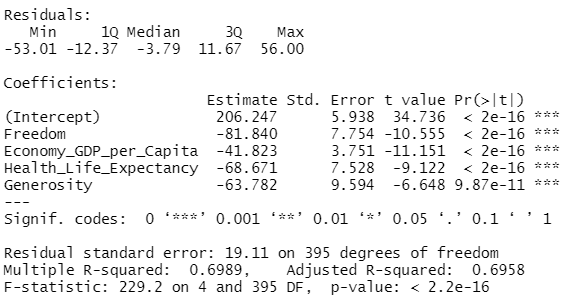


Figure 11.Explorarea heterogeneitatii in sectiunea temporala

* Model OLS - model clasic de regresie liniara

Table 36. Model OLS



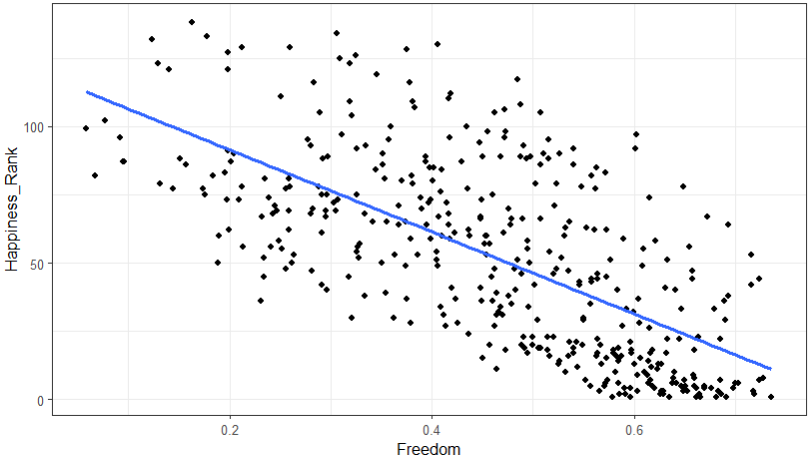
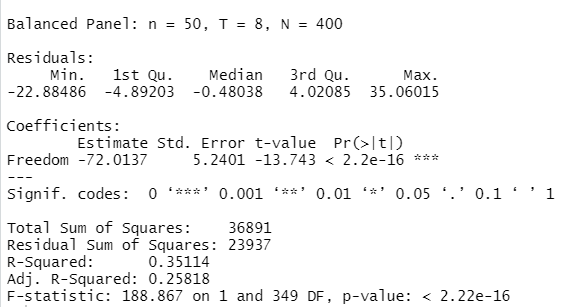


Figure 12.Grafic OLS

* Model FE (cu efecte fixe)

Freedom este cea mai semnificativă, vom estima din nou modelul ținând cont doar de freedom.

Table 37. Model FE



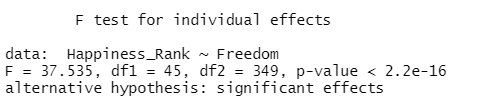
P-value < 0,05 => modelul este ok, toți coeficienții din model sunt diferiti de zero.

* Alegerea celei mai adecvate variante de model prin testarea între regresie OLS vs fixed effects panel model.

H0: FE

H1: OLS

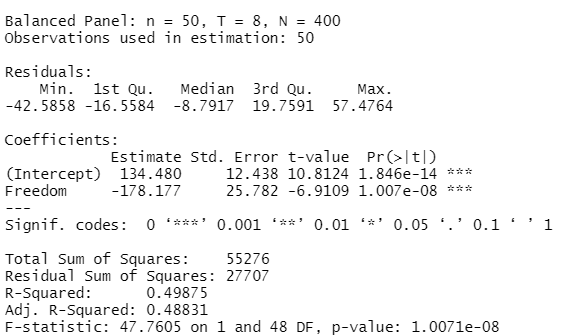
Table 38. Test FE



P-value < 0.05 => se recomandă model de panel data FE.

* Model cu efecte aleatorii RE (random effects)

Table 39.Model RE

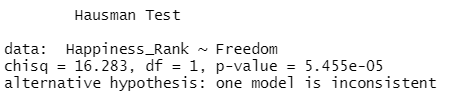


* Testul Hausmann il utilizam pentru a decide intre FE si RE

H0: model cu efecte random

H1: model cu efecte fixe

Table 40. Test Hausman



P-value <0.05 => se recomandă model cu efecte fixe .

*Indicatorii de bonitate*

Pentru aflarea calității ajustării ,calculăm indicatorii de bonitate:

**a. Coeficientul de determinație**

R2 = R-squared = 0,6989= > modelul explică 69,89% din variația rangului de fericire.

**b. Coeficientul de nedeterminație**

K2 = 1- R2 = 0,3011 => acțiunea altor factori neincluși în model și cuprinși în eroare explică aproximativ 30,11% din rangul de fericire.

**c. Coeficientul de determinație ajustat**

Adjusted R-squared= 0,6958. Rezultă că aproximativ 29.41% din variația rangului de fericire este explicată de libertate , PIB pe cap de locuitor, speranța de viață a sănătății și generozitate.

*Testare ipoteze model*

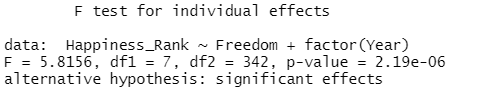
În urma testului Hausman am decis sa utilizam modelul FE.

**Testarea efectelor fixe in timp**

Ipoteze: H0: nu sunt necesare efectele fixe în timp

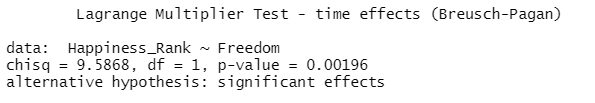
H1: sunt necesare efectele fixe în timp

Table 41.Testare efecte fixe în timp - Ftest



P-value < 0.05 => se recomandă folosirea efectelor fixe în timp

Table 42.Testare efecte fixe în timp - LM



P-value=0.0019 <0.05 => se recomandă folosirea efectelor fixe în timp

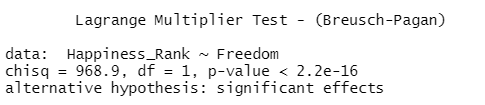
**Testarea efectelor aleatorii cu Breusch-Pagan Lagrange Multiplier**

Testul ne ajută sa decidem între RE și OLS

H0: variațiile în timp sunt 0

H1: variațiile în timp sunt diferite de 0

Table 43.Testare efecte aleatorii



P-value < 0.05 => respingem ipoteza nulă variațiile în timp sunt diferite de 0 => efectele aleatorii sunt adecvate a.î. există diferențe semnificative între tări.

**Testarea dependenței transversale**

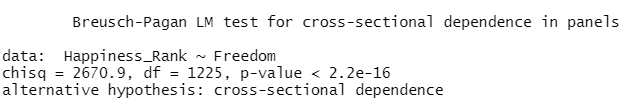
Testarea dependentei transversale folosind testul Breusch-Pagan LM și testul Parasan CD

Ipoteze teste:

H0: rezidurile între entități nu sunt corelate

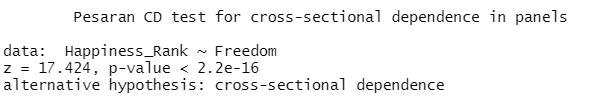
H1: rezidurile între entități sunt corelate

Table 44.Testare dependență transfersală



P-value < 0.05 => există dependență transversală

Table 45. Testare dependență transversală - Pesaran



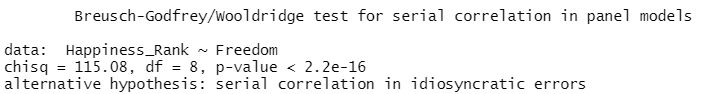
P-value < 0.05 => există dependență transversală și nu corectam pentru ca avem panel mic de 8 perioade.

**Testarea autocorelarii - Breusch-Godfrey**

H0: nu există autocorelate

H1: autocorelarea este prezentă

Table 46.Testare autocorelare



P-value <0.05 => avem autocorelare dar fiind panelul mic o vom ignora.

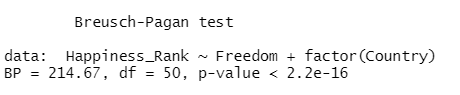
**Testarea heteroschedasticitatii**

Testarea heteroschedasticitatii cu testul Breusch-Pagan

H0: homoschedasticitate

H1: heteroschedasticitate

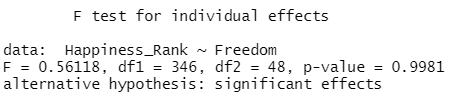
Table 47.Testare heteroschedasticitate



Deoarece p-value <0.05 => avem heteroschedasticitate

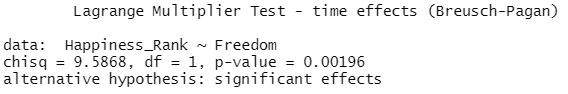
**Testarea efectelor random**

Table 48. Testare efecte random - F test



P-value=0.9 > 0.05 => nu se recomanda efecte random.

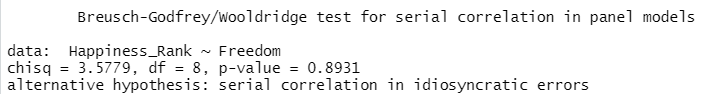
Table 49.Testare efecte random - Breusch-Pagan LM



P-value=0.0019 < 0.05 => recomanda efecte random

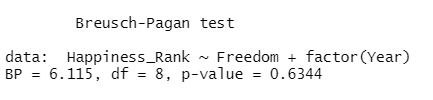
Este incertă situația efectelor random și vom accepta ca nu vom recomdanda efecte random. Testul Hausman nu recomandă efectele random dar vom aplica testele pentru corectarea autocorelării și heteroschedasticității.

Table 50.Corecție autocorelare



P-value=0.8931 > 0.05 => nu exista autocorelare

Table 51. Corectare heteroschedasticitate



P-value=0.6344 > 0.05 => nu exista heteroschedasticitate

3.3. Concluzii

În urma prelucrărilor efectuate asupra datelor s-a obținut un model econometric de regresie liniară cu o bonitate medie care surprinde modul în care libertatea influențează rangul de fericire pentru setul de date de tip panel, ce cuprind țările din Europa. Coeficientul de determinare confirmă faptul ca modelul explică 69,89% din variația rangului de fericire.

Modelul rezultat în urma estimării este următorul:

*RANG FERICIRE = 206 – 81 \* LIBERTATE - 41 \* PIB PE CAP DE LOCUITOR - 68 \* SPERANȚA SĂNĂTĂȚII - 63 \* GENEROZITATE*

Bibliografie

<https://divinea.com/blog/en/wine-history-and-the-digital-transformation/>

<https://www.researchgate.net/publication/266042753_Quality_Attributes_of_Wine_Products_an_Explorative_Study_of_Consumers'_Buying_Motivation_through_a_Means-End_Chains_Approach>

<https://www.sciencedirect.com/science/article/pii/B9780128018132000082>

<https://ro.wikipedia.org/wiki/Sulfat>

<https://ro.differbetween.com/article/difference_between_volatile_and_nonvolatile_acids>

<https://link.springer.com/article/10.1007/s11205-014-0747-y>

<https://media-01.imu.nl/storage/heart4happiness.nl/2097/wp/2019/03/WHR19.pdf#page=13>