Academia de Studii Economice

Facultatea de Cibernetică, Statistică și Informatică Economică

**PROIECT ECONOMETRIE**

**Modelul multifactorial de regresie liniară**

**Profesor coordonator: Proiect realizat de :**

Simona Apostu Bălan Diana-Iuliana,

Gupa 1075, seria C

Înregistrați pentru cel puțin 30 de unitati, valorile specifice ale unor caracteristici (X1, X2 si Y) ȋntre care există o legătură logică. Se cer următoarele:

**1. Prezentarea problemei**

Se presupune că între variabilele suprafața locuibilă, salariu și număr de salariați există o dependență liniară. Pentru a se verifica această ipoteză se înregistrează următoarele date privind suprafața locuibilă, salariul și număr de salariați în anul 2018 pentru cele 42 de unități teritoriale (județe) ale țării:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nr.Crt** | **Judet** | **Suprafața locuibilă** | **Salariu** | **Nr.Salariați** |
| 1 | Bihor | 564109 | 2087 | 175795 |
| 2 | Bistrita-Nasaud | 280106 | 2044 | 70635 |
| 3 | Cluj | 704759 | 3026 | 253827 |
| 4 | Maramures | 463354 | 2127 | 111202 |
| 5 | Satu Mare | 334678 | 2173 | 84467 |
| 6 | Salaj | 213816 | 2177 | 52184 |
| 7 | Alba | 328311 | 2375 | 96781 |
| 8 | Brasov | 551183 | 2609 | 193784 |
| 9 | Covasna | 203534 | 2098 | 52641 |
| 10 | Harghita | 304280 | 2059 | 71206 |
| 11 | Mures | 538329 | 2459 | 143011 |
| 12 | Sibiu | 400110 | 2626 | 142498 |
| 13 | Bacau | 591035 | 2270 | 115322 |
| 14 | Botosani | 385046 | 2174 | 57813 |
| 15 | Iasi | 791210 | 2691 | 178415 |
| 16 | Neamt | 447102 | 2083 | 88728 |
| 17 | Suceava | 626789 | 2118 | 112022 |
| 18 | Vaslui | 380001 | 2161 | 56948 |
| 19 | Braila | 294143 | 2119 | 72876 |
| 20 | Buzau | 419829 | 2170 | 85785 |
| 21 | Constanta | 676215 | 2377 | 184562 |
| 22 | Galati | 509471 | 2290 | 119110 |
| 23 | Tulcea | 197754 | 2298 | 47368 |
| 24 | Vrancea | 324545 | 2083 | 61191 |
| 25 | Arges | 585730 | 2562 | 162599 |
| 26 | Calarasi | 288043 | 2203 | 44999 |
| 27 | Dambovita | 496173 | 2243 | 86506 |
| 28 | Giurgiu | 269279 | 2340 | 36355 |
| 29 | Ialomita | 258931 | 2162 | 45983 |
| 30 | Prahova | 725609 | 2449 | 186641 |
| 31 | Teleorman | 341440 | 2079 | 56476 |
| 32 | Ilfov | 473445 | 2871 | 158567 |
| 33 | Municipiul Bucuresti | 1827810 | 3666 | 1035124 |
| 34 | Dolj | 630911 | 2458 | 137462 |
| 35 | Gorj | 319919 | 2387 | 77333 |
| 36 | Mehedinti | 245120 | 2184 | 47593 |
| 37 | Olt | 400802 | 2343 | 75381 |
| 38 | Valcea | 353188 | 2104 | 82188 |
| 39 | Arad | 419360 | 2388 | 139595 |
| 40 | Caras-Severin | 275063 | 2142 | 54985 |
| 41 | Hunedoara | 388600 | 2088 | 112027 |
| 42 | Timis | 701499 | 2922 | 258287 |

**Tabel 1 Date privind suprafața locuibilă, salariul și numărul de salariați în anul 2018 pentru cele 42 de județe ale țării**

**Sursa datelor:** <http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table>

**2. Definirea modelului de regresie**

**2.1. Forma, variabilele și parametrii modelului de regresie**

Modelul de regresie multifactorial are în vedere dependența dintre variabilele X (variabile exogene sau factoriale) și variabila Y (variabilă endogenă sau rezultativă).

Forma generală a modelul poate fi exprimată prin următoarea ecuație:

y= βo + β1x1+ β2x2 + ε

unde Y reprezintă variabila endogenă sau rezultativă, βo și β1 reprezintă parametrii ecuației de regresie și ε reprezintă componenta reziduală (eroare aleatoare).

Variabilele modelului sunt: variabila dependentă y ce reprezintă suprafața locuibilă măsurată în număr de persoane, x1 variabila independentă salariu măsurat în lei și x2 reprezintă variabilă independentă număr de salariați măsurat în număr de persoane.

**2.2.Aproximarea grafică a modelului legăturii dintre variabile**

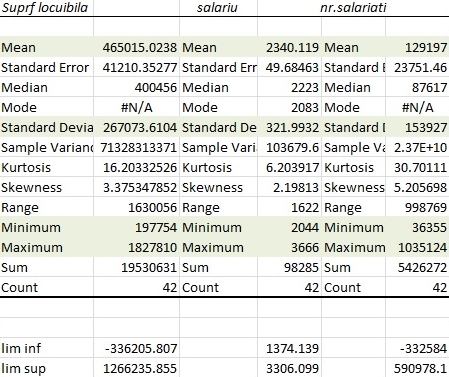
Pe baza datelor preluate, vom realiza două grafice care să ne arate legătura dintre variabile:

Figura 1. Suprafața locuibilă(nr pers) cu salariul (LEI)

Figura 2. Suprafața locuibilă(nr pers) cu nr.salariați

**3. Analiza preliminară a datelor**

**3.1. Verificarea preliminară a datelor**

****

Regula celor trei sigma:

ϵ(±3)

-336205.807<y<1266235.855 Exista 1 outlier.

: ϵ(±3)

1374.139<x1<3306.099 Exista 1 outlier

: ϵ(±3)

-332585<x2<590978.1 Exista 1 outlier

**3.1.Factorii independenți (ipoteza de multicoliniaritate)**

Pentru a testa ipoteza de multicoliniaritate, am calculat corelația dintre variabila x1=salariu și x2=număr salariați, folosind funcția din Excel Correl. Valorea rezultată o vom compara cu cea a lui R Square=0,88.

CORREL(x1,x2)=0.821216. Valoarea calculată este mai mica decât R Square => nu exista multicoliniaritate.

**4. Estimarea parametrilor modelului**

**4.1. Estimarea punctuală a parametrilor (MCMMP)**

Estimarea parametrilor βo, β1, β2 ce rezultă din tabelul ANOVA:

|  |  |
| --- | --- |
|  | *Coefficients* |
| Intercept | 129622.0682 |
| x1 | 59.06588342 |
| x2 | 1.526133188 |

bo =129622.0682

Interpretare

b0 = intercept, termen liber sau constantă. Nu are interpretate din punct de vedere economic, dar este valoarea pe care o poatea lua y atunci când x nu este semnificativ diferit de 0.

b1=59.06588342

Interpretare

b1 este coeficientul lui x1 (salariu), semnul îmi arată direcția legăturii => legătură directă. Când salariul crește cu 59.06 lei ,suprafața locuibilă (măsurată în nr de persoane), crește cu o persoană.

b2=1.526133188

b2 este coeficientul lui x2(număr salariați) semnul îmi arată direcția legăturii => legătură directă. Când numărul de salariați crește cu 1,5 persoane, suprafața locuibilă crește cu o persoană.

**4.2. Estimarea parametrilor prin interval de încredere**

|  |  |  |  |
| --- | --- | --- | --- |
| *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| -215665.0469 | 474909.2 | -215665 | 474909.2 |
| -102.9246085 | 221.0564 | -102.925 | 221.0564 |
| 1.187272401 | 1.864994 | 1.187272 | 1.864994 |

βo ϵ[-215665.0469, 474909.2]

Cu o probabilitate de 95%, βo poate lua valori de la -215665 până la 474909.

β1 ϵ[-102.9246085, 221.0564]

Cu o probabilitate de 95%, β1 poate lua valori de la -102.9 până la 221.

β2ϵ[1.187272401, 1.864994]

Cu o probabilitate de 95%, β2 poate lua valori de la 1.18 până la 1.86.

**5. Testarea semnificației corelației si a parametrilor modelului de regresie**

**5.1. Testarea parametrilor modelului de regresie**

Vom compara P-value pentru βo, β1, β2 cu pragul de semnificație ∝ = 0.05:

|  |
| --- |
| *P-value* |
| 0.452223 |
| 0.465218 |
| 3.35E-11 |

Ipoteze: βo: H0: βo=0

H1: βo ≠0

P-value βo=0.452223>∝ = 0.05 rezultă că estimatorul βo nu este semnificativ din punct de vedere statistic. Așadar, acceptăm ipoteza H0, nu avem suficiente dovezi să acceptăm H1 => βo=0

Ipoteze: β1: H0: β1=0

H1: β1 ≠0

P-value β1=0.465218>∝ = 0.05 rezultă că estimatorul βo nu este semnificativ din punct de vedere statistic. Așadar, acceptăm ipoteza H0, nu avem suficiente dovezi să acceptăm H1 => β1=0

Ipoteze: β2: H0: β2=0

H1: β2 ≠0

P-value β2=3.35E-11< ∝ = 0.05 rezultă că estimatorul β2 este semnificativ din punct de vedere statistic. Așadar, respingem H0, acceptăm H1 => β2 ≠0

**5.2. Testarea semnificației corelației (validitatea modelului)**

Prin compararea lui Significance F și a pragului de semnificație ∝ = 0.05 rezultă că:

Significance F=8.50E-19 < ∝ = 0.05 => modelul este valid.

Coeficientul de corelație Multiple R= 0.938 ∈ [0,1] => legătură puternică.

Coeficientul de determinare R Square= 0.881∈ [0,1] ne arată 88,1% din variația suprafeței locuibile este explicata de variața salariului și a numărului de salariați.

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

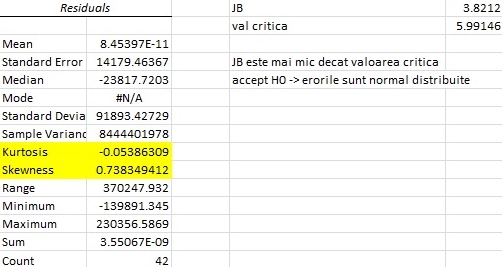
**6.1. Testarea normalității erorilor**

Pentru a teste normalitatea erorilor, vom lua în considerare următoarele ipoteze:

H0: erorile sunt normal distribuite

H1: erorile nu sunt normal distribuite

Statistica descriptivă:



Pe baza datelor din statistica descriptivă, am folosit formula:

JB=n(Skewness^2/6+Kurtosis^2/24)=3.8212

JB<valoarea critică => accept H0, resping H1 , erorile sunt normal distribuite.

**6.1. Testarea ipotezei de homoscedascititate**

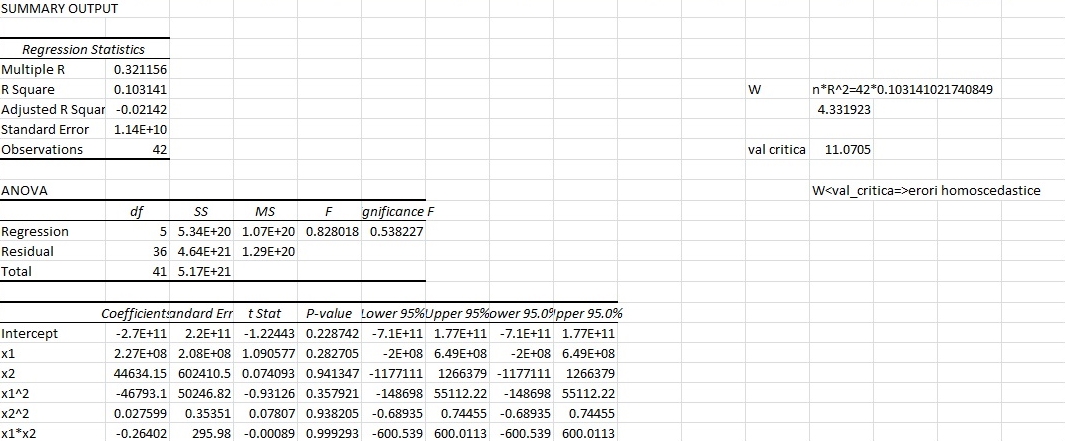
Pentru a testa ipoteza de homoscedascititate, am realizat următorul tabel:



Ipotezele:

H0: erori=homoscedastice

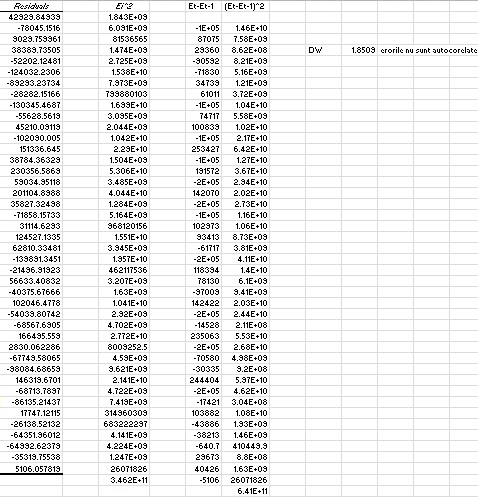
H1: erori=heteroscedastice



Am calcul LM=n\*R Square din regresia auxiliară

LM=4,331923 < valoarea critică => accept H0, resping H1 => erori homoscedastice

**6.3. Testarea ipotezei de autocorelare a erorilor**



DW=

DW=1.85095 erorile nu sunt autocorelate

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

𝐲̂= b0 + b1∙ x1 + b2 ∙ x2

Estimarea valorii suprafetei locuibile dacă ar mai exista un județ, iar salariul ar fi de 2300 lei, numărul de salariați de 21 450.

𝐲̂=129622+1.526\*2300+59.066 \* 21450

𝐲̂=129622+135851.8+32732

𝐲̂=298205,8 persoane

Daca ar mai exista un judet,iar salariul ar fi de 2300 lei si numarul de salariati de 21450 de persoane, atunci suprafata locuibila masurata in nr de persoane ar fi de 298205,8 persoane.

**8.Concluzii**

Conform modelului prezentat, se poate afirma ca intre suprafata locuibila si salariu exista o legatura directa, la fel si intre numarul de salariati si suprafata locuibila pentru cele 42 de judete analizate. Acest lucru poate fi observant atat din semnul coeficientilor, cat si din graficele din Figura 1, respectiv Figura 2 prezentate in capitolul 2.2.

De asemenea, se poate concluziona ca modelul de regresie explica in proportie de 88,16% variatia suprafetei locuibile. (R Square=0,881612).

Sursa datelor este Tempo [ http://statistici.insse.ro:8077/tempo-online/].