**Sustainable Development of Railway Transport: Assessing the Impact of Electrified Lines on Greenhouse Gas Emissions using ….**

Avem două baze de date (tabelul 1 si tabelul 2) acronimul pentru indicatorul din tabelul 1 e RTLEL (variabila independentă) . Acronimul pentru tabelul 2 este GHGET( variabila dependentă).

1. **Trebuie să calculăm coeficientul de corelație Pearson pentru perechile RTLEL/GHGET pentru fiecare țară din UE prezentă în cele două tabele**.

Formula este:

The sample Pearson`s r is calculated with the following formula (1):

Exemplu pentru România:

|  |  |  |
| --- | --- | --- |
| **An** | **RTLEL** | **GHGET** |
| 2012 | 4020 | 4328758 |
| 2013 | 4029 | 4572861 |
| 2014 | 4029 | 4377449 |
| 2015 | 4030 | 4717626 |
| 2016 | 4030 | 5280591 |
| 2017 | 4030 | 5610208 |
| 2018 | 4029 | 5686712 |
| 2019 | 4029 | 5935289 |
| 2020 | 4034 | 5484805 |
| 2021 | 4035 | 5774825 |

Aplicați testul de corelație Pearson în Python:

corr, p\_value = stats.pearsonr(RTLEL, GHGET)

r= 0.58152412 (calculat cu Excel: Data Analysis, Correlation)

1. **Calculăm apoi coeficientul de determinare**

r2 = r\*r = 0.338170302

1. P-value

Excelul face asta cu Data Analysis, Regression (vezi mai jos)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R (coeficientul r) | 0.581524 |  |  |  |  |  |  |  |
| R Square | 0.33817 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.255442 |  |  |  |  |  |  |  |
| Standard Error | 531925.5 |  |  |  |  |  |  |  |
| Observations | 10 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 1.16E+12 | 1.16E+12 | 4.087702 | 0.07784 |  |  |  |
| Residual | 8 | 2.26E+12 | 2.83E+11 |  |  |  |  |  |
| Total | 9 | 3.42E+12 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | -3.6E+08 | 1.8E+08 | -1.99297 | 0.081397 | -7.7E+08 | 56205840 | -7.7E+08 | 56205840 |
| km | 90091.32 | 44559.81 | 2.021807 | 0.07784 | -12663.8 | 192846.4 | -12663.8 | 192846.4 |

P-value = 0.077840229

Dpdv statistic avem un "test bilateral" sau "test cu două cozi". T-Distribution

Level of confidence: LOC=95%;

Level of significance: α = 0,05

Number of observations: n=10;

Degree of freedom: Dof=8.

Acesta este un mod general de a face acest lucru:

1. Importați pachetul necesar în Python:

python

import scipy.stats as stats

1. Utilizați datele furnizate:

python

ani = [2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021]

RTLEL = [4020, 4029, 4029, 4030, 4030, 4030, 4029, 4029, 4034, 4035]

GHGET = [4328758, 4572861, 4377449, 4717626, 5280591, 5610208, 5686712, 5935289, 5484805, 5774825]

1. Aplicați testul de corelație Pearson:

python

corr, p\_value = stats.pearsonr(RTLEL, GHGET)

1. Afișați rezultatul:

python

print(f"Correlation coefficient: {corr}")

print(f"P-value: {p\_value}")

Dacă p-value este mai mic decât 0,05 (sau alt nivel de semnificație pe care l-ați ales), atunci puteți respinge ipoteza nulă și puteți spune că există o corelație semnificativă statistic între cele două seturi de date. Dacă p-value este mai mare decât 0,05, nu puteți respinge ipoteza nulă și nu puteți spune că există o corelație semnificativă statistic între cele două seturi de date.

**Testul Breusch-Pagan:** Acesta este un test pentru heteroscedasticitate în modelele de regresie. Testează dacă variabilele explicative în modelul de regresie au o varianță egală.

Codul pentru Testul Breusch-Pagan în limbajele de programare populare ca R și Python sunt următoarele:

**1. În R folosind pachetul lmtest:**

Pentru a efectua testul Breusch-Pagan în R, puteți folosi pachetul lmtest. Dacă nu l-ați instalat deja, faceți-o cu:

R

install.packages("lmtest")

Apoi puteți utiliza funcția bptest() pentru a efectua testul:

R

library(lmtest)

# Creează un model de regresie liniară

model <- lm(y ~ x1 + x2, data = your\_data)

# Efectuează testul Breusch-Pagan

test\_result <- bptest(model)

print(test\_result)

**2. În Python folosind pachetul statsmodels:**

Pentru a efectua testul Breusch-Pagan în Python, puteți utiliza pachetul statsmodels. Dacă nu l-ați instalat deja, faceți-o cu:

bash

pip install statsmodels

Apoi puteți utiliza funcția het\_breuschpagan() pentru a efectua testul:

python

import numpy as np

import statsmodels.api as sm

from statsmodels.compat import lzip

from statsmodels.stats.diagnostic import het\_breuschpagan

# Datele dvs., presupunând că sunt stocate în numpy arrays

y = np.array(your\_y\_data)

X = np.column\_stack((your\_x1\_data, your\_x2\_data))

X = sm.add\_constant(X) # Adăugați o coloană de 1 pentru intercept

model = sm.OLS(y, X).fit()

bp\_test = het\_breuschpagan(model.resid, model.model.exog)

labels = ['LM Statistic', 'LM-Test p-value', 'F-Statistic', 'F-Test p-value']

print(dict(zip(labels, bp\_test)))

În ambele cazuri, trebuie să înlocuiți your\_data, your\_y\_data, your\_x1\_data, și your\_x2\_data cu seturile dvs. de date. Dacă p-value este sub un nivel de semnificație (de obicei 0.05), atunci respingeți ipoteza nulă de homoscedasticitate, indicând prezența heteroscedasticității. Heteroscedasticitatea se referă la situația în care varianța erorilor (sau reziduurilor) dintr-un model de regresie nu este constantă pe întreaga gamă de valori ale variabilei independente. În termeni simpli, variabilitatea erorii se schimbă pe măsură ce valoarea variabilei explicative se schimbă.

Testul Breusch-Pagan se face doar dacă p-value este mai mic decât 0,05.

Pentru țările cu relații semificative din punct devedere statistic între cei doi indicatori (RTLEL și

GHGET) generăm ecuațiade regresie:

The least squares regression line, often simply called the regression line, is the line that best fits a set of data points in such a way that it minimizes the sum of the squared vertical distances from each data point to the line.

The equation developed is of the form y = mx + b, where **m** is the slope of the regression line (or the regression coefficient), and b is where the line intersects the y-axis. The equation for the regression line can be found using the least squares method, where

m = (n(Σxy) − ΣxΣy)/(nΣx2 − (Σx)2)

and b = (Σy − mΣx)/n.

**Table 1: Railway transport - length of electrified lines, Total, Kilometre**

| **Country** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Belgium | 3055 | 3055 | 3088 | 3088 | 3086 | 3092 | 3095 | 3100 | 3099 | 3127 |
| Bulgaria | 2862 | 2869 | 2861 | 2859 | 2869 | 2870 | 2870 | 2869 | 2871 | 3001 |
| Czechia | 3217 | 3216 | 3216 | 3237 | 3236 | 3237 | 3235 | 3231 | 3236 | 3234 |
| Denmark | 621 | 621 | 621 | 621 | 621 | 626 | 730 | 730 | 776 | 803 |
| Germany | : | : | : | 20726 | : | : | : | : | 21100 | : |
| Estonia | 132 | 132 | 132 | 132 | 132 | 132 | 132 | 138 | 138 | 225 |
| Ireland | 145 | 145 | 145 | 145 | 145 | 145 | 53 | 53 | 53 | 53 |
| Greece | 438 | 437 | 494 | 494 | 520 | 532 | 679 | 731 | 731 | 731 |
| Spain | 8786 | 9275 | 9223 | 9717 | 10383 | 10122 | 9840 | 10252 | 10419 | 10428 |
| France | 16116 | 15858 | 16031 | 15987 | 16097 | 16052 | 16053 | 16067 | 16013 | 16054 |
| Croatia | 984 | 985 | 970 | 970 | 970 | 970 | 970 | 970 | 970 | 994 |
| Italy | 11931 | 11969 | 11996[[1]](#footnote-1) | 12010[[2]](#footnote-2) | 12023 | 12022 | 12018 | 12016 | 12065 | 12160 |
| Latvia | 250 | 248 | 251 | 251 | 251 | 251 | 251 | 251 | 251 | 251 |
| Lithuania | 122 | 122 | 122 | 122 | 122 | 152 | 152 | 152 | 152 | 152 |
| Luxembourg | 262 | 262 | 262 | 262 | 262 | 254 | 254 | 254 | 254 | 254 |
| Hungary | 2982 | 2969 | 2968 | 2963 | 3018 | 3103 | 3069 | 3111 | 3111 | 3221 |
| Netherlands | 2266 | 2266 | 2307 | 2302 | 2314 | 2310 | 2224 | 2224 | 2225 | 2264 |
| Austria | 3852 | 3854 | 3874 | 3905 | 3926 | 3946 | 3951 | 3976 | 3992 | 4003 |
| Poland | 11920 | 11868 | 11830 | 11865 | 11874 | 11854 | 11858 | 11982 | 12113 | 12101 |
| Portugal | 1630 | 1629 | 1630 | 1639 | 1639 | 1639 | 1639 | 1696 | 1696 | 1791 |
| Romania | 4020 | 4029 | 4029 | 4030 | 4030 | 4030 | 4029 | 4029 | 4034 | 4035 |
| Slovenia | 500 | 500 | 500 | 500 | 500 | 610 | 610 | 610 | 610 | 610 |
| Slovakia | 1586 | 1586 | 1586 | 1587 | 1587 | 1588 | 1587 | 1587 | 1585 | 1585 |
| Finland | 3172 | 3172 | 3256 | 3262 | 3270 | 3331 | 3330 | 3331 | 3349 | 3359 |
| Sweden | 8194 | 8214 | 8232 | 8235 | 8184 | 8189 | 8217 | 8185 | 8184 | 8186 |

Source: Eurostat, (2023a).

**Table 2. Greenhouse gases from land transport and transport via pipelines, tonne**

| **TIME** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Belgium | 1880439 | 1834173 | 1750518 | 1751846 | 1701727 | 1754342 | 1876459 | 2008703 | 1739308 | 1957128 |
| Bulgaria | 5763323 | 5104633 | 5734809 | 6356719 | 6882955 | 7070646 | 6795813 | 6852090 | 6610878 | 7080793 |
| Czechia | 6232944 | 6005703 | 6287916 | 6635017 | 7353413 | 7763259 | 7460101 | 7659382 | 7190234 | 8098257 |
| Denmark | 3876941 | 3384525 | 2918216 | 3292504 | 3492884 | 2946113 | 3551202 | 3686458 | 3561188 | 3174225 |
| Germany | 20012494 | 21106313 | 18834670 | 18572176 | 18137088 | 17781302 | 17298328 | 16928542 | 14616476 | 14620010 |
| Estonia | 790873 | 1056811 | 875355 | 764427 | 791955 | 889371 | 908882 | 884355 | 843328 | 925535 |
| Ireland | 3018713 | 2501222 | 2397746 | 2330277 | 2632301 | 2960697 | 3099636 | 2644644 | 2597638 | 2751337 |
| Greece | 3143800 | 3176100 | 3429756 | 3130787 | 3140372 | 3812364 | 4068643 | 4238054 | 3752871 | 3851014 |
| Spain | 22073990 | 22115918 | 22159116 | 23519643 | 24095535 | 25040761 | 24634245 | 24934264 | 22958535 | 26146615 |
| France | 20077488 | 18913937 | 17791508 | 17180546 | 17157107 | 16829881 | 15441212 | 16095828 | 15146572 | 16836074 |
| Croatia | 689307 | 929100 | 858788 | 869450 | 881467 | 1086478 | 1021417 | 1105878 | 934992 | 929926 |
| Italy | 16927787 | 16954664 | 17541535 | 16345243 | 16291174 | 15956220 | 16991027 | 16794781 | 15062759 | 17317789 |
| Cyprus | 160147 | 128816 | 126220 | 138905 | 148934 | 163000 | 164558 | 176220 | 156927 | 157662 |
| Latvia | 1836662 | 1772579 | 1924090 | 2148089 | 2017229 | 2087631 | 2086874 | 2198027 | 1740816 | 1797888 |
| Lithuania | 2691607 | 2789452 | 2946134 | 3395708 | 3657428 | 4585176 | 5857903 | 6219656 | 8168482 | 8187658 |
| Luxembourg | 268035 | 241000 | 209000 | 210219 | 193282 | 204760 | 185605 | 170311 | 171026 | 174170 |
| Hungary | 3464763 | 3450827 | 3714076 | 3714588 | 3367129 | 3402693 | 3302844 | 3131445 | 2565300 | 2750760 |
| Malta | 34444 | 45380 | 49871 | 38877 | 55898 | 56583 | 62909 | 65644 | 54759 | 54966 |
| Netherlands | 6007180 | 5839631 | 5705387 | 5722135 | 5776024 | 5773280 | 5991772 | 5887826 | 5784661 | 5786866 |
| Austria | 3445055 | 3633606 | 3468317 | 3075020 | 2758765 | 3229495 | 3589678 | 3646667 | 3877399 | 4019458 |
| Poland | 8567642 | 8365630 | 8556609 | 8985438 | 10503282 | 17531837 | 18014623 | 7711995 | 7590066 | 8123050 |
| Portugal | 4119632 | 4025406 | 4160335 | 4103880 | 4095667 | 4007818 | 3884975 | 3831240 | 3262611 | 3507046 |
| Romania | 4328758 | 4572861 | 4377449 | 4717626 | 5280591 | 5610208 | 5686712 | 5935289 | 5484805 | 5774825 |
| Slovenia | 541964 | 547509 | 561110 | 576475 | 574860 | 624461 | 711104 | 759573 | 733113 | 777846 |
| Slovakia | 2950991 | 1885357 | 1302359 | 1958906 | 2121593 | 2073979 | 2165557 | 2203221 | 1384841 | 1567267 |
| Finland | 4785924 | 4145504 | 3440000 | 3576671 | 4239489 | 4132805 | 4111463 | 3917017 | 3744516 | 3627357 |
| Sweden | 2981477 | 2728112 | 2627611 | 2414329 | 2159053 | 2029772 | 1909004 | 1834105 | 1662037 | 1665700 |

Source: Author's representation of data provided by Eurostat (2023c).

1. Average 2013-2016 [↑](#footnote-ref-1)
2. Average 2014-2016 [↑](#footnote-ref-2)