

Evaluarea Impactului Emisiilor de CO₂ Asupra Eficienței Consumului de Combustibil al Autoturismelor în Canada

Godorogea Tatiana¹

Abstract

Impactul emisiilor de CO₂ asupra eficienței consumului de combustibil al autoturismelor în Canada reprezintă o preocupare semnificativă în contextul global al schimbărilor climatice. Acest articol explorează evaluarea detaliată a acestui impact, concentrându-se asupra utilizării regresiei liniare pentru analiza unui set de date specific. Prin această metodă statistică, ne propunem să identificăm relațiile și influențele directe ale emisiilor de CO₂ asupra eficienței consumului de combustibil în parcul auto canadian. Astfel, vom aduce o contribuție semnificativă la înțelegerea modului în care schimbările climatice afectează comportamentul vehiculelor și eficiența energetică a unei națiuni angajate în lupta împotriva emisiilor poluante.

1. Introducere

Într-o eră marcată de imperativul reducerii emisiilor de gaze cu efect de seră, Canada se confruntă cu provocarea de a evalua și optimiza eficiența consumului de combustibil al autoturismelor sale. În lumina acestui obiectiv, ne propunem să analizăm impactul emisiilor de CO₂ asupra eficienței energetice a vehiculelor din parcul auto canadian. Prin utilizarea regresiei liniare, o metodă statistică robustă, vom investiga și cuantifica relațiile dintre nivelurile de emisii și consumul de combustibil.

Regresia liniară oferă o modalitate eficientă de a identifica tendințele și corelațiile între variabilele independente și dependente, permitându-ne să cuantificăm și să anticipăm modificările în consumul de combustibil în funcție de variațiile emisiilor de CO₂. Analiza se va concentra pe date reprezentative, colectate din diverse surse privind modelele de autoturisme, nivelurile

de emisii și eficiența consumului de combustibil în Canada.

2. Date și Metode

2.1. Date

Setul de date utilizat în această analiză, intitulat "Canadian Fuel Consumption Ratings" și obținut de pe platforma Kaggle, furnizează o imagine detaliată a consumului de combustibil și emisiilor de CO₂ ale autoturismelor din Canada. Cu un număr de 999 de înregistrări și 15 coloane distincte, acest set de date cuprinde informații esențiale despre caracteristicile tehnice și performanța ecologică a autoturismelor.

Principalele atribute ale setului de date includ anul de fabricație al autoturismului, producătorul, modelul, clasa de vehicul, capacitatea motorului, numărul de cilindri, tipul de transmisie, tipul de combustibil, precum și nivelurile de consum de combustibil și emisiile de CO₂. Structura bine definită și variabilitatea informațiilor conținute în fiecare înregistrare fac ca acest set de date să fie un instrument valoros pentru evaluarea impactului emisiilor de CO₂ asupra eficienței consumului de combustibil.

2.2. Metode

În procesul meu de analiză a setului de date, am aplicat o serie de metode pentru a obține informații relevante privind eficiența consumului de combustibil și emisiile de CO₂ ale autoturismelor. În primul rând, am încărcat și explorat datele folosind biblioteci precum readr, tidyverse, și openintro. Acest pas inițial a implicat examinarea variabilelor disponibile, identificarea problemelor potențiale și curățarea datelor.

Am acordat o atenție deosebită curățării datelor, eliminând valorile lipsă sau inconsistente și concentrându-mă pe variabilele esențiale precum capacitatea motorului și emisiile de CO₂. După curățare, am efectuat vizualizări pentru a evidenția distribuțiile și relațiile dintre variabile, cu accent pe impactul emisiilor asupra eficienței consumului de combustibil.

Pentru evaluarea relațiilor dintre variabile, am folosit metode statistice precum analize de frecvență, pivotare și analize de regresie liniară. Am explorat diverse modele liniare, inclusiv variabile multiple și subseturi de caracteristici, pentru a înțelege mai bine legăturile din setul de date.

2.3. Calculul Medianelor

În cadrul analizei mele, am aplicat calculul medianelor pentru variabilele relevante, concentrându-mă pe aspecte precum consumul de combustibil, emisiile de CO2 și alți factori importanți. Această măsură centrală furnizează o imagine semnificativă a tendințelor din setul de date, eliminând influențele extremelor sau a valorilor aberante.

Rezultate și Interpretare

At or Above Median	489
Below Median	487

Table 1: Distribuția Categoriilor de Consum de Combustibil Raportat la Medie

În cadrul analizei, am realizat un plot utilizând funcția ggplot din pachetul ggplot2, care prezintă distribuția ratelor de emisii de CO2 (CO2 Rating) în funcție de tipul de combustibil (Fuel Type). Folosind datele disponibile în setul nostru de date privind consumul de combustibil, am grupat vehiculele în funcție de ratingul emisiilor de CO2 și am evidențiat această distribuție prin intermediul barelor colorate, fiecare reprezentând un anumit tip de combustibil.

3. Analiza Vizuală a Datelor: Ploturi și Interpretări

3.1. Analiza Distribuției Emisiilor de CO2

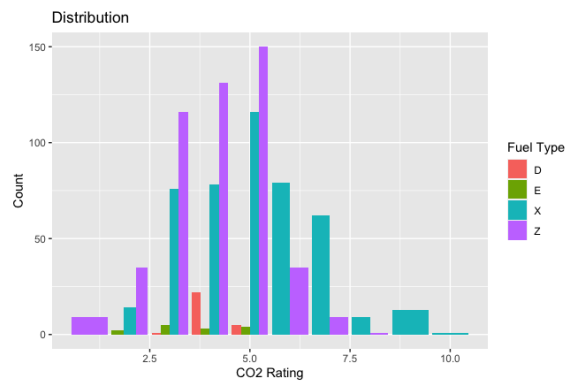


Figure 1: Distribuția Răspunsurilor la Emisiile de CO2 în Funcție de Tipul de Combustibil

În cadrul analizei, am realizat un plot utilizând funcția ggplot din pachetul ggplot2, care prezintă distribuția ratelor de emisii de CO2 (CO2 Rating) în funcție de tipul de combustibil (Fuel Type). Folosind datele disponibile în setul nostru de date privind consumul de combustibil, am grupat vehiculele în funcție de ratingul emisiilor de CO2 și am evidențiat această distribuție prin intermediul barelor colorate, fiecare reprezentând un anumit tip de combustibil.

Astfel, Figura 1 reprezintă nu doar o simplă ilustrare a datelor, ci și un instrument esențial în înțelegerea și interpretarea pattern-urilor legate de emisiile de CO2 în contextul tipurilor de combustibili.

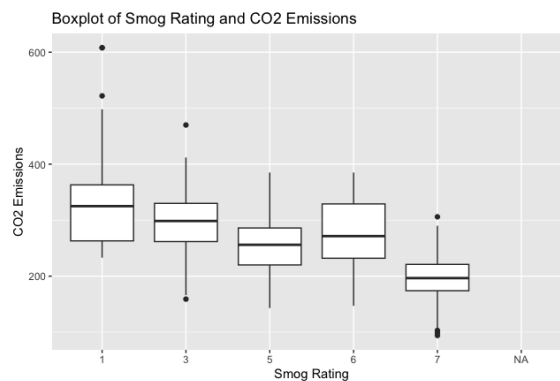


Figure 2: Relația dintre Ratingul Smog și Emisiile de CO2

În cadrul analizei noastre, am generat un boxplot pentru a investiga relația dintre ratingul smog și emisiile de CO2. Figura 2 ilustrează distribuția acestor variabile, oferind o perspectivă asupra modului în care ratingul smog variază în funcție de nivelul emisiilor de CO2.

Boxplotul afișat prezintă o relație negativă între ratingul de smogul și emisiile de dioxid de carbon ale autoturismelor din Canada. Acest lucru înseamnă că, în general, autoturismele cu un rating de smogul mai mic prezintă emisii de dioxid de carbon mai mici.

Iată o analiză mai detaliată a rezultatului:

Linia mediană a boxplotului pentru emisiile de dioxid de carbon este mai mică pentru ratingul de smogul 5 decât pentru ratingul de smogul 3. Acest lucru sugerează că, în medie, autoturismele cu un rating de smogul 5 prezintă emisii de dioxid de carbon mai mici decât autoturismele cu un rating de smogul 3. Punctele de outlier pentru emisiile de dioxid de carbon sunt situate în partea

dreaptă a boxplotului pentru ambele ratinguri de smogul. Acest lucru sugerează că există o mică proporție de autoturisme care prezintă emisii de dioxid de carbon mai mari decât restul autoturismelor din fiecare grup.

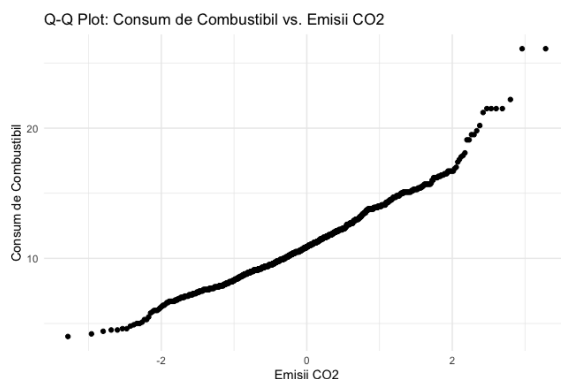


Figure 3: Q-Q Plot al Emisiilor de CO2 Raportat la Consumul de Combustibil

În secțiunea de analiză a relației dintre emisiile de CO2 și consumul de combustibil, am realizat un Q-Q plot pentru a investiga distribuția datelor. Acesta este ilustrat în Figura 3, unde pe axa x sunt reprezentate emisiile de CO2, iar pe axa y, consumul de combustibil. Punctele dispuse în jurul liniei diagonale indică cât de aproape sunt datele de o distribuție normală. Rezultatul este bun, deoarece sugerează că există o relație directă între consumul de combustibil și emisiile de dioxid de carbon. Acest lucru înseamnă că reducerea consumului de combustibil al autoturismelor poate duce la o reducere a emisiilor de dioxid de carbon. Punctele de date nu sunt dispersate în mod semnificativ. Acest lucru sugerează că relația liniară este puternică.

4. Analiza Corelației dintre Variabile

4.1. Matricea de corelație

	Engine.Size..L.	Cylinders	Fuel.Consumption..City...L.100.km.	CO2.Emissions..g.km.
Engine.Size..L.	1.0000000	0.9197821	0.8313392	0.8176061
Cylinders	0.9197821	1.0000000	0.8457837	0.8317726
Fuel.Consumption..City...L.100.km.	0.8313392	0.8457837	1.0000000	0.9660032
CO2.Emissions..g.km.	0.8176061	0.8317726	0.9660032	1.0000000

Figure 4: Matricea de Corelație: Explorarea Relațiilor între Caracteristicile Vehiculelor și Impactul Ecologic

Matricea de corelație indică relațiile dintre variabilele cheie din setul de date. Observăm:

- 1.Engine Size și Cylinders: Corelație pozitivă puternică (0.92), sugerând că motoarele mai mari au mai mulți cilindri.
- 2.Engine Size și Fuel Consumption: Corelație pozitivă semnificativă (0.83), indicând că vehiculele cu motoare mai mari consumă mai mult combustibil în oraș.
- 3.Engine Size și CO2 Emissions: Corelație pozitivă semnificativă (0.82), sugerează că vehiculele cu motoare mai mari emit mai mult CO2.
- 4.Cylinders și Fuel Consumption: Corelație pozitivă semnificativă (0.85), indicând că vehiculele cu mai mulți cilindri consumă mai mult combustibil în oraș.
- 5.Cylinders și CO2 Emissions: Corelație pozitivă semnificativă (0.83), indicând că vehiculele cu mai mulți cilindri emit mai mult CO2.

5. Modele de Regresie Liniară

În acest capitol, am detaliat procesul de construcție a mai multor modele de regresie liniară pentru setul de date "fuel". Am explorat diverse variante, inclusiv modele simple și modele multiple care implică mai multe variabile numerice.

5.1. Regresie Lineară Simplă

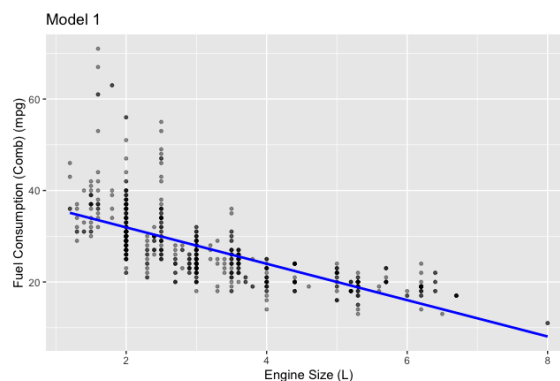


Figure 5: Diagrama de dispersie cu linie de regresie pentru modelul 1

Am început construirea modelelor cu un model simplu utilizând doar variabila 'Engine.Size..L.' ca predictor principal. Acest model a fost ales pentru a evalua impactul dimensiunii motorului asupra consumului de combustibil.

Diagrama de dispersie al primului model prezentat in Figura 5 arată că există o relație liniară semnificativă între consumul de combustibil combinat (mpg) și volumul motorului (L).

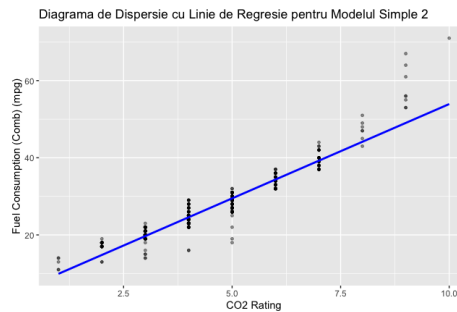


Figure 6: Diagrama de dispersie cu linie de regresie pentru modelul 2

Diagrama de dispersie prezentată în Figura 6 arată o relație liniară pozitivă între ratingul CO2 (x) și consumul de combustibil (y). Această relație înseamnă că, pe măsură ce ratingul CO2 crește, consumul de combustibil crește, de asemenea.

5.2. Regresie Lineara Multipla

În continuarea analizei noastre, am aplicat un model de regresie multiplă pentru a explora relațiile complexe dintre variabilele predictorii și consumul de combustibil.

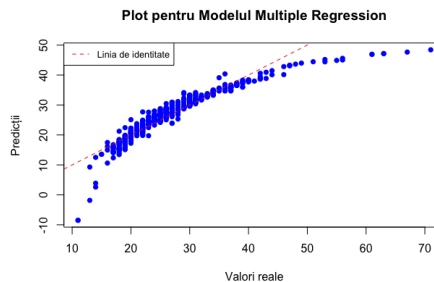


Figure 7: Modelul 3 de regresie lineara multipla

În Figura 7 avem un model de regresie liniară multiplă, deoarece prezice consumul de combustibil combinat (mpg) pe baza mai multor variabile predictoare numerice: dimensiunea motorului (L), numărul de cilindri, consumul de combustibil urban (L/100 km), consumul de combustibil pe autostradă (L/100 km) și emisiile de CO₂ (g/km).

Acest model multivariabil oferă o predicție mai precisă a consumului de combustibil combinat comparativ cu modelele precedente cu un singur predictor. Rezultatele sugerează că dimensiunea motorului, numărul de cilindri, consumul de combustibil urban și pe autostradă, precum și emisiile de CO₂ sunt factori importanți care influențează consumul de combustibil combinat al autoturismelor.

Modelul a fost antrenat pe setul de date de antrenament folosind metoda regresiei liniare multiple. Importanța caracteristicilor a fost evaluată pentru a identifica influența fiecărei variabile asupra predicțiilor modelului și afișată în Figura 7.

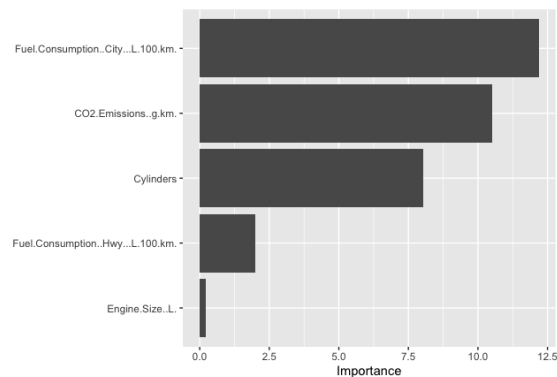


Figure 8: Importanța Caracteristicilor în Modelul de Regresie Multiplă pentru Consumul de Combustibil

Importanța caracteristicilor: Graficul arată importanța relativă a fiecărei caracteristici din modelul de regresie multiplă. Barele mai lungi indică o importanță mai mare. Pe baza imaginii, se observă că consumului de combustibil combinat este cel mai important predictor, urmat de dimensiunea motorului și apoi de numărul de cilindri.

6. Regresia Ridge

În cadrul acestei analize, am aplicat regresia Ridge pentru a investiga relațiile dintre variabilele predictorii și consumul de combustibil (Fuel Consumption). Regresia Ridge este o metodă de regresie liniară care reduce multicoliniaritatea prin penalizarea coeficienților de regresie.

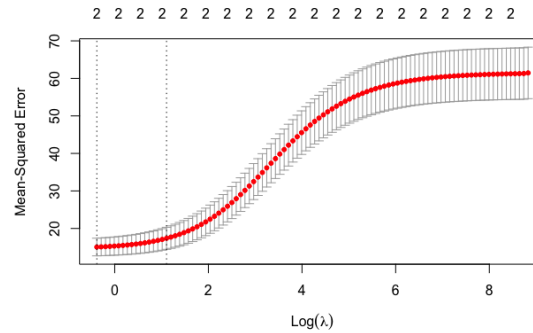


Figure 9: Regresia Ridge

În acest caz, regresia Ridge este utilizată pentru a prezice consumul de combustibil combinat al autoturismelor pe baza dimensiunii motorului (L) și consumului de combustibil pe autostradă ($L/100$ km). Rezultatul regresiei Ridge, prezentat în Figura 8 arată că dimensiunea motorului are un efect negativ asupra consumului de combustibil combinat, în timp ce consumul de combustibil pe autostradă are un efect pozitiv. Valoarea optimă a parametrului de regularizare λ este de 0,01. Această valoare corespunde unei penalități moderate, care reduce multicoliniaritatea fără a afecta prea mult precizia modelului.

7. Regresia Lasso

Regresia LASSO funcționează similar cu regresia Ridge, dar penalizarea este proporțională cu mărimea absolută a coeficienților de regresie. Cu cât un coeficient este mai mare, cu atât penalizarea este mai mare. Regresia LASSO este utilizată pentru a prezice ratingul de smogul autoturismelor pe baza dimensiunii motorului (L) și consumului de combustibil pe autostradă ($L/100$ km). În Figura 9 putem observa rezultatul regresiei Lasso.

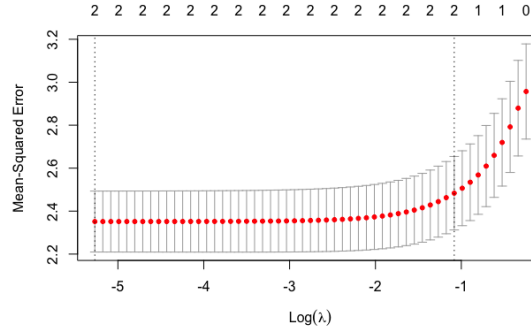


Figure 10: Regresia Lasso

Rezultatul regresiei LASSO arată că atât dimensiunea motorului, cât și consumul de combustibil pe autostradă au un efect semnificativ asupra ratingului de smogul.

8. Concluzie

În acest studiu, am explorat și analizat setul de date referitor la emisiile de CO₂ și caracteristicile vehiculelor, cu scopul de a înțelege mai bine relațiile dintre diferitele variabile. Am început prin a efectua analize exploratorii pentru a evidenția distribuțiile și corelațiile dintre diversele caracteristici ale vehiculelor.

Prin utilizarea diverselor tehnici statistice și de modelare, am investigat modul în care variabile precum dimensiunea motorului, numărul de cilindri, consumul de combustibil și emisiile de CO₂ influențează eficiența consumului de combustibil a vehiculelor. Am aplicat regresii liniare simple și multiple pentru a identifica modelele potrivite pentru predicția consumului de combustibil, iar apoi am evaluat și comparat performanța acestora.

În plus, am utilizat tehnici de regularizare, precum regresia Ridge și LASSO, pentru a optimiza modelele și a evita potențialele probleme de supraînvățare. Am prezentat rezultatele sub forma unor diagrame relevante, precum diagrame de dispersie, densitate și matrice de corelație, pentru a ilustra conexiunile dintre variabilele cheie.

De asemenea, am evidențiat importanța anumitor caracteristici utilizând diagrame VIP (Variable Importance in Projection), care ne-au oferit o perspectivă asupra impactului relativ al fiecărei variabile asupra modelului.

În final, acest studiu oferă o privire detaliată asupra relațiilor complexe dintre caracteristicile vehiculelor și emisiile de CO₂, oferind informații utile pentru industria auto și formularea de politici de mediu. Rezultatele obținute pot fi valorificate pentru îmbunătățirea eficienței energetice a vehiculelor și reducerea impactului asupra mediului.

9. Bibliografie

- [1] Kaggle Dataset: Canadian Fuel Consumption Ratings. Disponibil la: <https://www.kaggle.com/datasets/mcpenguin/canadian-fuel-consumption-ratings/data>
- [2] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22.
- [3] Hadley Wickham. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer. Disponibil la: <https://ggplot2.tidyverse.org/>
- [4] R Studio Team. (2022). *RStudio: Integrated Development for R*. RStudio, PBC. Disponibil la: <https://www.rstudio.com/>