

AGH

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE
Wydział Zarządzania

***Badanie czynników wpływających
na liczbę osób poszkodowanych
w wypadkach przy pracy w podregionach
w Polsce w 2021r.***

Autor:	<i>Arkadiusz Florek</i>
Kierunek studiów:	<i>Informatyka i Ekonometria</i>
Przedmiot:	Ekonometria
Prowadzący:	dr Paweł Zając

Kraków, 2023r.

Spis Treści

CEL PROJEKTU	2
HIPOTEZY BADAWCZE	3
OPIS DANYCH.....	4
STATYSTYKI OPISOWE.....	5
WSTĘPNA ANALIZA MODELU	8
METODA KROKOWA-WSTECZNA	10
METODA HELLWIGA.....	11
WYBÓR OSTATECZNEJ POSTACI MODELU	12
OPIS I TESTOWANIE WŁASNOŚCI MODELU	13
WSPÓŁCZYNNIK DETERMINACJI	13
EFEKT KATALIZY.....	13
NORMALNOŚĆ ROZKŁADU SKŁADNIKA LOSOWEGO.....	14
ISTOTNOŚĆ ZMIENNYCH	15
OBSERWACJE ODSTAJĄCE	15
TEST SERII	16
TEST RESET.....	17
TESTOWANIE HETEROSKEDASTYCZNOŚCI.....	17
TEST CHOWA.....	18
WSPÓŁLINIOWOŚĆ	18
KOINCYDENCJA	19
INTERPRETACJA PARAMETRÓW MODELU	19
PREDYKCJA WRAZ Z 95% PRZEDZIAŁEM UFNOŚCI	19
PODSUMOWANIE	20
LITERATURA.....	21
ŹRÓDŁA INTERNETOWE	21
SPIS RYSUNKÓW	21
SPIS TABEL	21

Cel projektu

Wypadki przy pracy, poza aspektem społecznym (zwłaszcza humanitarnym), mają znaczny wymiar ekonomiczny, zarówno dla przedsiębiorstw, osób poszkodowanych i ich rodzin, jak i całego społeczeństwa. W odniesieniu do przedsiębiorstw, wypadki przy pracy mogą skutkować podwyższeniem składki na społeczne ubezpieczenie wypadkowe, a także zwiększać absencję chorobową, wymuszającą wprowadzanie zastępstw lub pracy w nadgodzinach, zmniejszając jej wydajność i jakość. Następstwami mogą być również zaburzenia toku produkcji, które podrażają jej koszt i negatywnie wpływają na wizerunek firmy.

Przez społeczne koszty wypadków przy pracy należy rozumieć łączne koszty wypadków ponoszone przez przedsiębiorstwa, poszkodowanych i ich rodziny oraz koszty przeniesione na społeczeństwo (np. koszty leczenia, świadczeń rentowych). Według najnowszych szacunków Międzynarodowej Organizacji Pracy społeczne koszty wypadków przy pracy wynoszą w krajach uprzemysłowionych około 4% PKB¹.



Rysunek 1 Krzywa przedstawiająca liczbę osób poszkodowanych

Jak możemy zauważyć na wykresie [rysunek 1], stworzonym na podstawie danych z GUS, od 2013 średnia liczba ofiar w wypadkach przy pracy utrzymuje się cały czas na podobnym poziomie (ponad 500 osób). Jest to zaskakująca obserwacja biorąc pod uwagę postępujący postęp technologiczny oraz zmieniające się zasady dotyczące bezpieczeństwa i higieny pracy, które mają na celu zmniejszać liczbę wypadków w pracy.

¹ Jan Rzepecki, *Społeczne koszty wypadków przy pracy*, „Bezpieczeństwo Pracy - Nauka i Praktyka” 5/2012, str.20

Biorąc pod uwagę wcześniejsze konkluzje dotyczące jak ważną kwestią(ekonomiczną, społeczną oraz humanitarną) jest liczba ofiar wypadków przy pracy, chciałbym w moim projekcie zbadać wpływ wybranych czynników na liczbę poszkodowanych w wypadkach przy pracy w podregionach w Polsce w 2021r. W kolejnych częściach przedstawię bliżej zmienne objaśniające, które wziąłem pod uwagę, wybiorę właściwy model oraz zweryfikuję postawione hipotezy.

Hipotezy badawcze

1. Wyższy stopień rejestrowanego bezrobocia skutkuje niższą liczbą poszkodowanych osób w wypadkach przy pracy.
2. Zwiększenie się liczby kotłowni przyczynia się istotnie do wyższej liczby ludzi ulegających wypadkom w pracy.
3. Czym większa liczba zawartych małżeństw tym większa liczba ofiar wypadków przy pracy.
4. Im większe nakłady inwestycyjne w przedsiębiorstwach przypadające na 1 mieszkańca tym większa liczba osób ulegających wypadkom w pracy.

Opis danych

Dane, które wykorzystuje w projekcie, pochodzą z zasobów Głównego Urzędu Statystycznego. Poniższy link przekieruje do strony internetowej GUS, z której zostały pobrane: <https://bdl.stat.gov.pl/bdl/dane/podgrup/temat>.

Dane przedstawiają jedną zmienną objaśnianą Y oraz $X_1 \dots X_4$ zmiennych objaśniających dla podregionów (wartości odstające zostały usunięte) - 64 obserwacji - w 2021r. Model ten można zapisać w następującej postaci:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \varepsilon$$

Zmienna objaśniania:

- Y - Liczba poszkodowanych osób w wypadkach przy pracy [osoba]

Potencjalne zmienne objaśniające:

- X_1 - Stopa bezrobocia, obliczono jako stosunek liczby bezrobotnych do liczby cywilnej ludności aktywnej zawodowo (bez osób odbywających czynną służbę wojskową itd.) [%]
- X_2 - Liczba kotłowni znajdujących się w gospodarce mieszkaniowej i komunalnej [szt.]
- X_3 - Liczba zawartych małżeństw
- X_4 - Nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca [zł]

Statystyki opisowe

Tabela 1 Statystyki opisowe

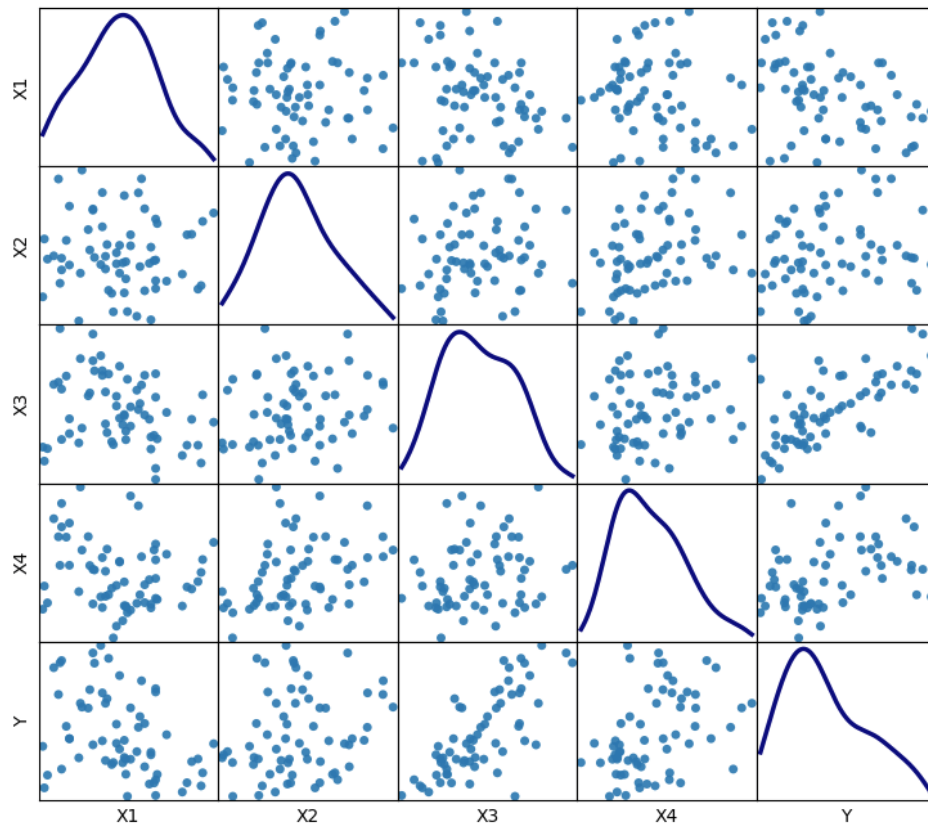
Zmienna Statystyka	Y	X₁	X₂	X₃	X₄
Średnia	915,95	6,5159	473,57	2198,2	4137,5
Mediana	824	6,5	451	2106	3846
Minimalna	298	1,6	187	763	1573
Maksymalna	1806	12,5	831	3906	7665
Odch.stand	410,84	2,7062	154,6	698,74	1349,2
Wsp.zmienności	0,44854	0,41533	0,32646	0,31787	0,3261
Skośność	0,50319	0,15355	0,31963	0,261	0,64832
Kurtoza	-0,82455	-0,55719	-0,41083	-0,53654	-0,11129

Źródło: Własne opracowanie na podstawie wyników programu Gretl

W tabeli 1 zostały przedstawione podstawowe statystyki zmiennych, które zostały wykorzystane w badaniu.

- Zmienna Y – Wartość minimalna wynosi 298, a wartość maksymalna 1806. W zmiennej objaśnianej Y średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 0.5. Typowa wielkość Y różni się od wartości przeciętnej średnio 410.84.
- Zmienna X₁ – Minimalna wartość stopy bezrobocia wynosi 1.6, a wartość maksymalna 6.5. Współczynnik skośności wynosi 0.15. Typowa wielkość X₆ różni się od wartości przeciętnej średnio o 2.71.
- Zmienna X₂ – Minimalna wartość wynosi 187, a wartość maksymalna 831. W zmiennej objaśniającej X₂ średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 0.31. Typowa wielkość X₂ różni się od wartości przeciętnej średnio o 154.6.
- Zmienna X₃ – Minimalna liczba małżeństw wynosi 763, a wartość maksymalna 3906. Współczynnik skośności wynosi 0.26. Typowa wielkość X₂ różni się od wartości przeciętnej średnio o 698.74.

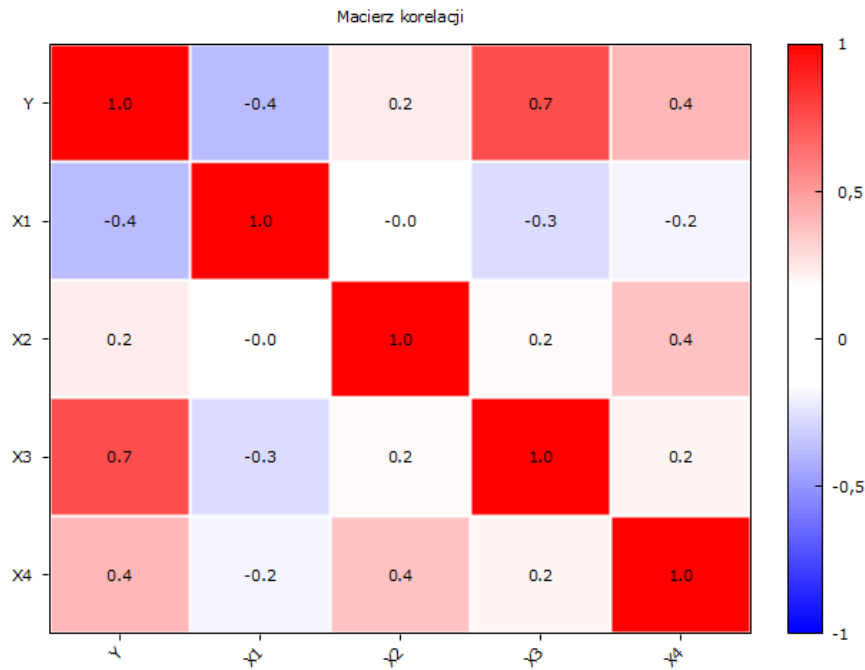
- Zmienna X_4 – Minimalna ilość nakładów inwestycyjnych w przedsiębiorstwach jest równa 1573, a wartość maksymalna 7665. W zmiennej objaśniającej X_3 średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 0.65. Typowa wielkość X_3 różni się od wartości przeciętnej średnio o 1349.2.



Rysunek 2 Wykres rozrzutu

Na wykresie można zauważyć ciekawe zależności:

- Y i X_1 odznaczają się ujemną korelacją, aczkolwiek jest dużo obserwacji odstających, które mogą zaburzać tę zależność
- Y i X_3 wykazują dość mocną dodatnią korelację, punkty skupiają się wzdłuż szybko rosnącej linii prostej
- Y i X_4 odznaczają się dodatnią korelacją
- Wykresy gęstości Y , X_3 , X_4 cechują się podobnym rozkładem (prawostronna asymetria), co może sugerować korelację między nimi



Rysunek 3 Macierz korelacji

Wzajemne korelacje zmiennych objaśniających oraz korelacje pomiędzy zmiennymi objaśniającymi X, a zmienną objaśnianą Y zostały przedstawione na rysunku 3. Wiemy, że z założeń budowania modelu, zmienne powinny być silnie skorelowane ze zmienną objaśnianą, natomiast słabo skorelowane między sobą. Wiele z tych zależności zostało wykryte i opisane przy okazji analizowania wykresu rozrzutu [rysunek 2], dlatego omówię te zależności, które były nieoczywiste i nie zostały rozważone, jak na przykład:

- X_2 jest słabo skorelowane ze zmienną objaśnianą Y i najprawdopodobniej zmienna ta zostanie wyeliminowana na następnych etapach estymacji
- Najwyższą korelacją na poziomie 0.7 ze zmienną objaśnianą Y odznacza się zmienna X_3 , czyli liczba zawartych małżeństw
- Potencjalne zmienne objaśniające nie wykazują korelacji między sobą, co sugeruje, że nie wystąpi problem współliniowości predyktorów

Wstępna Analiza modelu

Analiza regresji jest jedną z najczęściej stosowanych technik do analizy danych wieloczynnikowych. Jej szeroka atrakcyjność i użyteczność wynika z conceptualnie logicznego procesu użycia równania do wyrażenia związku między zmienną będącą przedmiotem zainteresowania (odpowiedzią) a zestawem powiązanych zmiennych predykcyjnych. Analiza regresji jest również interesująca ze strony teoretycznej ze względu na elegancką matematykę i dobrze rozwiniętą teorię statystyczną. Skuteczne wykorzystanie regresji wymaga zarówno zastosowania teorii, jak i praktycznych rozwiązań problemów, które zwykle pojawiają się, gdy technika ta jest stosowana z danymi ze świata rzeczywistego².

W celu przeanalizowania modelu i oszacowania parametrów wykorzystam metodę najmniejszych kwadratów (KMNK). Ogólną postać modelu można zapisać następująco:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \varepsilon$$

Ideą KMNK jest znalezienie takich wartości nieznanego wektora parametrów α , który minimalizuje sumę kwadratów reszt, czyli różnic pomiędzy wartościami obserwowanymi, a teoretycznymi. Po oszacowaniu parametrów modelu ekonometrycznego:

$$\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2 + \dots + \hat{\alpha}_4 X_4$$

Po próbach zmiany zmiennej X_2 na logarytmy i kwadraty tych zmiennych model nie wykazywał poprawy w parametrach, dlatego usunąłem zmienną i stworzyłem model z trzema zmiennymi objaśniającymi.

² Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, *Introduction to Linear Regression Analysis 5th Edition*, New Jersey 2012, s.15

Tabela 2 Model nr 1

<i>Zmienna</i> <i>Statystyka</i>	Współczynnik	Błąd standardowy	t-studenta	Wartość p
const	-73,4994	180,406	-0,4074	0,6852
X₁	-23,8889	12,4426	-1,920	0,0597
X₃	0,384993	0,048188	7,989	5,70e-11
X₄	0,072221	0,024455	2,953	0,0045

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Z tabeli 2, można odczytać, że największą wartość p-value ma X_1 i wynosi ona 0.0597 (wartość większa niż poziom istotności o wartości 5%), co oznacza, że można wnioskować, iż zmienna X_1 jest statystycznie nieistotna i jest kandydatką do wyeliminowania w modelu, co potwierdza statystyka t-studenta wynosząca -1.92 (wartość krytyczna jest równa około 2), więc można uznać, że zmienna ma statystycznie nieistotny wpływ na zmienną objaśnianą Y , czyli jest nieistotna dla modelu.

Dla reszty zmiennych objaśniających w modelu, czyli X_3, X_4 wartość p-value jest mniejsza od wartości poziomu istotności, więc są to zmienne statystycznie istotne, co potwierdza statystyka t-studenta, która jest większa od wartości krytycznej, więc wnioskuję, że zmienne mają statystycznie istotny wpływ na zmienną objaśnianą Y .

Tabela 3 Statystyki modelu 1

Średn.aryt.zm.zależnej	915,9524	Odch.stand. zm.zależnej	410,8409
Suma kwadratów reszt	3742475	Błąd standardowy reszt	251,8567
Wsp.determ.R^2	0,642382	Skorygowany R^2	0,624198
F(3,59)	35,32678	Wartość p dla testu F	3,36e-13
Logarytm wiarygodności	-435,6450	Kryt.inform.Akaike'a	879,29
Kryt. Bayes. Schwarza	887,8625	Kryt. Hannana - Quinna	882,6616

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Muszę zredukować ilość zmiennych, przy użyciu metody Hellwiga oraz metody krokowej-wstecznej.

Metoda krokowa-wsteczna

Metoda wsteczna została wprowadzona na początku lat 60-tych (Marill & Green, 1963). Jest to jedno z głównych podejść regresji krokowej. W statystyce metoda wsteczna jest metodą dopasowywania modeli regresji, w której wybór zmiennych predykcyjnych regresji odbywa się za pomocą automatycznej procedury. Obejmuje ona zaczynanie od wszystkich zmiennych kandydujących, testując usunięcie każdej zmiennej przy użyciu wybranego kryterium dopasowania modelu, usunięcie zmiennej (jeśli w ogóle), której utrata powoduje najbardziej nieistotne statystycznie pogorszenie dopasowania modelu i powtórzenie tej procedury do momentu, gdy nie będzie można usunąć kolejnych zmiennych bez statystycznie znaczącej utraty dopasowania. Celem selekcji jest zredukowanie zestawu zmiennych predykcyjnych do tych, które są niezbędne. Zasadniczo selekcja pomaga określić poziom ważności każdej zmiennej predykcyjnej. Pomaga również w ocenie efektów po statystycznym wyeliminowaniu innych zmiennych predykcyjnych³.

Chcąc zweryfikować, czy zbiór zmiennych można odrzucić przy opisie zmiennej objaśnianej można posłużyć się testem na odrzucenie pominiętych zmiennych, wykorzystując statystykę F. Test ten weryfikuje zasadność usunięcia z modelu wybranej zmiennej lub zmiennych⁴.

H_0 : Parametry dla wskazanych zmiennych są równe 0

H_1 : Parametry dla wskazanych zmiennych nie są równe 0

W modelu w pierwszej kolejności wykonuje test pominiętych zmiennych dla zmiennej X_1 .

Dla X_1 wartość statystyki $F(1,59) = 3.686$ z wartością $p = 0.0597$. W takim razie nie ma podstaw do odrzucenia hipotezy zerowej, co oznacza usunięcie zmiennej X_1 z modelu.

³ Changxia Shao, Boehringer Ingelheim, *Backward Selection-a way to final model*, Philadelphia 2019, s.1

⁴ <https://mat.ug.edu.pl/~olanowel/pliki/e10.pdf>

Tabela 4 Model nr. 2

<i>Zmienna</i> <i>Statystyka</i>	Współczynnik	Błąd standardowy	t-studenta	Wartość p
const	-308,867	135,284	-2,283	0,026
X₃	0,407942	0,047716	8,549	5,70e-12
X₄	0,079294	0,024711	3,209	0,0021

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Tabela 5 Statystyki modelu 2

Średn.aryt.zm.zależnej	915,9524	Odch.stand. zm.zależnej	410,8409
Suma kwadratów reszt	3976291	Błąd standardowy reszt	257,4326
Wsp.determ.R²	0,620039	Skorygowany R²	0,607374
F(2,60)	48,95546	Wartość p dla testu F	2,47e-13
Logarytm wiarygodności	-437,5540	Kryt.inform.Akaike'a	881,108
Kryt. Bayes. Schwarza	887,5374	Kryt. Hannana - Quinna	883,6367

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Pominięcie zmiennej poprawiło jedno kryterium informacyjne modelu. Porównując tabelę 5 z tabelą 3 można zauważyć, że R^2 i skorygowane R^2 uległo tylko niewielkiemu pogorszeniu w stosunku do poprzedniego modelu.

Pozostałe zmienne objaśniające w modelu są istotne.

Metoda Hellwiga

Metoda Hellwiga została opracowana przez Zdzisława Hellwiga w 1969 roku. Metoda ta pozwala na dobór zmiennych objaśniających do badania ekonometrycznego. Zmienne wybrane do liniowego modelu ekonometrycznego powinny być silnie skorelowane ze zmienną objaśnianą i ewentualnie słabo skorelowane ze zmienną zależną. Idea stojąca za tą metodą opiera się na kryterium liczbowym, które pozwala wybrać najlepszą kombinację zmiennych spośród potencjalnych kombinacji zmiennych objaśniających branych pod uwagę⁵.

⁵ Gałecka Małgorzata, Smolny Katarzyna, *Evaluation of theater activity using Hellwig's method*, Białystok 2018, s.41

Przy użyciu kodu własnego autorstwa w programie Gretl otrzymałem największą integralną pojemność informacyjną równą 0.59528 dla zmiennych objaśniających: $\{X_3, X_4\}$.

Ostateczny model otrzymany przy pomocy dwóch metod jest taki sam.

Wybór ostatecznej postaci modelu

Po próbach transformacji modelu korzystając z logarytmów, odwrotności i kwadratów zmiennych nie udało się stworzyć lepszego modelu niż ten opisany i wybrany metodą krokową-wsteczną oraz metodą Hellwiga we wcześniejszych rozdziałach.

$$Y = \alpha_0 + \alpha_3 X_3 + \alpha_4 X_4 + \varepsilon$$

Opis i testowanie własności modelu

Tabela 6 Finalna forma modelu

<i>Zmienna</i> <i>Statystyka</i>	Współczynnik	Błąd standardowy	t-studenta	Wartość p
const	-308,867	135,284	-2,283	0,026
X₃	0,407942	4,77e-02	8,549	5,70e-12
X₄	0,079294	0,024711	3,209	0,0021

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Tabela 7 Statystyki finalnego modelu

Średn.aryt.zm.zależnej	915,9524	Odch.stand. zm.zależnej	410,8409
Suma kwadratów reszt	3976291	Błąd standardowy reszt	257,4326
Wsp.determ.R²	0,620039	Skorygowany R²	0,607374
F(2,60)	48,95546	Wartość p dla testu F	2,47e-13
Logarytm wiarygodności	-437,5540	Kryt.inform.Akaike'a	881,108
Kryt. Bayes. Schwarza	887,5374	Kryt. Hannana - Quinna	883,6367

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Współczynnik determinacji

Współczynnik determinacji wynosi 0.620039, co oznacza, że model w 62% wyjaśnia zmienność liczby osób poszkodowanych w wypadkach przy pracy. Biorąc pod uwagę ilość zmiennych i ilość obserwacji, model wyjaśnia zmienność liczbę ofiar w wypadkach w 60.74%. Różnica między tymi wskaźnikami nie jest duża, czyli jest wysokie zaufanie do interpretacji R².

Efekt katalizy

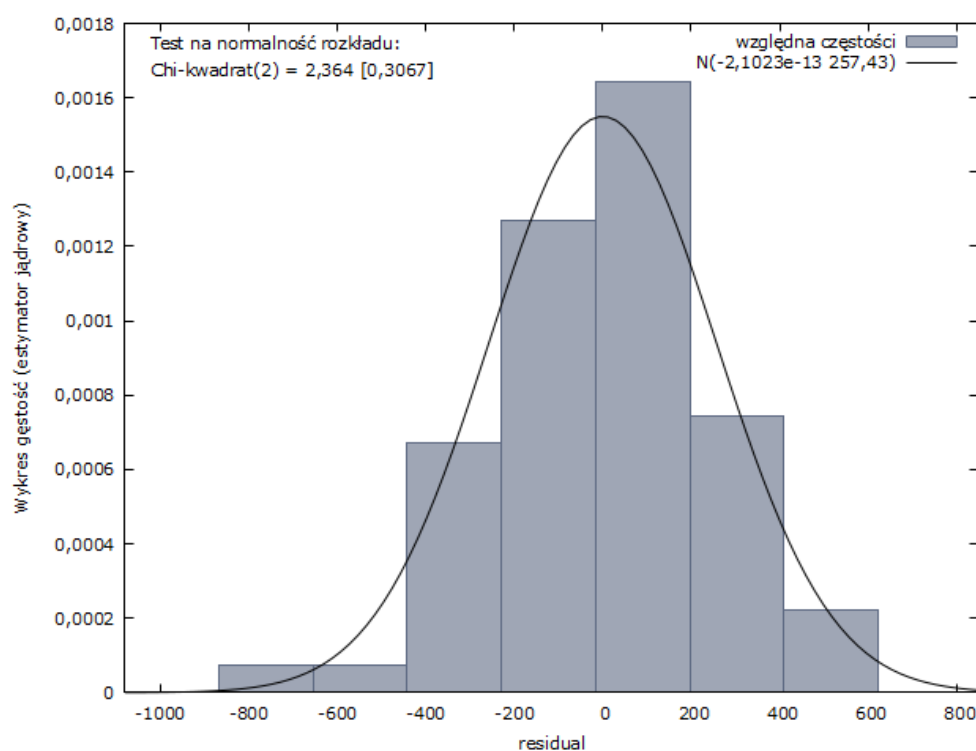
Przy wykorzystaniu kodu własnego autorstwa w programie Gretl, mogę wnioskować, że efekt katalizy nie występuje dla danych zmiennych objaśniających. Natężenie efektu katalizy dla tych zmiennych wynosi: 0.024756.

Normalność rozkładu składnika losowego

Jednym z warunków poprawności modelu ekonometrycznego jest rozkład normalny składnika losowego. Istnieje wiele testów statystycznych, które pozwalają zweryfikować ten warunek. Posłużę się testem Doornika-Hansena, który wykonam w Gretlu i sprawdzę czy składnik losowy pochodzi z rozkładu normalnego.

H_0 : Składnik losowy ma rozkład normalny

H_1 : Składnik losowy nie posiada rozkładu normalnego



Rysunek 4 Test normalności rozkładu reszt

Na podstawie rysunku 4 wnioskuję, że wartość p-value w teście wynosi 0.3067, co jest większą wartością od przyjętego poziomu istotności, który wynosi 5%. Z tego powodu nie ma podstaw do odrzucenia hipotezę zerowej - składnik losowy ma rozkład normalny.

Istotność zmiennych

$$H_0 : \alpha_j = 0$$

$$H_1 : \alpha_j \neq 0$$

Wszystkie zmienne objaśniające oraz wyraz wolny mają p-value mniejsze od poziomu istotności, co oznacza, że są statystycznie istotne oraz ich wartość t-studenta jest większa od wartości krytycznej, więc wnioskuję, że mają one statystycznie istotny wpływ na zmienną objaśnianą Y.

Obserwacje odstające

Tabela 8 Obserwacje odstające

<i>Statystyka</i> <i>Indeks</i>	Reszty	Leverage	Influence	DFFITS
10	-408,56	0,144*	-68,763	-0,716
14	-326,94	0,116*	-42,892	-0,493
27	147,71	0,119*	19,956	0,224
30	149,30	0,101*	16,805	0,204
41	-8,54	0,113*	-1,0869	-0,012

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Podregion 30(podregion miasto Wrocław) jest obserwacją nietypową, odstającą od reszty, ze względu, że w samym mieście Wrocław na rynku popularne są sektory takie jak przemysł, budownictwo, przetwórstwo przemysłowe⁶, które są sektorami szczególnie narażonymi na urazy oraz wypadki podczas pracy. Pominięcie obserwacji dźwigniowej dla podregionu Miasto Wrocław, mimo jego nietypowej natury, może prowadzić do zniekształcenia ogólnego obrazu dotyczącego wypadków przy pracy w regionie. Warto zachować tę obserwację, aby lepiej zrozumieć i monitorować specyficzne wyzwania związane z bezpieczeństwem pracy w tym obszarze.

⁶ <https://www.gowork.pl/poradnik/17/pracodawcy/najwieksi-pracodawcy-we-wroclawiu/>

Podregion 10(podregion rybnicki) jest obserwacją nietypową, ponieważ jest to region składający się głównie z populacji w wieku produkcyjnym, co przekłada się na to, że jest statystyczna większa szansa wypadku przy większej liczbie osób pracujących oraz jest to region skupiający się na pracy fizycznej, co przekłada się na większą liczbę wypadków.

Podregion 14(podregion Gorzowski) jest obserwacją odstającą od reszty, ze względu na to, że rejon skupiający się na rolnictwie, który nie jest narażony na wypadki przy pracy oraz w tym regionie znajduje się Gorzów Wielkopolski, który jest bardzo szybko rozwijającym się miastem zatrudniającym głównie w sektorach słabo narażonych na urazy oraz wypadki.

Test serii

W celu poprawnego wykonania testu liczby serii porządkuje dane rosnąco według wartości Y, a następnie wykonuje odpowiednie polecenia w Gretlu.

Liczba serii dla zmiennej 'uhat1' = 25.

Test niezależności oparty na liczbie dodatnich i ujemnych serii.

H_0 : Postać modelu jest dobrze dobrana

H_1 : Postać modelu nie jest dobrze dobrana⁷

Wartość p-value tego testu wynosi 0.0567798 przy dwustronnym obszarze krytycznym i jest to wartość większa od przyjętego poziomu istotności. Zatem nie mam podstaw do odrzucenia hipotezy zerowej, że postać modelu jest dobrze dobrana.

⁷ Materiał z ćwiczeń z ekonometrii

Test RESET

Przy użyciu testu Ramseya RESET sprawdzę poprawność analityczną modelu.

H0: Wybór postaci analitycznej modelu jest prawidłowy

H1: Wybór postaci analitycznej modelu nie jest prawidłowy⁸

1. Test RESET na specyfikację (kwadrat i sześćcian zmiennej)

Statystyka testu: $F = 1.211$,

z wartością $p = P(F(2,58) > 1.211) = 0.333$

2. Test RESET na specyfikację (tylko kwadrat zmiennej)

Statystyka testu: $F = 0.3501$,

z wartością $p = P(F(1,59) > 0.3501) = 0.556$

3. Test RESET na specyfikację (tylko sześćcian zmiennej)

Statystyka testu: $F = 0.1613$,

z wartością $p = P(F(1,59) > 0.1613) = 0.689$

Wszystkie trzy testy mają p-value większe od przyjętego poziomu istotności, czyli nie mam podstaw do odrzucenia hipotezy zerowej, która mówi, że postać analityczna modelu jest prawidłowa. Postać funkcyjna modelu jest poprawna.

Testowanie heteroskedastyczności

Heteroskedastyczność składnika losowego modelu jest odstępstwem od klasycznych założeń MNK. Oznacza to, że choć składniki losowe są wzajemnie nieskorelowane, to mają różne wariancje. W konsekwencji otrzymujemy estymator parametrów modelu, który pozostaje estymatorem nieobciążonym, liniowym, i zgodnym, ale nie jest estymatorem najefektywniejszym w klasie estymatorów liniowych i nieobciążonych. Heteroskedastyczność jest zjawiskiem, którego nie powinno być w modelu. W celu zbadania czy występuje w badanym modelu wykorzystam test White'a⁹.

⁸ Materiał z ćwiczeń z ekonometrii

⁹ Materiał z ćwiczeń z ekonometrii

H0: Heteroskedastyczność reszt nie występuje

H1: Heteroskedastyczność reszt występuje

Statystyka testu wynosi 10.259 , a wartość p-value 0.0682 , czyli nie ma podstaw do odrzucenia hipotezy zerowej, która mówi, że heteroskedastyczność reszt nie występuje w modelu.

Test Chowa

Test Chowa jest najczęściej stosowanym testem weryfikującym hipotezę o stabilności parametrów modelu. Pozwala stwierdzić, czy oszacowane parametry strukturalne otrzymane na podstawie próby są stabilne.

H0: Parametry modelu są stabilne

H1: Parametry modelu nie są stabilne¹⁰

Test Chowa na zmiany strukturalne przy podziale próby w obserwacji 32 $F(3, 57) = 1.284$ z wartością p równą 0.2886, czyli większa od ustalonego poziomu istotności, czyli nie ma podstaw do odrzucenia hipotezy zerowej - parametry modelu są stabilne.

Współliniowość

Współliniowość jest cechą zbioru danych statystycznych, wykorzystanych do szacowania parametrów modelu ekonometrycznego. Współliniowość w modelu jest zjawiskiem bardzo niepożądanym i może prowadzić do sfalszowania wyników estymacji.

Żeby zbadać współliniowość w modelu wykonam ocenę współliniowości VIF(Variance Infaltion Factors¹¹) w Gretlu.

Wskaźnik stanu cond dla każdej ze zmiennej oraz dla wyrazu wolnego w modelu jest mniejszy niż 10, stąd mogę wnioskować, że nie ma dowodów na nadmierną współliniowość w modelu¹².

¹⁰ Materiał z ćwiczeń z ekonometrii

¹¹ <https://gretl.sourceforge.net/gretl-help/cmdref.html#vif>

¹² <https://mat.ug.edu.pl/~olanowel/pliki/e10.pdf>

Koincydencja

Model jest koincydentny, jeśli przy każdej zmiennej objaśniającej znak parametru strukturalnego w modelu jest taki sam jak znak współczynnika korelacji ze zmienną objaśnianą.

Korelacja między zmiennymi objaśniającymi a zmienną objaśnianą jest dodatnia, tak jak parametry zmiennych objaśniających, co oznacza że model jest koincydentny.

Interpretacja parametrów modelu

Gdy liczba zawartych małżeństw wzrasta o 1 jednostkę, to liczba osób poszkodowanych w wypadkach przy pracy wzrasta średnio o 0.41 osób

Gdy nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca wzrastają 1 zł, to liczba osób poszkodowanych w wypadkach przy pracy wzrasta średnio o 0.08 osób

Dla przykładu - jeśli nakłady inwestycyjne na 1 mieszkańca wzrosną o 1000 zł, to średnia liczba osób poszkodowanych wzrośnie o 80 osób.

Predykcja wraz z 95% przedziałem ufności

Rzeczywista liczba osób poszkodowanych różni się od prognozy wyznaczonej z modelu o 25.78%. Prognoza nie jest akceptowalna¹³.

Wyznaczono prognozę liczby ofiar wypadków dla podregionu pierwszego. Prognozowa punktowa wyniosła 1244.67 osób. Z prawdopodobieństwem 95% prognozowana liczba osób będzie większa niż 710, a mniejsza od 1778. Rzeczywista wartość liczby osób poszkodowanych wynosi 813.

¹³ https://m6.pk.edu.pl/materialy/mp/MP_02_bledy_prognozy.pdf

Podsumowanie

Model spełnia wszystkie założenia metody estymacji, jednak wyznaczone prognozy są obarczone dużym błędem, czyli model nie nadaje się do prognozowania.

Zweryfikuje postawione na początku hipotezy:

1. Wyższa stopa rejestrowanego bezrobocia skutkuje niższą liczbą poszkodowanych osób w wypadkach przy pracy.

Hipoteza fałszywa. Zmienna mówiąca o stopie bezrobocia nie jest istotna w modelu.

2. Zwiększenie się liczby kotłowni przyczynia się istotnie do wyższej liczby ludzi ulegających wypadkom w pracy.

Hipoteza nieprawdziwa. Zmienna nie była istotna w modelu.

3. Czym większa liczba zawartych małżeństw tym większa liczba ludzi poszkodowanych w wypadkach przy pracy.

Hipoteza prawdziwa. Gdy liczba zawartych małżeństw wzrasta o 1 jednostkę, to liczba ofiar wypadków przy pracy wzrasta średnio o 0.41 osób.

4. Im większe nakłady inwestycyjne w przedsiębiorstwach przypadające na 1 mieszkańca tym większa liczba osób ulegających wypadkom w pracy.

Hipoteza prawdziwa. Gdy nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca wzrastają 1 zł, to liczba osób poszkodowanych w wypadkach przy pracy wzrasta średnio o 0.08 osób

Literatura

1. Rzepecki Jan, *Spoleczne koszty wypadków przy pracy*, „Bezpieczeństwo Pracy - Nauka i Praktyka” 5/2012, str.20
2. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, *Introduction to Linear Regression Analysis 5th Edition*, New Jersey 2012, s.15
3. Changxia Shao, Boehringer Ingelheim, *Backward Selection-a way to final model*, Philadelphia 2019, s.1
4. Gałęcka Małgorzata, Smolny Katarzyna, *Evaluation of theater activity using Hellwig's method*, Białystok 2018, s.41
5. Zając Paweł, Materiały z ćwiczeń z ekonometrii

Źródła internetowe

1. <https://mat.ug.edu.pl/~olanowel/pliki/e10.pdf>
2. <https://www.gowork.pl/poradnik/17/pracodawcy/najwieksi-pracodawcy-we-wroclawiu/>
3. <https://gretl.sourceforge.net/gretl-help/cmdref.html#vif>
4. https://m6.pk.edu.pl/materiały/mp/MP_02_bledy_prognozy.pdf

Spis rysunków

RYSUNEK 1 KRZYWA PRZEDSTAWIAJĄCA LICZBĘ OSÓB POSZKODOWANYCH	2
RYSUNEK 2 WYKRES ROZRZUTU	6
RYSUNEK 3 MACIERZ KORELACJI.....	7
RYSUNEK 4 TEST NORMALNOŚCI ROZKŁADU RESZT	14

Spis Tabel

TABELA 1 STATYSTYKI OPISOWE	5
TABELA 2 MODEL NR 1	9
TABELA 3 STATYSTYKI MODELU 1	9
TABELA 4 MODEL NR. 2	11
TABELA 5 STATYSTYKI MODELU 2	11
TABELA 6 FINALNA FORMA MODELU	13
TABELA 7 STATYSTYKI FINALNEGO MODELU	13
TABELA 8 OBSERWACJE ODSTAJĄCE	15