POLITECHNIKA RZESZOWSKA
im. IGNACEGO ŁUKASIEWICZA

**Projektowanie modeli łączenia źródeł danych**


Draguła Bartłomiej, Dereń Adrian

166643, 166724


Inżynieria i Analiza Danych

Grupa projektowa 1.

09.02.2023

## Spis treści

## Wstęp teoretyczny

Analiza statystyczna jest ważnym narzędziem w procesie podejmowania decyzji i formułowania wniosków na podstawie danych. W szczególności, analiza rozkładów normalnych jest szeroko stosowana w różnych dziedzinach, w tym w ekonomii, medycynie i inżynierii, w celu opisywania i porównywania danych. W projekcie będziemy analizować dane z Głównego Urzędu Statystycznego (GUS) dotyczące przeciętnych cen niektórych produktów z nabiału. Celem projektu jest zbadanie, czy ceny te można opisać za pomocą rozkładu normalnego oraz porównanie tych cen z cenami innych produktów. Aby wykonać tę analizę, będziemy wykorzystywać różne narzędzia statystyczne, w tym testy normalności, takie jak test Shapiro-Wilka, test Goldfelda-Quandta, test Jarqua-Bera oraz wizualizacje danych. Oczekujemy, że wyniki projektu pozwolą nam na uzyskanie lepszego zrozumienia przeciętnych cen produktów z nabiału i pomogą w formułowaniu wniosków dotyczących trendów i zmienności tych cen. W dalszej części tego wprowadzenia omówimy szczegółowo metodologię projektu i założenia dotyczące oczekiwanych wyników.

## Opis danych

W ramach projektu z Projektowania Modeli Łączenia źródeł danych, zdecydowaliśmy się wybrać dane dotyczące przeciętnych cen producentów niektórych wyrobów spożywczych na rynku krajowym. Produkty, które wybraliśmy to produkty będące przykładami kategorii tłuszczy oraz nabiałów. Dane, z których korzystamy to ceny roczne poczynając od roku 2010, na 2021 kończąc. Skupiliśmy się na dziesięciu różnych produktach:

- Cena słoniny - za 1 kg,
- Cena smalcu wieprzowego jadalnego paczkowanego w kostkach 250g – za 1 kg,
- Cena oleju jadalnego rafinowanego rzepakowego konfekcjonowanego – za 1 l,
- Cena masła świeżego o zawartości tłuszczu ok. 82,5% - za 1kg,
- Cena mleko krowiego o zawartości tłuszczu 1,5-2%, w opakowaniu z folii - za 1l,
- Cena jaj kurzych świeżych - za 1szt.,
- Cena mleka krowiego o zawartości tłuszczu 3-3,5% (przedłużony okres trwałości, UHT), w opakowaniu kartonowym - za 1l,
- Cena śmietany o zawartości tłuszczu 18% - za 1l,
- Cena sera twarogowego półtłustego - za 1kg,
- Cena sera dojrzewającego "Gouda" - za 1kg.

Jako kolumnę Y wybraliśmy natomiast wskaźnik cen towarów i usług konsumpcyjnych, czyli oczywiście wskaźnik inflacji.

| Produkt | Zmiana ceny produktu na przestrzeni lat | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
| | [zł] | [zł] | [zł] | [zł] | [zł] | [zł] | [zł] | [zł] | [zł] | [zł] | [zł] | [zł] |
| Wskaźnik cen towarów i usług konsumpcyjnych (pot. inflacja) | 102,6 | 104,3 | 103,7 | 100,9 | 100,0 | 99,1 | 99,4 | 102,0 | 101,6 | 102,3 | 103,4 | 105,1 |
| słonina - za 1kg | 3,86 | 4,32 | 5,80 | 5,72 | 5,17 | 4,45 | 4,36 | 4,46 | 4,50 | 5,17 | 6,17 | 6,02 |
| smalec wieprzowy jadalny paczkowany w kostkach 250g - za 1kg | 4,56 | 5,18 | 6,50 | 6,35 | 6,15 | 5,72 | 5,58 | 6,00 | 6,11 | 6,56 | 7,78 | 7,60 |
| olej jadalny rafinowany rzepakowy konfekcjonowany - za 1l | 4,23 | 5,29 | 5,50 | 5,21 | 5,24 | 5,46 | 4,84 | 4,97 | 4,50 | 4,43 | 4,66 | 5,49 |
| masło świeże o zawartości tłuszczu ok. 82,5% - za 1kg | 15,05 | 16,91 | 15,95 | 17,23 | 17,05 | 15,15 | 16,19 | 22,85 | 23,89 | 21,00 | 19,17 | 22,20 |
| mleko krowie o zawartości tłuszczu 1,5-2%, w opakowaniu z foli | 1,67 | 1,67 | 1,68 | 1,70 | 1,81 | 1,74 | 1,66 | 1,71 | 1,75 | 1,78 | 1,85 | 1,96 |
| jaja kurze świeże - za 1szt. | 0,27 | 0,26 | 0,41 | 0,32 | 0,29 | 0,3 | 0,26 | 0,35 | 0,32 | 0,3 | 0,32 | 0,37 |
| mleko krowie o zawartości tłuszczu 3-3,5% (przedłużony okres tr | 1,80 | 1,98 | 2,04 | 2,13 | 2,15 | 1,89 | 1,85 | 2,05 | 2,05 | 2,07 | 2,25 | 2,43 |
| śmietana o zawartości tłuszczu 18% - za 1l | 5,47 | 5,56 | 5,59 | 5,70 | 5,75 | 5,57 | 5,52 | 6,09 | 6,64 | 6,86 | 7,04 | 7,47 |
| ser twarogowy półtłusty - za 1kg | 8,66 | 8,97 | 9,10 | 9,23 | 9,58 | 9,22 | 9,01 | 9,16 | 9,22 | 9,27 | 9,49 | 9,87 |
| ser dojrzewający "Gouda" - za 1kg | 12,84 | 14,01 | 14,19 | 15,71 | 14,80 | 12,00 | 12,52 | 14,86 | 13,80 | 14,49 | 15,05 | 16,59 |

## Eliminacja Quasi-stałych

Eliminacja quasi-stałych to metoda stosowana w modelowaniu i analizie danych, której celem jest usunięcie zmiennych, które są bardzo słabo związane ze zmienną objaśnianą. Zmienne te nazywane są quasi-stałymi, ponieważ ich wartości są prawie stałe w kontekście innych zmiennych. Eliminacja quasi-stałych jest używana, ponieważ takie zmienne mogą mieć negatywny wpływ na jakość modelu i wyników analizy. Mogą powodować nadmierną złożoność modelu i zaburzenie interpretacji wyników, co może prowadzić do błędnych wniosków. Metoda eliminacji quasi-stałych może być wykonywana na kilka sposobów, w tym za pomocą analizy korelacji, testów statystycznych i wizualizacji danych. Po usunięciu quasi-stałych, pozostałe zmienne są bardziej związane ze zmienną objaśniającą i pozwalają na bardziej skuteczne i jasne modelowanie oraz analizę danych.

| ELIMINACJA QUASI-STAŁYCH | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Y | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 |
| 2010 | 102,6 | 3,86 | 4,56 | 4,23 | 15,05 | 1,67 | 0,27 | 1,80 | 5,47 | 8,66 | 12,84 |
| 2011 | 104,3 | 4,32 | 5,18 | 5,29 | 16,91 | 1,67 | 0,26 | 1,98 | 5,56 | 8,97 | 14,01 |
| 2012 | 103,7 | 5,80 | 6,50 | 5,50 | 15,95 | 1,68 | 0,41 | 2,04 | 5,59 | 9,10 | 14,19 |
| 2013 | 100,9 | 5,72 | 6,35 | 5,21 | 17,23 | 1,70 | 0,32 | 2,13 | 5,70 | 9,23 | 15,71 |
| 2014 | 100,0 | 5,17 | 6,15 | 5,24 | 17,05 | 1,81 | 0,29 | 2,15 | 5,75 | 9,58 | 14,80 |
| 2015 | 99,1 | 4,45 | 5,72 | 5,46 | 15,15 | 1,74 | 0,3 | 1,89 | 5,57 | 9,22 | 12,00 |
| 2016 | 99,4 | 4,36 | 5,58 | 4,84 | 16,19 | 1,66 | 0,26 | 1,85 | 5,52 | 9,01 | 12,52 |
| 2017 | 102,0 | 4,46 | 6,00 | 4,97 | 22,85 | 1,71 | 0,35 | 2,05 | 6,09 | 9,16 | 14,86 |
| 2018 | 101,6 | 4,50 | 6,11 | 4,50 | 23,89 | 1,75 | 0,32 | 2,05 | 6,64 | 9,22 | 13,80 |
| 2019 | 102,3 | 5,17 | 6,56 | 4,43 | 21,00 | 1,78 | 0,3 | 2,07 | 6,86 | 9,27 | 14,49 |
| 2020 | 103,4 | 6,17 | 7,78 | 4,66 | 19,17 | 1,85 | 0,32 | 2,25 | 7,04 | 9,49 | 15,05 |
| 2021 | 105,1 | 6,02 | 7,60 | 5,49 | 22,20 | 1,96 | 0,37 | 2,43 | 7,47 | 9,87 | 16,59 |
| WSP. ZM. | - | 14,42% | 13,07% | 8,45% | 16,13% | 3,47% | 13,48% | 6,36% | 9,36% | 2,59% | 7,83% |
| WAR KRYT | 10% | 14,88% | 14,09% | 8,58% | 16,28% | 4,91% | 13,78% | 8,10% | 11,07% | 3,23% | 8,90% |

Jak widać, w naszym przypadku po eliminacji Quasi – stałych oraz ustaleniu wartości krytycznej jako 10%, z dziesięciu zmiennych pozostaje nam dokładnie pięć. Są to te, których współczynnik zmienności był większy od wspomnianych 10%.

## Metoda analizy macierzy współczynników korelacji

Analiza macierzy współczynników korelacji to jedna z metod stosowanych do identyfikacji wzajemnych powiązań między zmiennymi w danych. Polega ona na określeniu współczynników korelacji między każdą parą zmiennych, a następnie na przedstawieniu tych wartości w formie macierzy. Współczynnik korelacji między dwoma zmiennymi mówi o stopniu, w jakim ich wartości są skorelowane. Wartość współczynnika korelacji może wynosić od -1 do 1. Wartość -1 oznacza ujemną korelację, co oznacza, że wzrost wartości jednej zmiennej powoduje spadek wartości drugiej. Wartość 1 oznacza dodatnią korelację, co oznacza, że wzrost wartości jednej zmiennej powoduje wzrost wartości drugiej. Wartość 0 oznacza brak korelacji między zmiennymi.

Metoda analizy macierzy współczynników korelacji

| | *Y* | *X1* | *X2* | *X4* | *X6* | *X8* |
|---|---|---|---|---|---|---|
| Y | 1 | 0,382461684 | 0,3548139 | 0,313465 | 0,42873264 | 0,486797 |
| *X1* | 0,3824617 | 1 | 0,9081276 | 0,188092 | 0,6352101 | 0,55614 |
| *X2* | 0,3548139 | 0,908127586 | 1 | 0,482989 | 0,60499182 | 0,798629 |
| *X4* | 0,3134648 | 0,188092197 | 0,482989 | 1 | 0,35469661 | 0,772636 |
| *X6* | 0,4287326 | 0,635210104 | 0,6049918 | 0,354697 | 1 | 0,362527 |
| *X8* | 0,4867967 | 0,556140496 | 0,7986294 | 0,772636 | 0,36252664 | 1 |

W naszym przypadku, wszystkie uzyskane wartości są wartościami dodatnimi. Wynika z tego, że przy jednoczesnym wzroście cen jednego produktu, zwiększa się również cena drugiego. Jest tak dlatego, że jak oczywiście wiadomo, ceny produktów spożywczych wzrastały na przestrzeni lat, co jest oczywiście rezultatem między innymi postępującej inflacji.

## Metoda Hellwiga

Metoda Hellwiga może być wykorzystywana w połączeniu z kombinacjami zmiennych objaśniających w celu lepszego zrozumienia zależności między zmiennymi i stopnia normalności ich rozkładów. W tym przypadku, po dodaniu kombinacji zmiennych objaśniających, możemy wykonać test Hellwiga i porównać wyniki z wynikami bez uwzględnienia tych zmiennych. Jeśli dodanie kombinacji zmiennych objaśniających poprawiło stopień normalności rozkładu, to możemy uznać, że te zmienne mają istotny wpływ na nasze dane i warto je uwzględnić w dalszej analizie. W przypadku, gdy wyniki nie uległy znaczącej poprawie, możemy uznać, że kombinacja ta nie ma istotnego wpływu na nasze dane i nie jest konieczna do uwzględnienia. W ten sposób, metoda Hellwiga w połączeniu z kombinacjami zmiennych objaśniających pozwala na lepsze zrozumienie zależności między zmiennymi i stopnia normalności ich rozkładów, co jest ważne przy wyborze odpowiedniego modelu statystycznego i dokonywaniu innych obliczeń.

Metoda Hellwiga

| | Kombinac | Indywidualne pojemności nośników informacji h | | | | | Integralne pojemnośći nośników informacji H |
|---|---|---|---|---|---|---|---|
| 1 | X1 | 0,1462769 | | | | | 0,14627694 |
| 2 | X2 | | 0,125892921 | | | | 0,125892921 |
| 3 | X4 | | | 0,09826 | | | 0,098260164 |
| 4 | X6 | | | | 0,183812 | | 0,183811675 |
| 5 | X8 | | | | | 0,23697107 | 0,236971067 |
| 6 | X1X2 | 0,0766599 | 0,065977203 | | | | 0,14263714 |
| 7 | X1X4 | 0,1231192 | | 0,082704 | | | 0,205823339 |
| 8 | X1X6 | 0,0894545 | | | 0,112409 | | 0,201863121 |
| 9 | X1X8 | 0,0939998 | | | | 0,15228128 | 0,246281108 |
| 10 | X2X4 | | 0,084891336 | 0,066258 | | | 0,151149522 |
| 11 | X2X6 | | 0,078438357 | | 0,114525 | | 0,192963349 |
| 12 | X2X8 | | 0,069993809 | | | 0,13175091 | 0,201744723 |
| 13 | X4X6 | | | 0,072533 | 0,135685 | | 0,208217719 |
| 14 | X4X8 | | | 0,055432 | | 0,1336829 | 0,189114576 |
| 15 | X6X8 | | | | 0,134905 | 0,17392032 | 0,308825332 |
| 16 | X1X2X4 | 0,0697813 | 0,052650264 | 0,0588 | | | 0,181231914 |
| 17 | X1X2X6 | 0,0575138 | 0,050094286 | | 0,082051 | | 0,189659443 |
| 18 | X1X2X8 | 0,0593592 | 0,046510611 | | | 0,10063449 | 0,206504284 |
| 19 | X1X4X6 | 0,0802264 | | 0,06369 | 0,092372 | | 0,236288348 |
| 20 | X1X4X8 | 0,0838632 | | 0,050114 | | 0,10175777 | 0,235735102 |
| 21 | X1X6X8 | 0,066752 | | | 0,09201 | 0,12350817 | 0,282270085 |
| 22 | X2X4X6 | | 0,060294097 | 0,05347 | 0,093796 | | 0,207559991 |
| 23 | X2X4X8 | | 0,055177027 | 0,043562 | | 0,09216128 | 0,19090059 |
| 24 | X2X6X8 | | 0,052376357 | | 0,093423 | 0,10965014 | 0,2554496 |
| 25 | X4X6X8 | | | 0,0461894 | 0,10704 | 0,11098504 | 0,264214481 |
| 26 | X1X2X4X6 | 0,0535532 | 0,063069179 | 0,0485049 | 0,070836 | | 0,235963117 |
| 27 | X1X2X4X8 | 0,0551497 | 0,039468008 | 0,0680606 | | 0,1113897 | 0,274067988 |
| 28 | X1X2X6X8 | 0,0471941 | 0,038014031 | | 0,070623 | 0,08720839 | 0,243039162 |
| 29 | X1X4X6X8 | 0,0614753 | | 0,0424372 | 0,078137 | 0,08805069 | 0,270100024 |
| 30 | X2X4X6X8 | | 0,043612719 | 0,0376429 | 0,079154 | 0,08077297 | 0,241182226 |
| 31 | X1X2X4X6 | 0,0444939 | 0,033175657 | 0,0351128 | 0,062153 | 0,06790134 | 0,242836346 |

# Estymacja parametrów

Estymacja parametrów to proces określania wartości nieznanych parametrów modelu na podstawie dostępnych danych. W statystyce, estymacja parametrów polega na doborze takich wartości parametrów, które najlepiej pasują do danych i pozwala na opisanie ich rozkładu. Po dokonaniu estymacji parametrów, ważne jest wyznaczenie ich błędów i przeprowadzenie testów statystycznych, aby ocenić ich jakość i stopień ufności w nasze wyniki. Wnioski z estymacji parametrów pozwalają na lepsze zrozumienie i opisanie danych. Wiedza o estymowanych parametrach pozwala na lepsze prognozowanie i podejmowanie decyzji opartych na danych.

Estymacja parametrów

| | Y | X1 | X2 | X4 | X6 | X8 | |
|---|---|---|---|---|---|---|---|
| 2010 | 102,6 | 3,86 | 4,56 | 15,05 | 0,27 | 5,47 | 1,00 |
| 2011 | 104,3 | 4,32 | 5,18 | 16,91 | 0,26 | 5,56 | 1,00 |
| 2012 | 103,7 | 5,80 | 6,50 | 15,95 | 0,41 | 5,59 | 1,00 |
| 2013 | 100,9 | 5,72 | 6,35 | 17,23 | 0,32 | 5,70 | 1,00 |
| 2014 | 100,0 | 5,17 | 6,15 | 17,05 | 0,29 | 5,75 | 1,00 |
| 2015 | 99,1 | 4,45 | 5,72 | 15,15 | 0,3 | 5,57 | 1,00 |
| 2016 | 99,4 | 4,36 | 5,58 | 16,19 | 0,26 | 5,52 | 1,00 |
| 2017 | 102,0 | 4,46 | 6,00 | 22,85 | 0,35 | 6,09 | 1,00 |
| 2018 | 101,6 | 4,50 | 6,11 | 23,89 | 0,32 | 6,64 | 1,00 |
| 2019 | 102,3 | 5,17 | 6,56 | 21,00 | 0,3 | 6,86 | 1,00 |
| 2020 | 103,4 | 6,17 | 7,78 | 19,17 | 0,32 | 7,04 | 1,0 |
| 2021 | 105,1 | 6,02 | 7,60 | 22,20 | 0,37 | 7,47 | 1,00 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 102,6 | 104,3 | 103,7 | 100,9 | 100,0 | 99,1 | 99,4 | 102,0 | 101,6 | 102,3 | 103,4 | 105,1 |
| 3,86 | 4,32 | 5,80 | 5,72 | 5,17 | 4,45 | 4,36 | 4,46 | 4,50 | 5,17 | 6,17 | 6,02 |
| 4,56 | 5,18 | 6,50 | 6,35 | 6,15 | 5,72 | 5,58 | 6,00 | 6,11 | 6,56 | 7,78 | 7,60 |
| 15,05 | 16,91 | 15,95 | 17,23 | 17,05 | 15,15 | 16,19 | 22,85 | 23,89 | 21,00 | 19,17 | 22,20 |
| 0,27 | 0,26 | 0,41 | 0,32 | 0,29 | 0,3 | 0,26 | 0,35 | 0,32 | 0,3 | 0,32 | 0,37 |
| 5,47 | 5,56 | 5,59 | 5,70 | 5,75 | 5,57 | 5,52 | 6,09 | 6,64 | 6,86 | 7,04 | 7,47 |
| 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |

XT

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |
| 3,86 | 4,32 | 5,80 | 5,72 | 5,17 | 4,45 | 4,36 | 4,46 | 4,50 | 5,17 | 6,17 | 6,02 |
| 4,56 | 5,18 | 6,50 | 6,35 | 6,15 | 5,72 | 5,58 | 6,00 | 6,11 | 6,56 | 7,78 | 7,60 |
| 15,05 | 16,91 | 15,95 | 17,23 | 17,05 | 15,15 | 16,19 | 22,85 | 23,89 | 21,00 | 19,17 | 22,20 |
| 0,27 | 0,26 | 0,41 | 0,32 | 0,29 | 0,3 | 0,26 | 0,35 | 0,32 | 0,3 | 0,32 | 0,37 |
| 5,47 | 5,56 | 5,59 | 5,70 | 5,75 | 5,57 | 5,52 | 6,09 | 6,64 | 6,86 | 7,04 | 7,47 |

X

| | X1 | X2 | X4 | X6 | X8 |
|---|---|---|---|---|---|
| 1,00 | 3,86 | 4,56 | 15,05 | 0,27 | 5,47 |
| 1,00 | 4,32 | 5,18 | 16,91 | 0,26 | 5,56 |
| 1,00 | 5,80 | 6,50 | 15,95 | 0,41 | 5,59 |
| 1,00 | 5,72 | 6,35 | 17,23 | 0,32 | 5,70 |
| 1,00 | 5,17 | 6,15 | 17,05 | 0,29 | 5,75 |
| 1,00 | 4,45 | 5,72 | 15,15 | 0,3 | 5,57 |
| 1,00 | 4,36 | 5,58 | 16,19 | 0,26 | 5,52 |
| 1,00 | 4,46 | 6,00 | 22,85 | 0,35 | 6,09 |
| 1,00 | 4,50 | 6,11 | 23,89 | 0,32 | 6,64 |
| 1,00 | 5,17 | 6,56 | 21,00 | 0,3 | 6,86 |
| 1,0 | 6,17 | 7,78 | 19,17 | 0,32 | 7,04 |
| 1,00 | 6,02 | 7,60 | 22,20 | 0,37 | 7,47 |

XTX

| | | | | | |
|---|---|---|---|---|---|
| 12 | 60 | 74,09 | 222,64 | 3,77 | 73,26 |
| 60,0 | 306,6 | 377,5 | 1118,273 | 19,0955 | 369,6544 |
| 74,09 | 377,50 | 466,53 | 1389,854 | 23,5501 | 457,9537 |
| 222,64 | 1 118,27 | 1 389,85 | 4240,251 | 70,5028 | 1378,1433 |
| 3,77 | 19,10 | 23,55 | 70,5028 | 1,2069 | 23,1431 |
| 73,26 | 369,6544 | 457,9537 | 1378,143 | 23,1431 | 452,7302 |

(XTX)^(-1)

| | | | | | |
|---|---|---|---|---|---|
| 11,869398 | -1,60296 | 2,530089715 | 0,100815 | -14,145798 | -2,7549204 |
| -1,6029569 | 2,054604 | -1,863297527 | 0,137624 | -4,8473573 | 0,295454034 |
| 2,5300897 | -1,8633 | 2,294086149 | -0,0541 | -0,439037 | -1,02147104 |
| 0,1008148 | 0,137624 | -0,054095983 | 0,040511 | -1,028444 | -0,14470829 |
| -14,145798 | -4,84736 | -0,439037038 | -1,02844 | 102,138141 | 4,600481079 |
| -2,7549204 | 0,295454 | -1,021471042 | -0,14471 | 4,60048108 | 1,445353967 |

| XTY | a |
|---|---|
| 6128,29 | 8962,1862 |
| 7566,475 | -72498,57 |
| 22737,64 | 49012,132 |
| 385,076 | -7427,505 |
| 7482,233 | 636108,55 |
| 1224,4 | -1737,405 |

PODSUMOWANIE - WYJŚCIE

*Statystyki regresji*

Model oszacowany

$$\hat{Y} = 8962{,}186\, x_1 - 72498{,}6\, x_2 + 49012{,}13\, x_4 - 7427{,}51\, x_6 + 636108{,}5\, x_8 - 1737{,}41$$

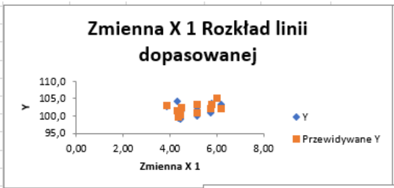
Wartościami, które na ten moment przydadzą nam się najbardziej, są wartości a, z których skorzystamy w jednym z kolejnych kroków. Ponadto korzystając z tych wartości, jesteśmy także w stanie wyprowadzić wzór na model oszacowany, który znajduje się w prawym dolnym rogu powyższego obrazu.

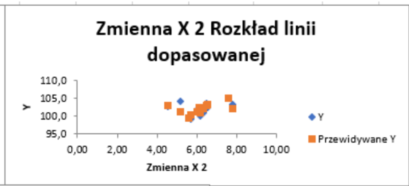
## Regresja

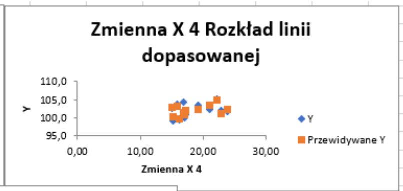
Korzystając z wbudowanej funkcji w MS Excel przeanalizowaliśmy danę pod kątem regresji. Po zaznaczeniu obszarów zawierających kolumnę Y oraz kolumny zawierające dane zmiennych objaśniających X1, X2, X4, X6 oraz X8, otrzymaliśmy pokazane poniżej podsumowanie.

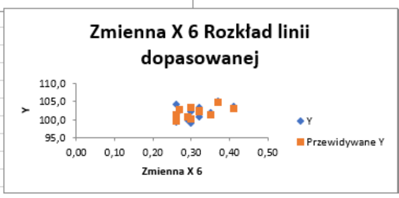
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| PODSUMOWANIE - WYJŚCIE | | | | | | | | |
| | | | | | | | | |
| *Statystyki regresji* | | | | | | | | |
| Wielokro | 0,7709559 | | | | | | | |
| R kwadra | 0,594373 | | | | | | | |
| Dopasow | 0,2563505 | | | | | | | |
| Błąd stan | 1,6593089 | | | | | | | |
| Obserwac | 12 | | | | | | | |
| | | | | | | | | |
| ANALIZA WARIANCJI | | | | | | | | |
| | *df* | *SS* | *MS* | *F* | *Istotność F* | | | |
| Regresja | 5 | 24,20683101 | 4,8413662 | 1,758383 | 0,25538738 | | | |
| Resztkow | 6 | 16,51983566 | 2,7533059 | | | | | |
| Razem | 11 | 40,72666667 | | | | | | |
| | | | | | | | | |
| | *Współczynnik* | *łąd standardow* | *t Stat* | *Wartość-p* | *Dolne 95%* | *Górne 95%* | *Dolne 95,0%* | *Górne 95,0%* |
| Przecięcie | 85,494648 | 5,716649709 | 14,955376 | 5,63E-06 | 71,5065105 | 99,48279 | 71,5065105 | 99,482786 |
| Zmienna | 3,2464271 | 2,378435106 | 1,3649425 | 0,221237 | -2,5733939 | 9,066248 | -2,5733939 | 9,0662482 |
| Zmienna | -4,8020958 | 2,513229203 | -1,9107274 | 0,104602 | -10,951746 | 1,347555 | -10,951746 | 1,3475545 |
| Zmienna | -0,1295869 | 0,333974234 | -0,3880146 | 0,711398 | -0,9467924 | 0,687619 | -0,9467924 | 0,6876186 |
| Zmienna | 20,342141 | 16,76954236 | 1,2130409 | 0,2707 | -20,691451 | 61,37573 | -20,691451 | 61,375733 |
| Zmienna | 4,2537158 | 1,994868834 | 2,1323286 | 0,076961 | -0,6275524 | 9,134984 | -0,6275524 | 9,134984 |

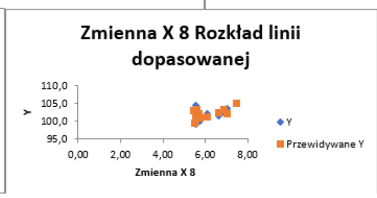
Ponadto wygenerowaliśmy również rozkład linii dopasowanych dla każdego X.



## Składniki resztowe

Składniki resztowe są to różnice pomiędzy wartościami faktycznymi a wartościami przewidywanymi przez model. Są one używane do oceny jakości modelu i jego zgodności z danymi.

SKŁADNIKI RESZTOWE - WYJŚCIE

| *Obserwacjc* | *zewidywane* | *ładniki resztowe* |
|---|---|---|
| 1 | 102,93822 | -0,33822131 |
| 2 | 101,39266 | 2,907340184 |
| 3 | 103,16194 | 0,538058502 |
| 4 | 102,09379 | -1,19378655 |
| 5 | 100,89442 | -0,89441799 |
| 6 | 100,30586 | -1,20585928 |
| 7 | 99,524832 | -0,12483248 |
| 8 | 101,22496 | 0,775042855 |
| 9 | 102,42109 | -0,8210928 |
| 10 | 103,33874 | -1,03873654 |
| 11 | 102,13626 | 1,263737623 |
| 12 | 104,96723 | 0,132767793 |

| *Obserwacjc* | *adniki resztowe* |
|---|---|
| 1 | -0,3382213 |
| 2 | 2,9073402 |
| 3 | 0,5380585 |
| 4 | -1,1937866 |
| 5 | -0,894418 |
| 6 | -1,2058593 |
| 7 | -0,1248325 |
| 8 | 0,7750429 |
| 9 | -0,8210928 |
| 10 | -1,0387365 |
| 11 | 1,2637376 |
| 12 | 0,1327678 |


| *Obserwacja* | *dniki resztowe* |
|---|---|
| 1 | -1,20586 |
| 2 | -1,19379 |
| 3 | -1,03874 |
| 4 | -0,89442 |
| 5 | -0,82109 |
| 6 | -0,33822 |
| 7 | -0,12483 |
| 8 | 0,132768 |
| 9 | 0,538059 |
| 10 | 0,775043 |
| 11 | 1,263738 |
| 12 | 2,90734 |


## Istotność statystyczna

Korzystając z wygenerowanej wcześniej tabeli, zawierającej statystyki regresji oraz bazując na naszych danych sprawdziliśmy również istotność statystyczną. W naszym przypadku k jest równe 5 jako, że jest to ilość zmiennych pozostałych po eliminacji Quasi – stałych. N oznacza natomiast kolumny zawierające kolejno poszczególne lata.

| ISTOTNOŚĆ STATYSTYCZNA F | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Statystyki regresji* | | | k | 5 | | | alfa | 0,05 |
| Wielokro | 0,7709559 | | n | 12 | | | s1 = k | 5 |
| R kwadra | 0,594372999 | | | | | | s2 = n-k-1 | 6 |
| Dopasow | 0,256350498 | | | | | | | |
| Błąd stan | 1,659308875 | | Femp | 1,758382941 | | | F | 0,202008 |
| Obserwac | 12 | | | | | | | |

## Weryfikacja modelu

SSE, SST i SSR to skróty od odpowiednio resztowej sumy kwadratów, całkowitej sumy kwadratów odchyleń i regresyjnej sumy kwadratów. Są to używane w statystyce do weryfikacji modeli regresji.

SSE (sum of squared errors) to suma kwadratów różnic między wartościami prognozowanymi przez model i rzeczywistymi wartościami.

SST (sum of squared totals) to suma kwadratów różnic między rzeczywistymi wartościami a średnią wartością danych. Jest to miernik zmienności danych.

SSR (sum of squared residuals) to część zmienności wyjaśniona przez model liniowy. Im większe SSR, tym lepsze pasowanie modelu do danych.

| WERYFIKACJA MODELU | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| t | Y | X1 | X2 | X4 | X6 | X8 | $\hat{y}$ | e(y-y^) | | | | a |
| 1 | 102,6 | 3,86 | 4,56 | 15,05 | 0,27 | 5,47 | 3074,79101 | -2 972,2 | 8 833 919,4 | 0,3 | 6574609,715 | 8962,186237 |
| 2 | 104,3 | 4,32 | 5,18 | 16,91 | 0,26 | 5,56 | -20219,639 | 20 323,9 | 413 062 498,8 | 5,1 | 429746577,3 | -72498,5662 |
| 3 | 103,7 | 5,80 | 6,50 | 15,95 | 0,41 | 5,59 | 39673,0621 | -39 569,4 | 1 565 734 417,2 | 2,8 | 1533691323 | 49012,13211 |
| 4 | 100,9 | 5,72 | 6,35 | 17,23 | 0,32 | 5,70 | -28826,963 | 28 927,9 | 836 821 245,1 | 1,3 | 860697919,6 | -7427,50515 |
| 5 | 100,0 | 5,17 | 6,15 | 17,05 | 0,29 | 5,75 | -16588,354 | 16 688,4 | 278 501 143,4 | 4,1 | 292377317,9 | 636108,5465 |
| 6 | 99,1 | 4,45 | 5,72 | 15,15 | 0,3 | 5,57 | 35321,4755 | -35 222,4 | 1 240 615 739,0 | 8,6 | 1211790732 | -1737,40534 |
| 7 | 99,4 | 4,36 | 5,58 | 16,19 | 0,26 | 5,52 | 1902,57106 | -1 803,2 | 3 251 425,9 | 6,9 | 1937330,565 | |
| 8 | 102,0 | 4,46 | 6,00 | 22,85 | 0,35 | 6,09 | 22030,0738 | -21 928,1 | 480 840 420,2 | 0,0 | 463083842,9 | |
| 9 | 101,6 | 4,50 | 6,11 | 23,89 | 0,32 | 6,64 | -3241,969 | 3 343,6 | 11 179 453,7 | 0,2 | 14082455,88 | |
| 10 | 102,3 | 5,17 | 6,56 | 21,00 | 0,3 | 6,86 | -21399,459 | 21 501,8 | 462 325 646,4 | 0,1 | 480054672,2 | |
| 11 | 103,4 | 6,17 | 7,78 | 19,17 | 0,32 | 7,04 | -8101,4518 | 8 204,9 | 67 319 592,9 | 1,9 | 74169000,57 | |
| 12 | 105,1 | 6,02 | 7,60 | 22,20 | 0,37 | 7,47 | 2504,1518 | -2 399,1 | 5 755 449,5 | 9,4 | 3973886,615 | |
| | | | | | | | | | 5 374 240 951,5 | 40,7 | 5372179669 | |
| | | | | | | | | | sse | sst | ssr | |

## Wyrazistość modelu

Obliczyliśmy również wyrazistość modelu poprzez podzielenie błędu standardowego przez średnią z wartości kolumny Y.

| WYRAZISTOŚĆ MODELU |
|---|
| 1,63% |

## Test symetrii

Test symetrii na podstawie liczby reszt dodatnich polega na ocenie, czy rozkład reszt jest symetryczny. W tym teście, liczba reszt dodatnich jest porównywana z liczbą reszt ujemnych, aby ustalić, czy model jest w stanie wyjaśnić symetrię danych. Jeśli liczba reszt dodatnich jest zbliżona do liczby reszt ujemnych, można uznać, że model jest w stanie wyjaśnić symetrię danych i jest to pozytywny wynik testu. Test symetrii na podstawie liczby reszt dodatnich jest ważny, ponieważ pomaga ocenić, czy model jest w stanie wyjaśnić symetrię danych. Jeśli wynik testu jest pozytywny, można uznać, że model jest wystarczająco dobry i można go użyć do prognozowania. W przeciwnym razie konieczne może być dostosowanie modelu, aby lepiej opisywał dane.

TEST SYMETRII

| e(y-y^) |
|---|
| -2 972,2 |
| 20 323,9 |
| -39 569,4 |
| 28 927,9 |
| 16 688,4 |
| -35 222,4 |
| -1 803,2 |
| -21 928,1 |
| 3 343,6 |
| 21 501,8 |
| 8 204,9 |
| -2 399,1 |

| Liczba reszt dodatnich |
|---|
| 6 |

| Liczba próby |
|---|
| 12 |

| Wartość statystyczna |
|---|
| 0 |

Rozkład jest idealnie symetryczny

```
=MODUŁ.LICZBY(E221/E224-0,5)/PIERWIASTEK((E221/E224*(1-E221/E224)/(E224-1)))
```

Liczba reszt dodatnich równa jest liczbie reszt ujemnych zatem wynik naszego działania jest równy zero. Stwierdzamy zatem, że rozkład jest symetryczny.

## Variance Inflation Factor

Korzystając z utworzonej tabel przeprowadziliśy również test sprawdzający współliniowość zmiennych. Obliczyliśmy to za pomocą dwóch metod, z których korzystaliśmy na zajęciach. Wynik, który wzięliśmy pod uwagę to wartość obliczona pierwszym sposobem. Zatem w przypadku, gdy VIF osiąga wartość mniejszą od 4, można założyć brak współliniowości predyktorów w modelu.

VIF

| X1 | | X2 | | X4 | | X6 | | X8 | |
|---|---|---|---|---|---|---|---|---|---|
| Średnia | 5 | Średnia | 6,1741667 | Średnia | 18,5533333 | Średnia | 0,31416667 | Średnia | 6,105 |
| Błąd stan | 0,2243036 | Błąd standar | 0,2623595 | Błąd stan | 0,91094551 | Błąd stan | 0,01305341 | Błąd stand | 0,2037136 |
| Mediana | 4,835 | Mediana | 6,13 | Mediana | 17,14 | Mediana | 0,31 | Mediana | 5,725 |
| Tryb | 5,17 | Tryb | #N/D | Tryb | #N/D | Tryb | 0,32 | Tryb | #N/D |
| Odchylen | 0,7770106 | Odchylenie s | 0,90884 | Odchylen | 3,1556078 | Odchylen | 0,04521833 | Odchyleni | 0,7056847 |
| Wariancja | 0,6037455 | Wariancja pr | 0,8259902 | Wariancja | 9,95786061 | Wariancja | 0,0020447 | Wariancja | 0,4979909 |
| Kurtoza | -1,4629371 | Kurtoza | 0,2974201 | Kurtoza | -1,2648567 | Kurtoza | 0,38249661 | Kurtoza | -0,725625 |
| Skośność | 0,2443903 | Skośność | 0,2382934 | Skośność | 0,59459906 | Skośność | 0,79877578 | Skośność | 0,9080485 |
| Zakres | 2,31 | Zakres | 3,22 | Zakres | 8,84 | Zakres | 0,15 | Zakres | 2 |
| Minimum | 3,86 | Minimum | 4,56 | Minimum | 15,05 | Minimum | 0,26 | Minimum | 5,47 |
| Maksimu | 6,17 | Maksimum | 7,78 | Maksimu | 23,89 | Maksimu | 0,41 | Maksimun | 7,47 |
| Suma | 60 | Suma | 74,09 | Suma | 222,64 | Suma | 3,77 | Suma | 73,26 |
| Licznik | 12 | Licznik | 12 | Licznik | 12 | Licznik | 12 | Licznik | 12 |

| | | | |
|---|---|---|---|
| VIF | 1. MET. | | 2,4653191 |
| | 2. MET. | X1 | 5,7369808 |
| | | X2 | 8,29364 |
| | | X4 | 13,286703 |
| | | X6 | 0,1369898 |
| | | X8 | 3,9689349 |

| Statystyki regresji | |
|---|---|
| Wielokro | 0,7709559 |
| R kwadra | 0,594372999 |
| Dopasow | 0,256350498 |
| Błąd stan | 1,659308875 |
| Obserwac | 12 |

ANALIZA WARIANCJI

| | df | SS | MS | F | Istotność F |
|---|---|---|---|---|---|
| Regresja | 5 | 24,20683 | 4,8413662 | 1,758382941 | 0,255387 |
| Resztkow | 6 | 16,51984 | 2,75330594 | | |
| Razem | 11 | 40,72667 | | | |

| | Współczynniki | d standardc | t Stat | Wartość-p | Dolne 95% | Górne 95% | olne 95,0% | Górne 95,0% |
|---|---|---|---|---|---|---|---|---|
| Przecięcie | 85,4946484 | 5,71665 | 14,9553764 | 5,62769E-06 | 71,50651 | 99,48279 | 71,50651 | 99,482786 |
| X1 | 3,246427113 | 2,378435 | 1,36494248 | 0,221236846 | -2,57339 | 9,066248 | -2,57339 | 9,0662482 |
| X2 | -4,802095816 | 2,513229 | -1,9107274 | 0,104602372 | -10,9517 | 1,347555 | -10,9517 | 1,3475545 |
| X4 | -0,129586863 | 0,333974 | -0,3880146 | 0,711397982 | -0,94679 | 0,687619 | -0,94679 | 0,6876186 |
| X6 | 20,34214112 | 16,76954 | 1,21304092 | 0,270699949 | -20,6915 | 61,37573 | -20,6915 | 61,375733 |
| X8 | 4,253715788 | 1,994869 | 2,13232856 | 0,076961427 | -0,62755 | 9,134984 | -0,62755 | 9,134984 |

## Test Goldfelda – Quandta

Test Goldfelda-Quandta jest jednym z testów heteroskedastyczności, który polega na sprawdzeniu, czy wahania w odpowiedzi są stałe dla wszystkich wartości zmiennych objaśniających. W przypadku heteroskedastyczności, czyli niestabilności wariancji, wahania odpowiedzi nie są stałe i zależą od wartości zmiennych objaśniających. Test Goldfelda-Quandta polega na porównaniu wartości statystyki GQ z kryterium kwantylowym dla odpowiedniego rozkładu. Jeśli wartość statystyki jest istotna na poziomie ufności, oznacza to, że występuje heteroskedastyczność. W takim przypadku nie należy stosować metod opartych na założeniu stałej wariancji, ponieważ wyniki byłyby niepoprawne. Wnioski z testu Goldfelda-Quandta są ważne dla badania heteroskedastyczności, ponieważ niestabilność wariancji może prowadzić do niepoprawnych wniosków i błędów w estymacji parametrów. W przypadku stwierdzenia heteroskedastyczności, może być konieczne zastosowanie innej metody, takiej jak transformacja danych lub modelowanie wariancji.

TEST GOLDFELDA - QUANDTA

| e(y-y^) | e^2 |
|---|---|
| -2 972,2 | 8 833 919,4 |
| 20 323,9 | 413 062 498,8 |
| -39 569,4 | 1 565 734 417,2 |
| 28 927,9 | 836 821 245,1 |
| 16 688,4 | 278 501 143,4 |
| -35 222,4 | 1 240 615 739,0 |
| -1 803,2 | 3 251 425,9 |
| -21 928,1 | 480 840 420,2 |
| 3 343,6 | 11 179 453,7 |
| 21 501,8 | 462 325 646,4 |
| 8 204,9 | 67 319 592,9 |
| -2 399,1 | 5 755 449,5 |

| e1 | e2 | e1^2 | e2^2 |
|---|---|---|---|
| -2 972,2 | -1 803,2 | 8 833 919,4 | 3 251 425,9 |
| 20 323,9 | -21 928,1 | 413 062 498,8 | 480 840 420,2 |
| -39 569,4 | 3 343,6 | 1 565 734 417,2 | 11 179 453,7 |
| 28 927,9 | 21 501,8 | 836 821 245,1 | 462 325 646,4 |
| 16 688,4 | 8 204,9 | 278 501 143,4 | 67 319 592,9 |
| -35 222,4 | -2 399,1 | 1 240 615 739,0 | 5 755 449,5 |

| | |
|---|---|
| suma e1^2 | 4 343 568 962,9 |
| suma e2^2 | 1 030 671 988,7 |
| liczba obserwacji 1szej proby | 6 |
| liczba obserwacji 2giej proby | 6 |
| liczba zmiennych | 5 |
| warainacja 1. | 720044782,3 |
| wariancja 2. | 170448531,9 |
| f(w2/w1) | 0,236719349 |
| wartosc krytyczna | 161,4476388 |

Nie występuje heteroskedastyczność

Na podstawie uzyskanych wyników możemy stwierdzić, że w naszym przypadku nie występuje heteroskedastyczność.

## Test Shapiro Wilka

Do przeprowadzenia testu Shapiro Wilka mumimy najpierw posortować dane, a następnie obliczyć potrzebne parametry. Następnei obliczamy statystykę W, do której policzyliśmy wcześniej licznik oraz mianownik tego właśnie ilorazu. Do kolejnego kroku potrzebujemy dwóch tabel, które zostały zamieszczone obok. Z pierwszej z nich uzyskujemy wartości zawarte w kolumnie ai, natomiast z drugiej możemy odczytać wartość W dla alfy równej 0,05 oraz W równego 12. Porównując obie wartości W, możemy zauważyć, że ta obliczona poprzez działania na danych z utworzonej przez nas tabeli jest mniejsza od tej, którą odczytaliśmy z tablic. Możemy zatem odrzucić hipotezę zerową.

TEST SHAPIRO WILKA

| t | e(y-y^) | sort | U | xi | xn-i+1 | xn-i-1-xi | ai | e-sred |
|---|---|---|---|---|---|---|---|---|
| 1 | -2 972,2 | -39 569,4 | -0,121158 | -39 569,4 | 28 927,9 | 68 497,2 | 0,5475 | -39 160,7 |
| 2 | 20 323,9 | -35 222,4 | 0,9798661 | -35 222,4 | 21 501,8 | 56 724,1 | 0,3325 | -34 813,7 |
| 3 | -39 569,4 | -21 928,1 | -1,8508173 | -21 928,1 | 20 323,9 | 42 252,0 | 0,2347 | -21 519,4 |
| 4 | 28 927,9 | -2 972,2 | 1,3865056 | -2 972,2 | 16 688,4 | 19 660,5 | 0,1586 | -2 563,5 |
| 5 | 16 688,4 | -2 399,1 | 0,8080407 | -2 399,1 | 8 204,9 | 10 603,9 | 0,0922 | -1 990,4 |
| 6 | -35 222,4 | -1 803,2 | -1,6453695 | -1 803,2 | 3 343,6 | 5 146,7 | 0,0303 | -1 394,5 |
| 7 | -1 803,2 | 3 343,6 | -0,0659076 | | | | | 3 752,2 |
| 8 | -21 928,1 | 8 204,9 | -1,0170529 | | | | | 8 613,5 |
| 9 | 3 343,6 | 16 688,4 | 0,1773381 | | | | | 17 097,0 |
| 10 | 21 501,8 | 20 323,9 | 1,0355324 | | | | | 20 732,6 |
| 11 | 8 204,9 | 21 501,8 | 0,4070926 | | | | | 21 910,4 |
| 12 | -2 399,1 | 28 927,9 | -0,0940702 | | | | | 29 336,5 |
| | | | | | | | SUMA | 5 372 236 940,1 |

| ŚREDNIA | ODCH. ST. | ŚRED. U |
|---|---|---|
| -408,7 | 21158,601 | 4,51028E-17 |

| | |
|---|---|
| Licznik wzoru | 977 308 806,9 |
| Mianownik wzoru | 5372236940 |
| Statystyka W | 0,181918411 |

| | |
|---|---|
| Poziom istotn | 0,05 |

| | |
|---|---|
| W(a,12) | 0,859 |

W < W(a, 12)
Zatem odrzucamy hipotezę zerową

| n = | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.7071 | 0.7071 | 0.6872 | 0.6646 | 0.6431 | 0.6233 | 0.6052 | 0.5888 | 0.5739 | 0.5601 | 0.5475 | 0.5359 | 0.5251 |
| a2 | | | 0.1677 | 0.2413 | 0.2806 | 0.3031 | 0.3164 | 0.3244 | 0.3291 | 0.3315 | 0.3325 | 0.3325 | 0.3318 |
| a3 | | | | | 0.0875 | 0.1401 | 0.1743 | 0.1976 | 0.2141 | 0.2260 | 0.2347 | 0.2412 | 0.2460 |
| a4 | | | | | | | 0.0561 | 0.0947 | 0.1224 | 0.1429 | 0.1586 | 0.1707 | 0.1802 |
| a5 | | | | | | | | | 0.0399 | 0.0695 | 0.0922 | 0.1099 | 0.1240 |
| a6 | | | | | | | | | | | 0.0303 | 0.0539 | 0.0727 |
| a7 | | | | | | | | | | | | | 0.0240 |

| n/p | 0,01 | 0,02 | 0,05 | 0,1 |
|---|---|---|---|---|
| 3 | 0,753 | 0,756 | 0,767 | 0,789 |
| 4 | 0,687 | 0,707 | 0,748 | 0,792 |
| 5 | 0,686 | 0,715 | 0,762 | 0,806 |
| 6 | 0,713 | 0,743 | 0,788 | 0,826 |
| 7 | 0,730 | 0,760 | 0,803 | 0,838 |
| 8 | 0,749 | 0,778 | 0,818 | 0,851 |
| 9 | 0,764 | 0,791 | 0,829 | 0,859 |
| 10 | 0,781 | 0,806 | 0,842 | 0,869 |
| 11 | 0,792 | 0,817 | 0,850 | 0,876 |
| 12 | 0,805 | 0,828 | 0,859 | 0,883 |

## Test Jarque – Bera

Przeprowadzamy teraz test Jarque – Bera. Obliczamy kolejne potęgi e, a następnie liczymy potrzebne wartości. Na ich podstawie stwierdzamy, że składniki losowe modelu mają rozkład normalny.

TEST JARQUE-BERA

| | e(y-y^) | e^2 | e^3 | e^4 |
|---|---|---|---|---|
| | -2 972,2 | 8 833 919,4 | -26 256 095 820,8 | 78 038 131 953 731,7 |
| | 20 323,9 | 413 062 498,8 | 8 395 057 051 643,7 | 170 620 627 893 916 000,0 |
| | -39 569,4 | 1 565 734 417,2 | -61 955 112 113 005,6 | 2 451 524 265 333 750 000,0 |
| | 28 927,9 | 836 821 245,1 | 24 207 450 150 293,8 | 700 269 796 230 411 000,0 |
| | 16 688,4 | 278 501 143,4 | 4 647 725 537 165,7 | 77 562 886 848 058 500,0 |
| | -35 222,4 | 1 240 615 739,0 | -43 697 433 466 121,6 | 1 539 127 411 883 040 000,0 |
| | -1 803,2 | 3 251 425,9 | -5 862 877 033,7 | 10 571 770 194 420,7 |
| | -21 928,1 | 480 840 420,2 | -10 543 904 217 449,3 | 231 207 509 739 859 000,0 |
| | 3 343,6 | 11 179 453,7 | 37 379 274 948,6 | 124 980 185 252 133,0 |
| | 21 501,8 | 462 325 646,4 | 9 940 814 697 000,4 | 213 745 003 341 947 000,0 |
| | 8 204,9 | 67 319 592,9 | 552 347 282 297,6 | 4 531 927 587 926 390,0 |
| | -2 399,1 | 5 755 449,5 | -13 807 621 530,8 | 33 125 199 250 629,5 |
| SUMY | -4 903,9 | 5 374 240 951,5 | -68 461 602 397 612,0 | 5 388 836 144 145 560 000,0 |

| | |
|---|---|
| SUMA (E^4)^2 | 29 039 554 988 449 500 000 000 000 000 000 000 000,0 |
| n | 12 |
| kurtoza | -0,543630832 |
| skośność | -0,625398788 |
| w | -2,35753E-24 |
| k | 1,85066E-28 |
| JB | 4,5 |
| chi^2 | 5,991464547 |

Nie mamy podstaw do odrzucenia hipotezy zerowej
Wynika z tego, że składniki losowe modelu mją rozkład normalny

## Badanie koincydencji

Test autokorelacji reszt jest testem stosowanym w statystyce, który służy do sprawdzenia, czy reszty regresji są autokorelowane. Autokorelacja oznacza, że wartości pomiaru są skorelowane z wcześniejszymi wartościami tego samego pomiaru. Test autokorelacji reszt jest ważnym krokiem w procesie budowania modelu regresji, ponieważ pozwala zidentyfikować, czy model jest właściwie skonstruowany i czy jego wyniki są wiarygodne. W naszym przypadku po sprawdzeniu elementów z tabel, trzy z nich mają ten sam znak.

Badanie koincydencji

| Y | X1 | X2 | X4 | X6 | X8 |
|---|---|---|---|---|---|
| 102,6 | 3,86 | 4,56 | 15,05 | 0,27 | 5,47 |
| 104,3 | 4,32 | 5,18 | 16,91 | 0,26 | 5,56 |
| 103,7 | 5,80 | 6,50 | 15,95 | 0,41 | 5,59 |
| 100,9 | 5,72 | 6,35 | 17,23 | 0,32 | 5,70 |
| 100,0 | 5,17 | 6,15 | 17,05 | 0,29 | 5,75 |
| 99,1 | 4,45 | 5,72 | 15,15 | 0,3 | 5,57 |
| 99,4 | 4,36 | 5,58 | 16,19 | 0,26 | 5,52 |
| 102,0 | 4,46 | 6,00 | 22,85 | 0,35 | 6,09 |
| 101,6 | 4,50 | 6,11 | 23,89 | 0,32 | 6,64 |
| 102,3 | 5,17 | 6,56 | 21,00 | 0,3 | 6,86 |
| 103,4 | 6,17 | 7,78 | 19,17 | 0,32 | 7,04 |
| 105,1 | 6,02 | 7,60 | 22,20 | 0,37 | 7,47 |

| | *Współczynniki* |
|---|---|
| Przecięcie | 85,494648 |
| X1 | 3,2464271 |
| X2 | -4,802096 |
| X4 | -0,129587 |
| X6 | 20,342141 |
| X8 | 4,2537158 |

| | |
|---|---|
| Y | 1 |
| *X1* | 0,382462 |
| *X2* | 0,354814 |
| *X4* | 0,313465 |
| *X6* | 0,428733 |
| *X8* | 0,486797 |

ZNAKI X1, X6 I X8 SIĘ ZGADZAJĄ

KIERUNEK ZALEŻNOŚCI MIĘDZY ZMIENNĄ OBJAŚNIAJĄCĄ ZE ZMIENNYMI OBJAŚNIAJĄCYMI X1, X6 I X8 ZGODNIE Z ZALEŻNOŚCIĄ WYNIKAJĄCĄ Z DANYCH EMPIRYCZNYCH

## Test autokorelacji reszt

Ostatnim już testem jest test autokorelacji reszt. Wprowadzamy kolumnę zawierającą wartości e oraz tworzymy taką samą, z tą różnicą, że przesuwamy elementy o jeden dalej. Obliczamy następnie różnice między tymi dwoma wartościami. Musimy teraz sprawdzić ich sumy oraz obliczyć statystykę DW. Następnie korzystając z tabeli wklejonej poniżej, odczytujemy wartości du oraz dl. Ostatnim już krokiem jest sprawdzenie warunków przyjęcia bądź odrzucenia postawionej hipotezy. Trzeci przypadek pokazuje jak zachowują się uzyskane w naszym przypadku liczby. Niestety test nie rozstrzyga autokorelacji w obszarze niekonkluzywności.

TEST AUROKORELACJI RESZT

| e(y-y^) | E-1 | ROZNICE |
|---|---|---|
| -2 972,2 | | |
| 20 323,9 | -2 972,2 | 23 296,1 |
| -39 569,4 | 20 323,9 | -59 893,3 |
| 28 927,9 | -39 569,4 | 68 497,2 |
| 16 688,4 | 28 927,9 | -12 239,5 |
| -35 222,4 | 16 688,4 | -51 910,7 |
| -1 803,2 | -35 222,4 | 33 419,2 |
| -21 928,1 | -1 803,2 | -20 124,9 |
| 3 343,6 | -21 928,1 | 25 271,6 |
| 21 501,8 | 3 343,6 | 18 158,2 |
| 8 204,9 | 21 501,8 | -13 296,9 |
| -2 399,1 | 8 204,9 | -10 603,9 |
| | -2 399,1 | |

| | |
|---|---|
| SUMA KWADRATOW ROZNIC | 1,4446E+10 |
| SUMA KWADRATOW RESZT | 5374240952 |
| STATYSTYKA DW | 2,68796985 |

| | |
|---|---|
| n | 12 |
| k | 5 |

| | |
|---|---|
| dl | 0,37956 |
| du | 2,50609 |
| 4 - du | 1,49391 |

| N | k=1 | | k=2 | | k=3 | | k=4 | | k=5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 6 | 0.61018 | 1.40015 | | | | | | | | |
| 7 | 0.69955 | 1.35635 | 0.46723 | 1.89636 | | | | | | |
| 8 | 0.76290 | 1.33238 | 0.55907 | 1.77711 | 0.36744 | 2.28664 | | | | |
| 9 | 0.82428 | 1.31988 | 0.62910 | 1.69926 | 0.45476 | 2.12816 | 0.29571 | 2.58810 | | |
| 10 | 0.87913 | 1.31971 | 0.69715 | 1.64134 | 0.52534 | 2.01632 | 0.37602 | 2.41365 | 0.24269 | 2.82165 |
| 11 | 0.92733 | 1.32409 | 0.75798 | 1.60439 | 0.59477 | 1.92802 | 0.44406 | 2.28327 | 0.31549 | 2.64456 |
| 12 | 0.97076 | 1.33137 | 0.81221 | 1.57935 | 0.65765 | 1.86397 | 0.51198 | 2.17662 | 0.37956 | 2.50609 |
| 13 | 1.00973 | 1.34040 | 0.86124 | 1.56212 | 0.71465 | 1.81593 | 0.57446 | 2.09428 | 0.44448 | 2.38967 |
| 14 | 1.04495 | 1.35027 | 0.90544 | 1.55066 | 0.76666 | 1.77882 | 0.63206 | 2.02955 | 0.50516 | 2.29593 |
| 15 | 1.07697 | 1.36054 | 0.94554 | 1.54318 | 0.81396 | 1.75014 | 0.68519 | 1.97735 | 0.56197 | 2.21981 |
| 16 | 1.10617 | 1.37092 | 0.98204 | 1.53860 | 0.85718 | 1.72773 | 0.73400 | 1.93506 | 0.61495 | 2.15672 |
| 17 | 1.13295 | 1.38122 | 1.01543 | 1.53614 | 0.89675 | 1.71009 | 0.77898 | 1.90047 | 0.66414 | 2.10414 |
| 18 | 1.15759 | 1.39133 | 1.04607 | 1.53525 | 0.93310 | 1.69614 | 0.82044 | 1.87189 | 0.70984 | 2.06000 |

- Hipoteza $H_0$ jest odrzucana jeżeli zachodzi nierówność $d<d_l$ i stwierdzamy istnienie istotnej dodatniej autokorelacji.
- zachodzenie nierówności $d_u<d<4-d_U$ test nie rozstrzyga kwestii autokorelacji gdyż jesteśmy w tzw. obszarze niekonkluzywności
- Gdy zachodzi nierówność $d_U<d<4-d_U$ nie mamy podstaw do odrzucenia hipotezy zerowej
- W przypadku $4-d_u<d<4-d_l$ test nie rozstrzyga kwestii autokorelacji gdyż jesteśmy w tzw. obszarze niekonkluzywności
- Gdy zachodzi $d>4-d_l$ stwierdzamy istnienie istotnej ujemnej autokorelacji.

| |
|---|
| 4 - du < du < DW |
| TEST NIE ROZSTRZYGA KWESTII AUTOKORELACJI ( JESTEŚMY W OBSZARZE NIEKONKLUZYWNOŚCI |