

**Wydział Zarządzania**

**Badanie czynników wpływających na średnie wynagrodzenie   
w podregionach w Polsce   
w 2021r.**

Przedmiot: Ekonometria

Prowadzący: dr Paweł Zając

*Arkadiusz Florek*

*Informatyka i Ekonometria*

*Studia stacjonarne*

# Cel projektu

Wynagrodzenie to doskonały sposób na weryfikację stanu gospodarki. Poprawa perspektyw biznesowych znajduje odzwierciedlenie w wyższym poziome pensji. Jej wzrost jest wynikiem lepszej sytuacji na rynku pracy(lepsza pozycja negocjacyjna pracowników) oraz dobrej sytuacji finansowej firm[[1]](#footnote-1). Można stwierdzić, że w wynagrodzenia to bardzo ważny czynnik w aktualnych społeczeństwach, przez co zostały obiektem badań, jednak nie można badać wynagrodzenia każdego człowieka osobno, ponieważ zależy od zbyt wielu czynników, dlatego potrzebowano znaleźć uniwersalny wskaźnik pozwalający badać to zjawisko.

Obraz zawierający tekst, linia, zrzut ekranu, Wykres

Opis wygenerowany automatycznieWskaźnikiem najczęściej używanym do opisu wynagrodzeń w Polsce jest przeciętne miesięczne wynagrodzenie brutto. W okresie od 2004 r. do 2012 r. przeciętne wynagrodzenie wzrosło z 2289,57 zł do 3521,67, czyli o 54%. Dynamika wzrostu była mocno zróżnicowana: w latach 2004–2006 nie przekraczała 5%, w latach 2007 i 2008 zwiększyła się, a w latach 2009–2011 utrzymywała się na wyraźnie niższym poziomie, nieznacznie przekraczającym wskaźnik wzrostu cen towarów i usług konsumpcyjnych[[2]](#footnote-2).

Rysunek 1 Krzywa przedstawiająca wartość średniego wynagrodzenia brutto w latach 2012-2021

Tak jak widać na powyższym wykresie [rysunek 1], stworzonym na podstawie danych z GUS, od 2012 roku średnie wynagrodzenie brutto zachowało swoją tendencje wzrostową i nic nie wskazuje, że ma to ulec zmianie.

W projekcie chce zbadać wpływ wybranych czynników na średnie wynagrodzenie w podregionach w Polsce w 2021 r. W kolejnych częściach przedstawię bliżej zmienne objaśniające, które wziąłem pod uwagę oraz wybiorę właściwy model i zweryfikuje postawione hipotezy.

# Hipotezy badawcze

1. Wyższy stopień zgłaszania wynalazków skutkuje wyższym średnim wynagrodzeniem
2. Wzrost stopy bezrobocia powoduje zmniejszenie się średniej pensji
3. Czy większa stopa osób poszkodowanych w wypadkach przy pracy tym mniejsze jest średnie wynagrodzenie
4. Czym mniejszy poziom emisji SO2 tym większa średnia pensja
5. Wysoki nakład inwestycyjny w przedsiębiorstwach przekłada się na wysokie średnie wynagrodzenie
6. Większa liczba przestępstw stwierdzonych przez policję powoduje spadek średniego wynagrodzenia

# 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 X1…X6 zmiennych objaśniających dla wszystkich podregionów - 73 obserwacji - w 2021 r. Model ten można zapisać w następującej postaci:

Zmienna objaśniania:

* Y - Średnia wartość wynagrodzenia brutto [zł]

Potencjalne zmienne objaśniające:

* X1 – Poziom emisji SO2 na 1 mieszkańca [kg]
* X2 – Przestępstwa stwierdzone przez Policję ogółem na 1000 mieszkańców[%]
* X3 - Nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca [zł]
* X4 – Poszkodowani w wypadkach przy pracy na 1000 pracujących [%]
* X5 – Liczba zgłoszeń wynalazków w UPRP(Urząd Patentowy RP) na 1 mln mieszkańców [%]
* X6 - Stopa bezrobocia, obliczono jako stosunek liczby bezrobotnych do liczby cywilnej ludności aktywnej zawodowo(bez osób odbywających czynną służbę wojskową itd.) [%]

# Statystki opisowe

Tabela 1 Statystyki opisowe

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Zmienna***  ***Statystyka*** | **Y** | **X1** | **X2** | **X3** | **X4** | **X5** | **X6** |
| **Średnia** | 5466,8 | 4,7178 | 20,994 | 4793,2 | 5,3579 | 72,816 | 6,8836 |
| **Mediana** | 5319,6 | 2,3 | 18,86 | 4217 | 5,34 | 37 | 6,7 |
| **Minimalna** | 4671,2 | 0,3 | 10,88 | 1573 | 3,32 | 5,5 | 1,6 |
| **Maksymalna** | 7687,6 | 79,6 | 46,95 | 19592 | 8,02 | 327,5 | 14,2 |
| **Odch**.**stand** | 643,13 | 10,028 | 6,293 | 2651,7 | 1,1645 | 86,616 | 3,01 |
| **Wsp**.**zmienności** | 0,11764 | 2,1255 | 0,29975 | 0,55322 | 0,21735 | 1,1895 | 0,43727 |
| **Skośność** | 1,4119 | 6,1174 | 1,5951 | 2,8864 | 0,18885 | 1,8236 | 0,37061 |
| **Kurtoza** | 1,5879 | 41,528 | 3,1578 | 12,077 | -0,66116 | 2,0328 | -0,30807 |

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 4671.2, a wartość maksymalna 5319.6. W zmiennej objaśnianej Y średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 1.411. Typowa wielkość Y różni się od wartości przeciętnej średnio 643.13.
* Zmienna X1 –Minimalna wartość emisji szkodliwych gazów wynosi 0.3,

a maksymalna 79.6. W zmiennej objaśniającej X1 średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 6.12. Typowa wielkość X1 różni się od wartości przeciętnej średnio o 10.03.

* Zmienna X2 – Minimalna wartość przyrostu naturalnego wynosi 10.88, a wartość maksymalna 18.86. W zmiennej objaśniającej X2 średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 1.6. Typowa wielkość X2 różni się od wartości przeciętnej średnio o 6.28.
* Zmienna X3 – Minimalna ilość nakładów inwestycyjnych w przedsiębiorstwach jest równa 1573, a wartość maksymalna 19592. W zmiennej objaśniającej X3 średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 2.87. Typowa wielkość X3 różni się od wartości przeciętnej średnio o 2651.7.
* Zmienna X4 – Minimalna liczba poszkodowanych w wypadkach przy pracy osób wynosi 3.32, a wielkość maksymalna 8.02. Współczynnik skośności wynoszący 0.19, wskazuje na słabą asymetrię prawostronną. Typowa wielkość X4 różni się od wartości przeciętnej średnio o 1.15.
* Zmienna X5 – Minimalna ilość zgłoszonych wynalazków jest równa 5.5, a wielkość maksymalna 327.5. W zmiennej objaśniającej X5 średnia jest większa od mediany, co wskazuje na wystąpienie asymetrii prawostronnej, potwierdza to współczynnik skośności wynoszący 1.81. Typowa wielkość X5 różni się od wartości przeciętnej średnio o 86.62.
* Obraz zawierający linia, diagram, kwadrat, Wielobarwność

  Opis wygenerowany automatycznieZmienna X6 – Minimalna wartość stopy bezrobocia wynosi 1.6, a wartość maksymalna 6.7. Współczynnik skośności wynosi 0,0.38. Typowa wielkość X6 różni się od wartości przeciętnej średnio o 3.01.

Rysunek 2 Wykres rozrzutu zmiennych

Na wykresie możemy zauważyć ciekawe zależności:

* Y i X2 odznaczają się dodatnią korelacją, aczkolwiek jest dużo obserwacji odstających
* Y i X3 wykazują dość mocną dodatnią korelację, punkty skupiają się wzdłuż szybko rosnącej linii prostej
* Na wykresie Y i X5 , można zaobserwować wykrywalną dodatnią korelację, jednak jest ona zaburzona przez dużą ilość obserwacji odstających, które osłabiają zależność między zmiennymi
* Y i X6 odznaczają się ujemną korelacją, jednak znowu dość spora ilość obserwacji odstających może zaburzyć tę zależność i osłabić korelację
* Pozostałe potencjalne zmienne objaśniające nie wykazują żadnych zależności względem zmiennej objaśnianej Y
* Pomiędzy X3 i X6 możemy zobaczyć widoczną silną ujemną korelację, więc trzeba uważać na współliniowość tych zmiennych
* X5 i X6 pomimo wielu obserwacji odstających mogą być skorelowane, więc trzeba zwrócić uwagę na współliniowość tych zmiennych
* Wykresy gęstości Y, X2,X3,X5,X6 cechują się podobnym rozkładem(silna prawostronna asymetria), co może sugerować korelację między nimi

Obraz zawierający tekst, zrzut ekranu, kwadrat, Prostokąt

Opis wygenerowany automatycznie

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 skupię się na tych zależnościach, które były nieoczywiste, jak na przykład:

* Zmienna X6 jest skorelowana ze zmienną Y, ale także ze zmiennymi X2, X3 i X5, co może powodować współliniowość w modelu
* X1 i X4 są słabo skorelowane ze zmienną objaśnianą Y i najprawdopodobniej zmienne te zostaną wyeliminowane na następnych etapach estymacji
* Najwyższą korelacją na poziomie 0.7 ze zmienną objaśnianą Y odznacza się zmienna X3, czyli nakłady inwestycyjne w przedsiębiorstwach

# Analiza modelu

Analiza regresji jest jedną z najczęściej stosowanych technik do analizy danych wieloczynnikowych. Jej szeroka atrakcyjność i użyteczność wynika z konceptualnie 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 rzeczywistego[[3]](#footnote-3).

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

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

Po próbach zmiany zmiennej X4 na logarytmy i kwadraty tych zmiennych model nie wykazywał poprawy w parametrach, dlatego usunąłem zmienne i stworzyłem model z pięcioma zmiennymi objaśniającymi.

Tabela 2 Model nr. 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Zmienna***  ***Statystyka*** | **Współczynnik** | **Błąd standardowy** | **t-studenta** | **Wartość p** |
| **const** | 4401,34 | 239,435 | 18,38 | 4,38e-28 |
| **X1** | 7, 49358 | 4,02531 | 1,862 | 0,0671 |
| **X2** | 21,9976 | 7,18116 | 3,063 | 0,0031 |
| **X3** | 0,114623 | 0,018267 | 6,275 | 2,79e-08 |
| **X5** | 2,56776 | 0,544519 | 4,716 | 1,24e-05 |
| **X6** | −19,2915 | 16,7792 | −1,150 | 0,2543 |

Źródło: Własne opracowanie na podstawie wyników programu Gretl

Z tabeli 2, możemy odczytać, że największą wartość p-value ma X6 i wynosi ona 0.2543(wartość mniejsza niż poziom istotności o wartości 5%), co oznacza, że możemy wnioskować, że zmienna X6 jest statystycznie nieistotna i jest kandydatką do wyeliminowana w modelu, co potwierdza statystyka t-studenta wynosząca -1.150(wartość krytyczna jest równa około 2), więc możemy uznać, że zmienna ma statystycznie nieistotny wpływ na zmienną objaśnianą Y, czyli jest nieistotna dla modelu.

Z tabeli 2, odczytujemy, że zmienna X1 ma wartość p-value na poziomie 0.0671, czyli jest większa niż wartość poziomu istotności, co sugeruje, że zmienna może być statystycznie nieistotna, co potwierdza statystyka t-studenta równa 1.862, ponieważ statystyka ta jest mniejsza niż wartość krytyczna(około 2), co oznacza, że powinniśmy uznać tą zmienną jako kandydatkę do odrzucenia z modelu, bo może być statystycznie nieistotna.

Dla reszty zmiennych objaśniających w modelu, czyli X2, X3, X5 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 wnioskujemy, że zmienne maja statystycznie istotny wpływ na zmienną objaśnianą Y.

|  |  |  |  |
| --- | --- | --- | --- |
| **Średn.aryt.zm.zależnej** | 5466,751 | **Odch.stand. zm.zależnej** | 643,1325 |
| **Suma kwadratów reszt** | 7604333 | **Błąd standardowy reszt** | 336,8939 |
| **Wsp.determ.R2** | 0,744655 | **Skorygowany R2** | 0,725599 |
| **F(5,67)** | 39,07798 | **Wartość p dla testu F** | 1,38E-18 |
| **Logarytm wiarygodności** | −525,2951 | **Kryt.inform.Akaike’a** | 1062,59 |
| **Kryt. Bayes. Schwarza** | 1076,333 | **Kryt. Hannana - Quinna** | 1068,067 |

1. Mariusz Nyk*, Macroeconomic conditions of pay in Polish economy in 1990-2007,* Łódź 2009, s. 29 [↑](#footnote-ref-1)
2. Wojciech Stefan Zgliczyński, *Wynagrodzenia w Polsce,* Warszawa 2013, s. 1 [↑](#footnote-ref-2)
3. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, *Introduction to Linear Regression Analysis 5th Edition*, New Jersey 2012, s.15 [↑](#footnote-ref-3)