



Jerzy Witold Wiśniewski

Uniwersytet Mikołaja Kopernika
Wydział Nauk Ekonomicznych i Zarządzania
Katedra Ekonometrii i Statystyki
Jerzy.Wisniewski@umk.pl

KORELACJA I REGRESJA W DECYZJACH KADROWYCH

Streszczenie: W praktyce zarządzania kadrami rzadko wykorzystuje się metody statystyczne oraz ekonometryczne do podejmowania decyzji. W pracy zaprezentowano możliwości ograniczania ryzyka błędnych decyzji kadrowych, wynikające z gromadzenia i przetwarzania informacji statystycznych o efektach pracy oraz cechach osobistych zatrudnionych. Służą temu narzędzia korelacji i regresji. Zwłaszcza stosowanie wielorównaniowego modelu ekonometrycznego może przyczynić się do racjonalizacji decyzji kadrowych.

Słowa kluczowe: korelacja, regresja, decyzje kadrowe, mikroekonometria.

Wprowadzenie

W praktyce gospodarczej jednym z najważniejszych obszarów decyzyjnych jest racjonalne gospodarowanie zasobami pracy, w tym trafne podejmowanie decyzji kadrowych. Rzadko wykorzystywane są narzędzia statystyki i ekonometrii do wspomaganie decyzji kadrowych w przedsiębiorstwie. Wsparcie analityczne wymaga gromadzenia odpowiednich danych statystycznych, które nie zawsze są dostępne w podmiocie gospodarczym. W związku z tym punktem wyjścia jest rejestracja odpowiednich informacji statystycznych o wydajności pracy pracowników oraz o ich cechach osobistych, które mogą oddziaływać na efektywność pracy. Celem niniejszej pracy jest wskazanie podstawowych sposobów wykorzystania narzędzi statystyki i ekonometrii do ograniczenia ryzyka błędnych decyzji zatrudnieniowych.

1. Kryteria decyzji kadrowych

Decyzję o zatrudnieniu kandydata na określone stanowisko podejmuje się z reguły intuicyjnie. Bierze się przy tym pod uwagę niektóre cechy osobiste kandydata na wakujące stanowisko, zakładając przy tym, że właśnie one będą odgrywać znaczącą rolę w „dobrze wykonywanych” zadaniach na stanowisku pracy. Tymczasem w rzeczywistości pracodawca chce zatrudnić kogoś, kto będzie pracował wydajnie¹ oraz dobrze jakościowo. Dlatego też zbieranie informacji o wydajności i jakości pracy osób zatrudnionych, w skojarzeniu z ich cechami osobistymi, pozwoli skutecznie wykorzystać dane statystyczne w racjonalizacji decyzji zatrudnieniowych. Dopiero powiązanie wydajności i jakości pracy zatrudnionego z jego cechami osobistymi pozwoli w miarę jednoznacznie wypowiedzieć się o pożądanych cechach osobistych i ich poziomach, w związku z rekrutacją na wakujące stanowisko pracy.

Wśród cech osobistych robotnika można wymienić każdą jego charakterystykę, która odróżnia go od innych pracowników. Wymienić tu można przykładowo:

- płeć,
- wiek,
- wyuczony zawód,
- wykształcenie,
- stan cywilny,
- stan rodzinny,
- miejsce zamieszkania,
- posiadany majątek.

Rozważyć też można wiele innych cech osobistych, jeżeli istnieją przesłanki wskazujące na ważność tych charakterystyk w kształtowaniu efektywności pracy robotników.

Najmniej skomplikowane jest liczbowe wyrażenie cechy charakteryzującej się występowaniem tylko dwóch jej wariantów. Przykładem takiej właściwości jest płeć robotnika. Istnienie dwóch wariantów powoduje możliwość wykorzystania tylko jednej zmiennej zero-jedynkowej, zdefiniowanej np. następująco:

$$x_{i1} = \begin{cases} 1, & \text{kobieta,} \\ 0, & \text{mężczyzna.} \end{cases}$$

Wyraźnie trudniejsze jest zmierzenie właściwości tzw. cech niemierzalnych [Wiśniewski, 2011; 2013a], gdy możliwe jest wyodrębnienie wielu wariantów danej cechy. Dobrym przykładem może być wykształcenie robotnika. W grę

¹ W każdym przypadku, gdy możliwy jest pomiar efektów pracy zatrudnionego, warto gromadzić informacje statystyczne. Ułatwiają one bowiem przyszłe procesy decyzyjne.

wchodzą rozmaite możliwości, jak np. wykształcenie podstawowe, zasadnicze zawodowe, średnie, wyższe licencjackie, wyższe magisterskie. W takim przypadku nie istnieje potrzeba równoczesnego wyodrębnienia zmiennej zero-jedynkowej dla każdego z wyróżnionych wariantów wykształcenia. W modelu należy uwzględnić jedynie zmienne dla poziomów wykształcenia, które mogą różnicować indywidualną wydajność pracy na danym stanowisku. Zdefiniować możemy przykładowo następujące zmienne zero-jedynkowe, reprezentujące rozpatrywaną cechę:

$$x_{i2} = \begin{cases} 1, & \text{gdy robotnik posiada wyksz. podstawowe,} \\ 0, & \text{w innych przypadkach;} \end{cases}$$

$$x_{i3} = \begin{cases} 1, & \text{gdy robotnik posiada wyksz. średnie,} \\ 0, & \text{w innych przypadkach;} \end{cases}$$

$$x_{i4} = \begin{cases} 1, & \text{gdy robotnik posiada wyksz. minimum średnie,} \\ 0, & \text{w innych przypadkach.} \end{cases}$$

Niektóre cechy mierzalne, w sensie wyników należących do skali ilorazowej, warto przekształcić na zmienne zero-jedynkowe. Przykładem takiej cechy jest wiek pracownika, którego wpływ na wydajność pracy jest specyficzny w różnych etapach trwania życia [Wiśniewski, 2009, pkt 4.3]. Wyróżnić można fazę początkową wzrostu wydajności z przyrostem wieku, która ma charakter krzywej logistycznej. W kolejnej fazie następuje stabilizacja wydajności w dość długim przedziale wieku. Wreszcie po przekroczeniu określonej granicy wieku następuje stopniowy spadek wydajności indywidualnej. Każdy ze wskazanych przedziałów wieku można wyrazić odpowiednią zmienną zero-jedynkową, by precyzyjnie opisać wpływ wieku na indywidualną wydajność pracownika.

W dalszej części pracy zaprezentowane zostaną procedury mierzące do skonstruowania narzędzi ułatwiających decyzje zatrudnieniowe w grupie robotników. W związku z tym zostaną wykorzystane dane liczbowe, zamieszczone w tabeli 1, przetwarzane dalej odpowiednimi metodami statystyki i ekonometrii.

Tabela 1. Wydajność i jakość pracy robotników oraz niektóre ich cechy osobiste

Nr robotnika (i)	Jakość wykonanej produkcji (y_{1i})	Wydajność pracy (w sztukach) (y_{2i})	Wykształcenie (x_{i1})	Płeć (x_{i2})	Staż pracy (x_{i3})	Czas dojazdu do firmy (x_{i4})
1	2	3	4	5	6	7
1	0	92	1	0	3	20
2	1	93	1	0	4	25
3	1	95	1	0	7	30
4	1	96	1	0	8	15
5	1	98	1	0	5	15

cd. tabeli 1

1	2	3	4	5	6	7
6	1	99	1	0	7	10
7	1	102	1	0	6	15
8	1	103	1	1	11	25
9	1	104	1	0	12	20
10	1	105	1	0	4	15
11	0	107	0	0	2	10
12	1	108	0	1	5	15
13	0	110	0	0	3	20
14	0	115	0	0	1	20
15	1	118	0	0	4	25
16	1	119	1	0	6	30
17	1	120	1	0	5	35
18	1	120	1	1	8	40
19	1	122	1	0	9	20
20	1	123	1	0	11	15
21	1	124	1	0	15	20
22	1	125	1	0	12	10
23	1	122	1	0	16	20
24	1	122	1	0	14	15
25	1	121	1	0	9	30
26	0	119	0	1	3	10
27	1	120	0	0	4	15
28	0	118	0	0	2	20
29	0	117	0	0	1	20
30	1	116	0	0	5	15
Σ	23	3353	20	4	202	595

* Dane statystyczne mają charakter umowny. Zostały oparte na rzeczywistych obserwacjach procesów gospodarczych. Należy pamiętać o specyfice każdego stanowiska pracy i odmienności ważnych i nieistotnych cech osobistych, kształtujących wydajność pracy i jakość wykonania produktów.

Źródło: Dane umowne analogiczne do rzeczywistych*.

2. Korelacja w decyzjach kadrowych

Analiza korelacji indywidualnej wydajności pracy (y_{2i}) z cechami osobistymi robotników pozwala na wstępną ocenę ich roli na danym rodzaju stanowiska pracy. Również ustalenie korelacji oraz asocjacji zmiennej² reprezentującej jakość wykonania produkcji przez robotników (y_{1i}), z ich cechami osobistymi – pozwala wstępnie ocenić znaczenie tychże cech w kształtowaniu jakości wykonywania produktów.

² O korelacji będziemy mówić w przypadku, gdy cecha osobista jest wyrażona liczbami należącymi do skali ilorazowej lub interwałowej. Natomiast w przypadku cechy mierzonej za pomocą zmiennej zero-jedynkowej będziemy mówić o jej asocjacji z jakością wykonania produkcji.

W tabeli 1 zmienne zero-jedynkowe zostały zdefiniowane następująco³:

- y_{1i} – oznacza jakość wykonania produkcji przez i -tego robotnika ($i = 1, \dots, 30$), przyjmująca wartość 1, gdy robotnik mieści się w ustalonej normie jakości oraz wartość zero, jeśli wykonuje nadmiernie dużo produktów o niedostatecznej jakości,
- x_{i1} – reprezentuje rodzaj wykształcenia robotnika, przyjmując wartość 1, jeśli posiada on wykształcenie zasadnicze zawodowe oraz 0 w przypadkach pozostałych,
- x_{21} – reprezentuje płeć robotników, przyjmując wartość 1 dla kobiet oraz 0 dla mężczyzn.

Staż pracy (x_{i3}) wyraża liczbę pełnych przepracowanych lat, natomiast czas trwania dojazdu do pracy (x_{i4}) wyrażony jest w minutach.

Tabela 2. Współczynniki korelacji pomiędzy parami zmiennych

Zmienna	(y_{1i})	(y_{2i})	(x_{i1})	(x_{i2})	(x_{i3})	(x_{i4})
(y_{1i})	1	0,0152	0,6130	-0,0155	0,6155	0,2017
(y_{2i})	0,0152	1	-0,1859	0,0025	0,2913	0,0935
(x_{i1})	0,6130	-0,1859	1	-0,1387	0,6416	0,2723
(x_{i2})	-0,1550	0,0025	-0,1387	1	0,0016	0,1422
(x_{i3})	0,6155	0,2913	0,6416	0,0160	1	0,0371
(x_{i4})	0,2017	0,0935	0,2723	0,1422	0,0371	1

Źródło: Obliczenia własne za pomocą pakietu eViews.

Tabela 2 zawiera współczynniki korelacji (asocjacji) wydajności pracy oraz jakości produkcji z cechami osobistymi. Dostrzega się silną dodatnią asocjację⁴ jakości wykonania (y_{1i}), z wykształceniem (x_{i1}), gdyż $r_{y_{1i}, x_{i1}} = 0,613$. Wartość statystyki t-Studenta dla tej wartości współczynnika asocjacji wynosi $t = 4,106$, co oznacza znaczącą jego wartość na poziomie istotności $\gamma < 0,001$. Wynik ten sugeruje, że robotnicy z wykształceniem zasadniczym zawodowym charakteryzują się większą częstością dobrej jakościowo produkcji w porównaniu z pozostałymi.

Dostrzega się również silną dodatnią korelację jakości wykonanej produkcji ze stażem pracy robotników, gdyż $r_{y_{1i}, x_{i3}} = 0,6155$. Empiryczna statystyka t-Studenta dla tej wartości współczynnika korelacji wynosi $t = 4,132$, co oznacza istotne statystycznie skorelowanie na poziomie istotności $\gamma < 0,001$. Oznacza to, że z wydłużaniem stażu pracy robotnik wykonuje lepszą jakościowo produkcję.

³ Fundamentem rozważań o pomiarze jest praca [Stevens, 1946].

⁴ Por. rozważania o asocjacji cech jakościowych w pracy [Wiśniewski, 2013a, chapter 3]. Pierwsze szerokie rozważania na ten temat znaleźć można w pracy [Wiśniewski, 1986, rozdział 3].

Indywidualna wydajność pracy (y_{2i}) charakteryzuje się słabym skorelowaniem z cechami osobistymi robotników. Najwyższy co do modułu współczynnik korelacji ($r_{y_{2i}, x_{i3}} = 0,2913$) generuje statystykę t-Studenta o wartości $t = 1,684$. Skorelowanie to można uznać za istotne z ryzykiem błędu pierwszego rodzaju, przekraczającym nieznacznie 10%.

Uzyskane wyniki obliczeń siły i kierunku skorelowania wydajności pracy oraz jakości wykonywanej produkcji wskazują ważne, czyli pożądane cechy osobiste. Ujawniają też cechy, które są obojętne z punktu widzenia efektywności wykonywanych czynności zawodowych.

3. Modele regresji we wspomaganii decyzji

Bardziej precyzyjnych informacji o związkach wydajności pracy oraz jakości wykonywania produkcji przez robotników z ich cechami osobistymi dostarczają modele regresji. W naszym przypadku przedmiotem rozważań będzie model ekonometryczny, składający się z dwóch następujących równań:

$$\begin{aligned} y_{1i} &= \alpha_{10} + \beta_{12}y_{2i} + \alpha_{11}x_{i1} + \alpha_{12}x_{i2} + \alpha_{13}x_{i3} + \alpha_{14}x_{i4} + \eta_{1i}, \\ y_{2i} &= \alpha_{20} + \alpha_{21}x_{i1} + \alpha_{22}x_{i2} + \alpha_{23}x_{i3} + \alpha_{24}x_{i4} + \eta_{2i}, \end{aligned} \quad (1)$$

gdzie:

α_{1i}, α_{2i} – parametry strukturalne przy zmiennych z góry ustalonych modelu ($i = 0, 1, 2, 3, 4$),

β_{12} – parametr strukturalny przy zmiennej łącznie współzależnej modelu,

η_{1i}, η_{2i} – składniki losowe równań modelu.

Dostrzega się rekurencyjność modelu. Wynika to z założenia o możliwości negatywnego oddziaływania wysokiej wydajności pracy na jakość wykonywanej produkcji. Zdarza się, że praca w nadmiernym tempie skutkuje powstawaniem zbyt dużej liczby tzw. braków. Model empiryczny pozwoli na weryfikację tej hipotezy.

Pierwsze z równań należy do klasy tzw. modelu Goldbergera. Wykorzystujemy liniową funkcję prawdopodobieństwa do opisu mechanizmu zmienności jakości produkcji robotników, w zależności od ich wydajności indywidualnej oraz cech osobistych. Posłużymy się do tego danymi statystycznymi zamiesz-

czonymi w tabeli 1. W wyniku estymacji parametrów modelu Goldbergera otrzymujemy następujące równanie empiryczne⁵:

$$\hat{y}_{li} = 0,505 - 0,003 y_{2i} + 0,244 x_{i1} + 0,006 x_{i2} + 0,047 x_{i3} + 0,007 x_{i4}, \quad (2)$$

(0,620)
(0,410)
(1,044)
(0,033)
(1,830)
(0,685)

$$R_1^2 = 0,472.$$

Powyższe równanie empiryczne zawiera wiele zmiennych objaśniających, które są statystycznie nieistotne. Zwraca uwagę ujemny znak przy ocenie parametru związanego z wydajnością pracy. Sygnalizuje on, że może pojawić się negatywny wpływ wysokiej wydajności pracy na jakość wykonywanych wyrobów. W obecnym stanie rzeczy oddziaływanie to jest statystycznie nieistotne.

Po wyeliminowaniu zmiennych, które okazały się statystycznie nieistotne w kolejnych iteracjach estymacyjnych, powstało następujące równanie empiryczne⁶:

$$\hat{y}_{li} = 0,284 + 0,333 x_{i1} + 0,039 x_{3i}, \quad (3)$$

(2,399)
(2,010)
(2,048)

gdzie:

$$R_1^2 = 0,460, \quad R_{1*}^2 = 0,5122.$$

Z równania (3) wynika, że robotnik posiadający wykształcenie zawodowe częściej wykonuje wyroby spełniające normy jakości w porównaniu z innymi pracownikami. Ponadto wydłużanie stażu pracy skutkuje lepszą jakościowo produkcją⁷. W związku z tym przy ewentualnym wakacie na stanowisku robot-

⁵ Parametry modelu Goldbergera oszacowano za pomocą klasycznej metody najmniejszych kwadratów. Metoda ta daje wystarczająco dokładne wyniki z punktu widzenia praktycznego wykorzystania modelu empirycznego. W swojej monografii A.S. Goldberger [1972] rekomenduje stosowanie do estymacji parametrów równania (4.15) ważonej metody najmniejszych kwadratów, z wagami opartymi na resztach otrzymanych z zastosowania klasycznej metody najmniejszych kwadratów do szacowania parametrów tegoż równania. Wazona metoda najmniejszych kwadratów daje szacunki parametrów o wyższej efektywności. Z praktycznego punktu widzenia, jeśli za pomocą klasycznej metody najmniejszych kwadratów uzyskane wyniki pozwalają wskazać zmienne statystycznie istotne, to dla celów decyzyjnych można poprzestać na takich obliczeniach. Pod ocenami parametrów równania zamieszczone są empiryczne statystyki t-Studenta.

⁶ Wartość współczynnika $R_{1*}^2 = 0,5122$ została obliczona po uwzględnieniu w obliczeniach granicznych wartości 0 lub 1, gdy teoretyczna wartość zmiennej objaśnianej dla danej obserwacji wykroczyła poza przedział [0,1].

⁷ Należy pamiętać, że nie odkrywamy tu uniwersalnej reguły. W każdym przypadku przedsiębiorstwa oraz w zależności od stanowiska pracy, mogą pojawiać się inne uwarunkowania jakości wykonania produktów.

niczym należy kierować się wartościami zmiennych x_{i1} oraz x_{i3} , charakteryzującymi kandydata na wolne stanowisko pracy.

Rozważmy teraz równanie empiryczne, opisujące mechanizm indywidualnej wydajności pracy na badanym typie stanowisk pracy:

$$\hat{y}_{2i} = 101,7 - 17,813 x_{i1} - 4,721 x_{i2} + 2,027 x_{i3} + 0,433 x_{i4}, \quad (4)$$

(19,091) (3,645) (0,950) (3,826) (1,816)

gdzie:

$$R_2^2 = 0,407.$$

Równanie (4) zawiera również zmienne objaśniające statystycznie nieistotne. Po redukcji otrzymujemy następujące równanie empiryczne:

$$\hat{y}_{2i} = 108,8 - 14,147 x_{i1} + 1,785 x_{i3}, \quad (5)$$

(33,007) (3,065) (3,376)

gdzie:

$$R_2^2 = 0,321.$$

Równanie (5) informuje, że jedynie dwie spośród cech osobowych robotników mają wpływ na ich indywidualną wydajność pracy. Pracownicy z wykształceniem zasadniczym zawodowym osiągnęli wydajność przeciętnie o ponad 14 sztuk niższą od robotników z innym wykształceniem. Przyrost stażu pracy robotnika o 1 rok skutkował wzrostem jego wydajności pracy średnio o blisko 1,8 sztuk. Pozostałe cechy osobiste okazały się neutralne.

Dysponując równaniem (5) możemy zrationalizować decyzje zatrudnieniowe, w sytuacji wakatów etatowych. Należy preferować kandydatów do pracy o długim stażu przy tego rodzaju zadaniach produkcyjnych. Możliwe jest ustalenie potencjalnej wydajności pracy kandydatów, co umożliwi uporządkowanie ich według potencjału wydajności⁸. Równanie empiryczne (5) może być też pomocne przy ustalaniu norm pracy na rozpatrywanym rodzaju stanowisk⁹.

⁸ Szczegóły procedury wyboru znaleźć można w pracy [Wiśniewski, 2009, pkt 4.3].

⁹ Por. w tej sprawie rozważania Cieślak [1965, rozdz. 6].

Podsumowanie

Wykorzystanie metod statystyki i ekonometrii w zarządzaniu kadrami spotyka się w praktyce gospodarczej niezwykle rzadko. Wynika to przede wszystkim z braku wiedzy właścicieli przedsiębiorstw oraz ich zarządców o korzyściach, możliwych do osiągnięcia z tego rodzaju wspomaganie decyzyjnego. Często przyczyną tego stanu rzeczy jest słabe wykształcenie ekonomistów, którzy trafiają do przedsiębiorstw.

Warto zwrócić uwagę na dokładność opisu zmienności jakości wykonywanej produkcji i wydajności pracy w równaniach empirycznych, mierzoną wartością współczynnika determinacji R^2 . W równaniach mikromodelu ekonometrycznego uzyskiwanie wartości $0,6 \leq R^2 \leq 0,8$ uznać należy za sukces badawczy. W równaniach opisujących indywidualną wydajność pracy lub jakość wykonania produkcji standardem są wartości w przedziale $0,25 \leq R^2 \leq 0,5$. Można oczekiwać, że im bardziej uzbrojony w technikę jest robotnik, tym mniejszą rolę w jego efektywności i jakości odgrywają cechy osobiste. Warto jednak znać rolę cech osobistych, ze wskazaniem istotnych statystycznie nawet wówczas, gdy kreują one tylko 25% zmienności wydajności pracy lub jakości wykonywania produktów.

Literatura

- Cieślak M. (1965), *Statystyczne problemy normowania pracy*, PWE, Warszawa.
- Goldberger A.S. (1972), *Teoria ekonometrii*, PWE, Warszawa.
- Stevens S.S. (1946), *On the Theory of Scales Measurement*, „Science”, Vol. 103, No. 2684.
- Wiśniewski J.W. (1986), *Ekonometryczne badanie zjawisk jakościowych. Studium metodologiczne*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Wiśniewski J.W. (2009), *Mikroekonometria*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń.
- Wiśniewski J.W. (2011), *Dilemmas of Economic Measurements in Weak Scales*, Folia Oeconomica Stetinensia, No. 10(18) 2011/2, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin, s. 50-59.
- Wiśniewski J.W. (2013): *Correlation and Regression of Economic Qualitative Features*, LAP LAMBERT Academic Publishing, Saarbrücken.

**CORRELATION AND REGRESSION
IN HUMAN RESOURCE DECISIONS**

Summary: In human resource management practice, statistical and econometric methods are rarely used for decision-making. This work presents the possible ways of reducing the risk of making incorrect staff decisions, resultant from collecting and processing the statistical data on the effects of work and the employees' personal qualities. Correlation and regression tools serve this purpose. Application of a multi-equations econometric model can especially contribute to rationalization of human resource decision-making.

Keywords: correlation, regression, human resource decision-making, microeconometrics.