

Jerzy CHOWANIEC, Robert KOBIERZEWSKI,  
Witold LEGUT

PROGNOZOWANIE WYDAJNOŚCI W POKŁADACH CIENKICH  
W OPARCIU O WYKORZYSTANIE REGRESJI KROKOWEJ

**Streszczenie.** W artykule przedstawiono metodę analizy regresji krokowej pozwalającej na wybór jak najmniejszej liczby nie skorelowanych ze sobą zmiennych objaśniających, a istotnie wpływających na kształtowanie się wydajności kopalni.

WSTĘP

Posiadanie wiarygodnych informacji o przebiegu procesu produkcyjnego jest konieczne dla przeprowadzenia oceny jego aktualnej działalności oraz dla podejmowania decyzji na okresy przyszłe. Podstawowy warunek zapewniający otrzymanie wiarygodnych ocen szacowanej wydajności został spełniony, gdyż otrzymane oceny oparte są na bogatym materiale empirycznym.

Zastosowanie analizy regresji krokowej doprowadziło do otrzymania zależności funkcyjnych opisujących w sposób zadowalający kształtowanie się wydajności węglowej w pokładach cienkich od zmiennych opisujących wydajność.

W artykule przeprowadzono analizę przyczynowo-skutkową oraz ocenę ilościową wydajności w pokładach cienkich, gdzie grubość waha się w granicach  $h \in < 0,50-1,00 \text{ m} >$  przy nachyleniu  $\alpha \in < 0^\circ-12^\circ >$ .

Pokład wybierany jest systemem ścianowym podłużnym na zawał z zastosowaniem obudowy stalowo-członowej (stojaki typu SHC-40 i stropnice SCGB-960/1500), urabianie - za pomocą struga węglowego typu SWS-4.

Do oceny ilościowej wartości oczekiwanej wydajności rzeczywistej  $Q$  wykorzystano dane za lata 1977-1979, znajdujące się w banku informacji i systemu IOS-8 i IOS-8M oraz przeprowadzono weryfikację ich przydatności metodą analizy regresji krokowej.

W ramach przeprowadzonej analizy zweryfikowano przydatność następujących zmiennych opisujących:

- $\alpha$  - kąt nachylenia pokładu - stopnie,
- $h$  - wysokość przodka (ściany) - dcm,
- $i$  - ilość współczynników korygujących normę zasadniczą,
- $t$  - czas (pracy) przebywania w przodku - min,
- $l$  - długość ściany - dcm.

Wartości współczynników korelacji poszczególnych zmiennych opisujących zaprezentowano w postaci macierzy korelacji

$$R = \begin{matrix} & \alpha & h & i & t & l & y \\ \alpha & \begin{bmatrix} 1 & - & - & - & - & - \\ 0,37 & 1 & - & - & - & - \\ 0,32 & -0,14 & 1 & - & - & - \\ 0,28 & -0,39 & 0,22 & 1 & - & - \\ 0,31 & 0,28 & -0,40 & -0,28 & 1 & - \\ 0,30 & 0,34 & 0,23 & -0,09 & 0,38 & 1 \end{bmatrix} \end{matrix}$$

#### ANALIZA WARIANCJI KROK-1

Wartości maksymalne statystyki V, wyznaczone z próby, wynoszą:

$$V_{\max} \{V_1, V_2, V_3, V_4, V_5\} = 0.12 = V_1$$

Do modelu regresji wielokrotnej wprowadzamy zmienną  $\alpha$ , jako najbardziej skorelowaną ze zmienną opisującą. Równanie regresji na tym etapie otrzyma postać:

$$\hat{Q} = -12308 + 152\alpha + \xi \dots \quad (1)$$

(42)

gdzie:

$$n = 28, \quad z = 12\%, \quad \xi = 15\%.$$

Wynik analizy wariancji dla równania (1) przedstawia tablica 1.

Tablica 1

Źródło	Stopień swobody	Suma kwadratów	Średni kwadrat	F obl.	F teoret. F (1.48, 0.05)
Ogółem	35	266318455		20.3	4.20
Regresja	1	160353140	23507296		
Reszta	28	105965315	1204031		

Jak z tablicy 1 wynika,  $F_{obl.} > F_{teor.}$ , co świadczy, że brak podstaw do odrzucenia zmiennej  $\alpha$ . Zmienna opisująca i zmienna  $\alpha$  są skorelowane na poziomie istotności  $q = 0,05$ . Zmienna  $\alpha$  wyjaśnia kształtowanie się wydajności  $\hat{Q}$  w 55%, która pozwala nam twierdzić, że równanie (model) (1) w sposób zadowalający opisuje kształtowanie się wielkości wydajności.

ANALIZA WARIANCJI KROK-2

Dla wprowadzenia nowej zmiennej korzystamy z tej samej metody co w kroku-1 i tak:

maksymalna wartość statystyki:

$$V_{\max} V_2, V_3, V_4, V_5 = 0,23 = V_2$$

Nową zmienną wprowadzoną do równania regresji jest zmienna h.

Model regresji na tym etapie ma postać:

$$\hat{Q} = 163082 - \underset{(58)}{371} + \underset{(31)}{121} h + \dots \quad (2)$$

Wyniki analizy wariancji dla modelu (2) przytoczono w tabelicy 2.

Tablica 2

Źródło	Stopień swobody	Suma kwadratów	Średni kwadrat	F obl.	F teor. F (2,26, 0,05)
Ogółem	380	266318331		19,04	3,37
Regresja	2	160316010	21507296		
Reszta	26	106002321	1104031		

Ponieważ  $F_{obl.} > F_{teor.}$ , nie ma podstaw do odrzucenia zmiennej h. Wartość F testu częściowego dla eliminacji zmiennej wprowadzonej do równania (2) w poprzednim kroku wynosi:  $F_{cz} = 404$ .

$$F_{cz} F(2,26, 0,05) = 3,37,$$

a więc nie eliminuje się zmiennej  $\alpha$  wprowadzonej w poprzednim kroku, co świadczy, że zmienna  $\alpha$  i h nie są silnie ze sobą skorelowane. Współczynnik korelacji  $\hat{r}(\alpha, h) = 0,37$  jest stosunkowo mały, co oznacza, że zmienne opisujące w modelu (2) w sposób istotny wyjaśniają kształtowanie się wydajności. Procent wyjaśnionych zmienności  $\hat{R}^2 = 62\%$ , czyli zmienne  $\alpha$  i h w sposób istotny opisują kształtowanie się wielkości wydajności.

ANALIZA WARIANCJI KROK-3

Maksymalna wartość statystyk  $V_{\max}$  wynosi:

$$V_{\max} \{V_3, V_4, V_5\} = 0,122 = V_4$$

Równanie regresji przejmie postać:

$$\hat{Q} = -7753 - 563a + 175h + 5i + \xi \dots \quad (3)$$

(124,2) (62,1) (3,2)

Wyniki analizy powyższej wariancji przedstawia tablica 3.

Tablica 3

Źródło	Stopień swobody	Suma kwadratów	Sredni kwadrat	F obl.	F teor. F (3,23, 0,05)
Ogółem	31	266315456,3		24,276	3,03
Regresja	3	160521889,6	53507296,5		
Reszta	23	105796566,6	2204095,1		

$F \text{ obl.} > F \text{ teor.}$ , nie ma więc podstaw do odrzucenia zmiennej opisującej  $i$ . Model (3) jest statystycznie istotny na poziomie istotności  $\alpha = 0,05$ . Procent wyjaśnionych zmienności  $\hat{R}^2 = 65\%$ . W tym przypadku możemy również pozytywnie zapatrywać się na rolę modelu w opisywaniu kształtowania się wydajności.

Wartość testu częściowego  $F_{\text{cz}} = 42$  i jest większa od wartości  $F(3,48, 0,05) = 3,03$ , co implikuje, że nie ruguje z modelu poprzednio wprowadzonej zmiennej niezależnej  $h$ . Współczynnik korelacji  $r$  dla zmiennych opisujących  $h$ ,  $i$  wynosi:

$$\hat{r}(h,i) = 0,14$$

#### ANALIZA WARIANCJI KROK-4

W czwartym kroku analizy wariancji wprowadzamy zmienną niezależną  $l$ , której wartość funkcji testowej jest większa od  $F \text{ teor.}$  dla stopni swobody 4 i 22.

Zmienna  $l$  wyjaśnia kształtowanie się wydajności w 75%. Model ten posiada jedynie 25% niewyjaśnionych zmienności.

#### ANALIZA WARIANCJI KROK-5

W piątym kroku analizy wariancji i ostatnim w naszym przykładzie wprowadzamy zmienną  $t$ .

Model regresji przyjmuje postać:

$$\hat{Q} = 13765 - \underset{(66)}{673\alpha} + \underset{(22)}{179 h} + \underset{(1,2)}{5 i} - \underset{(2)}{8 l} - \underset{(8)}{17 t} + \xi \dots \quad (4)$$

Wyniki wariancji dla równania (4) przedstawia tablica 4.

Tablica 4

Zródło	Stopień swobody	Suma kwadratów	Średni kwadrat	F obl.	F teor. F (5,21, 0,05)
Ogółem	24	266318456,3		26,917	2,62
Regresja	5	198479175,8	39695835,1		
Reszta	21	678392804,6	1474766,9		

Ponieważ  $F \text{ obl.} > F \text{ teor.}$  brak podstaw do odrzucenia wprowadzonej zmiennej  $t$ . Procent wyjaśnionych zmienności  $\hat{R}^2 = 87\%$ . Wartość testu częściowego  $F_{cz} = 14$  i jest większa od wartości  $F(5,21, 0,05) = 2,62$ , co decyduje, że ruguje się z modelu poprzednio wprowadzonej zmiennej niezależnej  $l$ . Współczynnik korelacji dla zmiennych opisujących  $l$  i  $t$  wynosi  $\hat{r}(l,t) = 0,28$ . Model (4) jest statystycznie istotny na poziomie istotności  $\alpha = 0,05$ . Procent wyjaśnionych zmienności wynosi  $\hat{R}^2 = 87\%$ .

Model posiada jedynie 13% nie wyjaśnionych zmienności, co upewnia nas, że model (4) w sposób poprawny opisuje kształtowanie się wydajności.

Przedziały ufności na poziomie 87% dla parametrów modelu wynoszą:

$$\begin{aligned} 13435 < \beta_0 < 14095 \\ -802 < \beta_1 < -544 \\ 135 < \beta_2 < 223 \\ 3,0 < \beta_3 < 7,0 \\ -11 < \beta_5 < -4 \\ -30 < \beta_4 < -3 \end{aligned}$$

Przedziały te nie zawierają zera, co świadczy o poprawności opisywanej rzeczywistości przez wartość parametrów.

Brak jest podstaw do wyeliminowania jakiegokolwiek zmiennej wprowadzonej do modelu.

Przedziały ufności na poziomie 87% dla prawdziwej wartości oczekiwanej wydajności przedstawia tablica 5.

Tablica 5

Estymator wartości oczekiwanej rzeczywistej wydajności Q	Wartość oczekiwanej rzeczywistej wydajności	Granica wartości oczekiwanej rzeczywistej wydajności	
		Dolna	Górna
$\hat{Q} (\alpha=8, h=77, \bar{i}=1,039, \bar{l}=1776, \bar{t}=377)$	5967	5637	6297
$\hat{Q} (\alpha_{\min}=3,0, \bar{h}=77, \bar{i}=1,039, \bar{l}=1776, \bar{t}=377)$	9889	9069	10710
$\hat{Q} (\alpha_{\max}=12, \bar{h}=77, \bar{i}=1,039, \bar{l}=1776, \bar{t}=377)$	3831	3305	4356
$\hat{Q} (h_{\min}=63, \bar{\alpha}=8, \bar{i}=1,039, \bar{l}=1776, \bar{t}=377)$	3417	2711	4123
$\hat{Q} (h_{\max}=96, \bar{\alpha}=8, \bar{i}=1,039, \bar{l}=1776, \bar{t}=377)$	9338	8449	10226
$\hat{Q} (i_{\min}=0,833, \bar{\alpha}=8, \bar{h}=77, \bar{l}=1776, \bar{t}=377)$	4917	4378	5455
$\hat{Q} (i_{\max}=1,311, \bar{\alpha}=8, \bar{h}=77, \bar{l}=1776, \bar{t}=377)$	7349	6699	7999
$\hat{Q} (l_{\min}=1473, \bar{\alpha}=8, \bar{h}=77, \bar{i}=1,039, \bar{t}=377)$	8434	7362	9506
$\hat{Q} (l_{\max}=1928, \bar{\alpha}=8, \bar{h}=77, \bar{i}=1,039, \bar{t}=377)$	4738	4132	5344
$\hat{Q} (t_{\min}=340, \bar{\alpha}=8, \bar{h}=77, \bar{l}=1776, \bar{i}=1,039)$	6612	6015	7208
$\hat{Q} (t_{\max}=420, \bar{\alpha}=8, \bar{h}=77, \bar{l}=1776, \bar{i}=1,039)$	5218	4554	5883

## ZAKOŃCZENIE

1. Zaprezentowana metoda regresji krokowej, zastosowana do analizy wydajności w pokładach cienkich, wykazała, że zmienne opisujące zastosowane w powyższym przykładzie wyjaśniają 87% zmienności.
2. Procent nie wyjaśnionych zmienności, który wynosi 13%, sugeruje potrzebę wprowadzenia bardziej złożonych funkcji oraz zweryfikowania wymagań dotyczących jednorodności badanych zbiorów danych statystycznych.
3. W celu lepszej estymacji zgodności funkcji regresji konieczne jest wprowadzenie do równań regresji dodatkowych zmiennych objaśniających, takich jak:
  - współczynnik mechanicznego wybierania,
  - współczynnik niezawodności maszyn i urządzeń,
  - współczynnik twardości węgla,
  - warunki stropowe i spągowe.
4. Na podstawie otrzymanych zależności można wyrazić pogląd, że zastosowana metoda doboru zmiennych objaśniających do modelu regresji jest słuszna i pozwala na uzyskanie poprawnej zgodności funkcji dla pokładów cienkich przy danym wyposażeniu i konfiguracji.
5. Omówiony przykład wskazuje na słuszność przyjętej zasady stosowania liniowych modeli addytywnych, gdyż ilustrują one w sposób przejrzysty związki przyczynowo-skutkowe.

## LITERATURA

- [1] DRAPER N.R., SMITH H.: Analiza regresji stosowana. PWN, Warszawa 1973.
- [2] PLATT Cz.: Problemy rachunku prawdopodobieństwa i statystyki matematycznej. Wyd. III. PWN, Warszawa 1978.
- [3] Instrukcja wypełniania dokumentów źródłowych systemu IOS Ministerstwa Górnictwa.
- [4] Bank Informacji Ministerstwa Górnictwa system IOS-8 i IOS-8M.

Recenzent: prof. dr hab. inż. Włodzimierz Sitko

Wpłynęło do Redakcji 3.03.1982 r.

ПРОГНОЗИРОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ В ТОНКИХ ПЛАСТАХ  
НА ОСНОВАНИИ ИСПОЛЬЗОВАНИИ ШАГОВОЙ РЕГРЕССИИ

Р е з ю м е

В статье описывается метод анализа шаговой регрессии позволяющей выбрать наименьшее количество некоррелированных друг с другом объясняющих переменных, оказывающих значительное влияние на производительность шахты.

FORECASTING THE OUTPUT IN THIN COAL BEDS BASED ON UTILIZATION  
OF STEPWISE REGRESSION

S u m m a r y

The article presents the method of stepwise regression analysis which makes it possible to choose the smallest number of non-correlated variables and substantially influencing the output of a coal-mine.