

## **PREDYKCJA ZA POMOCĄ METODY NAJMNIEJSZYCH KWADRATÓW**

**Streszczenie:** W artykule autor we wstępie przedstawia rys historyczny a następnie zapoznaje czytelnika z metodą najmniejszych kwadratów. Na początku zaprezentowano problem wyboru postaci analitycznej modelu, uznając według literatury za priorytet wybór jak najmniej skomplikowaną postać tego modelu. Następnie przedstawiono sposób obliczania procesu predykcji za pomocą modelu analitycznego tej metody, korzystając z ustalonych wzorów. Nie wykonano jednak obliczeń dla przypadku danych statystycznych przedstawionych w tabeli z uwagi na dość skomplikowany sposób ich wyliczenia. Przedstawiono także sposób obliczenia współczynników równania regresji za pomocą oprogramowania Excel. Z uwagi na ograniczone możliwości tego oprogramowania, nie wykonano postaci graficznej uzyskanego równania regresji. W dalszej części zaprezentowano uniwersalne narzędzie statystyczne jakim jest program Statistica. Dzięki niemu przedstawiono metodę interpretacji przedstawionych danych statystycznych za pomocą metody najmniejszych kwadratów w celu uzyskania predykcji tych danych. Pokazano dwa przykładowe wykresy jako rozwiązanie przedstawionego problemu predykcji. Na końcu przedstawiono zwięzłe wnioski dotyczące predykcji za pomocą metody najmniejszych kwadratów.

Słowa kluczowe: predykcja, metoda najmniejszych kwadratów, model, regresja, zmienna.

## **PREDICTION USING THE METHOD OF THE SMALLEST SQUARES**

**Abstract:** In the introduction of the article the author presents a historical outline and acquaints the reader with the method of the smallest squares. The problem of choosing the model's analytic figure is shown at the beginning and according to literature, the priority of the choice as the least complicated figure of this model was adopted. The way of calculation of prediction process was introduced using the analytic model of this method, using settled formulas. However, calculations were not made for cases of statistical data introduced in the table due to difficult way of their calculation. The way of the coefficients calculation of the regress equation using Excel software was also outlined. Because of the limited possibilities of this software, the graphic figure of the received regress equation was not presented. In the further part, the universal statistical tool i.e. the Statistica programme was shown. Thanks to this programme, the method of statistical data interpretations was introduced using the method of the smallest squares in order to obtain the prediction of these data. There were showed two example graphs as the solution introduced of the prediction problem. Concise conclusions relating to prediction using the method of the smallest squares were included at the end of the article.

Keywords: prediction, method of the smallest squares, model, regress, variable.

### **1. Wstęp**

W fizyce, technice oraz w innych naukach ścisłych spotykamy się zazwyczaj ze ścisłymi zależnościami występującymi między określonymi wielkościami. Zależności te można wyrazić odpowiednimi wzorami matematycznymi. Oprócz ścisłych związków spotyka się w praktyce również takie, które przy istnieniu pewnej ogólnej tendencji zmian jednej wielkości w zależności od drugiej, powstają jako mniej lub bardziej uchwytne odchylenia od tejże tendencji.

W przypadku istnienia tego rodzaju związku statystycznego poszczególnych wielkości, przed badaczem pojawia się zadanie znalezienia ogólnej tendencji właściwej danemu procesowi lub zjawisku, przy czym często trzeba wyrazić tę tendencję wzorem matematycznym. Do rozwiązywania tego typu zadań w statystyce często stosuje się uniwersalną metodę określoną jako metoda najmniejszych kwadratów.

Zgodnie z [2] metoda najmniejszych kwadratów została wprowadzona przez Legendre'a w 1805 roku. Gauss, który twierdził, że używał jej od 1794 roku, wsparł ją w 1809 założeniem o normalnym rozkładzie błędów. Od niego pochodzi też nazwa tej metody: Methode der Kleinsten Fehlerquadrate.

Początkowo była stosowana do obliczeń geodezyjnych, stała się podstawą teorii pomiarów błędów, używanej początkowo w astronomii i geodezji, obecnie we wszystkich pomiarach fizycznych. Legła też u podstaw statystyki.

Leżący u podstaw tej metody postulat Legendre'a nie wynika z żadnej ścisłej matematycznej teorii. Jakkolwiek poczyniono wiele prób, by udowodnić jego słuszność i uzasadnić stosowanie, to wszystkie one się nie powiodły. Pozostaje nadal tym, czym był od początku – założeniem matematyka o genialnej intuicji.

Wszystkich czytelników tego artykułu, chcących poszerzyć swoją wiedzę z zakresu procesu predykcji oraz rodzajów stosowanych metod predykcji odsyłam do wcześniejszych moich artykułów opublikowanych w latach 2012÷2013 w Problemach Techniki Uzbrojenia.

## 2. Wybór postaci analitycznej modelu

Zgodnie z [4] ogólnie biorąc, wybór postaci analitycznej modelu nie podlega żadnym, określonym z góry i jednocześnie uznanym przez wszystkich regułom. Teoria nie dostarcza zwykle decydujących argumentów za przyjęciem określonej postaci modelu, a analiza danych empirycznych rzadko kiedy pozwala zdecydowanie opowiedzieć się za przyjęciem jednej a odrzuceniem innych postaci analitycznych. Dotyczy to w szczególności modeli wielowymiarowych.

Dokonując wyboru analitycznej postaci modelu należy wziąć pod uwagę problem estymacji parametrów. Otóż nie opłaca się wybierać zawitych formalnie postaci modelu, gdyż wiąże się to z komplikacjami natury estymacyjnej. A skoro i tak nie potrafimy znaleźć funkcji, która w miarę dokładnie obrazowałaby powiązania między zmienną niezależną a wyróżnionymi zmiennymi zależnymi, to nie ma sensu narażać się na stosowanie skomplikowanych procedur estymacyjnych.

Zgodnie z [1] w przypadku opracowywania danych statystycznych metodą najmniejszych kwadratów stosuje się zazwyczaj przy modelowaniu pewną formą standardową, a mianowicie poszukiwaną zależność przedstawia się przy pomocy wielomianu określonego stopnia. Tak więc rozpatrując zależność pewnej wielkości  $y$  od innej  $x$  możemy mieć wielomian o postaci:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \quad (1)$$

gdzie:  $a_0, a_1, a_2$  itd. są współczynnikami wyliczonymi według danych rzeczywistych.

Problem polegający na tym, ile wyrazów tego szeregu należy uwzględnić w określonym przypadku rozwiązuje się na podstawie ogólnych wyobrażeń o charakterze badanego zjawiska, a także w oparciu o wyniki materiałów rzeczywistych.

W prostszym przypadku, gdy przyrost  $\Delta y$  jest w przybliżeniu proporcjonalny do przyrostu  $\Delta x$ , można przestać na wielomianie obejmującym tylko dwa pierwsze wyrazy, czyli można uznać związek liniowy:

$$y = y_0 + \Delta y_1 \Delta x \quad (2)$$

Jeśli są podstawy aby oczekiwać że wartość  $y$  będzie progresywnie wzrastać lub maleć wraz ze wzrostem  $x$  (czyli że w pewnym zakresie wystąpi maksimum lub minimum wartości  $y$ ), należy włączyć także trzeci wyraz  $\Delta y_2 \Delta x^2$ , w czego powstanie wielomian drugiego stopnia:

$$y = y_0 + \Delta y_1 \Delta x + \Delta y_2 \Delta x^2 \quad (3)$$

Możliwe są przypadki, gdy równanie drugiego stopnia okaże się niewystarczające i zajdzie potrzeba wprowadzenia czwartego wyrazu, czyli potrzeba zastosowania wielomianu trzeciego stopnia:

$$y = y_0 + \Delta y_1 \Delta x + \Delta y_2 \Delta x^2 + \Delta y_3 \Delta x^3 \quad (4)$$

Niekiedy bardziej dogodnie jest budowanie wielomianu, w którym zmienną niezależną jest odwrotność  $x$ . Równanie liniowe względem  $1/x$  przedstawia sobą hiperbolę względem  $x$ .

$$y = y_0 + \Delta y_1 \frac{1}{x} \quad (5)$$

Osobliwością tej krzywej jest jej asymptotyczne dążenie do linii poziomej. Możliwe jest również stosowanie wielomianów wyższych rzędów z wyrazem  $1/x^2$  np.:

$$y = y_0 + \Delta y_1 \frac{1}{x} + \Delta y_2 \frac{1}{x^2} \quad (6)$$

Zmienna w wielomianie asymptotycznym może być nie tylko wartość odwrotności wielomianu asymptotycznego  $1/x$  lub  $1/x^2$ , ale także parametry. Można zbudować

W przypadku konstrukcji modelu matematycznego zjawiska obejmującego związki statystyczne dwóch, trzech lub większej ilości cech, stosuje się zazwyczaj wielomiany określonego stopnia względem odpowiednich zmiennych. Jeżeli z góry wiadomo, że dwa zmienniki nie są ze sobą skorelowane, wówczas można je pominąć, a liniowa zależność  $y$  od dwóch zmiennych  $x_1$  i  $x_2$  wyrazić się równaniem:

$$y = \Delta y_1 x_1 + \Delta y_2 x_2 \quad (7)$$

Równanie typu  $Y = y_0 + \Delta y_1 x + \Delta y_2 x^2 + \Delta y_3 x^3 + \dots$  ustalone metodą najmniejszych kwadratów nazywa się równaniem regresji, a jego współczynniki – współczynnikami regresji. Użycie słowa regresja nie jest być może najbardziej trafne, ale termin ten rozpowszechnił się i jego zmiana spowodowałaby wiele trudności.

### 3. Podstawowa zasada metody najmniejszych kwadratów

Poznanie metody najmniejszych kwadratów przedstawiono poniżej na przykładzie prostego przypadku zależności pewnej wielkości  $y$  od innej wielkości  $x$  zwanej zmienną niezależną. Zadanie polega na wyszukaniu równania w postaci określonego wielomianu względem  $x$ . Weźmy wielomian drugiego stopnia:

$$y = y_0 + \Delta y_1 x + \Delta y_2 x^2 \quad (8)$$

Zadanie sprowadza się do odnalezienia wartości liczbowych występujących tu współczynników  $y_0, \Delta y_1$  i  $\Delta y_2$ . Można je rozwiązać posiadając dwa równoległe szeregi obserwacji:

$$x_1, x_2, x_3, \dots, x_n \quad (9)$$

$$a_0, a_1, a_2, \dots, a_n \quad (10)$$

W przypadku obecności związku statystycznego pomiędzy  $X$  i  $Y$  nie będzie możliwości znalezienia takich wielkości  $a_0, a_1$  i  $a_2$  przy których, wszystkie punkty rzeczywiste znalazłyby się na krzywej określonej równaniem paraboli. Inaczej mówiąc — dla dowolnych

wielkości współczynników  $a_0, a_1$  i  $a_2$  niektóre empiryczne wielkości  $Y$  będą wykazywać,

odchylenia od  $Y$  wyliczonego przy pomocy równania. Najlepszym rozwiązaniem zadania jest oczywiście dobór takich wartości  $a_0, a_1$  i  $a_2$ , przy których układ odchyłeń  $Y$  będzie zbliżony do minimum. W celu wyrażenia tego ostatniego postulatu (minimalizacji odchyłeń) wzorem matematycznym, należy ustalić zasady pomiaru stopnia niezgodności rzeczywistych

wartości  $Y$  z wyliczonymi  $Y$ . U podstawy metody najmniejszych kwadratów leży zasada,

zgodnie z którą stopień ustalonej niezgodności jest mierzony sumą kwadratów odchyłeń  $Y$  od  $Y$ , dlatego rozwiązanie zadania polega na znalezieniu takich  $a_0, a_1$  i  $a_2$ , przy których:

$$\sum (Y - \hat{Y})^2 = \min \quad (11)$$

Stąd właśnie pochodzi określenie „metoda najmniejszych kwadratów”. Rozwiązanie tego

zadania nie przedstawia większych trudności. Podstawmy do sumy kwadratów  $\sum (Y - \hat{Y})^2$  wartości  $\hat{Y} = a_0 + a_1 X + a_2 X^2$ , gdzie  $a_0, a_1$  i  $a_2$  są na razie nieznanne. W rezultacie

otrzymujemy funkcję:

$$Q(a_0, a_1, a_2) = \sum (Y - a_0 - a_1 X - a_2 X^2)^2 \quad (12)$$

W celu znalezienia wartości  $a_0, a_1$  i  $a_2$ , przy których osiąga swoje minimum, należy przyrównać poszczególne pochodne tych zmiennych do zera i rozwiązać otrzymany w ten sposób układ równań:

$$\begin{aligned} \frac{\partial Q}{\partial a_0} &= -2 \sum (Y - a_0 - a_1 X - a_2 X^2) = 0 \\ \frac{\partial Q}{\partial a_1} &= -2 \sum (Y - a_0 - a_1 X - a_2 X^2) X = 0 \\ \frac{\partial Q}{\partial a_2} &= -2 \sum (Y - a_0 - a_1 X - a_2 X^2) X^2 = 0 \end{aligned} \quad (13)$$

Po wykonaniu prostych przekształceń otrzymujemy układ trzech równań, zwany układem równań normalnych:

$$\begin{aligned} a_0 \sum 1 + a_1 \sum X + a_2 \sum X^2 &= \sum Y \\ a_0 \sum X + a_1 \sum X^2 + a_2 \sum X^3 &= \sum XY \\ a_0 \sum X^2 + a_1 \sum X^3 + a_2 \sum X^4 &= \sum X^2 Y \end{aligned} \quad (14)$$

gdzie:  $n$  — liczba obserwacji.

Chociaż metoda najmniejszych kwadratów daje gwarancje, że układ odchyłeń  $Y$  od  $\hat{Y}$  jest liczbowa wartość sumy kwadratów jako taka. W przypadku niewielkiej wartości  $\sum (Y - \hat{Y})^2$

$Y$  wyliczone według równania regresji jest bliskie rzeczywistego  $Y$  i co za tym idzie,

tendencja scharakteryzowana równaniem regresji dość ściśle odzwierciedla zmiany

rzeczywistych  $\hat{Y}$ . W przypadku większej wartości  $\sum(\hat{Y} - Y)^2$  (oczywiście z taką samą ilością obserwacji) rzeczywiste  $\hat{Y}$  różnią się bardziej od teoretycznego  $Y$  i w bardzo niskim stopniu odpowiadają ustalonej prawidłowości. Wyliczenie sumy kwadratów odchyłeń nie stanowi żadnej trudności. Podstawiając do równania regresji zaobserwowane wartości zmiennych

niezależnych  $X$ , możemy wyliczyć wszystkie teoretyczne wartości  $Y$ . Następnie, znajdując różnice  $\hat{Y} - Y$ , wyprowadzając ich kwadraty oraz sumując je możemy określić sumę

kwadratów odchyłeń. Istnieje jednak sposób pozwalający dość szybko wykonać wyliczenie

$\sum(Y - \hat{Y})^2$ , pomijając wyliczenia pojedynczych odchyłeń.

#### 4. Excel jako narzędzie do predykcji

Metodą najmniejszych kwadratów możemy dzisiaj rozwiązać każde zadanie bardzo szybko, posługując się oprogramowaniem Excel, które znajduje się w pakiecie Office Windows. Rozwiązanie klasycznej metody najmniejszych kwadratów, czyli tzw. prostej regresji liniowej obsługuje funkcja „REGLINP”. Funkcja ta oblicza linię prostą na podstawie istniejących danych. Oblicza ona statystykę dla linii prostej, która najlepiej pasuje do zadanych danych źródłowych. Mając dane wyniki pewnych badań w postaci określonej tablicy danych, możemy za pomocą kilku przycisków otrzymać wynik tej regresji tzn. wartości współczynników  $a$  i  $b$  prostej  $Y = aX + b$  gdzie współczynnik  $a$  określa jej prostą, a  $b$  jest jej wy截em. Dodatkowo funkcja ta zwraca nam tablicę wartości między innymi: podaje nam gotowe

wyniki dokładności wyznaczania wartości współczynników  $a$  i  $b$  czyli wielkości błędów wartości  $a$  i  $b$  oraz np. resztkową sumę kwadratów.

Oprogramowanie Excel umożliwia także obliczenie krzywej wykładniczej dzięki funkcji „REGEXP” oraz obliczenie wielokrotnej regresji liniowej, stosując funkcję „REGLINW”, która dokonuje predykcji wartości, pasujących do prostej trendu liniowego. Funkcję tę stosuje się w przypadku, kiedy istnieje zależność prostoliniowa pomiędzy każdą ze zmiennych niezależnych  $X_1, X_2, X_3$ , a zmienną zależną  $Y$ . Funkcja zwraca tablicę wartości regresji liniowej określonych zmiennych niezależnych względem zmiennej zależnej. Można także dopasować krzywą wielomianową za pomocą tego typu regresji wobec tej samej zmiennej podniesionej do różnych potęg np. do kwadratu lub do sześciastu.

Istnieje także możliwość obliczenia przewidywanego np. wzrostu wykładniczego, używając określonych wyników danych stosując funkcję „REGEXPW”, która to funkcja zwraca wartości zmiennej zależnej dla serii nowych wartości określonych na przebiegu czasu. Funkcja ta dokonuje predykcji wartości, które krzywej wykładniczej.

Kolejną funkcją, która służy do predykcji wartości jest funkcja „REGLINX”, która oblicza lub dokonuje predykcji wartości na podstawie dostępnych wartości istniejących. Wartość predykcji danej wartości jest to wartość  $\hat{Y}$  dla danej wartości  $X$ . Znaną są dwie wartości  $X_1$  i  $X_2$ , a nowa wartość jest obliczana za pomocą regresji liniowej najmniejszych kwadratów. Ta funkcja może być przydatna w predykcji np. poziomu wymagań dotyczących zapasów składowanych środków bojowych.

#### 5. Praktyczny sposób obliczenia metody najmniejszych kwadratów

Do obliczenia praktycznego metody najmniejszych kwadratów posłużono się danymi źródłowymi przedstawionymi w tabeli 1. Zawiera ona dane dotyczące liczby partii zbadanych zapalników typu DC-1 do roku 2010 włącznie w rozbiciu na decyzje dodatnie i ujemne oraz czas składowania tych zapalników wyrażony w latach, a także frakcję decyzji dodatnich uzyskiwanych podczas badań diagnostycznych. Do decyzji dodatnich zaliczono decyzje „B5” i „B3”, które to decyzje oznaczają bezpieczną predykcję przydatności badanych zapalników w kolejnych pięciu lub trzech latach ich składowania. Do decyzji ujemnych zaliczono wszystkie pozostałe decyzje diagnostyczne. Zapalniki tego typu są stosowane w granatach nasadkowych typu PGN-60 oraz KGN.

Zgodnie z [2], jeśli pomiędzy dwiema wielkościami występuje zależność liniowa (a niemal zawsze może tak być, wystarczy tylko odpowiednio dobrać osie wykresu) regresja liniowa jest prostą metodą wyznaczania parametrów dopasowanej prostej. Uzyskane parametry dopasowania mogą następnie służyć do wyznaczania szukanej wielkości.

Parametry prostej określonej równaniem  $y = a + bx$  można wyznaczyć przy użyciu zależności:

$$a = \frac{n \sum y - \sum y \sum x}{n^2 - (\sum x)^2} \quad (15)$$

$$b = \frac{1}{n} \frac{\sum yx - (\sum y \sum x)}{\sum x^2 - (\sum x)^2 / n} \quad (16)$$

gdzie:  $n$  - liczba zbadanych partii;

$x_i$  - czas składowania;

$y_i$  - frakcja decyzji dodatnich.

Błędy wyznaczonych wartości  $a$  i  $b$  określone są wzorami:

$$S_a = \frac{n \sum y^2 - (\sum y)^2}{\sqrt{\sum x^2 - (\sum x)^2 / n}} \quad (17)$$

$$S_b = \frac{1}{n} \sqrt{\frac{(n-2) [\sum yx - (\sum y \sum x)]^2}{\sum x^2 - (\sum x)^2 / n}} \quad (18)$$

W celu obliczenia tych wielkości należy sporządzić dodatkową tabelę zawierającą obliczenia cząstkowe, co oczywiście stwarza możliwości pomyłki. Zamiast pracować liczyć sumy i kwadraty jako obliczenia cząstkowe można posłużyć się funkcją „REGLINP” w Excel-u i program automatycznie oblicza te wielkości. W wyniku obliczeń otrzymaliśmy:  $a = -0,327$  i  $b = 54,347$  oraz dokładności wyznaczenia współczynników  $a$  i  $b$  wynoszące

$S_a = 0,506$  i  $S_b = 13,572$ . Otrzymaliśmy więc funkcję postaci:

$$y = -0,327x + 54,347 \quad (19)$$

Jeśli chcemy precyzyjnie przedstawić tę zależność, czyli funkcję frakcji decyzji dodatnich zapalników typu DC-1 w zależności od czasu składowania, to otrzymamy linię prostą, którą obliczona linia regresji. Przecięcie osi OY w punkcie  $a$  i przebiegać będzie według opadającej. Taka postać krzywej regresji oznacza, że frakcja decyzji dodatnich jest w zasadzie stała dla zapalników DC-1 w całym okresie badawczym. Wartość ujemna współczynnika nachylenia prostej, zwanego współczynnikiem regresji, wynika prawdopodobnie z wieku składowanych zapalników. Zachodzące w zapalnikach DC-1 procesy starzenia mogą powodować obniżenie ich trwałości, czyli mogą wpływać na zmniejszenie ich bezpieczeństwa i niezawodności działania.

Tabela 1. Zapalniki typu DC-1

Czas składowania (lata)	Decyzje dodatnie	Decyzje ujemne	Liczba partii zbadanych	Frakcja decyzji dodatnich [%]
10	7	3	10	70,00

11	7	4	11	63,64
12	2	6	8	25,00
13	2	6	8	25,00
14	3	6	9	33,33



15	4	4	8	50,00
16	7	8	15	46,67
17	9	4	13	69,23
18	5	4	9	55,56
19	6	4	10	60,00
20	2	2	4	50,00
21	2	5	7	28,57
22	4	6	10	40,00
23	6	5	11	54,55
24	4	4	8	50,00
25	3	4	7	42,86
26	4	8	12	33,33
27	4	2	6	66,67
28	3	3	6	50,00
29	2	8	10	20,00
30	2	4	6	33,33
31	2	5	7	28,57
32	4	2	6	66,67
33	4	1	5	80,00
34	1	0	1	100,00
36	2	1	3	66,67
37	1	4	5	20,00
38	0	2	2	0,00
39	0	1	1	0,00
40	1	0	1	100,00
41	0	1	1	0,00
Suma	103	117	220	

## 6. Program Statistica jako narzędzie do predykcji

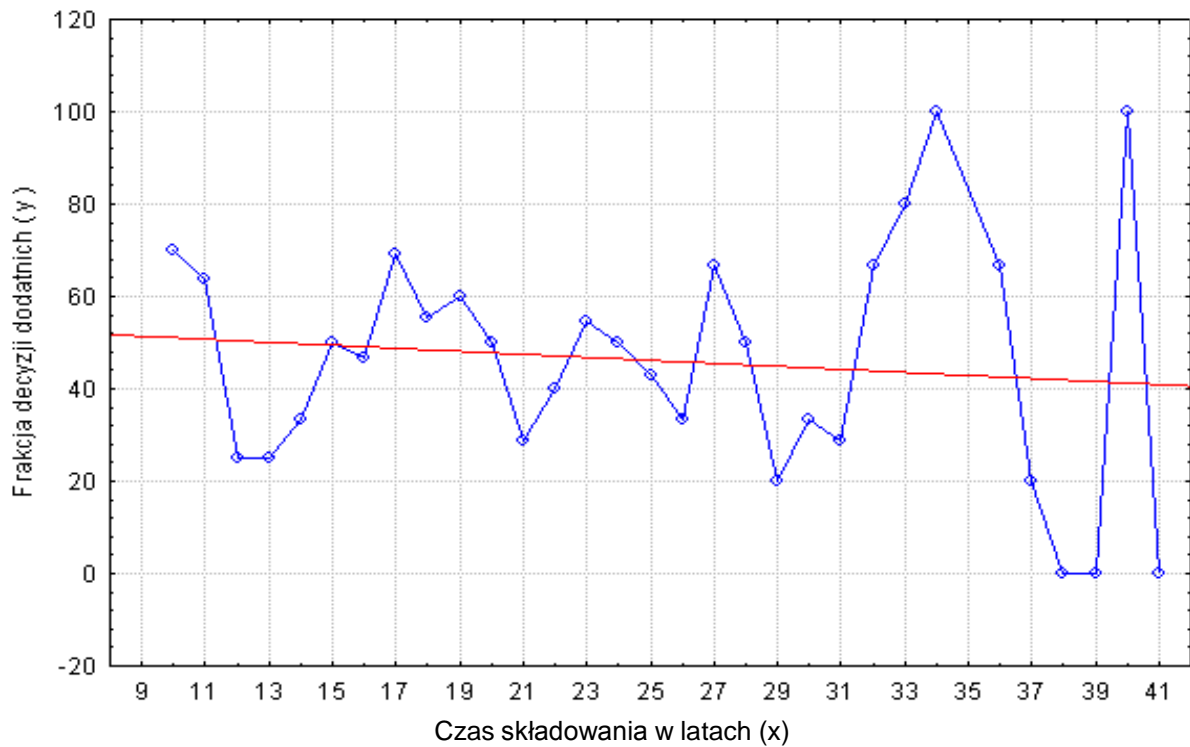
Program Statistica jest typowym oprogramowaniem matematycznym, który posiada bardzo rozbudowany element statystyczny. Narzędzie to umożliwia obliczenie każdego rodzaju regresji bez względu na rodzaj danych empirycznych.

Po utworzeniu w oprogramowaniu Statistica projektu zawierającego dane z tabeli 1, w menu programu wybieramy nakładkę „Wykresy”, w której wybieramy „Wykresy 2W” i następnie „Wykresy liniowe (zmienne)”. Otrzymamy wykres zależności frakcji decyzji dodatnich w zależności od czasu składowania zapalników DC-1. Na wykresie dodatkowo zaznaczona jest linia regresji, a w opisie wykresu podana jest wartość tej linii regresji. Jak widać z rys. 1, narzędzie to jest lepsze od Excel-a pod tym względem, ponieważ wszystkie potrzebne informacje są automatycznie obliczane i pokazywane na wykresie.

Możemy również wykonać wykres rozrzutu zadanych danych empirycznych z zaznaczonym przedziałem ufności na poziomie zadanym przez badacza czyli np. 95%. Wykres ten został przedstawiony na rys. 2. Celem procedury regresyjnej jest dopasowanie linii do zadanych punktów empirycznych. Program tak dobiera równanie tej linii, że suma kwadratów odległości punktów na wykresie rozrzutu od linii regresji będzie minimalna. Dzięki takiej

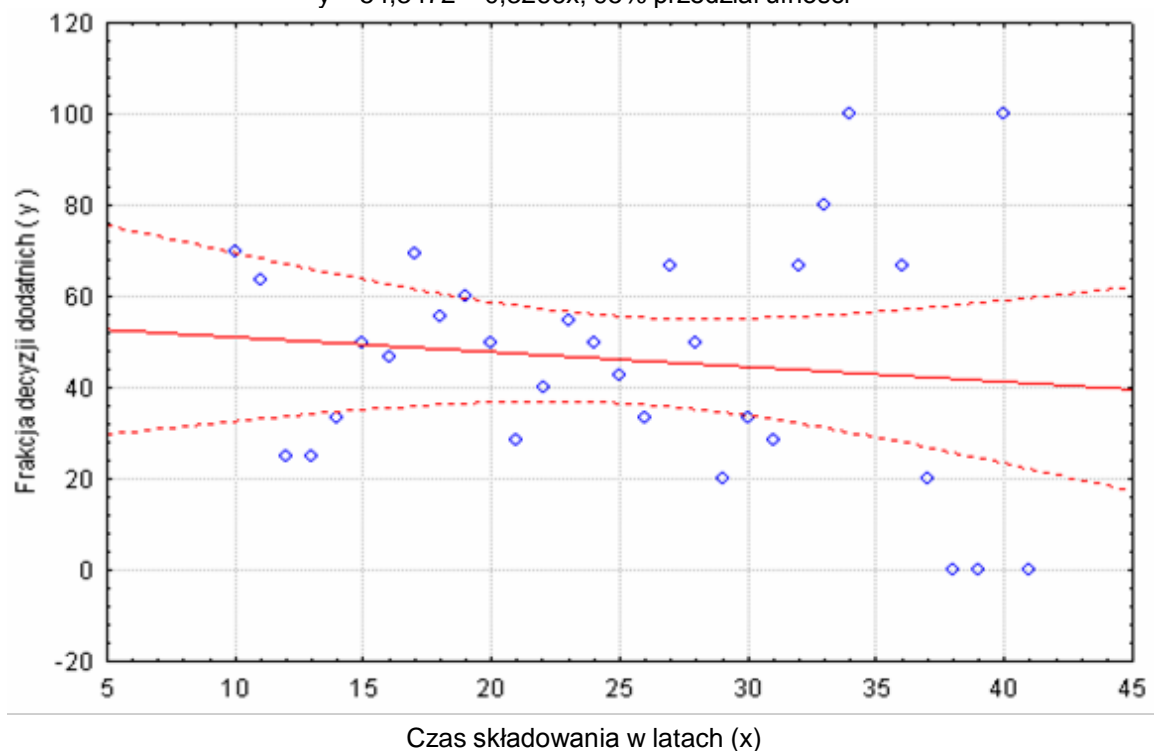
Frakcja decyzji dodatnich – zapalniki DC-1

Linia regresji  $y = 54,3472 - 0,3266x$



**Rys. 1. Frakcja decyzji dodatnich dla zapalników DC-1**

Wykres rozrzutu – zapalnik DC-1  
 $y = 54,3472 - 0,3266x$ ; 95% przedział ufności



**Rys. 2. Wykres rozrzutu dla zapalników DC-1 z określonym przedziałem ufności**

własności ta ogólna procedura jest nazywana oceną najmniejszych kwadratów w oprogramowaniu Statistica.

Narzędzie Statistica zostało przedstawione w artykule jako sposób obliczenia linii regresji metodą najmniejszych kwadratów. Przykład ten pokazał nam najprostszą metodę interpretacji graficznej oraz analitycznej uzyskiwania liniowej regresji prostej. Szerokie spektrum tego programu umożliwia znalezienie bardzo wielu funkcji, zależności rozkładów, jakie są poszukiwane przez badaczy w celu znalezienia odpowiedzi na podstawowe pytania dotyczące charakteru zmian określonego zjawiska fizycznego.

## 7. Wnioski

Ogólnie można powiedzieć, że metoda najmniejszych kwadratów polega w pierwszym stadium na konstrukcji uproszczonego modelu badanego zjawiska czyli na odpowiednim doborze badanych czynników i ustaleniu odpowiadających im wskaźników ilościowych. Drugie stadium to budowa wzoru matematycznego, wyrażającego najbardziej charakterystyczne cechy zjawiska, czyli matematyczne jego modelowanie. Matematyczny model zjawiska jest zazwyczaj budowany w oparciu o posiadane przez badacza pewne ogólne wyobrażenia dotyczące charakteru badanego zjawiska. Taki model porównuje się następnie z posiadanymi informacjami statystycznymi oraz przy pomocy specjalnych kryteriów ustala się ich przydatność zarówno w ujęciu ogólnym, jak też w poszczególnych fragmentach.

W przypadku opracowywania danych statystycznych, metodę najmniejszych kwadratów stosuje się zazwyczaj przy modelowaniu pewną formą standardową, a mianowicie poszukiwaną zależność przedstawia się przy pomocy wielomianu określonego stopnia. Problem polegający na tym, który stopień tego wielomianu należy zastosować w określonym przypadku rozwiązuje się na podstawie ogólnych wyobrażeń o charakterze badanego zjawiska, a także w oparciu o wyniki rzeczywiste badań.

W artykule scharakteryzowano ogólną zasadę stosowania metody najmniejszych kwadratów. Na przykładzie wyników badań zapalników granatnikowych typu DC-1 przedstawiono sposób jej zastosowania dzięki oprogramowaniom Excel oraz Statistica. Oczywiście wybór zastosowanego do obliczeń oprogramowania należy do badacza, jednakże możliwości użytkowe oraz uzyskiwane informacje statystyczne, bez potrzeby wykonywania dodatkowych żmudnych obliczeń, wskazują na oprogramowanie Statistica, które jest narzędziem pracy umożliwiającym otrzymywanie pełnych informacji statystycznych. Wadą takiego rozwiązania jest oczywiście koszt, jaki należy ponieść, aby takie oprogramowanie posiadać. Dużo łatwiejszym rozwiązaniem jest stosowanie standardowego oprogramowania Excel, które jest załączone w pakiecie Office, a który to pakiet można nabyć za znacznie niższą cenę. W tym drugim przypadku nie uzyska się jednak pełnych informacji statystycznych oraz nie można mieć wszystkich postaci graficznych rozwiązywanego problemu badawczego.

## Literatura

- [1] W. Pieriegunow – *Metoda najmniejszych kwadratów i jej zastosowanie* – Państwowe Wydawnictwo Ekonomiczne, Warszawa 1967 r.
- [2] Wikipedia – *Wolna Encyklopedia* 2013 r.
- [3] Statistica 9 – Statsoft Polska 2009 r. – oprogramowanie komputerowe
- [4] Witkowska A., Witkowski M. – *Podstawy ekonometrii* – Państwowe Wydawnictwo Naukowe, Warszawa 2005 r.
- [5] A. Luszniwicz, T. Słaby – *Statystyka z pakietem komputerowym Statistica* – Wydawnictwo C.H. Beck, Warszawa 2003 r.

- [6] S. Kot, J. Jakubowski, A. Sokołowski – *Statystyka* – Wydawnictwo Difin, Warszawa 2011 r.
- [7] A. Stanisław – *Przystępny kurs statystyki* – Statsoft Polska, Kraków 2007 r.
- [8] M. Rabiej – *Statystyka z programem Statistica* – Wydawnictwo Helion, Gliwice 2012 r.