Maciej **Gonet**

ZROZUMIEĆ EXCELA

OBLICZENIA I WYKRESY

Helion

Redaktor prowadzący: Małgorzata Kulik

Projekt okładki: Studio Gravite / Olsztyn
Obarek, Pokoński, Pazdrijowski, Zaprucki

Grafika na okładce została wykorzystana za zgodą Shutterstock.com

Wydawnictwo HELION
ul. Kościuszki 1c, 44-100 GLIWICE
tel. 32 231 22 19, 32 230 98 63
e-mail: *helion@helion.pl*
WWW: *http://helion.pl* (księgarnia internetowa, katalog książek)

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
http://helion.pl/user/opinie/zrexow
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

ISBN: 978-83-283-5693-1

Printed in Poland.

Spis treści

Rozdział 7.

Regresja liniowa jednowymiarowa

Zakres zastosowania regresji liniowej

Eksperymenty, których rezultatem są duże zbiory danych liczbowych reprezentujących zależności pomiędzy zmiennymi wejściowymi a wynikami, staramy się zwykle opracować w postaci wykresów lub równań. Postać równań może wynikać z przesłanek teoretycznych, lecz najczęściej stosuje się do opisu typowe, stosunkowo proste równania funkcji liniowej, wielomianowej, potęgowej, wykładniczej lub logarytmicznej. Najczęściej stosowanym kryterium dopasowania równania do danych doświadczalnych jest **minimalna suma kwadratów** różnic między wartościami wynikowymi zmierzonymi a wyznaczonymi z równania. Kryterium to stanowi podstawę metody regresji, której szczególnym przypadkiem jest regresja liniowa.

Regresja liniowa nie ogranicza się do zależności liniowej między wynikami a danymi wejściowymi, lecz obejmuje wszystkie równania, które są liniowe ze względu na występujące w nich współczynniki. Jeżeli równanie opisujące wyniki eksperymentu przedstawimy w postaci $\boldsymbol{Y} = f(\boldsymbol{X}, \boldsymbol{A})$, gdzie $\boldsymbol{Y}$ reprezentuje wielkości wynikowe, $\boldsymbol{X}$ — zmienne wejściowe, a $\boldsymbol{A}$ — współczynniki, to równanie jest liniowe ze względu na współczynniki, jeśli pochodna $\partial Y / \partial A$ nie zależy od $\boldsymbol{A}$. Do tej kategorii zaliczymy więc zarówno równania typu:

$$y = a_0+a_1x_1+a_2x_2+a_3x_3$$

w których wielkość wynikowa zależy od trzech niezależnych zmiennych wejściowych, jak i równania typu:

$$y = a_0+a_1x+a_2x^2+a_3x^{-1}+a_4\ln x$$

gdzie zależność od jednej zmiennej wejściowej przedstawiona jest za pomocą równania nieliniowego ze względu na x, ale liniowego ze względu na współczynniki. Możliwe są też równania typu:

$$y = a_0+a_1x_1+a_2x_2+a_{12}x_1x_2+a'_{12}x_1/x_2$$

reprezentujące zależność od dwóch zmiennych wejściowych.

Istnieje także kategoria równań, które wprawdzie nie są liniowe, ale dają się przekształcić do postaci liniowej przez odpowiednią zamianę zmiennych. Do tej grupy zaliczamy na przykład równanie wykładnicze:

$$y = ae^{bx}$$

Po obustronnym zlogarytmowaniu przyjmie ono postać:

$$\ln y = \ln a + bx$$

która jest już liniowa ze względu na współczynniki. Pamiętać jednak należy, że zastosowanie metody regresji do tak przekształconego równania prowadzi do innych wartości współczynników, niż gdyby brać pod uwagę równanie w pierwotnej postaci. Czym innym jest suma kwadratów odchyleń liczona dla y, a czym innym ta suma liczona dla $\ln y$. W tym drugim przypadku relatywnie zmniejsza się waga punktów pomiarowych, w których wartości y były większe, a zwiększa się waga punktów z małymi wartościami y. Niemniej metodę linearyzacji równań stosuje się dość powszechnie, gdy tylko jest to możliwe.

Podobnie daje się linearyzować zależność potęgowa:

$$y = ax^b$$

Po obustronnym zlogarytmowaniu otrzymujemy:

$$\ln y = \ln a + b \ln x$$

czyli równanie liniowe, ale — podobnie jak poprzednio — nie dla y, lecz dla $\ln y$.

Kolejny typ równania podlegający linearyzacji to dwuparametrowa funkcja homograficzna:

$$y = \frac{ax}{x+b}$$

Po inwersji równanie przekształca się do postaci:

$$\frac{1}{y} = \frac{1}{a} + \frac{b}{a} \cdot \frac{1}{x}$$

która reprezentuje liniową formę zależności $1/y = f(1/x)$.

Inne możliwe sposoby linearyzacji tego równania to:

$$\frac{x}{y} = \frac{b}{a} + \frac{1}{a}x\,, \quad y = a - b\frac{y}{x} \quad \text{lub} \quad xy = ax - by$$

Pierwszy reprezentuje liniową formę zależności $x/y = f(x)$, drugi $y = f(y/x)$, a trzeci $xy = f(x, y)$ bez wyrazu wolnego. Do każdej z tych zależności można dołączyć zależność odwrotną, która też jest liniowa.

Zastosowanie każdej z form linearyzacji prowadzi do innych wartości współczynników a i b.

Linearyzacji podlega również funkcja Gaussa, ważna w rachunku prawdopodobieństwa i statystyce matematycznej, ale także w obróbce danych doświadczalnych, gdyż wykresy otrzymywane w różnych technikach badawczych przyjmują często kształty zbliżone do krzywej Gaussa. Krzywą Gaussa można opisać równaniem z trzema parametrami, A, B i m:

$$y = Ae^{-B(x-m)^2}$$

Po zlogarytmowaniu otrzymamy zależność:

$$\ln y = \ln A - B(x-m)^2 = (\ln A - Bm^2) + 2Bmx - Bx^2$$

Jest to równanie kwadratowe ze względu na x. Można do niego zastosować regresję liniową wielowymiarową.

To najbardziej typowe przykłady linearyzacji. W podobny sposób linearyzuje się bardziej złożone funkcje, na przykład zależność potęgowo-wykładniczą:

$$y = ax^b e^{cx}$$

lub $$y = \frac{ax^n}{x+b}$$

pod warunkiem że n jest znaną stałą, a nie współczynnikiem do wyznaczenia.

Istnieje jednak wiele zależności ważnych z punktu widzenia obróbki danych doświadczalnych, które nie dają się linearyzować i wymagają nieco innego podejścia, zwanego regresją nieliniową. Przykłady takich zależności to:

$$y = a(1-e^{-bx})$$

$$y = \frac{ax^{1/2}}{1+bx^{1/2}} + cx$$

W tym rozdziale ograniczę się do omówienia przypadków regresji liniowej jednej zmiennej, które da się sprowadzić do najprostszej zależności:

$$y = a+bx$$

W następnym rozdziale omówione będą pozostałe przypadki regresji liniowej, a w kolejnym — regresja nieliniowa.

Ocena statystyczna danych pomiarowych i jakości dopasowania równania regresji

Charakterystyka zbiorów danych pomiarowych

Poszukując zależności między wielkościami ***Y*** i ***X***, zgromadziliśmy dane pomiarowe w postaci zbioru N par wartości (x_i, y_i). Zbiór ten można scharakteryzować za pomocą różnych wielkości zagregowanych, to znaczy obliczonych dla wszystkich punktów zbioru łącznie. Wielkości te mogą być obliczane osobno dla każdej zmiennej lub mogą obejmować obie zmienne równocześnie.

Dla każdej zmiennej można obliczyć wartość średnią (średnia arytmetyczna) według wzorów:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

W tym tekście kreska nad symbolem oznacza wartość średnią. W Excelu służy do tego funkcja `ŚREDNIA`, która może przyjmować dowolną liczbę argumentów w postaci stałych, adresów komórek i zakresów oraz nazw odnoszących się do tych obiektów.

Do określenia liczby N można użyć funkcji `ILE.LICZB`, przydatnej przy dużej liczbie pomiarów i wtedy, gdy ich wyniki są rozmieszczone w kilku miejscach w arkuszu.

Jedną z miar rozproszenia danych jest ich odchylenie kwadratowe, czyli suma kwadratów odchyleń zmiennej od wartości średniej:

$$SS_x = \sum_{i=1}^{N}(x_i - \overline{x})^2 \qquad SS_y = \sum_{i=1}^{N}(y_i - \overline{y})^2$$

W Excelu wielkość tę oblicza funkcja `ODCH.KWADRATOWE`.

Ponieważ odchylenie kwadratowe rośnie ze wzrostem liczby punktów, to aby można było porównywać różne zbiory danych, często się je standaryzuje, dzieląc przez liczbę danych. Otrzymuje się wielkość zwaną **wariancją**, która funkcjonuje w dwóch postaciach — **wariancji próbki**, gdy w mianowniku jest N–1, i **wariancji populacji**, gdy w mianowniku jest N:

$$S_{xpr} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1} \qquad S_{ypr} = \frac{\sum_{i=1}^{N}(y_i - \overline{y})^2}{N-1}$$

$$S_{xpop} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N} \qquad S_{ypop} = \frac{\sum_{i=1}^{N}(y_i - \overline{y})^2}{N}$$

W Excelu obliczenia te wykonują funkcje `WARIANCJA` i `WARIANCJA.POPUL`. Począwszy od Excela 2010, zaleca się używanie nowych nazw: `WARIANCJA.PRÓBKI` i `WARIANCJA.POP`, jakkolwiek stare nazwy nadal mogą być stosowane.

Pierwiastek z wariancji ma ten sam wymiar co wielkość mierzona i może być miarą odchylenia wartości zmierzonej w punkcie od średniej. Analogicznie do wariancji stosowane są dwie wielkości — **odchylenie standardowe próbki**, gdy w mianowniku jest N–1, i **odchylenie standardowe populacji**, gdy w mianowniku jest N:

$$s_{xpr} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} \qquad s_{ypr} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \overline{y})^2}{N-1}}$$

$$s_{xpop} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N}} \qquad s_{ypop} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \overline{y})^2}{N}}$$

W Excelu używa się funkcji `ODCH.STANDARDOWE` i `ODCH.STANDARD.POPUL`. Począwszy od Excela 2010, zaleca się stosowanie nowych nazw: `ODCH.STANDARD.PRÓBKI` i `ODCH.STAND.POPUL`, jakkolwiek stare nazwy nadal mogą być używane.

Rzadziej stosowaną miarą odchylenia zbioru danych od wartości średniej jest **odchylenie średnie**, liczone na podstawie sumy wartości bezwzględnych odchyleń:

$$S\acute{S}_x = \frac{\sum_{i=1}^{N} | x_i - \overline{x} |}{N}$$

W Excelu obliczenia te wykonuje funkcja `ODCH.ŚREDNIE`.

Ocena zależności między zbiorami danych pomiarowych

Kowariancja

Omówione powyżej wielkości charakteryzują oddzielnie zbiory wartości x i y. Aby ocenić zależność między tymi wielkościami, należy powiązać ich wartości we wspólnym agregacie. Pierwszą z takich wielkości jest **kowariancja**, występująca — podobnie jak wariancja — w dwóch odmianach: **kowariancji z próbki** i **kowariancji populacji**. Znając definicję wariancji, łatwo się domyślić, jak jest zbudowana ta funkcja:

$$S_{xypr} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{N-1}$$

$$S_{xypop} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{N}$$

Kowariancja służy do oceny, czy pomiędzy zmiennymi x i y istnieje zależność liniowa. Jeżeli zmienne są niezależne, kowariancja jest bliska zeru. Wariancje i kowariancję można wyrazić równoważnie także w inny sposób:

$$S_{xpop} = \frac{\sum_{i=1}^{N} x_i^2}{N} - \overline{x}^2 \qquad S_{ypop} = \frac{\sum_{i=1}^{N} y_i^2}{N} - \overline{y}^2 \qquad S_{xypop} = \frac{\sum_{i=1}^{N} x_i y_i}{N} - \overline{x} \cdot \overline{y}$$

Stosując konsekwentnie oznaczenie średniej arytmetycznej w postaci kreski nad symbolem, można te wzory zapisać jeszcze prościej:

$$S_{xpop} = \overline{x^2} - \overline{x}^2 \qquad S_{ypop} = \overline{y^2} - \overline{y}^2 \qquad S_{xypop} = \overline{xy} - \overline{x} \cdot \overline{y}$$

Należy pamiętać, że średnia wartość iloczynów to nie to samo co iloczyn średnich. W starszych wersjach Excela kowariancję obliczymy za pomocą funkcji `KOWARIANCJA`. Oblicza ona kowariancję populacji. Począwszy od Excela 2010, dostępne są dwie funkcje, `KOWARIANCJA.PRÓBKI` i `KOWARIANCJA.POPUL`, ale można też używać starych nazw funkcji w dotychczasowym znaczeniu.

Współczynnik korelacji

Do oceny zależności liniowej między zmiennymi x i y wykorzystuje się często kowariancję znormalizowaną poprzez podzielenie jej przez iloczyn odchyleń standardowych. Współczynnik ten nosi nazwę **współczynnika korelacji Pearsona**, zwany jest często krótko współczynnikiem korelacji R.

$$R = \frac{S_{xypop}}{S_{xpop} S_{ypop}} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2)(\overline{y^2} - \overline{y}^2)}} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}(y_i - \overline{y})^2}}$$

Współczynnik ten może przyjmować wartości z przedziału [–1, +1]. Wartości skrajne odpowiadają doskonałej korelacji liniowej, odpowiednio malejącej lub rosnącej.

W Excelu współczynnik korelacji oblicza się za pomocą funkcji `WSP.KORELACJI` lub `PEARSON`.

Równie często, a może nawet częściej, zamiast współczynnika korelacji oblicza się jego kwadrat. W Excelu służy do tego funkcja `R.KWADRAT`.

Zasady doboru współczynników w równaniu regresji

Jeżeli chcemy opisać zależność pomiędzy zmiennymi y i x w sposób sformalizowany, czyli w postaci równania ze współczynnikami liczbowymi (**równania regresji**), musimy określić kryterium doboru wartości tych współczynników. Jak już wspominałem, kryterium jakości dopasowania równania regresji do danych pomiarowych stanowi najczęściej minimalna wartość sumy kwadratów różnic między wartościami y zmierzonymi a wartościami $\hat{y}$ obliczonymi z równania regresji. Tę sumę określa się najczęściej nazwą **sumy resztkowej** i oznacza symbolem SS_{resz}:

$$SS_{resz} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

N oznacza tu liczbę danych odnoszących się do zmiennej y.

Do oceny jakości dopasowania niezbędna jest też tzw. **liczba stopni swobody,** $f = N - l$, czyli różnica między liczbą pomiarów N a liczbą współczynników l, które trzeba wyznaczyć. Jeżeli dysponujemy tylko tyloma wynikami pomiarów, ile współczynników jest do wyznaczenia, liczba stopni swobody $f = 0$, wtedy również $SS_{\mathrm{resz}} = 0$, ale nie ma możliwości statystycznej oceny jakości otrzymanego równania. Na przykład przez dwa punkty zawsze można przeprowadzić prostą, ale gdyby dodać trzeci punkt pomiarowy, to nie wiadomo, czy też leżałby na tej prostej. Miarą jakości dopasowania równania jest stosunek sumy resztkowej SS_{resz} do liczby stopni swobody f, zwany **wariancją resztkową**.

Pierwiastek z wariancji resztkowej zwany jest **średnim błędem oszacowania wartości *y*** s_y. To kolejna charakterystyka statystyczna równania regresji. Wartość tę umieszcza się czasem na wykresach w postaci słupków błędu wartości y. Jest ona obliczana ze wzoru:

$$s_y = \sqrt{\frac{SS_{resz}}{N - l}} = \sqrt{\frac{SS_{resz}}{f}}$$

Każdy z wyznaczonych współczynników w równaniu regresji obarczony jest pewnym błędem, który również można oszacować. Im większa jest wartość współczynnika (liczona jako wartość bezwzględna) w stosunku do błędu jego oszacowania, tym bardziej istotny jest dany współczynnik. Współczynniki mało istotne, tzn. takie, których błąd jest relatywnie duży, powinny być usuwane z równania regresji. Pomocne jest tu **kryterium statystyczne *t* Studenta**, które określa graniczną wartość tego stosunku, zależną od liczby stopni swobody powtarzalności pomiarów oraz od przyjętego tzw. poziomu istotności.

Suma resztkowa i wariancja resztkowa nie wystarczają jednak do pełnej charakterystyki danych pomiarowych, bo jeżeli wartości y zmieniają się wyraźnie od punktu do punktu (zależność jest stroma), to i suma resztkowa będzie stosunkowo duża. Jeżeli natomiast zmienność y jest niewielka (zależność jest płaska), to i suma resztkowa będzie mała, chociaż równanie niekoniecznie musi być dobrze dobrane. Dlatego sensownie jest porównać sumę resztkową z sumą kwadratów odchyleń wyników w punktach pomiarowych odniesionych do ogólnej średniej z wszystkich wyników pomiarów. Tę ogólną sumę kwadratów wokół średniej pomniejszoną o sumę resztkową nazwano **sumą regresyjną**. Oblicza się ją ze wzoru:

$$SS_{regr} = \sum_{i=1}^{N} (y_i - \overline{y})^2 - SS_{resz} = SS_y - SS_{resz}$$

W Excelu (od edycji 2003) zależność tę stosuje się tylko do pełnego równania regresji (gdy wyraz wolny jest obliczany). Kiedy wyraz wolny jest równy zeru, stosowany jest zmodyfikowany wzór (kwestię tę wyjaśniam dokładnie dalej). Sumy kwadratów resztkowa i regresyjna są porównywane w dwojaki sposób. Na ich podstawie oblicza się tzw. **współczynnik determinacji** *D*:

$$D = \frac{SS_{regr}}{SS_{regr} + SS_{resz}} = 1 - \frac{SS_{resz}}{SS_y}$$

Wielkość ta jest bardzo często wykorzystywana do oceny jakości równania regresji. Jeżeli dane nie są obarczone dużymi błędami pomiarów, a równanie regresji zostało dobrze dobrane, jej wartość jest bardzo bliska 1. Wartości poniżej 0,9 świadczą już o stosunkowo mało dokładnym równaniu. Nie jest to jednak całkiem jednoznaczne — w pomiarach fizycznych (szerzej w naukach ścisłych) współczynniki determinacji przyjmują zwykle wartości dużo bliższe 1 niż na przykład w naukach medycznych, rolniczych czy ekonomicznych.

Inną charakterystyką statystyczną wyliczaną na podstawie powyższych sum kwadratów jest stosunek *F* wariancji regresji do wariancji resztkowej. Jest to wielkość, którą porównuje się z **kryterium statystycznym Fishera-Snedecora**:

$$F = \frac{SS_{regr}/(l-1)}{SS_{resz}/f} \qquad \text{lub} \qquad F = \frac{SS'_{regr}/l}{SS_{resz}/f}$$

Pierwszy wariant wzoru stosuje się, gdy stała w równaniu regresji jest obliczana, a drugi — gdy stała jest równa zeru. Jednak w tym drugim przypadku suma regresyjna obliczana jest ze zmodyfikowanego wzoru (zobacz dalej). Wartość *F* powinna być możliwie jak największa, a w każdym razie większa od wartości granicznej, określonej w zależności od liczby stopni swobody licznika i mianownika oraz przyjętego poziomu istotności.

Regresja liniowa w Excelu

Znaczenie współczynników w równaniu regresji i sposoby ich obliczania

W arkuszu kalkulacyjnym dostępnych jest kilka funkcji, które służą do wyznaczania równań regresji. Ponadto dostępne są narzędzia do analizy regresji powiązane z kreatorem wykresów (linia trendu) oraz niezależne, wchodzące w skład dodatku *Analysis Toolpak*, których nie będę tu omawiał.

Najprostszy przypadek dotyczy dopasowania funkcji liniowej jednej zmiennej. Jeśli poszukiwana zależność da się opisać równaniem:

$$y = ax + b$$

to zastosowanie metody najmniejszej sumy kwadratów (określanej dalej skrótowo jako NK) prowadzi do następujących wzorów na współczynniki a i b:

$$a = \frac{S_{xypop}}{S_{xpop}} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$b = \overline{y} - a\overline{x} = \frac{\overline{y} \cdot \overline{x^2} - \overline{x} \cdot \overline{xy}}{\overline{x^2} - \overline{x}^2}$$

$$R^2 = \frac{S_{xypop}^2}{S_{xpop} S_{ypop}} = \frac{(\overline{xy} - \overline{x} \cdot \overline{y})^2}{(\overline{x^2} - \overline{x}^2)(\overline{y^2} - \overline{y}^2)} = a^2 \frac{\overline{x^2} - \overline{x}^2}{\overline{y^2} - \overline{y}^2} = a^2 \frac{S_{xpop}}{S_{ypop}}$$

$$SS_{reszt} = \sum_{i=1}^{N}(y_i - \overline{y})^2 - \frac{\left(\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})\right)^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2} = (1 - R^2)\sum_{i=1}^{N}(y_i - \overline{y})^2 = \sum_{i=1}^{N}(y_i - \overline{y})^2 - a^2\sum_{i=1}^{N}(x_i - \overline{x})^2$$

$$s_y = \sqrt{\frac{SS_{resz}}{N-2}} = \sqrt{\frac{SS_{resz}}{f}}$$

Funkcje NACHYLENIE, ODCIĘTA, R.KWADRAT i REGBŁSTD

W Excelu można wykorzystać cztery funkcje o podobnej składni, które umożliwiają obliczenie współczynników a i b, współczynnika determinacji D, który w tym przypadku jest równy kwadratowi współczynnika korelacji R^2, oraz standardowego błędu oceny wartości $\boldsymbol{y}$ s_y. Są to funkcje:

`NACHYLENIE(`***znane_y***`;` ***znane_x***`)` — oblicza współczynnik a

`ODCIĘTA(`***znane_y***`;` ***znane_x***`)` — oblicza współczynnik b

`R.KWADRAT(`***znane_y***`;` ***znane_x***`)` — oblicza współczynnik R^2

`REGBŁSTD(`***znane_y***`;` ***znane_x***`)` — oblicza oszacowanie błędu s_y

Znane_y i ***znane_x*** to tablice lub zakresy komórek o identycznych rozmiarach zawierające dane liczbowe. Jeżeli zakresy lub tablice zawierają dane nieliczbowe lub puste komórki, te wartości są ignorowane (jeśli jedna z wartości x lub y jest nieprawidłowa, ignorowana jest cała para). Jeżeli rozmiary tablic lub zakresów są różne, otrzymujemy komunikat błędu `#N/D!`. Orientacja ***znane_y***

i ***znane_x*** jest obojętna, ale zakresy muszą być spójne. Kolejność punktów ***znane_y*** i ***znane_x*** ustala się, odczytując te tablice lub zakresy wierszami. ***Znane_x*** nie musi być uporządkowane. Zakresy lub tablice ***znane_y*** i ***znane_x*** muszą zawierać przynajmniej dwa różne punkty.

Pomiędzy funkcjami wymienionymi wcześniej istnieją zależności:

```
NACHYLENIE(znane_y; znane_x) = KOWARIANCJA(znane_y; znane_x)/ WARIANCJA.POPUL(znane_x)
ODCIĘTA(znane_y; znane_x) = ŚREDNIA(znane_y)–ŚREDNIA(znane_x)* NACHYLENIE(znane_y; znane_x)
R.KWADRAT(znane_y; znane_x) = NACHYLENIE(znane_y; znane_x) * NACHYLENIE(znane_x; znane_y)
```

Zależności te są spełnione pod warunkiem, że wszystkie dane są poprawne, to znaczy zakresy ***znane_y*** i ***znane_x*** mają takie same rozmiary i zawierają wyłącznie liczby.

Funkcja REGLINP

Jeśli w równaniu regresji występuje więcej wyrazów, a zarazem więcej współczynników, albo jeśli wyraz wolny (odcięta) jest równy zeru, albo jeśli potrzebny jest pełniejszy statystyczny opis zbioru danych — należy skorzystać z funkcji tablicowej:

```
REGLINP(znane_y; znane_x; czy_stała; czy_stat)
```

Znane_y i *znane_x*, jak poprzednio, reprezentują zbiory danych pomiarowych zmiennych *y* i *x*, ale teraz obszary te nie muszą mieć identycznych rozmiarów, jednak muszą być dopasowane. Jeżeli tablica ***znane_y*** jest pojedynczą kolumną, każda kolumna zbioru *znane_x* jest interpretowana jako oddzielna zmienna. Analogicznie jeżeli tablica ***znane_y*** jest pojedynczym wierszem, każdy wiersz zbioru *znane_x* jest interpretowany jako oddzielna zmienna.

Tablica *znane_x* może zawierać jeden lub więcej zbiorów zmiennych. Jeżeli używana jest tylko jedna zmienna, ***znane_y*** i *znane_x* mogą być obszarami lub tablicami o dowolnym kształcie, ale o jednakowych wymiarach.

Jeżeli stosuje się więcej niż jedną zmienną *x*, ***znane_y*** musi być wektorem (tzn. zakresem lub tablicą o wysokości jednego wiersza bądź o szerokości jednej kolumny). Tablice ***znane_y*** i *znane_x* muszą być w całości wypełnione danymi liczbowymi (**nie mogą zawierać pustych komórek ani danych nieliczbowych** — inaczej niż w przypadku funkcji NACHYLENIE, ODCIĘTA, R.KWADRAT i REGBŁSTD), ponadto jeśli *znane_x* reprezentuje kilka zmiennych, musi stanowić **zakres spójny**, czyli musi być zawarte w jednym obszarze lub jednej tablicy.

Jeżeli pominie się *znane_x*, argument ten jest przyjmowany jako tablica {1;2;3;...} lub {1\2\3\...}, która ma taką samą wielkość jak ***znane_y***.

Funkcja toleruje sytuację, w której liczba punktów jest zbyt mała do jednoznacznego wyznaczenia współczynników regresji. Przyjmuje wówczas zerowe współczynniki dla zmiennych, dla których nie jest możliwe wyznaczenie jednoznacznego wyniku.

Czy_stała jest to wartość logiczna określająca, czy stała *b* ma być równa 0, czy obliczana normalnie. Jeżeli *czy_stała* ma wartość PRAWDA lub jest pominięta, *b* jest obliczane normalnie. Jeżeli *czy_stała* ma wartość FAŁSZ, *b* jest równe 0, a wartości *a* są dopasowywane do równania $y = ax$ lub $y = \sum a_i x_i$.

> Przy dodawaniu linii trendu do wykresu punktowego można zadeklarować dowolną wartość stałej *b* w równaniu regresji. W przypadku funkcji `REGLINP` można jedynie zadeklarować wartość `0`. Jeżeli znana wartość stałej *b* jest różna od zera, należy od zmierzonych wartości *y* odjąć *b* i w charakterze zmiennej ***znane_y*** użyć tej różnicy, nadając równocześnie argumentowi *`czy_stała`* wartość `FAŁSZ`.
>
> Jeżeli linia regresji ma przechodzić przez punkt stały o współrzędnych `(`***`x0`***`,`***`y0`***`)`, należy postąpić analogicznie, to znaczy wywołanie funkcji powinno mieć postać:
>
> `REGLINP(`***`znane_y – y0`***`; `***`znane_x – x0`***`; FAŁSZ; `***`czy_stat`***`)`
>
> a równanie linii regresji:
>
> ***`y = y0 + a*(x – x0).`***

`Czy_stat` jest to wartość logiczna określająca, czy ma być dodatkowo obliczana statystyka regresji. Jeżeli *`czy_stat`* ma wartość `PRAWDA`, funkcja `REGLINP` daje w wyniku dodatkowe statystyki regresji, tak że uzyskiwana tablica ma postać $\{a_n; a_{n-1}; \ldots; a_1; a_0 \setminus s_{an}; s_{an-1}; \ldots; s_{a1}; s_{a0} \setminus D; s_y \setminus F; f \setminus SS_{regr}; SS_{resz}\}$ (separatory w konwencji Excela 2007–). Wynik składa się z pięciu wierszy i tylu kolumn, ile współczynników jest obliczanych. W pierwszym wierszu jest lista współczynników, ale **w kolejności odwrotnej** w stosunku do kolejności kolumn z danymi, zakończona wyrazem wolnym a_0. W drugim wierszu są średnie błędy współczynników z pierwszego wiersza. Od trzeciego do piątego wiersza tylko po dwie wartości — kolejno: współczynnik determinacji *D*, średni błąd oszacowania wartości *y* s_y, statystyka Fishera-Snedecora *F*, obok liczba stopni swobody wariancji resztkowej *f* i wreszcie w piątym wierszu suma regresyjna SS_{regr} i suma resztkowa SS_{resz}. Znaczenie tych wielkości i sposób ich obliczania podałem wcześniej, w punkcie „Zasady doboru współczynników w równaniu regresji”. Jeżeli *`czy_stat`* ma wartość `FAŁSZ` lub jest pominięty, funkcja `REGLINP` daje w wyniku tylko współczynniki $\{a_n; a_{n-1}; \ldots; a_1; a_0\}$. Wyniki mieszczą się w jednym wierszu.

Funkcja `REGLINP` jest funkcją tablicową, więc wymaga wcześniejszego zaznaczenia miejsca na wynik i zatwierdzenia klawiszami *Ctrl+Shift+Enter*. Chociaż funkcja w swojej pełnej wersji może wyświetlać wyniki w pięciu wierszach, nie jest konieczne wyświetlanie ich wszystkich. Można na przykład zaznaczyć miejsce na wyświetlenie tylko trzech pierwszych wierszy, a zrezygnować z dwu ostatnich.

W przypadku równania z jedną zmienną współczynnik nachylenia i odciętą można odczytać z formuł:

Nachylenie *a*: `= INDEKS(REGLINP(`***`znane_y`***`;`***`znane_x`***`); 1)`

Punkt przecięcia z osią *y b*: `= INDEKS(REGLINP(`***`znane_y`***`;`***`znane_x`***`); 2)`

Relacja między współczynnikiem determinacji a kwadratem współczynnika korelacji

Należy zwrócić uwagę, że relacja:

$$\sum_{i=1}^{N}(y_i - \bar{y})^2 = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

jest spełniona tylko wtedy, gdy wartości funkcji $\hat{y}$ są wyznaczane z pełnego równania prostej z wyrazem wolnym, a współczynniki są obliczane zgodnie z zasadą NK.

Człony tego wzoru są interpretowane kolejno jako: całkowite zróżnicowanie zmiennej *y*, zróżnicowanie zmiennej *y* objaśnione modelem i zróżnicowanie zmiennej *y* nieobjaśnione modelem.

W oparciu o ten wzór określa się współczynnik determinacji *D* jako stosunek sum:

$$D = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} = 1 - \frac{SS_{resz}}{SS_y}$$

Po spełnieniu warunków, o których była mowa wyżej, prawdziwa jest także równość:

$$D = \frac{\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2}$$

Uwzględniając wartości współczynników *a* i *b* obliczone metodą NK, otrzymujemy:

$$D = \frac{\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} = \frac{\sum_{i=1}^{N}(ax_i + b - \overline{y})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} = \frac{\sum_{i=1}^{N}(ax_i - a\overline{x})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} = a^2\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} = R^2$$

Zatem w tym przypadku współczynnik determinacji jest równy kwadratowi współczynnika korelacji.

Należy zauważyć, że według podanej definicji wartość **współczynnika determinacji zależy** od przyjętego **równania modelowego** opisującego dane, natomiast wartość **współczynnika korelacji zależy** tylko od samych **danych**.

Współczynnik korelacji między zmienną y a funkcją liniową x

Załóżmy, że mamy zbiór par wartości (*x*, *y*). Dla tego zbioru możemy określić współczynnik korelacji pomiędzy wartościami *x* i *y*. Możemy obliczyć go za pomocą funkcji `WSP.KORELACJI`, a jego kwadrat za pomocą funkcji `R.KWADRAT`. Przekształcimy teraz wartości *x* w wartości $\hat{y}$, korzystając z dowolnej funkcji liniowej:

$$\hat{y} = ax + b$$

Obliczymy współczynnik korelacji między wartościami y i $\hat{y}$:

$$R(y,\hat{y}) = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Mamy:

$$\hat{y}_i = ax_i + b \qquad \text{i} \qquad \bar{\hat{y}} = a\bar{x} + b$$

Zatem:

$$\hat{y}_i - \bar{\hat{y}} = (ax_i + b) - (a\bar{x} + b) = a(x_i - \bar{x})$$

Po podstawieniu otrzymamy:

$$R(y,\hat{y}) = \frac{a\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{a^2\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} = \mathrm{sgn}(a)R(x,y)$$

Widzimy zatem, że współczynnik korelacji dowolnej funkcji liniowej zmiennej x ze zmienną y jest taki sam jak współczynnik korelacji samych zmiennych x i y z dokładnością do znaku a. Kwadraty współczynników korelacji są równe. Zależność ta obowiązuje dla dowolnych a i b, a nie tylko tych wyznaczonych metodą NK.

Równanie ze znanym wyrazem wolnym lub bez niego

Przypadkiem szczególnym jest model liniowy ze stałym współczynnikiem b (prosta przechodzi przez punkt stały) — wtedy zastosowanie metody NK prowadzi do poniższego wzoru na współczynnik kierunkowy a_N:

$$a_N = \frac{\sum_{i=1}^{N} x_i y_i - b\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N} x_i(y_i - b)}{\sum_{i=1}^{N} x_i^2} = \frac{\overline{xy} - b\bar{x}}{\overline{x^2}}$$

Model procesu może nie zawierać w ogóle wyrazu wolnego, czyli $b = 0$, a prosta wówczas przechodzi przez początek układu współrzędnych. W tym przypadku wzór upraszcza się do:

$$a_N = \frac{\overline{xy}}{\overline{x^2}}$$

Jeżeli rozpatrujemy pełne równanie prostej $y = ax + b$, to zgodnie z logiką narzucenie warunku $b = 0$ powoduje na ogół pogorszenie jakości dopasowania równania regresji do danych i w związku z tym należy oczekiwać niższych (w każdym razie nie wyższych) wartości współczynnika determinacji D. W pewnych przypadkach można otrzymać nawet wartości ujemne. Dotyczy to zwykle sytuacji, gdy dane są nieskorelowane lub słabo skorelowane, tzn. współczynnik korelacji R jest równy 0 lub bliski zeru. Równocześnie współczynnik korelacji i jego kwadrat zachowują tę samą wartość (zależą tylko od danych, a nie od postaci równania). W każdym razie jeśli narzucimy wartość któregokolwiek współczynnika lub dobierzemy współczynniki według algorytmu innego niż metoda NK, zamiast relacji $D = R^2$ mamy $D < R^2$.

Niescentrowany współczynnik determinacji

Aby uniknąć ujemnych wartości współczynnika D, w przypadku równania **bez wyrazu wolnego** zaproponowano użycie zmodyfikowanego współczynnika, zwanego w literaturze **niescentrowanym współczynnikiem determinacji** D_N:

$$D_N = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} y_i^{\,2}}$$

Wzór ten wynika z relacji:

$$\sum_{i=1}^{N} y_i^{\,2} = \sum_{i=1}^{N} \hat{y}_i^{\,2} + \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Jest ona spełniona tylko wtedy, gdy wartości funkcji $\hat{y}$ są wyznaczane z równania prostej bez wyrazu wolnego, a współczynnik a_N jest obliczany zgodnie z zasadą NK.

Wzór opisujący D_N można uogólnić na przypadek dowolnej (lecz ustalonej) wartości wyrazu wolnego b:

$$D_N = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - b)^2}$$

Jeżeli wzór ten zastosujemy do modelu liniowego z ustaloną wartością wyrazu wolnego b i ze współczynnikiem a_N określonym według metody NK, otrzymamy:

$$D_N = \frac{\left(\sum_{i=1}^{N} x_i (y_i - b) \right)^2}{\sum_{i=1}^{N} x_i^{\,2} \sum_{i=1}^{N} (y_i - b)^2} = \frac{(\overline{xy} - b\overline{x})^2}{\overline{x^2} \cdot (\overline{y^2} - 2b\overline{y} + b^2)}$$

W szczególnym przypadku, gdy $b = 0$, mamy:

$$D_N = \frac{\left(\sum_{i=1}^{N} x_i y_i\right)^2}{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2} = \frac{(\overline{xy})^2}{\overline{x^2} \cdot \overline{y^2}}$$

Z powyższych zależności widać, że tak zdefiniowany współczynnik będzie przyjmował wartości z przedziału [0, 1]. Należy jednak zauważyć, że w przypadku arbitralnego równania modelowego współczynnik D_N również może przyjmować wartości ujemne.

Związek między niescentrowanym a tradycyjnym współczynnikiem determinacji

Związek między niescentrowanym (przy $b = 0$) a tradycyjnym współczynnikiem determinacji można określić, porównując wzory:

$$D = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2} = 1 - \frac{SS_{resz}}{SS_y} \qquad \text{i} \qquad D_N = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} y_i^2} = 1 - \frac{SS_{resz}}{SS'_y}$$

Sumę z mianownika pierwszego wzoru można zapisać jako:

$$\sum_{i=1}^{N} (y_i - \overline{y})^2 = \sum_{i=1}^{N} y_i^2 - N\overline{y}^2 \quad \text{czyli} \quad SS_y = SS'_y - N\overline{y}^2$$

Z porównania wzorów wynika zależność:

$$\frac{1}{1-D} = \frac{1}{1-D_N} - \frac{N\overline{y}^2}{SS_{resz}}$$

Niescentrowany współczynnik korelacji

Można zdefiniować wielkość zwaną **niescentrowanym współczynnikiem korelacji** R_N według wzoru:

$$R_N = \frac{\overline{xy}}{\sqrt{\overline{x^2} \cdot \overline{y^2}}} = \frac{\sum_{i=1}^{N} x_i y_i}{\sqrt{\sum_{i=1}^{N} (x_i)^2 \sum_{i=1}^{N} (y_i)^2}}$$

Postać wzoru jest analogiczna do wzoru definiującego współczynnik korelacji *R*, lecz jego sens jest inny. Podobnie jak w przypadku *R* wartość tego współczynnika mieści się w przedziale [−1, +1]. Wartości skrajne odpowiadają doskonałej zależności proporcjonalnej, odpowiednio ze współczynnikiem ujemnym lub dodatnim. Jeżeli wyraz wolny nie jest równy zeru, należałoby zastosować wzór:

$$R_N = \frac{\sum_{i=1}^{N} x_i (y_i - b)}{\sqrt{\sum_{i=1}^{N} x_i^{\,2} \sum_{i=1}^{N} (y_i - b)^2}} = \frac{\overline{xy} - b\overline{x}}{\sqrt{\overline{x^2}} \cdot \sqrt{\overline{y^2} - 2b\overline{y} + b^2}}$$

Definicja współczynnika R_N zależy od wartości *b*, czyli od postaci równania regresji. Jeżeli współczynnik proporcjonalności a_N jest wyznaczony według metody NK, spełniona jest zależność $D_N = R_N^2$.

Równanie ze znanym współczynnikiem kierunkowym

Analogicznie gdyby w modelu liniowym wartość współczynnika *a* była ustalona, wartość współczynnika b_M wyznaczona metodą NK wyniosłaby:

$$b_M = \frac{\sum_{i=1}^{N} y_i - a\sum_{i=1}^{N} x_i}{N} = \frac{\sum_{i=1}^{N} (y_i - ax_i)}{N} = \overline{y} - a\overline{x}$$

W takim przypadku obowiązuje relacja:

$$\sum_{i=1}^{N} (y_i - ax_i)^2 = \sum_{i=1}^{N} (\hat{y}_i - ax_i)^2 + \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Na jej podstawie można zdefiniować jeszcze inny zmodyfikowany współczynnik determinacji D_M, specyficzny dla tej postaci równania:

$$D_M = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - ax_i)^2}$$

Po podstawieniu wartości współczynnika b_M obliczonej metodą NK otrzymujemy:

$$D_M = \frac{\sum_{i=1}^{N} (\hat{y}_i - ax_i)^2}{\sum_{i=1}^{N} (y_i - ax_i)^2} = \frac{\sum_{i=1}^{N} (ax_i + b_N - ax_i)^2}{\sum_{i=1}^{N} (y_i - ax_i)^2} = \frac{\sum_{i=1}^{N} b_N^{\,2}}{\sum_{i=1}^{N} (y_i - ax_i)^2} = \frac{(\overline{y} - a\overline{x})^2}{\overline{(y - ax)^2}} = R_M^2$$

Zmodyfikowany współczynnik korelacji R_M, który mógłby mieć zastosowanie w tym przypadku, wyraża się wzorem:

$$R_M = \frac{\overline{y} - a\overline{x}}{\sqrt{\overline{\overline{(y - ax)^2}}}}$$

Ten przypadek nie został uwzględniony w Excelu; przytaczam go, aby zwrócić uwagę, że do różnych postaci równania regresji można dopasować odpowiednio zmodyfikowane definicje współczynników korelacji i determinacji, jednakże nie ma pomiędzy nimi prostych zależności, które umożliwiałyby porównania.

Pomiędzy współczynnikami można zapisać relacje:

$$\frac{D}{a^2} = \frac{\overline{x^2} - \overline{x}^2}{\overline{y^2} - \overline{y}^2} = \frac{\overline{(x - \overline{x})^2}}{\overline{(y - \overline{y})^2}}, \qquad \frac{D_N}{a_N^2} = \frac{\overline{x^2}}{\overline{(y - b)^2}} \quad \text{oraz} \quad \frac{D_M}{a^2} = \frac{(\overline{y}/a - \overline{x})^2}{\overline{(y - ax)^2}}$$

Specyfika działania funkcji REGLINP przy braku wyrazu wolnego

W opisie działania funkcji `REGLINP` podano, że gdy wyraz wolny ma być równy **zeru**, zamiast współczynnika determinacji $D = R^2$ oblicza się **niescentrowany** współczynnik determinacji $D_N = R_N^2$. Do obliczania wartości F i R^2 zamiast sumy regresyjnej liczonej według wzoru:

$$SS_{regr} = \sum_{i=1}^{N} (y_i - \overline{y})^2 - SS_{resz} = SS_y - SS_{resz}$$

wykorzystuje się wielkość:

$$SS'_{regr} = \sum_{i=1}^{N} y_i^2 - SS_{resz} = SS'_y - SS_{resz}$$

Dotyczy to wersji 2003+. W wersji 2000 zawsze liczony jest tradycyjny współczynnik determinacji.

Przyjęcie takiego rozwiązania jest co najmniej dyskusyjne, gdyż prowadzi do różnych niekonsekwencji, o których mowa poniżej. Ponadto rozwiązanie to jest niezgodne z przyjętym w kreatorze wykresów. Gdy na wykresie, do którego dodano linię trendu, zażądamy wyświetlenia R^2, otrzymujemy w istocie wartości współczynnika determinacji D obliczone ze wzoru:

$$D = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2}$$

niezależnie od tego, czy wartość stałej *b* jest obliczana czy ustalona. Wartości *D* czasem bywają ujemne, ale nie powinno to stwarzać problemów interpretacyjnych, poza tym, że niefortunnie określono ten współczynnik mianem R^2. Należy pamiętać, że na wykresie można zażądać dowolnej wartości współczynnika *b*, a nie tylko wartości zerowej, jak w przypadku funkcji `REGLINP`. Co wtedy należałoby wyświetlić jako *D*? Gdyby chcieć utrzymać konwencję przyjętą w funkcji `REGLINP`, należałoby użyć podanej wcześniej uogólnionej definicji niescentrowanego współczynnika determinacji:

$$D_N = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - b)^2}$$

W wersji 2000 funkcja `REGLINP` daje taką samą wartość współczynnika determinacji jak kreator wykresu, do którego dodano linię trendu.

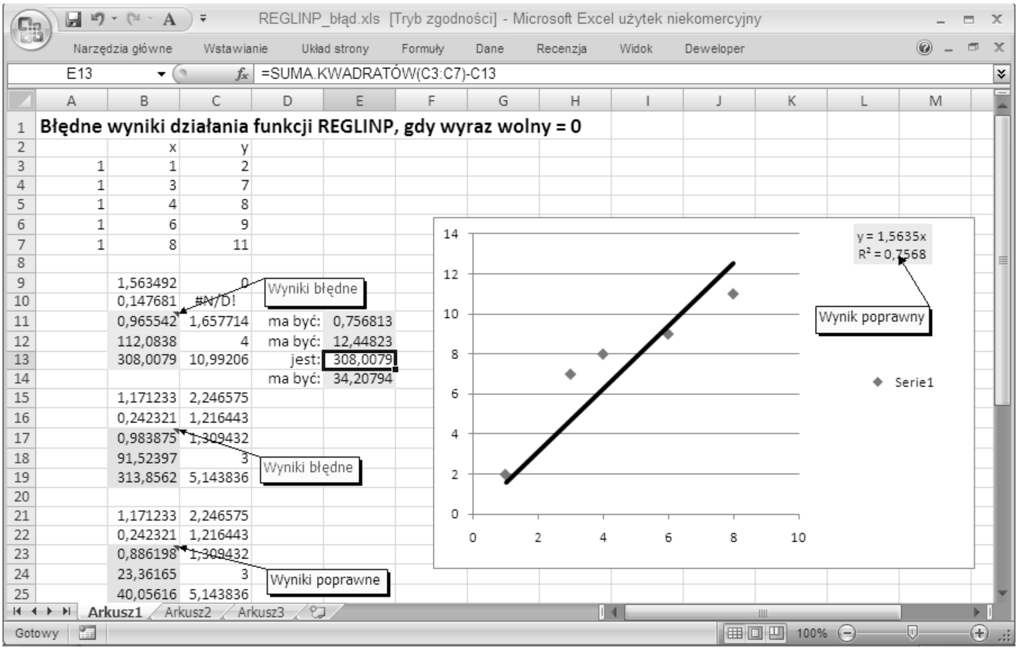
Na rysunku 7.1 pokazano, na czym polega (od wersji 2003) problem z funkcją `REGLINP`. Obliczenia przy użyciu tej funkcji wykonano trzykrotnie. Raz jako ***znane_x*** użyto jednej kolumny danych, z argumentem *`czy_stała`* o wartości `FAŁSZ` — wówczas funkcja zwraca (moim zdaniem niesłusznie) niescentrowany współczynnik determinacji, natomiast na wykresie współczynnik determinacji obliczony jest poprawnie. W drugim przypadku dodano drugą kolumnę *x* złożoną z samych jedynek i obliczono dwa współczynniki nadal z argumentem *`czy_stała`* o wartości `FAŁSZ` — tu również uzyskano niescentrowany współczynnik determinacji. W trzecim przypadku użyto jednej kolumny danych jako ***znane_x***, z argumentem *`czy_stała`* o wartości `PRAWDA` — i uzyskano klasyczny współczynnik determinacji (poprawnie). Obliczone współczynniki równania prostej w drugim i trzecim przypadku są liczbowo równe, gdyż dotyczą faktycznie tych samych danych, różnica polega tylko na innym sposobie obliczania wskaźników statystycznych.

Reasumując, uważam, że sposób wyliczenia współczynnika determinacji zastosowany w kreatorze wykresów jest bardziej poprawny i spójny logicznie, a podejście zastosowane w funkcji `REGLINP` od wersji 2003 jest niekonsekwentne, prowadzi bowiem do niejednoznaczności i niezbyt logicznych wniosków dotyczących jakości dopasowania wzoru.

Nietypowe zastosowania funkcji REGLINP

Rozwiązanie układu równań liniowych

Kwestii rozwiązywania układów równań liniowych poświęciłem początek poprzedniego rozdziału. Przedstawiłem tam typowe sposoby rozwiązania tego zagadnienia: wykorzystanie rachunku macierzowego lub dodatku *Solver*. W matematyce znane są również inne sposoby osiągnięcia tego celu, takie jak na przykład zastosowanie wzorów Cramera czy metody eliminacji Gaussa, których użycie w Excelu wymaga bardziej złożonych obliczeń, dlatego ich omówienie pominąłem. Skoro jednak znamy już funkcję `REGLINP`, warto powrócić do tego zagadnienia. Istotą działania funkcji `REGLINP` jest przecież rozwiązanie układu równań liniowych, tylko tych równań jest



RYSUNEK 7.1. *Przykład ilustrujący nieprawidłowe wartości parametrów statystycznych podawane przez funkcję REGLINP w przypadku, gdy wyraz wolny jest równy zeru (od wersji 2003)*

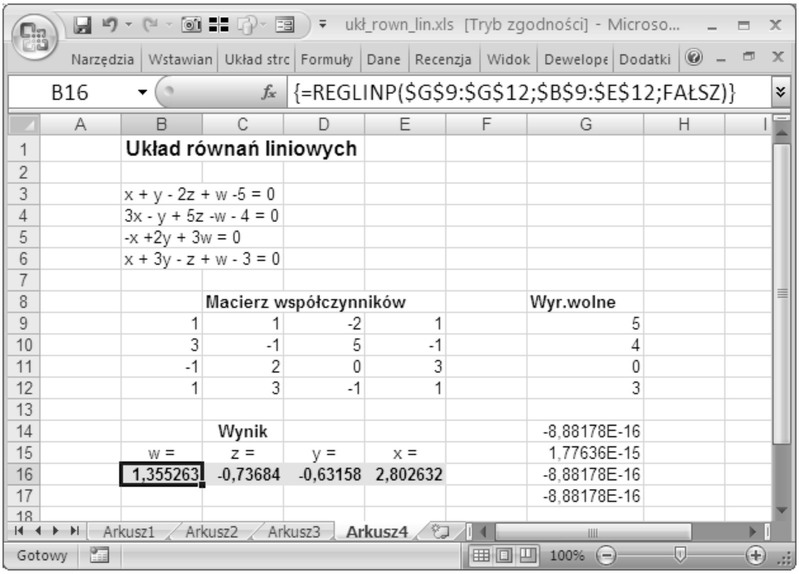
z reguły więcej niż niewiadomych, więc funkcja `REGLINP` nie może znaleźć dokładnego rozwiązania układu wszystkich równań, lecz szuka rozwiązania przybliżonego, najlepiej dopasowanego do danych. Jeśli jednak liczba równań jest równa liczbie niewiadomych, to może istnieć rozwiązanie dokładne (o ile układ jest oznaczony). Dla funkcji `REGLINP` jest to przypadek nietypowy, bo liczba stopni swobody jest równa zeru, więc funkcja nie może obliczyć żadnych statystyk regresji, ale same współczynniki, które odpowiadają w tym przypadku naszym niewiadomym — obliczyć może. Jedyne, o czym musimy pamiętać, to odwrócona kolejność niewiadomych w wynikach. Jak już pisałem, jest to specyfika funkcji `REGLINP`, którą zresztą też można wykorzystać (zobacz następny podpunkt).

Jeśli mamy do rozwiązania układ równań (w zapisie macierzowym):

$$\boldsymbol{A} \times \boldsymbol{X} = \boldsymbol{B}$$

i chcemy go rozwiązać za pomocą funkcji `REGLINP`, musimy potraktować macierz współczynników $\boldsymbol{A}$ jako ***znane_x***, natomiast wektor wyrazów wolnych $\boldsymbol{B}$ jako ***znane_y***. Należy przyjąć, że funkcja nie oblicza wyrazu wolnego, dlatego trzeci argument funkcji `REGLINP` powinien być równy `FAŁSZ` (lub `0`), a czwarty argument można pominąć (statystyka jest w tym przypadku nieistotna). Wyniki otrzymamy w wierszu (przy tradycyjnym sposobie rozwiązania była to kolumna) i w kolejności odwróconej w stosunku do kolejności kolumn w macierzy $\boldsymbol{A}$.

Na przykład jeśli mamy układ czterech równań z czterema niewiadomymi x, y, z, w i kolejność kolumn w macierzy $\boldsymbol{A}$ odpowiada tej kolejności niewiadomych, to w wyniku funkcja `REGLINP` zwróci nam wiersz z wynikami w kolejności: *w*, *z*, *y*, *x*. Ilustrację tego sposobu rozwiązania stanowi rysunek 7.2. W obliczeniach wykorzystano ten sam układ równań, który był pokazany na rysunku 6.1.



RYSUNEK 7.2. *Rozwiązanie układu równań liniowych za pomocą funkcji REGLINP*

Odwracanie kolejności elementów w zakresie lub tablicy

To zagadnienie również było już wcześniej poruszane. W rozdziale 6. pokazałem, jak można je rozwiązać za pomocą funkcji `INDEKS` i `PRZESUNIĘCIE`. Temat będzie jeszcze kontynuowany w rozdziale 20. Z przedstawionych opisów wynika, że stosunkowo łatwo można zapisać formułę generującą rozwiązanie sekwencyjnie (formuła wymaga kopiowania do kolejnych komórek). Natomiast uzyskanie rozwiązania w formie wielokomórkowej formuły tablicowej zwracającej od razu cały wynik przydatny do dalszych obliczeń jest trudniejsze i wymaga użycia sposobów niestandardowych.

Tymczasem, jak wiadomo, funkcja `REGLINP` została tak napisana, że zwraca wyniki w odwróconej kolejności. Zapiszmy przykładowy układ równań w postaci:

$$1{*}x + 0{*}y + 0{*}z = 2$$

$$0{*}x + 1{*}y + 0{*}z = 5$$

$$0{*}x + 0{*}y + 1{*}z = 1$$

W formalnym zapisie macierzowym można to zapisać jako:

$$\boldsymbol{I} \times \boldsymbol{X} = \boldsymbol{B}$$

I oznacza macierz jednostkową rzędu równego liczbie niewiadomych. Oczywistym rozwiązaniem tego równania jest:

$$\boldsymbol{X} = \boldsymbol{B}$$

Możemy zastosować do tego układu równań metodę opisaną w poprzednim podpunkcie. Jeśli jako argumentu ***znane_x*** użyjemy macierzy jednostkowej, otrzymamy wynik w postaci wektora ***B***, czyli ***znane_y***, w odwróconej kolejności. Ponadto, ponieważ macierz jednostkowa jest symetryczna, wektor przetwarzany (***znane_y***) może być zarówno w układzie pionowym, jak i poziomym. Natomiast wynik otrzymamy zawsze w układzie poziomym. W razie potrzeby można zastosować do niego transpozycję. Wykorzystanie tej metody ogranicza się wyłącznie do danych **liczbowych**.

W Excelu 2013+ sprawa jest prosta, bo istnieje funkcja generująca macierz jednostkową i wystarczy zapisać:

```
= REGLINP(znane_y; MACIERZ.JEDNOSTKOWA(ILE.NIEPUSTYCH(znane_y)); FAŁSZ)
```

We wcześniejszych wersjach programu trzeba wygenerować macierz jednostkową odpowiednią formułą [A4].

Jeśli ***znane_y*** składa się z wartości unikatowych, wystarczy taki zapis:

```
= REGLINP(znane_y; (znane_y = TRANSPONUJ(znane_y))*1; FAŁSZ)
```

Gdy przewidujemy możliwość wystąpienia duplikatów, przydatne może być zdefiniowanie i nazwanie tablicy bazowej:

```
wd : = WIERSZ(ADR.POŚR("1:"&ILE.NIEPUSTYCH(znane_y)))
```

i wykorzystanie jej w formule analogicznej do poprzedniej:

```
= REGLINP(znane_y; (wd = TRANSPONUJ(wd))*1; FAŁSZ)
```

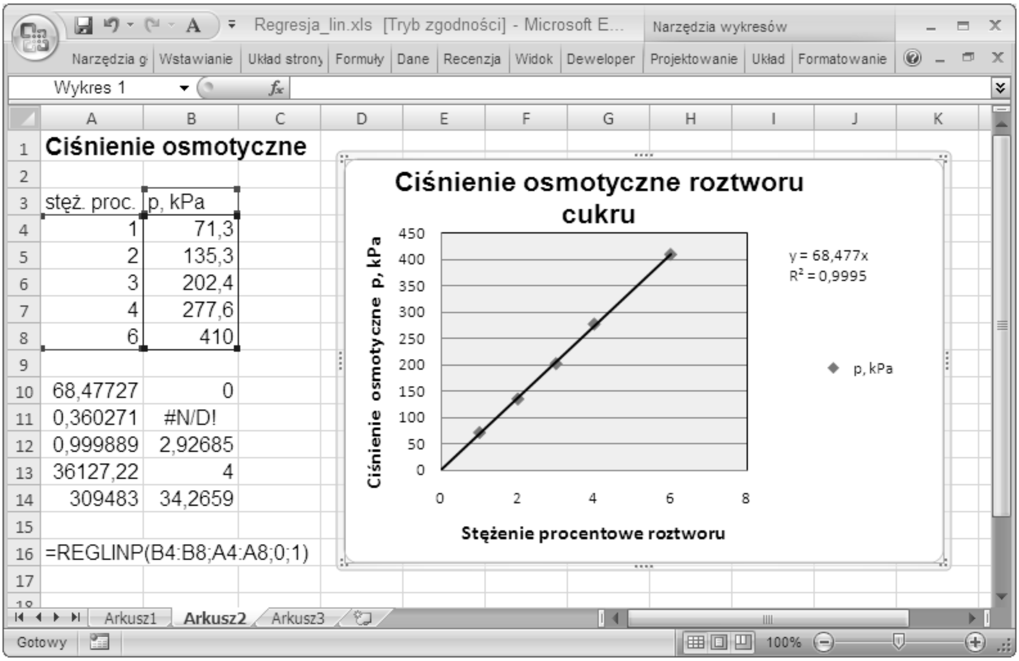
Wyniki uzyskane przy zastosowaniu funkcji `MACIERZ.JEDNOSTKOWA` są tablicami **TPW**, natomiast wyrażenia zawierające funkcję `TRANSPONUJ` wymagają w przypadku wykorzystania do dalszych obliczeń zatwierdzenia tablicowego.

Przykłady wyznaczania równań regresji

Przykład 7.1.

Dana jest zależność ciśnienia osmotycznego roztworu cukru od jego stężenia procentowego. Wykonamy wykres tej zależności i dopasujemy linię trendu — na wykresie wyświetlimy jej równanie i kwadrat współczynnika korelacji. Wykres powinien przechodzić przez początek układu współrzędnych.

Współczynnik proporcjonalności i współczynnik determinacji można wyznaczyć również za pomocą funkcji `REGLINP`. Niestety, jak już pisałem, w tym przypadku od wersji 2003 funkcja zwraca kwadrat niescentrowanego współczynnika korelacji zamiast tradycyjnego współczynnika. Z kolei w wersji 2007 Excela w przypadku korelacji wielomianem niekiedy nie jest wyświetlany współczynnik przy najwyższej potędze ***x***. (We wcześniejszych wersjach ten błąd nie występował). Dane i wyniki pokazano na rysunku 7.3. Dane do przykładu pochodzą z książki [10].



RYSUNEK 7.3. *Zależność ciśnienia osmotycznego roztworu cukru od jego stężenia procentowego jako przykład regresji liniowej*

Przy prowadzeniu obliczeń dotyczących regresji, począwszy od wersji 2007, można dodatkowo wykorzystać tabele Excela wraz z odwołaniami strukturalnymi. Nie uzyskamy przez to jakichś szczególnych ułatwień, lecz unikniemy kopiowania w kolumnach, które będzie się odbywać automatycznie, i nieco szybciej wprowadzimy nagłówki kolumn niż nazwy odpowiednich zakresów. Ponadto jeśli w jednym arkuszu chcemy umieścić kilka podobnych obliczeń, łatwiej zachować podobne schematy oznaczeń ze względu na to, że nagłówki kolumn w różnych tabelach mogą się powtarzać, a nazwy zakresów muszą być unikatowe w arkuszu.

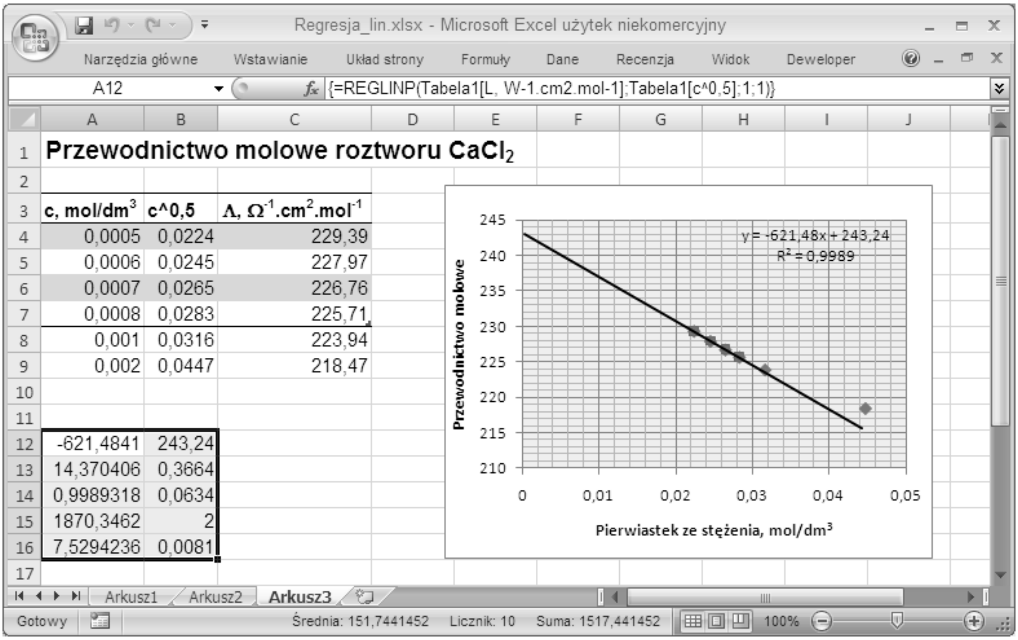
Przykład 7.2.

Dana jest zależność przewodnictwa molowego rozcieńczonych roztworów chlorku wapnia od stężenia molowego w temperaturze 25°C. (Dane pochodzą z książki [10]). Zgodnie z teorią rozcieńczonych roztworów elektrolitów przewodnictwo molowe powinno być liniową funkcją pierwiastka ze stężenia:

$$\Lambda = \Lambda_0 - k\sqrt{C}$$

Należy zweryfikować to równanie, obliczając jednocześnie jego współczynniki. Jeżeli chcemy wyznaczyć współczynniki na podstawie równania linii trendu z wykresu, to trzeba przygotować w arkuszu kolumnę z wartościami pierwiastka ze stężenia i wykorzystać tę kolumnę jako zmienną niezależną. W standardowym zestawie funkcji opisujących linię trendu nie przewidziano bowiem zależności tego typu. Korzystając z funkcji `REGLINP`, nie musimy przygotowywać kolumn pomocniczych. Można zastosować notację tablicową i jako *znane_x* podać `PIERWIASTEK(`***stężenie***`)` lub ***stężenie***`^0,5`.

Dane i wyniki pokazano na rysunku 7.4. Zależność liniowa jest dobrze spełniana przez cztery pierwsze punkty, dlatego tylko te punkty wykorzystano do obliczeń.



RYSUNEK 7.4. *Zależność przewodnictwa molowego rozcieńczonych roztworów chlorku wapnia od stężenia molowego w temperaturze 25 °C jako przykład regresji liniowej ze zmianą zmiennej niezależnej*

Przykład 7.3.

Adsorpcję par na stałych adsorbentach opisuje się często równaniem izotermy Langmuira:

$$x = \frac{abp}{1 + bp}$$

gdzie:

x — ułamek wagowy określający ilość substancji zaadsorbowanej na jednostkę masy sorbenta;

p — cząstkowa prężność pary substancji adsorbowanej w równowadze z adsorbentem.

Równanie izotermy Langmuira jest nieliniowe i przed jego opracowaniem metodą regresji liniowej trzeba je przekształcić do postaci liniowej względem nowych zmiennych.

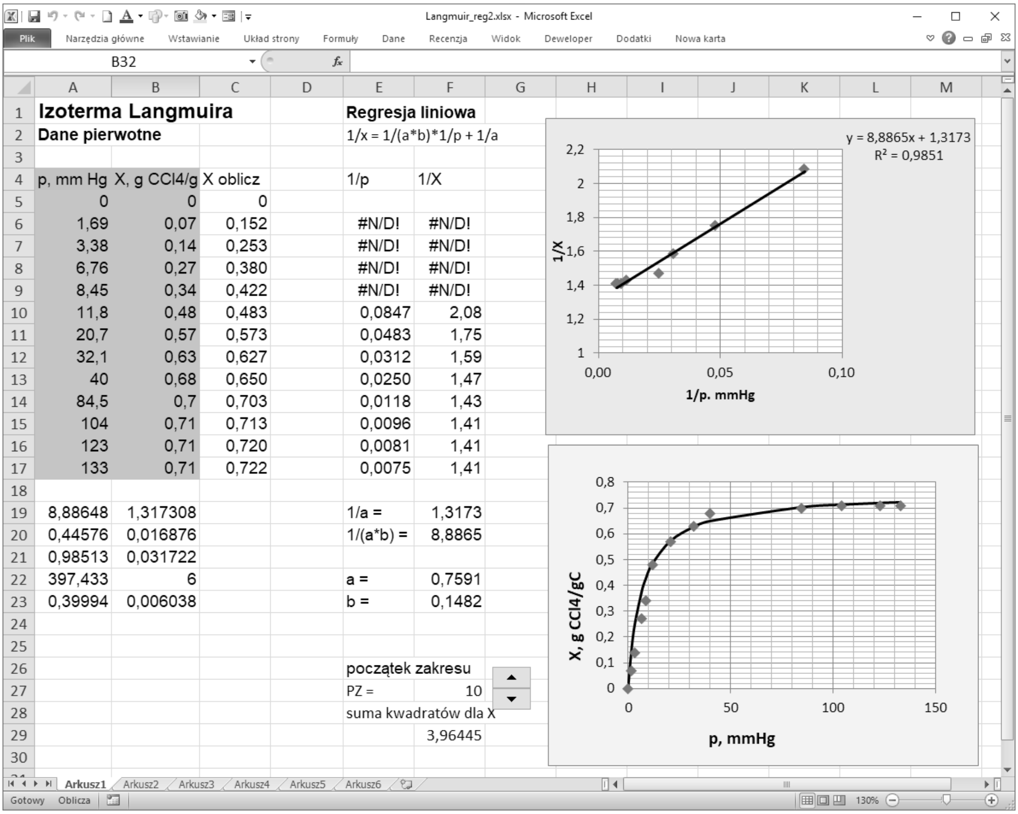
Sposoby przekształcenia podałem na początku rozdziału. Podobnie jak w poprzednim przykładzie, można albo przekształcić zmienne w arkuszu w sposób jawny, wykonać wykres linearyzowany i odczytać współczynniki z równania linii trendu, albo zapisać przekształcenie zmiennych w argumentach funkcji `REGLINP` i odczytać współczynniki z wyniku funkcji.

Należy pamiętać, że przekształcenie zmiennych zmienia wagi poszczególnych punktów pomiarowych, na przykład po obliczeniu $1/x$ małe wartości x zyskują na znaczeniu, a duże tracą.

Po wstawieniu obliczonych współczynników do pierwotnego równania jakość dopasowania równania ze względu na początkowe zmienne może być niezadowalająca. W takim przypadku pomaga czasem rezygnacja z kilku skrajnych punktów.

Przykładowe wyniki obliczeń przedstawiono na rysunku 7.5. Dane do tego przykładu pochodzą z książki [7]. W zaprezentowanym przykładzie przy obliczaniu współczynników równania linearyzowanego pominięto cztery pierwsze punkty, aby uzyskać zadowalające wizualnie dopasowanie linii trendu do punktów z końca zakresu — po powrocie do pierwotnych zmiennych. Aby można było łatwo modyfikować zakresy punktów branych pod uwagę w obliczeniach, zdefiniowano komórkę przechowującą numer wiersza z początkowymi danymi. Komórce tej nadano nazwę ***PZ*** (komórka `F27`). Wszystkie adresy potrzebne do zdefiniowania zakresów roboczych są konstruowane z użyciem funkcji `ADR.POŚR`, której argumentem jest adres w postaci tekstowej. Tekst ten może być złożony z elementów oznaczających kolumnę i wiersz. Alternatywnie można użyć funkcji `ADRES`, która również pozwala na konstrukcję adresu w formie tekstowej z numerów wiersza i kolumny. Jeżeli na przykład chcemy odwołać się do komórki z kolumny `B` (jest to druga kolumna) i z wiersza o numerze pobranym z komórki ***PZ***, to możemy użyć jednego z wyrażeń:

`= ADR.POŚR("B"&`***PZ***`)` lub `= ADR.POŚR(ADRES(`***PZ***`;2))`



RYSUNEK 7.5. *Stopień adsorpcji par czterochlorku węgla na węglu aktywnym jako przykład regresji według równania linearyzowanego*

Funkcja `ADR.POŚR` wymaga argumentu tekstowego, dlatego użyto operatora połączenia tekstów `&`. Zawartość komórki ***PZ*** jest liczbą, ale zostanie automatycznie skonwertowana na tekst. Funkcja `ADRES` [A3] wymaga co najmniej dwu argumentów liczbowych: pierwszym jest numer wiersza, drugim numer kolumny. Można podać jeszcze argumenty opcjonalne, ale nie będą nam tu potrzebne.

Do wyznaczenia wartości zmiennych pomocniczych $1/p$ i $1/X$ użyto wyrażeń typu:

```
= JEŻELI(WIERSZ()>=PZ;1/A6;BRAK())
```

Funkcja `WIERSZ()` bez argumentów zwraca numer wiersza z miejsca wywołania. Wyrażenie oblicza wartości $1/p$, w tym przypadku `1/A6`, jeżeli wywołanie jest w wierszu o numerze ***PZ*** lub większym, w przeciwnym razie zwraca kod `#N/D!`.

Współczynniki w linearyzowanym równaniu regresji ($1/ab$ i $1/a$) obliczano ze wzoru:

```
= REGLINP(1/ADR.POŚR("B"&PZ):B17;1/ADR.POŚR("A"&PZ):A17;PRAWDA;PRAWDA)
```

Jest to oczywiście formuła tablicowa.

W analogiczny sposób można przeprowadzić obliczenia dla pozostałych wzorów linearyzujących izotermę Langmuira.

Przykład 7.4.

Często dane doświadczalne zawierają punkty, co do których można mieć wątpliwości, czy wartości zostały poprawnie zmierzone. Chcielibyśmy móc sprawdzić, jak zmieniłyby się wyniki regresji po wyeliminowaniu wątpliwych punktów. Można te punkty trwale wyeliminować, ale korzystniej byłoby mieć możliwość ich tymczasowego (odwracalnego) zamaskowania. Gdy korzystamy z regresji (linii trendu) na wykresie, sprawa jest dość prosta: wystarczy przenieść wartość y do sąsiedniej kolumny, pozostawiając pustą komórkę — równanie linii trendu zostanie wyznaczone z pominięciem usuniętych punktów. Jednak kiedy obliczamy współczynniki równania regresji lub wartości prognozowane za pomocą funkcji, sprawa nie jest już taka prosta. Funkcje `REGLINP` i `REGLINW` (oraz podobne) nie tolerują pustych komórek w zakresach danych, natomiast funkcje `NACHYLENIE`, `ODCIĘTA`, `R.KWADRAT` i `REGLINX` radzą sobie z tym problemem doskonale, ale za to zakres ich zastosowania nie jest tak szeroki jak wymienionych wcześniej funkcji. (Opisy funkcji `REGLINW` i `REGLINX` podałem w następnym rozdziale).

Jeśli chcemy zamaskować niektóre punkty w zakresie danych funkcji `REGLINP` lub `REGLINW`, musimy zastosować procedurę bardziej złożoną i utworzyć nowe ciągłe zakresy danych, w których zamaskowane punkty będą pominięte.

Zakresom pierwotnych danych nadajemy nazwy ***x*** i ***y***. Zakładamy, że dane są zapisane w kolumnach. Wartości y, które mają być zamaskowane, przenosimy do sąsiedniej kolumny. Sprawdzamy liczebność danych (`ILE.LICZB(y)`) i tworzymy listę z numerami wierszy w zakresie ***y***. Wykorzystujemy do tego znaną konstrukcję z funkcją `WIERSZ` (tablicę bazową):

```
z : = WIERSZ(ADR.POŚR("1:"&ILE.LICZB(y)))
```

Ustalamy położenie zamaskowanych punktów w zakresie ***y***. W tym celu użyjemy funkcji `PODAJ.POZYCJĘ`. Pierwszej pustej komórki szukamy za pomocą formuły tablicowej:

```
= PODAJ.POZYCJĘ("";y&"";0)
```

Alternatywnie można użyć formuły zwykłej z funkcją `INDEKS`:

```
= PODAJ.POZYCJĘ("";INDEKS(y&"";);0)
```

Pusta komórka musi być poszukiwana jako pusty tekst, po przekształceniu zakresu ***y*** w teksty. Służy do tego operacja łączenia zakresu ***y*** z pustym tekstem.

Aby znaleźć położenie kolejnych pustych komórek, trzeba zmodyfikować zakres do przeszukania tak, by zaczynał się za znalezioną pustą komórką. Odpowiednia formuła (tablicowa) może mieć postać:

```
= PODAJ.POZYCJĘ(""; PRZESUNIĘCIE(y;p;0)&""; 0)+p
```

W tej formule ***p*** oznacza odwołanie (względne) do pozycji poprzedniej pustej komórki. Wzór można kopiować w dół lub w prawo; każde jego użycie wskazywać będzie kolejne puste komórki. Gdy uzyskamy wynik większy od długości zakresu ***y***, będzie to oznaczać koniec zakresu i zarazem brak dalszych pustych komórek.

Wzór można też zmodyfikować tak, aby po osiągnięciu końca zakresu sygnalizował błąd:

```
= JEŻELI.BŁĄD(PODAJ.POZYCJĘ(""; PRZESUNIĘCIE(y;p;0;ILE.WIERSZY(y)–p)&""; 0)+p; "koniec")
```

Dalej można wykorzystać funkcję `MACIERZ.ILOCZYN`. Jeśli dane ***x*** i ***y*** są w kolumnach, to numery punktów, które chcemy zamaskować, umieścimy w wierszu i nazwiemy ten wiersz ***wp***. Zdefiniujemy nową tablicę bazową (pionową) opartą na numerach punktów zamaskowanych i nadamy jej nazwę ***wz***:

```
wz : = WIERSZ(ADR.POŚR("1:"&ILE.LICZB(wp)))
```

Po zdefiniowaniu formuły indeksowej:

```
fi : = z+MACIERZ.ILOCZYN(––(z>wp–TRANSPONUJ(wz)); wz^0)
```

wartości współrzędnych punktów niezamaskowanych otrzymamy ze wzorów:

```
= INDEKS(x; fi)
= INDEKS(y; fi)
```

Te wyniki mają postać tablic **TDW**. Piszę o tym dokładniej w rozdziale 20. Aby użyć ich jako argumentów funkcji `REGLINP` lub `REGLINW`, należy wykorzystać jeden ze sposobów podanych we wspomnianym rozdziale.

Według sposobu 1. można użyć zapisu:

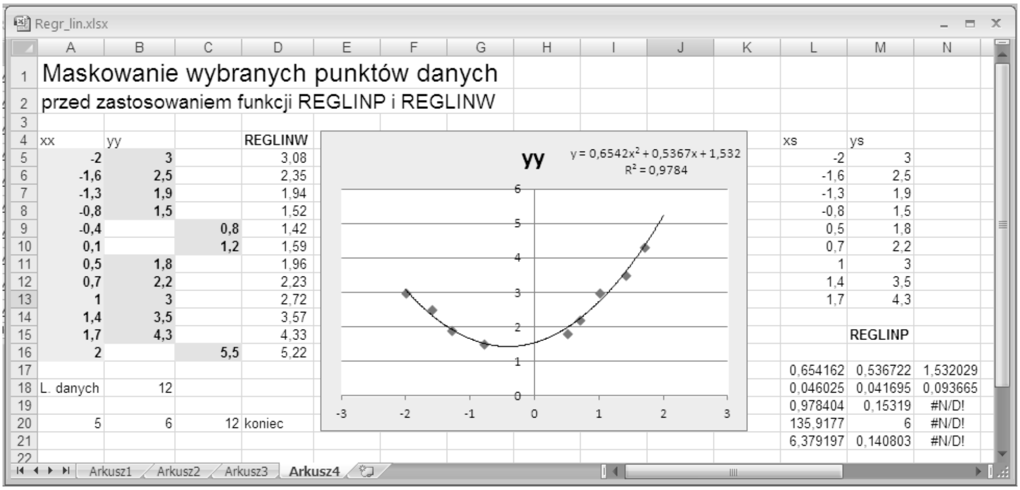
```
= REGLINP(+INDEKS(y; fi); INDEKS(x; fi)^{1;2}; ; PRAWDA)
```

Według sposobu 2. należy obudować wyrażenie indeksowe ***fi*** funkcją `L` (lub `N` w wersji 2010+):

```
= REGLINP(INDEKS(y; L(fi)); INDEKS(x; L(fi))^{1;2}; ; PRAWDA)
```

Funkcja `INDEKS` w takiej formie zwraca wynik pełnowartościowy, który może zostać użyty jako argument funkcji `REGLINP`, ale lepiej nadać formule nazwę, by potem móc ją wykorzystać również do zdefiniowania wykresu.

Przykład obliczeń wykonanych w ten sposób pokazano na rysunku 7.6.



RYSUNEK 7.6. *Maskowanie wybranych punktów pomiarowych w celu umożliwienia obróbki danych za pomocą funkcji REGLINP i REGLINW*

Można ten problem rozwiązać również inaczej. Po usunięciu wybranych punktów z zakresu ***y*** tworzymy listę indeksów pozostałych (nieusuniętych) punktów. Użyjemy do tego formuły, której można nadać nazwę ***zs***:

```
zs : = MIN.K(JEŻELI(y<>"";WIERSZ(y)–KOMÓRKA("wiersz";y)+1);
↪WIERSZ(ADR.POŚR("1:"&ILE.LICZB(y))))
```

Formuła ta zastosowana jako argument funkcji `INDEKS` utworzy listę współrzędnych nieusuniętych punktów. Listy tej można użyć w funkcji `REGLINP`. Przykładowo wywołanie funkcji może mieć postać:

```
= REGLINP(INDEKS(y;zs)*1;INDEKS(x;zs)^{1;2};;PRAWDA)
```

Wynikiem funkcji mają być współczynniki równania kwadratowego. Pierwszy argument funkcji `REGLINP` jest pomnożony przez 1, aby wyeliminować błąd wynikający z zastosowania funkcji `INDEKS` z argumentem tablicowym. W przypadku drugiego argumentu rolę tę spełnia podniesienie do potęgi.

Przykład 7.5.

Do problemu przedstawionego w poprzednim przykładzie można podejść również z drugiej strony. Zamiast usuwać niepotrzebne punkty można stworzyć listę numerów (indeksów) punktów, które chcemy pozostawić do analizy lub do prezentacji na wykresie. Sposób ten będzie wygodniejszy, gdy mamy stosukowo dużo punktów do dyspozycji i chcemy wybrać tylko niektóre. Wyboru można dokonać tradycyjnie, za pomocą funkcji INDEKS, ale takich przykładów było już dużo, więc tym razem wykorzystamy pseudozakresy. Wyrażenia te omawiam dość szczegółowo w rozdziale 21. Celem jest uzyskanie formuły pełnowartościowej, którą można wykorzystać zarówno do zdefiniowania wykresu, jak i do obliczeń za pomocą funkcji z grupy REGLINP.

Załóżmy, że pierwotne dane znajdują się w kolumnach, których pierwsze komórki oznaczono nazwami ***px*** i ***py***. Utworzymy dodatkową kolumnę, ***wyb*** (można ją zdefiniować z zapasem), przeznaczoną do wpisywania indeksów punktów, które mają być uwzględnione w obliczeniach. Pierwszą komórkę tego zakresu oznaczymy jako ***pwyb***. Należy teraz zdefiniować w *Menedżerze nazw* kilka formuł pomocniczych. Pierwsza definiuje listę wybranych punktów:

lista := PRZESUNIĘCIE(***pwyb***;;;ILE.NIEPUSTYCH(***wyb***))–1

Zakładam, że indeksy punktów w kolumnie ***wyb*** będą liczone od 1, jak w funkcji INDEKS, ponieważ jednak zamierzam wykorzystać funkcję PRZESUNIĘCIE, która liczy od 0, stąd potrzeba odjęcia jedynki w powyższym wzorze.

Jak już wspomniałem, do utworzenia roboczych (przefiltrowanych) tablic wartości ***x*** i ***y*** wykorzystamy pseudozakresy bazujące na funkcji PRZESUNIĘCIE. Mamy do wyboru trzy równoważne formuły, analogiczne do obu zmiennych. Oto te trzy warianty dla zmiennej ***x*** (w przypadku zmiennej ***y*** wystarczy zamienić ***px*** na ***py***):

x.1 := L(PRZESUNIĘCIE(***px***;***lista***;0))

x.2 := SUMA.JEŻELI(PRZESUNIĘCIE(***px***;***lista***;0);"<1E100")

x.3 := SUMY.CZĘŚCIOWE(9;PRZESUNIĘCIE(***px***;***lista***;0))

W wariancie drugim należy podać warunek, który będzie spełniony przez wszystkie wybrane punkty.

Tak zdefiniowane tablice ***x*** i ***y*** mogą zostać wykorzystane zarówno w formułach, jak i przy określaniu danych do wykresów.

Jeszcze jeden sposób rozwiązania tego problemu przedstawiam w następnym rozdziale, w podrozdziale „Regresja danych o niejednakowej dokładności”.

Skorowidz

A

B

C

D

F

N

O

P

R

S

Ś

T

U

Z


PROGRAM
PARTNERSKI
GRUPY HELION
1. ZAREJESTRUJ SIĘ
2. PREZENTUJ KSIĄŻKI
3. ZBIERAJ PROWIZJĘ
Zmień swoją stronę WWW w działający bankomat!
Dowiedz się więcej i dołącz już dzisiaj!
http://program-partnerski.helion.pl
GRUPA
Helion

Obliczenia i wykresy w Excelu? Nic trudnego!

- **Wykonuj obliczenia**
- **Twórz wykresy**
- **Zrozum Excela**

Na rynku nie brakuje książek opisujących obsługę i zastosowania arkusza kalkulacyjnego MS Excel, żadna jednak nie wprowadzi Cię w tę tematykę tak skutecznie jak ta! Omiń rafy i białe plamy dokumentacji, skorzystaj z doświadczenia autora i śmiało wkrocz w świat niesamowitych możliwości Excela.

Książka przedstawia sposoby tworzenia i zastosowania wykresów w Excelu. Przybliża zagadnienie zaokrąglania liczb, a także prezentuje metody przeprowadzania w arkuszu kalkulacyjnym różnego rodzaju obliczeń matematycznych, w tym różniczkowania, całkowania, rozwiązywania równań i układów równań oraz przetwarzania liczb zespolonych.

- **Graficzna reprezentacja danych**
- **Zaokrąglanie liczb i obliczenia przybliżone**
- **Równania i układy równań nieliniowych**
- **Regresja i interpolacja**
- **Obliczanie całek oznaczonych**
- **Równania różniczkowe zwyczajne i cząstkowe**
- **Szeregi potęgowe i kombinatoryka**
- **Konwersja liczb i jednostek**
- **Liczby i funkcje zespolone**
- **Przetwarzanie tekstów, dat i czasów**

Dowiedz się, jak wykorzystać Excela do obliczeń i wizualizowania wyników!

HELION SA
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 98 63
helion@helion.pl

ISBN 978-83-283-5693-1
9 788328 356931