

Rozdział 4.

Dopasowywanie krzywych

Nie ma chyba nic powszechniejszego w pracy inżyniera niż dopasowywanie danych uzyskanych na drodze eksperymentu do funkcji analitycznej, i na odwrót. Pozwala to inżynierom i naukowcom chociażby na wyznaczanie linii trendu i na badanie przebiegu badanych zjawisk. Czasami dopasowywanie krzywych jest dość proste (choćby w przypadku funkcji liniowych), ale niejednokrotnie zdarza się, że ich wyznaczenie jest nie lada sztuką.

Jeżeli z jakiegoś względu nie można znaleźć odpowiedniej krzywej pasującej do danego zbioru punktów, zawsze można zastosować mechanizm interpolacji liniowej i spróbować zagęścić liczbę punktów, dzięki czemu krzywą będzie łatwiej znaleźć. Jednak w większości przypadków Excel umożliwi użytkownikowi znalezienie właściwej krzywej. Istnieje na to kilka różnych sposobów: wykorzystanie regresji liniowej i wielomianowej, rysowanie linii trendu za pomocą standardowych narzędzi dostępnych w Excelu. Można również zaprzęgnąć do pracy narzędzie Solver umożliwiające w sposób automatyczny znalezienie wyniku.

Pliki referencyjne

Wszystkie przykłady zamieszczone w tym rozdziale dostępne są na serwerze pod adresem <ftp://ftp.helion.pl/przyklady/excinz.zip>. Po rozpakowaniu archiwum należy znaleźć plik o nazwie *Roz_4.xlsx*. Ponadto archiwum zawiera plik *Odp_4.xlsx* z odpowiedziami do pytań kontrolnych znajdujących się na końcu tego rozdziału.

4.1. Regresja liniowa

W Excelu dostępne są wbudowane funkcje do obliczania regresji liniowej. Są to funkcje o nazwach:

- ◆ NACHYLENIE — funkcja zwraca nachylenie wykresu regresji liniowej,
- ◆ REGLINP — dopasowuje linię prostą na podstawie istniejących danych,

- ◆ REGLINX — prognozuje wartości, na podstawie dostępnych danych
- ◆ REGLINW — prognozuje wartości, które pasują do prostej trendu liniowego.

Do wyznaczania prostej aproksymującej za pomocą mechanizmu regresji liniowej służy więc funkcja o nazwie REGLINP.



Wskazówka

W wersji 2003 programu Microsoft Excel opisywane tutaj funkcje dostępne są dopiero po zainstalowaniu dodatku Analysis ToolPak. Dotyczy to również funkcji do obliczania regresji wykładniczej. W wersji 2007 funkcje te zostały zaimplementowane bezpośrednio do arkusza.

Do pokazania wbudowanych mechanizmów regresji posłużymy się danymi pomiarowymi uzyskanymi w wyniku badania przyspieszenia ziemskiego. Szerzej o tym problemie traktuje rozdział 9. Dane uzyskane w wyniku eksperymentu przedstawione są na rysunku 4.1.

Rysunek 4.1.

Tabela arkusza Excel z danymi uzyskanymi dzięki eksperymentalnemu wyznaczaniu przyspieszenia ziemskiego

	A	B	C	D
1	Pomiary przyspieszenia ziemskiego			
2				
3	t	x	v	g
4	s	m	m/s	m/s ²
5	0,0	0,00000	0,00000	0,00000
6	0,1	0,05005	1,04082	9,10490
7	0,2	0,19855	1,99478	9,97443
8	0,3	0,44487	2,98817	9,89329
9	0,4	0,88811	4,04995	11,34242
10	0,5	1,33152	5,02198	8,09806
11	0,6	1,87255	5,90978	9,65791

Z lekcji fizyki ze szkoły średniej wiemy, że przyspieszenie ziemskie jest przyspieszeniem stałym. W związku z tym prędkość w danym momencie wyraża się wzorem: $V = V_p + gt$, gdzie g oznacza przyspieszenie ziemskie. Właśnie dzięki temu wiemy, jaki charakter powinna mieć prosta, dzięki czemu będziemy w stanie zdiagnozować, czy dane zostały zebrane w sposób poprawny. Oczywiście w rzeczywistości narzędzia regresji używa się w przypadku, kiedy nie znamy równania opisującego nasze dane eksperymentalne i chcemy wyznaczyć równanie prostej będącej przybliżeniem rzeczywistej charakterystyki.

Aby przeprowadzić regresję liniową przy wykorzystaniu jednej z dostępnych funkcji Excela, przygotowujemy następujący arkusz:

- ◆ w komórce A1 wpisujemy tytuł Regresja liniowa;
- ◆ w komórce A3 wpisujemy t [s];
- ◆ w komórce B3 wpisujemy V [m/s];
- ◆ w komórkach A4:A10 oraz B4:B10 wpisujemy kolejne wartości czasu oraz prędkości zgodnie z danymi pokazanymi na rysunku 4.1;
- ◆ w komórce D3 wpisujemy tytuł Współczynniki regresji;

- ♦ w komórkach *D4* i *E4* wpisujemy odpowiednio *A* oraz *B* — będą to współczynniki funkcji regresji $y=ax+b$;
- ♦ zaznaczamy komórki *D5* i *E5* i wpisujemy formułę tablicową $=\text{REGLINP}(B4:B10;A4:A10;\text{PRAWDA};\text{FAŁSZ})$; aby wprowadzić ją do obu wierszy, musimy wcisnąć kombinację klawiszy *Ctrl+Shift+Enter*;
- ♦ dokonujemy formatowania komórek, dodajemy obramowanie, kolorujemy komórki, wyrównujemy itp.

Gotowy arkusz powinien wyglądać tak, jak ten pokazany na rysunku 4.2.

Rysunek 4.2.

Arkusz z obliczonymi współczynnikami regresji

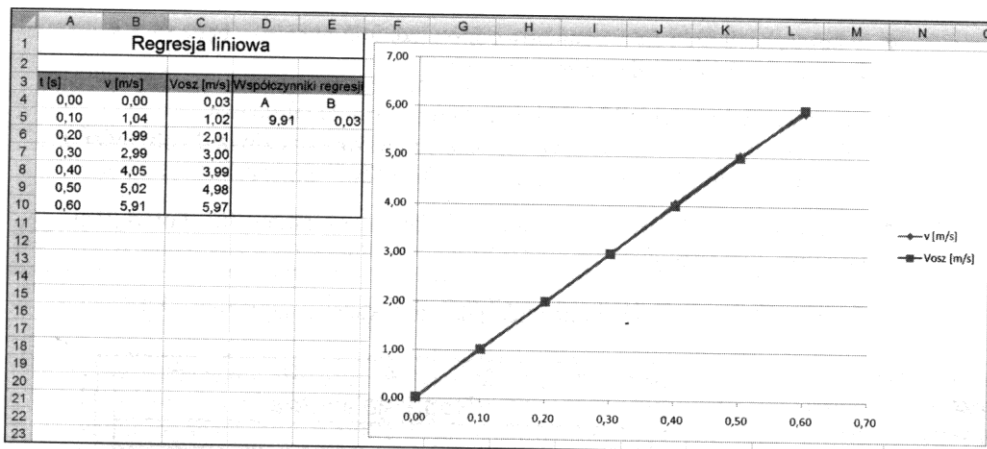
	A	B	C	D	E
1	Regresja liniowa				
2					
3	t [s]	v [m/s]		Współczynniki regresji	
4	0,00	0,00		A	B
5	0,10	1,04		9,91	0,03
6	0,20	1,99			
7	0,30	2,99			
8	0,40	4,05			
9	0,50	5,02			
10	0,60	5,91			
11					

Funkcja **REGLINP** ma cztery argumenty:

- ♦ Pierwszym z nich jest tablica znanych wartości *Y*.
- ♦ Drugim jest tablica znanych wartości *X*.
- ♦ Trzecim jest stała przyjmująca wartość **PRAWDA** lub **FAŁSZ**. Wybranie opcji **PRAWDA** oznacza, że wartość współczynnika *B* będzie niezerowa. **FAŁSZ** oznacza, że wymuszona zostanie wartość $B=0$.
- ♦ Czwarty argument również przyjmuje wartość **PRAWDA** lub **FAŁSZ**. Wybranie opcji **PRAWDA** spowoduje, że dodatkowo obliczone zostaną statystyki regresji, czyli standardowe wartości błędu dla kolejnych współczynników, współczynnik wyznaczenia, standardowy błąd oceny wartości *Y*, stopień swobody oraz regresyjna i resztkowa suma kwadratów. Oczywiście jeżeli wybralibyśmy tę opcję, tablica wyników powinna mieć wymiar 2×5 , a nie 2×2 .

Dodajmy teraz kolumnę *Vosz*, w której wyliczymy wartości prędkości przy wykorzystaniu wyliczonych wartości współczynników regresji liniowej. Następnie sporządzimy wykres (rysunek 4.3), na którym sprawdzimy, w jakim stopniu dokładna jest nasza regresja.

- ♦ W komórce *C3* wpisujemy *Vosz* [m/s].
- ♦ W komórce *C4* wpisujemy formułę $=\$D\$5*\$A4+\$E\$5$ i kopiujemy ją do komórek *C5:C10*.
- ♦ Na podstawie danych z komórek *A4:C10* rysujemy wykres punktowy (*X Y*).



Rysunek 4.3. Wykres funkcji przyspieszenia przy swobodnym spadaniu ciała

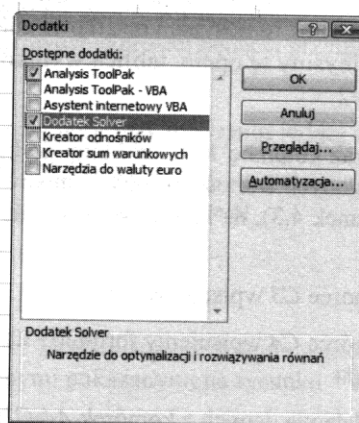
Zgodnie z oczekiwaniami obie zależności są w przybliżeniu liniowe. Co ważne, regresja liniowa doskonale odwzorowuje dane eksperymentalne, tak więc może nam swobodnie posłużyć do oszacowania wartości, poza tymi, które już znamy. Wystarczyłoby przedłużyć prostą regresji i odczytać odpowiednie wartości. Można również wykorzystać wbudowaną funkcję Excela o nazwie `REGLINW`.

Regresję można przeprowadzić również inaczej. W tym celu należy uruchomić dodatek Analysis ToolPak (jeśli jeszcze nie jest włączony). Aby to uczynić, należy wykonać następującą sekwencję czynności:

- ◆ kliknąć przycisk *Office* (znajdujący się w lewym górnym rogu Excela),
- ◆ kliknąć przycisk *Opcje programu Excel* i przejść do opcji *Dodatki*,
- ◆ w oknie po prawej stronie kliknąć przycisk *Przejdź*, a następnie zaznaczyć dodatek *Analysis ToolPak* i kliknąć przycisk *OK*.

W tym miejscu można włączyć również inne dodatki, takie jak Solver, o którym mowa będzie w dalszej części tego rozdziału.

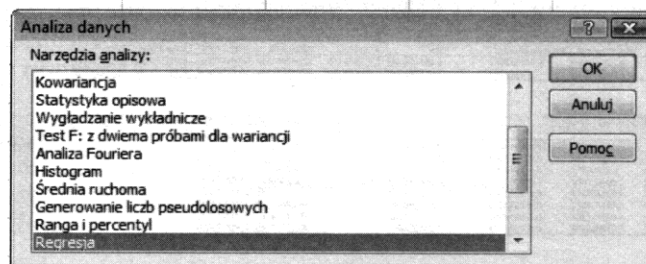
Rysunek 4.4.
Włączanie dodatku
Analysis ToolPak



Gdy dodatek Analysis ToolPak jest już aktywny (włączony), możemy przejść do przeprowadzenia regresji liniowej. Aby ją wykonać, należy przejść na wstążkę *Dane* i kliknąć przycisk *Analiza danych*. Pojawi się okno widoczne na rysunku 4.5, z którego wybieramy opcję *Regresja* i klikamy przycisk *OK*.

Rysunek 4.5.

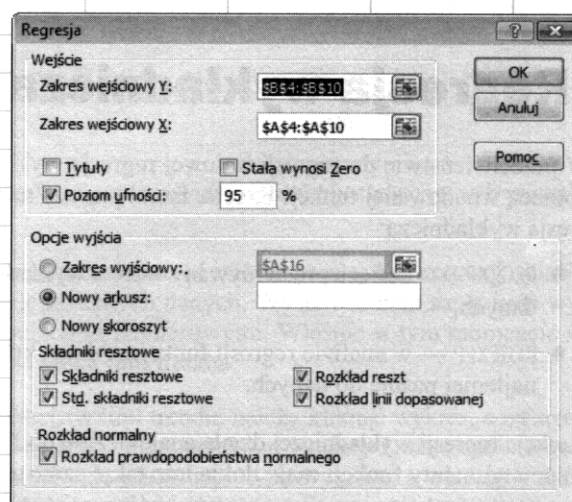
Okno wyboru typu analizy danych



Spowoduje to wyświetlenie kolejnego okna, tym razem o nazwie *Regresja*, w którym musimy wprowadzić zakres danych *X* oraz *Y* i zaznaczyć opcje regresji, które chcemy zastosować. Ilustruje to rysunek 4.6.

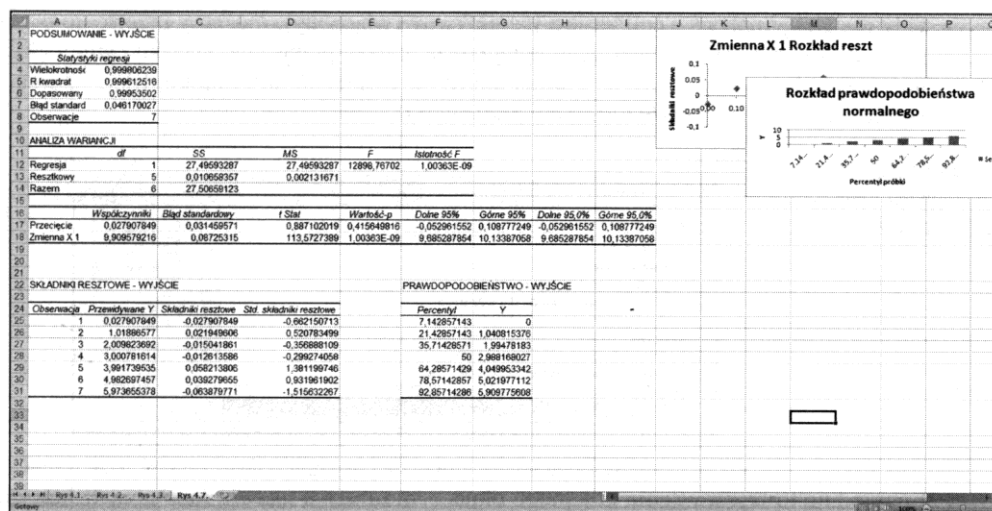
Rysunek 4.6.

Wprowadzenie danych wejściowych regresji liniowej



W wyniku działania tego mechanizmu powstanie nowy arkusz z wyliczonymi współczynnikami regresji oraz składnikami resztowymi. Ponadto wyrysowane zostaną wykresy rozkładu reszt oraz prawdopodobieństwa normalnego. Całość arkusza pokazana jest na rysunku 4.7.

W obu przypadkach wartość współczynników po zaokrągleniu jest taka sama. Wynika to stąd, iż w obu przypadkach wykorzystany został dokładnie ten sam algorytm obliczeń. Wybór metody zależy więc tak naprawdę od tego, czy potrzebujemy jedynie współczynników regresji, czy potrzebna jest nam rozbudowana analiza wyników przeprowadzonej regresji liniowej.



Rysunek 4.7. Arkusz Excela będący wynikiem analizy regresji liniowej przeprowadzonej za pomocą dodatku Analysis ToolPak

4.2. Regresja wykładnicza

W przeciwieństwie do regresji liniowej regresja wykładnicza możliwa jest jedynie za pomocą wbudowanej funkcji Excela. Excel posiada następujące funkcje związane z regresją wykładniczą:

- ◆ REGEXPW — oblicza przewidywany wzrost wykładniczy, używając istniejących danych,
- ◆ REGEXPP — w analizie regresji funkcja oblicza krzywą wykładniczą, która najlepiej pasuje do danych.

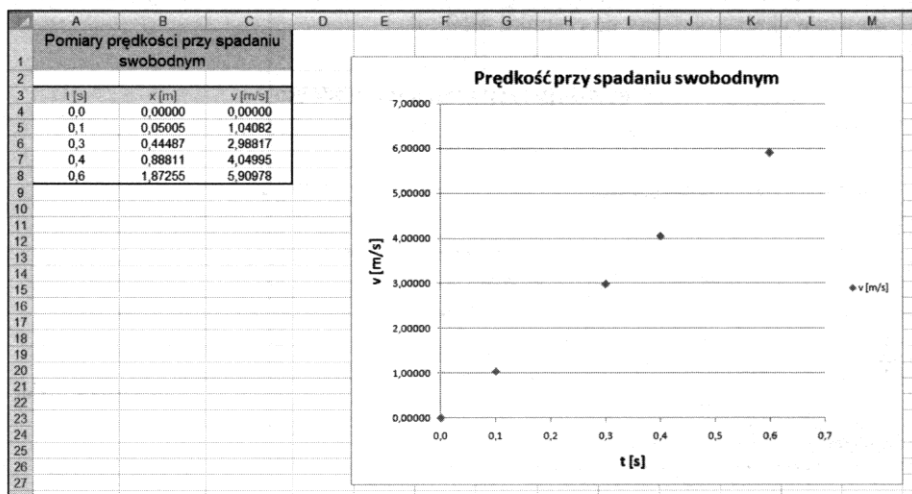
Funkcja regresji wykładniczej działa analogicznie do funkcji regresji liniowej. Poszczególne argumenty funkcji mają dokładnie takie samo znaczenie. Jedyna różnica polega na tym, że funkcja regresji ma postać ogólną $y = ba^x$, gdzie a i b są współczynnikami regresji.

Ze względu na fakt, iż regresja wykładnicza nie może być przeprowadzana, jeżeli któraś z wartości jest zerowa, w formule dane znane (X oraz Y) nie mogą zawierać wartości zerowej. Jeżeli tak się stanie, Excel zwróci błąd LICZBA!.

Korzystając z mechanizmów regresji wykładniczej, należy mieć świadomość, że dopasowanie krzywej będzie tym lepsze, im bardziej wykładniczy jest charakter danych wejściowych.

4.3. Linie trendu

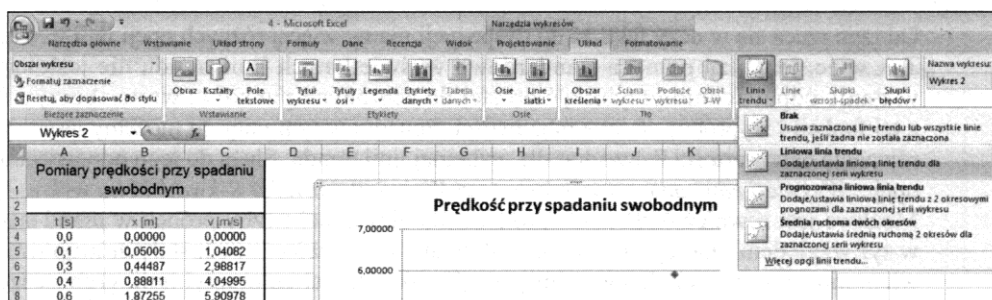
Excel daje również możliwość definiowania linii trendu, kiedy wykres jest już gotowy. Ponownie rozważmy wykorzystany wcześniej przykład z obliczaniem prędkości przy spadaniu swobodnym. Załóżmy, że dysponujemy mniejszą liczbą punktów pomiarowych, które rozmieściliśmy na wykresie punktowym. Tabela z danymi oraz wykres przedstawione są na rysunku 4.8.



Rysunek 4.8. Wyniki pomiarów prędkości przy swobodnym spadaniu ciała oraz wykres punktowy prezentujący je na osi czasu

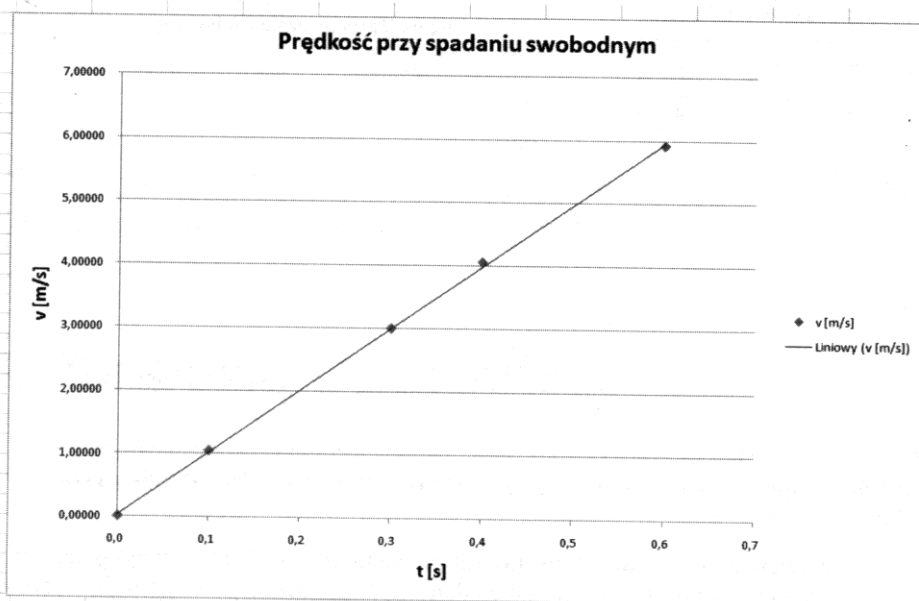
Często zdarza się, że nie interesuje nas równanie prostej lub krzywej opisującej trend wyznaczony na podstawie zebranych danych. Wystarcza nam sama linia wyrysowana na wykresie pomiędzy punktami pomiarowymi. Właśnie w tym momencie należy sięgnąć do mechanizmu rysowania linii trendu.

Aby skorzystać z dobrodziejstw linii trendu, należy kliknąć wykres, a następnie przejść na wstążkę *Układ*. Znajduje się na niej przycisk rozwijany *Linia trendu*. Jego kliknięcie pokaże nam dostępne linie trendu dla danego typu wykresu, a właściwie danych, które się na nim znajdują. Ilustruje to rysunek 4.9.



Rysunek 4.9. Opcje linii trendu dostępne dla wykresu punktowego prędkości przy spadaniu swobodnym

W tym przypadku dostępne są trzy linie trendu: liniowa linia trendu, prognozowana linia trendu oraz średnia ruchoma dwóch okresów. Wybór opcji zależy będzie oczywiście od naszych intencji. Jeżeli chcemy pokazać jedynie trend w zakresie zebranych danych pomiarowych, wówczas musimy wybrać pierwszą z możliwości. Jeśli chcemy na podstawie trendu wyznaczyć wartość spoza przedziału znanych wartości, wówczas musimy wybrać opcję numer dwa. Natomiast jeżeli zamierzamy maksymalnie przybliżyć wzorzec danych i wygładzić fluktuacje zebranych danych, to najlepszym wyborem jest opcja średniej ruchomej. W naszym przypadku mamy za mało danych, by ją wykorzystać, dlatego skorzystamy z pierwszej możliwości. Wynik jej działania przedstawia wykres pokazany na rysunku 4.10.

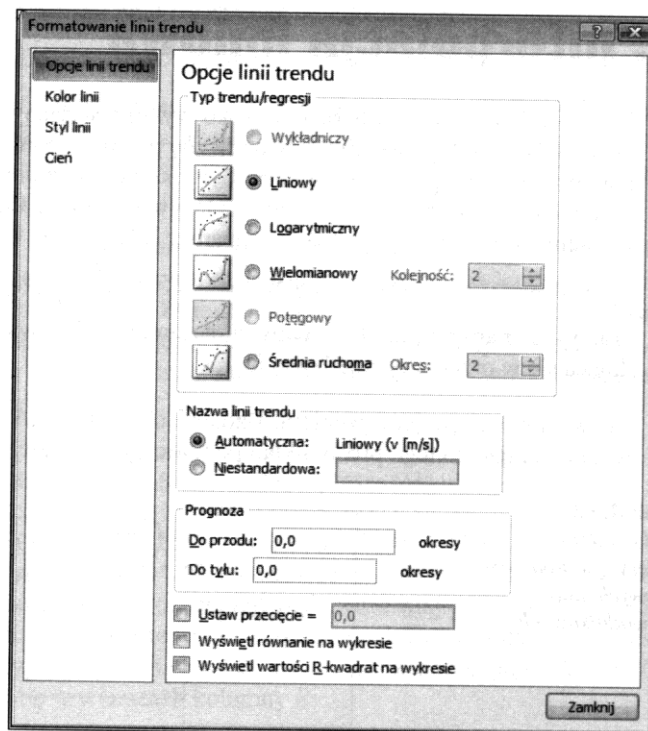


Rysunek 4.10. Wyrysowana linia trendu

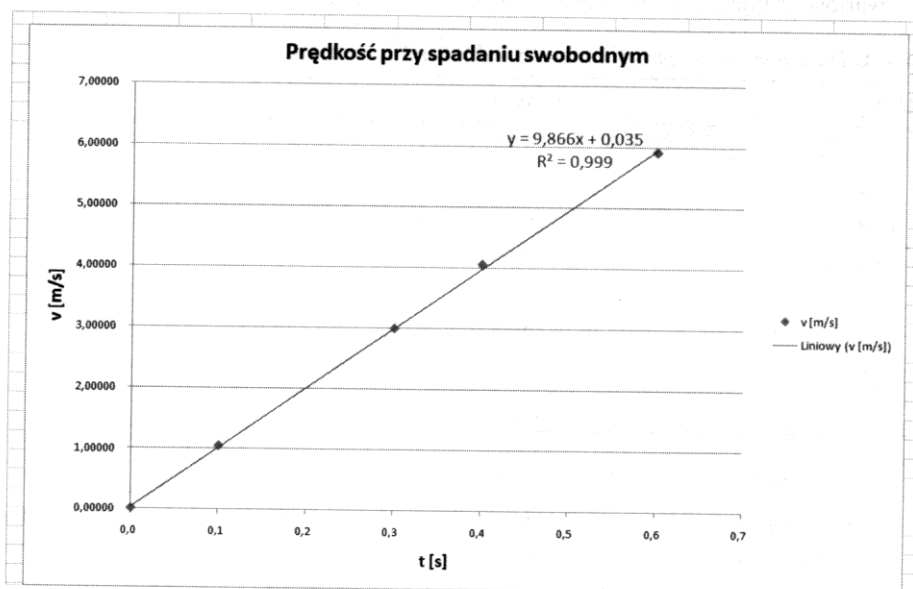
To jednak nie wszystkie możliwości, jakie niosą ze sobą linie trendu. Ostatnią z opcji, która dostępna jest po kliknięciu przycisku *Linia trendu* znajdującego się na wstążce *Układ*, jest *Więcej opcji linii trendu*. Wyświetla ona okno *Formatowanie linii trendu* pokazane na rysunku 4.11. Służy ono do personalizowania linii trendu.

Od razu rzuca się w oczy fakt, że do dyspozycji mamy więcej różnych opcji linii trendu. Tak więc, jeżeli za pomocą standardowo wyświetlanych linii trendu nie jesteśmy w stanie odpowiednio przybliżyć prostej do zebranych punktów, warto skorzystać z tej funkcji i wypróbować inne możliwości. Ponadto w tym miejscu możemy umieścić równanie oraz wartości R-kwadrat dla wybranej linii trendu. Służą do tego odpowiednio opcje: *Wyświetl równanie na wykresie* oraz *Wyświetl wartości R-kwadrat na wykresie*. Możliwe jest również wykonanie prognoz dla określonej liczby próbek w przód oraz w tył względem znanego przedziału. Ponadto korzystając z pozostałych grup ustawień, możemy spersonalizować linię trendu: określić jej kolor, styl linii, grubość itp.

Rysunek 4.11.
Okno Formatowanie
linii trendu



Rysunek 4.12. przedstawia poprzedni wykres z linią trendu z naniesionymi równaniami krzywej oraz błędem R-kwadrat, a także z kilkoma ustawieniami formatującymi linię trendu.



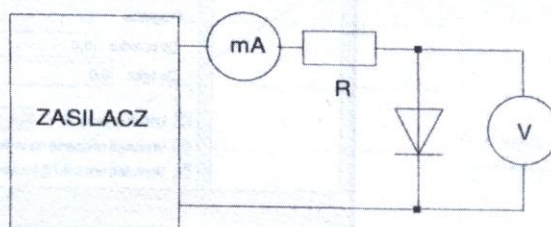
Rysunek 4.12. Sformatowana linia trendu

4.4. Interpolacja liniowa

Interpolacja liniowa jest bardzo szeroko wykorzystywanym mechanizmem przybliżania danych. Jeżeli posiadamy pewną liczbę punktów pomiarowych i chcemy obliczyć punkty pośrednie po to, aby wykres był jeszcze dokładniejszy, wówczas interpolacja liniowa jest doskonałym narzędziem. Oczywiście im znane punkty pomiarowe są bliżej siebie, tym lepiej, gdyż błąd związany z interpolacją będzie mniejszy. Interpolacja liniowa wyraża się następującym wzorem: $y = \frac{(x-x_2)}{(x_1-x_2)}y_1 + \frac{(x-x_1)}{(x_2-x_1)}y_2$, gdzie x_1 , x_2 , y_1 oraz y_2 są znanymi punktami danych, natomiast x jest wartością interpolowaną znajdującą się w przedziale między x_1 i x_2 .

Rozważmy następujący problem. Badamy charakterystyki prądowo-napięciowe diod, wykorzystując do tego prosty układ pomiarowy pokazany na rysunku 4.13.

Rysunek 4.13.
Układ do badania
charakterystyk prądowo-
napięciowych diod
półprzewodnikowych



W czasie naszego eksperymentu zebraliśmy pewną liczbę danych, które zestawione są w tabeli 4.1. Jak widać, pomiary zostały przeprowadzone krokowo w przedziale od 1,6 do 1,74 V. Naszym zamiarem jest ujednolicenie tabeli tak, aby posiadać dane dla wartości napięcia zwiększającego się co 0,01 V, a nie co 0,02 V.

Tabela 4.1. Dane pomiarowe przy badaniu charakterystyki prądowo-napięciowej diody

U [V]	I [mA]
1,6	3,98
1,62	5,19
1,64	6,16
1,66	8,58
1,68	10,9
1,7	13,09
1,72	16,07
1,74	18,05

Aby zrealizować to zadanie, musimy:

- ◆ W komórce A1 wprowadzić tytuł arkusza: Interpolacja danych do charakterystyk prądowo-napięciowych diody.
- ◆ W komórkach A3:B11 wprowadzić dane z tabeli 4.1.

- ♦ W komórce D3 wpisać tytuł Interpolacja liniowa.
- ♦ W komórkach D4:F4 wpisać kolejne tytuły kolumn U [V], Indeks, I [mA].
- ♦ W komórkach D5:D11 wpisać kolejne wartości napięcia, dla których chcemy interpolować wartość prądu. Pamiętajmy, że powinny one znajdować się pośrodku przedziału kolejnych znanych danych. W naszym przypadku będą to odpowiednio wartości: 1,61; 1,63; 1,65; 1,67; 1,69; 1,71 oraz 1,73.
- ♦ W komórce E5 wpisać formułę: =PODAJ.POZYCJĘ(D5;\$A\$4:\$A\$11). Pozwoli nam ona na odczytanie pierwszego numeru wiersza z zadanego zakresu, w którym wartość napięcia zmierzonego jest mniejsza lub równa szukanemu napięciu. Dzięki temu będziemy mogli łatwo określić pozycje wartości x_1 , x_2 , y_1 oraz y_2 , które używane są we wzorze interpolacyjnym. Następnie kopiujemy formułę do komórek.
- ♦ W komórce F5 wpisać formułę:

$$=((D5-INDEXS(\$A\$4:\$A\$11;E5+1))/(INDEXS(\$A\$4:\$A\$11;E5)-INDEXS(\$A\$4:\$A\$11;E5+1))*INDEXS(\$B\$4:\$B\$11;E5)+((D5-INDEXS(\$A\$4:\$A\$11;E5))/(INDEXS(\$A\$4:\$A\$11;E5+1)-INDEXS(\$A\$4:\$A\$11;E5))*INDEXS(\$B\$4:\$B\$11;E5+1)$$
 Jest to bezpośrednie przepisanie wzoru na interpolację podanego wcześniej. Użyta funkcja INDEXS pozwala na odczytanie wartości o wskazanym indeksie z podanego zakresu. W naszym przypadku jest to wartość zwrócona przez funkcję PODAJ.POZYCJĘ, która znajduje się w wierszach kolumny E.

Wynik działania tego arkusza widoczny jest na rysunku 4.14. Dzięki aproksymacji zagęściliśmy ilość punktów danych, co pozwoliło nam na dużo dokładniejsze wyrysowanie wykresu charakterystyki napięciowo-prądowej badanej diody.

Rysunek 4.14.

Arkusz Excela
do liczenia wartości
za pomocą
interpolacji liniowej

	A	B	C	D	E	F
	Interpolacja danych do charakterystyk prądowo-napięciowych diody					
1						
2						
3	U [V]	I [mA]		Interpolacja liniowa		
4	1,6	3,98		U [V]	Indeks	I [mA]
5	1,62	5,19		1,61	1	4,59
6	1,64	6,16		1,63	2	5,68
7	1,66	8,58		1,65	3	7,37
8	1,68	10,9		1,67	4	9,74
9						

Nic nie stoi na przeszkodzie, aby w kolumnie F, w miejscu, gdzie wykorzystywana jest wartość wyliczona przez funkcję PODAJ.POZYCJĘ, użyć zagnieżdżenia funkcji. Spowodowałoby to jednak znaczne jej rozbudowanie, a przez to jeszcze bardziej zmniejszyłoby czytelność całej formuły. To z kolei groziłoby większym prawdopodobieństwem popełnienia błędu podczas wpisywania formuły i większymi problemami z jego znalezieniem.

4.5. Podsumowanie

W rozdziale tym poznaliśmy różne metody dopasowywania krzywych do zebranych danych. Metody te są szczególnie użyteczne w przypadku, kiedy posiadamy jedynie ograniczony zbiór danych wejściowych zebranych metodami eksperymentu.

Jeżeli nie zależy nam na wyliczaniu wartości pośrednich, wówczas mechanizmy rysowania linii trendu są najbardziej optymalnym wyjściem. Pozwalają bowiem szybko, przy pomocy kilku kliknięć, wyrysować krzywą dopasowaną do naszych danych i określić jej równanie. Jeżeli zależy nam jedynie na wyznaczeniu współczynników równania opisującego krzywą dopasowaną do danych, wówczas z pomocą przychodzą wbudowane mechanizmy regresji liniowej oraz wielomianowej.

W przypadku gdy mamy do czynienia z bardziej złożonymi funkcjami lub mamy zbyt mało punktów danych, aby poprawnie wyrysować jakąkolwiek krzywą, możemy odwołać się do mechanizmów interpolacji i oszacować brakujące dane.

4.6. Zadania do samodzielnego wykonania

1. Utwórz arkusz Excela zawierający wyliczenia wielomianu $y = 2x^3 + 3x^2 - 5x + 100$ dla kilku losowych wartości x , a następnie wyrysuj wykres punktowy z odpowiednią linią trendu.
2. Utwórz arkusz Excela zawierający wyliczenia funkcji $y = e^{3x}$ dla kilku losowych wartości x , po czym, korzystając z mechanizmu regresji, określ współczynniki krzywej dopasowanej. Następnie wykreśl obie funkcje (analityczną oraz uzyskaną w wyniku regresji) i sprawdź, jakim błędem obarczone są dane regresji.
3. W literaturze dość łatwo można znaleźć wzór określający szybkość wzrostu populacji zwierzęcej, który opisany jest zależnością: $N = Ce^{Bt}$, gdzie:
 - ♦ N — populacja w chwili t ,
 - ♦ C — populacja początkowa,
 - ♦ B — wskaźnik urodzeń,
 - ♦ t — czas.

Wykonaj obliczenia dla kilku różnych wskaźników urodzeń oraz kilku równomiernie oddalonych od siebie momentów dla każdego wskaźnika. Spróbuj dopasować krzywą za pomocą różnych trendów (wielomianowego, logarytmicznego itp.). Sprawdź, czy jest to możliwe dla jakiegokolwiek innego trendu poza trendem wykładniczym.

4. Dla danych z zadania 3. dokonaj interpolacji danych dla momentów znajdujących się w połowie wybranego przedziału czasu. Wykreśl odpowiednie wykresy i porównaj je z tymi, które uzyskałeś w wyniku zastosowania linii trendu.