

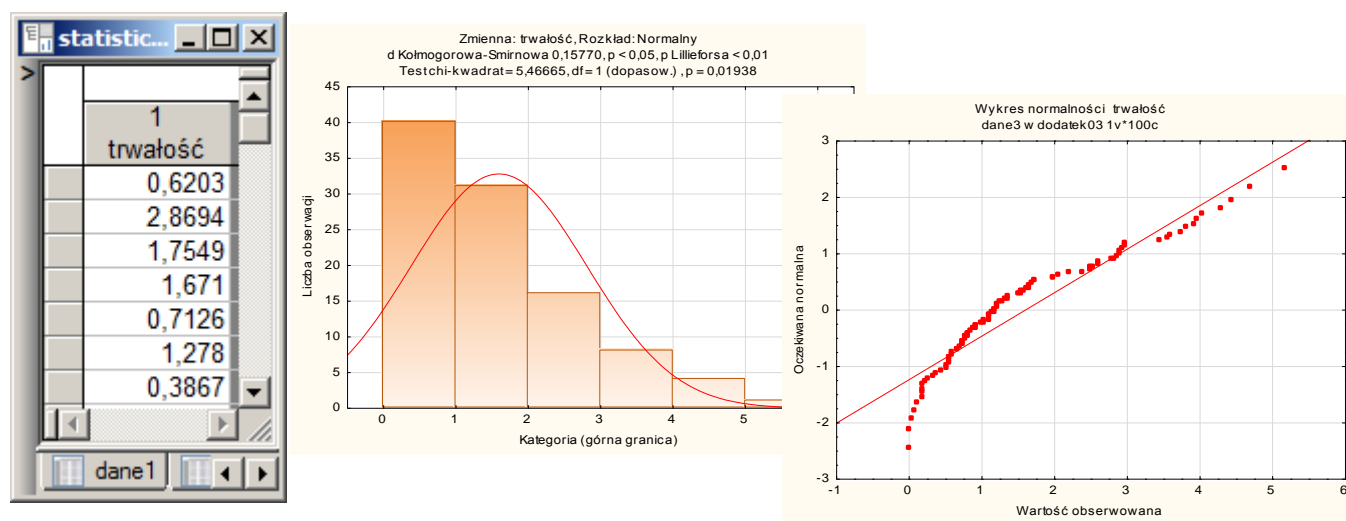
7. TRANSFORMACJA ROZKŁADU DO ROZKŁADU NORMALNEGO

Większość analiz statystycznych dotyczy zmiennych o rozkładzie normalnym. W praktyce rozkład zmiennych często odbiega od rozkładu normalnego – w takim przypadku przed wykonaniem odpowiedniej analizy konieczne jest takie przekształcenie zmiennych aby po wykonaniu przekształcenia ich rozkład był bliski normalnemu. O transformacji, którą można zastosować decyduje rozkład zmiennej, w wielu przypadkach do normalizacji zmiennej można zastosować przekształcenie Boxa-Coxa. Przekształcenie to to właściwie cała rodzina przekształceń obejmującą (w zależności od wartości parametru λ tego przekształcenia), przekształcenia potęgowe oraz logarytmiczne. W STATISTICE przekształcenie Boxa-Coxa można wykonać z poziomu okna **Przekształcenie Boxa-Coxa**, dostępnego z menu **Dane** (opcja **Dane** jest widoczna jeżeli bieżącym elementem skoroszytu jest arkusz).

Przykład

Wykorzystując zapisane w arkuszu *dane3* wyniki 100 pomiarów trwałości narzędzia stosowanego w pewnym procesie technologicznym w trakcie 6 umownych jednostek czasu.

Przeprowadzona analiza (histogram, wykres normalności i testy zgodności) wykazała, że rozkład trwałości nie jest rozkładem normalnym.



dodatek03.stw - Zmienna: trwałość, Rozkład: Normalny (dane3 w dodatek03.stw)

Zmienna: trwałość, Rozkład: Normalny (dane3 w dodatek03.stw)
d Kolmogorowa-Smirnowa 0,15770, p < 0,05, p Lillieforsa < 0,01
Chi-kwadrat = 5,46665, df = 1 (dopasow.), p = 0,01938

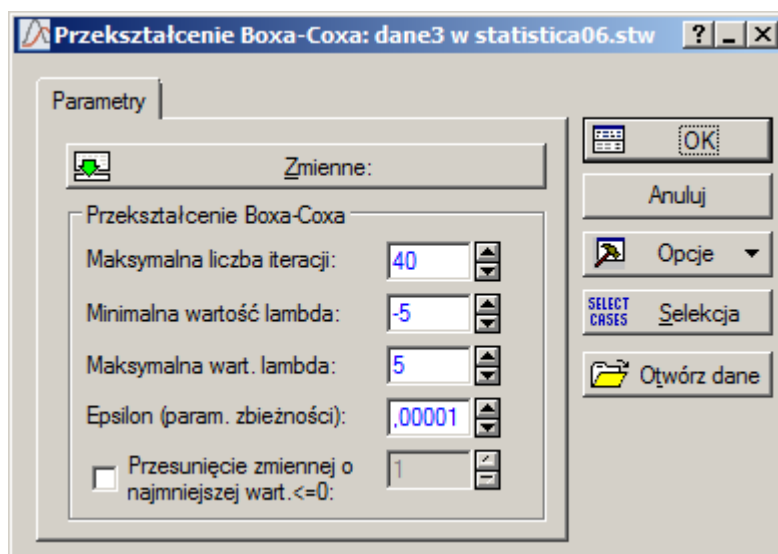
Górna Granica	Obserw. Liczność	Skumulow. Obserw.	Procent Obserw.	Skumul. % Obserw.	Oczekiwana Liczność
<= 1,00000	40	40	40,00000	40,0000	31,2645
2,00000	31	71	31,00000	71,0000	31,8174
3,00000	16	87	16,00000	87,0000	24,5425
4,00000	8	95	8,00000	95,0000	9,9836
5,00000	4	99	4,00000	99,0000	2,1373
<nieskończoność	1	100	1,00000	100,0000	0,2545

Zmienna: trwałość, Rozkład: Normalny (dane3 w dodatek03.stw)

Do normalizacji rozkładu zmiennej wykorzystane zostało przekształcenie Boxa – Coxa. Poszukiwanie optymalnej wartości parametru λ jest realizowane w programie przy pomocy *metody złotego podziału*.

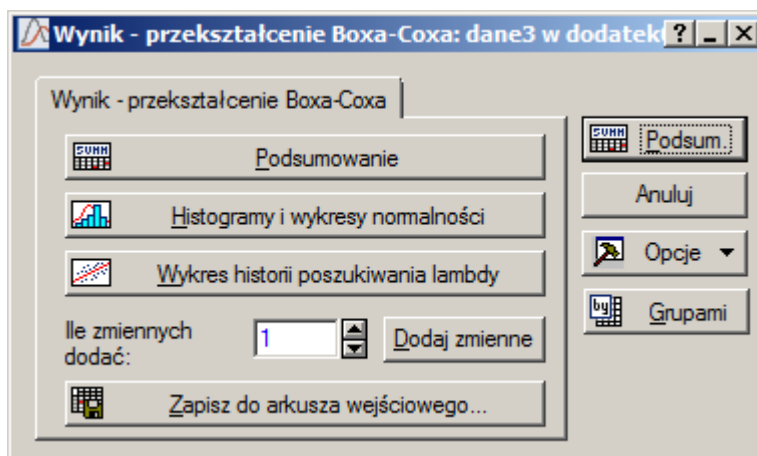
W oknie **Przekształcenie Boxa-Coxa** można zmienić domyślne parametry tej metody:

- maksymalną liczbę iteracji (domyślnie: 40),
- przedział w którym poszukiwany jest parametr λ (domyślnie: od -5 do 5),
- parametr zbieżności (domyślnie 0,00001).



Sposób działania metody polega na iteracyjnym zawężaniu pierwotnego przedziału poszukiwań. Metoda kończy swoje działanie po zawężeniu przedziału do długości równej parametrowi zbieżności lub po osiągnięciu maksymalnej liczby iteracji.

Po zaakceptowaniu okna parametrów przyciskiem **OK**, w kolejnym oknie



można:

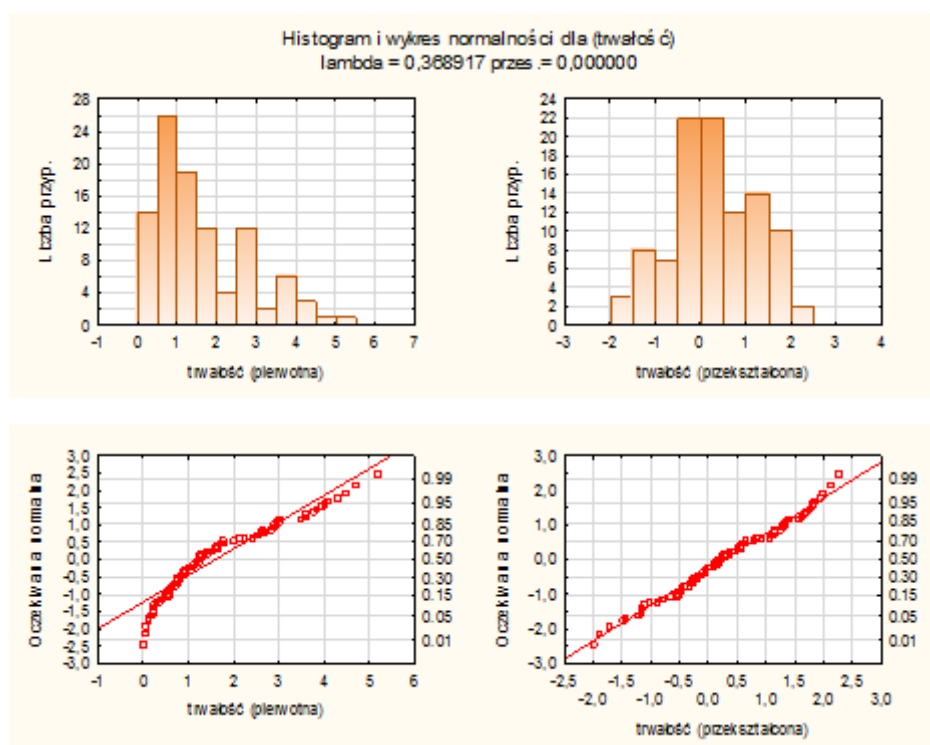
- wygenerować arkusz podsumowania zawierający między innymi znaną wartość parametru λ (przycisk **Podsumowanie**),
- oglądnąć przebieg poszukiwań (przycisk **Wykres historii poszukiwania lambda**),
- przeanalizować jakość znalezionej odpowiedzi porównując histogramy i wykresy normalności zmiennej pierwotnej i zmiennej przekształconej (przycisk **Histogramy i wykresy normalności**),
- zapisać przekształconą zmienną w arkuszu tak aby można ją było wykorzystać do dalszych analiz (przycisk **Zapisz do arkusza wejściowego**).

Z arkusza zawierającego podsumowanie wynika, że optymalna wartość parametru λ (dla której *funkcja największej wiarygodności* osiąga wartość maksymalną) wynosi $\lambda = 0,368919$. Dodatkowo, w arkuszu podsumowania umieszczone jest wyrażenie pozwalające na przekształcenie zmiennej (po skopiowaniu do arkusza zawierającego przekształcaną zmienną).



Statystyki (pomiar2 w lab04_1_rozw.stw)							
Przekształcona(e) zmienna(e)	Lambda	Przes.	Śred.	Odch. stand.	Dolna granica przedz. ufn.	Górna granica przedz. ufn.	Formuła przekształcenia Boxa-Coxa
pomiar	0,396475	0,00	,282148	1,181058	0,171002	0,634181	$((1+(0,396475))^{-1})/(0,396475)$

Analiza histogramu i wykresu normalności wskazuje, że po zastosowaniu znalezionej przekształcenia rozkład przekształconej zmiennej można uznać za normalny.



Zmodyfikowane wartości danych można zapisać w tym samym arkuszu, w którym zapisane są dane pierwotne – w ten sposób możliwe będzie wykonywanie dalszych analiz, w których konieczne jest spełnienie założenia o normalności rozkładu. Dane mogą zostać zapisane, o ile w arkuszu wejściowym znajduje się dodatkowa kolumna (tzn. dodatkowa zmienna).

Zmienną tą można w arkuszu utworzyć przed wykonaniem przekształcenia, możliwe jest także uzupełnienie arkusza z poziomu okna wyników przekształcenia. Nową zmienną do arkusza wejściowego można dodać po wybraniu przycisku **Dodaj zmienne** – zmienna otrzyma nazwę wygenerowaną automatycznie przez program. Po uzupełnieniu arkusza można zapisać przekształcone dane w arkuszu. Po naciśnięciu przycisku **Zapisz do arkusza wejściowego** należy wskazać w jakiej zmiennej arkusza zostaną zapisane zmodyfikowane dane. W tym celu, w części **Zmienne** należy wskazać zmienną, która przyjmie wartości wyznaczone na podstawie przekształcenia (tutaj jest to wygenerowana automatycznie zmienna **ZmPrz2**) i nacisnąć przycisk **Przypisz**. Przekształcone dane zostaną zapisane w arkuszu po zamknięciu okna przyciskiem **OK**.



The first screenshot shows the 'Przypisywanie przekształconych zmiennych do nowych zmiennych' dialog box. The 'Przekształcone' list contains 'trwałość'. The 'Zmienne' list contains '1 - trwałość' and '2 - ZmPrz2', both with 'Typ' set to 'Podw. precyzji'. Buttons for 'OK', 'Anuluj', and 'Przypisz' are visible.

The second screenshot shows the same dialog box after the transformation. The 'Przekształcone' list is empty, and the 'Przypisane przekształcone' list contains 'trwałość'. The 'Przypisane zmienne' list contains '2 - ZmPrz2'.

The third screenshot shows a data viewer window titled 'statistica06.stw - dan...'. It displays two columns: '1 trwałość' and '2 ZmPrz2'. The data values are as follows:

1 trwałość	2 ZmPrz2
0,6203	-0,4378503
2,8694	1,28842788
1,7549	0,62501138
1,671	0,5652673
0,7126	-0,3185129
1,278	0,25673786
0,3867	-0,801457

Analiza wykonana przy pomocy opcji **Statystyka/Rozkłady i symulacja/Dopasuj rozkład** wskazuje, że nie można odrzucić hipotezy o normalności rozkładu zmiennej **ZmPrz2**.

The screenshot shows the 'Wyniki dopasowania rozkładów: dane3 w statistica06.stw' dialog box. The 'Zmienne' dropdown is set to 'ZmPrz2'. The 'Dostosuj rozkłady' button is visible. The 'Użyj dostosowanych rozkładów' checkbox is unchecked. The results table is as follows:

Rozkład	d K-S	K-S p	Stat. AD	p AD	Chi-kwadrat	Chi kwadrat p
Mieszanka Gaus...	0,055932	0,895680	0,319106	0,923189	11,980000	0,017501
Normalny (locatio...	0,061166	0,826131	0,426092	0,822327	11,760000	0,162238
Johnson SB(typ ...	0,065748	0,755145	0,446254	0,801620	N/A	N/A
Uogólniony warto...	0,071046	0,667029	0,488044	0,758735	14,400000	0,044507
Trójkątny(min,ma...	0,093343	0,327729	1,979221	0,094402	16,160000	0,023694
Log-normalny (sk...	N/A	N/A	N/A	N/A	N/A	N/A
Nałożony normal...	N/A	N/A	N/A	N/A	N/A	N/A
...	N/A	N/A	N/A	N/A	N/A	N/A

Buttons at the bottom include 'Macierz korelacji', 'Podsumowanie rozkładu', and 'Zapisz dopasowanie'.

