

PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu nr 309
RESEARCH PAPERS
of Wrocław University of Economics No. 309

Spółeczno-gospodarcze aspekty statystyki

Redaktorzy naukowi

**Zofia Rusnak
Edyta Mazurek**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Joanna Szynal

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Kopiowanie i powielanie w jakiegokolwiek formie wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2013

ISSN 1899-3192

ISBN 978-83-7695-398-4

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Tadeusz Bednarski: Rola Jerzego Sławy-Neymana w kształtowaniu metod statystycznej analizy przyczynowości	11
Filip Borowicz: Ocena możliwości uzupełnienia danych BAEL informacjami ze źródeł administracyjnych w celu dokładniejszej analizy danych o bezrobociu	19
Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna: Przydatność testów nieparametrycznych Kruskala-Wallisa i mediany w długoterminowej ocenie parametrów kruszyw melafirowych	27
Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna: Karty kontrolne w ocenie jakości kruszyw dla budownictwa drogowego.....	42
Czesław Domański: Uwagi o procedurach weryfikacji hipotez z brakuącą informacją.....	54
Stanisław Heilpern: Zależne procesy ryzyka.....	62
Artur Lipieta, Barbara Pawelek, Jadwiga Kostrzewska: Badanie struktury wydatków w ramach wspólnej polityki UE z wykorzystaniem analizy korespondencji.....	78
Agnieszka Marciniuk: Dwa sposoby modelowania stopy procentowej w ubezpieczeniach życiowych	90
Beata Bieszk-Stolorz, Iwona Markowicz: Model nieproporcjonalnej intensywności Coxa w analizie bezrobocia	114
Edyta Mazurek: Statystyczna analiza podatku dochodowego od osób fizycznych.....	127
Katarzyna Ostasiewicz: Awersja do nierówności w modelowaniu użytkowania dóbr wspólnych.....	159
Piotr Peternek: Porównanie kart kontrolnych indywidualnych pomiarów uzyskanych z wykorzystaniem uogólnionego rozkładu lambda oraz krzywych Johnsona.....	179
Małgorzata Podogrodzka: Starzenie się ludności a płodność w Polsce w latach 1991-2010 – ujęcie regionalne	192
Renata Rasińska, Iwona Nowakowska: Jakość życia studentów w aspekcie znajomości wskaźników zrównoważonego rozwoju	203

Maria Rosienkiewicz, Jerzy Detyna: Analiza efektywności metod wyboru zmiennych objaśniających do budowy modelu regresyjnego	214
Jerzy Śleszyński: National Welfare Index – ocena nowego miernika rozwoju trwałego i zrównoważonego	236
Maria Szmuksta-Zawadzka, Jan Zawadzki: Wykorzystanie oszczędnych modeli harmonicznych w prognozowaniu na podstawie szeregów czasowych o wysokiej częstotliwości w warunkach braku pełnej informacji.....	261
Anna Zięba: O możliwościach wykorzystania metod statystycznych w badaniach nad stresem	278

Summaries

Tadeusz Bednarski: Role of Jerzy Sława-Neyman in statistical inference for causality	18
Filip Borowicz: Assessing the possibility of supplementing the Polish LFS data with register records for more detailed unemployment data analysis.	26
Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna: Usefulness of nonparametric Kruskal-Wallis and median tests in long-term parameters assessment of melaphyre crushed rocks	41
Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna: Control charts in the assessment of aggregates quality for road construction.....	53
Czesław Domański: Some remarks on the procedures of the verification of hypotheses under incomplete information.....	61
Stanisław Heilpern: Dependent risk processes	77
Artur Lipieta, Barbara Pawelek, Jadwiga Kostrzewska: Study of the structure of expenditure under the EU's common policy using correspondence analysis	89
Agnieszka Marciniuk: Two ways of stochastic modelling of interest rate in life insurances	113
Beata Bieszk-Stolorz, Iwona Markowicz: The Cox non-proportional hazards model in the analysis of unemployment.....	126
Edyta Mazurek: Statistical assessment of Personal Income Tax	158
Katarzyna Ostasiewicz: Inequality aversion in modeling the use of common pool resources	178
Piotr Peternek: Comparison of control charts of individual measurements based on general Lambda distribution and Johnson curves.....	191
Małgorzata Podogrodzka: The ageing of the population and fertility in Poland in the years 1991-2010 by voivodeships.....	202
Renata Rasińska, Iwona Nowakowska: Students' life quality in terms of knowledge of sustainable development indicators	213

Maria Rosienkiewicz, Jerzy Detyna: Efficiency analysis of chosen methods of explanatory variables selection within the scope of regression model construction.....	235
Jerzy Śleszyński: <i>National Welfare Index</i> – assessment of a new measure of sustainable development.....	260
Maria Szmuksta-Zawadzka, Jan Zawadzki: The application of harmonic models in forecasting based on high frequency time series in condition of lack of full information.....	277
Anna Zięba: About statistical methods in the study on stress	284

Piotr Peternek

Uniwersytet Ekonomiczny we Wrocławiu

PORÓWNANIE KART KONTROLNYCH INDYWIDUALNYCH POMIARÓW UZYSKANYCH Z WYKORZYSTANIEM UOGÓLNIONEGO ROZKŁADU LAMBDA ORAZ KRZYWYCH JOHNSONA

Streszczenie: W problematyce statystycznego sterowania procesem założenie o rozkładzie normalnym przy konstruowaniu kart kontrolnych jest powszechnie obowiązujące. Brak takiego założenia skutkuje potrzebą korzystania ze specyficznych rozwiązań. W pracy przedstawione zostanie porównanie kart kontrolnych wykorzystujących krzywe Johnsona z kartami kontrolnymi skonstruowanymi z wykorzystaniem uogólnionego rozkładu lambda.

Słowa kluczowe: karty kontrolne, krzywe Johnsona, uogólniony rozkład lambda

1. Wstęp

Zastosowanie kart kontrolnych w problematyce zarządzania jakością odbywa się przy założeniu normalności rozkładu badanych cech. Tymczasem w rzeczywistości założenie to jest bardzo często nie spełnione. Janacek i Meikle [1997] wskazali, że na 10 badanych przez nich cech, związanych z realizowanym projektem w przemyśle lotniczym, aż 7 było dalekich od rozkładu normalnego. Podobny wynik uzyskali Alloway i Raghavachari [1991]. Korzystanie z kart kontrolnych nie odbywa się jednak bez zwracania uwagi na założenia, stosuje się bowiem „zabiegi” mające na celu zapewnienie rozkładu normalnego badanych cech. Jednym ze sposobów na uzyskanie założonego rozkładu jest dokonywanie matematycznych przekształceń badanych cech, jeszcze innym – stosowanie kart kontrolnych specyfikowanych do danego rozkładu. Najczęściej (powszechnie) wykorzystywanym sposobem gwarantującym zbliżenie się do rozkładu normalnego jest pobieranie prób wieloelementowych. Schilling i Nelson [1976] pokazali, że wystarczy pobierać próby 4-elementowe, by zapewnić rozkład normalny badanych cech. Są jednak takie sytuacje, w których nie można pobierać próbek kilkuelementowych i pobierane próby są pojedyncze (indywidualne). Z taką sytuacją ma się do czynienia na przykład, gdy rozpatruje się produkcję indywidualną (na zamówienie), czy też w przypadku badań niszczących.

Celem tej pracy jest porównanie kart kontrolnych indywidualnych pomiarów uzyskanych z wykorzystaniem krzywych Johnsona z kartami skonstruowanymi z wykorzystaniem uogólnionego rozkładu lambda (*Generalized Lambda Distribution*, GLD). Obie karty wykorzystują podejście kwantylowe bazujące na statystykach pozycyjnych. Przy tym karta wykorzystująca GLD uzyskiwana jest w ten sposób, że aproksymuje się wartości obserwacji za pomocą odpowiedniej funkcji, a następnie na jej podstawie uzyskuje się granice kart kontrolnych. W przypadku korzystania z krzywych Johnsona poszukuje się takiej funkcji transformującej, w wyniku której dane ulegną przekształceniu na rozkład normalny w ten sposób, że zgadzać się będą skumulowane prawdopodobieństwa dla poszczególnych kwantyli. W pracy przedstawiono wyniki badań symulacyjnych przeprowadzone z wykorzystaniem dwóch rodzajów rozkładów: normalnego oraz chi-kwadrat.

2. Krzywe Johnsona – podstawy teoretyczne

Problem transformacji zmiennych, tak by miały one rozkład normalny, dyskutowany jest w literaturze od XIX wieku. Podejście Pearsona i koncepcja całego systemu krzywych Pearsona dała podstawy do poszukiwań takich przekształceń, które zapewniłyby pożądane właściwości transformowanych danych. W 1947 roku Johnson opublikował pracę, która stała się fundamentem dla koncepcji krzywych Johnsona. Jego podejście bazowało na transformacji postaci:

$$z = \gamma + \eta k_i(x; \lambda; \varepsilon), \quad (1)$$

gdzie: z jest zmienną o standaryzowanym rozkładzie normalnym, γ i η są parametrami, natomiast funkcja k_i występuje w trzech alternatywnych postaciach, zwanych:

– rozkładem S_U :

$$k_1(x; \lambda; \varepsilon) = \arcsin h \left(\frac{x - \varepsilon}{\lambda} \right), \quad (2)$$

– rozkładem S_B :

$$k_2(x; \lambda; \varepsilon) = \ln \left(\frac{x - \varepsilon}{\lambda + \varepsilon - x} \right), \quad (3)$$

– rozkładem S_L

$$k_3(x; \lambda; \varepsilon) = \ln \left(\frac{x - \varepsilon}{\lambda} \right). \quad (4)$$

Przekształcenie S_L nazywane jest przekształceniem lognormalnym, S_B jest rozkładem ograniczonym na $(\varepsilon; \lambda + \varepsilon)$, natomiast S_U jest rozkładem nieograniczonym. Procedurę estymacji parametrów transformacji należy poprzedzić wyborem odpowiedniego przekształcenia k_i . Johnson [1949] w swej pracy opisuje sposoby wyboru transformacji bazujące na trzecim i czwartym momencie oraz odpowiednie metody estymacji. Sugerowane w pracy Johnsona metody rozwijane są w późniejszych latach. Wybór rodzaju krzywej, metody estymacji czy też tabele parametrów krzy-

wych znaleźć można m.in. w następujących pracach: [Wheeler 1980]; Bukac 1972; Johnson, Kitchen 1971]. W 1980 roku Slifker oraz Shapiro zaproponowali prostą procedurę wyboru krzywej oraz przygotowali wzory umożliwiające estymację parametrów transformacji. Proponowana przez Slifkera i Shapiro procedura bazuje na porównaniu prawdopodobieństw skumulowanych dla ustalonych wartości pochodzących z rozkładu normalnego z prawdopodobieństwami empirycznymi uzyskanymi dla badanych danych. Polega ona na wyborze takiej dodatniej wartości z pochodzącej z rozkładu normalnego standaryzowanego, aby możliwe było ustalenie wartości $(-3z, -z, z, 3z)$ i skonstruowanie 3 równych przedziałów pokrywających znaczny „obszar” standaryzowanego rozkładu normalnego. Odpowiadające wartościom $(-3z, -z, z, 3z)$ wartości $x_{-3z}; x_{-z}; x_z; x_{3z}$ w rozkładzie sprzed transformacji są podstawą do wyboru odpowiedniej krzywej k_i . Autorzy pracy wykazują bowiem (por. [Slifker, Shapiro 1980]), że wyboru krzywej k_i można dokonać za pomocą wyrażenia:

$$\frac{mn}{p^2}, \quad (5)$$

gdzie: $m = x_{3z} - x_z; n = x_{-z} - x_{-3z}; p = x_z - x_{-z}$,

ponieważ: $\frac{mn}{p^2} < 1$ dla przekształcenia S_B ; $\frac{mn}{p^2} > 1$ dla przekształcenia S_U oraz $\frac{mn}{p^2} = 1$ dla transformacji S_L .

W przedstawionej wyżej procedurze postępowania pewną kontrowersję wzbudzić może arbitralny wybór wartości z . Autorzy pracy sugerują wybór z poniżej wartości 1 ze względu na trudności z odnalezieniem punktów (kwantyli) odpowiadających wartościom $-3z$ oraz $+3z$. Autorzy sugerują przyjmowanie z w okolicach liczby 0,5 (sami przyjmują $z = 0,524$) i wskazują, że im większa próbka tym z powinno być bliższe wartości 1. Wydaje się, że należałoby zbadać wpływ wyboru wartości z na uzyskiwane rezultaty.

Kontynuując omawianie procedury wyboru adekwatnej postaci funkcji transformacji, należy przedstawić sposób uzyskiwania wartości $x_{-3z}; x_{-z}; x_z; x_{3z}$ odpowiadających wartościom $(-3z, -z, z, 3z)$. Dla każdej wartości $Z = \{-3z; -z; z; 3z\}$ liczy się wartość dystrybuanty rozkładu normalnego standaryzowanego P_Z . Prawdopodobieństwa te stają się podstawą do obliczenia kwantyli dla danych empirycznych;

korzysta się przy tym ze znanego wzoru: $\frac{\left(i - \frac{1}{2}\right)}{n} = P_Z \Leftrightarrow i = n \cdot P_Z + \frac{1}{2}$. Wartość i zwykle okazuje się być wartością niecałkowitą, stąd zazwyczaj dokonuje się interpolacji odpowiedniego kwantyla. Tak wyliczone kwantyle stają się podstawą do wyznaczenia wartości m, n oraz p , a w konsekwencji – wyboru rodzaju transformacji.

Jak omówiono wyżej, w pracy Slifkera i Shapiro zaprezentowano estymatory parametrów dla poszczególnych transformacji. Estymatory te przedstawione są dla

wartości parametrów m, n i p . Należy zwrócić uwagę, że wartości te są w rzeczywistości funkcjami kwantyli $x_{-3z}, x_{-z}, x_z, x_{3z}$. Poniżej przedstawione zostały estymatory jedynie dla transformacji S_B (pozostałe szacunki znaleźć można w oryginalnej pracy Slifkera i Shapiro):

$$\gamma = \eta \arcsin \left[\frac{\left(\frac{p-p}{n-m} \right) \left[\left(1 + \frac{p}{m} \right) \left(1 + \frac{p}{n} \right) - 4 \right]^{0.5}}{2 \left(\frac{p-p}{m-n} - 1 \right)} \right], \quad (6)$$

$$\varepsilon = \frac{x_z + x_{-z}}{2} - \frac{\lambda}{2} + \frac{p \left(\frac{p-p}{n-m} \right)}{2 \left(\frac{p-p}{m-n} - 1 \right)}, \quad (7)$$

$$\lambda = \left[\frac{p \left\{ \left[\left(1 + \frac{p}{m} \right) \left(1 + \frac{p}{n} \right) - 2 \right]^2 - 4 \right\}^{0.5}}{\left(\frac{p-p}{m-n} - 1 \right)} \right] > 0, \quad (8)$$

$$\eta = \frac{z}{\arcsin h \left(\frac{1}{2} \left[\left(1 + \frac{p}{m} \right) \left(1 + \frac{p}{n} \right) \right]^{0.5} \right)} > 0. \quad (9)$$

3. Uogólniony rozkład lambda – podstawy teoretyczne

Przez uogólniony rozkład lambda rozumie się funkcję postaci:

$$Q(p) = x_p = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2}, \quad (10)$$

gdzie: x_p oznacza kwantyl rozkładu rzędu p , λ_1 to parametr położenia, λ_2 to parametr skali, natomiast λ_3 i λ_4 to parametry kształtu.

Rozkład ten został zaproponowany przez Ramberga i Schmeistera [1974], jako uogólnienie przekształcenia proponowanego przez Tukeya w 1960 roku. Wykorzystanie do aproksymacji danych uogólnionego rozkładu lambda w formie zaproponowanej przez Ramberga i Schmeistera wymaga wykorzystania odpowiednich metod estymacji parametrów. Literatura wskazuje tu na kilka zasadniczych metod. Pierwsza z nich to metoda momentów, przedstawiona przez Ramberga i Schmeistera [1974], druga to metoda najmniejszych kwadratów, zaproponowana i udowodniona

przez Ozturka i Dale'a w 1985 roku. Natomiast trzecia, najmłodsza z prezentowanych w literaturze metoda, to metoda gwiazdna (*Starship method*) przedstawiona przez Kinga i MacGillivraya w 1999 roku. Wady i zalety tych metod zostały omówione np. w pracy [Lakhany, Mausser 2000]. Także w 2000 roku Kanji i Artif przedstawili tzw. kwantylowy rozkład logistyczny (*Quantile Logistic Distribution*), który połączył założenie o rozkładzie logistycznym z przekształceniem Ramberga i Schmeistera. Umożliwiło to konstrukcję kart kontrolnych dla pojedynczych pomiarów. Zasadniczą wadą pracy Kanji i Artifa jest założenie o rozkładzie logistycznym badanej zmiennej, przyjęte ze względu na łatwość estymacji parametrów tak przekształconego rozkładu lambda (zob. [Kanji, Artif 2000]). W niniejszej pracy, podobnie jak zrobiono to w pracy Kanji i Artifa [2000], wykorzystano klasyczną metodę najmniejszych kwadratów. Takie postępowanie w przypadku korzystania z oryginalnego podejścia Ramberga i Schmeistera nie jest jednak pozbawione wad, ponieważ nie gwarantuje uzyskania rozwiązania optymalnego. Przyjęty został sposób optymalizacji nieliniowej, pozwalający znaleźć rozwiązanie w ekstremum lokalnym. Niezbędna zatem będzie analiza uzyskanych rozwiązań z uwzględnieniem dobroci dopasowania danych do wartości teoretycznych. Optymalizacji dokonywano numerycznie z wykorzystaniem modułu Solver zawartego w pakiecie Excel. Posłużono się klasycznym rozumieniem metody najmniejszych kwadratów, tzn. minimalizowano sumę kwadratów różnic pomiędzy wartościami empirycznymi a wartościami teoretycznymi:

$$FC = \sum_{i=1}^n e_i^2 = \sum \left(x_i - \hat{Q} \left(\frac{r}{n+1}, \lambda_1, \lambda_2, \lambda_3, \lambda_4 \right) \right)^2 \longrightarrow \min. \quad (11)$$

Należy zauważyć, że w miejscu prawdopodobieństw w rozkładzie lambda pojawiły się częstości, a także, co oczywiste i niezmiernie ważne, wartości empiryczne były uporządkowane od najmniejszych do największych. Umożliwiło to porównania dla poszczególnych kwantyli.

4. Wyniki badań symulacyjnych

W celu weryfikacji możliwości zastosowania omawianych w punkcie 1 i 2 metod postanowiono przeprowadzić badania symulacyjne. Początkowo planowano przeprowadzenie badania z wykorzystaniem rozkładu normalnego, Cauchy'ego oraz chi-kwadrat, jednakże ograniczono się wyłącznie do rozkładu normalnego i chi-kwadrat. To ograniczenie spowodowane było zbyt słabym dopasowaniem uogólnionego rozkładu lambda do danych pochodzących z rozkładu Cauchy'ego.

Symulacje rozpoczęto od wygenerowania danych z rozkładu normalnego o wartości oczekiwanej 10 oraz wariancji 4 oraz rozkładu chi-kwadrat o 10 stopniach swobody. Arbitralnie zdecydowano o generowaniu po 40 obserwacji w ramach tych rozkładów. Wybór tak niewielkiej próby podyktowany był aspektami praktycznymi; w praktycznych zastosowaniach kart kontrolnych nie należy spodziewać się prób

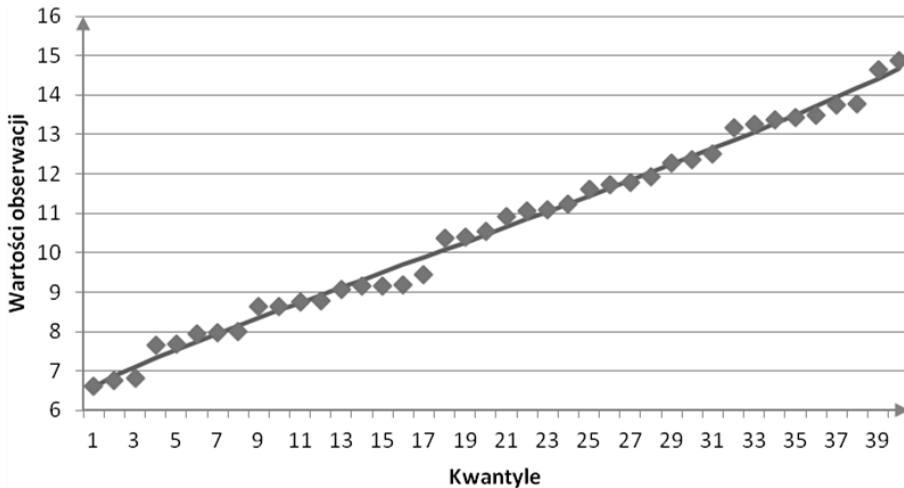
o znacznej liczności – wynika to z charakteru danych. Dla tak wygenerowanych danych, potraktowanych jako dane uregulowane, skonstruowano karty kontrolne. Wykorzystano najpierw uogólniony rozkład lambda, a następnie z krzyw Johnsona.

Optymalizacja funkcji celu (11) dla danych pochodzących z rozkładu normalnego pozwoliła oszacować parametry GLD wynoszące: $\lambda_1 = 7,075$; $\lambda_2 = 0,126$; $\lambda_3 = 0,024$; $\lambda_4 = 0,879$, dla których funkcja celu (11) przyjęła wartość $\sum_{i=1}^n e_i^2 = 1,79$.

Dopasowanie funkcji dla danych empirycznych przedstawiono na rys. 1, a także w tab. 3, gdzie zamieszczono wartości statystyki testu zgodności chi-kwadrat. Uznając takie dopasowanie za wystarczająco dobre do uzyskania granic karty kontrolnej, wykorzystano do ich obliczenia funkcję:

$$Q(p) = x_p = 7,075 + \frac{p^{0,024} - (1-p)^{0,879}}{0,126}. \quad (12)$$

Posłużono się tu klasycznym rozumieniem kart kontrolnych, a ich granice ustalono w zwyczajowo przyjętym „miejscu”, czyli ustalono je w ramach trzech odchyłeń standardowych. Oznacza to, że dolną granicę kontrolną (DLK) uzyskuje się, wprowadzając do funkcji (12) w miejsce prawdopodobieństwa wartość $p = 0,00135$, natomiast dla uzyskania górnej linii kontrolnej (GLK) $p = 0,99865$. Wynikiem takich działań jest uzyskanie karty kontrolnej o szerokości 8,41 (DLK = 6,4, natomiast GLK = 14,81). Zauważmy tu, że dla rozkładu $N(10; 2)$ teoretyczny przedział trzysigmowy ma szerokość 12 i jest przedziałem: (4; 16). Oznacza to, że tak uzyskana karta kontrolna, jako że jej szerokość jest węższa od przedziału teoretycznego, będzie dość często generowała sygnały o rozregulowaniu. Zauważmy bowiem, że



Rys. 1. Dopasowanie GLD do danych z rozkładu normalnego

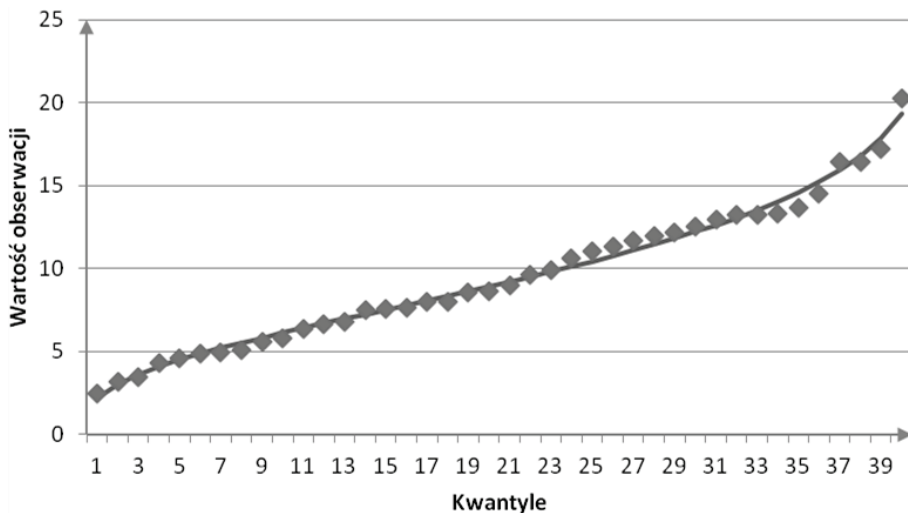
Źródło: opracowanie własne.

prawdopodobieństwo znalezienia się obserwacji pochodzących z rozkładu $N(10; 2)$ poza dolną lub górną linię kontrolną wynosi $p = 0,044$ przy zakładanym $0,0027$.

W sposób analogiczny postąpiono dla danych pochodzących z rozkładu chi-kwadrat. W wyniku przeprowadzonej procedury optymalizacyjnej uzyskano parametry uogólnionego rozkładu lambda: $\lambda_1 = 6,369$; $\lambda_2 = 0,05$; $\lambda_3 = 0,067$; $\lambda_4 = 0,287$, a następnie dolną i górną linię kontrolną, równą odpowiednio $1,47$ oraz $20,45$. Dopasowanie GLD do danych określone za pomocą funkcji celu (11) wynosiło teraz:

$$\sum_{i=1}^n e_i^2 = 6,45, \text{ a graficznie przedstawiono je na rys. 2. Podobnie jak to było dla roz-}$$

kładu normalnego, szerokość karty kontrolnej – wynosząca w tym przypadku $18,98$ – jest węższa niż teoretyczny trzysigmowy przedział zmienności ($1,58; 28,79$), którego szerokość wynosi $27,2$. Warte zauważenia jest przesunięcie w dół obydwu granic kart kontrolnych w stosunku do teoretycznego przedziału zmienności. Prawdopodobieństwa znalezienia się punktu pochodzącego z rozkładu chi-kwadrat z 10 stopniami swobody poza obliczonymi granicami kontrolnymi wynosi $0,026$, co oznacza, że jest około 10-krotnie większe niż założone.



Rys. 2. Dopasowanie GLD do danych z rozkładu chi-kwadrat

Źródło: opracowanie własne.

Po uzyskaniu granic kart kontrolnych z wykorzystaniem uogólnionego rozkładu lambda przystąpiono do procedury konstrukcji kart alternatywnych wykorzystujących krzywe Johnsona. Skorzystano tu z prezentowanej wcześniej procedury, proponowanej przez Slifkera i Shapiro. Zgodnie z tą procedurą i sugestiami autorów przyjęto wartość $z = 0,5438$. Odpowiadające wartościom $-3z$, $-z$, z oraz $3z$ wartości dystrybuanty standaryzowanego rozkładu normalnego wynosiły odpowiednio:

$P_{-3z} = 0,05; P_{-z} = 0,2918; P_z = 0,7082; P_{3z} = 0,95$. Dla tak obliczonych prawdopodobieństw, zgodnie z wzorem: $i = n \cdot P_z + \frac{1}{2}$, obliczono numery poszczególnych kwantyli dla danych empirycznych, a następnie same kwantyle. Numery kwantyli oraz kwantyle zaprezentowano w tab. 1. Na tej podstawie, zgodnie z procedurą, obliczono kryterium wyboru krzywej zgodnie z wzorem (5): $\frac{mn}{p^2} = \frac{2,04 \cdot 1,999}{3,39^2} = 0,35 < 1$.

Wartość mniejsza od 1 wskazuje na krzywą S_B . Zatem – zgodnie ze wzorami (6), (7), (8), (9) – obliczono parametry transformacji: $\eta = 0,684; \lambda = 8,909; \varepsilon = 6,027; \gamma = -0,013$. Na ich podstawie zapisano przekształcenie: $z = -0,013 + 0,684 \ln\left(\frac{x - 6,027}{8,909 + 6,027 - x}\right)$, które pozwala uzyskać wartości granic kontrolnych.

Poszukuje się zatem takich wartości x , które odpowiadają wartościom z równym -3 oraz 3 . Po dokonaniu obliczeń otrzymano DLK = 6,14 oraz GLK = 14,83. Po porównaniu tego wyniku z teoretycznym przedziałem zmienności (4; 16) okazuje się, że szerokość tej karty, podobnie jak było to w przypadku karty kontrolnej bazującej na GLD, jest mniejsza. Jednak prawdopodobieństwo znalezienia się wartości poza granicami kontrolnymi wynosi tu 0,035, jest więc niższe niż w przypadku uzyskanej wcześniej karty kontrolnej bazującej na GLD. Karta wykorzystująca krzywe Johnsona jest zatem kartą mniej czułą i generowałaby mniej fałszywych sygnałów o rozregulowaniu.

Tabela 1. Numery kwantyli i kwantyle z rozkładu empirycznego pochodzącego z rozkładu normalnego, odpowiadające prawdopodobieństwom z rozkładu normalnego standaryzowanego

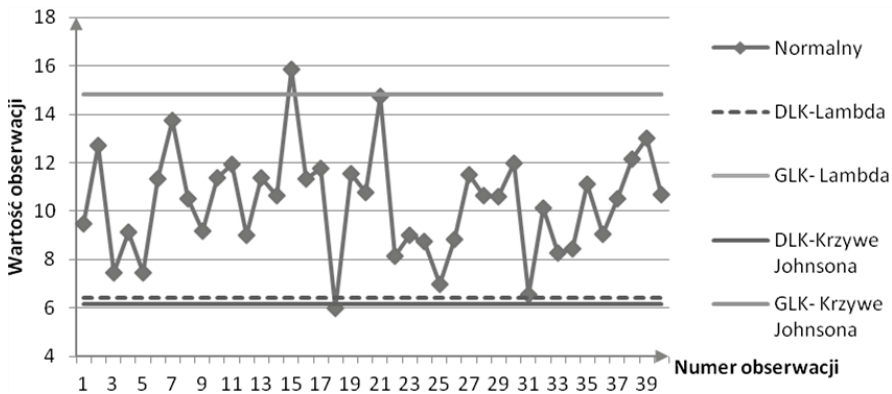
Numer kwantyla		Kwantyl
38,5	x_{3z}	14,21152
28,828	x_z	12,21275
12,172	x_{-z}	8,822107
2,5	x_{-3z}	6,778145

Źródło: obliczenia własne.

Analogicznie uzyskano kartę kontrolną dla danych generowanych z rozkładu chi-kwadrat. Powtórzono procedurę postępowania ustaloną przez Slifkera i Shapiro i uzyskano odpowiednie kwantyle (tab. 2), kryterium wyboru krzywej $\frac{mn}{p^2} = 0,534 < 1$, a następnie samo przekształcenie: $z = 0,2897 + 0,908 \ln\left(\frac{x - 1,2357}{19,1076 + 1,2357 - x}\right)$.

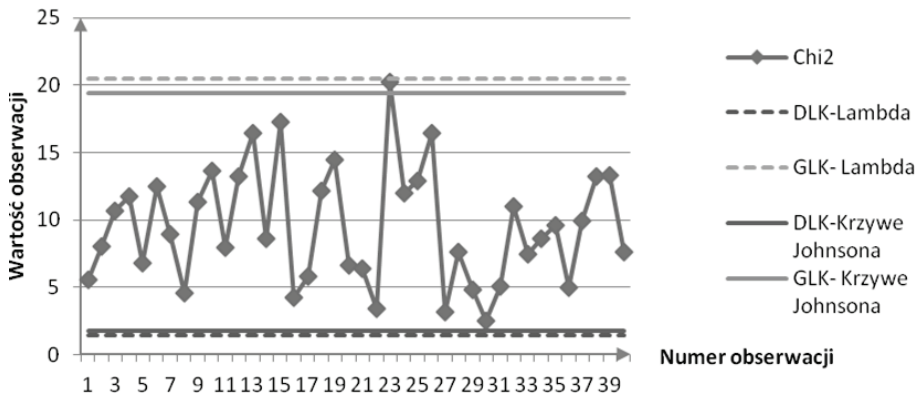
Granice karty kontrolnej uzyskane na podstawie prezentowanego przekształcenia wyliczono jako DLK = 1,733 oraz GLK = 19,423. Po raz kolejny szerokość karty

kontrolnej okazała się być mniejsza od teoretycznego przedziału (1,58; 28,79), a prawdopodobieństwo uzyskania fałszywego sygnału o rozregulowaniu wynosi w tym przypadku 0,037 i jest większe niż analogiczny wynik uzyskany dla karty wykorzystującej GLD. Potwierdzeniem tak uzyskanych prawdopodobieństw są wykresy kart kontrolnych przedstawione na rysunkach 3 i 4.



Rys. 3. Karty kontrolne dla danych pochodzących z rozkładu normalnego

Źródło: obliczenia własne.



Rys. 4. Karty kontrolne dla danych pochodzących z rozkładu chi-kwadrat

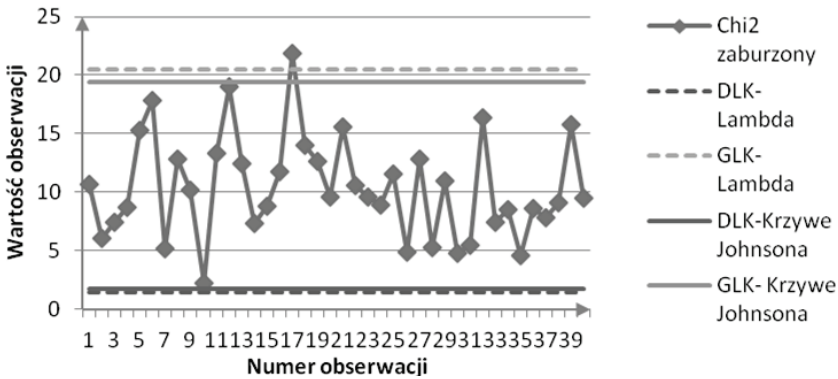
Źródło: obliczenia własne.

Tak uzyskane karty kontrolne poddano kolejnej weryfikacji. Tym razem wykorzystano dane zaburzone. Wygenerowano kolejne obserwacje z tych samych rozkładów, które zakłócono poprzez dodanie wartości wygenerowanych z rozkładu $N(0, 2)$. Wyniki tych symulacji przedstawiono na rysunkach 5 i 6. W przypadku rozkładu chi-kwadrat tylko 1 obserwacja znalazła się poza wyznaczonymi granicami kontrol-

Tabela 2. Numery kwantyli i kwantyle z rozkładu empirycznego pochodzącego z rozkładu chi-kwadrat, odpowiadające prawdopodobieństwom z rozkładu normalnego standaryzowanego

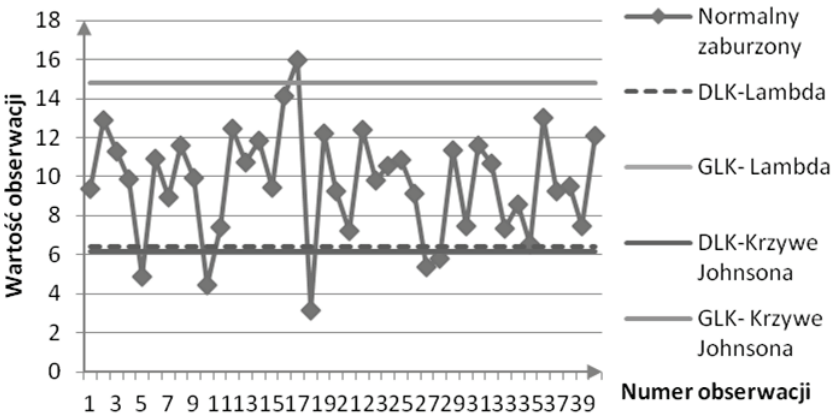
Numer kwantyla		Kwantyl
38,5	x_{3z}	16,83469
28,828	x_z	12,14031
12,172	x_{-z}	6,670417
2,5	x_{-3z}	3,265393

Źródło: obliczenia własne.



Rys. 5. Karty kontrolne dla zaburzonych danych pochodzących z rozkładu chi-kwadrat

Źródło: obliczenia własne.



Rys. 6. Karty kontrolne dla zaburzonych danych pochodzących z rozkładu normalnego

Źródło: obliczenia własne.

Tabela 3. Wartości statystyki chi-kwadrat dla danych pochodzących z rozkładu normalnego

Klasy	Obserwowane n_i	n_i krzywe Johnsona	n_i GLD
Do 8,5	8	10,09294	9,558771
8,5-10,5	11	9,792559	10,15013
10,5-12,5	12	9,872583	9,829311
Powyżej 12,5	9	10,24192	10,46179
Suma	40	40	40
	Wartość statystyki chi-kwadrat	1,1919	1,008975

Źródło: obliczenia własne.

Tabela 4. Wartości statystyki chi-kwadrat dla danych pochodzących z rozkładu chi-kwadrat

Klasy	Obserwowane n_i	n_i krzywe Johnsona	n_i GLD
Do 5	7	6,477074	6,267168
5-8	10	9,472782	10,09335
8-11	7	9,218327	9,714144
11-14	11	7,730981	7,043294
Powyżej 14	5	7,100837	6,882048
Suma	40	40	40
	Wartość statystyki chi-kwadrat	2,609229	3,582333

Źródło: obliczenia własne.

nymi. Dla zaburzonego rozkładu normalnego dla obu kart kontrolnych obserwacji poza granicami kontrolnymi było 6 na 40 wygenerowanych danych, co stanowiło 15%. Nietrudno przy tym policzyć, że w przypadku zmiennej losowej pochodzącej z rozkładu normalnego $N(10; 2)$, do którego dodano rozkład $N(0; 2)$, prawdopodobieństwo znalezienia się poza granicami kontrolnymi wynosi: dla karty kontrolnej generowanej przez GLD ok. 0,146, a dla karty uzyskanej z wykorzystaniem krzywej Johnsona ok. 0,13. Obydwie karty zachowały się zatem poprawnie, generując adekwatną i zgodną z teorią liczbę sygnałów o rozregulowaniu.

5. Zakończenie

Przeprowadzone analizy możliwości zastosowania kart kontrolnych konstruowanych z wykorzystaniem uogólnionego rozkładu lambda czy też krzywych Johnsona nie wskazują jednoznacznie na wyższość któregoś z podejść. W niektórych sytuacjach lepsze okazują się być karty korzystające z osiągnięć Johnsona, innym razem lepsze są karty wykorzystujące GLD. Warto zauważyć jest jednak to, że wykorzystanie tego rodzaju kart w przypadku procesu uregulowanego generowało wyższą

niż zakładana liczbę fałszywych sygnałów o rozregulowaniu. Dla procesu rozregulowanego liczba sygnałów o rozregulowaniu w przypadku danych generowanych z rozkładu normalnego zgadzała się z teoretyczną liczbą takich sygnałów. W pracy przedstawiono zbyt małą liczbę symulacji, by rekomendować korzystanie z konkretnej karty kontrolnej, wręcz odwrotnie, już dla dwóch rozkładów wyniki nie są jednoznaczne. Wydaje się zatem, że można stwierdzić, iż nie należy rezygnować z żadnego z podejść. Potencjalnie rekomendacją do stosowania wybranej karty kontrolnej mogłoby być dopasowanie wybranego rozkładu (uzyskiwanego z GLD czy krzywych Johnsona), nazwijmy go tu teoretycznym, do danych empirycznych na przykład za pomocą testu zgodności (tabela 3 i 4). Jednakże i tu odpowiedź nie jest oczywista – na przykład dla danych generowanych z rozkładu normalnego lepsze dopasowanie, mierzone wartością statystyki chi-kwadrat, miał rozkład wykorzystujący GLD, a tymczasem lepsza w sensie prawdopodobieństwa uzyskania fałszywego sygnału o rozregulowaniu była karta wykorzystująca krzywe Johnsona. Podsumowując uzyskane wyniki, należy stwierdzić, że należy przeprowadzić dodatkowe badania, aby dać jednoznaczną rekomendację dla którejś z kart. Bez wątplenia w przypadku procedury Slifkera i Shapiro należy zbadać wpływ wartości z na wyniki dopasowania. Natomiast w przypadku estymacji parametrów GLD można poszukiwać innych metod optymalizacji funkcji celu.

Literatura

- Alloway J.A., Raghavachari M., *Control chart based on the Hodges-Lehmann estimator*, Journal of Quality Technology 1991, ss. 336-347.
- Bukac J., *Fitting SB curves using symmetrical percentile points*, Biometrika 1972, 59, ss. 688-690.
- Janacek G.J., Meikle S.E., *Control Charts Based on Medians*, Journal of Royal Statistical Society 1997, series D, ss. 19-31.
- Johnson N.L., *System of frequency curves generated by methods of translation*, Biometrika 1949, 36, ss. 149-176.
- Johnson N.L., Kitchen J.O., *Tables to facilitate fitting SB curves II: Both terminals known*, Biometrika 1971, 58, ss. 657-663.
- Kanji G.K., Osama Kasan Artif, *Median rankit control chart by the quantile approach*, Journal of Applied Statistics 2000, ss. 757-770.
- King R.A., MacGillivray H.L., *A starship estimation method for the generalized Lambda distribution*, Australian and New Zealand Journal of Statistic 1999, 41(3), ss. 353-374.
- Lakhany A., Mausser H., *Estimating the Parameters of the Generalized Lambda Distribution*, Algo Research Quarterly 2000, vol. 3 no. 3, ss. 47-58.
- Ozturk A., Dale R., *Least squares estimation of the parameters of the Generalized Lambda Distribution*, Technometrics 1985, 27 (1), ss. 81-84.
- Peternek P., *Wybrane karty kontrolne indywidualnych pomiarów*, [w:] S. Forlicz (red.) *Zastosowanie metod ilościowych w ekonomii i zarządzaniu*, CeDeWu 2012, ss. 249-260.
- Ramberg J.S., Tadikmalla P.R., Dudewicz E.J., Mykytka E.F., *A Probability Distribution and its Uses in Fitting Data*, Technometrics 1979, v. 21, ss. 201-214.

- Ramberg J.S., Schmeiser B.W., *An approximate method for generating asymmetric random variable*, Communications of the ACM 1974, ss. 78-82.
- Schilling E.G., Nelson P.R., *The effect of non-normality on the control limits of \bar{x} charts*, Journal of Quality Technology 1976, ss. 183-188.
- Slifker J.F., Shapiro S.S., *The Johnson System: Selection and Parameter Estimation*, Technometrics 1980, v. 22, ss. 239-246.
- Wheeler R.E., *Quantile estimators of Johnson Curve parameters*, Biometrika 1980, 67, ss. 725-728.

COMPARISON OF CONTROL CHARTS OF INDIVIDUAL MEASUREMENTS BASED ON GENERAL LAMBDA DISTRIBUTION AND JOHNSON CURVES

Summary: In the statistical process control the assumption of normal distribution is generally applicable while constructing control charts. If the assumption cannot be used, specific methods must be applied. This paper presents the comparison of control charts using Johnson curves with control charts made using general Lambda distribution.

Keywords: control charts, Johnson curves, general Lambda distribution.