

Jacek M. CZAPLICKI  
Politechnika Śląska, Gliwice

## TESTOWANIE HIPOTEZY O WYSTĘPOWANIU ELEMENTÓW NIETYPOWYCH W CIĄGU OBSERWACJI OPISUJĄCYM NIEZAWODNOŚĆ URZĄDZEŃ GÓRNICZYCH

**Streszczenie.** W badaniach niezawodności urządzeń górniczych, prowadzonych w rzeczywistych warunkach eksploatacji, pojawiają się niekiedy takie wartości obserwowanej cechy, które w sposób ewidentny odbiegają od pozostałych. Powstaje problem, czy te, nietypowe wartości, są właściwością tej cechy czy nie? Zagadnienie występowania elementów nietypowych (odstających, oddalonych) może dotyczyć pojawienia się realizacji wyraźnie *za dużych* jak i *za małych* i mieści się w szeroko pojętym problemie jednorodności danych. Autor rozważa to zagadnienie wskazując na metody weryfikacji przynależności tych realizacji do badanego zbioru, prezentując na przykładach odpowiednie procedury postępowania sprawdzającego.

## ON OUTLIERS IN RELIABILITY INVESTIGATION OF MINE EQUIPMENT

**Summary.** In reliability investigations of mine equipment, especially those carried out in real exploitation conditions, the problem of outlying observations can occur. Immediately in such a case, a question comes into being whether this strange outcome is the property of observed feature or not. The problem of untypical data concerns values significantly greater as well as lower than usually and it can be reckoned to the problem of data homogeneity in a wide sense. The author considers this problem indicating methods lying in statistical field and outside of it. Particular procedures are presented and practical examples are shown in the paper.

### 1. Wprowadzenie

Jednorodność danych w statystyce matematycznej bywa rozumiana różnie w zależności od przyjętego kryterium tej jednorodności. Problemem wchodzącym w ten zakres, jest

występowanie bądź nie występowanie elementów ciągu obserwacji, które – pod względem wartości – odbiegają od innych elementów. Są to tzw. elementy odstające, zwane także *nietypowymi* bądź *oddalonymi*<sup>1</sup>. Ich wyłowienie czasami może być dość łatwe, czasami może przysporzyć sporo kłopotów.

Częstym (w badaniach niezawodnościowych) przypadkiem – jeżeli chodzi o zagadnienie jednorodności – jest zaobserwowanie bardzo długo trwającej naprawy, której czas znacznie odbiega od innych napraw.

Pierwszym, najważniejszym etapem w takim przypadku powinna być wnikliwa analiza przyczyn i skutków zaszłego zdarzenia – a zatem analiza pozastatystyczna.

Istnieje obszerna klasa obiektów technicznych, dla których istnieje możliwość wystąpienia takiego uszkodzenia, które zniszczy ten obiekt bądź go poważnie zdegraduje; zniszczeniu może także ulec część otoczenia tego obiektu (np. przejechanie krańcowych poziomów przez naczynie wyciągowe i uderzenie w belki odbojowe czy zerwanie się liny nośnej wyciągu). Mówimy wówczas, że zaszło zdarzenie katastroficzne, a więc zdarzenie wyjątkowe, którego częstość pojawienia się i czas trwania są odbiegające od analogicznych wartości dla innych uszkodzeń. Wynika stąd, że nie można łączyć zwykłych uszkodzeń z uszkodzeniami typu katastroficznego – są to dwa różne typy uszkodzeń. A zatem, analizując zapisy jesteśmy na ogół w stanie wyśledzić te stany, które *nie pasują* do danej rodziny stanów. Na szczęście zdarzenia takie pojawiają się rzadko, a projektanci, konstruktorzy i eksploataotorzy starają się w określony sposób zminimalizować prawdopodobieństwo pojawienia się takiego zdarzenia.

Nieco bardziej skomplikowana sytuacja jest wtedy, gdy czas danego stanu odbiega ewidentnie od innych, a brak jest informacji w zapisach, dlaczego tak jest. Wówczas pomocne będzie sięgnięcie do metod statystycznych, które pozwolą ocenić, z jak rzadkim przypadkiem mamy do czynienia.

Celem niniejszej pracy jest rozważenie problemu weryfikacji elementów odstających zarówno od strony teoretycznej, jak i praktycznej. Podane zostaną przykłady z jakimi autor zetknął się w swojej praktyce badań niezawodności urządzeń górniczych.

---

<sup>1</sup> Mamy różne definicje elementu odstającego, patrz np. A. Zeliaś [5], M. Czekala [1]. Według Czekala „obserwację  $y$  z ciągu zmiennych losowych  $(X_1, X_2, \dots, X_n)$  nazywamy nietypową (odbiegającą w górę), jeżeli spełnione są dwa warunki: a)  $y = \max(X_1, X_2, \dots, X_n)$ , b)  $P\{\max(X_1, X_2, \dots, X_n) \geq y\} \leq \alpha$ , gdzie  $\alpha$  jest ustalonym z góry (zwykle małym) prawdopodobieństwem”.

## 2. Teoria i praktyka

Przyjmijmy, że próba została pobrana spośród elementów będących mieszaniną dwóch różnych populacji. Niech zmienne w tych populacjach różnią się silnie pomiędzy sobą pod względem wartości przeciętnej oraz jedna z tych populacji jest znacznie mniej liczna aniżeli druga. W tej sytuacji mamy prawo oczekiwać, że w naszej próbie może pojawić się wartość znacznie odbiegająca od innych wartości. Pojawienie się jej zmusza nas do postawienia hipotezy o niejednorodności ze względu na interesującą nas cechę (por. [1]).

Mamy zatem pobraną próbę:

$$\{X_1, X_2, \dots, X_k, \dots, X_n\}.$$

Podejrzewamy, że element  $k$ -ty jest nietypowy<sup>2</sup>. Zakładamy, że przeciętna wartość pozostałych elementów:

$$E(X_{i \neq k}) = \mu_i = \mu.$$

- ▲ Formułujemy hipotezę:  $H_0: E(X_i) = \mu$  mówiącą, że wszystkie elementy pochodzą z jednej populacji.
- ▲ Wobec hipotezy:  $H_1: \mu_k \neq \mu \wedge \mu_{i \neq k} = \mu$  głoszącej, że element  $k$ -ty ma istotnie różną wartość przeciętną.

Dalszy etap postępowania zależy od informacji, jaką posiadamy na temat rozkładu prawdopodobieństwa badanej zmiennej. Rozpatrzmy najpierw przypadek, gdy rozkład ten jest rozkładem wykładniczym (a zatem tym najbardziej rozpowszechnionym w zastosowaniach inżynierii niezawodności) i rozkładem Erlanga, a następnie rozkładem normalnym. Zastanowimy się potem, co zrobić, gdy brak jest jakiegokolwiek informacji w tym względzie.

---

<sup>2</sup> W większości przypadków ten element odstający to wartość maksymalna zaobserwowana w próbie.

## 2.1. Rozkład wykładniczy badanej zmiennej

Rozpatrzmy zatem, jako pierwszy, przypadek, gdy wiemy – czy to z poprzednich badań, czy z rozważań teoretycznych, że rozkład opisujący zmienną jest rozkładem wykładniczym z parametrem  $\lambda$ .

Zauważmy, że – jeżeli chodzi o ten typ rozkładu – jesteśmy w szczególnej sytuacji. Intensywny rozwój teorii niezawodności w drugiej połowie lat sześćdziesiątych i siedemdziesiątych ubiegłego wieku doprowadził do znaczącego wzrostu liczby probabilistycznych narzędzi badawczych, do badania statystycznych właściwości procesu eksploatacji obiektów technicznych. Stąd, do weryfikacji hipotezy o jednorodności próby, w sensie przez nas rozpatrywanym, możemy zastosować kilka różnych testów; różny będzie sposób podejścia do tego samego zagadnienia. Rozpatrzmy najpierw sprawdzenie naszego przypuszczenia obliczając wartości średnie z próby.

Wiadomo, że iloczyn  $2\lambda \sum_{i=1, i \neq k}^n X_i$  ma rozkład  $\chi^2$  o  $2(n-1)$  stopniach swobody, co umownie zapiszemy jako:

$$2\lambda \sum_{i=1}^n X_i : \chi^2(2(n-1)).$$

Wprowadzając estymator średniej na podstawie  $(n-1)$ -elementowej próby w postaci:

$$\bar{X}_{n-1} = \frac{1}{n-1} \sum_{i=1, i \neq k}^n X_i,$$

możemy zapisać, że:

$$2\lambda(n-1)\bar{X}_{n-1} : \chi^2(2(n-1)).$$

Zakładając prawdziwość hipotezy  $H_0$  i rozpatrując stosunek ilorazów łatwo stwierdzić, że:

$$\frac{2\lambda n \bar{X}_n}{2n} : \frac{2\lambda(n-1)\bar{X}_{n-1}}{2(n-1)} : F(2n, 2(n-1)),$$

gdzie  $\bar{X}_n$  jest estymatorem średniej z uwzględnieniem elementu odstającego, a  $F(2n, 2(n-1))$  oznacza rozkład *F Snedecora* o  $2n, 2(n-1)$  stopniach swobody. Upraszczając otrzymujemy:

$$\frac{\bar{X}_n}{X_{n-1}} : F(2n, 2(n-1)).$$

Hipotezę  $H_0$  należy odrzucić, gdy dla założonego poziomu istotności iloraz lewej strony powyższego równania jest większy aniżeli kwanty rzędu  $\alpha$  rozkładu  $F$ , tzn.:

$$\frac{\bar{X}_n}{X_{n-1}} > F_\alpha(2n, 2(n-1)).$$

Inny sposób podejścia do weryfikacji hipotezy dotyczącej istnienia elementu odstającego prezentował przeszło siedemdziesiąt lat temu *R. A. Fisher*, choć wtedy zagadnienie elementów odstających w samodzielnej formie jeszcze nie istniało. W swojej pracy [3] zaprezentował test dla jednoczesnego sprawdzenia hipotezy o wykładniczości z istnieniem elementu próby o wyraźnie większej wartości od pozostałych elementów. Zaproponował rozpatrzenie statystyki:

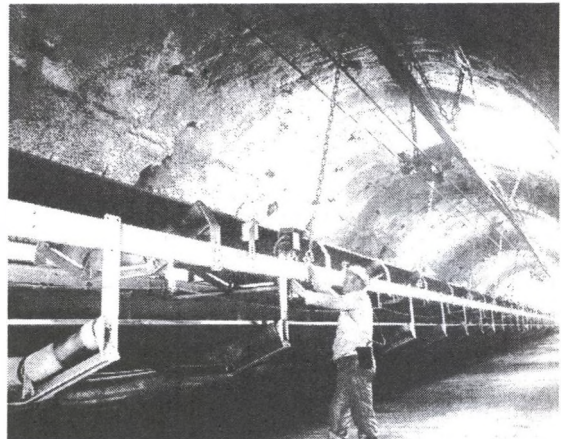
$$\eta = \frac{X_k}{\sum_{i=1}^n X_i}.$$

W cytowanej pracy przedstawione zostały wartości krytyczne dla tej statystyki, pozwalające na weryfikację sformułowanego przypuszczenia.

Rozważmy przykład.

■ W pewnej głębinowej kopalni notowano czasy pracy pomiędzy sąsiednimi naprawami przenośnika taśmowego, transportującego węgiel na głównej drodze odstawczej danego poziomu wydobywczego. Otrzymano następujący ciąg  $n = 16$  czasów:

910, 170, 280, 790, 660, 1210, 510,  
180, 6310, 140, 220, 440, 870, 380,  
60, 1340 h.



Podejrzenie wzbudził 9 element ciągu, odstający istotnie od innych czasów i wynoszący 6310 h. Istniało podejrzenie, że wystąpiła luka w zapisach – *może jednak coś w międzyczasie się uszkodziło i nie zostało to odnotowane?*<sup>3</sup>.

Jak wykazały liczne, wieloletnie badania niezawodności górniczych przenośników taśmowych, czasy pracy tych urządzeń prawie zawsze można opisać rozkładem wykładniczym.

Średnie, obliczone dla wszystkich elementów ciągu oraz dla elementów ciągu z wyłączeniem elementu odstającego, były równe:

$$\bar{X}_{n-1} = 544 \text{ h}, \quad \bar{X}_n = 904 \text{ h}.$$

Jak widać, różnica jest ogromna.

Stosunek tych średnich, czyli sprawdzian hipotezy  $H_0$  wynosi:

$$\frac{\bar{X}_n}{\bar{X}_{n-1}} = 1,66.$$

Wartość krytyczna, odczytana z tablic rozkładu  $F$  *Snedecora* dla stopni swobody właściwych naszemu przykładowi i poziomu istotności  $\alpha = 0,05$ , wynosi [6]:

$$F_{\alpha=0,05}(32,30) = 1,82.$$

Ponieważ wartość empiryczna jest mniejsza od krytycznej, brak jest podstaw do odrzucenia weryfikowanej hipotezy – nie mamy podstaw do wyeliminowania tego elementu z dalszej analizy.

Rozważmy obecnie jak przedstawia się weryfikacja naszego przypuszczenia w świetle procedury zaproponowanej przez *Fishera*. Wartość statystyki:

$$\eta = \frac{X_k}{\sum_{i=1}^n X_i} = \frac{6310}{14470} = 0,44.$$

Natomiast wartość krytyczna odczytana z tablic [2 lub 3] wynosi:

$$\eta(\alpha=0,05; n=16)=0,33.$$

Są zatem podstawy do odrzucenia sprawdzanego przypuszczenia, głoszącego, że dane pochodzą z populacji scharakteryzowanej rozkładem wykładniczym i że istnieje element wyraźnie większy od pozostałych. A zatem, i tym razem otrzymaliśmy ten sam wynik wnioskowania statystycznego.

<sup>3</sup> Przypadek, że coś nie zostało odnotowane nie jest całkiem bezpodstawny. Autorzy monografii [2 s.230] opisują przypadek, gdy rejestrujący uszkodzenia maszyny cyfrowej poszedł na urlop i powstała przerwa w rejestracji, skutkująca zanotowaniem *elementu odstającego*.



## 2.2. Rozkład Erlanga badanej zmiennej

Nasze rozważania możemy rozszerzyć na przypadek, gdy zaobserwowane wartości można opisać rozkładem Erlanga rzędu  $k$ . Sprawa jest elementarnie prosta. Wystarczy zauważyć, że tym razem iloczyn  $2\lambda \sum_{i=1, i \neq k}^n X_i$  ma rozkład  $\chi^2$  o  $2k(n-1)$  stopniach swobody.

Prowadząc analogiczne rozumowanie jak poprzednio dochodzimy do stwierdzenia, że :

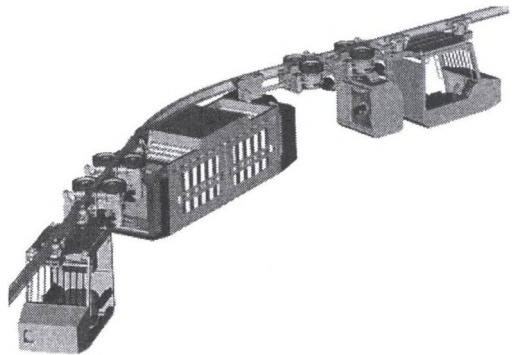
$$\frac{\bar{X}_n}{\bar{X}_{n-1}} > F_{\alpha}(2kn, 2k(n-1)).$$

Rozważmy przykład.

W pewnej głębokiej kopalni węgla ROW notowano w minutach czasu napraw spalinowej lokomotywy podwieszanej, służącej do przezbierania ścian wydobywczych. Otrzymano następujący ciąg  $n=24$  czasów:

90, 70, 180, 90, 60, 60, 280, 240, 190, 140,  
90, 270 60, 100, 180, 310, 60, 1850, 110,  
60, 240, 70, 120, 80.

Uwagę zwracał 18. czas wyraźnie odstający od innych czasów i wynoszący aż 1850 min. W pierwszym etapie analizy brak było opisu, co uległo uszkodzeniu i dlatego ta naprawa trwała aż tak długo.



Udało się jedynie wyeliminować przypadek, że lokomotywa czekała na jakieś części zamienne.

Wcześniejsze badania niezawodności tej lokomotywy pozwoliły zidentyfikować rozkład, który dobrze opisywał czasy napraw. Okazał się, że jest to rozkład gamma z parametrem kształtu bliskim 2. Przyjęto więc, że rozkład Erlanga rzędu drugiego ( $k=2$ ) nadal będzie dobrze opisywał nowe dane. Średnie czasy naprawy lokomotywy były równe:

$$\bar{X}_n = 208,3 \text{ min} \quad \text{i} \quad \bar{X}_{n-1} = 137 \text{ min.}$$

Ich iloraz wynosił:  $\frac{\bar{X}_n}{\bar{X}_{n-1}} = 1,52.$

Wartość krytyczna równa się [6]:  $F_{0,05}(96,92) \approx 1,4.$

Obliczona przez nas wartość empiryczna wyraźnie przewyższa wartość krytyczną. Są zatem podstawy do odrzucenia hipotezy, głoszącej jednorodność całej próbki. Element odstający powinien być wyłączony z dalszej analizy.

Późniejsze dociekania, dotyczące przyczyn tak długiej naprawy, pozwoliły ustalić, że zdarzył się wypadek (deformacja górotworu i przemieszczenie toru jezdnego) i lokomotywa wykoleiła się. Uszkodzeniu uległo szereg elementów. Zaszło zatem zdarzenie rzadkie z gatunku zdarzeń katastroficznych.

### 2.3. Rozkład normalny badanej zmiennej

Rozpatrzmy teraz przypadek, gdy rozrzut zaobserwowanych wartości można opisać rozkładem normalnym. Zanim przejdziemy do samej analizy, należy powiedzieć kilka zdań na temat stosowalności tego rozkładu w zagadnieniach górniczych inżynierskich.

Prawdą jest, że wiele rozkładów empirycznych czasów wykonywanych faz cykli użytkowania maszyn, jak również wielkości charakterystycznych realizacji wydobywania kopaliny ma kształt symetrycznej krzywej dzwonowej. Mowa tu o rozkładach czasów faz cyklu pracy ładowarek, wozów odstawczych, wozów wiertniczych itd. Strumień urobku, przepływający przez systemy przenośnikowe, to strumień w przybliżeniu gaussowski, a rozkład sumarycznego strumienia masy, dostarczonego pod szyb systemami odstawy, jest w przybliżeniu normalny. Rozrzut przypadkowych błędów pomiarowych różnych wielkości jest oczywiście też normalny.

Obszar określoności zmiennej w rozkładzie normalnym to cała oś liczb rzeczywistych od  $-\infty$  do  $+\infty$ . W zastosowaniach inżynierskich jest to oczywiście nie do przyjęcia; wielkości fizyczne, którymi się posługujemy mają swoje naturalne granice. Takie podejście, jakie jest prezentowane w teorii, może być wzięte pod uwagę jedynie jako model graniczny. Bardzo często lewostronną, naturalną granicą, bywa zero. Zastosowań z uwzględnieniem wartości ujemnych dla opisu zmienności analizowanych wielkości nie jest relatywnie tak dużo, dlatego wymienione zostaną takie parametry jak obciążenie, naprężenie, siła hamowania, przyspieszenie czy moment dynamiczny. Wielkości rozpatrywane w zagadnieniach niezawodnościowych, a także opisujące urobioną, przetransportowaną, wydobytą, przerobioną czy zwałowaną masę są z reguły opisane na nieujemnym zbiorze liczb rzeczywistych (np. czasy stanów). Jeżeli chodzi o środki transportu, to uwzględnienie także prawostronnej granicy jest niezbędne dla uzyskania wiarygodnych wyników obliczeń, związanych z tymi urządzeniami przemieszczania mas. Przykłady te świadczą jednoznacznie,



że w praktyce inżynierskiej, a w górnictwie w szczególności, mają zastosowanie rozkłady normalne ucięte i to obustronnie. W przypadku natomiast zastosowania, czy to rozkładu jednostronnie uciętego czy nieuciętego, niezbędna jest ocena wielkości popełnianego błędu, z racji aplikacji takiego modelu. W przypadku gdy błąd jest pomijalnie mały, wówczas można zastosować taki rozkład.

Przystąpmy obecnie do analizy jednorodności. Przypomnijmy, że weryfikujemy hipotezę  $H_0$ , głoszącą, iż wszystkie elementy próby pochodzą z jednej populacji, mimo iż jeden element ciągu ma wartość zdecydowanie różną od innych elementów ciągu.

Do weryfikacji hipotezy możemy użyć statystyki:

$$T_k = \frac{X_k - \bar{X}_{n-1}}{S_{n-1}} \sqrt{\frac{n}{n-1}}$$

gdzie

$$S_{n-1} = \sqrt{\frac{1}{n-2} \sum_{i=1, i \neq k}^n (X_i - \bar{X}_{n-1})^2}$$

jest estymatorem odchylenia standardowego obserwowanej zmiennej, bez elementu odstającego. Jeżeli hipoteza  $H_0$  jest prawdziwa, to zmienna losowa  $T_k$  ma rozkład *Studenta* z  $(n-2)$  stopniami swobody (w przypadku prawdziwości hipotezy  $H_1$  zmienna losowa  $T_k$  ma niecentralny rozkład *Studenta*). Hipotezę sprawdzaną należy odrzucić na poziomie istotności  $\alpha$ , gdy zajdzie nierówność:

$$T_k \geq t_{\alpha}(n-2),$$

gdzie  $t_{\alpha}(n-2)$  to kwanty rzędu  $\alpha$  rozkładu *t Studenta* o  $n-2$  stopniach swobody.

Rozważmy przykład.

■ W kopalni stożkowej *Nchanga Open Pit* w Zambii notowano czasy czystego załadunku (sam czas załadunku bez uwzględniania czasu manewrowania wozu i czasu manewrowania koparki) wozów odstawczych typu *Lectra Haul* o udźwigu użytecznym 120 t przez koparkę łyżkową typu *P & H*.



Relacja pomiędzy pojemnością łyżki a pojemnością skrzyni wozu tak dobrano, by do załadowania wystarczyły tylko 3 łyżki. Zanotowany ciąg 29 czasów przedstawiał się następująco:

1,5; 1,8; 1,6; 1,9; 2,6; 2,2; 2,1; 2,0; 1,8; 2,0; 5,3; 2,3; 2,1; 2,5; 2,1;

2,7; 1,9; 1,7; 1,6; 2,4; 2,3; 2,2; 2,9; 2,0; 1,8; 1,9; 2,7; 1,6; 1,5.

Czas podany był w minutach. Zwrócono uwagę na jedenasty czas, wyraźnie dłuższy od pozostałych, wynoszący 5,3 min. Wiedząc, że rozkład normalny (bo z takim rozkładem mamy zwykle do czynienia, opisując czas fazy cyklu pracy wywrotki) jest odporny na odstawanie, stwierdzono, że czas ten należy wyłączyć z dalszych rozważań.

Sprawdźmy słuszność powziętej decyzji, posługując się uprzednio omówionym testem statystycznym. Mamy:

$$\bar{X}_{n-1} = 2,06 \text{ min} \quad S_{n-1} = 0,38 \text{ min} \quad \text{i} \quad T_k = 8,6.$$

Założmy tym razem poziom istotności  $\alpha = 0,02$ . Wartość krytyczna, odczytana z tablic dla rozkładu *Studenta* [6], to:

$$t_{\alpha=0,05}(27) = 2,5.$$

Jak widać, wartość empiryczna znacznie przekracza wartość krytyczną. Autor zadał sobie trud, by dojść do przyczyny zwiększenia tego czasu. Okazało się, że operator koparki załadowywał wtedy duży głaz, odstrzelony w ostatnich robotach przygotowawczych.

#### 2.4. Brak informacji o rozkładzie

Pozostaje nam rozpatrzyć przypadek, w którym nie mamy żadnej informacji na temat rozkładu badanej zmiennej i nie jesteśmy także w stanie wysunąć hipotezy na ten temat, mimo iż dysponujemy właśnie pobraną próbą. Może tak być w przypadku, gdy próba nie jest zbyt liczna bądź zastosowane testy identyfikacji rozkładu dają niejednoznaczną odpowiedź.

W takim przypadku przypuszczalnie najstarszą praktykowaną zasadą jest zastosowanie tzw. reguły  $3\sigma$ . Według tej zasady obszarem odrzucenia jest domena wszystkich tych wartości  $X$ , dla których prawdziwa jest następująca nierówność:

$$X > \bar{X}_{n-1} + 3S_{n-1}.$$

Choć trudno jest jednoznacznie określić, jakie jest dokładnie prawdopodobieństwo zajścia powyższego zdarzenia (zależy to bowiem od typu rozkładu z jakim mamy tu do czynienia) wiadomo jednak, że prawdopodobieństwo to jest bardzo małe, rzędu co najwyżej kilku procent. Mocniejszym stwierdzeniem, w tym względzie, jest tutaj oszacowanie, niestety tylko graniczne, podane w nierówności *Czebyszewa*, według którego:

$$P\{|X - E(X)| \leq k\sigma_x\} \geq 1 - k^{-2},$$

gdzie:  $E(X), \sigma_x$  to wartość oczekiwana i odchylenie standardowe zmiennej losowej  $X$ , natomiast  $k > 0$ . Jeżeli przyjmiemy  $k=3$ , to według nierówności *Czebyszewa* oszacowanie dolne prawdopodobieństwa wynosi ok. 0,9.

W statystyce znanych jest szereg dalszych testów, pozwalających na wytypowanie elementów odstających. Rozpatrzmy następujący test, typujący elementy odstające.

**Test  $Q$ .** Dla weryfikacji hipotezy o elemencie odstającym w tym teście niezbędne jest uporządkowanie zaobserwowanej próby w ciąg, załóżmy, niemalejący. Przyjmijmy, że otrzymaliśmy ciąg w postaci:

$$\{y_1 \leq y_2 \leq \dots \leq y_n < X_k\}.$$

Budujemy statystykę:

$$Q = \frac{X_k - y_n}{X_k - y_1}.$$

Jest to empiryczna wartość statystyki  $Q$ . Porównuje się ją z wartością krytyczną [4], gdyż zachodzi równanie:

$$P\{Q \geq Q_\alpha(n)\} = \alpha.$$

W przypadku gdy:

$$Q \geq Q_\alpha(n),$$

wówczas weryfikowaną hipotezę odrzuca się na poziomie istotności  $\alpha$ .

### 3. Podsumowanie

Zaprezentowane tu metody zaczerpnięte ze statystyki matematycznej pozwalają ocenić, z jak rzadkim przypadkiem mamy do czynienia. Do badającego należy decyzja, czy odrzucić element odstający czy go włączyć do dalszej analizy. Odnotujmy na koniec rozważań następującą prawidłowość. W statystyce matematycznej znamy wiele rozkładów prawdopodobieństwa, charakteryzujące się różnymi właściwościami. Jedną z interesujących własności jest tzw. *odporność na odstawanie*. Chodzi właśnie o możliwość pojawienia się wartości nietypowej w pobranej próbie z danego rozkładu. Już w latach siedemdziesiątych ubiegłego wieku odporność ta była zdefiniowana w sensie normatywnym i przebadano szereg rozkładów pod tym kątem. Rozkładem odpornym na odstawanie okazał się np. rozkład normalny, natomiast rozkład gamma (a zatem i wykładniczy) i rozkład *Weibulla* rozkładami nieodpornymi. Oznacza to, że może pojawić się wartość nietypowa, a próba będzie jednorodna. Pamiętajmy, że zawsze – jeżeli to tylko możliwe – należy przy analizowaniu wartości nietypowej sięgnąć do źródeł dokładnego opisu, co wtedy się stało i tam szukać odpowiedzi na pytanie, czy element ten można zaliczyć do badanej populacji czy nie.

#### LITERATURA

1. Czekala M.: Grupy obserwacji nietypowych – test wielowymiarowy. w: „Postępy ekonometrii”. Praca zbiorowa pod redakcją A. S. Barczaka. AE, Katowice 2004. str. 15÷21.
2. Gnienenko B. W., Bielajew J. K., Sołowiew A. D.: Metody matematyczne w teorii niezawodności. WNT, Warszawa 1968.
3. Fisher R. A.: Test of significance in harmonic analysis. Proc. Roy. Soc. A 125, 1929, s.54-59.
4. Szepe R.: Radiometria stosowana. WNT, Warszawa 1967.
5. Zeliaś A.: Metody wykrywania obserwacji nietypowych w badaniach ekonomicznych. Wiadomości Statystyczne, nr 8, 1996.
6. Zieliński R.: Tablice statystyczne. PWN, Warszawa 1972.

Recenzent: Prof. dr hab. Leon Dziembała