

**Magdalena Chmieleńska**

Uniwersytet Ekonomiczny w Katowicach

# ZASTOSOWANIE ESTYMACJI JĄDROWEJ DO MONITOROWANIA PROCESU O NIEZNANYM ROZKŁADZIE

## Wprowadzenie

Karta kontrolna jest prostym i efektywnym narzędziem statystycznej kontroli jakości. Pozwala ona dokonać oceny poprawności procesu, ale pod warunkiem spełnienia podstawowych jej założeń. W przypadku naruszenia założeń klasyczne karty kontrolne mogą dawać fałszywe sygnały, dla procesu uregulowanego, bądź nie zasygnalizować lub zasygnalizować z pewnym opóźnieniem fakt rozregulowania procesu.

W opracowaniu przedstawiono metodę wyznaczania linii kontrolnych na podstawie oszacowanej w drodze estymacji jądrowej gęstości zmiennej losowej. Metoda ta jako metoda nieparametryczna nie wymaga spełnienia założeń odnośnie do rozkładu badanej charakterystyki, więc może być stosowana dla danych o dowolnym, nieznanym rozkładzie.

## 1. Karta kontrolna

Karta kontrolna jest instrumentem umożliwiającym statystyczne monitorowanie procesu wytwórczego. Dostarcza ona informacji o jego przebiegu, jak również przyczynia się do podnoszenia jakości produkcji. Jej cechami charakterystycznymi są prostota konstrukcji i skuteczność.

Klasyczne karty kontrolne opierają się na założeniu, iż zmienna diagnostyczna obserwowana podczas monitorowania procesu ma rozkład normalny z wartością oczekiwaną  $\mu$  i odchyleniem standardowym  $\sigma$  (Kończak, 2007). Zakładają one również, iż pomiary w kolejnych okresach czasowych są niezależne. Karta kontrolna jest wykresem służącym do rejestracji wyników bieżącej

kontroli jakości produktów. Wykreśla się na niej linię centralną – wyznaczającą wartość oczekiwaną monitorowanej charakterystyki, linie kontrolne (górną i dolną) oraz linie ostrzegawcze (górną i dolną).

Linie kontrolne są ustawione na takim poziomie, aby prawdopodobieństwo pojawienia się wartości poniżej dolnej lub powyżej górnej linii kontrolnej, przy procesie uregulowanym statystycznie, było odpowiednio małe. Klasycznie, przy założeniu normalności rozkładu badanej cechy, linie te są oddalone od linii centralnej o trzy odchylenia standardowe monitorowanego procesu wytwórczego. Linie ostrzegawcze są umieszczane z reguły na poziomie linii centralnej plus lub minus  $2\sigma$  (Grant, 1972). Ze względu na naruszenie założenia o normalności rozkładu badanej charakterystyki coraz częściej wyznacza się położenie linii kontrolnych na podstawie wartości odpowiednich kwantyli znanych rozkładów teoretycznych, które mogą być wykorzystywane do modelowania danych rzeczywistych.

Układ punktów na karcie kontrolnej w relacji do poziomu linii kontrolnych i ostrzegawczych stanowi podstawę do wnioskowania o poprawności przebiegu badanego procesu produkcyjnego. W literaturze jest rozważanych wiele sygnałów świadczących o rozregulowaniu procesu, natomiast w opracowaniu jako sygnał rozregulowania procesu uznaje się tylko punkt ponad górną bądź pod dolną linią kontrolną.

Karty kontrolne są charakteryzowane przez ARL (przeciętna długość serii do wystąpienia sygnału). Dla procesu uregulowanego statystycznie oznacza się ją jako  $ARL_0$ , a dla procesu rozregulowanego jako  $ARL_1$ .  $ARL_0$  powinna przyjmować wysokie wartości. Stanowi ona odwrotność prawdopodobieństwa wystąpienia błędu I. rodzaju, tzn. odwrotność prawdopodobieństwa wystąpienia sygnału ( $p_s$ ) dla pojedynczej obserwacji (próby):

$$ARL_0 = \frac{1}{\alpha} = \frac{1}{p_s} \quad (1)$$

Przeciętna długość serii do wystąpienia sygnału dla procesu rozregulowanego statystycznie wyraża się wzorem:

$$ARL_1 = \frac{1}{1 - \beta} \quad (2)$$

gdzie:

$\beta$  – prawdopodobieństwo niewykrycia rozregulowania w pierwszej próbie po wystąpieniu rozregulowania.

## 2. Estymacja jądrowa

Estymacja jądrowa jest metodą szacowania funkcji gęstości, która wymaga zdefiniowania. Niech  $f(x)$  będzie gęstością rozkładu zmiennej  $X$  w populacji, taką, że:

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \quad (3)$$

oraz niech  $X_1, \dots, X_n$  będzie próbą prostą wylosowaną z tej populacji. Estymatorem jądrowym gęstości  $f(x)$  nazywamy funkcję w postaci:

$$f_n(x; a_n) = \frac{a_n}{n} \sum_{i=1}^n K[a_n(x - X_i)], \quad (4)$$

gdzie  $(a_n)_{n \in \mathbb{N}}$  to ciąg liczb dodatnich rozbieżnych do nieskończoności i taki, że

$a_n = o(n)$ , tzn.  $\lim_{n \rightarrow \infty} \frac{a_n}{n} = 0$ , oraz  $K(x)$  – funkcja spełniająca warunki:

$$\int_{-\infty}^{\infty} |K(x)|^2 dx < \infty, \quad (5)$$

$$K(x) = K(-x) \text{ dla } x \in (-\infty; \infty), \quad (6)$$

$$\int_{-\infty}^{\infty} K(x) dx = 1, \quad (7)$$

$$\sup_{-\infty < x < \infty} |K(x)| \leq A < \infty, \quad (8)$$

$$\int_{-\infty}^{\infty} x^i K(x) dx = 0 \text{ dla } i = 1, \dots, s-1, \quad (9)$$

$$\int_{-\infty}^{\infty} x^s K(x) dx \neq 0, \quad (10)$$

$$\int_{-\infty}^{\infty} x^s K(x) dx < \infty, \quad (11)$$

gdzie  $s$  jest ustaloną liczbą naturalną (Domański, Pruska, 2000). Estymator jądrowy gęstości zależy od wyboru postaci jądra i od wyboru ciągu  $a_n$ .

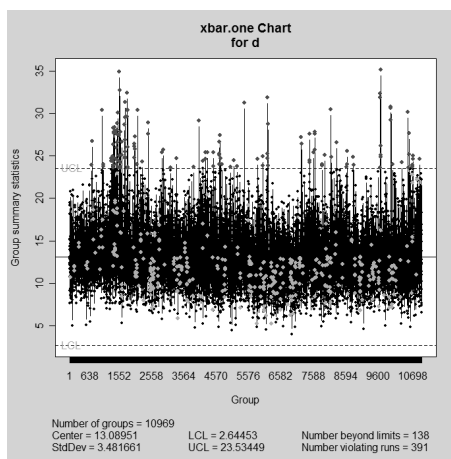
### 3. Opis danych

Prezentowane w artykule rozważania prowadzono z wykorzystaniem udostępnionych danych gromadzonych przez pewien zakład produkcyjny reprezentujący branżę motoryzacyjną prowadzący swą działalność na terenie województwa śląskiego. Analizowane dane pochodzą z kontroli międzyoperacyjnej pewnej charakterystyki produkowanego tam elementu. Udostępnione dane charakteryzują się tygodniowym horyzontem czasowym. Obejmują one informacje o prowadzonych kontrolach w okresie od 10.09.2012 do 15.09.2012. Łącznie stanowi to 10969 rekordów.

W omawianym okresie nie stwierdzono rozregulowania procesu produkcyjnego, gdyż wszystkie uzyskane podczas przeprowadzonej kontroli wartości mieściły się w wyznaczonych dla tej badanej charakterystyki normach. Analizowane dane można więc traktować jako dane pochodzące z procesu o prawidłowym przebiegu. Z tego względu dane te mogą posłużyć do wyznaczenia granic kontrolnych badanej charakterystyki.

### 4. Analiza poprawności procesu produkcyjnego

Analizę procesu produkcyjnego należy rozpocząć od sprawdzenia (najczęściej na podstawie danych historycznych) założeń karty kontrolnej. W omawianym przypadku zgodność rozkładu empirycznego danych z rozkładem normalnym została zweryfikowana za pomocą testu Kołmogorowa-Smirnowa z poprawką Lillieforsa oraz testu Shapiro-Wilka. Na ich podstawie należy odrzucić weryfikowaną hipotezę o normalności rozkładu badanej cechy.



Rys. 1. Klasyczna karta kontrolna wykreślona dla analizowanych danych

Ze względu na brak normalności rozkładu badanej cechy klasyczne karty kontrolne, wymagające zgodności empirycznego rozkładu danych z rozkładem normalnym, mogą prowadzić do błędnej oceny poprawności przebiegu procesu produkcyjnego. Rysunek 1 przedstawia klasyczną kartę kontrolną wykreśloną dla analizowanych danych (położenie linii centralnej i kontrolnych ustalono na podstawie wartości średniej i odchylenia standardowego wyliczonych dla dostępnych danych). Układ punktów na tej karcie wskazuje, iż proces nie przebiega w sposób prawidłowy, gdyż 138 spośród zarejestrowanych pomiarów osiąga wartości przekraczające wartość górnej linii kontrolnej. Prawdopodobieństwo sygnału wynosi więc  $p_s = 0,0128$ , a  $ARL_0 = 79$ . Dla porównania należy tutaj podać, iż dla klasycznej karty kontrolnej, przy spełnieniu jej założeń, wynosi  $ARL_0 = 370$ . Oznacza to, iż ze względu na brak normalności rozkładu badanej cechy klasyczne karty kontrolne, wymagające zgodności rozkładu danych z rozkładem normalnym, mogą prowadzić do błędnej oceny poprawności przebiegu procesu produkcyjnego.

#### 4.1. Nieklasyczne metody wyznaczania granic kontrolnych

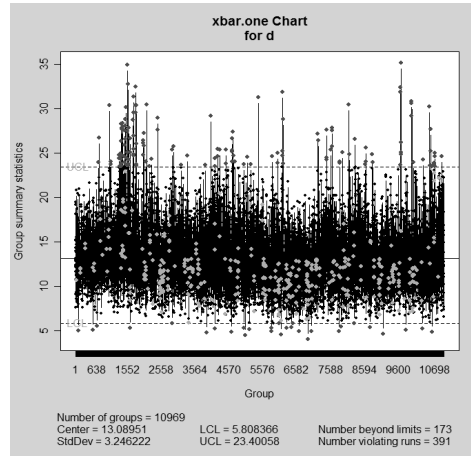
Ze względu na niebezpieczeństwo błędnej oceny prawidłowości przebiegu procesu produkcyjnego dla danych o rozkładzie empirycznym istotnie różnym od rozkładu normalnego granice kontrolne na karcie powinny zostać wyznaczone w inny sposób.

W opracowaniu rozważano dwa możliwe nieklasycznie sposoby wyznaczania linii kontrolnych. Pierwszy bazuje na kwantylach teoretycznego rozkładu zbliżonego kształtem do rozkładu empirycznego. Drugi opiera się natomiast na kwantylach oszacowanej w drodze estymacji jądrowej funkcji gęstości danych empirycznych.

W przypadku analizowanej cechy jest możliwe zastosowanie wielu znanych, teoretycznych rozkładów aproksymujących empiryczny rozkład danych. Jako przykład został wybrany rozkład gamma. Rozkład o parametrach  $\alpha = 15,201$ ,  $\beta = 0,861$  najlepiej spośród rozkładów z rodziny gamma przybliża rozkład monitorowanej charakterystyki. Przeprowadzony test zgodności Kołogomorowa-Lillieforsa odrzuca jednak hipotezę o zgodności obu rozkładów.

Karta kontrolna oparta na kwantylach rozkładu gamma o zadanych parametrach, prezentowana na rysunku 2, została wykreślona w celu zobrazowania konsekwencji braku zgodności obu rozkładów. Jej analiza skłania do wniosku, iż przybliżenie rozkładu empirycznego badanej charakterystyki rozkładem gamma i wnioskowanie o prawidłowości przebiegu tego procesu na podstawie tego teoretycznego rozkładu, prowadzi również do błędnej oceny procesu wytwórczego.

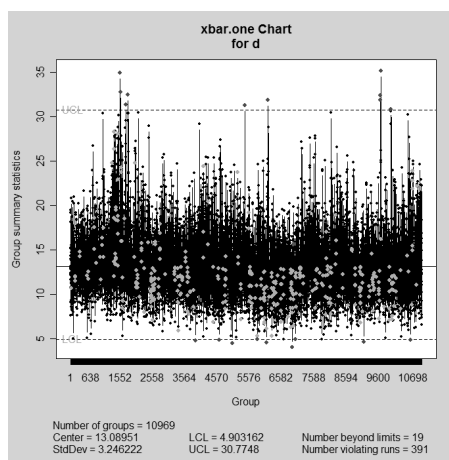
Karta ta wskazuje, iż analizowany proces jest rozregulowany statystycznie, o czym wskazują liczne (173) przekroczenia linii kontrolnych.



Rys. 2. Karta kontrolna wykreślona dla analizowanych danych, gdzie linie kontrolne wyznaczano na podstawie kwantyli funkcji gęstości rozkładu gamma

Ze względu na nieuzasadnione wnioskowanie o prawidłowości przebiegu procesu produkcyjnego na podstawie znanych rozkładów teoretycznych (rozkład normalny, gamma) uzasadnione jest stosowanie kart nieparametrycznych czy kart nieograniczonych postacią rozkładu (distribution – free charts) (Chakraborti, van der Laan, van de Wiel, 2004, s. 443-462). Proponowana metoda nie musi zakładać żadnego szczególnego rozkładu. Opiera się ona na oszacowaniu nieznannej funkcji gęstości empirycznego rozkładu danych metodą estymacji jądrowej (Wegman, 1972, s. 533-546).

Rysunek 3 przedstawia kartę kontrolną, gdzie granice kontrolne wyznaczono na podstawie wartości odpowiednich kwantyli oszacowanej metodą estymacji jądrowej zmiennej losowej. Na jej podstawie można wysnuć wniosek, iż proponowana metoda dobrze sprawdza się przy wyznaczaniu granic kontrolnych dla danych o innym niż normalny rozkładzie prawdopodobieństwa, gdyż ustala je na takim poziomie, iż prawdopodobieństwo uzyskania sygnału o rozregulowaniu w przypadku procesu uregulowanego statystycznie jest rzędu  $p_s = 0,001$  (dokładna wartość to  $p_s = 0,00087$ ).  $ARL_0$  wynosi więc  $ARL_0 = 577$ .



Rys. 3. Karta kontrolna wykreślona dla analizowanych danych, gdzie linie kontrolne wyznaczone na podstawie kwantyli funkcji gęstości rozkładu oszacowanej metodą estymacji jądrowej

Tabela 1 przedstawia podstawowe informacje o analizowanych metodach wykorzystanych do monitorowania procesu produkcyjnego na podstawie dostępnych danych. Proponowana metoda oparta na kwantylach funkcji gęstości rozkładu oszacowanej metodą estymacji jądrowej cechuje się najdłuższą spośród analizowanych metod przeciętną długością serii do wystąpienia sygnału. Największa wartość  $ARL_0$  i najmniejsza liczba przekroczeń granic kontrolnych wynika bezpośrednio z położenia linii kontrolnych. Należy zauważyć, iż w tym przypadku rozstęp pomiędzy granicami kontrolnymi jest największy.

Tabela 1

Wartości granic kontrolnych i liczba ich przekroczeń dla analizowanych danych w zależności od zastosowanej metody wyznaczania poziomu linii kontrolnych

Metoda wyznaczania granic kontrolnych	Klasyczna karta $\bar{x}$	Karta oparta na kwantylach rozkładu gamma	Karta oparta na kwantylach funkcji gęstości rozkładu oszacowanej metodą estymacji jądrowej
Liczba przekroczeń granic kontrolnych	138	179	19
$ARL_0$	79,49	63,40	577,32
LCL	2,64	5,81	4,9
UCL	23,53	23,4	30,77

Należy ponadto podkreślić, iż poziom, na jakim są ustalone linie kontrolne, wynika z przyjętego rozkładu stanowiącego podstawę karty kontrolnej. I tutaj ogromne znaczenie ma dobre dopasowanie oszacowanej funkcji gęstości zmiennej losowej do empirycznego rozkładu danych, otrzymane metodą estymacji jądrowej.

#### 4.2. Weryfikacja poprawności proponowanej metody wyznaczania granic kontrolnych na podstawie kwantyli rozkładu oszacowanej metodą estymacji jądrowej zmiennej losowej

W celu weryfikacji poprawności proponowanej metody wyznaczania linii kontrolnych na podstawie kwantyli funkcji gęstości rozkładu empirycznego oszacowanej w drodze estymacji jądrowej przeprowadzono symulacje komputerowe. Procedura symulacyjna obejmująca dwa warianty przebiegała etapowo. Wariant pierwszy zakładał proces przebiegający prawidłowo. W pierwszym kroku z udostępnionych danych losowano  $n$ -elementową próbkę, na podstawie której wyznaczano poziomy linii kontrolnych. Rozważano omówione trzy metody wyznaczania granic kontrolnych, tzn. metodę klasyczną i metody oparte na wartościach kwantyli rozkładu gamma o zadanych parametrach i rozkładu uzyskanego w wyniku estymacji jądrowej. Następnie wykreślano karty kontrolne z zadanymi granicami kontrolnymi i dla dostępnych danych rzeczywistych zliczano liczbę przekroczeń linii kontrolnych, prawdopodobieństwo sygnału oraz  $ARL_0$ . Procedurę powtarzano 10000-krotnie, uśredniając wyniki. Linie kontrolne wyznaczano dla próbek liczących 100, 500, 1000, 2000, 4000 i 5000 elementów.

W wariacie drugim, zakładającym rozregulowanie procesu, podobnie jak w wariacie pierwszym, posłużono się  $n$ -elementowymi próbkami do wyznaczania poziomu linii kontrolnych zgodnie z algorytmami analizowanych metod. Następnie generowano  $N = 10969$  wartości z rozkładu normalnego o parametrach  $\mu = 11$  i  $\sigma = 2,9$ . Wartości te nanoszono na przygotowane karty kontrolne, następnie zliczano liczbę przekroczeń linii kontrolnych, obliczano prawdopodobieństwo wystąpienia sygnału i  $ARL_1$ . Procedurę powtarzano 10000-krotnie, uśredniając wyniki.

Tabela 2

$ARL_0$  dla analizowanych danych w zależności od zastosowanej metody wyznaczania poziomu linii kontrolnych i liczebności próbki wstępnej (proces uregulowany)

Liczebność próbki $n$	Klasyczna karta $x$	Karta oparta na kwantylach rozkładu gamma	Karta oparta na kwantylach funkcji gęstości rozkładu oszacowanej metodą estymacji jądrowej
10969	79,49	63,40	577,32
100	69,54	24,52	45,05
500	76,86	53,01	181,34
1000	78,68	62,41	260,36
2000	79,31	70,66	350,67
4000	79,45	78,48	422,21
5000	79,47	81,58	438,41

Tabela 2 przedstawia przeciętną długość serii do wystąpienia sygnału dla analizowanych danych w zależności od sposobu wyznaczania linii kontrolnych i liczebności próbki wstępnej. Na jej podstawie można wysnuć wniosek, iż nie-



zależnie od zastosowanej metody, wraz ze wzrostem liczebności próbki wstępnej  $ARL_0$  rośnie. Przy założeniu prawidłowego przebiegu procesu produkcyjnego oznacza to, iż wraz ze wzrostem liczebności próbki wzrasta dokładność analizowanych metod wyznaczania linii kontrolnych.

Tabela 3

$ARL_1$  dla analizowanych danych w zależności od zastosowanej metody wyznaczania poziomu linii kontrolnych i liczebności próbki wstępnej (proces rozregulowany)

Liczebność próbki $n$	Klasyczna karta $\bar{x}$	Karta oparta na kwantylach rozkładu gamma	Karta oparta na kwantylach funkcji gęstości rozkładu oszacowanej metodą estymacji jądrowej
100	318,77	12,53	18,09
500	454,58	22,74	36,52
1000	474,56	25,57	41,39
2000	487,34	28,27	46,45
4000	494,23	30,14	50,06
5000	498,00	30,40	50,70

Na podstawie tabeli 2 można również wysnuć wniosek, iż próbka wstępna licząca 2000 elementów zapewnia  $ARL_0$  dla proponowanej metody na poziomie 350, co jest wynikiem porównywalnym z  $ARL_0$  dla klasycznej karty kontrolnej przy spełnieniu jej założeń. Ponadto z tabeli tej wypływa wniosek, iż dalszy wzrost wielkości próbki wstępnej powoduje coraz to mniejszy przyrost  $ARL_0$ . Z tego względu można uznać, iż w przypadku proponowanej metody próbka wstępna powinna liczyć minimum 2000 elementów.

Tabela 3 przedstawia wartość  $ARL_1$  dla danych pochodzących z rozkładu normalnego o parametrach  $\mu = 11$  i  $\sigma = 2,9$  w zależności od sposobu wyznaczania linii kontrolnych i liczebności próbki wstępnej. Jej analiza skłania do wniosku, iż klasyczna karta kontrolna, w której położenie linii centralnej, kontrolnych i ostrzegawczych determinuje wartość średnia i odchylenie standardowe uzyskane z próbki wstępnej (w przypadku analizowanych danych kształtują się one na poziomie  $\bar{x} = 13,09$  i  $s = 3,48$ ), niezależnie od wielkości próbki wstępnej, jest nieczuła na wprowadzone zaburzenie procesu (duże wartości  $ARL_1$ ). Pozostałe dwie metody zareagowały prawidłowo na to rozregulowanie, przy czym przeciętna długość serii do wystąpienia sygnału dla karty opartej na kwantylach rozkładu gamma jest niższa, niezależnie od wielkości próbki wstępnej, niż dla karty opartej na kwantylach funkcji gęstości rozkładu oszacowanej metodą estymacji jądrowej. Niskie wartości  $ARL_1$  w przypadku procesu rozregulowanego statystyczne świadczą o wrażliwości karty kontrolnej na rozregulowanie procesu i tym samym o jej skuteczności.

## Podsumowanie

Prezentowane w artykule rozważania na temat monitorowania procesów produkcyjnych cechujących się innym niż normalny rozkładem badanej charakterystyki skłaniają do wniosku, iż stosowanie klasycznego podejścia opartego na często naruszonym założeniu o normalności rozkładu badanej charakterystyki prowadzi do błędnej oceny poprawności monitorowanego procesu.

Dla danych o rozkładzie empirycznym istotnie różnym od znanych teoretycznych rozkładów istnieje jednak metoda wyznaczania granic kontrolnych, która zapewnia, iż sygnały świadczące o rozregulowaniu procesu w przypadku procesu uregulowanego statystycznie pojawiają się na karcie kontrolnej z założonym niskim prawdopodobieństwem.

Metoda ta opiera się na wartościach odpowiednich kwantyli wyznaczonej w drodze estymacji jądrowej funkcji gęstości rozkładu empirycznego. Proponowany sposób monitorowania poprawności przebiegu procesu produkcyjnego o nieznanym rozkładzie badanej charakterystyki posiada pożądane cechy dobrych kart kontrolnych. Proponowana karta szybko reaguje na wzrost wadliwości procesu produkcyjnego, jak również nie sygnalizuje zbyt często jego rozregulowania dla procesu przebiegającego prawidłowo.

## Literatura

- Chakraborti S., Laan P. van der, Wiel M.A. van de (2004): *A Class of Distribution-free Control Charts*. „Journal of the Royal Statistical Society”, No. 3, Vol. 53, s. 443-462.
- Domański Cz., Pruska K. (2000): *Nieklasyczne metody statystyczne*. PWE, Warszawa.
- Grant E.L. (1972): *Statystyczna kontrola jakości*. PWE, Warszawa.
- Kończak G. (2007): *Metody statystyczne w sterowaniu jakością produkcji*. Wydawnictwo Akademii Ekonomicznej, Katowice.
- Wegman E.J. (1972): *Nonparametric Probability Density Estimation: I. A Summary of Available Methods*. „Technometrics”, No. 3, Vol. 14, s. 533-546.

## THE USE OF THE KERNEL ESTIMATION FOR MONITORING PROCESSES WITH UNKNOWN DISTRIBUTIONS

### Summary

The control chart is a tool of statistical quality control, which is widely used in factories. The fulfillment of its basic assumptions ensures faultless assessing the monitored process. Infringements the assumptions of classical control charts can cause false signals

in the case of a regulated process, either lack of signal or the signal delayed in time, when process is out-of-control. Incorrect assessment of the accuracy of the manufacturing process is of course the economic impact.

In this paper, based on actual data an attempt to determine control limits for the manufacturing process of the distribution of the controlled characteristics, which is significantly different from a normal distribution, was taken. The result of this work is the method of determining the control limits based on the quantile of a random variable estimated by the kernel estimation. The article pays attention to the economic consequences of infringements the assumptions of classical control charts.