ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 156, 2002

*Karol Andrzejczak**

WYBRANE METODY ANALIZY CENZUROWANYCH CZASÓW ZDATNOŚCI PRODUKTÓW

Streszczenie. Badanie czasu zdatności produktów, z różnych powodów, może być ograniczone w czasie. W takich przypadkach mogą pojawić się tzw. obserwacje cenzurowane, dla których nie są znane dokładne czasy zdatności. Można powiedzieć jedynie, że czasy zdatności pewnych produktów są większe od ich czasów monitorowania. Celem tej pracy jest przedstawienie wybranych metod analizy cenzurowanych czasów zdatności obejmujących:

- opisywanie czasów zdatności;
- estymację funkcji niezawodności, intensywności uszkodzenia oraz gęstości prawdopodobieństwa;
- dopasowywanie do danych rozkładów zdatności;
- porównywanie czasów zdatności dla dwóch lub większej liczby partii produktów;
- modele regresji.

Przedstawione metody początkowo były rozwijane i stosowane w naukach medycznych i biologicznych. Obecnie międzynarodowe konwencje i unormowania wymagają stosowania metod statystycznych, np. w kontroli jakości do badania czasu zdatności określonych produktów. Ponadto podane metody statystyczne są coraz częściej stosowanym narzędziem nie tylko w naukach społecznych, ekonomicznych czy inżynierskich, lecz również przez zarządzających firmami produkcyjnymi i usługowymi, szczególnie zaś przez zarządzających firmami ubezpieczeniowymi.

Zakres przedstawionych w tej pracy metod statystycznych, dotyczących danych cenzurowanych, jest ograniczony do tych, które są już dostępne w pakietach statystycznych i można je stosować z wykorzystaniem technik komputerowych. Ze względu na szybki rozwój i rozpowszechnianie oprogramowania specjalistycznego ograniczamy się do tych metod statystycznych, które są dostępne w module „Analiza przeżycia i regresja dla danych uciętych" pakietu STATISTICA oraz procedur „Kaplan", „Lifetab" i „Paramod" pakietu STATGRAPHICS.

Prezentowane w tej pracy metody można zastosować do podstawowego problemu poznawczego w badaniach inżynierskich dotyczących analizy cenzurowanych czasów zdatności, jakim jest rozstrzygnięcie – czy pewne zmienne towarzyszące wpływają na czas zdatności badanego produktu. Metody te są oparte na konstrukcji modelu regresji, w którym czas zdatności ma rozkład zależny od innych zmiennych. Zmienne, od których

* Dr, Instytut Matematyki, Politechnika Poznańska.

zależy czas zdatności, są oznaczone jako wektor kolumnowy $z = [z_1, ..., z_m]^T$, który może oznaczać, w zastosowaniach inżynierskich, warunki eksploatacji, konserwacji, przechowywania, technologiczne itp. W tej pracy wektor ten nazywamy wektorem warunków eksploatacji produktu. Z kilku powodów problemu badania zależności czasu zdatności produktu od warunków eksploatacji nie można rozwiązać stosując proste techniki regresji wielokrotnej. Po pierwsze, czas zdatności produktu przeważnie nie ma rozkładu normalnego - co poważnie narusza założenie regresji wielokrotnej przeprowadzanej zwykłą metodą najmniejszych kwadratów. Po drugie, występuje problem wykorzystania obserwacji cenzurowanych.

Słowa kluczowe: czas zdatności, cenzurowanie, testowanie, regresja, intensywność uszkodzenia, statystyki, estymator, Statistica, Statgraphics.

I. ZAŁOŻENIA I OZNACZENIA

Przyjmujemy, że modelem probabilistycznym czasu zdatności produktu pochodzącego z pewnej populacji jest nieujemna zmienna losowa T typu ciągłego. Niech $f(t)$ oznacza gęstość prawdopodobieństwa. Dystrybuanta $F(t)$ i funkcja niezawodności $S(t)$ są określone równościami:

$$F(t) = P(T \leqslant t) = \int_0^t f(x)dx, \quad S(t) = P(T \geqslant t) = \int_t^\infty f(x)dx \tag{1}$$

Intensywnością uszkodzenia (*hazard function*) lub intensywnością utraty zdatności produktu nazywamy funkcję $h(t)$ określoną jako granica:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t} \tag{2}$$

Iloczyn $h(t)\Delta t$ jest przybliżonym prawdopodobieństwem utraty zdatności produktu w przedziale $[t, t + \Delta t)$, pod warunkiem, że produkt był zdatny do chwili t. Intensywność uszkodzenia ma szczególne znaczenie w badaniu rozkładu czasu zdatności produktu. Bezpośrednio z określenia intensywności uszkodzenia i prawdopodobieństwa warunkowego wynikają następujące własności:

$$h(t) \geqslant 0 \tag{3}$$

$$h(t) = \frac{f(t)}{S(t)} \tag{4}$$

$$\int_t^\infty h(t)dt = \infty \tag{5}$$

$$S(t) = \exp\left(-\int_0^t h(u)du\right), \quad t \geqslant 0 \tag{6}$$

Z podanych własności wynika, że funkcja intensywności uszkodzenia stanowi w zastosowaniach jakościową informację pomocną w wyborze modelu czasu zdatności. Funkcja występująca w wykładniku własności (6) jest nazywana funkcją wiodącą. Funkcja ta jest oznaczana przez $H(t)$, czyli dla $t \geqslant 0$

$$H(t) = \int_0^t h(u)du \tag{7}$$

Funkcja wiodąca jest miarą wyczerpywania się „zapasu zdatności" produktu. Między podanymi funkcjami zachodzą związki (patrz Bobrowski (1985)):

$$S(t) = \exp(-H(t)), \quad H(t) = \ln\frac{S(0)}{S(t)}, \tag{8}$$

Zależności między charakterystykami funkcyjnymi czasu zdatności pozwalają na uzyskiwanie odpowiednich informacji o czasie zdatności produktu pochodzącego z danej ich populacji.

Obserwację czasu zdatności T nazywamy prawostronnie cenzurowaną lub krótko cenzurowaną, jeżeli $T > M$, gdzie M oznacza deterministyczny lub losowy czas monitorowania zdatności produktu. Oczywiście, jeżeli $T \leqslant M$, to w czasie monitorowania zdatności produktu zaobserwujemy dokładny jego czas zdatności.

II. ESTYMACJA ROZKŁADÓW

Dla celów predykcyjnych pożądana jest znajomość rozkładu czasu zdatności produktów pochodzących z pewnej ich populacji. W zastosowaniach inżynierskich do najczęściej stosowanych rozkładów, w modelowaniu czasów zdatności, należą rozkłady: wykładniczy, liniowy intensywności uszkodzenia, Weibulla i Gompertza. Zasadniczo funkcje intensywności uszkodzenia, w szczególności przekształcenia logarytmiczne funkcji intensywności uszkodzenia, podanych rozkładów teoretycznych są liniowymi funkcjami czasów zdatności lub logarytmów czasów zdatności. Zatem funkcje intensywności uszkodzenia można wyrazić w kategoriach funkcji regresji liniowej:

- dla rozkładu wykładniczego $h(t) = 1/\alpha$ (9)
- dla rozkładu liniowo-wykładniczego $h(t) = \alpha + \beta t$ (10)

– dla rozkładu Weibulla $h(t) = (\beta/\alpha)(t/\alpha)^{\beta-1}$ (11)

– dla rozkładu Gompertza $h(t) = \exp(\alpha + \beta t)$ (12)

Estymacja parametrów tych rozkładów jest oparta na metodzie regresji liniowej ważonych najmniejszych kwadratów. Minimalizowana jest wielkość:

$$WSS = \Sigma\left(w_i(y_i - a - bx_i)^2\right) \tag{13}$$

dla wag $w_i = \frac{1}{v_i}$ oraz $w_i = n_i \cdot h_i$,

gdzie:

v_i – wariancja estymatora intensywności uszkodzenia,

h_i – szerokość i-tego przedziału,

n_i – liczba produktów narażonych na utratę zdatności w i-tym przedziale.

W szczególności dla $w_i = 1$ otrzymujemy nieważone najmniejsze kwadraty.

Wynik dopasowywania modelu do danych empirycznych można przedstawić graficznie w postaci wykresów funkcji niezawodności, gęstości prawdopodobieństwa lub intensywności uszkodzenia nałożonych na ich empiryczne odpowiedniki. Wykresy te pozwalają na wizualną ocenę jakości dopasowania modelu oraz na sumaryczne porównanie pomiędzy grupami dla danych pogrupowanych.

Jedną z najstarszych metod analizy danych cenzurowanych dotyczących czasu zdatności jest tzw. tablica trwania zdatności. Stanowi ona rozbudowaną tablicę rozkładu liczebności. Przedział obserwacji jest dzielony na przedziały. Dla każdego przedziału są obliczane:

– odsetek monitorowanych produktów, które utraciły zdatność;

– odsetek monitorowanych produktów, które nie utraciły zdatności;

– skumulowany odsetek produktów, które nie utraciły zdatności oraz błędy standardowe tego estymatora;

– empiryczna gęstość prawdopodobieństwa będąca oceną prawdopodobieństwa utraty zdatności w danym przedziale obliczona na jednostkę czasu i błąd standardowy tego estymatora;

– empiryczna intensywność utraty zdatności produktu i odpowiedni błąd standardowy tego estymatora.

Dla otrzymania rzetelnych oszacowań trzech głównych funkcji, tj. niezawodności, gęstości prawdopodobieństwa i intensywności uszkodzenia zalecane jest za Lee (1980), aby minimalna wielkość próby wynosiła 30. Oceny tablicy trwania zdatności oblicza się według standardowych formuł opisanych w literaturze: Lawless (1982), Lee (1980), Nelson (1982). Dane stabelaryzowane określają trzy zmienne zawierające informacje o początkach przedziałów, liczbach obserwacji cenzurowanych i liczbach produktów, które utraciły zdatność w kolejnych przedziałach czasowych.

Nieparametrycznej oceny empirycznej funkcji niezawodności można dokonać stosując estymator Kaplana–Meiera (1958) limitu iloczynowego. Estymatorem Kaplana–Meiera funkcji niezawodności nazywamy statystykę

$$\tilde{S}(t) = \prod_{j:t_j < t} \frac{n-j}{(n-j+1)^{\delta(j)}} \tag{14}$$

gdzie n to całkowita liczba monitorowanych czasów zdatności, $\delta(j)$ to stała, która wynosi 1, jeśli j-ty czas zdatności nie jest cenzurowany, a 0, jeśli jest cenzurowany.

Przewaga metody limitu iloczynowego nad metodą tablic trwania zdatności polega na tym, że otrzymane estymatory nie zależą od sposobu grupowania danych na przedziały czasowe. W przypadku danych pogrupowanych wstępnej identyfikacji rozkładów i interpretacji różnic pomiędzy funkcjami niezawodności dla grup można dokonać na podstawie wykreślenia estymatorów Kaplana–Meiera empirycznych funkcji niezawodności dla analizowanych grup. Jeżeli przedziały tablicy trwania zdatności zawierają po co najwyżej jednej obserwacji, to metoda Kaplana–Meiera i metoda tablic trwania życia są identyczne.

III. PODSTAWOWE TESTY DLA CENZUROWANYCH CZASÓW ZDATNOŚCI

Do podstawowych testów porównania czasów zdatności produktów pochodzących z dwóch populacji przeprowadzanych na podstawie danych zawierających obserwacje cenzurowane należą:
- uogólnienia Gehana testu Wilcoxona;
- test Coxa–Mantela;
- test F Coxa;
- test logarytmiczny rang;
- uogólnienie Peto i Peto testu Wilcoxona;
- test Mantela–Haenszela.

Statystyka Gehana jest obliczana za pomocą procedury Mantela (1967). Sposób wyznaczania statystyki F-Coxa i Coxa–Mantela jest podany przez Coxa (1972). Test log-rank i uogólnienie Peto i Peto są podane przez Peto i Peto (1972). Efektywność tych testów przy małych próbach nie jest do końca zbadana i rzetelne wyniki można uzyskać przy odpowiednio dużych próbach. Problem mocy tych testów jest przedstawiony przez Lee (1980). Test Coxa–Mantela i test logarytmiczny rang są mocniejsze od testu Gehana, bez względu na cenzurowanie, gdy próby są pobierane z populacji o rozkładzie wykładniczym lub Weibulla. Test Mantela–Haenszela (1959) służy do porównywania dwóch populacji, przy kontroli wpływu zmiennej kategoryzującej. Test ten opiera się na analizie tablic 2×2 podzielonych względem innej zmiennej kategoryzującej. Test umożliwia określenie czy dwie zmienne uwzględnione w tablicach 2×2 są ze sobą powiązane po wykluczeniu wpływu zmiennej kategoryzującej.

Testy dla wielu prób są rozwinięciami:
- uogólnienia Gehana testu Wilcoxona;
- uogólnienia Peto i Peto testu Wilcoxona;
- testu log-rank.

Porównując wiele prób, najpierw każdemu czasowi zdatności przypisujemy punkty za pomocą procedury Mantela (1967); następnie opierając się na sumach (dla każdej grupy) tych punktów obliczamy wartości chi-kwadrat. W przypadku tylko dwóch grup test ten jest równoważny uogólnieniu Gehana testu Wilcoxona.

Ideę konstrukcji testów przedstawiamy na przykładzie modelu proporcjonalnej intensywności uszkodzenia. W modelu proporcjonalnej intensywności uszkodzenia przyjmujemy założenie, że iloraz warunkowych intensywności uszkodzenia

$$h(t|\mathbf{z}_1)/h(t|\mathbf{z}_2) \tag{15}$$

dla dwóch różnych warunków eksploatacji produktu $\mathbf{z}_1$, $\mathbf{z}_2$ jest niezależny od czasu t. Klasa modeli czasów zdatności z tą własnością jest oznaczana *PH* (*proportional hazards*). Dla tej klasy czasów zdatności warunkową intensywność uszkodzenia produktu $h(t|\mathbf{z})$, przy warunkach eksploatacji $\mathbf{z}$, możemy przedstawić w postaci

$$h(t|\mathbf{z}) = h_0(t)g(\mathbf{z}) \tag{16}$$

Występująca w (16) funkcja $h_0(t)$ nazywa się bazową intensywnością uszkodzenia. Bazowa intensywność uszkodzenia jest intensywnością uszkodzenia produktu, dla którego $g(\mathbf{z}) = 1$. W ogólności obydwie funkcje h_0 i g mogą zawierać nieznane parametry. Model *PH* jest bardzo ogólny, gdyż nie jest on oparty na jakichkolwiek założeniach dotyczących rozkładu czasu zdatności produktu. Do klasy modeli *PH* należą m. in. rozkłady: Weibulla, wykładniczy, log-normalny i gamma. W przypadku, gdy funkcja g zawiera nieznane parametry $\beta_1, \ldots, \beta_m$, warunkową intensywność uszkodzenia przy warunkach eksploatacji $\mathbf{z}$ zapisujemy w postaci

$$h(t|\mathbf{z}) = h_0(t)g(\mathbf{z}; \beta_1, \ldots, \beta_m) \tag{17}$$

Przykładowo warunkowa gęstość prawdopodobieństwa zmiennej losowej T o rozkładzie Weibulla ma postać:

$$f(t|\mathbf{z}) = \frac{\gamma}{\alpha(\mathbf{z})}\left(\frac{t}{\alpha(\mathbf{z})}\right)^{\gamma-1} \exp\left(-\left(\frac{t}{\alpha(\mathbf{z})}\right)^{\gamma}\right), \ t \geqslant 0 \tag{18}$$

a warunkowa intensywność uszkodzenia przy danych warunkach eksploatacji $\mathbf{z}$ przyjmuje postać

$$h(t|\mathbf{z}) = \frac{\gamma}{\alpha(\mathbf{z})}\left(\frac{t}{\alpha(\mathbf{z})}\right)^{\gamma-1}, \ t \geqslant 0 \tag{19}$$

gdzie α jest parametrem skali zależnym od warunków eksploatacji $\mathbf{z}$, a γ jest parametrem kształtu tego rozkładu. Warunkowa wartość oczekiwana oraz warunkowa wariancja czasu zdatności T są funkcjami zależnymi od warunków eksploatacji produktu:

$$E(T|\mathbf{z}) = \alpha(\mathbf{z})\Gamma(1/\gamma)/\gamma, \quad Var(T|\mathbf{z}) = \alpha^2(\mathbf{z})(2\Gamma(2/\gamma) - (1/\gamma)\Gamma^2(1/\gamma))/\gamma \tag{20}$$

Iloraz intensywności uszkodzenia przy różnych warunkach eksploatacji $\mathbf{z} = \mathbf{z}_1$, $\mathbf{z} = \mathbf{z}_2$ jest następujący:

$$\frac{h(t|\mathbf{z}_1)}{h(t|\mathbf{z}_2)} = \left(\frac{\alpha(\mathbf{z}_2)}{\alpha(\mathbf{z}_1)}\right)^{\gamma} \tag{21}$$

i jak widać nie zależy od czasu t, czyli rozkład ten należy do klasy rozkładów PH.

Szczególnie użyteczną klasą modeli czasów zdatności typu PH są modele Coxa. Modele PH Coxa są to modele czasów zdatności, których intensywność utraty zdatności można przedstawić w postaci:

$$h(t|z_1, ..., z_m) = h_0(t) \cdot \exp(\beta_1 z_1 + ... + \beta_m z_m) \tag{22}$$

gdzie funkcja h jest intensywnością uszkodzenia przy danych wartościach m zmiennych kodujących przypadek $(z_1, z_2, ..., z_m)$ i czasu zdatności t. Czynnik $h_0(t)$ jest intensywnością bazową. Parametry $\beta_1, ..., \beta_m$ są współczynnikami regresji.

Funkcja wiarogodności dla modeli z proporcjonalnymi intensywnościami uszkodzenia ma postać

$$L(\beta_1, ..., \beta_m; h_0(t)) =$$
$$= \prod_{i=1}^{n} (h_0(t_i)\exp(z_{i1}\beta_1 + ... + z_{im}\beta_m) S_0(t_i)^{\exp(z_{i1}\beta_1 + ... + z_{im}\beta_m)})^{\delta_i} (S_0(t_i)^{\exp(z_{i1}\beta_1 + ... + z_{im}\beta_m)})^{1-\delta_i} \tag{23}$$

gdzie t_i jest obserwowanym czasem zdatności albo czasem cenzurowanym dla i-tego produktu, δ_i jest zmienną wskaźnikową, przyjmującą wartość 1, jeżeli t_i jest czasem zdatności oraz 0, jeżeli t_i jest czasem cenzurowanym. Przyjmując dla funkcji wiarogodności szczególną postać, otrzymamy pełny parametryczny model *PH*. Na przykład dla bazowej intensywności

$$h_0(t) = \gamma\alpha^{-1}(t/\alpha)^{\gamma-1} \tag{24}$$

otrzymujemy w pełni sparametryzowany model Weibulla czasu zdatności. W konstrukcji testu nie zakładamy żadnej szczególnej postaci dla funkcji intensywności. Oczywiście, jeśli dane pochodzą ze szczególnego modelu, np. Weibulla, to stosując ogólne podejście zamiast odpowiedniego parametrycznego modelu, traci się na efektywności.

Dla przejrzystości prześledzimy najpierw ideę konstrukcji testów na przykładzie modelu *PH* Coxa dla dwóch populacji produktów. W tym przypadku jest rozważana hipoteza zerowa

$$H_0: S_1(t) = S_2(t) \tag{25}$$

o równości funkcji niezawodności dla dwóch populacji. Z założenia warunkowa funkcja niezawodności ma postać:

$$S(t|\mathbf{z}) = S_0(t)^{exp(\beta_1 z_1 + \ldots + \beta_m z_m)} \tag{26}$$

gdzie $\beta = [\beta_1, \ldots, \beta_m]^T$ jest wektorem parametrów związanych z warunkami eksploatacji $\mathbf{z}$.

W przypadku dwóch populacji do weryfikacji hipotezy zerowej są wykorzystane obserwacje czasów zdatności produktów z zaznaczeniem za pomocą zmiennej z przyjmującej wartości 0 lub 1, z której populacji produkt pochodzi. W przyjętym modelu *PH* Coxa intensywności uszkodzenia produktów eksploatowanych w dwóch różnych warunkach przyjmują postać

$$h_1(t) = h_0(t), \quad h_2(t) = h_0(t)e^{\beta} \tag{27}$$

i rozkłady czasów zdatności są identyczne wtedy i tylko wtedy, gdy $\beta = 0$. Jest to równoważne założeniu, że $S_2(t) = S_1(t)^{exp\beta}$ i sprawdzenie, czy $\beta = 0$ oznacza testowanie postawionej hipotezy zerowej przeciw hipotezie alternatywnej

$$H_A: S_2(t) = S_1(t)^{\delta} \quad \text{dla } \delta \neq 1 \tag{28}$$

zwanej rodziną Lehmanna czasów zdatności.

Niech N_1, N_2 będą liczebnościami prób pochodzących z dwóch populacji produktów eksploatowanych w dwóch różnych warunkach W_1 i W_2 oraz niech $N = N_1 + N_2$. Dopuszczamy obserwacje cenzurowane. W dalszym ciągu dwie próby traktujemy jako jedną próbę łączną, przy czym korzystamy z dodatkowej zmiennej z, zwanej zmienną fikcyjną, do kodowania informacji, z której populacji obserwacja pochodzi. Przyjmujemy konwencję, że jeżeli obserwacja pochodzi z pierwszej próby, to $z = 0$, jeżeli z drugiej próby, to $z = 1$. Niech $t_{(1)} < ... < t_{(k)}$ oznacza k różnych zaobserwowanych czasów zdatności w łącznej próbie oraz niech d_i oznacza liczbę niezdatnych produktów w chwili $t_{(i)}$, natomiast d_{1i} oraz d_{2i} oznaczają liczby niezdatnych produktów w chwili $t_{(i)}$ eksploatowanych w warunkach W_1 albo W_2. Ponadto niech n_{1i} i n_{2i} oznaczają odpowiednio zmniejszające się liczebności monitorowanych czasów zdatności produktów, tj.

$$n_{ji} = Card\{l \in \{1, ..., N\}: t_{jl} \geqslant t_{(i)}\}, \ j = 1, 2 \tag{29}$$

gdzie t_{jl} oznacza czas zdatności l-tego produktu z j-tej próby.

Niech $n_{1i} + n_{2i} = n_i$ i $d_{1i} + d_{2i} = d_i$. Przy przyjętych oznaczeniach i założeniach logarytm wiarogodności dla parametru β

$$\log L(\beta) = \beta \Sigma d_{2i} - \Sigma d_i \log(n_{1i} + n_{2i} e^{\beta}) \tag{30}$$

Stąd

$$U(\beta) = \frac{\partial \log L}{\partial \beta} = \Sigma d_{2i} - \Sigma \frac{d_i n_{2i} e^{\beta}}{n_{1i} + n_{2i} e^{\beta}} \tag{31}$$

$$I(\beta) = \frac{-\partial^2 \log L}{\partial \beta^2} = \Sigma \frac{d_i n_{1i} n_{2i} e^{\beta}}{(n_{1i} + n_{2i} e^{\beta})^2} \tag{32}$$

Z równania wiarogodności $U(\beta) = 0$ można otrzymać estymator parametru β. Wnioski dotyczące parametru β można uzyskać stosując metodę ilorazu wiarogodności, przyjmując, że

$$\tilde{\beta} \approx N(\beta, I(\tilde{\beta})^{-1}) \quad \text{lub} \quad U(\beta) \sim N(0, \ I(\beta)) \tag{33}$$

W szczególności, przy założeniu hipotezy $H_0: \beta = 0$, statystyka

$$Z = \frac{U(0)}{(I(0))^{1/2}} \tag{34}$$

ma w przybliżeniu standaryzowany rozkład normalny.

W teście porównania czasów zdatności produktów eksploatowanych w m różnych warunkach definiujemy wektor $m-1$ zmiennych zero-jedynkowych $\mathbf{z}=[z_1, ..., z_{m-1}]$ kodujący numer populacji, z której obserwowany produkt pochodzi. Wektor $\mathbf{z}=\mathbf{0}$ oznacza, że obserwacja pochodzi z m-tej populacji, natomiast wektor $[0, ..., 0, 1_i, 0, ..., 0]$ dla $i=1, ..., m-1$ oznacza, że obserwacja pochodzi z i-tej populacji. W modelu Coxa czasów zdatności mamy

$$S_1(t)=S_0(t)^{\delta_1}, ..., S_{m-1}(t)=S_0(t)^{\delta_{m-1}}, \; S_m(t)=S_0(t), \text{ gdzie } \delta_i=\exp\beta_i \tag{35}$$

Test równości m rozkładów czasów zdatności jest równoważny temu, że $[\beta_1, ..., \beta_{m-1}]=\mathbf{0}$... Służy on do wykrywania odstępstw od równości m rozkładów typu PH. Zakładamy, że obserwacje czasów zdatności dotyczą ustalonej liczby $N=N_1+...+N_m$ produktów eksploatowanych w m różnych warunkach. Część obserwacji może być cenzurowana. Obserwacje dokonuje się z zadaną dokładnością Δ. Niech $t_{(1)}<...<t_{(k)}$ będą czasami zdatności wszystkich badanych produktów. Ponadto niech

– $n_i=card\{l\in\{1, ..., N\}: t_l \geqslant t_{(i)})$ oznacza liczbę produktów zdatnych do chwili $t_{(i)}$;

– $d_i=card\{l\in\{1, ..., N\}: t_{(i-1)}<t_l\leqslant t_{(i)}\}$, $t_{(0)}\equiv 0$;

– $n_{ri}=card\{l\in W_r: t_l\geqslant t_{(i)}\}$ oznacza liczbę produktów eksploatowanych w warunkach W_r, które są zdatne do chwili $t_{(i)}$;

– $d_{ri}=card\{l\in W_r: t_{(i-1)}<t_{(i)}\leqslant t_{(i)}\}$, $t_{(0)}\equiv 0$ oznacza liczbę produktów eksploatowanych w warunkach W_r, które utraciły zdatność w i-tym przedziale czasowym.

Elementy wektora pochodnych oraz macierzy informacyjnej przyjmują postać:

$$U_r(\mathbf{0})=\left(\frac{\partial \log L}{\partial\beta_r}\right)\Big|_{\beta=0}=\sum_{i=1}^{k}\left(d_{ri}-\frac{d_i n_{ri}}{n_i}\right) \quad r=1, ..., m-1 \tag{36}$$

$$I_{rs}(\mathbf{0})=\sum_{i=1}^{k} d_i\frac{n_{ri}}{n_i}\left(\delta_{rs}-\frac{n_{ri}}{n_i}\right) \quad r, s=1, ..., m-1 \tag{37}$$

gdzie $\delta_{rs}=1$ dla $r=s$ lub 0 dla $r\neq s$ jest deltą Kroneckera. Jeżeli hipoteza $\beta=\mathbf{0}$ jest prawdziwa, to statystyka $\mathbf{U}=[U_1(\mathbf{0}), ..., U_{m-1}(\mathbf{0})]^T$ jest zbieżna do wielowymiarowego rozkładu normalnego z wektorem wartości oczekiwanych $\mu=\mathbf{0}$ i macierzą kowariancji $\mathbf{I}(\mathbf{0})$. Do sprawdzenia hipotezy $\beta=\mathbf{0}$ o równości wszystkich funkcji niezawodności produktów eksploatowanych w różnych warunkach, przeciw hipotezie alternatywnej będącej zaprzeczeniem hipotezy zerowej można wykorzystać statystykę $\mathbf{U}^T\mathbf{I}(0)^{-1}\mathbf{U}$ mającą w przybliżeniu rozkład $\chi^2_{(m-1)}$. Test można przeprowadzać wtedy, gdy wskaźnik cenzurowanych czasów zdatności w próbie jest mały. W przeciwnym przypadku sprawdzanie hipotezy odbywa się z wykorzystaniem macierzy informacyjnej

$$I_{rs}(0) = \sum_{i=1}^{k} \frac{d_i(n_i - d_i)}{(n_i - 1)n_i} \left(\delta_{rs} - \frac{n_{si}}{n_i} \right) \quad r, s = 1, \ldots, m-1 \tag{38}$$

Test m populacji był omawiany przez kilku autorów z różnych punktów widzenia. Pierwszym, który zaproponował przedstawione podejście, był Mantel (1967).

IV. MODELE REGRESJI DLA CENZUROWANYCH CZASÓW ZDATNOŚCI

Niech T będzie ciągłą zmienną losową reprezentującą czas zdatności pewnego produktu. Niech $\mathbf{z} = [z_1, \ldots, z_m]$ będzie wektorem warunków eksploatacji, tj. wektorem zmiennych stowarzyszonych z badanym czasem zdatności. Przy wprowadzonych oznaczeniach warunkową intensywność uszkodzenia produktu dla modelu ciągłego czasu zdatności produktu T możemy przedstawić równością

$$h(t|\mathbf{z}) = \frac{-S'(t|\mathbf{z})}{S(t|\mathbf{z})} \tag{39}$$

Do podstawowych modeli regresji stosowanych w analizie cenzurowanych czasów zdatności należą:

a) model regresji wykładniczej,

b) normalny model regresji liniowej,

c) logarytmiczno-normalny model regresji liniowej,

d) model proporcjonalnej intensywności uszkodzenia Coxa,

e) model *PH* Coxa za zmiennymi warunkującymi zależnymi od czasu.

Model regresji wykładniczej. Jest to model parametryczny, w którym zakładamy, że czas zdatności badanych wyrobów ma rozkład wykładniczy uwarunkowany wartościami zmiennych $z_1, z_2, \ldots, z_m$. Warunkowa gęstość prawdopodobieństwa tego modelu wyrażana jest wzorem

$$f(t|\mathbf{z}) = \theta_z^{-1} \exp\left(\frac{-t}{\theta_z} \right), \quad r > 0 \tag{40}$$

Wektor $\mathbf{z}$ określa warunki eksploatacji produktu, parametr θ wyraża warunkowy oczekiwany czas zdatności

$$\theta_z = E(T|\mathbf{z}) \tag{41}$$

Możliwe są różne postaci funkcyjne dla parametru θ. Dla celów praktycznych bardzo użyteczną postacią jest

$$\theta_z = \exp(\beta_1 z_1 + \ldots + \beta_m z_m) \tag{42}$$

Adekwatność tego modelu ocenia się na podstawie funkcji logarytmu wiarogodności dla modelu z wszystkimi estymatorami parametrów (L_1) i logarytmu wiarogodności modelu, w którym do wszystkich zmiennych objaśniających podstawimy zero (L_0). Jeden ze sposobów sprawdzenia założenia wykładniczości modelu podany przez Lawlessa (1982) polega na wykreśleniu resztkowych czasów zdatności względem standardowej wykładniczej statystyki pozycyjnej α. Jeśli założenie wykładniczości jest spełnione, wszystkie punkty na wykresie powinny ułożyć się w przybliżeniu wzdłuż linii prostej. Model regresji dotyczący czasu zdatności dla cenzurowanych danych o rozkładzie wykładniczym jako pierwsi opisali Feigl i Zelen (1965). Algorytm otrzymywania estymatorów największej wiarogodności wykorzystuje metodę Newtona-Raphsona i jest przedstawiony w pracy Lagakos i Kuhns (1978).

Modele normalnej i log-normalnej regresji liniowej. W modelach tych zakładamy, że czas zdatności lub logarytm czasu zdatności ma rozkład normalny. Czas zdatności można wówczas przedstawić jako:

$$T = \alpha + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_m z_m + \varepsilon \tag{43}$$

i zastosować metody typowej regresji wielokrotnej. Normalny model regresji jest szczególnie przydatny, ponieważ wiele zbiorów danych można przekształcić tak, by dały aproksymacje rozkładu normalnego. W tym sensie jest to bardzo ogólny w pełni parametryczny model. Oszacowania parametrów dla wielu różnych ukrytych rozkładów przeżycia są rozważane przez Schneidera (1986). W celu otrzymania estymatorów największej wiarogodności dla parametrów modelu jest stosowany algorytm minimalizacji wartości oczekiwanej omówiony przez Coxa i Oakesa (1984).

W przypadku danych pogrupowanych można zastosować analizę warstwową do testowania hipotezy, czy identyczne modele regresji są właściwe dla różnych grup, tzn. czy zależność między zmiennymi warunkującymi a zdatnością jest identyczna w różnych grupach. Analiza ta polega na dopasowaniu odpowiedniego modelu regresji oddzielnie w ramach każdej grupy. Suma logarytmów wiarogodności tych analiz odzwierciedla logarytm wiarogodności modelu z różnymi wyrazami wolnymi i różnymi współczynnikami regresji w różnych grupach. Następnie jest dopasowywany model regresji do wszystkich danych, tj. z pominięciem przynależności do grupy i jest obliczany logarytm wiarogodności tego całkowitego dopasowania. Różnica między logarytmami wiarogodności jest wykorzystana do testowania istotności np. za pomocą statystyki chi-kwadrat.

Model *PH* Coxa. W ogólnym modelu *PH* Coxa nie czyni się żadnych założeń co do kształtu bazowej funkcji intensywności uszkodzenia. Ponieważ zachodzi równość

$$S(t|\mathbf{z}) = \exp\left(-\int_0^t h(u|\mathbf{z})du\right) \tag{44}$$

więc dla klasy modeli *PH* warunkowa funkcja niezawodności, przy warunkach eksploatacji $\mathbf{z}$, wyraża się wzorem:

$$S(t|\mathbf{z}) = S_0(t)^{\exp(z_1\beta_1 + \ldots + z_m\beta_m)} \qquad 45)$$

gdzie

$$S_0(t) = \exp\left(-\int_0^t h_0(u)du\right) = \exp(-H_0(t)) \tag{46}$$

jest bazową niezawodnością, a funkcja $H_0(t)$ jest bazową funkcją wiodącą. Chlebowska (1999) wykazała, że funkcje niezawodności z rodziny *PH* są uporządkowane, tj. dla dwóch produktów eksploatowanych w warunkach $\mathbf{z}_1$, $\mathbf{z}_2$ dla wszystkich t albo $S(t|\mathbf{z}_1) \geqslant S((t|\mathbf{z}_2)$, albo odwrotnie. Podstawowym zadaniem, jakie się tu pojawia, to estymacja wektora parametrów β i bazowej funkcji niezawodności na podstawie danych uzyskanych z cenzurowanej obserwacji czasów zdatności n produktów, z uwzględnieniem m różnych warunków eksploatacji $W_1, W_2, \ldots, W_m$. W modelach *PH* zmienne stanowiące warunki eksploatacji mają multiplikatywny wpływ na intensywność uszkodzenia.

Przypuśćmy, że obserwacja czasów zdatności n produktów daje próbę losową złożoną z k różnych czasów $t_{(1)} < \ldots < t_{(k)}$ i $n-k$ czasów cenzurowanych. Zbiór

$$R_i = \{l : t_l \geqslant t_{(i)}\} \tag{47}$$

przedstawia zbiór tych produktów, które były zdatne i obserwacje ich czasów zdatności nie są cenzurowane. Do estymacji wektora parametrów β Cox (1972) zaproponował funkcję:

$$L(\beta_1, \ldots, \beta_m) = \prod_{i=1}^{k} \frac{\exp(z_{i1}\beta_1 + \ldots + z_{im}\beta_m)}{\sum_{l \in R_i} \exp(z_{l1}\beta_1 + \ldots + z_{lm}\beta_m)} \tag{48}$$

gdzie $[z_{i1}, ..., z_{im}]$ jest wektorem warunków eksploatacji produktu, który stracił zdatność w chwili $t_{(i)}$. Funkcja Coxa nie zależy od funkcji $h_0(t)$ i jest wykorzystywana do wyznaczania estymatora $\tilde{\beta}$. Funkcja Coxa nie jest funkcją wiarogodności w zwykłym sensie, ale jest traktowana, jakby nią była. W szczególności przy odpowiednich warunkach maksymalizacja funkcji Coxa prowadzi do estymatora $\tilde{\beta}$ asymptotycznie normalnego z macierzą kowariancji wyestymowaną za pomocą macierzy drugich pochodnych zlogarytmowanej funkcji Coxa.

Z powodu zaokrągleń, grupowania danych lub dyskretnego monitorowania czasów zdatności obserwacje dotyczące zmiennych ciągłych mogą zawierać powtórzenia. Jeśli jest wiele powtórzeń, to parametry modelu można estymować metodami dyskretnymi lub oprzeć się na informacjach pogrupowanych w modelu ciągłym. Jeśli występuje relatywnie mało powtórzeń obserwacji, to do estymacji parametrów jest stosowana tzw. zmodyfikowana funkcja Coxa

$$L(\beta_1, ..., \beta_m) = \prod_{i=1}^{k} \frac{\exp(s_{i1}\beta_1 + ... + s_{im}\beta_m)}{\left(\sum_{l \in R_1} \exp(z_{l1}\beta_1 + ... + z_{lm}\beta_m)\right)^{d_i}} \tag{49}$$

gdzie d_i jest liczbą czasów zdatności równych $t_{(i)}$, natomiast s_{ij} jest sumą warunków eksploatacji z_{ij} dla tych produktów. Jeżeli przez D_i oznaczymy zbiór tych produktów, które utraciły zdatność w chwili $t_{(i)}$, to $d_i = cardD_i$ oraz

$$s_{ij} = \sum_{l \in D_j} z_{il}, \quad i = 1, ..., k, \quad j = 1, ..., m \tag{50}$$

Jeśli nie występują powtórzenia, tj. $d_i = 1$ dla $i = 1, ..., n$, to zmodyfikowana funkcja Coxa (49) sprowadza się do zwykłej funkcji Coxa (48). Pierwsze pochodne zlogarytmowanej zmodyfikowanej funkcji Coxa przyjmują postać:

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^{k} \left(s_{ir} - d_i \sum_{l \in R_i} z_{lj} e^{z_{1l}\beta_1 + ... + z_{lm}\beta_m} \Big/ \sum_{l \in R_i} z_{lj} e^{z_{1l}\beta_1 + ... + z_{lm}\beta_m} \right), \quad r = 1, ..., m \tag{51}$$

Elementy macierzy informacyjnej są postaci:

$$I_{rs}(\beta) = \frac{-\partial^2 \log L}{\partial \beta_r \partial \beta_s} =$$

$$= \sum_{i=1}^{k} d_i \left(\sum_{l \in R_i} z_{lr} z_{ls} e^{z_l\beta} \Big/ \sum_{l \in R_i} e^{z_l\beta} - \left(\sum_{l \in R_i} z_{lr} e^{z_l\beta} \right) \left(\sum_{l \in R_i} z_{ls} e^{z_l\beta} \right) \Big/ \left(\sum_{l \in R_i} e^{z_l\beta} \right)^2 \right) \tag{52}$$

gdzie $r, s = 1, ..., m$. Do rozwiązania układu równań uproszczonej wiarogodności cząstkowej

$$\partial \log L/\partial \beta_r = 0 \quad (r = 1, ..., m) \tag{53}$$

jest wykorzystywany algorytm Breslowa oparty na metodzie Newtona-Raphsona podany w książce L a w l e s s a (1982). Obliczenie wartości oczekiwanej $\mathbf{I}_{rs}(\beta)$ nie jest możliwe bez znajomości mechanizmu cenzurowania obserwacji, a nawet jeżeli mechanizm ten jest znany, ich wyznaczenie może być trudne lub pozostać wręcz niemożliwe. Najprostszym podejściem jest przyjęcie założenia, że wektor ocen parametrów $\tilde{\beta}$ ma rozkład normalny z wektorem wartości oczekiwanych β i macierzą kowariancji $\mathbf{I}(\tilde{\beta})^{-1}$. Na własności asymptotycznej normalności estymatorów największej wiarogodności jest oparta statystyka Walda mająca rozkład chi-kwadrat. Wnioskowanie może być również oparte na metodach ilorazu wiarogodności. Trzecia możliwość wnioskowania oparta jest na wektorze

$$\mathbf{U}(\beta) = \left[\frac{\partial \log L}{\partial \beta_1}, ..., \frac{\partial \log L}{\partial \beta_m}\right]^T \tag{54}$$

który dla dużych prób ma w przybliżeniu rozkład normalny z wektorem wartości oczekiwanych $\mathbf{0}$ i macierzą kowariancji $\mathbf{I}(\beta)$.

Obliczając wielkości zlogarytmowanej funkcji Coxa, jej pochodne i składowe macierzy wygodnie jest uporządkować czasy zdatności i cenzurowane czasy od największego do najmniejszego i zastosować zmienną wskazującą, czy obserwacje są czasami zdatności, czy cenzurowanymi czasami. Obserwację cenzurowaną oznaczamy gwiazdką. Gdy wszystkie czasy są różne, to uporządkowane czasy można zapisać w sposób:

$$\{t^*_{(k),i}\},\ t_{(k)},\ \{t^*_{(k-1),i}\},\ t_{(k-1)},\ ...,\ \{t^*_{(1),i}\},\ t_{(1)},\ \{t^*_{(0),i}\} \tag{55}$$

gdzie $\{t^*_{(j),i}\}$ jest zbiorem cenzurowanych czasów z przedziału $[t_{(j)}, t_{(j-1)})$ i dla wygody przyjmujemy, że $t_{(0)} = 0$ i $t_{(k+1)} = \infty$. Ponieważ zbiór R_i w chwili $t_{(i)}$ zawiera zbiór R_{i+1} w chwili $t_{(i+1)}$ plus te produkty, które utraciły zdatność lub obserwacje czasu zdatności zostały przerwane w $[t_{(i)}, t_{(i+1)})$, więc funkcja Coxa i jej pochodne mogą być obliczone rekurencyjnie poczynając od R_k i kończąc na R_1.

Model *PH* Coxa ze zmiennymi warunkującymi zależnymi od czasu. W wersji podstawowej modelu *PH* Coxa przyjmujemy, że wektor $\mathbf{z}$ określający warunki eksploatacji jest ustalony i znany dla każdego monitorowanego produktu. Uogólnieniem tego modelu jest przyjęcie założenia, że wektor $\mathbf{z}$ jest zależny od czasu. Piszemy wówczas $\mathbf{z} = \mathbf{z}(t)$.

Intensywność uszkodzenia w chwili t przyjmuje teraz postać

$$h(t|z_1(t), \ldots, z_m(t)) = h_0(t)\exp(z_1(t)\beta_1 + \ldots + z_m(t)\beta_m) \tag{56}$$

Zmienna $\mathbf{z}(t)$ może być stochastyczna lub deterministyczna. Model *PH* Coxa ze zmiennymi warunkującymi zależnymi od czasu stosujemy dla danych skategoryzowanych. Model ten pozwala dopasować intensywność uszkodzenia oddzielnie dla każdej grupy przy użyciu analizy warstwowej. W ten sposób z góry dopuszcza się różne funkcje intensywności awarii w każdej z grup. Założenie proporcjonalności może nie być spełnione. Zmienne warunkujące można zdefiniować jako funkcje czasu. Na przykład badając dwie populacje tego samego wyrobu otrzymanego przy zastosowaniu różnych technologii można zastosować zmienną grupującą próbę na dwie grupy wprowadzając kody 1 i 0 wskazujące, którą technologią był produkowany wyrób (patrz *Statistica for Windows* (1997), Vol. III). Szacowaniu poddawany jest wówczas następujący warunkowy model proporcjonalnej intensywności uszkodzenia:

$$h(t|z) = h_0(t) \cdot \exp(\beta_1 z + \beta_2 z \ln t) \tag{57}$$

Warunkowa intensywność uszkodzenia w chwili t jest funkcją zmiennej warunkującej i czasu. Model ten można wykorzystać do testowania założenia proporcjonalności. Jeżeli parametr β_2 jest istotny statystycznie, np. jest co najmniej dwa razy taki, jak jego błąd standardowy, to można powiedzieć, że wpływ zmiennej warunkującej na czas zdatności zależy od czasu i w konsekwencji, założenie proporcjonalności nie jest spełnione.

W celu odzwierciedlenia odpowiedniego przekształcenia zmiennej warunkującej jest stosowana zmodyfikowana wiarogodność cząstkowa. Częściowa funkcja największej wiarogodności dla parametrów β jest teraz postaci:

$$L(\beta_1, \ldots, \beta_m) = \prod_{i=1}^{k} \frac{\exp(z_{i1}(t_{(i)})\beta_1 + \ldots + z_{im}(t_{(i)})\beta_m)}{\sum_{l \in R_i} \exp(z_{l1}(t_{(l)})\beta_1 + \ldots + z_{lm}(t_{(l)})\beta_m)} \tag{58}$$

Wiarogodność cząstkowa dla danego zbioru parametrów jest sumą geometryczną wiarogodności przypadków. Aby obliczyć wiarogodność dla danego układu parametrów dla każdego przypadku i, muszą zostać uwzględnione wszystkie przypadki z czasami przeżycia większymi lub równymi, jak ten dla przypadku i. Zatem modele dopasowania ze zmiennymi warunkującymi zależnymi od czasu mogą wymagać rozbudowanych obliczeń, szczególnie wtedy, gdy zbiór danych zawiera wiele przypadków.

Zmienne warunkujące zależne od czasu są definiowane za pomocą wyrażeń arytmetycznych niekoniecznie zawierających odwołania do czasu. Można określić kilka funkcji dwóch lub więcej zmiennych objaśniających. Może to być wygodną metodą do oceny modeli dla danych zebranych w eksperymencie wieloczynnikowym. Dla każdego czynnika można w zbiorze danych utworzyć zmienną w celu zdefiniowania pożądanych kontrastów. Określając zmienne warunkujące dla modelu regresji *PH*, można wpisać odpowiednie mnożenia definiujące człony interakcyjne. Na przykład załóżmy, że czynniki A i B mają po dwa poziomy. Wszystkim jednostkom przydzielono wartości -1 albo $+1$ odpowiednich zmiennych kodujących w zależności od przypisanego poziomu czynnika. Przy takim kodowaniu zmienne A i B można określić jako zmienne warunkujące, a iloczyn AB jako zmienną towarzyszącą do testowania interakcji między tymi czynnikami w eksperymencie. W niektórych przypadkach można postawić hipotezę, że efekt jednej lub więcej zmiennych warunkujących na intensywność uszkodzenia jest funkcją nieciągłą w czasie. Na przykład możemy określić zmienną warunkującą zależną od czasu jako $z \cdot (T \leqslant t_0)$, gdzie$(T \leqslant t_0)$ jest wyrażeniem logicznym. Parametr przy tej zmiennej warunkującej zależnej od czasu odnosi się do efektu zmiennej z jedynie do czasu t_0, np. czasu gwarancji produktu.

Wnioski. Podane modele regresji pozwalają uzyskiwać estymatory parametrów metodą największej wiarogodności. Model *PH* Coxa oraz model regresji wykładniczej są szacowane za pomocą metody iteracyjnej Newtona–Raphsona. Parametry największej wiarogodności dla modeli regresji normalnej i log-normalnej są szacowane za pomocą algorytmu minimalizacji wartości oczekiwanej. Ogólna istotność modelu regresji jest oceniana za pomocą testu chi-kwadrat obliczanego na podstawie logarytmów wiarogodności odpowiedniego modelu zerowego oraz modelu dopasowanego. Do oceny adekwatności modelu można wykorzystać asymptotyczne błędy standardowe oraz empiryczną macierz kowariancji estymatorów parametrów. Wstępnej oceny adekwatności modelu można dokonać na podstawie różnorodnych wykresów. W przypadku modeli *PH* można wykreślić funkcje niezawodności dla różnych wartości zmiennych niezależnych. Dla modelu regresji wykładniczej można wykreślić reszty względem wykładniczej statystyki pozycyjnej, reszty względem przewidywanych czasów zdatności oraz reszty względem zlogarytmowanych wartości obserwowanych czasów zdatności. Dla modelu normalnego i log-normalnego można wykreślić wykresy rozrzutu wartości obserwowanych względem prognozowanych czasów zdatności, wartości przewidywanych czasów zdatności względem resztkowych czasów zdatności oraz normalny wykres prawdopodobieństwa reszty. W przypadku danych nie cenzurowanych do dopasowania dowolnego rodzaju modelu regresji można stosować metody estymacji nieliniowej. Jeśli czasy zdatności są traktowane jako zmienne binarne, to do badania ciągłych zmiennych objaśniających można zastosować modele regresji probit i logit. Jeśli istnieje możliwość kategoryzacji czasów zdatności do dwóch lub większej liczby przedziałów, to do oceny powiązania różnych kategoryzujących zmiennych niezależnych można wykorzystać ogólny model log-liniowy.

V. PODSUMOWANIE

Przedstawiony przegląd wybranych metod analizy danych dotyczących cenzurowanych czasów zdatności produktów obejmuje tablicę czasu zdatności, metody parametryczne, nieparametryczną metodę estymacji Kaplana–Meiera limitu iloczynowego oraz modele regresji. Zwrócona jest uwaga na możliwości wizualizacji wyników estymacji i komputerowe wspomaganie przedstawionych analiz. Daje to często starym metodom nowe możliwości zastosowań. Podkreślona jest szczególna rola, jaką w zastosowaniach inżynierskich odgrywają rozkłady wykładniczy, liniowy intensywności uszkodzenia, Weibulla i Gompertza. Wskazane są testy porównania czasów zdatności dla danych cenzurowanych.

W związku z ograniczonymi możliwościami stosowania zwykłych technik regresji wielokrotnej, w analizie czasów zdatności z cenzurowanymi obserwacjami, omówione są szczególne modele regresji. Podane modele regresji pozwalają uzyskiwać estymatory parametrów metodą największej wiarogodności.

Metody analizy cenzurowanych czasów zdatności mają szerokie możliwości praktycznych zastosowań, tym bardziej że w obliczeniach możemy skorzystać z pakietów statystycznych, takich jak STATISTICA (1997), STATGRAPHICS (1993) i in. W szczególności przedstawione metody statystyczne można zastosować w kontroli jakości do badania czasu zdatności lub wyznaczaniu okresu gwarancji przez określone produkty.

LITERATURA

Bobrowski D. (1985), *Modele i metody matematyczne teorii niezawodności w przykładach i zadaniach*, WNT, Warszawa.

Chlebowska M. (1999), *Metody statystyczne dla rozkładów ciągłych typu PH*, Politechnika Poznańska (praca magisterska).

Cox D. R. (1972), *Regression Models and Life Tables*, „Journal of the Royal Statistical Society", **34**, 187–202.

Cox D. R., Oakes D. (1984), *Analysis of Survival Data*. Chapman & Hall, New York.

Feigl P., Zelen M. (1965), *Estimation of Exponential Survival Probabilities with Concomitant Information*, „Biometrics", **21**, 826–838.
Kaplan E. L., Meier P. (1958), *Nonparametric Estimation From Incomplete Observations*, „Journal of the American Statistical Association", **53**, 457–481.
Lagakos S. W., Kuhns M. H. (1978), *Maximum Likelihood Estimation for Censored Exponential Survival Data with Covariates*, „Applied Statistics", **27**, 190–197.
Lawless J. F. (1982), *Statistical Models and Methods for Lifetime Data*, Wiley, New York.
Lee E. T. (1980), *Statistical Methods for Survival Data Analysis*, Lifetime Learning, Belmont.
Mantel N. (1967), *Ranking Procedures for Arbitrarily Restricted Observations*, „Biometrics", **23**, 65–78.
Mantel N., Haenszel W. (1959), *Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease*, „Journal of the National Cancer Institute", **22**, 719–748.
Nelson W. (1982), *Applied Life Data Analysis*, Wiley, New York.
Peto R., Peto J. (1972), *Asymptotically Efficient Rank Invariant Procedures*, J. R. Stat. Soc. A, **135**, 185–206.
Schneider H. (1986), *Truncated and Censored Samples From Normal Distribution*, Marcel Dekker, New York.
Statistica for Windows (1997), Vol. III: *Statistics II*, Statsoft, Inc.
Statgraphics. Reference Manual (1993), Manugistics.

Karol Andrzejczak

SELECTED METHODS OF ANALYSIS OF CENSORED LIFETIMES OF GOODS

(Summary)

The study of goods lifetimes, for various reasons, might be time-restricted. In such cases the so called censored observation might appear, for which the exact lifetimes are not known. We only might say that lifetimes of certain goods are longer than the monitoring time. The aim of this work is to describe selected methods of censored lifetimes analysis, comprising:

- lifetime description,
- estimation of survival function, hazard function and probability density,
- fitting distributions to lifetime data,
- comparison of lifetimes for two or more lots of goods,
- regression models.

The described methods were at first developed and applied for medical and biological sciences. Nowadays, international conventions and regulations demand the application of statistical methods e.g. in quality control for the study of lifetime of certain goods. Moreover, statistical methods are frequently used as a tool in social sciences, economics and engineering, and also by managements of various companies, especially insurance companies.

The range of the statistical methods for censored data described in this work is limited to those present in statistical packages. Due to a rapid development of statistical software we limit our methods to those present in the Survival Analysis module of STATISTICA, and such procedures like: Kaplan, Lifetab and Paramod of STATGRAPHICS.