ANALÝZA PŘEŽÍVÁNÍ, COXŮV MODEL

Věra Lánská, IKEM, Praha

Příspěvek je zaměřen na statistickou analýzu dob přežívání. Bude uveden přehled základních statistických metod. Hlavní důraz bude kladen na vyšetřovaní vlivu tak zvaných vysvětlujících proměnných . Tento přístup je velmi často užíván v medicinské statistice, i když oblast využití je jistě daleko širší. Základním předpokladem je existence nezáporné veličiny s odhadnutelným rozptylem, často cenzorovaná zprava.

1.ÚVOD

Předmětem analýzy přežívaní je vyšetřovaní jedné nebo více skupin objektů, kde pro každý jednotlivý objekt máme dánu událost (selhaní) , která nastává po nějakém časovém intervalu (doba do selhání, doba přežití). Příklady doby do selhání zahrnují dobu životnosti součástek, dobu přežívání pacientů v klinických experimentech, dobu trvání stávek, délku doby nezaměstnanosti, dobu potřebnou k vykonání nějakého specifického úkolu a další. Problémy, které se nejčastěji řeší, jsou odhad pravděpodobnostního rozložení doby přežití v dané skupině nebo statistické porovnání dob přežití v několika skupinách. Pokud máme pro každého jedince k dispozici vektor vysvětlujících proměnných, můžeme analyzovat jejich vztah k době přežívání. K jednoznačnému určení doby do selhání potřebujeme znát jednoznačně počátek, měřítko času musí být shodné pro všechny jedince a okamžik selhání musí být také jednoznačně určen. Možnost výskytu neúplných pozorování zabraňuje užití klasických statistických metod. Tato pozorování nám dávají pouze informaci, že k selhání do nějaké doby nedošlo (cenzorovaní). Důležitý předpoklad je, že selhání neovlivňují cenzorování. Existuje několik typů cenzorování, v následujícím výkladu budeme předpokladat náhodné cenzorování. Vedle speciálních metod, které budou uvedeny v příspěvku, existuje mnoho jiných, které lze také využít (kontingenční tabulky, diskriminační analýza, logistická regrese nebo regresní analýza pokud nemáme cenzorovaná pozorování). Programové vybavení je z velké části obsaženo v souborech programů BMDP, SAS nebo GLIM.

2. PARAMETRICKÉ MODELY

Uvažujme nezápornou náhodnou veličinu T, která popisuje dobu do selhaní. Funkcí přežití rozumíme $S(t)=P(T\geq t)$. Budeme předpokládat, že existuje hustota rozložení $f(t) = -S'(t)$. Intenzita selhání (riziková funkce) je definovaná vztahem

$$\lambda(t) = \lim_{h\to 0+} \frac{P(t\leq T<t+h \mid t<T)}{h} = \frac{f(t)}{S(t)} .$$

Mezi funkcí přežití a intenzitou selhání platí

$$S(t) = \exp(-\int_0^t \lambda(s)\,ds) .$$

Intenzita selhání popisuje okamžité riziko selhání a je velmi často používana v analýze přežívání.

Nyní uvedeme některá pravděpodobnostní rozložení vhodná pro doby do selhání:

(i) exponenciální rozložení, $\lambda(t) = \lambda$

(ii) Gamma rozložení, intenzita má komplikované vyjádření s 2 parametry

(iii) Weibullovo rozložení, $\lambda(t) = k\lambda\,(\lambda t)^{k-1}$

(iv) log normální rozložení, intenzita není monotónní a má dva parametry

(v) inversní Gaussovo rozložení, dvouparametrické, není vhodné pro cenzory

Při odhadování neznámých parametrů v parametrických modelech se nejčastěji užívá metoda maximální věrohodnosti. Nechť $f(t;\theta)$ je hustota rozložení doby do selhaní. Vektor neznámých parametrů uvažujeme ve tvaru $\theta' = (\omega', \lambda')$, kde vektor parametrů ω nás zajímá a vektor λ obsahuje rušivé parametry. Jedinec, u kterého dojde k selhání v okamžiku t přispívá k funkci věrohodnosti členem $f(t;\theta)$ a jedinec, který je cenzorován v okamžiku c přispívá členem $S(c;\theta)$. Funkce věrohodnosti pro n nezávislých jedinců s indexem i má tvar

$$L(\theta) = \prod_{\text{necenzor.}} f(t_i;\theta) \prod_{\text{cenzor.}} S(c_i;\theta) .$$

Tedy pro logaritmus věrohodnostní funkce platí vztah

$$l(\theta) = \sum_{\text{necenzor.}} \log f(t_i;\theta) + \sum_{\text{cenzor.}} \log S(c_i;\theta) .$$

Označíme-li $x_i = \min(t_i, c_i)$ a $\Lambda(t;\theta) = \int_0^t \lambda(s;\theta)\,ds$, pak dostaneme

$$l(\theta) = \sum_{\text{necenzor.}} \log \lambda(x_i;\theta) + \sum_{i=1}^{n} \log S(x_i;\theta) = \sum_{\text{necenzor.}} \log \lambda(x_i;\theta) - \sum_{i=1}^{n} \Lambda(x_i;\theta).$$

Tento vztah pro logaritmickou věrohodnostní funkci ukazuje jednu z výhod zavedení intenzity selhání. Odhady neznámých parametrů maximalizují $l(\theta)$. Testovaní hypotéz o neznámých parametrach je založeno na asymptotických vlastnostech maximálně věrohodných odhadů. Uvedeme tři základní statistiky pro testovaní hypotézy $H: \omega = \omega_0$.

(i) věrohodnostní poměr $W(\omega_0) = 2\,[\,l(\hat\omega, \hat\lambda) - l(\omega_0, \hat\lambda_{\omega_0})\,]$

(ii) Waldova statistika $W_1(\omega_0) = (\hat\omega - \omega_0)' V_\omega^{-1}(\hat\omega, \hat\lambda)\,(\hat\omega - \omega_0)$, kde V_ω je submatice inverzní informační matice

(iii) skórová funkce $W_2(\omega_0) = U'(\omega_0)\, V(\omega_0, \hat\lambda_{\omega_0})\, U(\omega_0)$, kde $U(\omega_0) = \frac{\partial}{\partial\omega} l(\omega, \lambda)\big|_{\omega=\omega_0,\ \lambda=\hat\lambda_{\omega_0}}$

Všechny tyto statistiky mají za platnosti hypotézy χ^2 rozložení pravděpodobnosti se stupni volnosti , které se rovnají dim ω_0. (Podrobná odvození lze najít např. v [1], [2], [5], [7].)

3. NEPARAMETRICKÉ METODY

V předchozí části jsme předpokládali, že známe třídu pravděpodobnostních rozložení, kterou se řídí doba do selhání. Nyní uvedeme dva základní neparametrické odhady pravděpodobnosti přežití. Klasická metoda, užívaná hlavně v epidemiologii a v pojišťovací matematice, se nazývá 'life-table'. Sledovaný časový úsek se rozdělí na pevné podintervaly $I_1 = \langle 0,\tau_1)$, ... $I_k = \langle\tau_{k-1}, \tau_k)$ (např. 0-6, 6-12, 12-24, 24 a více). Označíme si

n_i = # živých jedinců na počátku I_i ,

d_i = # selhaní během I_i,

l_i = # ztracených během I_i ,

w_i = # ukončení bez selhání během I_i,

p_i = P(přežití I_i | na počátku I_i je naživu),

$w_i + l_i$ = # cenzorovaných pozorování v intervalu I_i.

Pro pravděpodobnost přežití platí

$$S(\tau_k) = P(T > \tau_k) = \prod_{i=1}^{k} p_i .$$

Metodou 'life-table' odhadneme p_i pomocí $\hat{p}_i = 1 - d_i/n_i'$, kde $n_i' = n_i - 1/2\,(w_i + l_i)$ se nazývá efektivní rozsah a odhad pravděpodobnosti přežití je pak

$$\hat{S}(\tau_k) = \prod_{i=1}^{k} \hat{p}_i .$$

Rozptyl odhadu je dán tzv. Greenwoodovou formulí

$$\mathrm{var}(\hat{S}(\tau_k)) = (\hat{S}(\tau_k))^2 \sum_{i=1}^{k} \frac{d_i}{n'_i(n'_i - d_i)} .$$

Tento odhad se nehodí pro malé výběry, kdy různou volbou intervalů můžeme dostat velmi odlišné výsledky.

Pro přesnější odhad funkce přežití se užívá Kaplan-Meieruv odhad (nazývaný též product-limit). Odpovídá předchozímu, kde za τ_i bereme i-té uspořádané pozorování. Tedy předpokládejme, že pozorujeme dvojice (Y_1, δ_1), ... (Y_n, δ_n), kde δ_i označuje indikátor selhání a platí $Y_{(1)} < \ldots < Y_{(n)}$ (nenastávají shody). Pak odhad funkce přežití má tvar

$$\hat{S}(t) = \prod_{y_{(i)} \le t} \left(1 - \frac{1}{n-i+1}\right)^{\delta_{(i)}} .$$

Pokud $\delta_{(n)} = 0$ je třeba dodefinovat $\hat{S}(t)$ pro $t > y_{(n)}$, nejčastěji se v tomto případě klade $\hat{S}(t) = 0$. Rozptyl odhadu je opět dán Greenwoodovou formulí. Dále lze dokázat, že $\hat{S}(t)$ má asymptoticky normální rozložení. Parametr polohy se nejčastěji odhaduje mediánem $\hat{S}(1/2)^{-1}$ a nebo odhadem střední hodnoty $\int_0^\infty \hat{S}(t)\,dt$.

Ke srovnání dvou a více výběrů existuje mnoho statistik. Pro jednoduchost budeme uvažovat pouze dva výběry a uvedeme dva základní testy. Nechť pro první výběr $T_1, \ldots, T_n$ (doby do selhaní) jsou i.i.d. s distribuční funkcí F_1; $C_1, \ldots, C_n$ jsou i.i.d. s distribuční funkcí G_1 (k nim příslušné okamžiky cenzorování). Pozorujeme dvojice (X_1, δ_1), ... (X_n, δ_n), kde $X_i = T_i \wedge C_i$ a $\delta_i = \chi(T_i \le C_i)$. Pro druhý výběr máme $U_1, \ldots, U_m$ s F_2 a $D_1, \ldots, D_m$ s G_2 a pozorujeme (Y_1, ε_1), ... (Y_m, ε_m). Obvyklá hypotéza je $H_0: F_1 = F_2$.

První test se nazývá Gehanův a je zobecněním Wilcoxonova (resp. Mann-Whitney) neparametrického testu na cenzorovaná pozorování. Testovací statistika má tvar

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}, \qquad \text{kde } U_{ij} = \begin{cases} 1 \text{ pro } t_i > u_j \\ 0 \quad \text{jinak.} \\ -1 \text{ pro } t_i > u_j \end{cases}$$

Druhý, Mantel-Haenzelův, je založen na posloupnosti čtyřpolních kontingenčních tabulek. Výběry sloučíme, uspořádáme a pro každé necenzorované pozorování vytvoříme tabulku pro počty v kategoriích skupina 1,2, selhal neselhal. Testovací statistika se rovná

$$MH = \frac{\sum (a_i - E_0(A_i))}{\sqrt{\sum \mathrm{var}_0 A_i}},$$

kde a_i je počet jedinců v levém horním políčku kontingenční tabulky u i-tého necenzorovaného pozorování.

Oba výše uvedené testy lze zapsat pomocí

$$\sum w_i (a_i - E_0(A_i)),$$

kde $w_i \equiv 1$ pro Mantel-Haenzelův a $w_i = n_i$ pro Gehanův test, kde n_i je počet jedinců v riziku v okamžiku i-tého uspořádaného selhání. Odtud je vidět, že Gehanův test klade větší váhu na počáteční selhání, zatímco MH test dává všem stejnou. Oba testy mají zobecnění pro více skupin. (Podrobnější literatura viz např. [2], [6], [9].)

4. REGRESNÍ MODELY

V následující části uvedeme některé modely, které jsou vhodné pro vyšetrování vlivu vysvětlujících faktorů (nezávisle proměnných) na dobu do selhání. Pro každého jedince máme dán p rozměrný vektor z vysvětlujících faktorů. Složky z reprezentují různé veličiny, jako např. způsob ošetření, vnitřní vlastnosti jedinců (pohlaví, věk) a vnější vlastnosti. Další složky vektoru z mohou být dodatečně vytvořeny; např. složky k vyšetřování interakcí mezi faktory. Vysvětlující faktory mohou být konstantní a nebo proměnné v čase. Často je výhodné definovat vektor z tak, aby $z = (0,\dots,0)$ odpovídal nějaké 'základní' množině podmínek. Modelování pak má dvě úrovně:

(a) model pro dobu do selhání při $z = (0,\dots,0)$;

(b) reprezentace změn při nenulovém z.

4.1 Parametrické regresní modely

Intenzita selhání $\lambda(t, z; \theta)$ je určena až na vektor neznámých parametrů $\theta = (\phi, \beta)$, kde ϕ ovlivňují intenzitu a β se vztahují k vysvětlujícím faktorům. Přesná pravidla, kdy aplikovat plně parametrický model neexistují, ale je vhodné mít na mysli následující doporučení:

(a) pro velmi složité situace je vhodná alespoň částečná parametrizace

(b) pro pouhé srovnání několika skupin s relativně velkým počtem pozorování je většinou nejlepší užít test na porovnání empirických funkcí přežití

(c) máme-li studovat relativně složité závislosti a data jsou cenzorovaná, pak proporcionální Coxův model je výhodný

(d) klademe-li hlavní důraz na vysvětlující faktory, odhad funkce přežití není důležitý

(e) jest třeba provést alespoň nějakou kontrolu správnosti parametrického vyjádření

(f) statistické závěry jsou citlivé na extremní hodnoty jak v prostoru z tak v prostoru dob do selhání

Analýza plně parametrických modelů je opět založena na metodě maximální věrohodnosti. Rovnice maximální věrohodnosti je obvykle řešena iterativní numerickou metodou. Hypotézy o parametrech se testují pomocí statistik uvedených v paragrafu 2. Nyní uvedeme některé nejčastěji užívané modely.

(i) exponenciální rozložení s intenzitou $\lambda(t, z; \theta) = \exp(\beta' z)$

(ii) Weibullovo rozložení s intenzitou $\lambda(t, z; \theta) = k\, t^{k-1} \exp(\beta' z)$

(iii) obecný proporcionální model s intenzitou $\lambda(t, z; \theta) = \lambda_0(t; \phi)\, h(\beta; z)$, kde λ_0 a h jsou známé a nezáporné funkce

(iv) aditivní model s intenzitou $\lambda(t, z; \theta) = \lambda_0(t; \phi) + h(\beta; z)$

(Podrobnější literaturu lze najít např. v [2], [6], [9].)

4.2 COXŮV REGRESNÍ MODEL

Tato část bude věnovaná proporcionálnímu semiparametrickému modelu s intenzitou

$$\lambda(t, z) = \lambda_0(t) \exp(\beta' z),$$

kde $\lambda_0(t)$ je neznámá nezáporná funkce a β je p-rozměrný vektor neznámých regresních koeficientů. Nejprve se budeme zabývat odhadem regresních koeficientů. Odvodíme funkci, kterou Cox nazval 'podmíněná' věrohodnost. S její pomocí získáme odhady regresních koeficientů, které mají stejné asymptotické vlastosti jako odhady maximálně věrohodné a jejichž efficience je poměrně velká. Předpokládejme, že v jednom časovém okamžiku nedojde k více událostem (nenastávají shody). Pro uspořádaná pozorování platí $y_{(1)} < y_{(2)} < \dots < y_{(n)}$, $\delta_{(i)}$ je indikátor selhání a $z_{(i)}$ je vektor vysvětlujících faktorů příslušný k $y_{(i)}$. $\mathcal{R}_{(i)}$ označuje množinu jedinců, kteří jsou v riziku těsně před okamžikem $y_{(i)}$ (tedy množina $\mathcal{R}_{(i)}$ obsahuje jedince, kteří příslušejí k časům $y_{(i)}, y_{(i+1)}, \dots, y_{(n)}$). Pro každé necenzorované pozorování platí

$$P(\text{dojde k selhání v } [\, y_{(i)}, y_{(i+1)}+\Delta y \,) \mid \mathcal{R}_{(i)}) = \sum_{j\in \mathcal{R}_{(i)}} e^{\beta' z_j} \; \lambda_0(y_{(i)})\Delta y$$

$$P(\text{ selže (i)-tý v čase } y_{(i)} \mid \text{jeden jedinec z } \mathcal{R}_{(i)} \text{ selže v čase } y_{(i)}) = \frac{\exp(\beta' z_{(i)})}{\sum_{j\in \mathcal{R}_{(i)}} \exp(\beta' z_j)} .$$

Vezmeme-li součin těchto podmíněných pravděpodobností, dostaneme tzv. 'podmíněnou' věrohodnost

$$L_C(\beta) = \prod_{\text{necenzor.}} \frac{\exp(\beta' z_{(i)})}{\sum_{j\in \mathcal{R}_{(i)}} \exp(\beta' z_j)} .$$

Při odhadování regresních koeficientů zacházíme s $L_C(\beta)$ jako s obvyklou věrohodnostní funkcí. Tedy k nalezení odhadů užijeme skorový vektor a výběrovou informační matici

$$U(\beta) = \frac{\partial}{\partial\beta} \log L_C(\beta) \quad \text{a} \quad i(\beta) = - \frac{\partial^2}{\partial\beta^2} \log L_C(\beta) .$$

Vektor $\hat{\beta}$ je řešením soustavy rovnic $U(\beta) = \underset{\sim}{0}$, která většinou vyžaduje iterativní metodu. Stejně jako pro maximálně věrohodné odhady platí , že $\hat{\beta}$ je asymptoticky normální se střední hodnotou β a rozptylem $i^{-1}(\beta)$. K testování hypotézy H_0: $\beta=\underset{\sim}{0}$ lze užít skórovou statistiku $U(\underset{\sim}{0})' \, i^{-1}(\underset{\sim}{0}) \, U(\underset{\sim}{0})$, která má za platnosti hypotézy χ^2 s p stupni volnosti. Jednotlivé složky skórového vektoru a informační matice mají pro $\beta=\underset{\sim}{0}$ jednodušší tvar

$$\frac{\partial}{\partial\beta_k} \log L_C(\underset{\sim}{0}) = \sum_{\text{necenzor.}} (z_{(i)k} - \bar{z}_{(i)k})$$

$$i_{kl}(\underset{\sim}{0}) = \sum_{\text{necenzor.}} \Big\{ \frac{1}{n_i} \sum_{j\in \mathcal{R}_{(i)}} (z_{jk} - \bar{z}_{(i)k}) \, (z_{jl} - \bar{z}_{(i)l}) \Big\},$$

kde $\bar{z}_{(i)} = n_i^{-1} \sum_{j\in \mathcal{R}_{(i)}} z_j$ a n_i = počet prvků množiny $\mathcal{R}_{(i)}$.

Speciálně pro p=1 a vysvětlující faktor , který označuje příslušnost do skupiny, tedy z_i=0 pro jedince z 1.skupiny a z_i=1 pro jedince z 2.skupiny. Výše uvedená statistika odpovídá Mantel-Haezelově za předpokladu , že nedochází ke shodám.

Dále uvedeme odhad funkce přežití S(t;z). V literatuře existuje mnoho přístupů, zde uvedeme pouze jeden. Funkci přežití vyjádříme ve tvaru

$$S(t;\, z) = \exp(-e^{\beta' z} \int_0^t \lambda_0(s)\, ds) = S_0(t)^{e^{\beta' z}} ,$$

a $S_0(t)$ odhadneme podle Breslowa pomocí

$$\hat{S}_0(t) = \prod_{y_{(i)}\le t} \Big(1 - \frac{\delta_{(i)}}{\sum_{j\in \mathcal{R}_{(i)}} \exp(\beta' z_j)} \Big).$$

Literatura týkající se Coxova modelu je např. [2], [6], [8], [9].

Nyní se budeme zabývat vztahem mezi 'podmíněnou' věrohodností a některými jinými typy věrohodností. Nejprve uvažujme marginální věrohodnost pro pořadí a předpokládejme, že nedochází ke shodám. Pozorujeme dvojice (X_1, δ_1), ... $,(X_n, \delta_n)$, kde k nim příslušná T_i mají distribuční funkci F_i s hustotou f_i . Vektor pořadí $R = (R_1, \ldots ,R_n)$ je vytvořen následujícím způsobem:

R_i = pořadí mezi necenzorovanými pozorováními při δ_i=1,

R_i = pořadí předchozího necenzorovaného při δ_i=0.

Marginální věrohodnost pro pořadí má pak vyjádření

$$p(r, \delta) = \int_{u_1 < \ldots < u_{n_{nec}}} \ldots \int \prod_{i=1}^{n_{nec}} \left\{ f_{u(i)}(u_i) \prod_{j \in C_{i,i+1}} [1 - F_j(u_i)] \right\} du_1 \ldots du_{n_{nec}}.$$

Dosadíme-li

$$F_i(t) = 1 - \exp(-e^{\beta' z_i} \int_0^t \lambda_0(s)\, ds\,),$$

dostaneme

$$L_C(\beta) = p(r, \delta).$$

Dále předpokládejme, že máme posloupnost dvojic náhodných veličin $(X_1, S_1, \ldots X_m, S_m)$. Nechť $y_{u(i)}$ označuje i-té uspořádané necenzorované pozorování, X_i obsahuje informaci o cenzorování v intervalu $[y_{u(i-1)}, y_{u(i)})$ a informaci o tom, že k selhání dojde v $y_{u(i)}$ a S_i obsahuje informaci o tom, že právě jedinec se $z_{u(i)}$ selže v $y_{u(i)}$.
Potom platí, že marginální věrohodnost $S_1, \ldots, S_m$ má tvar

$$p(S_1, \ldots, S_m | \beta) = \prod_{i=1}^{m} p(S_i | S_1, \ldots, S_{i-1}; \beta).$$

Podmíněná věrohodnost $S_1, \ldots, S_m$ dáno $X_1, \ldots, X_m$ má tvar

$$p(S_1, \ldots, S_m \mid X_1, \ldots, X_m; \beta) = \prod_{i=1}^{m} p(S_i | S_1, \ldots, S_{i-1}; X_1, \ldots, X_m; \beta).$$

Plná věrohodnost se rovná

$$p(S_1, \ldots, S_m,\ X_1, \ldots, X_m; \beta) = \prod_{i=1}^{m} p(X_i, S_i \mid X_1, \ldots, X_{i-1}; S_1, \ldots, S_{i-1}; \beta) =$$

$$= \prod_{i=1}^{m} p(X_i \mid X_1, \ldots, X_{i-1}, S_1, \ldots, S_{i-1}; \beta) \prod_{i=1}^{m} p(S_i \mid X_1, \ldots, X_{i-1}, X_i; S_1, \ldots, S_{i-1}; \beta).$$

Cox v [15] dokázal, že platí

$$L_C(\beta) = \prod_{i=1}^{m} p(S_i \mid X_1, \ldots, X_{i-1}, X_i; S_1, \ldots, S_{i-1}; \beta).$$

Proto se L_C také nazývá částečná věrohodnost.

Před použitím Coxova modelu je vhodné ověřit, zda jsou splněny předpoklady proporcionality. Obvykle se používá test grafický, který je vhodné ve sporných případech doplnit testem numerickým. Nyní uvedeme hlavní myšlenku grafického testu. Vektor vysvětlujících faktorů napíšeme ve tvaru $z' = (z_1, \ldots, z_k, z_{k+1})'$, kde prvních k složek splňuje předpoklad proporcionality. Intenzita pro i-tého jedince má tvar

$$\lambda(t, z_{ij}, j=1, \ldots, k+1) = \mu_0(t, z_{i\,k+1}) \exp(\beta' z_i).$$

Za platnosti předpokladu proporcionality platí

$$\mu_0(t, z_{i\,k+1}) = \lambda_0(t) \exp(\beta_{k+1} z_{k+1}).$$

Vysvětlující faktor z_{k+1} se kategorizuje do p úrovní. Potom platí

$$\lambda(t, z_{ij}, j=1, \ldots, k+1) = \lambda_{0s}(t) \exp(\beta' z_i),\ s=1, \ldots, p.$$

Za platnosti hypotézy dostaneme

$$\lambda_{0s+1}(t) = \lambda_{0s}(t)\, e^{\alpha_{s+1}},\ s=1, \ldots, p.$$

Označíme-li

$$\Lambda_s(t) = \int_0^t \lambda_{0s}(u)\,du,\ s=1,\dots,p,$$

pak grafy $\log \hat{\Lambda}_s(t)$ $s=1,\dots,p$ versus t mají za platnosti předpokladu proporcionality konstantní diference $\alpha_2, \dots, \alpha_p$.

Poměrně jednoduchý numerický test lze najít v [12] a nebo v [24].

5. LITERATURA

Knihy.

[1] Barlow, R. E. a Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. New York: Holt,Rinehart and Winston.

[2] Cox, D. R. a Oakes, D. (1985). Analysis of Survival Data. London, New York: Chapman and Hall.

[3] Elandt-Johnson, R. C. a Johnson, N.L. (1980). Survival Models and Data Analysis. New York: Wiley

[4] Gill, R.D. (1980). Censoring and Stochastic Integrals. Mathematical Centre Tracts, no. 124. Amsterdam: Mathematische Centrum.

[5] Gross, A. J. a Clark, V. A. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. New York: Wiley.

[6] Kalbfleisch, J. D. a Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data. New York:Wiley.

[7] Lawless, J. F. (1982). Statistical Models and Methods for Lifetime Data. New York: Wiley.

[8] Mike, V. a Stanley, K.E. (eds) (1982). Statistics in Medical Research. New York: Wiley.

[9] Miller, R. G. (1981). Survival Analysis. New York: Wiley.

[10] Nelson, W. (1972). Applied Life Data Analysis. New York: Wiley.

Články.

[11] Aalen, O. (1978). Nonparametric inference for a family of counting processes. Ann. Statist., 6, 701-726.

[12] Andersen, P. K. (1982). Testing goodness of fit of Cox's regression and life model. Biometrics, 38, 67-77.

[13] Andersen, P. K. a Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. Ann. Statist., 10, 1100-1120.

[14] Cox, D. R. (1972). Regression models and life tables (with discussion). J. R. Statist. Soc., B, 34, 187-220.

[15] Cox, D. R. (1975). Partial likelihood. Biometrika, 62, 269-276.

[16] Cox, D. R. (1979). A note on the graphical analysis of survival data. Biometrika, 66, 188-190.

[17] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. J. Am. Statist. Assoc.,72, 557-565.

[18] Harrington, D. P. a Fleming, T. R. (1982). A class of rank test procedures for censored survival data. Biometrika, 69, 553-566.

[19] Hollander, M. a Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. Biometrics, 35, 393-401.

[20] Kay, R. (1979). Some further asymptotic efficiency calculations for survival data regression models. Biometrika, 66, 91-96.

[21] Lagakos, S. W. (1981). The graphical evaluation of explanatory variables in proportional hazard regression models. Biometrika, 68, 93-98.

[22] Lustbader, E. D. (1980). Time-dependent covariates in survival analysis. Biometrika, 67, 697-698.

[23] Miller, R. G. a Halpern, J. (1982). Regression with censored data. Biometrika, 69, 521-531.

[24] Nagelkerke, N. J. D., Oosting, J. a Hart, A. A. M. (1984). A simple test for goodness of fit of Cox's proportional hazard model. Biometrics, 40, 483-486.

[25] Schmee, J. a Hahn, G. J. (1979). A simple method for regression analysis with censored data. Technometrics, 21, 417-432.

[26] Tsiatis, A. A. (1981). A large sample study of Cox's regression model. Ann. Statist., 9, 93-108.