

PROGRAM PRO ANALÝZU RELATIVNÍHO PŘEŽITÍ

Josef Tvrdlík, Jíří Vaňha, KHS, KÚNZ Ostrava

1. ÚVOD

V příspěvku je popsán programový systém pro analýzu relativního přežití /1,2/. Tento programový systém byl získán z Finnish Cancer Registry ve verzi pro počítač VAX/VMS (zdrojový tvar ve FORTRANU 77) a na našem pracovišti upraven pro počítače SMEP s operačním systémem DOS RV (nebo RSX). Úpravy se týkaly snížení paměťových nároků programů a nahrazení některých příkazů (např. NAMELIST), které Fortran 77 pod DOS RV nezná. Nyní je tento programový systém běžně užíván v analýze dat z krajského registru nádorových onemocnění.

2. RELATIVNÍ PŘEŽITÍ

Při využití běžných metod analýzy přežití /3,4/, které jsou součástí známých statistických programových systémů, např. /5/ se v některých úlohách dostáváme do jistých obtíží:

- výsledky jsou citlivé na to, jak spolehlivě je určena příčina selhání (t.j. většinou úmrtí) tj. jak jsou klasifikováni pacienti;
- pomocí metod /3,4/ může popsat přežití jedné nebo více sledovaných skupin, případně je navzájem porovnat, avšak není k dispozici informace o tom, jak sledovaná příčina úmrtí ovlivňuje délku přežití vzhledem k "normální" populaci.

Tyto nevýhody odstraňuje (alespoň částečně) metoda analýzy relativního přežití. Idea analýzy relativního přežití pochází snad z 50. let. Jejím základem je porovnání délky přežití sledované skupiny s přežitím "normální" skupiny ve stejných podmínkách, t.j. ve stejném období, na stejném místě, stejného pohlaví a věkové struktury. Různé metody zavedení relativního přežití jsou uvedeny v /6/. Zde uvedeme jen základní pojmy potřebné pro užití programů.

Předpokládáme, že doba sledování je rozdělena do w intervalů $\langle i, i+1 \rangle$, kde $i = 0, 1, \dots, w-1$. Pozorovaná kumulativní pravděpodobnost dožití času i je

$${}_iP_0 = \prod_{j=0}^{i-1} p_j \quad (1)$$

kde $p_j = 1 - q_j$, q_j je pozorovaná pravděpodobnost úmrtí v j -tém intervalu, počítaná jako podíl zemřelých v j -tém intervalu, ku počtu pacientů v riziku v j -tém intervalu, t.j. ${}_iP_0$ je veličina užívaná v obvyklých tabulkách přežití /3/.

Relativní přežití je poměr

$${}_iR_0 = {}_iP_0 / {}_iP_0^* \quad (2)$$

kde ${}_iP_0^*$ je očekávaná pravděpodobnost přežití. Existuje však několik způsobů, jak může být tato očekávaná pravděpodobnost definována :

- sprůměrováním přes všechny osoby :

$${}_iP_{01}^* = \sum_{h=1}^l {}_iP_0^*(h) / l_0, \quad (3)$$

kde ${}_iP_0^*(h) = \prod_{j=0}^{i-1} p_j^*(h)$, $p_j^*(h)$ je pravděpodobnost, že osoba z "normální"

populace analogická h-tému pacientu ve věku, pohlaví a obloží přežije j-tý interval sledování; l_0 je počet pacientů na začátku studie. Veličiny $p_j(h)$ jsou roční očekávané pravděpodobnosti, které jsou zjišťovány statistickým úřadem a vydávány jeho populační tabulky přežití. Užití veličiny ${}_1p_{01}^*$ z rov. (3) vyžaduje stejnou délku sledování u všech pacientů. Hakulinen /5/ kromě toho ukázal, že užití ${}_1p_{01}^*$ může vést k vychýleným odhadům ${}_1r_0$.

b) z "průměrného přežití v intervalech

$${}_1p_{02}^* = \prod_{j=0}^{i-1} p_{j2}^* \quad (4)$$

$$\text{kde } p_{j2}^* = \sum_{n=1}^{l_j} p_j^*(h) / l_j .$$

Ve srovnání s (3) může být nyní rozdílná délka doby sledování pacientů, avšak očekávaná kumulativní pravděpodobnost přežití ${}_1p_{02}^*$ je závislá na pozorovaných počtech úmrtí v jednotlivých intervalech.

c) konstrukcí očekávané tabulky přežití pro výběr z "normální" populace s vlastnostmi (věk, pohlaví, atd.) shodnými se sledovanou skupinou - podrobnosti viz Hakulinen /6/. Tato metoda nemá nevýhody předchozích dvou a dává nestranný odhad relativního kumulativního přežití ${}_1r_0$.

Nechť k_j je počet pacientů s potenciální délkou sledování $\geq j$ ($j = 0, 1, \dots, w-1$) a k_{j+1} počet pacientů s délkou sledování $\geq j+1$; $k_j - k_{j+1}$ je pak počet potenciálně ztracených (withdrawals) během intervalu $\langle j, j+1 \rangle$. Očekávaný počet sledovaných žijících pacientů v čase j je lán

$$l_j^* = \sum_{h=0}^{k_j} {}_j p_{0j}^*(h), \quad (5)$$

počet ztracených pacientů z žijících v intervalu $\langle j, j+1 \rangle$ je

$$w_j^* = \sum_{h=k_{j+1} + 1}^{k_j} {}_j p_{0j}^*(h) \cdot \sqrt{{}_j p_{0j}^*(h)} \quad (6)$$

Očekávaný počet pacientů umírajících v intervalu $\langle j, j+1 \rangle$ z pacientů, u kterých by potenciální doba sledování rovněž končila v tomto intervalu

$$d_j^* = \sum_{h=k_{j+1} + 1}^{k_j} {}_j p_{0j}^*(h) [1 - \sqrt{{}_j p_{0j}^*(h)}] \quad (7)$$

Celkový očekávaný počet umírajících v intervalu $\langle j, j+1 \rangle$ je pak

$$d_j^* = \sum_{h=1}^{k_{j-1}} \left\{ {}_j p_{0j}^*(h) [1 - p_j^*(h)] \right\} + d_j^* \quad (8)$$

Očekávanou roční pravděpodobnost přežití je možno počítat buď podle Chiangu - viz /1/ jako

$$p_{j3}^* = \frac{1}{4} (1_j^* - \frac{1}{2} n_j^*)^{-1} \left\{ -\frac{1}{2} d_j^* + \left[\frac{1}{4} (d_j^*)^2 + 4(1_j^* - \frac{1}{2} n_j^*) \cdot \right. \right. \\ \left. \left. \cdot (1_{j+1}^* + \frac{1}{2} w_j^*) \right]^{\frac{1}{2}} \right\}^2 \quad (9)$$

kde $n_j^* = w_j^* + d_j^*$, nebo podle Berksona a Gage - viz /1/

$$p_{j3}^* = 1 - d_j^* / (1_j^* - \frac{1}{2} w_j^*) \quad (10)$$

Očekávaná kumulativní pravděpodobnost přežití (survival rate) se určí obvyklým způsobem

$${}_1P_0^* = \prod_{j=0}^{i-1} p_{j3}^* \quad (11)$$

Přežití sledované skupiny lze kromě uvedených ročních a kumulativních pravděpodobností přežití vztažených k časovým interválům (a odpovídajících relativních přežití) charakterizovat i snížením délky přežití vzhledem k "normální" populaci. Předpokládejme, že po jisté době sledování c je možno považovat pacienty za vyléčené, t. zn., že $p_j = p_j^*$ pro $j \geq c$. Očekávaná doba dožití pro pacienty v čase j ($j \geq c$) je pak

$$e_j^* = \sum_{h=1}^{1_j} e_j^*(h) / 1_j$$

kde $e_j(h)$ je očekávaná délka dožití pro osobu z normální populace, která je téhož pohlaví, věku a žije ve stejném kalendářním období jako h -tý pacient ze sledované skupiny přežívající čas j . Očekávaná délka dožití $e_j^*(h)$ je součástí populačních tabulek přežití vyřádaných centrálními statistickými úřady. Střední délka dožití v čase 0 je

$$e_0 = \frac{1}{2} + \sum_{i=1}^{j-1} {}_iP_0 + jP_0 \left(\frac{1}{2} + e_j^* \right)$$

- detailně viz /7/. Sledovanou skupinu pak můžeme charakterizovat ztrátou v očekávané délce dožití $e_0^* - e_0$ nebo relativní ztrátou $(e_0^* - e_0) / e_0^*$.

3. STRUKTURA PROGRAMOVÉHO SYSTÉMU

Programový systém je tvořen třemi programy :

- a) **TABULATION** - tento program zpracovává vstupní data o pacientech s údaji o věku, pohlaví, délce sledování a stavu pacienta a vytváří z nich výstupní soubor s tabulkami přežití. Počet intervalů tabulek je vstupním parametrem. Pro výpočet relativního přežití využívá soubor s údaji o ročních pravděpodobnostech přežití a střední délkou dožití "normální" populace. Data o pacientech jsou v sekvenčním formátovaném souboru, údaje jsou uloženy jako celá čísla ve znakovém tvaru (fortranská formátová konverze I). Formát vstupních dat a počet proměnných se zadává ve vstupních parametrech. Rovněž je možné zadat pravidla výběru pacientů pro zpracování v případě, že chceme vybrat jen některou skupinu pacientů (např. jen některou diagnózu, věkovou skupinu a pohlaví). Lze zadat max. 5 proměnných podmiňujících výběr. Při jednom průchodu je možné vytvořit až 10 tabulek přežití. Výstupem z programu je soubor

těchto tabulek a stručný protokol o běhu programu (počet vstupních vět, odmítnutých vět apod.).

- b) PRINTING - tisk tabulek přežití vygenerovaných programem TABULATION a popisné statistiky vztahené ke zpracovávané skupině (ztráta v očekávaném dožití atd.). Parametricky lze ovládat výběr tabulek, jež mají být vtištěny a výběr tištěných veličin v tabulkách (různé typy odhadů relativního přežití - viz odst. 2, asymptotické odhady standardních chyb atd.). Součástí tištěného výstupu je i semigrafické zobrazení zadaných veličin charakterizujících průběh přežívání.
- c) GROUPTEST - program počítá testy potřebné pro porovnání dvou a více sledovaných skupin (testované hypotézy jsou uvedeny v odstavci 4). Vstupem programu může být i soubor vytvořený jinými prostředky než programy TABULATION a PRINTING.

V původní finské verzi programu je zalávání vstupních parametrů řešeno formou NAMELISTŮ, takže hodnoty parametrů se zadávají ve tvaru přiřazovacích příkazů : <klíčové slovo> = <hodnota>. Ve verzi pro SNEP byl tento způsob nahrazen vstupem volným formátem. Ten je pro zadávání pro uživatele poněkud nepřehledný (parametry jsou poziční). Proto pro zalávání parametrů uživatelem byl vytvořen interaktivní programový prostředek.

4. TESTOVANÉ HYPOTÉZY

Testováním hypotéz o relativním přežití se podrobně zabývá nezávňý článek /8/. Při porovnávání dvou a více sledovaných skupin (k) je nulovou hypotézu možno vyjádřit

$$H_0 : r_{ki} = r_{li} \quad \text{pro } k = 2, 3, \dots, m \\ i = 0, 1, \dots, w-1$$

t.j., že relativní přežití v odpovídajících časových intervalech je ve všech skupinách stejné.

Tuto hypotézu je možno testovat proti různým alternativám:

$$H_1 : r_{ki} = r_{li}^{c_k} \quad \text{pro } k = 2, 3, \dots, m \\ i = 0, 1, \dots, w-1$$

t.zv. proporcionální rizika, t.j. stejné uspořádání skupin podle rizika ve všech časových intervalech

$$H_1' : [r_{ki} / (1-r_{ki})] / [r_{li} / (1-r_{li})] = R_k$$

t.zv. konstantní odds ratios ve všech skupinách.

$$H_2 : r_{ki} \text{ bez omezení}$$

t.j. všeobecná neshoda skupin.

Program GROUPTEST provádí následující testy :

$$H_0 \text{ vs } H_1' ; H_0 \text{ vs } H_2$$

$$H_0, H_1, H_2 \text{ hierarchicky ;}$$

Všechny užité testové statistiky vycházejí z principu maximální věrohodnosti a asymptoticky mají rozdělení χ^2 . Podrobněji viz /8/. Lze je užít nejen pro testování relativního přežití, ale také pro pozorované pravidelnosti přežití.

Závislost hodnot jednotlivých statistik na přežívání "normální" populace ilustruje následující tabulka. V úloze byly porovnávány dvě skupiny pacientů se stejnou diagnózou. První číslo v každém sloupci je hodnota statistiky, druhé znamená stupně volnosti. Označení sloupců FIN, ČSR, SMK znamená, že při výpočtu byly jako tabulky přežití "normální" populace postupně užity údaje o populaci Finska, ČSR a Severomoravského kraje.

	FIN		ČSR		SMK		POZOROVANÉ	
H_0 vs H_2	28.7	9	29.1	9	29.6	6	46.7	14
H_0 vs H_1	26.0	1	26.6	1	28.4	1	38.8	1
H_1 vs H_2	2.8	8	2.6	8	1.2	5	7.9	13

H_0 vs H_1	26.3	1	26.9	1	26.7	1	39.1	1
H_0 vs H_2	18.1	1	18.8	1	19.8	1	34.8	1

Vidíme, že hodnoty statistik pro relativní přežití jsou vesměs nižší než pro pozorované pravděpodobnosti přežití. Pro tato data výsledek testů nezávisel na tom, která populace byla určena jako referenční. Ve všech testech byla zamítnuta nulová hypotéza a naopak nezamítnuta hypotéza proporcionálního rizika proti všeobecné neshodě.

5. ZÁVĚR

Cílem příspěvku bylo upozornit na jednu dostupnou metodu a programové vybavení pro analýzu přežití. Programy je možno získat z Finnish Cancer Registry, SF-00170 Helsinky 17; verzi pro SMEP na pracovišti autorů tohoto příspěvku po vyžádání souhlasu od finských autorů programu.

Literatura :

1. Hakulinen, T., Abeywickrama, K.H., A computer program for relative survival analysis, Computer Programs in Biomedicine 19, 197-207 (1985)
2. Hakulinen, T., Abeywickrama, K.H., Söderman, B., A computer program package for cancer survival studies, Finnish Cancer Registry, Helsinky, 1987
3. Cutler, S.J., Ederer F., Maximum utilization of the life - table method in analyzing survival, J. Chron. Dis. 8, 699-713 (1958)
4. Kaplan, E.L., Meier P., Nonparametric estimation from incomplete observations, J. Amer. Stat. Assoc. 53, 457-481 (1958)
5. Dixon, W.J. (ed.), BMDP Statistical Software, UCLA Press, 1985
6. Hakulinen, T., Cancer Survival Corrected for Heterogeneity in patient withdrawal, Biometrics 38, 933-942 (1982)
7. Hakama M., Hakulinen, T., Estimating the expectation of life in cancer survival studies with incomplete follow-up information, J.Chron. Dis. 30, 585-595 (1977)
8. Hakulinen, T., Tenkanen L., Abeywickrama K., Päiväranta L., Testing Equality of relative survival patterns based on aggregated data, Biometrics 43, 313-325 (1987)