

StatSoft



Analýza přežití

V tomto článku bychom vám rádi představili jedno z nejstarších odvětví statistiky nazývané analýza přežití (anglicky *survival analysis*).



Jak již název napovídá, nejtypičtější a nejčastější využití najdeme při analýze lékařských dat. Také asi nikoho nepřekvapí, že počátky tohoto typu analýz, datované do 17. století, se týkaly právě analýz lidské úmrtnosti. Neznamena to ale, že bychom se s touto analýzou nemohli setkat i jinde. Na analýzu přežití můžeme narazit i pod názvy jako analýza doby trvání (*duration analysis*) v ekonomii, analýza spolehlivosti (*reliability analysis*) v inženýrských aplikacích nebo analýza historie událostí (*event history analysis*) v sociologii.

Představte si, že máte skupinu pacientů trpících nějakou nemocí a zajímá vás, jak odhadnout, jak dlouho bude pacient s takovou nemocí naživu či jaké je jeho riziko úmrtí. Toto je typická úloha analýzy přežití. Jak už to ve statistice bývá, na to, abychom dokázali předpovědět dobu přežití, či předpovědět riziko pro daného pacienta v daný čas, je potřeba mít nasbíraná data, o která se můžeme opřít. Jaká data můžeme v popsané situaci mít? U každého pacienta budeme mít dobu léčení a údaj, jestli a případně kdy pacient zemřel.

Pokud nás zajímá, jestli událost (úmrtí) u pacienta nastala či nenastala, pak se jedná o úlohu logistické regrese, pokud nás ale zajímá čas, kdy událost nastala, pak se jedná právě o analýzu přežití.

Ale zpět k našim datům. Jak tedy odhadnout dobu přežití. První, co by nás asi napadlo, by bylo použít jenom ty pacienty, u kterých událost nastala a z těchto odhadnout dobu přežití, u těch, u kterých

událost nenastala, přece čas úmrtí nemáme k dispozici a v klasické analýze bychom se k takovýmto pozorováním chovali jako k chybějícím pozorováním. Avšak tato úvaha není správná, ačkoli nemáme u pacientů, kteří ještě žijí, údaj o době úmrtí, víme u nich jistě, že doba bude delší, než kterou nyní sledujeme. Takovouto informaci bychom jistě měli při svých analýzách využít.

Právě popsanému typu dat se říká cenzorovaná data. Máme údaje (výskyt událostí) do dané doby, ale co se děje v době delší, už nevíme, doby do události máme tedy u některých pozorování useknuté nebo jinak řečeno cenzorované.

Důvodů, proč se nám mohou vyskytnout cenzorovaná data, máme mnoho.

- Cenzorování může nastat například z důvodu ukončení studie nebo potřeby provést statistickou analýzu – určitě není možné čekat s analýzou, až nastanou události u všech pozorování, některé časy jsou tedy cenzorované
- Pozorování musíme z nějakého důvodu ze studie vyloučit nebo se sama vyloučila ze studie (pacient se sám rozhodl ukončit léčbu, přestěhoval se, ztratili jsme na něj kontakt,...)
- U daného pozorování nastala jiná událost, která znemožnila další sledování (sledujeme úmrtí na rakovinu a pacient nám zemře při pádu ze schodů,...)

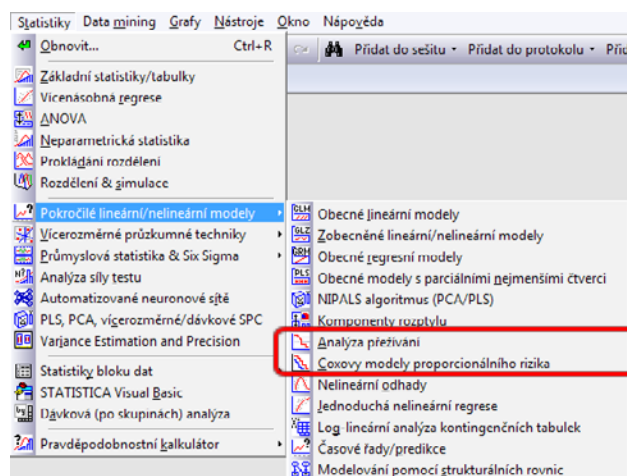
Jak jste sami mohli posoudit, výskyt cenzorovaných dat není vůbec ojedinělou situací, což dokládají i následující příklady dat z nejrůznějších oblastí, u kterých se bez analýzy přežití neobejdeme.

- **Lékařství:** doba do úmrtí, doba do znovuobjevení příznaků nemoci...
- **Finance/obchod:** doba do transakce cenného papíru na burze, doba do splacení úvěru, doba do defaultu, doba do reakce na reklamní nabídku, doba do splacení pohledávky, doba úspěšného setrvání pobočky na trhu
- **Pojišťovnictví:** úmrtnostní tabulky, doba do likvidace pojistné události...
- **Inženýrské aplikace:** čas funkčnosti nějakého zařízení...
- **Sociologie/společenské vědy:** doba nezaměstnanosti, analýza životních drah (doba do sňatku, trvání manželství, doba do narození prvního dítěte), doba nutná k zvládnutí komunikace v anglickém jazyce...

Toto byl samozřejmě pouze úvod na seznámení s problematikou, obor analýza přežití jde dále než jen k odhadu času události či rizika, běžnými postupy jsou například testování shodnosti funkcí přežití pro dvě a více skupin nebo zahrnutí proměnných, které mohou funkci přežití ovlivňovat.

Více o analýze přežití se můžete dozvědět například na našem odborném kurzu [Analýza přežití](#), případně v elektronické učebnici [StatSoft Electronic Statistics Textbook](#).

Na závěr bychom vám ukázali, kde v softwaru STATISTICA můžete metody analýzy přežití najít, abyste si je mohli ihned vyzkoušet.



A přidáme ještě výčet pojmů, které s uvedeným tématem souvisí: funkce přežití, riziková funkce (*hazard function*), Kaplanův-Meierův odhad, úmrtnostní tabulka (*life table*), Nelsonův-Aalenův odhad, logrankový test, Gehanovo zobecnění Wilcoxonova testu, AFT model, Coxův model proporcionálních rizik, Schonfeldova rezidua,...