

Oponentský posudek diplomové práce

Sylva Šmoldasová: Modely analýzy přežití s neproporcionálním rizikem a jejich aplikace

Předložená diplomová práce představuje v prvních čtyřech kapitolách ucelený a přiměřeně detailní úvod do problematiky modelování přežívání včetně Coxova modelu proporcionálních rizik a ověřování předpokladu proporcionality (kapitola 2), modelů používaných při porušení předpokladu proporcionality (kapitola 3) a nástrojů pro výběr a hodnocení těchto modelů (kapitola 4). Závěrečná kapitola je pak ukázkou použití těchto metod při analýze dat. Tato ukáзка obsahuje také četné vstupy a výstupy softwaru R, které jsou rozebrány a okomentovány.

Přestože mám k práci řadu připomínek, považuji ji celkově za zdařilou, doporučuji ji k obhajobě a navrhuji hodnotit ji stupněm B.

Připomínky k teoretické části:

- s. 20: Nesouhlasím s první větou kapitoly 1.2.2.
- s. 20, ř. -1: Nutno opravit interval (též 1.23 na str. 22) a konstatovat závislost na volbě Δt , tj. ve skutečnosti máme $\hat{f}_{\Delta t}(t)$.
- s. 21, ř. 6: Nesouhlasím, že „odhad hustoty je [...] pravděpodobnost ...“.
- s. 21, ř. -10: Není zřejmé, o jakou křivku jde.
- s. 24, vz. (1.31) Opravit argument funkce h .
- s. 26 vz. funkce přežití: Plete se t a t_i , vz. rizikové funkce: plete se t_i a $t_{(i)}$, též v dalších vztazích na stránce.
- s. 26, ř. -2: Bylo by dobré okomentovat vztah pro $S(t)$, to by nám mohlo napovědět, jak vztah zapsat správně. Např.

$$S(t_5) = \frac{S(t_5)}{S(t_4)} \cdot \frac{S(t_4)}{S(t_3)} \frac{S(t_3)}{S(t_2)} \frac{S(t_2)}{S(t_1)} \frac{S(t_1)}{S(t_0)}, \quad \text{kde } S(t_0) = 1.$$

- s. 28, vz (1.36): Má být $i = 1$.
- s. 28, ř. -2: Spíše *odhad* než *výpočet*, též dále.
- s. 28, 2.2.5: Možná by bylo vhodné upozornit, že průměrná doba dožití není narozdíl od obou ostatních charakteristik jediným číslem, ale záleží na čase, ze kterého vycházíme, viz definice (1.42).
- s. 29, vz. 1.41: Jde o odhad?
- s. 29, vz. 1.42: Poslední rovnost by stála za komentář.
- s. 30, ř. 8: Takto zavedený medián není dobře definován (např. v čase 0 nedošlo k události u 50 % subjektů).
- s. 30, ř. -13: A co když poslední není cenzorován, ale více než polovina je cenzorovaných?
- s. 31, ř. -2: $\hat{S}(t) = 0,5$ neplatí pro žádnou hodnotu t .

- s. 34, ř. 3: Není jasné, co se myslí kvalitativní proměnnou (indikátor?).
- s. 36, vz. (2.16): Opět $t_{(i)}$ + je matoucí, že j je index složky, podle které derivujeme a zároveň sčítací index.
- s. 36, ř. -5: Hodnota z závisí na l , což značení opomíjí.
- s. 36, ř. -3 a -10: Jde o interval spolehlivosti pro $\hat{\beta}_j$, nebo β_j ?
- s. 37, ř. -6: Proč „včetně“?
- s. 40, ř. 1: Vyjádření „pokud bychom nebrali v úvahu žádné vysvětlující proměnné“ není přesné. Představme si jediný regresor $x \in \{0, 1\}$. Co pak představuje $h_0(t)$ a jak by vypadala $h(t)$, pokud bychom nebrali v úvahu žádné regresory?
- s. 40, ř. -12: Rovnoběžnost se netýká pevného času t .
- s. 47, vz. (3.6): Index j zbytečný. Může beta záviset na r ? Podobně vz. (3.8).
- s. 51, H_0 : Takto určitě nulovou hypotézu neformulujeme. Předpokládáme, že M_2 je podmodelem modelu M_1 (ujasněte si, co to znamená) a testujeme platnost podmodelu.
- s. 51, vz. (4.3): Vysvětlete přesněji l_{M_1} , l_{M_2} . Viz též připomínka ke str. 57.
- s. 51, ř. -9: Tvrzení o podmnožinách není přesné.
- s. 51: Beta je vektor a má být tedy sázeno tučně ($u_j(\boldsymbol{\beta})$, $i_j(\boldsymbol{\beta})$).
- s. 51, vz. (4.6): $l : \mathbb{R}^k \rightarrow \mathbb{R}^1$, není tedy zřejmé, co je $u_j(0)$ a $i_j(0)$.
- s. 53, ř. 1: Hypotéza se týká teoretických, nikoliv odhadovaných pravděpodobností.
- s. 53, ř. 8: Kterého t se hypotéza týká? (Též na ř. -11, s. 55, ř.-4).
- s. 53, ř. 10: Významnost není záležitostí hypotézy. Navíc nejde o rozdíl mezi léky, ale rozdíl v přežívání mezi skupinami pacientů léčenými těmito léky.
- s. 53, vz. (4.9): E_k jsou odhady (za předpokladu H_0).
- s. 54, vz. (4.11): Bylo by zajímavé vidět, v čem se tato statistika liší od (4.10).
- s. 57, ř. 10 a 15: Logaritmus funkce je funkce, zde jde ale o hodnotu v určitém bodě. Kterém? Též s. 58, vz. (4.15).
- s. 61, ř. 13: Lépe $C \in (0.6, 0.7)$.
- s. 61, ř. 14: Konkordance.
- s. 61, vz. (4.20): Zasloužilo by komentář, proč takto počítané C je odhadem pravděpodobnosti (4.18).
- s. 62, ř. 10: Platí $d = 2C - 1$, viz (4.20).
- s. 62, ř. -10: Popis „ $Y_i(t)$ představuje pozorovanou hodnotu pro i -tý subjekt“ je nedostatečný. Není jasné, o jakou hodnotu se jedná.
- s. 116: Citace článku Grambsch, Therneau (1994) chybně uvádí tři autory.

Připomínky k praktické části:

- s. 65: Mělo by být uvedeno, odkud data pocházejí.
- s. 68, ř. -15: Nerozumím, proč vyloučíte OSevent, ale ponecháte CSSevent a dobu strávenou ve studii.
- s. 70, ř. -11: To je matoucí pro interpretaci výsledků na následujících stranách. Není zřejmé, čeho by věk měl či neměl být prediktorem.
- s. 73, obr. 5.5: Měla by být uvedena jednotka na y -ové ose (měsíce). Podobně u řady dalších grafů (od obr. 5.7 dále).
- s. 77, ř. 1–3: Vůbec není zřejmé, odkud se vzala tato interpretace.
- s. 77, obr. 5.9: Obrázek svádí k dezinterpretaci „chemoterapie zkracuje život“. Poslední měsíce přitom ukazují, že je možná lepší takovéto grafy vůbec nepublikovat, aby k podobným dezinterpretacím nedocházelo. Jsem velmi rád, že je toto v dalším textu rozebráno, viz str. 78–79.
- s. 79: U obrázku pro stupeň III mělo být nastaveno stejné rozmezí hodnot na x -ové ose jako u předchozích dvou obrázků.
- s. 81, ř. 2 textu: Zde by se mělo psát „odhadnutá pravděpodobnost“.
- s. 83, ř. -5: Marně jsem v textu pátral, proč se vykresluje log-log funkce. V odstavci o ověřování proporcionality (s. 39–41) se píše přímo o funkci $S(t)$. Prosím o doplnění úvahy.
- s. 86: Není jasné, proč (zda) byl předpoklad proporcionality ověřován jen pro regresor CHT, o dalších regresorech se píše až na s. 89.
- s. 94, ř. -10: Nemá velký smysl hodnotit významnost u samotných regresorů, které vstupují do interakce.
- s. 95, ř. -8: U interpretace HR je třeba opatrnosti. Riziko úmrtí je vázáno vždy k nějakému časovému intervalu. Zde však porovnáváme spíše hazard rate – míru úmrtnosti, která je konstantně větší po celou dobu.
- s. 96, ř. 6: Slovu „jistě“ bych se vyhnul.
- s. 97, ř. 3: Ovlivňují odhady parametrů, nikoliv odhadované parametry.
- s. 99, ř. 4: Má být 0.8, nikoliv 0.08.
- s. 102, obr. 5.21: V popisu grafu by mělo být vysvětleno označení skupin 1 a 2, aby je čtenář nemusel hledat v textu. Podobně u obr. 5.22 by mělo být popsáno, co je vlevo a co vpravo. Též na dalších obrázcích.
- s. 103, ř. 3: Grayův test mohl/měl být uveden v teoretické části práce.
- s. 109, ř. 2: Cenzorování neznamena, že „nemáme žádnou informaci“.
- s. 114, ř. 10: Ověření předpokladu proporcionality rizik bylo popsáno v kapitole 2.

V Olomouci 8. 6. 2020
Ondřej Vencálek