

Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

M7985 Analýza přežití II

Zadání domácích úloh
1.část

Stanislav Katina, Iveta Selingerová

katina@math.muni.cz, selingerova@math.muni.cz

Příklad 1. Uvažujte arkus sinus odmocninovou škálu funkce přežití, tj. $\arcsin\left(\sqrt{S(t)}\right)$.

- (i) Odvodte $100 \times (1 - \alpha)\%$ intervaly spolehlivosti Waldova typu pro funkci přežití a kumulativní rizikovou funkci.
- (ii) Naprogramujte v \mathbb{R} funkci pro výpočet $100 \times (1 - \alpha)\%$ intervalů spolehlivosti Waldova typu funkce přežití v časech úmrtí.
- (iii) Naprogramujte v \mathbb{R} funkci pro výpočet $100 \times (1 - \alpha)\%$ intervalů spolehlivosti Waldova typu kumulativní rizikové funkce v časech úmrtí.
- (iv) Naprogramujte v \mathbb{R} funkci pro výpočet $100 \times (1 - \alpha)\%$ pásu spolehlivosti funkce přežití mezi nejmenším a největším časem události. Uvažujte Nairův i Hallův–Walnerův pás spolehlivosti.
- (v) Naprogramujte v \mathbb{R} funkci pro výpočet $100 \times (1 - \alpha)\%$ pásu spolehlivosti kumulativní rizikové funkce mezi nejmenším a největším časem události. Uvažujte Nairův i Hallův–Walnerův pás spolehlivosti.
- (vi) Uvažujte data *Diabetes* (soubor `diabetes.txt`), zvlášť muže a ženy. Pro každou skupinu zobrazte odhad funkce přežití spolu s 95% empirickými intervaly spolehlivosti Waldova typu a oběma typy 95% pásů spolehlivosti.
- Interval spolehlivosti zobrazte jako vertikální úsečky ve tvaru písmene „I“.
 - Pásky zobrazte jako schodovité křivky odlišené barevně pro jednotlivé metody.
 - Zvolte vhodný typ odhadu funkce přežití a příslušného rozptylu.
- (vii) Uvažujte data *Diabetes* (soubor `diabetes.txt`), zvlášť muže a ženy. Pro každou skupinu zobrazte odhad kumulativní rizikové funkce spolu s 95% empirickými intervaly spolehlivosti Waldova typu a oběma typy 95% pásů spolehlivosti.
- Interval spolehlivosti zobrazte jako vertikální úsečky ve tvaru písmene „I“.
 - Pásky zobrazte jako schodovité křivky odlišené barevně pro jednotlivé metody.
 - Zvolte vhodný typ odhadu kumulativní rizikové funkce a příslušného rozptylu.
- (viii) Uvažujte data *Diabetes* (soubor `diabetes.txt`), zvlášť muže a ženy. Pro každou skupinu spočítejte odhad pětiletého a osmiletého přežití spolu s 95% empirickými intervaly spolehlivosti Waldova typu. Uveďte také příslušné hodnoty 95% pásů spolehlivosti.
- (ix) Uvažujte data *Diabetes* (soubor `diabetes.txt`), zvlášť muže a ženy. Pro každou skupinu spočítejte odhad pětiletého a osmiletého kumulativního rizika spolu s 95% empirickými intervaly spolehlivosti Waldova typu. Uveďte také příslušné hodnoty hranic 95% pásů spolehlivosti.

Příklad 2. Naprogramujte v \mathbb{R} funkce pro výpočet

- (i) průměrného času přežití,
- (ii) odhad rozptylu průměrného času přežití,
- (iii) $100 \times (1 - \alpha)\%$ intervalu spolehlivosti Waldova typu střední hodnoty času přežití.

Uvažujte data *Diabetes* (soubor `diabetes.txt`), zvlášť muže a ženy. Pro každou skupinu spočítejte průměrný čas přežití spolu s 95% empirickým intervalem spolehlivosti Waldova typu střední hodnoty času přežití. Výsledky porovnejte s hodnotami, které získáte pomocí knihovny `survival`. Zvolte vhodný typ odhadu funkce přežití a příslušného rozptylu.

Příklad 3. Uvažujte data ze souboru `prostate.xlsx`. Popis těchto dat najdete v souboru `popis_prostata.pdf`. Uvažujte dvě skupiny pacientů – léčení a neléčení. Jako léčení pacienti jsou považováni ti, kteří dostávali aspoň 1 mg diethylstilbestrolu denně. Zajímáme se o úmrtí těchto pacientů na karcinom prostaty a na kardiovaskulární onemocnění s ohledem na léčbu.

- (i) Cílem je porovnat přežití léčených a neléčených pacientů pro každou událost (úmrtí na karcinom prostaty, na kardiovaskulární onemocnění a z ostatních příčin). Spočítejte a zobrazte doplněk Kaplanova–Meierova odhadu $(1 - \hat{S}_{KM}(t))$ a odhad kumulativní incidenční funkce $(\widehat{CIF}(t))$ samostatně pro každou událost dle léčebných skupin. Oba přístupy $(1 - \hat{S}_{KM}(t))$ a $\widehat{CIF}(t)$ porovnejte, okomentujte případné rozdíly. Interpretujte výsledky CIF s ohledem na léčbu pacientů a sledovanou událost.
- (ii) Naprogramujte v \mathbb{R} funkci pro výpočet odhadu rozptylu odhadu kumulativní incidenční funkce (CIF) v časech událostí.
- (iii) Naprogramujte v \mathbb{R} funkci pro výpočet $100 \times (1 - \alpha)\%$ intervalů spolehlivosti Waldova typu pro kumulativní incidenční funkci v časech událostí pro různé škály.
- (iv) Zobrazte $\widehat{CIF}(t)$ z (i) včetně intervalů spolehlivosti s použitím vhodné škály.

Příklad 4. Uvažujte data *Diabetes* (soubor *diabetes.txt*), pacienty rozdělte do třech skupin dle věku – 0–60 let, 61–70 let a 71 a více let. Naprogramujte v \mathbb{R} funkce pro testování rovnosti tří a více funkcí přežití.

- (i) Zobrazte do jednoho grafu odhady funkce přežití podle věkové skupiny. Uvažujte jak Kaplanův–Meierův typ odhadu, tak Breslowův typ odhadu.
- (ii) Zobrazte do jednoho grafu odhady kumulativní rizikové funkce podle věkové skupiny. Uvažujte jak Kaplanův–Meierův typ odhadu, tak Nelsonův–Aalenův typ odhadu.
- (iii) Testujte $H_0 : \lambda_1(t) = \lambda_2(t) = \lambda_3(t)$ proti $H_1 : \exists$ aspoň jedno $i < j, \lambda_i(t) \neq \lambda_j(t)$ pomocí zevšeobecněného Wilcoxonova testu, Cox–Mantelova testu, Tarone–Ware testu, Mantel–Haenszelův testu a Peto–Peto–Wilcoxonova testu.
- (iv) Testy, u kterých je to možné, proveďte také pomocí knihovny *survival*.

Příklad 5. Uvažujte data *Diabetes* (soubor `diabetes.txt`) a postupně exponenciální, Weibullovo, log-normální a log-logistické rozdělení.

- (i) Zapište tvar funkce přežití pro každé rozdělení, tento vztah následně upravte tak, abyste získali výraz, který je lineárně závislý na $\ln(t)$, tj. $g(S(t)) = a + b\ln(t)$. Pomocí této závislosti graficky ověřte, zda je možné čas úmrtí po operaci ve skupinách s diabetem a bez modelovat těmito rozděleními. Obě skupiny zobrazte do jednoho grafu. Pomůcka: Na ose x bude $\ln(t)$ a na ose y bude $g(\hat{S}(t))$ dle typu rozdělení, kde použijte vhodný neparametrický odhad funkce přežití. Získané body proložte přímkou pomocí lineární regrese.
- (ii) Pomocí funkce `survreg` odhadněte parametry každého rozdělení pro čas úmrtí po operaci ve skupinách s diabetem a bez. Získané odhady porovnejte s výsledky lineární regrese z (i).
- (iii) Označme $S_0(t)$ funkci přežití pro skupinu pacientů bez diabetu a $S_1(t)$ pro skupinu pacientů s diabetem. Zapište vztah mezi funkcemi přežití $S_0(t)$ a $S_1(t)$ při platnosti modelu zrychleného času s využitím akceleračního parametru θ . Dále opět s využitím akceleračního parametru θ odvoďte vztah mezi kvantily, kde t_{0p} je p -tý kvantil pro skupinu bez diabetu a t_{1p} je p -tý kvantil pro skupinu s diabetem. Pomocí kvantil–kvantilového grafu graficky ověřte vhodnost modelu zrychleného času pro popis vlivu přítomnosti diabetu na čas úmrtí po operaci a odhadněte akcelerační parametr.
- (iv) Odhadněte akcelerační parametr modelu zrychleného času pro popis vlivu přítomnosti diabetu pomocí funkce `survreg` a výsledek porovnejte s výsledky z (iii).

Příklad 6. Uvažujte data *Diabetes* (soubor `diabetes.txt`). Modelujte závislost času úmrtí po operaci na vysvětlujících proměnných pomocí následujících variant Coxova modelu proporcionálních rizik:

1. Jako vysvětlující proměnnou uvažujte pouze věk.
2. Jako vysvětlující proměnnou uvažujte pouze přítomnost diabetu.
3. Jako vysvětlující proměnné uvažujte věk a přítomnost diabetu.
4. Jako vysvětlující proměnné uvažujte věk, přítomnost diabetu a interakci mezi přítomností diabetu a věkem.
5. Uvažujte stratifikovaný Coxův model, kde stratifikační proměnná je přítomnost diabetu (tj. `strata(diab)`) a vysvětlující proměnnou je věk.
6. Uvažujte stratifikovaný Coxův model stejně jako v předchozím bodě, ale navíc uvažujte i interakční člen.
7. Uvažujte Coxovy modely s věkem jako vysvětlující proměnnou samostatně pro skupiny s diabetem a bez.

Pro každou variantu zapište tvar Cozova modelu pomocí rizikové funkce a odhadněte parametry těchto modelů pomocí funkce `coxph` z knihovny `survival`. Spočítejte poměry rizik (hazard ratio) včetně příslušného 95% intervalu spolehlivosti a interpretujte výsledky jednotlivých modelů. Pro každou variantu otestujte významnost jednotlivých parametrů a celého modelu.

Uvažujte všechny modely, ve kterých je zahrnut věk, a spočítejte poměr rizik pro věk včetně příslušného 95% intervalu spolehlivosti samostatně pro skupiny pacientů s diabetem a bez. Výsledky dle jednotlivých modelů porovnejte, interpretujte a okomentujte.

Pomůcka: Varianční matice pro koeficienty získáte z výstupu funkce `coxph` pomocí `$var`. Předpokládejte Waldův typ intervalů spolehlivosti.