

Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

M7985 Analýza přežití

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 desetiletého 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 desetiletého 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.xls`. 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 vhodně zvolenou škálou.