

UNIVERZITA PALACKÉHO V OLMOUCI
PŘÍRODOVEDECKÁ FAKULTA

BAKALÁRSKA PRÁCA

Základy analýzy prežívania



Katedra matematickej analýzy a aplikácií matematiky

Vedúci bakalárskej práce: Mgr. Ondřej Vencálek Ph.D.

Vypracovala: Gabriela Garajová

Študijný program: B1103 Aplikovaná matematika

Študijný odbor: Aplikovaná štatistika

Forma štúdia: prezenčná

Rok odovzdania: 2022

BIBLIOGRAFICKÁ IDENTIFIKÁCIA

Autor: Gabriela Garajová

Názov práce: Základy analýzy prežívania

Typ práce: Bakalárska práca

Pracovisko: Katedra matematickej analýzy a aplikácií matematiky

Vedúci práce: Mgr. Ondřej Vencálek Ph.D.

Rok obhajoby práce: 2022

Abstrakt: Cieľom bakalárskej práce je zoznámiť sa so základnými metódami analýzy prežívania, a vyskúšať ich použitie na praktickom príklade. Na dátach obašujúcich údaje o pacientoch s určitou chorobou sme využitím Coxovho modelu, funkcií prežitia, a testu ich rovnosti našli faktory, ktoré najviac vplývajú na dobu prežitia. Z týchto faktorov je na záver vytvorený optimálny model modelujúci dobu prežitia pacientov.

Kľúčové slová: analýza prežívania, Coxov model, Kaplan-Meirov model

Počet strán: 57

Počet príloh: 1

Jazyk: slovenský

BIBLIOGRAPHICAL IDENTIFICATION

Author: Gabriela Garajová

Title: Basics of survival analysis

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: Mgr. Ondřej Vencálek Ph.D.

The year of presentation: 2022

Abstract: The aim of this bachelor's thesis is to get acquainted with the methods of survival analysis, and apply them in use case. We worked with data containing information about patients suffering from particular disease. By using Cox model, survival functions and equality test, we found out several factors that impact on survival time. These factors were also used to develop an optimal model that describes survival time.

Key words: survival analysis, Cox model, Kaplan-Meier model

Number of pages: 57

Number of appendices: 1

Language: Slovak

Prehlásenie

Prehlasujem, že som bakalársku prácu spracovala samostatne pod vedením pána Mgr. Ondřeje Vencálka Ph.D., Ph.D. a všetky použité zdroje som uviedla v zozname literatúry.

V Olomouci dňa

.....

podpis

Obsah

Úvod	8
1 Analýza prežívania	9
1.1 Cenzurovanie dát	9
1.2 Funkcia prežitia	12
1.3 Funkcia hazardu	13
1.3.1 Vzťah medzi funkciou hazardu a funkciou prežitia	14
1.3.2 Typy funkcie hazardu	15
2 Odhad funkcie prežívania	19
2.1 Tabuľky prežitia	19
2.2 Kaplan-Meierov odhad funkcie prežívania	21
2.2.1 Interval spoľahlivosti	22
2.2.2 Medián prežitia	22
3 Porovnávanie dvoch alebo viacerých funkcií prežitia	24
3.1 Log-rank test	24
3.1.1 Log-rank test pre dve skupiny	25
3.1.2 Log-rank test pre 3 a viac skupín	26
3.2 Alternatívy log-rank testu	27
3.3 Výber testu	29
4 Coxov model	30
4.1 Funkcia prežitia	31
5 Praktické využitie teórie analýzy prežívania	32
5.1 Dátová sada a popis premenných	32
5.2 Popisná štatistika	34
5.3 Modelovanie funkcie prežitia	37
5.3.1 Kaplan-Meierov odhad funkcie prežitia bez rozdelenia na skupiny	38
5.3.2 Odhad funkcie prežitia pomocou Coxovho modelu bez rozdelenia na skupiny	40
5.3.3 Porovnanie funkcií prežitia podľa rôznych vysvetľujúcich pre- menných	41
5.3.4 Optimálny Coxov model	50

Záver	53
Literatura	55
A	

Pod'akovanie

Touto cestou by som sa rada poďakovala môjmu vedúcemu práce Mgr. Ondřeji Vencálkovi Ph.D. za všetkú pomoc, odborné rady, cenné pripomienky a čas, ktorý mi venoval. Vďaka partí i mojej rodine, ktorá ma celú dobu podporovala.

Úvod

Analýza prežívania je odvetvím štatistiky zaoberajúcim sa sledovaním subjektov v čase až do doby výskytu určitej udalosti. Udalosť, ktorej výskyt očakávame, býva vopred stanovená najmä podľa tohto, v akom odvetví života sa pohybujeme, a aký subjekt pozorujeme. Uvedme si niekoľko príkladov.

S analýzou prežívania sa môžeme stretnúť v každodennom živote, kde napríklad očakávame závalu na aute, a teda sledujeme čas do výskytu poruchy od jeho zakúpenia. Rôzne situácie zo života môžeme zahrnúť do jednotlivých oblastí ako sú technické odbory (čas od výroby súčiastky do jej zlyhania), sociológia (doba trvania manželstva do rozvodu, doba do ukončenia pracovného pomeru), ekonómia (doba od získania pôžičky do splatenia, doba do zamestnania sa po škole), či medicína.

Práve v oblasti medicíny a lekárstva je analýza prežívania najčastejšie používaná. V tomto prípade sledujeme čas do výskytu udalosti v skupine pacientov. Túto skupinu pacientov môžeme pozorovať od ich narodenia, alebo od diagnózy choroby do času úmrtia. Avšak koniec pozorovania nie je nutne úmrtie, taktiež to môže byť doba do vyliečenia, hospitalizácie, návratu choroby. Voľba udalosti, ktorú považujeme za koniec pozorovania pacientov závisí od situácie a problému ktorý riešime v daný moment.

V práci si predstavíme základy analýzy prežívania, ako je jej obecná charakteristika a použitie. V prvej časti uvedieme základné pojmy analýzy prežívania, a taktiež metódy pre odhad funkcie prežitia. Ako sme si uviedli, analýza prežívania má využitie v mnohých odvetviach, avšak pre lepšie znázornenie sa budeme v texte zaoberať práve odvetvím medicíny, kde sledujeme spomínanú dobu prežitia pacienta. Následne si v druhej časti práce ukážeme praktické využitie v tejto oblasti a porovnáme jednotlivé metódy používané v analýze prežívania. Cieľom práce bude porovnať jednotlivé odhady funkcie prežitia, a nájsť faktory ovplyvňujúce dobu prežitia pacientov s určitou chorobou.

Kapitola 1

Analýza prežívania

Analýza prežívania je jednou z odvetví štatistiky, ktorá sa zaoberá sledovaním subjektov v čase až do doby výskytu určitej udalosti, prípadne zmenou ich stavu. Subjekt začíname pozorovať v čase kedy nastane vstupná udalosť (tzv. starting point), a subjekt pozorujeme až do doby výskytu nami sledovanej udalosti (tzv. ending point). [8] Začiatok a koniec pozorovania sa môže líšiť od skúmaného odvetvia. Analýza prežívania sa vyskytuje najmä v oblasti medicíny a lekárstva, kedy sledujeme čas od diagnostiky choroby, resp. narodenia až do času úmrtia. Koniec pozorovania nie je nutne úmrtie, taktiež to môže byť doba do vyliečenia, hospitalizácie, návratu choroby a pod.

Cielom analýzy prežívania je najčastejšie odhadnúť funkcie prežitia, porovnať ich, a taktiež odhadnúť vplyv sledovaných vysvetľujúcich premenných na dobu prežitia.

1.1. Cenzurovanie dát

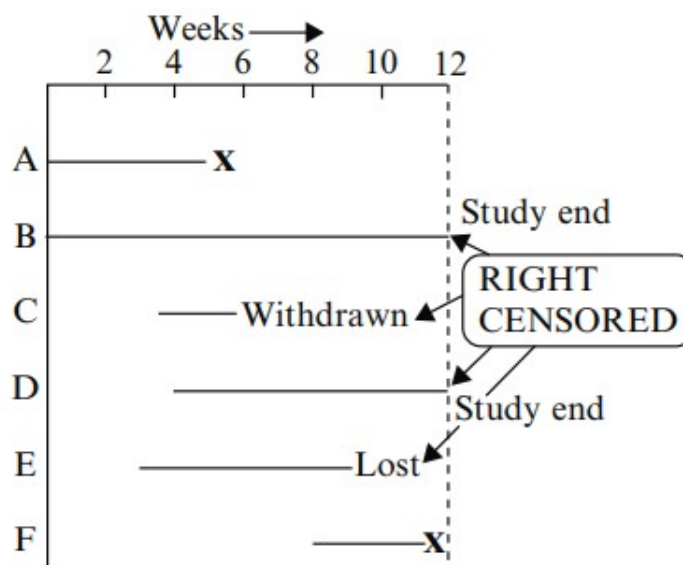
V mnoho štúdiách zaoberajúcich sa analýzou prežívania je často časovo náročné čakať, kým nastane pozorovaná udalosť u každého subjektu, a teda sú ukončené skôr. Cenzurovanie dát sa objavuje v prípade, kedy nemáme kompletnú informáciu o čase prežitia niektorých subjektov.

Doba prežitia je teda cenzurovaná vtedy, keď u subjektu nastane pozorovaná udalosť mimo dobu, kedy štúdium prebieha. [1] Sledovaná udalosť u subjektu môže nastať pred započatím samotnej štúdie. Taktiež keď subjekt určitý čas pozorujeme, avšak až do konca štúdie u neho definovaná udalosť nenastane (môže nastať po ukončení štúdie, alebo nemusí nastať vôbec). V oblasti lekárstva sa cenzurovanie objavuje i v prípade straty kontaktu s pacientom či úmrtia z inej ako sledovanej príčiny. Čas preži-

tia subjektov, u ktorých sledovaná udalosť nenastala, nazývame cenzurovaný.

Pri cenzurovaní dát môžu nastať tri rôzne situácie, a to:

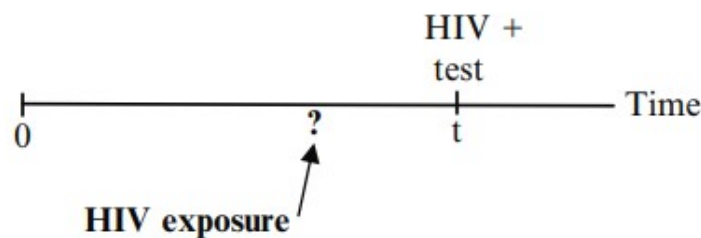
1. Údaj je cenzurovaný sprava (right censoring): Tento typ cenzurovania nastáva v prípade, kedy u subjektu nastane pozorovaná udalosť až po skončení štúdie, prípadne, kedy nemáme kompletnú informáciu o čase prežitia. [3] Cenzurovanie sprava sa v praxi vyskytuje najčastejšie.



Obr. 1.1: Cenzurovanie dát sprava [3]

Princíp cenzúrovania sprava je znázornený v Obr. 1.1. Naľavo je začiatok štúdie a vpravo prerušovanou čiarou koniec štúdie. Pozorované subjekty (pacienti) sú označení písmenami A až F. Vodorovné priamky znázorňujú čas, po aký boli jednotliví pacienti pozorovaní. Dva prípady, kedy je na konci priamky X, znázorňujú, že nastala skúmaná udalosť. V štyroch prípadoch sa vyskytuje cenzurovanie sprava. U pacientov B a D došlo k cenzúre z dôvodu ukončenia štúdie, a stanovená udalosť mohla nastať až po štúdiu. V prípadoch C a E došlo k cenzúre z dôvodu odstúpenia od štúdie a straty kontaktu.

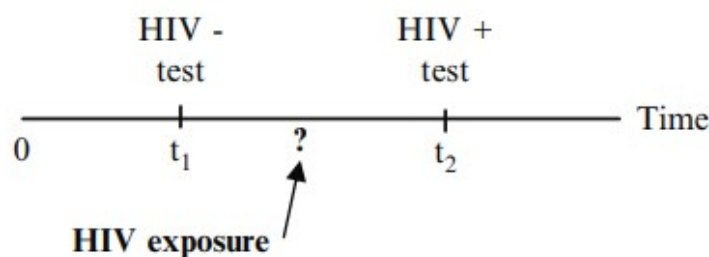
2. Údaj je cenzurovaný zľava (left censoring)- Nastáva v prípade, kedy pozorovaná udalosť nastane pred začiatkom štúdie.



Obr. 1.2: Cenzurovanie dát zľava [3]

Vhodným príkladom môže byť situácia, kedy sledujeme ľudí do času, kým majú pozitívny HIV test (viz Obr.1.2). Prvým pozitívnym testom môžeme spozorovať nastanie udalosti, ktorú je nákaza vírusom HIV. Avšak v tomto prípade nevieme jednoznačne určiť presný čas, kedy došlo k vystaveniu vírusu a teda nevieme presne určiť, kedy sa udalosť vyskytla.

3. Údaj je cenzurovaný intervalovo (interval censoring)-Vieme, že pozorovaná udalosť sa nachádza medzi dvomi rôznymi hodnotami, t.j. v intervale. Napríklad, že sledovaná osoba zomrela niekedy medzi dvomi kontrolnými vyšetreniami.



Obr. 1.3: Intervalové cenzurovanie dát. [3]

V tomto prípade, človek podstúpil dva HIV testy, pričom prvý test bol negatívny a druhý test už pozitívny. Nevieme ale presne určiť, kedy došlo k nastaniu udalosti, teda nakazenie HIV vírusom. Vieme iba, že nákaza nastala v intervale medzi časom t_1 a t_2 (viz Obr.1.3)

Predošlé prípady sú konkrétnymi realizáciami troch typov cenzúrovania.

Prvým je cenzurovanie 1. typu, kde má štúdia dopredu stanovený čas trvania, a subjekty sú pozorované rovnakú dobu. V prípade, že sa u jedinca udalosť nevyskytne, je

údaj cenzurovaný sprava.

Ďalej sa uvádza cenzurovanie 2. typu, kde štúdia končí n -tým výskytom stanovenej udalosti.

V treťom prípade ide o náhodné cenzurovanie, kedy cenzurovaný čas subjektu je nezávislý na dobe udalosti. Práve v tomto type sa vyskytuje cenzurovanie dát v prípade straty kontaktu s pacientom, odmietnutie pacienta pokračovať v štúdiu či ukončenie štúdie. [7]

Sledovanú veličinu (dobu do výskytu udalosti) značíme písmenom Y a pozorujeme n subjektov. C_i , $i=1, \dots, n$ značíme veličinu, ktorá môže nadobúdať len hodnoty 0 a 1. V prípade, že udalosť nastala, sa C_i rovná 0. Naopak, v prípade že pozorovaná udalosť Y behom štúdie nenastala a údaj bol cenzurovaný je C_i rovné 1. [1]

1.2. Funkcia prežitia

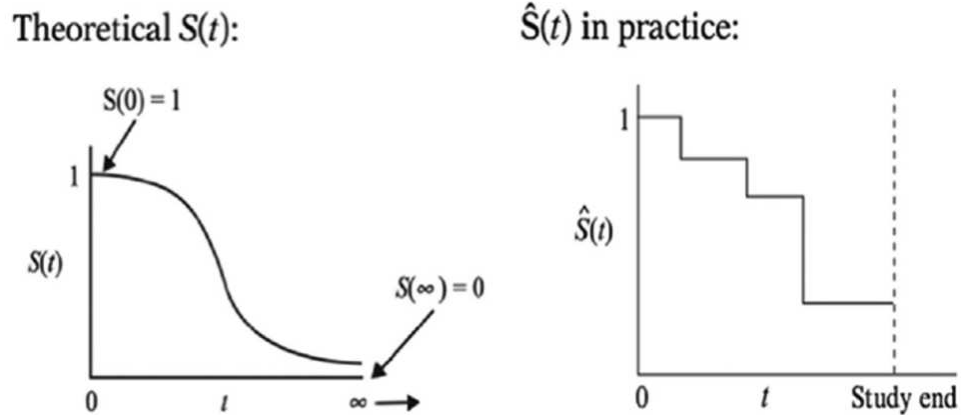
Čas, označovaný ako doba prežitia od začiatku pozorovania až do výskytu stanovenej udalosti, prípadne cenzúry, značíme písmenom T . T je náhodná veličina, ktorá môže nadobúdať hodnoty väčšie, alebo rovné nule. Konkrétna realizácia veličiny T je t .

Odhad funkcie prežitia je jedným z hlavných cieľov analýzy prežívania. Funkcia prežitia (survival function, survivor function) udáva pravdepodobnosť, že daný subjekt žije dlhšie ako konkrétny čas t . Značíme ju $S(t)$ a je definovaná vzťahom $S(t)=P(T>t)$, kde T je náhodná veličina udávajúca čas do nastatia udalosti, a t je daný čas. [1] Teoreticky je funkcia prežitia zobrazená ako hladká krivka, nazývajúca sa krivka prežitia, ktorá má nasledujúce vlastnosti:

- Krivka je nerastúca.
- V čase $t=0$, kedy boli subjekty zaradené do štúdie, platí $S(t)=S(0)=1$. To znamená, že na začiatku štúdie u nikoho nenastala pozorovaná udalosť a teda pravdepodobnosť prežitia v čase $t=0$ je rovná 1.
- V čase $t=\infty$, $S(t)=S(\infty)=0$. A teda, teoreticky, ak by sme mali štúdiu, ktorá nie je časovo obmedzená, na konci by pozorovaná udalosť nastala u všetkých

subjektov. To znamená, že pravdepodobnosť prežitia klesne až k 0.

Všetky vyššie uvedené vlastnosti sa týkajú teoretickej funkcie prežitia, v praxi sa ale vyskytuje schodovitá funkcia prežitia. Je to z toho dôvodu, že funkcia prežitia sa mení iba v čase, kedy nastane sledovaná udalosť. Pre lepšiu názornosť je rozdiel vo funkciách prežitia uvedený v obrázku.



Obr. 1.4: Teoretická funkcia prežitia, a funkcia vyskytujúca sa v praxi [3]

Hlavný rozdiel vo vlastnostiach medzi uvedenými funkciami je, že čas štúdie sa v praxi nevyskytuje nekonečný. A teda môže dôjsť k prípadu, že v čase ukončenia štúdie nastala u všetkých subjektov pozorovaná udalosť (viz Obr.1.4). Z toho dôvodu odhadovaná funkcia prežitia nemusí byť na konci štúdie nutne rovná nule. [3]

1.3. Funkcia hazardu

Funkcia hazardu hovorí o aktuálnom riziku výskytu stanovenej udalosti v časovom okamihu t , avšak s podmienkou, že sa u daného subjektu udalosť ešte nevyskytla. Na rozdiel od funkcie prežitia, funkcia hazardu sa zameriava na zlyhanie/ výskyt udalosti, teda napríklad výskyt úmrtia. Pričom funkcia prežitia sa zameriava na nezlyhanie/ udalosť sa ešte nevyskytla, napríklad prežitie. [3] Funkciu hazardu značíme $h(t)$ a je definovaná nasledujúcim vzťahom.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (1.1)$$

Môžeme teda vidieť, že hodnota funkcie v čase t hovorí o pravdepodobnosti výskytu udalosti v intervale $(t, t + \Delta t)$, za podmienky že sa u subjektu udalosť do času t nevyskytla. [5] Celkové riziko výskytu pozorovanej udalosti od začiatku pozorovania subjektu až do času t nazývame kumulatívna riziková funkcia a je definovaná ako:

$$H(t) = \int_0^t \lambda(x) dx \quad (1.2)$$

Hlavné charakteristiky funkcie hazardu sú:

- funkcia je vždy nezáporná, $h(t) \geq 0 \forall t$,
- funkcia nieje obmedzená zhora,

1.3.1. Vzťah medzi funkciou hazardu a funkciou prežitia

Medzi funkciou hazardu $h(t)$ a funkciou prežitia $S(t)$ existuje vzájomný vzťah. Ak poznáme funkciu prežitia, vieme ňou vyjadriť funkciu hazardu. To isté platí i v opačnom prípade. Môžeme si všimnúť, že vzorci pre výpočet rizikovej funkcie (1.1) sa vykytuje podmienená pravdepodobnosť. Keď použijeme vzťah vyplývajúci z definície o podmienenej pravdepodobnosti¹, a dosadíme ho do čitateľa funkcie hazardu dostaneme nasledujúci výraz:

$$P(t < T \leq t + \Delta t | T > t) = \frac{P((t < T \leq t + \Delta t) \cap (T > t))}{P(T > t)} = \frac{P(t < T \leq t + \Delta t)}{P(T > t)} \quad (1.3)$$

Funkcia prežitia bola definovaná v kapitole 1.2, avšak túto definíciu si môžeme ďalej rozvinúť do tvaru $S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$.

Kde $F(t) = P(T \leq t)$ je distribučná funkcia, ktorá vyjadruje pravdepodobnosť, že hodnota náhodnej veličiny T neprekročí konkrétny čas t . Teda čas prežitia bude menší alebo rovný t .

Hustota pravdepodobnosti vyjadruje pravdepodobnosť výskytu stanovenej udalosti v čase t a získame ju vzťahom $f(t) = \frac{dF(t)}{dt} = \frac{d(1-S(t))}{dt} = -\frac{dS(t)}{dt} = -dS'(t)$.

Alebo je taktiež definovaná ako

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t} \cdot [2] \quad (1.4)$$

¹ $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Ďalej s využitím distribučnej funkcie $F(t)$ a funkcie prežitia $S(t)$, je možné výraz (1.3) rozšíriť do tvaru

$$\frac{P(t < T \leq t + \Delta t)}{P(T > t)} = \frac{F(t + \Delta t) - F(t)}{S(t)} \quad (1.5)$$

Tento vzťah následne dosadíme do definície funkcie hazardu

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\frac{F(t+\Delta t)-F(t)}{S(t)}}{\Delta t} = \frac{P(t < T \leq t + \Delta t)}{S(t) \Delta t} = \frac{f(t)}{S(t)}. \quad (1.6)$$

Z derivácie zloženej funkcie a taktiež pravidla pre deriváciu logaritmu môžeme predchádzajúci vzťah napísať v tvare

$$h(t) = -\frac{d(\ln(S(t)))}{dt}. \quad (1.7)$$

Tento výsledok následne dosadíme do vzťahu (2) definujúceho kumulatívnu funkciu hazardu. Po úprave následne dostaneme vzťah

$$H(t) = -\ln(S(t)) \quad (1.8)$$

a vidíme, že dostávame vzorec ktorý vyjadruje súvislosť medzi kumulatívnu funkciou hazardu a funkciou prežitia. Vzťah ešte môžeme ďalej upraviť do tvaru

$$S(t) = e^{-\int_0^t h(x) dx}. \quad (1.9)$$

[2]

1.3.2. Typy funkcie hazardu

V tejto časti si ukážeme tri názorné príklady, ako môže funkcia hazardu vyzerieť. Jednotlivé prípady sú pritom dané rozdelením veličiny T . Bežne sa používajú exponenciálne, Weibullovo a logaritnicko-normálne rozdelenie.

- Exponenciálne rozdelenie

Tento typ rozdelenia je charakterizovaný konštantným tempom funkcie hazardu viz Obr. 1.5. V prípade veľkej hodnoty λ je riziko výskytu udalosti vysoké, a doba

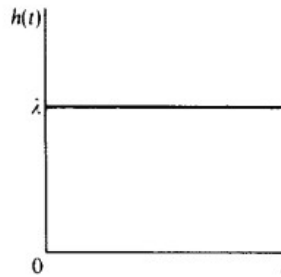
prežitia nízka. V opačnom prípade, kedy je λ nízke číslo, je riziko výskytu udalosti nízke, a o niečo dlhšia doba prežitia. [5]

Pre prípad kedy $t \geq 0$, a $\lambda > 0$ platí:

Funkcia prežitia je v tvare: $S(t) = e^{-\lambda t}$

Hustota pravdepodobnosti: $f(t) = \lambda e^{-\lambda t}$

V predchádzajúcej kapitole bol odvodený vzťah pre funkciu hazardu (rovnica 6), ktorý keď použijeme v tomto, a všetkých nasledujúcich prípadoch dostávame funkciu hazardu v tvare: $h(t) = \lambda$



Obr. 1.5: Funkcia hazardu pre exponenciálne rozdelenie [5]

Prípad konštantného tempa funkcie hazardu je napríklad pozorovanie zdravej osoby, ktorej sa zdravotný stav nemení.

- Weibullovo rozdelenie

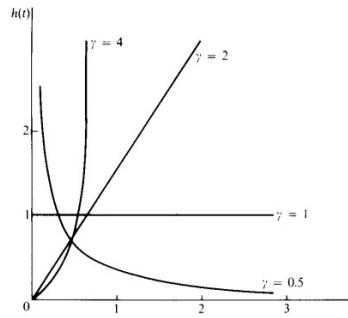
Je zobecnenie exponenciálneho rozdelenie, avšak nepoužíva konštantnú funkciu hazardu. Funkcia závisí na dvoch parametroch, a to γ , a λ , kde λ je parameter škálovania (v Obr. 1.6 je nastavený na $\lambda = 1$).

Pre $\gamma, \lambda > 0$, a $t > 0$ platí:

Funkcia hazardu: $h(t) = \lambda \gamma (\lambda t)^{\gamma-1}$

Funkcia prežitia: $S(t) = e^{-(\lambda t)^\gamma}$

Hustota pravdepodobnosti: $f(t) = \lambda \gamma (\lambda t)^{\gamma-1} e^{-(\lambda t)^\gamma}$



Obr. 1.6: Funkcia hazardu pre Weibullovo rozdelenie [5]

Parameter γ udáva tvar krivky. V prípade kedy $\gamma > 1$, funkcia hazardu rastie, a naopak, kedy je $\gamma < 1$, funkcia hazardu klesá, viz Obr. 1.6 Pre prípad, kedy $\gamma = 1$, dostávame spomínané exponenciálne rozdelenie.

Príkladom kedy $\gamma > 1$, môžu byť pacienti s rakovinou, ktorým nezaberá liečba a s dlhšou dobou rastie riziko úmrtia. V prípade $\gamma < 1$ si môžeme predstaviť pacientov, ktorí podstúpili operáciu. Čím je dlhšia doba od operácie, tým sa znižuje riziko úmrtia.[5]

- Logaritmicko-normálne rozdelenie

Najjednoduchší prípad logaritmicko normálneho rozdelenia je definovaný ako rozdelenie premennej, ktorej logaritmus sa riadi normálnym rozdelením. [5] V tomto prípade je funkcia hazardu najskôr rastúca do doby, kým dosiahne maximum, a potom klesajúca (viz Obr. 1.7).

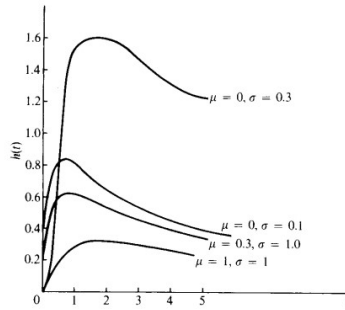
Tento prípad môže nastať pre pacientov s tuberkulózou, kedy na začiatku choroby riziko rastie a potom postupne klesá.

Pre $\sigma > 0$ a $t \geq 0$ platí:

Funkcia hazardu: $h(t) = \frac{f(t)}{S(t)}$. [9]

Funkcia prežitia: $S(t) = 1 - \Phi\left(\frac{\ln(t)}{\sigma}\right)$, kde Φ značí hodnotu kumulatívnej distribučnej funkcie normálneho rozdelenia.

Hustota pravdepodobnosti: $f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\ln t - \mu)^2\right]$. [10]



Obr. 1.7: Funkcia hazardu pre Logaritmicke-normálne rozdelenie [5]

Rozdelenie závisí na dvoch parametroch, a to μ a σ , ktoré sú škálovacie parametre. Taktiež je kladne zošíkmené, a čím väčšia je hodnota σ , tým väčšie zošíkmenie nastane.

Kapitola 2

Odhad funkcie prežívania

V analýze prežívania je častou úlohou čo najlepšie odhadnúť dobu prežitia - dobu do výskytu určenej udalosti. Následne potom skúmame rôzne faktory, ktoré môžu mať výrazný vplyv na dĺžku tejto doby. Existuje niekoľko spôsobov ako odhadnúť funkciu prežiti. V tejto kapitole sa im budeme venovať.

2.1. Tabuľky prežitia

Táto metóda je označovaná za najstarší a najjednoduchší prístup ako odhadnúť funkciu prežitia. Je taktiež nazývaná aj ako aktuárska metóda [7]. Jej postup je založený na rozdelení doby prežitia do intervalov, ktorých počet (budeme označovať J) je stanovený dopredu. Tieto intervaly sú najčastejšie rovnakej dĺžky, avšak nie nutne musí byť dĺžka všetkých intervalov rovnaká. Vo väčšine prípadov je veľkosť intervalu jeden rok, no pri dlhších štúdiách sa vyskytujú i viacročné intervaly, napríklad 5 či 10 ročné.

Počet, koľkokrát nastala pozorovaná udalosť v j -tom intervale, značíme d_j , kde $j=1, \dots, J$. Ďalej R_j je počet subjektov v rizikovej skupine na začiatku j -tého intervalu. A c_j je počet subjektov, ktorí boli v j -tom interavale cenzúrovaný. Odhad funkcie prežitia je definovaný nasledovne:

$$\hat{S}(t_j) = \prod_{i=1}^j \hat{p}_i, \quad j = 1, \dots, J \quad (2.1)$$

kde,

$$\hat{p}_j = 1 - \frac{d_j}{R_j - \frac{c_j}{2}} \cdot [2] \quad (2.2)$$

Z definícií môžeme vidieť, že krivku prežitia odhadujeme ako súčin odhadnutých pravdepodobností prežitia v jednotlivých intervaloch $j=1, \dots, J$. Pre cenzorované subjekty však predpokladáme, že v j -tom intervale ostávajú v priemere polovicu doby intervalu.

Pre názornosť, si môžeme použitie vysvetliť na jednoduchom prípade. Na začiatku máme 100 pacientov, ktorí patria do rizikovej skupiny (teda $R_1 = 100$). Ďalej budeme predpokladať, že v priebehu celej štúdie nedošlo k cenzúre pacientov ($c_j = 0, \forall j$), a vieme, že počas prvého intervalu zomrelo 15 pacientov ($d_1 = 15$). Potom pravdepodobnosť prežitia prvého intervalu (I_1), za podmienky že pacient žil na začiatku intervalu je podľa vzorca (2.2) $\hat{p}_1 = 1 - \frac{15}{100} = 0,85$. Teda je 85% pravdepodobnosť, že pacient prežije prvý interval. Následne je v rizikovej skupine na začiatku druhého intervalu 85 pacientov ($R_2 = 85$), a počas tohto intervalu zomrelo 5 pacientov. Potom pravdepodobnosť že pacient prežije druhý interval (I_2) je $\hat{p}_2 = 1 - \frac{5}{85} = 0,94$. Tieto pravdepodobnosti sa však týkajú samostatných intervalov.

Pokiaľ však chceme zistiť pravdepodobnosť, že pacient prežije druhý interval od začiatku štúdie, vyjdeme z podmienenej pravdepodobnosti, kde do podmienky zahrnieme že pacient prežil prvý interval, a násobíme ju pravdepodobnosťou podmienky. Z vety o podmienenej pravdepodobnosti teda platí $P(I_1 \cap I_2) = P(I_2|I_1) \cdot P(I_1) = \hat{p}_2 \hat{p}_1$. Nakoniec v krátkom príklade bude $\hat{S}(t_2) = \hat{p}_2 \hat{p}_1 = 0,94 * 0,85 = 0,799$. Teda je 80% pravdepodobnosť, že sa pacient od začiatku štúdie dožije do konca druhého intervalu. Túto úvahu je však možné rozšíriť na ľubovoľný počet intervalov, pokiaľ máme k dispozícii pacientov v riziku.

Typy tabuliek prežitia

- Kohortové tabuľky prežitia (*cohort life table*). Tento typ tabuliek znázorňuje pravdepodobnosť úmrtia subjektov z danej kohorty¹ počas ich života. Na začiatku intervalu je počet subjektov v rizikovej skupine, ktorí prežili (neboli cenzúrovaný, ani nezomreli) v predchádzajúcom intervale.
- Aktuálne tabuľky prežitia (*current life table*). V tomto prípade tabuľky znázorňujú súčasnú pravdepodobnosť úmrtia pre rôzne vekové skupiny v danom roku. Na začiatku j -tého intervalu je uvažovaná skupina ľudí v určitom veku, ktorá partí

¹Kohorta je skupina ľudí, ktorí sú pozorovaní počas celej štúdie

do rizikovej skupiny. Uvažovaná skupina subjektov je úplne odlišná od rizikovej skupiny v predchádzajúcom intervale. [7]

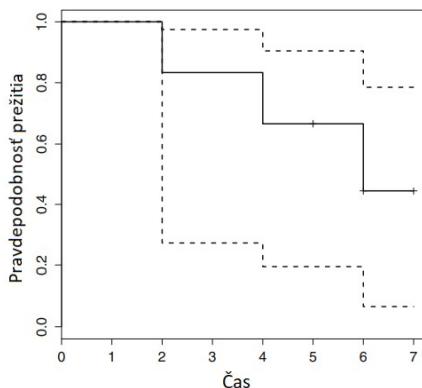
2.2. Kaplan-Meierov odhad funkcie prežívania

Kaplan-Meierov odhad je najznámejšia a najpoužívanejšia metóda pre odhad funkcie prežitia. Patrí k neparametrickým metódam, pretože nepredpokladá určitý známy typ rozdelenia veličiny T , a pri odhade funkcie prežitia nestačí iba odhadnúť parametre daného typu rozdelenia. Odhad hodnôt funkcie prežitia sa počíta vo všetkých okamihoch, kedy došlo k pozorovanej udalosti.

Časy, kedy došlo k sledovanej udalosti, označíme ako $t_1 < t_2 < \dots < t_k$, a sú zoradené vzostupne. Ďalej $n_i, i=1, \dots, k$ je počet subjektov v riziku v čase t_i , čo znamená, že subjekty patria do skupiny, u ktorých sa zatiaľ nevykystla žiadna udalosť. A počet subjektov, u ktorých nastala pozorovaná udalosť v čase t_i , značíme d_i . [6] Odhad funkcie prežitia je v tvare:

$$\hat{S}(t) = \prod_{i:t_i < t} \left(1 - \frac{d_i}{n_i}\right) \quad (2.3)$$

Kde $1 - \frac{d_i}{n_i}$ značí odhad pravdepodobnosti prežitia subjektu času t_i (v tomto čase u neho nenastane sledovaná udalosť), za podmienky, že sa dožije tohto času t_i . [2] Podmienka, že sa do času t_i dožije, znamená, že v časoch t_1 až t_{i-1} u neho pozorovaná udalosť nenastala, a taktiež ani v intervale (t_{i-1}, t_i) .



Obr. 2.1: Kaplan-Meierov odhad funkcie pežitia s 95% intervalom spoľahlivosti. [6]

Ako môžeme vidieť v Obr. 2.1, tvar funkcie je schodovitý, kde k poklesu odhadu

dochádza vždy v okamihu, kedy nastane pozorovaná udalosť, čo zníži odhad pravdepodobnosti prežitia. Každá zvislá čiarka značí cenzúrovaný údaj. Z čoho vyplýva, že subjekt ktorý bol cenzúrovaný, nepatrí do skupiny s výskytom udalosti (nastal by pokles funkcie), ale do skupiny subjektov v riziku.

2.2.1. Interval spoľahlivosti

V Obr. 2.1 je prerušovanou čiarou zvýraznený 95% interval spoľahlivosti pre odhad funkcie prežitia. Avšak pre funkciu prežitia je možné konštruovať ľubovoľný $100(1-\alpha)\%$ interval spoľahlivosti, ktorý má tvar

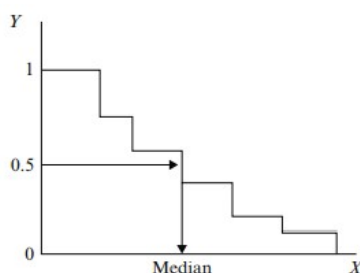
$$[\hat{S}(t) \pm z_{1-\alpha/2} \sqrt{\text{var}(\hat{S}(t))}] \quad (2.4)$$

kde $z_{1-\alpha/2}$ značí $100(1-\alpha/2)\%$ kvantil štandardizovaného normálneho rozdelenia. Ako môžeme vidieť, pre výpočet intervalu spoľahlivosti je potrebný rozptyl $\hat{S}(t)$, ktorého odhad dostaneme použitím Greenwoodovho vzorca, ktorý vyzerá nasledovne

$$\text{var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} [8] \quad (2.5)$$

2.2.2. Medián prežitia

Pri Kaplan-Meierovom odhade funkcie prežitia sa často používa medián prežitia. Medián definuje dobu, ktorú pacient prežije s pravdepodobnosťou 0.5. V prípade, kedy je viac ako 50% dát cenzorovaných, nie je možné medián prežitia konštruovať. [5].



Obr. 2.2: Kaplan-Meierov odhad s mediánom funkcie pežitia . [3]

V obrázku 2.2 je znázornený medián prežitia, z čoho môžeme vidieť, že udáva

hodnotu na x-ovej osi, teda čas. V praxi sa nám to môže hodiť najmä v prípadoch, kedy porovnáваме viac funkcií prežitia, a ako veľmi sa od seba dané skupiny líšia.

Kapitola 3

Porovnávanie dvoch alebo viacerých funkcií prežitia

V praxi sa často vyskytujú prípady, kedy doba prežitia, napríklad po diagnostike choroby, alebo len bežnom zákroku, závisí na rôznych faktoroch. Ako faktory môžeme uvažovať všetko od nášho spôsobu života, pohlavia, až po vekovú skupinu či spôsob liečby. Avšak niektoré z nich sú naozaj kľúčové v rôznych situáciách. Preto sa pomocou týchto faktorov porovnávajú dve, alebo i viac funkcií prežitia, a testujeme, ktoré rozdiely sú štatisticky významné.

K tomu existuje niekoľko testov, ako sú napríklad: log-rank test, Tarone-Ware test, Wilcoxonov test a Fleming-Harrington test. [1] Tieto testy porovnávajú Kaplan-Meierove odhady funkcie prežitia pre rôzne úrovne faktorov, a testujú, či sú funkcie prežitia pri rôznych úrovniach faktoru rovnaké, alebo je medzi nimi významný rozdiel.

V tejto kapitole si uvedieme niektoré zo spomenutých testov.

3.1. Log-rank test

Log-rank test je jedným z najpoužívanejších a najjednoduchších testov, ktorý sa používa na overenie predpokladu o rovnosti funkcií prežitia pre rôzne skupiny subjektov. Taktiež sa označuje i pojmom Mantel-Haenszel test [4]. Jedná sa v podstate o chi-kvadrát test pre veľké rozsahy pozorovaní [3]. Pomocou log-rank testu overujeme zhodu dvoch funkcií prežitia (pre dve skupiny), no dá sa použiť aj pre porovnávanie väčšieho počtu skupín ako dú dve.

3.1.1. Log-rank test pre dve skupiny

V tomto prípade testujeme hypotézu:

$$H_0 : S_1(t) = S_2(t) \quad \forall t > 0$$

$$H_A : S_1(t) \neq S_2(t) \quad \forall t > 0,$$

kde $S_1(t)$ a $S_2(t)$ sú funkcie prežitia pre prvú a druhú skupinu.

Na začiatku predpokladáme, že máme n rôznych usporiadaných časov pre obe skupiny, kedy nastala pozorovaná udalosť. Pre každý čas t_j môžeme následne zostaviť kontingenčnú tabuľku 2x2 (Tabuľka 3.1), ktorá obsahuje pozorované udalosti v danom čase v oboch skupinách. Ďalej musíme zistiť počet výskytov pozorovanej udalosti (d_{ij}) a počet subjektov v riziku n_{ij} , kde i značí skupinu. [3]. V oboch skupinách sa počet subjektov môže líšiť, a v niektorých štúdiách môže byť dokonca i rovnaký.

Skupina	Počet udalostí	Počet bez udalosti	Počet v riziku
1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
2	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Celkom	d_j	$n_j - d_j$	n_j

Tabuľka 3.1: Tabuľka pre test rovnosti funkcií prežitia pre dve skupiny v j -tom čase. [4]

Následne rozšírime kontingenčnú tabuľku o očakávané četnosti, a taktiež o rozdiel medzi pozorovanými a odhadnutými četnosťami. Hodnoty uvažujeme pre jednotlivé skupiny v konkrétnom čase t_j . Očakávané počty pre obe skupiny v j -tom čase vypočítame ako:

$$e_{1j} = \left(\frac{n_{1j}}{n_{1j} + n_{2j}} \right) (d_{1j} + d_{2j}), \quad (3.1)$$

$$e_{2j} = \left(\frac{n_{2j}}{n_{1j} + n_{2j}} \right) (d_{1j} + d_{2j}), \quad (3.2)$$

kde prvá zátvorka označuje časť v rizikovej skupine zo všetkých subjektov, a druhá zátvorka značí počet udalostí v oboch skupinách v j -tom čase. [3]

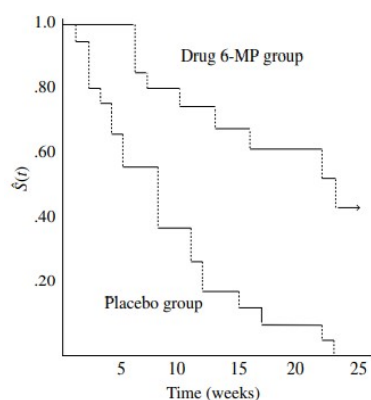
Testová štatistika pomocou ktorej budeme overovať hypotézu je

$$\text{Log-rank statistika} = \frac{(O_i - E_i)^2}{\text{Var}(O_i - E_i)}, i = 1, 2 \quad (3.3)$$

kde O_i a je súčet pozorovaných udalostí, a E_i je súčet očakávaných udalostí v i -tej skupine. Rozptyl je v tvare

$$\text{var}(O_i - E_i) = \sum_j \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)} \quad (3.4)$$

Testová štatistika má asymptoticky χ^2 rozdelenie s jedným stupňom voľnosti a porovnávame ju s jeho kritickou hodnotou.[4] Za platnosti H_0 by počet výskytov udalosti v oboch skupinách mal byť približne rovnaký, a očakávané četnosti sú blízke pozorovaným četnostiam. Naopak väčší rozdiel medzi výskytom udalosti značí rozdiel medzi skupinami.



Obr. 3.1: Porovnanie dvoch funkcií prežitia pre pacientov s leukémiou. [11]

Na obrázku vidíme názorný príklad, ako môžu vyzerat' odhady funkcií prežitia pre dve rôzne skupiny. Príklad je zamerný na pacientov s leukémiou, kde jednej skupine bol podaný placebo liek (vidíme, že funkcia prežitia klesá rýchlejšie), a druhej skupine bol podaný liek 6-MP (funkcia má menej "schodov", klesá pomalšie).

3.1.2. Log-rank test pre 3 a viac skupín

Ako už bolo spomenuté, log-rank test je možné použiť aj pre testovanie ekvivalenčnosti viac ako dvoch funkcií prežitia. H_0 bude v tomto prípade predpokladať, že všetky funkcie prežitia budú rovnaké.

Opäť pracujeme s n rôznymi časmi, kedy došlo k sledovanej udalosti, a tentokrát máme počet skupín $G \geq 2$, ktoré budeme porovnávať. Potom e_{ij} je počet očakávaných výskytov pozorovanej udalosti v i -tej skupine v j -tom čase, a má rovnaký tvar ako

v predchádzajúcom prípade, s jediným rozdielom, že počítame očakávané četnosti pre každú zo skupín. Ďalej n_{ij} je počet subjektov v riziku a d_{ij} je pozorovaný počet výskytov udalosti, v oboch prípadoch uvažujeme i -tu skupinu v j -tom čase. Celkové súčty tentokrát môžeme vyjadriť vzorcami $n_j = \sum_{i=1}^G n_{ij}$ a $d_j = \sum_{i=1}^G d_{ij}$.

Keďže porovnávame viac ako dve skupiny, používame len miesto rozptylu aj kovariancie jednotlivých rozdielov pozorovaných a odhadnutých počtov udalostí pre všetky skupiny.

$$O_i - E_i = \sum_{j=1}^n (d_{ij} - e_{ij}), \quad i = 1, \dots, G \quad (3.5)$$

$$\text{var}(O_i - E_i) = \sum_{j=1}^n \frac{n_{ij}(n_j - n_{ij})d_{ij}(n_j - d_j)}{n_j^2(n_j - 1)} \quad (3.6)$$

$$\text{Cov}(O_i - E_i, O_k - E_k) = \sum_{j=1}^n \frac{-n_{ij}n_{kj}d_j(n_j - d_j)}{n_j^2(n_j - 1)} \quad (3.7)$$

Potom d je vektor rozdielov $O_i - E_i$, rovný $d = (O_1 - E_1, \dots, O_{G-1} - E_{G-1})$.

A V je variančná matica s rozptylmi $\text{Var}(O_i - E_i)$ na digonále, a kovarianciami $\text{Cov}(O_i - E_i, O_k - E_k)$ mimo diagonálu, kde $i, k = 1, \dots, G - 1$. Nakoniec dostaneme testovú štatistiku pre porovnávanie viacerých funkcií prežitia.

$$\text{Log - rank statistika} = d^t V^{-1} d \quad (3.8)$$

a má asymptoticky χ^2 rozdelenie o $G-1$ stupňoch voľnosti. [3]

3.2. Alternatívy log-rank testu

Log-rank test je vhodné použiť v prípade, kedy rozdiel medzi funkciami prežitia narastá. Avšak keď dochádza ku kríženiu funkcií, je lepšie použiť jeden z alternatívnych testov k log-rank testu. Nakoľko práve tento test nie je schopný správne posúdiť rozdiel medzi krivkami.

Alternatívne testy, ktoré je možné použiť, sú zobecnený Wilcoxonov test (nazývaný aj ako Gehan test, resp. Breslow test), Tarone-Ware test, Fleming-Harrington test a Peto test. Všetky tieto testy sú založené na princípe log-rank testu a pracujú s očakávanými a pozorovanými počtami výskytu udalosti.

Testovú štatistiku používame rovnakú pre všetky testy, a je v tvare:

$$\text{Statistika pre skupinu } i = \frac{\sum_j w_j (d_{ij} - e_{ij})^2}{\text{var}(\sum_j w_j (d_{ij} - e_{ij}))}, \quad i = 1, \dots, G \quad (3.9)$$

kde i značí skupinu, a j znázorňuje čas v ktorom nastala pozorovaná udalosť. Z tohto vzorca môžeme vidieť, že pokiaľ za w_j zvolíme 1, dostávame testovú štatistiku pre log-rank test. Alternatívne testy sa líšia v tom, že pri každom z nich volíme inú váhu w_j . [2] Používané váhy pre jednotlivé testy su uvedené v Tabuľke 3.2.

Test	Váhy w_j
Wilcoxon test	n_j
Tarone-Ware test	$\sqrt{n_j}$
Peto test	$\hat{S}(t_j)$
Fleming-Harrington test	$\hat{S}(t_{(j-1)})^p (1 - \hat{S}(t_{(j-1)}))^q$

Tabuľka 3.2: Tabuľka udávajúca váhy pre testy o zhode funkcií prežitia [3].

Zobecnený Wilcoxonov test je rozšírený klasický Wilcoxonov test pre cenzorované dáta [4]. Váhy sú v tomto prípade rovné počtu subjektov v riziku, čo znamená, že váhy sú vyššie pre rozdiely nachádzajúce sa na začiatku doby pozorovania.[8] Postupujúcim časom budú subjekty ubúdať, či už z dôvodu cenzúrovania alebo výskytu udalosti, preto budú neskôr váhy menšie.

Testová štatistika pre Tarone-Ware test je podobná tej z Wilcoxonovho testu, pretože taktiež pracuje s počtom subjektov v riziku, avšak s jeho druhou odmocinou. A teda aj v tomto prípade sú váhy vyššie na začiatku pozorovanej doby.

V prípade Peto testu sa ako váhová funkcia používa upravený odhad celkovej funkcie prežitia značený ako \tilde{S} . Details ohľadom použitého odhadu je možné nájsť na stránke¹, resp. v knihe [3] na str.74.

V testovej štatistike Fleming-Harringtonovho testu je použitý Kaplan-Meierov odhad funkcie prežitia. Váhy pre časové okamihy ,kedy nastala udalosť, sú výsledky Kaplan-Meierovho odhadu funkcie prežitia v čase t_{j-1} s váhou p , a $1-\hat{S}(t_{j-1})$ s váhou q , pre $p, q \geq 0$. V prípade, kedy je $p > q$, sú váhy v tomto teste vyššie pre skoršie

¹https://cran.r-project.org/web/packages/survminer/vignettes/Specifying_weights_in_log-rank_comparisons.html#peto-petos-modified-survival-estimate

výskyty udalosti. Naopak, kedy je $p < q$ sú váhy vyššie pre neskoršie výskyty udalosti. Nakoniec keď je $p, q = 0$, sú váhy rovné 1, a z testu sa stane log-rank test.[12]

Všetky testové štatistiky spomenutých testov v tabuľke 3.2 majú χ^2 rozdelenie o $G-1$ stupňoch voľnosti, kde G je počet porovnávaných skupín. Testy sú taktiež vhodné pre cenzurované dáta.

3.3. Výber testu

Všetky zo spomínaných testov s rôznymi váhami by mali mať relatívne približné výsledky, ktoré by viedli k rovnakému záveru (platnosť/zamietnutie nulovej hypotézy). Najčastejšie používaný a uvádzaný je log-rank test. Pokiaľ majú testy rôznu štatistickú významnosť, je lepšie interpretovať výsledky viacerých testov, pretože pomocou toho môžeme vidieť, kde približne sa funkcie prežitia môžu líšiť. Napríklad keď použijeme log-rank test a H_0 zamietame, tak to znamená že je významný rozdiel medzi funkciami, avšak nevieme kde presne. Ak k tomu použijeme i zobecnený Wilcoxonov test, s dôrazom na začiatok štúdie, a H_0 budeme taktiež zamietat', vieme určiť, že rozdiel je štatisticky významný na začiatku štúdie. Je nutné, aby test nebol zvolený na základe toho, aký výsledok by bol vhodný pre nás (používateľa testu). Test je preto dobré voliť podľa uváženia, typu štúdie a potrebnej veľkosti dôrazu na jednotlivé výskyty udalosti, aby sme správne zachytili rozdiel medzi skupinami. [4]

Kapitola 4

Coxov model

V predchádzajúcom prípade sa používali doba prežitia (resp. výskyt udalosti), a kategoriálne premenné pri porovnávaní skupín. Avšak často k týmto faktorom máme viac informácií o stave subjektu, ktorého môžu popisovať i spojité premenné.

Napríklad, v prípade onkologických pacientov môžeme mať informáciu o kategoriálnych faktoroch ako sú pohlavie a typ životného štýlu (fajčenie, pitie alkoholu), ale aj o spojitých premenných ako sú množstvo určitého typu buniek vo vzorku, hodnota kyslíku v krvi a mnoho ďalších.

Tieto všetky premenné môžu výrazne ovplyvniť dobu prežitia pacienta, a označujeme ich ako nezávislé premenné. Práve pomocou Coxovho modelu vieme zahrnúť väčšie množstvo faktorov do modelovania funkcie prežitia. Coxov model patrí do skupiny semiparametrických metód, a je taktiež nazývaný aj ako Coxov model proporcionálneho hazardu [1].

Tento názov sa používa z toho dôvodu, že model je definovaný pomocou rizikovej funkcie. Riziková funkcia so zahrnutým vplyvom faktorov je definovaná ako

$$h(t, z_1, \dots, z_p) = h_0(t)e^{\beta_1 z_1 + \dots + \beta_p z_p} = h_0(t)e^{\beta_1 z_1} \dots e^{\beta_p z_p} \quad (4.1)$$

Výraz sa skladá z dvoch častí, kde prvá je h_0 a nazýva sa základný hazard [1], ktorý je spoločný pre celý súbor subjektov a značí očakávané riziko za predpokladu, že nulovosti všetkých uvažovaných faktorov. Druhá časť je vplyv p faktorov z_1, \dots, z_p , ktoré sú označované ako vysvetľujúce premenné. Veľkosť ich vplyvu je vyjadrená regresnými parametrami β_1, \dots, β_p , ktorých hodnotu dopredu nepoznáme a musíme ich odhadnúť [11].

Tieto regresné parametre udávajú zmenu rizika výskytu sledovanej udalosti, a teda o koľko sa zvýši hodnota funkcie hazardu za jednotkové zvýšenie konkrétnej premennej s predpokladom nemenných ostatných faktorov. Pokiaľ je regresný parameter kladný, značí to, že riziko výskytu sledovanej udalosti u subjektu je vyššie s vyššou hodnotou konkrétnej vysvetľujúcej premennej. Naopak, so zápornou hodnotou vysvetľujúcej premennej je menšie riziko výskytu udalosti, hovoríme, že vysvetľujúca premenná má s vyššou hodnotou protektívny účinok [2].

Z predchádzajúcej rovnice je možné vyjadriť výraz

$$\frac{h(t, z_1, \dots, z_p)}{h_0(t)}, \quad (4.2)$$

ktorý sa nazýva relatívne riziko (resp. podiel hazardu).[1] Relatívne riziko vyjadruje rast alebo pokles rizika, ktorý je spôsobený vplyvom vysvetľujúcich premenných z_1, \dots, z_p . Hodnota e^{β_i} je označovaná ako HR (hazard ratio), a rozlišujeme 3 situácie

- $\beta_i = 0$, značí, že i -tý faktor nemá žiaden vplyv na dobu prežitia,
- $\beta_i > 0$ (ekvivalentne $e^{\beta_i} > 1$), značí zvýšenie vplyvu i -tého faktoru, čo spôsobí nárast rizika výskytu udalosti, a zníženie doby prežitia
- $\beta_i < 0$ (ekvivalentne $e^{\beta_i} < 1$), značí zníženie vplyvu i -tého faktoru, čo spôsobuje pokles rizika výskytu udalosti, a zvýšenie doby prežitia. [14]

4.1. Funkcia prežitia

Tak ako v prípade Kaplan-Meiera, aj tu konštruujeme funkciu prežitia, avšak s využitím Coxovho modelu. Odhad funkcie prežitia využíva vysvetľujúce premenné, a jej predpis je v tvare

$$S(t, Z) = S_0(t)^{\exp\{\sum_{i=1}^p \beta_i z_i\}} [13] \quad (4.3)$$

kde $Z = (z_1, \dots, z_p)$ je vektor vysvetľujúcich premenných. A keďže β_i nepoznáme, nahradíme ich odhadnutou hodnotou $\hat{\beta}_i$, taktiež aj $S_0(t)$ hodnotou $\hat{S}_0(t)$.

Tieto odhady často v praxi dostávame pomocou využitia softwarových programov ako sú R, Statistica a mnoho ďalších.

Funkcia prežitia odhadnutá s Coxovým modelom je taktiež schodovitá funkcia.

Kapitola 5

Praktické využitie teórie analýzy prežívania

Ako bolo spomínané v teoretickej časti, najčastejšie sa s analýzou prežívania stretávame práve v oblasti medicíny, kedy chceme zistiť dobu prežitia od výskytu udalosti (najčastejšie diagnózy určitou chorobou), a faktory ovplyvňujúce túto dobu. Cieľom v nasledujúcich kapitolách bude vyskúšať si, a ukázať jednotlivé techniky analýzy prežívania, ktoré boli spomenuté vyššie, v praxi. Zároveň taktiež porovnať faktory, ktoré by mohli mať najväčší vplyv na dobu prežitia v rôznych skupinách pacientov.

5.1. Dátová sada a popis premenných

Dátová sada s ktorou budem pracovať obsahuje údaje o pacientoch, ktorým bol diagnostikovaný určitý druh choroby. Prvých pár riadkov dát je vyobrazených v Tabuľke 5.1.

ODBER	VEK	POHLAVIE	FAJČENIE	FAJČENIEa	ALKOHOL	ALKOHOLa	pS	grading	KONTROLA	FSTAV	HPV.RNA	Tc_PD1p
1968-04-19	59	1	1	2	1	1	4	2	1978-06-15	1	1	39.402723
1968-03-08	73	2	1	2	2	2	3	3	1973-03-17	1	1	53.744156
1968-01-04	50	1	1	1	1	1	4	3	1969-11-09	4	0	2.317807
1968-04-25	42	1	1	1	1	1	4	1	1971-04-20	1	1	33.900029
1968-05-04	59	1	1	1	1	1	2	1	1974-12-28	1	0	1.311571
1968-05-10	55	1	1	1	1	1	4	2	1975-07-04	5	0	17.238691

Tabuľka 5.1: Ukážka dát

V ukážke môžeme vidieť prvých 6 pacientov, z celkovo pozorovaných 97, a taktiež premenné v stĺpcoch, ktoré bližšie charakterizujú stav pacientov. Dve premenné sú dátumové, jedna premenná je kvantitatívna a desať premenných je kvalitatívnych.

Pre každého pacienta máme údaj o dátume odberu vzorku (stĺpec ODBER), ktorý značí, že od tohoto dátumu bol pacient a jeho stav pozorovaný. Druhý údaj s dátumom je v stĺpci KONTROLA, a značí posledný dátum kontroly pacienta a jeho aktuálneho stavu.

Ďalej máme nasledujúce údaje:

1. VEK
2. POHLAVIE:
 - (a) 1-muž
 - (b) 2-žena

Nasledujúce stĺpce sú kódované ako 1-áno, 2-nie.

1. FAJČENIE: Údaj o tom, či daný človek pred diagnózou fajčil, alebo nie.
2. FAJČENIEa: Stĺpec udáva či pacient po dignóze fajčil, alebo nefajčil.
3. ALKOHOL: Informácia, či pacient pred diagnostikovaním choroby pil alkohol, alebo nie.
4. FAJČENIEa: Udáva, či pacient po diagnóze konzumoval alkohol, alebo nie.

Premenné s iným kódovaním.

1. PS,GRADING: Obe premenné značia v akom štádiu bola rakovina zachytená. Sú škálované 1-5, pričom 1 značí najľahšie a 5 najhoršie štádium.
2. FSTAV: Údaj znázorňuje stav pacienta v čase poslednej kontroly. Môže sa vyskytnúť šesť prípadov.
 - (a) 1-pacient žije bez známkov tumoru
 - (b) 2-pacient žije s tumorom
 - (c) 3-pacient žije, ale jeho stav nieje známy
 - (d) 4-pacient zomrel s tumorom
 - (e) 5-pacient zomrel bez známkov tumoru

(f) 6-pacient zomrel z neznámej príčiny

3. HPV.RNA: Značí, či má pacient rakovinu vírusového, alebo nevírusového pôvodu.

(a) 0-nie

(b) 1-áno

Posledná premenná je spojitá (Tc_PD2p), a udáva výsledok laboratórneho vyšetrenia, kde hodnota značí množstvo buniek určitého typu vo vzorku.

5.2. Popisná štatistika

Aby sme si vedeli predstaviť, ako skupina pacientov vyzerá, je dobré si predstaviť aspoň základné informácie o pacientoch.

- Medzi pacientami je väčší počet mužov, a to 78 (80.41%), zatiaľčo žien len 19 (19.59%). Vekové zloženie skupín je približne rovnaké, pretože priemerný vek žien je 59 rokov a mužov 56. Avšak vek mužov sa pohybuje v rozmedzí od 29 do 84 rokov, zatiaľ čo vek žien má menší rozsah, a to 31 až 73 rokov.
- Celkovo sa u 45 pacientoch vyskytla rakovina vírusového pôvodu, a u 52 pacientoch šlo o rakovinu nevírusového pôvodu. Rozdelenie podľa pohlavia a rakoviny vírusového pôvodu si môžeme pozrieť v Tabuľke 5.2, kde vzhľadom k nepomeru veľkosti skupín sú udávané aj relatívne početnosti zastúpení skupín.

	nevírusový pôvod	vírusový pôvod
muži	43 (55 %)	35 (54 %)
ženy	9 (47 %)	10 (53 %)

Tabuľka 5.2: Výskyt pôvodu rakoviny rozdelený podľa pohlavia.

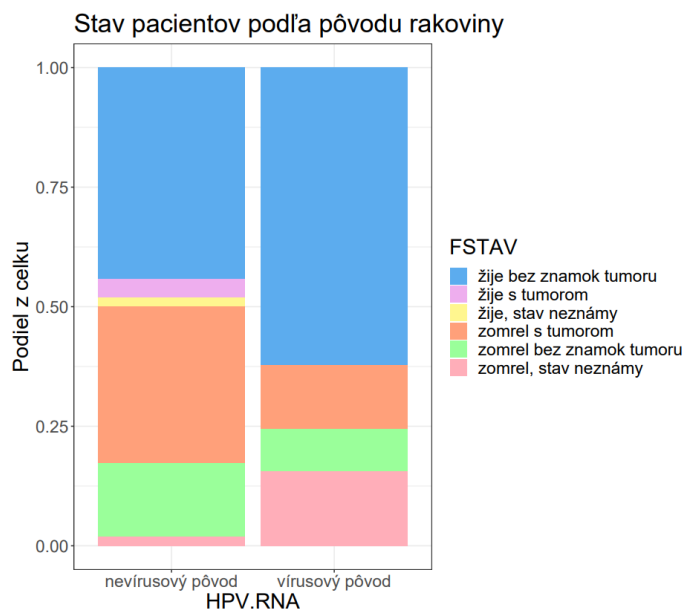
Vidíme, že u mužov prevláda rakovina nevírusového pôvodu (55 %), a u žien prevláda skôr rakovina vírusového pôvodu (53 %). To, či má pohlavie alebo pôvod rakoviny vplyv na dobu prežitia budeme skúmať v ďalšom texte.

- Rozlišujeme šesť rôznych kategórií pacientov podľa ich stavu. Celkové počty sú uvedené v Tabuľke 5.3.

Stav	počet
žije bez známkov tumoru	51
žije s tumorom	2
žije, stav neznámy	1
zomrel s tumorom	23
zomrel bez známkov tumoru	12
zomrel, stav neznámy	8

Tabuľka 5.3: Četnostná tabuľka uvádzajúca stav pacientov.

Ako môžeme vidieť, najväčší počet pacientov žije bez známkov tumoru, avšak počet pacientov, ktorí zomreli s tumorom je taktiež celkom vysoký. Je zaujímavé sa pozrieť i na informáciu, koľko pacientov má v jednotlivých skupinách rakovinu vírusového pôvodu.



Obr. 5.1: Rozdelenie podľa pôvodu choroby.

V Obrázku 5.1, vidíme že väčšie percento pacientov s rakovinou vírusového pôvodu žije bez známkov tumoru (62 % oproti 44 %). Naopak, väčšie percento pacientov (33 % oproti 13 %), u ktorých sa rakovina vírusového pôvodu nevyskytla,

zomrelo s vyskytujúcim sa tumorom.

- Až 80 % pacientov uviedlo, že pred diagnostikovaním choroby fajčilo. Z celkového počtu mužov to bolo 83 %, a z celkového počtu žien to bolo 68 %. To, ako sa celkový počet a celkové percento zmenilo po zistení diagnózy je v Tabuľke 5.4.

Fajčenie pred diagnózou/ po diagnóze	áno	nie
áno	46 (47%)	32 (33%)
nie	0 (0%)	19 (20%)

Tabuľka 5.4: Porovnanie počtu pacientov, ktorí fajčili pred/po dignóze

Percentuálne zatúpenie pacientov, ktorí fajčili pred diagnózou, sa po diagnóze zmenilo z 80 % na 47 %, čo je takmer o polovicu. Naopak, ani jeden pacient po diagnóze nezačal fajčiť. Z hľadiska pohlavia, percento mužov, ktorí prestali fajčiť, je 30 % (zmena z 83 % na 53 %), a percento žien, ktoré prestali fajčiť, je 42 % (zo 68 % na 26 %).

- Takmer rovnaký počet pacientov ako v predchádzajúcom prípade uviedlo, že pije alkohol, a to 79 %. Z mužov to bolo až 87 %, a zo žien 47 %. To, ako sa tieto hodnoty zmenili, je uvedené v Tabuľke 5.5.

Alkohol pred diagnózou/ po diagnóze	áno	nie
áno	64 (66%)	13 (13%)
nie	0 (0%)	20 (21%)

Tabuľka 5.5: Porovnanie počtu pacientov, ktorí konzumovali alkohol pred/po dignóze.

V prípade alkoholu sa celkové percento nezmenilo tak veľmi ako v predchádzajúcom prípade. Množstvo pacientov, ktorí prestali s konzumáciou alkoholu, je 13 % (z celkových 79 % na 66 %). Takisto aj v tomto prípade žiaden z pacientov nezačal s konzumáciou alkoholu. Na toto množstvo sa môžeme pozrieť taktiež z hľadiska pohlavia, kde prestalo piť alkohol 16 % mužov. Zaujímavé ale je, že ani jedna žena s pitím alkoholu po diagnóze neprestala.

- Pred diagnostikovaním choroby fajčilo a pilo alkohol 66 pacientov, po diagnostikovaní len polovica z toho počtu. Čiže neskôr ostalo presne 33 pacientov, ktorí zároveň fajčili a konzumovali alkohol.

K ďalšej analýze bude potrebné pridať ďalšiu premennú, ktorá bude obsahovať časový údaj o tom, ako dlho sme jednotlivých pacientov pozorovali. Momentálne, ako sme v Tabuľke 5.1 videli, boli k dispozícii iba konkrétne dátumy začiatku a konca pozorovania pacientov. K analýze prežívania a vytváraniu jednotlivých grafov je vhodné doplniť údaj časového rozmedzia týchto dátumov. Vytvorená premenná uvádza dĺžku pozorovania pacientov v rokoch.

5.3. Modelovanie funkcie prežitia

V tejto kapitole si ukážeme odhad jednotlivých funkcií prežitia Kaplan-Meierovou metódou, a porovnáme ich s vytvorenými odhadmi pomocou Coxovho modelu.

K samotnému odhadu funkcie je potrebné poznať dobu prežitia pacientov (udávanú napríklad v dňoch, týždňoch, rokoch). Dobu prežitia v našich dátach máme udanú dátumom odberu a dátumom poslednej kontroly. Pomocou týchto dvoch stĺpcov určíme dobu prežitia jednotlivých pacientov v rokoch. Taktiež budeme potrebovať rozlíšiť u pacientov či došlo k pozorovanej udalosti alebo nie. Za pozorovanú udalosť chápeme, že pacient zomrel. Preto jednotlivým pacientom priradíme hodnotu podľa premennej FSTAV tak, že tým, u ktorých udalosť nastala, priradíme 1. Naopak pacientom, u ktorých sa udalosť nevykytla, priradíme 0. Nakoniec k Tabuľke 5.1 pridáme dva opísané stĺpce znázornené v Tabuľke 5.6.

diff_years	STAV
10.15	0
5.02	0
1.85	1
2.98	0
6.65	0
7.15	1

Tabuľka 5.6: Rozšírenie dátovej tabuľky.

Po rozdelení pacientov podľa výskytu udalosti dostávame počet pacientov bez udalosti 54, a počet pacientov s udalosťou (ktorí zomreli) 43.

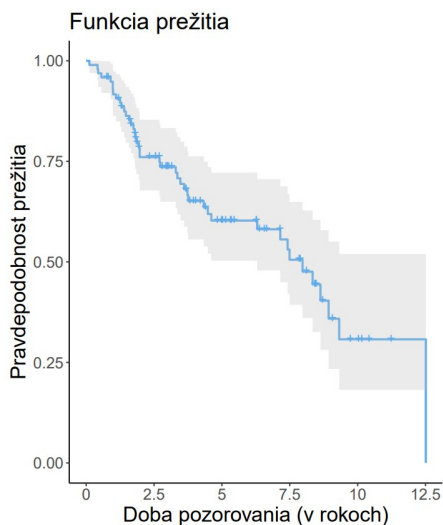
Ako prvú odhadneme základnú funkciu prežitia spoločnú pre všetkých pacientov, bez ohľadu na hodnotu akejkoľvek vysvetľujúcej premennej. Pre odhad použijeme Kaplan-Meierovu metódu. Túto funkciu na záver porovnáme s funkciou prežitia odhadnutou pomocou Coxovho modelu pre významné premenné. Ďalej sa pozrieme na funkcie prežitia vytvorené s použitím jednotlivých premenných, odhadnuté pomocou Coxovho modelu a Kaplan-Meierovej metódy.

5.3.1. Kaplan-Meierov odhad funkcie prežitia bez rozdelenia na skupiny

Najskôr si vykreslíme základnú funkciu prežitia pre získané dáta, ktoré zatiaľ nebudeme nijak rozdeľovať. Ako prvé si môžeme vypísať vzostupne usporiadané časy priežitia.

0.12	0.42	0.44	0.56	0.76+	0.81+	0.90	0.98	0.98	0.99
1.10	1.19+	1.25	1.30	1.30+	1.40	1.46	1.58	1.61+	1.61+
1.65	1.65+	1.74	1.79	1.80+	1.81	1.83+	1.85	1.88+	1.94
1.95+	1.97	1.97	2.33+	2.55+	2.69+	2.70	2.73	2.80+	2.94+
2.98+	2.99+	3.04+	3.15+	3.30	3.36	3.47	3.61	3.67+	3.68+
3.73	3.76	3.82+	3.95+	4.03+	4.20+	4.33	4.39+	4.48	4.61
4.83+	4.98+	5.02+	5.14+	5.31+	5.34+	5.45+	6.26+	6.27+	6.30
6.38+	6.56+	6.65+	7.14+	7.15	7.42	7.49	7.84+	7.87+	7.88+
7.97	8.11+	8.34	8.42+	8.44+	8.47+	8.63	8.71+	8.93	9.07+
9.32	9.73+	10.03+	10.15+	10.42+	11.23+	12.51			

Pri určitých časoch je pripísané +, ktoré značí, že nedošlo k sledovanej udalosti, a teda máme cenzúrované pozorovanie. Tento údaj je v stĺpci STAV označený číslom 0. Ďalej pomocou funkcie `survfit` v programe R odhadneme funkciu prežitia, ktorá vyzera nasledovne.



Obr. 5.2: Kaplan-Meierova funkcia prežitia s 95% intervalom spoľahlivosti.

V Obrázku 5.2 je odhadnutá funkcia prežitia, v ktorej každý schod znázorňuje pacienta, u ktorého nastala pozorovaná udalosť. Taktiež na obrázku môžeme vidieť 95% interval spoľahlivosti, ktorý je znázornený šedou farbou okolo funkcie prežitia. Na funkcii sa nachádzajú i spomínané +, kedy dochádza k cenzúre pacienta. K vykreslenej funkcii prežitia si môžeme prehľadne vypísať aj základné charakteristiky. Tie v programe R vypíšeme pomocou funkcie `summary`, do ktorej vložíme odhadnutý model funkcie prežitia.

time	n.risk	n.event	n.censor	estimate	std.error	conf.high	conf.low
0.12	97	1	0	0.9896907	0.0103628	1.0000000	0.9697921
0.42	96	1	0	0.9793814	0.0147322	1.0000000	0.9515066
0.44	95	1	0	0.9690722	0.0181389	1.0000000	0.9352253
...
10.15	4	0	1	0.3074666	0.2675261	0.5194176	0.1820033
10.42	3	0	1	0.3074666	0.2675261	0.5194176	0.1820033
11.23	2	0	1	0.3074666	0.2675261	0.5194176	0.1820033
12.51	1	1	0	0.0000000	Inf	NA	NA

Tabuľka 5.7: Charakteristiky funkcie prežitia pre celý súbor pacientov.

V grafe funkcie prežitia nie vždy presne vidíme, v ktorých časoch nastane sledovaná udalosť, resp. u koľkých pacientov, prípade parametre intervalu spoľahlivosti. Preto je možné sa na tieto údaje pozrieť bližšie tak, ako sú uvedené v Tabuľke 5.7, kde sú vypísané prvé a posledné tri údaje (rozšírená verzia tabuľky sa nachádza v

Dodatku A). V tabuľke je ako prvý čas, kedy nastane pozorovaná udalosť, a následne stĺpec **n.event** udáva počet udalostí, koľko v tento čas nastalo. Ďalej máme **n.risk**, čo znamená počet pacientov v riziku. Taktiež sa v tabuľke nachádza počet cenzúrovaných pacientov (**n.censor**) v danom čase. **Estimate** je samotný odhad funkcie prežitia, a po vynásobení 100 udáva odhad percenta pacientov, ktorí prežijú dlhšie ako konkrétny čas. V posledných dvoch stĺpcoch sú hodnoty 95 % intervalu spoľahlivosti pre odhad funkcie prežitia.

Na záver vidíme, že posledný čas je 12.5 roka. V tomto čase zostal jediný pacient v riziku, u ktorého nastala sledovaná udalosť. Po tomto čase i v hodnote estimate vidíme, že žiaden z pacientov a jeho stav nebol evidovaný po tomto čase, a teda 0 pacientov prežije tento čas. Avšak pacienti, ktorí boli cenzúrovaní, môžu stále tento čas prežiť, no taký dlhý údaj o nich nemáme.

5.3.2. Odhad funkcie prežitia pomocou Coxovho modelu bez rozdelenia na skupiny

Videli sme odhad základnej funkcie prežitia, ktorá nezohľadňuje hodnoty iných premenných ako je doba prežitia. Avšak o pacientoch máme viac informácií (vstupných premenných), ktoré boli počas ich doby sledovania zozbierané, a môžu mať na dobu prežitia vplyv. To, ktoré vstupné premenné sú naozaj významné a najviac ovplyvňujú dobu prežitia nevieme presne určiť len z grafu funkcie, a pohľadu na zozbierané dáta. K zisteniu významnosti premenných, a ako prežívanie ovplyvňujú využijeme Coxov model.

Najskôr pre každú premennú zvlášť zistíme významnosť v Coxovom modeli. Následne vyberieme významné premenné a bližšie ich zanalyzujeme. Vstupné premenné a ich významnosť v základom Coxovom modeli s konkrétnou jednou premennou sú uvedené v Tabuľke 5.8.

vstupná premenná	$\hat{\beta}$	odhad HR	p -hodnota
vek	-0.02	0.98	0.223
pohlavie	0.07	1.07	0.8700
fajčenie	-1.01	0.36	0.0400
fajčenie po diag	-1.07	0.34	0.0020
alkohol	-0.24	0.79	0.5600
alkohol po diag	-0.22	0.81	0.5200
HPV.RNA	-1.24	0.29	0.0005
bunky	-0.006	0.99	0.0880

Tabuľka 5.8: Výsledky významnosti premenných v jednotlivých modeloch.

V tabuľke sú výsledky testu o nulovosti jednotlivých premenných a teda ich vzťahu k celkovému prežitiu. Uvedené sú odhady parametru β , ktorých hodnota odpovedá nárastu, resp. poklesu rizika výskytu udalosti. Ďalej máme hodnotu odhadu hazard ratio ($\text{HR}, \exp(\hat{\beta})$), ktorá popisuje zmenu rizika pri jednotkovej zmene vstupnej premennej. Rozhodovať o významnosti sa budeme na základe p -hodnoty. Z porovnania p -hodnoty s hladinou významnosti $\alpha = 0.05$ je zrejmé, že významné premenné vstupujúce do modelu sú vírusový pôvod choroby, fajčenie pred a po diagnóze.

Interpretáciu odhadu β a HR si predvedieme na spojitaj premennej vek. Záporná hodnota odhadu β značí menšie riziko výskytu udalosti, a hodnota odhadu HR 0.98 udáva zmenu výskytu udalosti pri jednotkovej zmene veku. Z týchto dvoch hodnôt vieme určiť, že jednotkový nárast veku znižuje riziko výskytu udalosti 0.98krát. Kategoriálne premenné si bližšie rozoberieme v nasledujúcom texte.

Najskôr si podrobne prejdeme jednotlivé vysvetľujúce premenné v modeli, a aký vplyv majú na dobu prežitia. Následne všetky významné premenné spojíme do jedného kompletného modelu.

5.3.3. Porovnanie funkcií prežitia podľa rôznych vysvetľujúcich premenných

V úvode sme si ukázali, ako môže vyzerat' základná funkcia prežitia pre celý súbor pacientov bez prihliadania na ďalšie faktory. V tejto časti si uvedieme príklady, kedy pacienti budú rozdelení na viac skupín, a jednotlivé funkcie prežitia porovnáme. Na príklady použijeme najmä tie premenné, ktoré vyšli v predchádzajúcom prípade

významné. Pre tieto premenné odhadneme funkcie prežitia podľa Kaplan-Meierovej metódy i pomocou Coxovho modelu. Taktiež si pomocou Kaplan-Meierovej metódy môžeme vykresliť i premenné, ktoré v predchádzajúcom prípade nevyšli ako štatisticky významné asociované s dĺžkou doby prežitia.

Vplyv pôvodu rakoviny na dobu prežitia

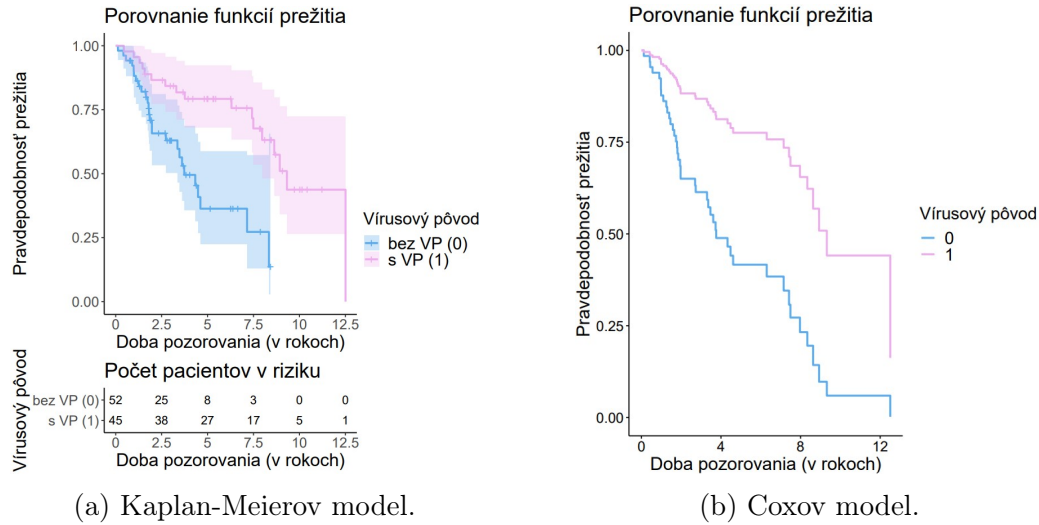
Ako prvá významná premenná v Coxovom modeli vyšla premenná týkajúca sa pôvodu rakoviny. Preto sa pozrieme na rozdiel funkcií prežitia medzi rakovinou vírusového pôvodu a rakovinou nevírusového pôvodu. Četnosti pre jednotlivé skupiny sú vyobrazené v Tabuľke 5.9.

HPV.RNA	celkový počet	výskyt udalosti	cenzerovaný pacienti	percento cenzúrovaných pacientov v skupine
áno (1)	45	17	28	62.22%
nie (0)	52	26	26	50%

Tabuľka 5.9: Rozloženie v skupinách podľa pôvodu rakoviny.

Vidíme, že zastúpenie v skupinách je takmer vyvážené. V prípade vírusového pôvodu zomrelo 17 pacientov, a v prípade nevírusového pôvodu 26. Graf pre odhadnuté funkcie prežitia Kaplan-Meierovou metódou je znázornený v Obr. 5.3a, kde je i zjednodušená tabuľka opisujúca vývoj funkcie prežitia, kde vidíme meniaci sa počet pacientov v riziku. V grafe sú znázornené aj 95% intervaly spoľahlivosti.

Hneď vedľa, v Obr. 5.3b, sú odhadnuté funkcie prežitia vytvorené Coxovým modelom.



Obr. 5.3: Funkcie prežitia podľa pôvodu rakoviny.

Vidíme, že postupom času narastá rozdiel medzi funkciami prežitia pre obe skupiny, v oboch grafoch. V Kaplan-Meierovom odhade vidíme čas prežitia dlhší pre jednu skupinu, a to pacientov s rakovinou s vírusovým pôvodom. Krivka pre týchto pacientov klesá pomalšie. Či sa aj v tomto prípade potvrdí štatistická významnosť premennej overíme. Výsledky sú uvedené v Tabuľke 5.10.

	Chi-sq	df	<i>p</i> -hodnota
Mantel-Heanszel test (Log-rank)	13.3	1	0.0003

Tabuľka 5.10: Test rovnosti funkcií prežitia na základe pôvodu rakoviny.

Významnosť sme overili pomocou log-rank testu, kde *p*-hodnota vyšla 0.0003, čo je menšia hodnota ako hladina významnosti $\alpha = 0.05$, na ktorej sme hypotézu testovali. Preto hypotézu o rovnosti funkcií prežitia zamietame.

Tento test porovnáme s výsledkom Coxovho modelu s jednou vstupnou premennou.

	$\hat{\beta}$	$se(\hat{\beta})$	<i>p</i> -hod	$\exp(\hat{\beta})$	testová štatistika	lower 95% KI	upper 95% KI
HPV.RNA=1	-1.2379	0.3564	0.0005	0.29	12.07	0.1442	0.5831

Tabuľka 5.11: Výsledky Coxovho modelu.

Ako bolo spomenuté v teoretickej časti, hodnota regresného parametru je určená pomerom rizík sledovanej udalosti v porovnávaných skupinách. Coxov model porovnáva

hodnoty rizika v druhej skupine vzhľadom k prvej skupine. Keďže je v našom prípade premenná HPV.RNA kódovaná 0,1, budeme výsledky interpretovať ako skupina pacientov s vírusovým pôvodom rakoviny (1), vzhľadom k skupine pacientov s rakovinou nevírusového pôvodu (0).

Pre premennú v modeli boli odhadnuté nasledujúce hodnoty: β , smerodatná odychlka pre β , Waldova štatistika, p -hodnota, hazard ratio ($\exp(\hat{\beta})$), a 95% interval spoľahlivosti pre HR.

V Kaplan-Meierovom modeli sa potvrdila štatistická významnosť, ktorú môžeme určiť i z výsledku Coxovho modelu. A to pomocou odhadu β , ktorý je záporný, čo značí menšie riziko výskytu udalosti pre skupinu pacientov s vírusovým pôvodom rakoviny. Následne hodnota $\exp(\hat{\beta})=0.29$ značí, že pokiaľ má pacient rakovinu vírusového pôvodu má v ktoromkoľvek okamihu 0.29krát nižšie riziko úmrtia oproti pacientom s rakovinou nevírusového pôvodu. Tento rozdiel je viditeľný i v grafoch, keďže funkcia prežitia pre pacientov s vírusovým pôvodom má až do konca doby pozorovania vyššie pravdepodobnosti prežitia.

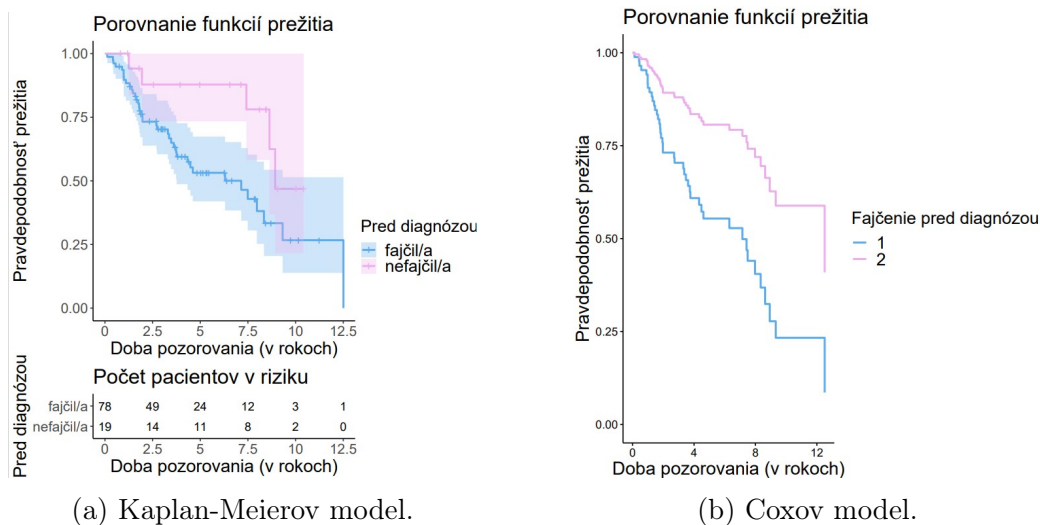
Vplyv fajčenia pred diagnostikovaním rakoviny na dobu prežitia

Taktiež fajčenie pred samotnou diagnostikou choroby vyšlo v Coxovom modeli s jednou vysvetľujúcou premennou významné. Práve z toho dôvodu si i túto premennú bližšie zanalyzujeme. Na jednotlivé početnosti v skupinách sa môžeme pozrieť v Tabuľke 5.12.

fajčenie	celkový počet	výskyt udalosti	cenzurovaný pacienti	percento cenzúrovaných pacientov v skupine
áno	78	38	40	51.28 %
nie	19	5	14	73.68 %

Tabuľka 5.12: Rozloženie v skupinách podľa fajčenia pred diagnózou.

Tentokrát je rozloženie v skupinách nevyrovnané, avšak v percentuálnom zastúpení vidíme, že väčšie percento pacientov bolo cenzúrovaných v prípade, kedy pacienti pred diagnostikovaním rakoviny nefajčili. Funkcie prežitia vytvorené Kaplan-Meierovou a Coxovou metódou sú v Obr. 5.4.



Obr. 5.4: Funkcie prežitia podľa fajčenia pred dignostikovaním rakoviny.

V tomto prípade vidíme trochu odlišné odhady funkcií prežitia. V prípade Kaplan-Meierovho odhadu sa ku koncu funkcie približujú. Taktiež pre skupinu pacientov, ktorí pred diagnostikou nefajčili, sú funkcie prežitia v oboch prípadoch trochu odlišné. To, či vyjde významný rozdiel v Kaplan-Meierovej funkcii prežitia, testujeme pomocou log-rank testu.

	Chi-sq	df	p-hodnota
Mantel-Heanszel test (Log-rank)	4.7	1	0.03

Tabuľka 5.13: Test rovnosti funkcií prežitia na základe fajčenia.

I tentokrát z výsledku vidíme, že rozdiel medzi funkciami prežitia je významný. Ďalej sa pozrieme na výsledok Coxovho modelu.

	β	$se(\beta)$	p-hod	$exp(\beta)$	testová štatistika	lower 95% KI	upper 95% KI
FAJČENIE=2	-1.0105	0.4822	0.0361	0.3640	4.39	0.1415	0.9367

Hodnota regresného parametru je záporná, čo spôsobuje pokles rizika výskytu udalosti a tým sa zvyšuje doba prežitia pre pacientov, ktorí pred diagnostikovaním rakoviny nefajčili. Pre túto skupinu je 0.364krát nižšie riziko úmrtia.

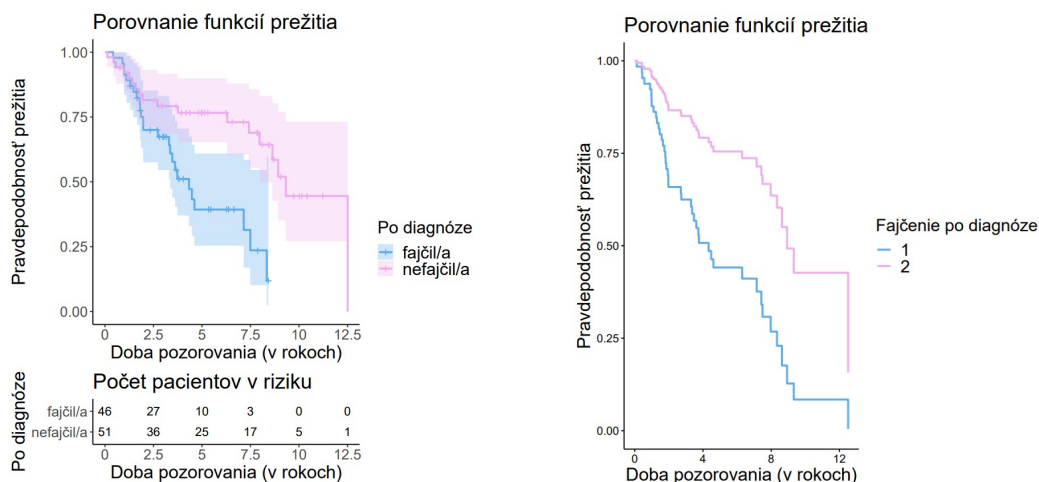
Vplyv fajčenia po diagnostikovaní rakoviny na dobu prežitia

Ďalšou významnou premennou, ktorú bližšie zanalyzujeme, je premenná obsahujúca údaj či pacienti po zistení diagnózy fajčili alebo nie. Opäť máme približne vyrovnané skupiny, a ich početnosti sú v nasledujúcej tabuľke.

fajčenie	celkový počet	výskyt udalosti	cenzurovaný pacienti	percento cenzúrovaných pacientov v skupine
áno	46	25	21	45.65%
nie	51	18	33	64.71%

Tabuľka 5.14: Rozloženie v skupinách podľa fajčenia po diagnóze.

V tomto prípade môžeme vidieť väčší počet pacientov v skupine, ktorí po diagnóze nefajčili, a taktiež aj menší počet u ktorých nastala smrť. Odhadnuté funkcie prežitia Kaplan-Meierovou metódou sú v Obr. 5.5a. Vedľa, v Obr. 5.5b, sú odhadnuté funkcie prežitia pomocou Coxovho modelu.



(a) Kaplan-Meierov model.

(b) Coxov model.

Obr. 5.5: Funkcie prežitia podľa fajčenia po diagnóze rakoviny.

V prípade Kaplan-Meierovho modelu sa od začiatku funkcie prežitia mierne krížili, no postupujúcim časom medzi nimi narastal rozdiel. Taktiež si môžeme všimnúť, že i 95% intervaly spoľahlivosti sa prekrývajú len minimálne, a opäť je čas prežitia pre jednu skupinu kratší. V Coxovom modeli je hneď narastajúci rozdiel medzi odhadnutými funkciami, a sú rovnako dlhé. Je to z toho dôvodu, že Coxov model má predpoklad

proporcionality hazardu, čo znamená, že riziko pacientov je vzájomne proporcionálne v každom časovom okamihu. To si môžeme všimnúť v Obr. 5.5b, kde výskyt udalosti (schod) nastane v oboch funkciách v rovnaký čas. Preto je tento model vhodný v situácii, kedy máme nevyvážený počet pozorovaní v skupinách. I tu vidíme, že funkcia prežitia pre pacientov, ktorí po diagnostike fajčili klesá rýchlejšie, a pravdepodobnosti prežitia sú nižšie v jednotlivých časových okamihoch ako v opačnej skupine. Štatistickú významnosť overíme taktiež pre Kaplan-Meierove funkcie prežitia.

	Chi-sq	df	p -hodnota
Mantel-Heanszel test (Log-rank)	10.3	1	0.001

Tabuľka 5.15: Test rovnosti funkcií prežitia na základe fajčenia.

Z testovania rovnosti odhadnutých funkcií prežitia môžeme na základe nízkej p - hodnoty povedať, že funkcie prežitia sú i v tomto prípade rozdielne. Na výsledok sa pozrieme i v Coxovom modeli.

	$\hat{\beta}$	$se(\hat{\beta})$	p -hod	$\exp(\hat{\beta})$	testová štatistika	lower 95% KI	upper 95% KI
FAJČENIEa=2	-1.0679	0.3454	0.002	0.3437	9.56	0.1747	0.6765

Opäť máme pri regresnom parametre $\hat{\beta}$ zápornú hodnotu, a teda pôjde o pokles rizika výskytu udalosti pre pacientov, ktorí po diagnostike rakoviny nefajčili, oproti tým, čo s fajčením pokračovali. Hodnota hazard ratio je rovná 0.3454, a preto ide o 0.35krát nižšie riziko úmrtia pre pacientov, ktorí po dignostike rakoviny nefajčili.

Keďže sme si prešli vysvetľujúce premenné, ktoré v Coxovom modeli s jednou premennou vyšli významné, môžeme sa pozrieť i na premenné, ktoré nevyšli významné. Konkrétne sa vykreslíme funkcie prežitia pre premenné pohlavie a konzumácia alkoholu po diagnostikovaní rakoviny.

Vplyv pohlavia na dĺžku prežitia

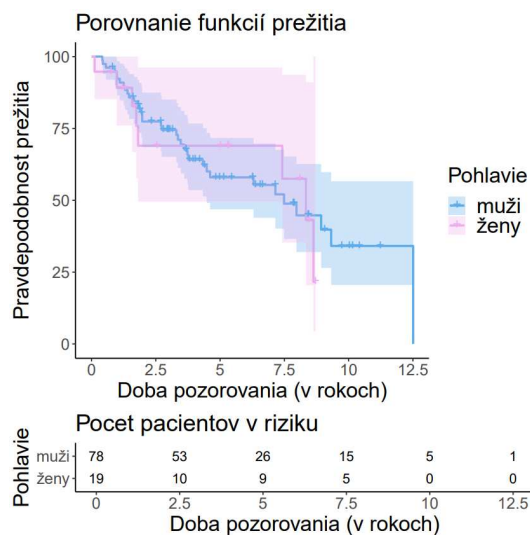
Prvé porovnanie, ktoré nás v každom prípade napadne, je podľa pohlavia. I keď je zastúpenie oboch pohlaví nevyrovnané, mohlo by mať vplyv na výskyt rakoviny a jej

priebeh. V Tabuľke 5.16 sú rozpísané počty pacientov, rozdelených do skupín podľa výskytu udalosti a cenzúry.

pohlavie	celkový počet	výskyt udalosti	cenzurovaný pacienti	percento cenzúrovaných pacientov v skupine
muži	78	35	43	55.13%
ženy	19	8	11	57.89%

Tabuľka 5.16: Rozloženie v skupinách podľa pohlavia.

Graf pre odhadnuté funkcie prežitia je znázornený v Obr. 5.6, kde je i zjednodušená tabuľka opisujúca vývoj funkcie prežitia. Taktiež vidíme aj 95% intervaly spoľahlivosti. Funkcie nie sú príliš odlišné, často sa prekrývajú, a jediný badateľný rozdiel je, že skupina mužov má dlhšiu dobu pozorovania, to však môže byť z toho dôvodu, že počet mužov v skupine pacientov prevláda.



Obr. 5.6: Kaplan-Meierove funkcie prežitia podľa pohlavia.

Významnosť rozdielu medzi funkciami prežitia overíme testovaním hypotézy o rovnosti odhadnutých funkcií.

	Chi-sq	df	p-hodnota
Mantel-Heanszel test (Log-rank)	0	1	0.9

Tabuľka 5.17: Test rovnosti funkcií prežitia na základe pohlavia.

Z testu je v tabuľke uvedená testová štatistika, počet stupňov voľnosti a následne p -hodnota, na základe ktorej test vyhodnocujeme. Keďže je p -hodnota=0.9, usudzujeme, že nie je signifikantný rozdiel v dĺžke prežitia na základe pohlavia.

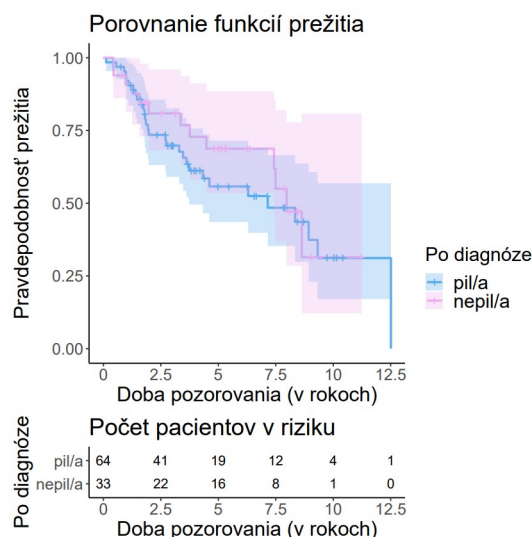
Vplyv konzumácie alkoholu po diagnostike rakoviny

Keďže sa potvrdil rozdiel medzi skupinou fajčiarov a nefajčiarov, ďalší faktor ovplyvňujúci dobu prežitia by mohol byť aj konzumácia alkoholu po zistení rakoviny. Taktiež sa najskôr pozrieme na jednotlivé početnosti v skupinách, počet výskytov udalosti, a počet cenzurovaných pacientov.

alkohol	celkový počet	výskyt udalosti	cenzurovaný pacienti	percento cenzúrovaných pacientov v skupine
áno	64	30	34	53.14%
nie	33	13	20	60.60%

Tabuľka 5.18: Rozloženie v skupinách podľa konzumácie alkoholu.

Tentokrát máme nevyvážené zastúpenie pacientov v skupinách, a počet pacientov, ktorí po diagnostikovaní rakoviny pili alkohol je takmer dvakrát tak veľký ako počet pacientov, ktorí po diagnostikovaní rakoviny nepili alkohol. Väčšia skupina pacientov ostala pri konzumácii alkoholu, a v tejto skupine nastala udalosť u tridsiatich pacientov (46.86%), v prípade kedy pacienti nepili alkohol zomrelo trinásť z nich (39.4%).



Obr. 5.7: Kaplan-Meierove funkcie prežitia podľa alkoholu po diagnóze rakoviny.

Medzi odhadnutými funkciami nie je narastajúci rozdiel. Až do približne siedmeho roku vidíme vyššie hodnoty pravdepodobnosti prežitia u skupiny pacientov, ktorí alkohol nepili. Avšak ku konci doby pozorovania sú funkcie prežitia podobné. Taktiež ich intervaly spoľahlivosti sa prekrývajú.

	Chi-sq	df	<i>p</i> -hodnota
Mantel-Heanszel test (Log-rank)	0.4	1	0.5

Tabuľka 5.19: Test rovnosti funkcií prežitia na základe konzumácie alkoholu.

Tentokrát vysoká *p*-hodnota značí, že nieje významný rozdiel medzi rozdielnymi skupinami pacientov podľa konzumácie alkoholu.

5.3.4. Optimálny Coxov model

V teoretickej časti bolo spomenuté, že použitie Coxovho modelu oproti Kaplan-Meierovej metóde má tú výhodu, že vieme do modelu zahrnúť i viac vstupných premenných naraz. V predchádzajúcom prípade sme si ukázali významnosť jednotlivých premenných v modeli. Tentokrát bude úlohou vytvoriť optimálny Coxov model, v ktorom bude zahrnutý väčší počet vstupných premenných, ktoré nám poskytnú viac informácií týkajúcich sa doby prežitia vzhľadom ku charakteristikám pacientov.

Analýzou vstupných premenných a ich interakcií bol vytvorený Coxov model, kde boli ponechané len významné premenné, resp. interakcie. Nájdenný optimálny Coxov model je uvedený v Tabuľke 5.20.

	$\hat{\beta}$	$se(\hat{\beta})$	<i>p</i> -hodnota	HR	lower 95% KI	upper 95% KI
HPV.RNA=áno	0.121	0.516	0.815	1.12	0.41	3.10
FAJČENIEa=nie	0.174	0.537	0.746	1.19	0.42	3.41
BUNKY	-0.013	0.005	0.019	0.99	0.97	0.99
VEK	-0.021	0.018	0.239	0.98	0.95	1.01
HPV.RNA=áno:FAJČENIEa=nie	-2.097	0.838	0.012	0.12	0.02	0.64
bunky:VEK	-0.0005	0.0002	0.015	0.99	0.99	0.99

Tabuľka 5.20: Optimálny Coxov model

V tabuľke sú pre premenné uvedené odhady parametru β , smerodatná odchylka pre $\hat{\beta}$, *p*-hodnota, hazard ratio ($exp(\hat{\beta})$), a 95% interval spoľahlivosti pre hazard ratio.

Znovu podľa p -hodnoty vieme určiť významnosť jednotlivých premenných.

Nasledujúce interpretácie výsledkov optimálneho Coxovho modelu uvažujeme pri fixných hodnotách ostatných premenných uvedených v modeli.

Významný interakčný člen vyšiel medzi vírusovým pôvodom rakoviny a fajčením po diagnostike rakoviny. V rizikovej funkcii sa teda bude vyskytovať vplyv premenných HPV.RNA, FAJČENIEa a ich interakcia. Táto závislosť je v modeli popísaná pomocou nasledujúcich členov:

$$\beta_1 * I_{[HPV.RNA=1]} + \beta_2 * I_{[FAJČENIE=2]} + \beta_3 * I_{[HPV.RNA=1]} * I_{[FAJČENIE=2]},$$

kde $I_{[podmienka]}$ je indikátorová funkcia, ktorá nadobúda hodnotu 1, ak je podmienka splnená, a hodnotu 0, ak nieje podmienka splnená. Po dosadení odhadov β_1 , β_2 a β_3 dostávame

$$0.121 * I_{[HPV.RNA=1]} + 0.174 * I_{[FAJČENIE=2]} - 2.09 * I_{[HPV.RNA=1]} * I_{[FAJČENIE=2]}.$$

Tento súčin je rovný:

- 0 pokiaľ má pacient nevírusový pôvod a fačí,
- 0.121 pokiaľ má pacient vírusový pôvod a fajčí,
- 0.174 pokiaľ má pacient nevírusový pôvod a nefajčí
- -1.795 pokiaľ má pacient vírusový pôvod a nefajčí.

Riziká týchto prípadov môžeme porovnať nasledujúcim spôsobom. Prvý prípad znázorňuje efekt pôvodu choroby vzhľadom na skupiny fajčiarov a nefajčiarov.

1a Uvažujeme pacientov ktorí fajčia. Tí, ktorí majú vírusový pôvod, majú podľa modelu $e^{0.121}=1.12$ krát vyššie riziko výskytu udalosti, oproti pacientom, ktorí majú nevírusový pôvod. Rozdiel však nie je signifikantný. Riziko teda môžeme považovať za rovnaké v oboch skupinách.

1b V tomto prípade uvažujeme skupinu nefajčiarov. Tí, ktorí majú vírusový pôvod majú nižšie riziko úmrtia $e^{0.174-2.09}=0.15$ krát.

2a Pacienti, ktorí majú nevírusový pôvod a nefajčia, majú $e^{0.175}=1.19$ krát vyššie riziko úmrtia oproti pacientom, ktorí majú taktiež nevírusový pôvod, ale fajčia.

I v tomto prípade rozdiel nieje významný, a teda riziko v oboch skupinách môžeme považovať za rovnaké.

- 2b Pacienti s vírusovým pôvodom rakoviny, ktorí nefajčia, majú $e^{0.121-2.09}=0.14$ krát nižšie riziko výskytu udalosti oproti pacientom, ktorí majú vírusový pôvod rakoviny a fajčia.

Druhá významná interakcia je medzi počtom buniek vo vzorku a vekom pacienta.

Ich závislosť je v modeli popísaná nasledovne:

$$\beta_4 * \text{BUNKY} + \beta_5 * \text{VEK} + \beta_6 * \text{BUNKY} * \text{VEK},$$

kde tentokrát za BUNKY a VEK dosádzame konkrétne hodnoty. Po dosadení odhadov

$\beta_4, \beta_5, \beta_6$ dostávame

$$-0.013 * \text{BUNKY} - 0.021 * \text{VEK} - 0.0005 * \text{BUNKY} * \text{VEK}.$$

Odhad $\beta_1 = -0.013$ udáva závislosť doby prežitia na počte buniek, ak je vek rovný jeho priemernej hodnote (57 rokov), nakoľko sme veličinu VEK na začiatku analýzy pre lepšiu interpretáciu centrovali. β_2 udáva závislosť doby prežitia na veku, ak je počet buniek rovný priemernej hodnote 40. Taktiež bola veličina BUNKY centrovaná pre lepšiu interpretáciu. Analogicky môžeme fixovať hodnotu veku na akejkoľvek hodnote a dostaneme závislosť na počte buniek, avšak regresný koeficient sa bude meniť podľa hodnoty zafixovanej vo veku. Napríklad pre vek=30 dostávame

$$-0.013 * \text{bunky} - 0.021 * 30 - 0.0005 * \text{bunky} * 30 = (-0.013 - 0.015) * \text{bunky} - 0.63,$$

regresný parameter je tak rovný -0.018, čo odpovedá zníženiu rizika $\exp(-0.018) = 0.98$ krát pri jednotkovej zmene počtu buniek u pacientov s vekom 30 rokov.

Môžeme však zafixovať i počet buniek, napr. 10, a skúmať rozdiely podľa veku.

$-0.013 * 10 - 0.021 * \text{VEK} - 0.0005 * 10 * \text{VEK} = (-0.021 - 0.005) * \text{VEK} - 0.13$, a regresný parameter je tak rovný -0.026, čo odpovedá zníženiu rizika $\exp(-0.026) = 0.97$ krát pri jednotkovej zmene veku u pacientov s počtom buniek vo vzorku 10.

Záver

Cieľom bakalárskej práce bolo zoznámiť sa s teóriou analýzy prežívania, jej praktickým využitím, a taktiež použitie v reálnom prípade. Využitie analýzy prežívania sme si ukázali najmä v praktickej časti, kde bolo hlavnou úlohou porovnať funkcie prežitia a nájsť faktory, ktoré by mohli ovplyvňovať dobu prežitia pacientov.

V prvej časti sme sa zoznámili s teóriou analýzy prežívania. Prvá kapitola obsahuje teoretický základ a vysvetlenie základných pojmov, ktoré sú názorne ukázané i pomocou krátkych príkladov a obrázkov. Tieto pojmy sú nevyhnutné na pochopenie témy a následnú prácu. V ďalšej kapitole sme sa zoznámili s rôznymi typmi odhadov funkcie prežívania, a príslušnými intervalmi spoľahlivosti. Následne sme si uviedli testy, pomocou ktorých funkcie prežívania porovnávame, a tým zisťujeme, či je rozdiel medzi funkciami štatisticky významný. V poslednej kapitole teoretickej časti sme si predstavili i semiparametrický prístup modelovania funkcie prežitia, a to Coxov model s jeho charakteristikou.

V druhej časti bakalárskej práce sme pracovali s reálnou situáciou, kedy využívame analýzu prežívania. Najskôr sme si predstavili dátovú sadu, ktorá obsahovala údaje o pacientoch s druhom rakoviny. Zastúpenie jednotlivých kategórií a rozdelenie pacientov sme si mohli podrobne prezrieť v kontingenčných tabuľkách. Následne bola odhadnutá základná funkcia prežitia pre všetkých pacientov pomocou Kaplan-Meierovej metódy. Taktiež sme si ukázali odhady funkcie prežitia pomocou Coxovho modelu, na základe ktorého sme určili významnosť jednotlivých vstupných premenných.

Z významných vstupných premenných boli vytvorené odhady funkcií prežitia pomocou Kaplan-Meierovej metódy i Coxovým modelom. Následne sme v grafoch videli rozdiely medzi použitými metódami. Ďalej sme odhadli zmenu rizika výskytu udalosti pre rôzne skupiny pacientov rozdelených podľa vstupnej premennej. Na záver bol vytvorený optimálny Coxov model, do ktorého boli zahrnuté vysvetľujúce premenné,

ktoré vzájomne najviac prispievajú k odhadu funkcie prežitia. Ukázalo sa, že premenné, ktoré najviac vplyvajú na dobu prežitia, sú počet buniek určitého typu vo vzorku krvi, vek, vírusový pôvod rakoviny a to, či pacient po diagnostikovaní rakoviny fajčil, alebo nie. Na dobu prežitia pacientov môžu pôsobiť i iné faktory, ktoré však v tejto práci neboli bližšie analyzované.

Prínos práce spočíva nielen v predstavení teoretického základu analýzy prežívania, ale aj ukážke jej praktického využitia v oblasti medicíny. Avšak nakoľko je táto štatistická metóda všestranná, je možné ju použiť i v akomkoľvek inom odvetví. Verím, že tak ako mne, tak aj čitateľovi táto práca predstavila ako veľmi je prínosné používať práve analýzu prežívania, a aké zaujímavé, niekedy aj kľúčové informácie môže poskytnúť.

Literatúra

- [1] PROCHÁZKA, Bohumír: *Biostatistika pro lékaře: principy základních metod a jejich interpretace s využitím statistického systému R*. V Praze: Univerzita Karlova v Praze, nakladatelství Karolinum, 2015. ISBN 978-80-246-2782-3.
- [2] HOLČÍK, Jiří, KOMENDA, Martin (eds.) a kol. *Matematická biologie: Matematická biologie e-learningová učebnice [online]. 1. vydání*. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-8095-9. <https://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat-aplikovana-analyza-preziti>
- [3] KLEINBAUM, David G. a Mitchel KLEIN.: *Survival analysis: a self-learning text*, 2005 ISBN 0387239189.
- [4] HOSMER, David W. a Stanley LEMESHOW: *Applied survival analysis: regression modeling of time to event data*. New York: Wiley, c1999. ISBN 0471154105.
- [5] T.LEE, Elisa, WANG, John Wenyu.: *Statistical Methods for Survival Data Analysis. Third Edition*. New Jersey, 2003 ISBN 9780471369974
- [6] MOORE, Dirk F.: *Applied Survival Analysis Using R*. USA ISBN 978-3-319-31243-9
- [7] MILLER, Rupert G., Gail GONG a Alvaro MUNOZ. *Survival analysis*. New York: Wiley, c1981, xi, 238 s. *Wiley classics library*. ISBN 0-471-25548-3.
- [8] CHOW, Shein-Chung a Jen-pei LIU. *Design and analysis of clinical trials: concepts and methodologies*. 2nd ed. Hoboken: John Wiley, c2004, xiii, 729 s. *Wiley series in probability and mathematical statistics*. ISBN 0-471-24985-8.
- [9] D Kurniasari et al 2019 J. Phys.: Conf. Ser. 1338 012036 [online]
- [10] NIST/SEMATECH e-Handbook of Statistical Methods. <https://www.itl.nist.gov/div898/handbook/eda/section3/eda3669.htm>
- [11] LE, Chap T. a Lynn E. EBERLY. *Introductory biostatistics. Second edition*. Hoboken, New Jersey: John Wiley & Sons, 2016. ISBN 9780470905401.
- [12] *Test equality of survivor functions* <https://www.stata.com/manuals13/stststest.pdf> [online]

- [13] KLEIN, John P. *Handbook of survival analysis / edited by John P. Klein ... [et al.]*. 2014. ISBN 9781466555662.
- [14] Cox, D.R. *Regression Models and Life Tables (with Discussion)*. *Journal of the Royal Statistical Society: Series B*, 34, 187-220, (1972)

Dodatok A

time	n.risk	n.event	n.censor	estimate	std.error	conf.high	conf.low
0.12	97	1	0	0.9896907	0.0103628	1.0000000	0.9697921
0.42	96	1	0	0.9793814	0.0147322	1.0000000	0.9515066
0.98	90	2	0	0.9271553	0.0286013	0.9806137	0.8766112
1.10	87	1	0	0.9060836	0.0328984	0.9664323	0.8495033
1.46	81	1	0	0.8631844	0.0408738	0.9351805	0.7967310
1.74	75	1	0	0.8301069	0.0466900	0.9096549	0.7575152
1.81	72	1	0	0.8075157	0.0506033	0.8917120	0.7312694
1.97	66	2	0	0.7605083	0.0588278	0.8534504	0.6776878
2.33	64	0	1	0.7605083	0.0588278	0.8534504	0.6776878
2.70	61	1	0	0.7480410	0.0611059	0.8432160	0.6636085
2.94	58	0	1	0.7355736	0.0633752	0.8328585	0.6496524
3.04	55	0	1	0.7355736	0.0633752	0.8328585	0.6496524
3.15	54	0	1	0.7355736	0.0633752	0.8328585	0.6496524
3.61	50	1	0	0.6800586	0.0745430	0.7870412	0.5876182
3.76	46	1	0	0.6511200	0.0806367	0.7626041	0.5559336
3.95	44	0	1	0.6511200	0.0806367	0.7626041	0.5559336
4.03	43	0	1	0.6511200	0.0806367	0.7626041	0.5559336
4.39	40	0	1	0.6352390	0.0843329	0.7494134	0.5384592
4.48	39	1	0	0.6189508	0.0882429	0.7358150	0.5206474
5.02	35	0	1	0.6026626	0.0921848	0.7220082	0.5030445
5.45	31	0	1	0.6026626	0.0921848	0.7220082	0.5030445
6.26	30	0	1	0.6026626	0.0921848	0.7220082	0.5030445
6.65	25	0	1	0.5811390	0.0990999	0.7057226	0.4785485
7.14	24	0	1	0.5811390	0.0990999	0.7057226	0.4785485
7.87	19	0	1	0.5053382	0.1278379	0.6492291	0.3933384
8.11	16	0	1	0.4756125	0.1414885	0.6276080	0.3604275
8.47	12	0	1	0.4439050	0.1574195	0.6043461	0.3260575
8.63	11	1	0	0.4035500	0.1840430	0.5788353	0.2813453
8.93	9	1	0	0.3587111	0.2185422	0.5505139	0.2337337
9.07	8	0	1	0.3587111	0.2185422	0.5505139	0.2337337
9.73	6	0	1	0.3074666	0.2675261	0.5194176	0.1820033
10.03	5	0	1	0.3074666	0.2675261	0.5194176	0.1820033
10.42	3	0	1	0.3074666	0.2675261	0.5194176	0.1820033
11.23	2	0	1	0.3074666	0.2675261	0.5194176	0.1820033
12.51	1	1	0	0.0000000	Inf	NA	NA

Tabuľka A.1: Rozšírená tabuľka opisujúca funkciu prežitia v Obr. 5.2