Univerzita Pardubice

Fakulta ekonomicko-správní

Modelování úmrtnosti v závislosti na rizikových faktorech

Bc. Kateřina Hájková

Diplomová práce

2009

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: 2008/2009

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Kateřina HÁJKOVÁ**

Studijní program: **N6209 Systémové inženýrství a informatika**

Studijní obor: **Pojistné inženýrství**

Název tématu: **Modelování úmrtnosti v závislosti na rizikových faktorech**

Zásady pro vypracování:

Význam modelování úmrtnosti v životním a důchodovém pojištění.
Kvantifikace vlivu rizikových faktorů v analýze přežití.
Logistický regresní model jako model přežití.
Kaplan Meierovy křivky přežití.
Coxův regresní model úmrtnosti.
Ukázka aplikace modelů úmrtnosti pomocí statistického programového balíku Statgraphics Centurion.

Rozsah grafických prací:
Rozsah pracovní zprávy:
Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:

BENJAMIN, B. – POLLARD, S. H.: The Analysis of Mortality and other actuarial statistics. Institute of Actuaries, Scotland, London 1993

CIPRA, T.: Pojistná matematika – teorie a praxe. Praha: EKOPRESS, Vydání II, 2006.

FIALA, T.: Výpočty aktuárské demografie v tabulkovém procesoru. Praha: Vysoká škola ekonomická, 2005.

Kleinbaum, D. G.: Logistic Regression: A Self – Learning Text. Springer-Verlag, New York, 1996.

Kleinbaum, D. G.: Survival analysis: A Self – Learning Text. Springer-Verlag, New York, 1996.

KOSCHIN, F.: Aktuárská demografie. Vysoká škola ekonomická v Praze, 2002.

SEKERKA, B. Matematické a statistické metody ve financování, cenných papírech a pojišťovnictví. Praha: Profess Consulting, 2002

Vedoucí diplomové práce: prof. RNDr. Viera Pacáková, Ph.D.
Ústav ekonomie

Datum zadání diplomové práce: 6. října 2008
Termín odevzdání diplomové práce: 1. května 2009

doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.

doc. Ing. Jiří Křupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 6. října 2008

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích dne 21. 4. 2009

Kateřina Hájková

Ráda bych touto cestou poděkovala vedoucí práce prof. RNDr. Viere Pacákové, Ph. D. za její odborné vedení, cenné připomínky, rady a vstřícnost.

Anotace

Tato diplomová práce se zabývá možností kvantifikace faktorů úmrtnosti. První kapitola se věnuje významu pojištění a jeho vývoji od počátků po současnost. Ve druhé kapitole je charakterizován význam a modelování úmrtnosti. Třetí, a to zásadní kapitola zpracovává logistický regresní model jakožto model přežití. V poslední kapitole je popsána analýza přežití, jsou zde definovány různé metody odhadu křivky přežití a dále Coxův regresní model úmrtnosti.

Klíčová slova

modelování úmrtnosti, úmrtnostní tabulky, logistický regresní model, Kaplan Meierovy křivky přežití, Coxův regresní model úmrtnosti

Title

Mortality rates modelling in dependence on risk elements

Annotation

This thesis deals with the possibility of quantifying the factors of mortality. The first chapter examines the importance of insurance and its development from its beginnings to the present. The second chapter is characterized by the importance and modeling of mortality. Third, a major chapter deals with logistic regression model as a model of survival. The last chapter describes the analysis of survival, there are defined the different estimation methods of survival curves and Cox regression model of mortality.

Keywords

mortality rates modelling, life tables, logistic regression model, Kaplan Meier survival curves, Cox regression model of survival

OBSAH

ÚVOD

Současná společnost se nachází v období transformace hospodářství a zásadních ekonomických reforem, které jsou doprovázené různými ekonomickými, politickými a kulturními změnami [1].

Pro osoby i podnikatelské subjekty vzniká spousta nových rizik, ale i nových možností, jak eliminovat negativní následky těchto rizik. Roste také význam komerčního pojištění při krytí rizik jako způsob uplatnění zodpovědnosti jednotlivých ekonomických subjektů na řešení následků náhodných událostí s nepříznivým ekonomickým dopadem [1].

V zemích Evropské Unie, ale i v ostatních zemích, které mají vyspělou tržní ekonomiku, se pojištění všeobecně člení na životní a neživotní. Do pojištění životního patří pojištění osob a jeho význam roste jak v návaznosti na stabilizaci a růst životní úrovně, tak i v souvislosti se stárnutím obyvatelstva [1].

Lidé berou životní pojištění jako zabezpečení sebe nebo své rodiny v nelehké životní situaci, stále však v České republice přetrvává chápání pojištění spíše jako investičního nástroje pro přilepšení si k důchodu místo nástroje na skutečné pojištění – tedy krytí rizik [2].

Pojištění jako zabezpečení rodiny v případě úrazu nebo smrti je vnímáno spíše lidmi z větších měst, naopak v menších obcích přetrvává pocit, že životní pojištění je nástrojem pro zajištění se na důchod [2].

Podle České asociace pojišťoven zahrnuje životní pojištění tyto základní druhy [18]:

- investiční životní pojištění,
- kapitálové životní pojištění,
- rizikové životní pojištění,
- důchodové pojištění,
- vkladové pojištění,
- univerzální (flexibilní pojištění),
- pojištění ve prospěch dítěte.

Pojistnou ochranu lze i rozšířit formou připojištění (tedy doplňkových pojištění) pro případ úrazu, vážných chorob, invalidity, zproštění od placení, pro případ nemoci – denní dávky při pracovní neschopnosti nebo pro případ pobytu v nemocnici [18].

I když jsou všichni lidé smrtelní, je okamžik, kdy smrt nastane, neznámý. Vzhledem k této skutečnosti je smrt náhodný jev, na který se vztahuje pojištění. Pro konkrétního pojištěného jedince je délka života neznámá, proto pojišťovna pro žádného ze svých pojištěnců nemůže přesně určit, jestli a kdy bude vyplaceno pojistné plnění, ke kterému se v případě smrti pojištěného jedince zavázala. V pojištění osob existují výpočty založené na určitém modelu platném pro velké soubory osob, které při správném použití zaručují splnitelnost závazků, které pojišťovna na sebe převzala. Tento postup je zjištěný dlouhodobými zkušenostmi a s výjimkou určitých věkových kategorií vyjadřuje kvantitativní zkušenost, že člověk je tím blíže smrti, čím je starší. Jinak řečeno, s rostoucím věkem se zvyšuje pravděpodobnost úmrtí [1].

Znalost měr úmrtnosti (latinsky mortality) v závislosti na věku a pohlaví, je velmi důležitá pro životní pojištění. Kromě toho se berou v úvahu i další faktory, jako je například sociální příslušnost, zaměstnání, geografická lokalizace, domácí prostředí, klimatické podmínky, sportovní aktivity, osobní návyky jako je kouření, konzumace alkoholu a jiné [7].

Zkušenosti životní pojišťovny se opírají o dlouhodobá pozorování a na základě nich se odhaduje budoucí průběh úmrtnosti. Tyto zkušenosti by se měly porovnávat se standartními úmrtnostními tabulkami. Kromě uvedených rizikových faktorů mají vliv na předpoklady úmrtnosti i možné změny vývoje úmrtnosti a jejich faktorů v budoucnu. Možností kvantifikace faktorů úmrtnosti se zabývá i tato diplomová práce. Cílem tedy je kvantifikovat vliv nejen základních, ale i dalších rizikových faktorů v analýze přežití, popsat a použít vhodné metody této kvantifikace. Pro jejich aplikaci použít vhodný programový balík.

1 VÝZNAM A HISTORIE POJIŠTĚNÍ

Následující kapitola 1 je zpracována dle [3], [4], [19], [20].

Pojištění patří mezi finanční služby. Předmětem této finanční služby je poskytnutí pojistné ochrany za úplatu. Pojištění je součástí infrastruktury ekonomiky a působí na průběh procesu reprodukce tak, že část finančních prostředků přesouvá na ta místa, kde jsou v určitém okamžiku zapotřebí (z hlediska výskytu nahodilých potřeb).

Z hlediska finanční kategorie pojištění reprezentuje tvorbu, rozdělování a užití pojistného fondu k uhrazení finančních potřeb ekonomických subjektů, které jsou náhodně vcelku odhadnutelné v jednotlivých případech výskytu.

Pojištění můžeme též zařadit do důchodové kategorie a to proto, že prostřednictvím něj dochází k důchodové stabilizaci ekonomických subjektů.

A nakonec z právního hlediska představuje pojištění právní vztah, ve kterém na sebe pojistitel přebírá závazek, že poskytne pojistné plnění pojištěnému v případě, že nastane nahodilá událost, která je blíže specifikována v pojistných podmínkách.

Člověk byl odpradávna sužován strachem, obavami i nejistotou. Proto vzniklo pojištění, které zabezpečuje pojištěným právo na výplatu peněžních prostředků k úhradě potřeb, které vznikají z negativních nahodilých událostí.

Úvahy nějakým způsobem se vyrovnávat s nahodilými událostmi jako jsou živelné katastrofy a osobní neštěstí, ke kterému patřily především úrazy a ztráta živitele, vznikaly již v hlubokém dávnověku.

První zmínky se datovaly již kolem roku 2 000 před n. l. a byly spojeny hlavně se vzájemným krytím ztrát (vzájemnostní pojištění), zejména v souvislosti s krytím výdajů na pohřby a taktéž v souvislosti s přepravou zboží (krytí ztrát mezi majiteli karavan při plánované cestě). U těchto prvních pojištění lze shledat tyto rysy:

- pojištění bylo koncentrováno především v uzavřených skupinách obyvatel,
- pojištění zahrnovalo hlavně řemeslníky a kupce, ale ne zemědělce, kteří počtem v té době převažovali,
- při pojišťování nebyl plně oddělen pojistitel a pojistník,

- pojištění mělo většinou vzájemnostní charakter, ale lze zde pozorovat i určité prvky komerčního pojištění (například u tzv. námořní půjčky – lze si ji představit jako kombinaci pojištění a úvěru, poněvadž obchodník si vzal při přepravě zboží námořní půjčku, která se rovnala výši ceny zboží, a pokud loď s nákladem dosáhla místa určení, obchodník půjčku vrátil, ale s vysokými úroky (až 36 %), a pokud loď do cíle nedoplula, půjčku nevracel).

Od této doby prošlo pojištění dlouhodobým vývojem, ve kterém lze pozorovat dvě základní tendence: vývoj od všeobecně formulované vzájemné pomoci k určité konkretizaci vzájemné pomoci s upřesněním okruhu pojistných událostí a vývoj od následného rozvrhu výdajů na pojistná plnění k praxi pravidelných pevně stanovených příspěvků.

Zárodky jsou dochovány ze starověku u kulturních národů, kde vznikají různorodá sociální zařízení, která lze považovat za předchůdce majetkových a životních pojištění. První zmínky dochované jako záznamy na hliněných destičkách o sdružení majitelů velbloudů jdou datovány přibližně z doby 2000 let před naším letopočtem. Tato sdružení chránila své členy proti ztrátám, které souvisely s provozem obchodu a dopravy pomocí karavan. Dále například staří Féničané vymysleli jakési dopravní pojištění lodí a nákladů.

Za další zmínky o pojišťovnictví můžeme považovat cechy, které představovaly hospodářská zřízení středověké Evropy. Z nich vznikly řemeslnické cechy, které si udržely své postavení až do 18. století. Cechy zakládaly četné mistrovské a tovaryšské pokladnice a bratrstva a tím byly opět dány základy pojišťování, poskytovaly tedy nejen právní ochranu, ale i provozovaly do určité míry požární pojištění a pojištění lodní dopravy.

Jako nejstarší pojistná smlouva z roku 1385 je dochována námořní pojistka vydaná notářem v Pise. První řádnou požární pojišťovnou, která vznikla v roce 1676 v Hamburku, byla Generální požární pokladna. V roce 1687 vzniká z kavárny Edvarda Lloyda, ve které se scházeli kupci, makléři a námořníci, největší pojišťovací systém LLOYDS. Na tomto základě vybudoval Lloyd největší informační centrum o pohybu lidí, zboží, o haváriích a jejich pojišťování. Takto vznikla největší pojišťovací „burza“ ve světě.

Historie pojišťovnictví na území současné České republiky se datuje od konce 17. století. V roce 1699 předložil Jan Kryštof Bořek promyšlený návrh na zřízení povinného požárního pojištění budov v Čechách. Tento návrh tkvěl v zavedení protipožárního fondu, do kterého by povinně přispívali všichni občané, kteří by si právě pořídili dům. Tento fond nakonec ale nebyl realizován. Roku 1777 byla zřízena pojišťovna proti škodám způsobených ohněm na polních zásobách, nábytku, nářadí a dobytku. Brzy však zanikla. Po roce 1822 zahájily na českém území činnost dvě „zahraniční“ pojišťovny sídlící ve Vídni a Terstu.

Císařskokrálovský privilegovaný, český, společný, náhradu škody ohněm svedené pojišťující ústav vznikl v roce 1827. Od něj se odvíjí tradice českého pojišťovnictví. Název byl později změněn na *První českou vzájemnou pojišťovnu*, která byla založena roku 1827 v Praze. Takřka současně vznikl v Brně *Císařsko-královský privilegovaný pospolný ústav pohořelý pro Moravu a Slezsko*, který byl následně přejmenován na *Moravsko-slezskou vzájemnou pojišťovnu*. Nejprve obě tyto pojišťovny provozovaly pouze požární pojištění nemovitostí. „První česká“ začala např. provozovat krupobitní pojištění a pojištění movitostí až roku 1864. Teprve až na 81. řádné valné hromadě dne 10. května 1909 se rozhodlo, že pojišťovna rozšíří svou činnost na nová pojišťovací odvětví, a to životní pojišťování, pojišťování proti vloupání, pojišťování zákonné odpovědnosti a úrazu.

Rozmach v zakládání nových pojišťoven byl zaznamenán ve druhé polovině 19. století. K významnějším ústavům patřily například *Asekurační spolek cukrovarníků* (1862), *Pražská městská pojišťovna* (1865), *Pojišťovací požární spolek sv. Florian* v Chebu (1868), *Praha, vzájemná pojišťovna* (1869), *Rolnická vzájemná pojišťovna* v Praze (1869). K jedné z nejvíce významných českých pojišťoven vůbec se řadí *Slavia, vzájemně pojišťovací banka, Praha* (1869). Dva roky poté vznikla *Plaňanská vzájemná pojišťovna* v Plaňanech (1871), dále byla založena velmi důležitá, a jak se později ukázalo, i úspěšná instituce v oblasti zajišťovnictví – *První česká zajišťovací banka* v Praze (1872).

V sedmdesátých a osmdesátých letech 19. století bylo české pojišťovnictví opravdu silné. Ekonomickou stabilitou *První české vzájemné pojišťovny* neotřásly například ani značné škody, jako byly například požáry mlýnů roku 1870 v Čejticích. Ani požár Národního divadla v Praze 12. srpna 1881, kdy byla poskytnuta největší náhrada škody za jednu pojistnou událost v devatenáctém století, pojišťovnu neohrozil. Částka 297 869 zlatých rakouské měny umožnila téměř ihned zahájit práce na jeho obnově.

Na přelomu 19. a 20. století vznikly další společnosti – např. *Hasičská vzájemná pojišťovna* v Brně (1900) a *Moravská dobytčí pojišťovna* v Brně (1902).

Za předpokladu odborných znalostí a politické prozíravosti vedoucích představitelů v oblasti pojišťovnictví se i přes těžké válečné roky 1914-1918 povedlo převést svěřené finanční prostředky pojistníků. Novou etapu československého pojišťovnictví bylo možno začít po vzniku samostatné Československé republiky. Činnost zahájilo mnoho dalších pojišťoven, např. *Čechoslavia, lidová pojišťovna* (1919), *Pojišťovna průmyslu kvasného* (1919), *Akciová dopravní a živelní pojišťovna v Praze* (1920), *Národní pojišťovna, a.s.* (1922), aj.

I přes útlum vývoje pojišťovnictví, který nastal za protektorátu v době 2. světové války, bylo v roce 1945 evidováno dohromady 733 pojišťoven, pojišťovacích spolků a zahraničních prezentací. Dekretem prezidenta republiky z 24. října 1945 došlo k jejich znárodnění a řízením pojišťovnictví byla pověřena Pojišťovací rada se sídlem v Praze. Od 1. ledna 1947 pak v Československu vzniklo pouze pět pojišťoven, národních podniků. Nakonec po únoru 1948 došlo ke zřízení pouze jediného ústavu – *Československé pojišťovny*, národního podniku. Na několik desítek let byl přerušen přirozený tržní vývoj pojišťovnictví. V souvislosti s novým federativním uspořádáním státu roku 1968 byly z jedné *Státní pojišťovny* vytvořeny dva samostatné subjekty, a to *Česká státní pojišťovna* se sídlem v Praze a *Slovenská štátna poisťovňa* se sídlem v Bratislavě, a to s účinností od 1. ledna 1969. Toto monopolní období trvalo až do počátku devadesátých let 20. století.

Právní rámec pro změny, které souvisely se zaváděním tržní ekonomiky a soukromého podnikání po roce 1989, vytvořily v oblasti pojišťovnictví zejména nové zákony o pojišťovnictví (v ČR zákon č. 185/1991Sb.). Udělením povolení Ministerstva financí ČR a SR jakožto dozorčích orgánů nad pojišťovnictvím mohly od té doby na území České a Slovenské federativní republiky podnikat v pojišťovnictví i další pojišťovny, v právní formě založené jako akciové společnosti, státní podniky a družstva. Rozdělením ČSFR k 1. 1. 1993 se vytvořily podmínky pro samostatný rozvoj českého pojistného trhu.

České pojišťovnictví se vyvíjí v tržním prostředí v souladu s normami Evropské unie. Značné kompatibility se podařilo dosáhnout v ekonomice pojištění (účetnictví, technické rezervy, solventnost). Obecně se rozšířila mezinárodní spolupráce jak na úrovni státu, tak i České asociace pojišťoven.

K 31. prosinci 2006 provozovalo pojišťovací činnost na českém trhu 49 pojišťoven.

2 ŽIVOTNÍ POJIŠTĚNÍ A VÝZNAM MODELOVÁNÍ ÚMRTNOSTI

Smyslem životního pojištění je pojistná ochrana pojištěné osoby. Při pojistné události je totiž snížena nebo úplně znemožněna schopnost pojištěného vydělávat peníze. A právě tyto chybějící finanční prostředky mohou být zčásti nebo i plně nahrazeny z pojištění. V případě pojistné události (úmrtí, závažné onemocnění, plný invalidní důchod, úraz ap.) pojišťovna vyplatí oprávněné osobě pojistné plnění. V životním pojištění jsou tedy kryta dvě základní rizika: riziko úmrtí a riziko dožití [3].

V matematice životního pojištění je zkombinována finanční matematika s matematickým modelováním úmrtnosti, protože pojistná událost v oblasti pojištění osob tkví v úmrtí nebo dožití se určitého věku. Stejně tak i pro výpočty v důchodovém pojištění je potřeba modelování úmrtnosti [5].

Z pojistně-matematického hlediska je možno úmrtnost charakterizovat tímto způsobem [5]:

- objevují se zde právě dva stavy a to „naživu“ a „zemřelý“, přičemž o příslušném stavu každého z pojištěných lze jednoznačně rozhodnout,
- přechod mezi těmito dvěma stavy může nastat pouze jedním směrem, označovaným jako úmrtí,
- okamžik úmrtí je náhodný (i když nezvratný) a lze ho popsat jen s použitím pravděpodobnostních nástrojů.

Prostřednictvím zmíněných pravděpodobnostních nástrojů je možno vytvořit pravděpodobnostní model úmrtnosti, který např. je schopen odpovědět na otázky typu [5]:

- s jakou pravděpodobností zemře třicetiletý muž před dožitím věku 40 (kolik úmrtí během 10 let lze očekávat ve vzorku 1 000 mužů ve věku 30 let),
- s jakou pravděpodobností zemře třicetiletý muž ve věku 40 let,
- s jakou pravděpodobností bude dvacetiletá žena s třicetiletým manželem za 10 let vdovou,
- kolik budoucích let života lze očekávat u muže ve věku 50 let apod.

Numerické odpovědi na takovéto otázky zkombinované s příslušnými finančními výpočty představují výpočetní metodiku pro pojištění osob.

2.1 Délka života jako spojitá náhodná veličina

Kapitola 2.1 je zpracována dle [5] a [6].

Model úmrtnosti je možno založit na náhodné veličině $\mathbf{T}_0$, která představuje délku života právě narozeného jedince, tj. dobu mezi věkem 0 a úmrtím. Jde se o délku života lidí, takže je $\mathbf{T}_0$ obvykle měřeno v rocích a může nabývat i neceločíselných hodnot na spojité časové ose (mluví se o spojité náhodné veličině). Protože s $\mathbf{T}_0$ jsou spjaty náhodné veličiny, které představují budoucí délku života v obecném věku x, je zde tedy představa, na níž je založen model úmrtnosti, jasná: jestliže náhodně vybereme jednoho jedince z velké skupiny x-letých, pak jeho budoucí délka života sice není známá, ale můžeme na ni pohlížet jako na náhodnou veličinu s odhadnutelným pravděpodobnostním rozdělením (v daném čase a pro dostatečně velkou skupinu x-letých, např. pro padesátileté muže v České republice v roce 2007 nebo pro třicetileté ženy pojištěné v roce 2007 u České pojišťovny).

Pravděpodobnostní rozdělení náhodné veličiny se obvykle popisuje pomocí distribuční funkce, v našem případě

$$F_0(t) = P(\mathbf{T}_0 \leq t)$$

(vzhledem ke spojitosti náhodné veličiny $\mathbf{T}_0$ ovšem platí $P(\mathbf{T}_0 \leq t) = P(\mathbf{T}_0 < t)$). Někdy se vedle výše uvedené distribuční funkce délky života zavádí také *funkce přežití* jako

$$S_0(t) = P(\mathbf{T}_0 > t) = 1 - F_x(t)$$

Jak již bylo řečeno, pro pojištění osob jsou zajímavější náhodné veličiny $\mathbf{T}_x$, které představují (budoucí) délku života ve věku x za té podmínky, že daný jedinec se dožil věku x. Tato podmínka způsobuje, že pravděpodobnostní rozdělení $\mathbf{T}_x$ nelze jednoduše založit na vztahu $\mathbf{T}_x = \mathbf{T}_0 - x$, ale distribuční funkci délky života ve věku x je nutno počítat s pomocí podmíněné pravděpodobnosti:

$$F_x(t) = P(\mathbf{T}_x \leq t) = P(\mathbf{T}_0 \leq x + t | \mathbf{T}_0 > x) = \frac{P(x < \mathbf{T}_0 \leq x + t)}{P(\mathbf{T}_0 > x)}$$
$$= \frac{F_0(x + t) - F_0(x)}{S_0(x)} = \frac{F_0(x + t) - F_0(x)}{1 - F_0(x)}$$

Podobně pro funkcí přežití ve věku x platí

$$S_x(t) = P(\mathbf{T}_x > t) = P(\mathbf{T}_0 > x + t | \mathbf{T}_0 > x) = \frac{P(\mathbf{T}_0 > x + t)}{P(\mathbf{T}_0 > x)} = \frac{S_0(x + t)}{S_0(x)}$$

Například $F_{50}(1) = 0{,}005884$ pro muže v České republice v roce 2007 představuje pravděpodobnost toho, že délka života 50letého muže bude kratší než jeden rok, tj. jiným slovy se jedná o pravděpodobnost toho, že se 50letý muž nedožije věku 51.

Protože práce s distribuční funkcí je zdlouhavá, zavádí matematika životního pojištění na jejich základě některé jednodušší symboly:

q_x – *pravděpodobnost úmrtí ve věku* x

je to pravděpodobnost toho, že jedinec, který je naživu ve věku x, zemře před dosažením věku $x + 1$:

$$q_x = F_x(1) = P(\mathbf{T}_x \leq 1),$$

p_x – *pravděpodobnost dožití ve věku* x

je to pravděpodobnost toho, že jedinec, který je naživu ve věku x, se dožije věku $x + 1$:

$$p_x = S_x(1) = P(\mathbf{T}_x > 1),$$

${}_tq_x$ – je to pravděpodobnost toho, že jedinec, který je naživu ve věku x, zemře před dosažením věku $x + t$ (pro $t = 1$ se zřejmě jedná o q_x):

$${}_tq_x = F_x(t) = P(\mathbf{T}_x \leq t),$$

${}_tp_x$ – je to pravděpodobnost toho, že jedinec, který je naživu ve věku x, se dožije věku $x + t$ (pro $t = 1$ se zřejmě jedná o p_x):

$${}_tp_x = S_x(t) = P(\mathbf{T}_x > t),$$

${}_{s|}q_x$ – je to pravděpodobnost toho, že jedinec, který je naživu ve věku x, zemře ve věku $x+s$ (pro $s=0$ se zřejmě jedná o q_x):

$$ {}_{s|}q_x = F_x(s+1) - F_x(s) = P(s < \mathbf{T}_x \le s+1) $$

${}_{s|t}q_x$ – je to pravděpodobnost toho, že jedinec, který je naživu ve věku x, se dožije věku $x+s$, ale zemře před dosažením věku $x+s+t$ (pro $t=1$ se zřejmě jedná o ${}_{s|}q_x$):

$$ {}_{s|t}q_x = F_x(s+t) - F_x(s) = P(s < \mathbf{T}_x \le s+t) $$

Velmi důležité jsou také některé globální charakteristiky náhodné veličiny $\mathbf{T}_x$, a to především její střední hodnota označovaná jako *střední délka života ve věku x* za podmínky, že daný jedinec se dožil věku x. Udává průměrný počet let života zbývajících jedinci ve věku x:

$$ \dot{e}_x = E(\mathbf{T}_x). $$

2.2 Intenzita úmrtnosti

Tato kapitola 2.2 je zpracována dle [5] a [6].

Pokud náhodná veličina $\mathbf{T}_0$ má pravděpodobnostní hustotu

$$ f_0(t) = \frac{\mathrm{d}}{\mathrm{d}\,t} F_0(t) = -\frac{\mathrm{d}}{\mathrm{d}\,t}\ {}_tp_0, $$

pak mají pravděpodobnostní hustotu také náhodné veličiny $\mathbf{T}_x$

$$ f_x(t) = \frac{\mathrm{d}}{\mathrm{d}\,t} F_x(t) = \frac{\mathrm{d}}{\mathrm{d}\,t} \frac{F_0(x+t) - F_0(x)}{1 - F_0(x)} = -\frac{\mathrm{d}}{\mathrm{d}\,t} \frac{{}_{x+t}p_0}{{}_xp_0} = -\frac{\mathrm{d}}{\mathrm{d}\,t}\ {}_tp_x. $$

V takovémto případě lze zavést další důležitý pojem, kterým je *intenzita úmrtnosti ve věku x*:

$$ \mu_x = \frac{f_0(x)}{{}_xp_0} = -\frac{1}{{}_xp_0} \cdot \frac{\mathrm{d}}{\mathrm{d}\,x}\ {}_xp_0 = -\frac{\mathrm{d}}{\mathrm{d}\,x} \ln\left({}_xp_0\right). $$

Intenzita úmrtnosti je základem konstrukce stochastických úmrtnostních tabulek a má tyto základní vlastnosti:

- zohledňuje spojitost lidského života,
- intenzita úmrtnosti ve věku x:

$$\mu_x = \lim_{\Delta x \to 0} \left\{ \frac{1}{\Delta x} \cdot P(T_0 \leq x + \Delta x | T_0 > x) \right)$$

- pravděpodobnost, že když se x-letá osoba dožije okamžiku x, umře v tomto okamžiku:

$\lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{P(x < T_0 \leq x + \Delta x)}{P(T_0 > x)} = \frac{1}{S_0(x)} \lim_{\Delta x \to 0} \frac{F_0(x+\Delta x) - F_0(x)}{\Delta x} = \frac{1}{S_0(x)} \cdot F_0^{'}(x) = \frac{1}{S_0(x)} [1 - S_0(x)]^{'} = -\frac{S_0^{'}(x)}{S_0(x)} = -[\ln S_0(x)]^{'} = -\frac{d}{dx} \ln \ {}_x P_0$.

2.3 Zákony úmrtnosti

Celá kapitola 2.3 včetně podkapitol 2.3.1, 2.3.2, 2.3.3 a 2.3.4 je zpracována dle [5] a [6].

Problematika zákonů úmrtnosti je zdůvodňována snahou modelovat lidskou úmrtnost prostřednictvím matematických vzorců. Jestliže se nám takovéto vzorce (křivky úmrtnosti) povedou v určitém čase pro určitou populaci (případně pojistný kmen) zkonstruovat včetně numerického odhadu jejich parametrů, potom nám vznikne účinný nástroj pro různé operace s praktickými úmrtnostními údaji, které jsou většinou vyjadřovány ve formě tzv. úmrtnostních tabulek. Křivky úmrtnosti odpovídající různým zákonům úmrtnosti, znamenají totiž úsporný popis obrovského množství údajů (např. pravděpodobností úmrtí pro jednotlivé věky). Poněvadž to jsou hladké křivky, jejich užití současně přispívá k vyrovnání (vyhlazování) úmrtnostních tabulek, které jsou jako každý statistický odhad zatíženy statistickými chybami způsobujícími nechtěné fluktuace kolem odhadovaných hodnot.

Přitom různé zákony úmrtnosti jsou právě charakterizovány svými typickými intenzitami úmrtnosti.

2.3.1 Konstantní intenzita úmrtnosti

Jestliže je intenzita úmrtnosti konstantní, tedy

$$\mu_x = \lambda,$$

pak z ${}_tp_x = exp\left(-\int_0^t \mu_{x+s} \mathrm{d}s\right)$ plyne

$${}_tp_x = e^{-\lambda t}$$

($F_x(t) = 1 - {}_tp_x = 1 - \exp(-\lambda t)$ je skutečně pravděpodobnostní distribuční funkce). Protože zde funkce přežití ${}_tp_x$ zřejmě nezávisí na věku x, je takovýto zákon úmrtnosti očividně nevhodný pro lidskou populaci. Zde totiž očekáváme růst intenzity úmrtnosti s věkem, jak je tomu u následujících zákonů úmrtnosti.

2.3.2 Moivrův zákon úmrtnosti

De Moivre navrhl rovnoměrné rozdělení délky života s pravděpodobnostní hustotou

$$f_x(t) = \frac{1}{\omega - x}, \qquad 0 < t < \omega - x,$$

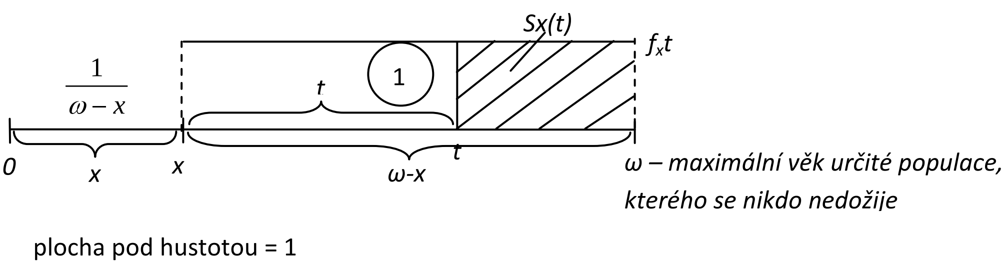
kde ω je stanovený nejvyšší věk pro uvažovanou populaci. Pak je zřejmě

$${}_tp_x = \frac{\omega - x - t}{\omega - x}, \qquad 0 < t < \omega - x$$

a intenzita úmrtnosti

$$\mu_x = \frac{1}{\omega - x}, \qquad 0 < x < \omega$$

je hyperbolicky rostoucí funkce.



Obrázek 1: Moivrův zákon úmrtnosti

2.3.3 Gompertzův zákon úmrtnosti

Gompertz použil exponenciálně rostoucí intenzitu úmrtnosti:

$$\mu_x = B \cdot c^x,$$

kde B > 0, c > 1 jsou parametry. Dle ${}_tp_x = exp\left(-\int_0^t \mu_{x+s}\mathrm{d}s\right)$ lze odvodit

$${}_tp_x = g^{c^x(c^t-1)},$$

kde $g = e^{-\frac{B}{\ln c}}$.

2.3.4 Makehamův zákon úmrtnosti

Tento zákon úmrtnosti ve srovnání s ostatními úmrtnostními křivkami podává obvykle nejlepší výsledky.

Makeham zobecnil Gompertzovu intenzitu úmrtnosti do tvaru:

$$\mu_x = A + B \cdot c^x,$$

s dalším parametrem A > 0. Podle ${}_tp_x = exp\left(-\int_0^t \mu_{x+s}\mathrm{d}s\right)$ je

$${}_tp_x = s^t \cdot g^{c^x(c^t-1)},$$

kde $s = e^{-A}$.

Makehamovu křivku lze odhadnout na základě hodnot intenzity úmrtnosti pro tři různé věky, např. pro hodnoty intenzity úmrtnosti mužů μ_{30} = 0,000 967, μ_{40} = 0,001 961 a μ_{50} = 0,005 575 (hodnoty byly vypočítány z úmrtnostních tabulek mužů za rok 2007). Odhadnutou křivku použijeme pro výpočet pravděpodobnosti dožití ${}_5p_{50}$.

Postup je následující:

$\mu_{30} = A + B \cdot c^{30}$

$\mu_{40} = A + B \cdot c^{40}$

$\mu_{50} = A + B \cdot c^{50} \;\Rightarrow A = \mu_{50} - B \cdot c^{50}$

$\mu_{50} - \mu_{40} = B \cdot c^{40}(c^{10} - 1) \Rightarrow B = \frac{\mu_{50}-\mu_{40}}{c^{40}(c^{10}-1)} = \frac{\mu_{50}-\mu_{40}}{c^{50}-c^{40}}$

$\mu_{40} - \mu_{30} = B \cdot c^{30}(c^{10} - 1)$

$$\frac{\mu_{50} - \mu_{40}}{\mu_{40} - \mu_{30}} = c^{10} \;\Rightarrow c = \left(\frac{\mu_{50} - \mu_{40}}{\mu_{40} - \mu_{30}}\right)^{\frac{1}{10}} = \left(\frac{0{,}005\,575 - 0{,}001\,961}{0{,}001\,961 - 0{,}000\,967}\right)^{\frac{1}{10}} \Rightarrow$$

$\boldsymbol{c = 1{,}137\,852}$

$$\boldsymbol{B} = \frac{0{,}005\,575 - 0{,}001\,961}{1{,}137\,852^{40}(1{,}137\,852^{10} - 1)} = \mathbf{0{,}000\,008}$$

$\boldsymbol{A} = \mu_{50} - B \cdot c^{50} = 0{,}005\,575 - 0{,}000\,008 \cdot 1{,}137\,852^{50} = \mathbf{0{,}000\,591}$

Z tohoto dostaneme

$s = \exp(-A) = \exp(-0{,}000\,591) = 0{,}999\,409,$

$g = \exp\left(-B/\ln(c)\right) = \exp\left(-0{,}000\,008/\ln\left(1{,}137\,852\right)\right) = 0{,}999\,939.$

Dle vzorce ${}_t p_x = s^t \cdot g^{c^x(c^t-1)}$ tedy je

$$\boldsymbol{{}_5 p_{50}} = s^5 \cdot g^{c^{50}(c^5-1)} = 0{,}999\,409^5 \cdot 0{,}999\,939^{1{,}137\,852^{50}(1{,}137\,852^5-1)}$$
$$= \mathbf{0{,}962\,739}.$$

2.4 Faktory a ukazatele úmrtnosti

Následující kapitola 2.4 včetně podkapitoly 2.4.1 je zpracována dle [9], [21] a [22].

Úmrtnost, která je jedním z hlavních demografických procesů, představuje spolu s porodností základní složku demografické reprodukce populací. Pokud doplníme úmrtnost nemocností, je úmrtnost jedním z klíčových ukazatelů, které vypovídají o zdravotním stavu populace. Zdravotní stav, nemocnost a úmrtnost jsou determinovány řadou faktorů.

V podstatě lze vyčlenit tři hlavní skupiny:

a) genetické faktory - např. skutečnost vyšší míry úmrtnosti mužů (ženy mají nižší úmrtnost a žijí déle, proto studujeme úmrtnost vždy odděleně za jednotlivá pohlaví),
b) ekologické faktory - např. klimatické podmínky, životní prostředí,
c) socioekonomické faktory:
    - individuální: životní úroveň, úroveň vzdělání, péče o zdraví a využívání preventivních opatření, stravovací návyky, výživa, fyzická aktivita,
    - vlivy prostředí: úroveň zdravotnictví, dostupnost a kvalita lékařské péče, rozvoj medicíny a lékařské techniky, systém zdravotní politiky, systém sociálního zabezpečení, ekonomická situace.

Počty zemřelých jsou statisticky evidovány v rámci určitého území. O každém zemřelém je v ČR (dle vyhlášky MZ ČR z roku 1987) vystaven formulář „List o prohlídce mrtvého“, který je vyplněn příslušným lékařem a zaslán na matriku. Na tomto formuláři se o zemřelém uvádí řada charakteristik (kromě identifikačních údajů i rodný stav, vzdělání apod.), zvláštní část je věnována lékařskému osvědčení o příčinách smrti. Matrika dle tohoto tiskopisu vypíše další formulář „Hlášení o úmrtí“ a zašle ho ke zpracování Českému statistickému úřadu. Ten shromažďuje a čtvrtletně publikuje absolutní i relativní údaje o zemřelých. V mezinárodním pohledu je tato evidence vykonávána Světovou zdravotnickou organizací.

2.4.1 Základní ukazatele úmrtnosti

- **Hrubá míra úmrtnosti** – nejjednodušší ukazatel, který vyjadřuje úroveň úmrtnosti, je definován jako počet zemřelých připadajících na 1 000 obyvatel (středního stavu). Nevýhodou tohoto ukazatele je skutečnost, že nebere v úvahu věková specifika dané populace, proto je např. pro srovnání několika územních celků (např. kraje, státy apod.) mnohdy používána standardizovaná úmrtnost.

  Př. V roce 2007 zemřelo v ČR 104 636 mužů a žen, počet obyvatel (k 31. 12. 2007) činil 10 381 130 osob. Hrubá míra úmrtnosti v roce 2007 tedy byla 10,08 zemřelých na 1 000 obyvatel.

- **Míra úmrtnosti dle věku** (věkově specifická úmrtnost) – vyjadřuje počet zemřelých ve věku x (respektive v dané věkové skupině – pěti či desetileté), vztahuje se k počtu obyvatel (středního stavu) v daném věku x (věkové skupině).

  Př. V roce 2007 zemřelo v ČR ve věku 50 – 54 let 2 810 mužů, počet mužů ve věku 50 – 54 let (k 31. 12. 2007) činil 373 028, míra úmrtnosti ve věku 50 – 54 let představovala tedy 7,53 zemřelých na 1 000 mužů dané věkové skupiny.

  S přibývajícím věkem úmrtnost stoupá, výjimku však představují nejnižší věkové kategorie a vyšší úmrtnost je také spojena s prvním rokem života. Dále ve věku 15 – 35 let se úmrtnost také mírně zvyšuje a to především v důsledku vnějších příčin smrti jako jsou například úrazy, nehody, dopravní nehody, otravy apod. Ve všech věkových kategoriích míra úmrtnosti mužů převyšuje úmrtnost žen (tzv. mužská neúmrtnost).

- **Střední délka života** (naděje dožití) – ukazatel, který vychází z úmrtnostních tabulek a který vyjadřuje počet let, které v průměru ještě osoba ve věku x prožije. Jedná se o ukazatel hypotetický, vychází z předpokladu zachování stávajících úmrtnostních poměrů a vyjadřuje úmrtnostní situaci v daném roce.

  Nejčastěji se udává střední délka života ve věku 0, tedy při narození, zvlášť za obě pohlaví.

Př.: V roce 2007 činila střední délka života mužů 73,67, u žen byla její hodnota 79,90. Znamená to, že právě narozený muž by se, za předpokladu přetrvání stávajících úmrtnostních poměrů po celý svůj život, dožil zhruba 73,6 let, právě narozená žena by se dožila 80 let.

- **Kojenecká úmrtnost** – vyjadřuje úmrtnost v prvním roce života, tedy v průběhu prvních 365 dní života dítěte.

- **Kvocient kojenecké úmrtnosti** je definován jako počet zemřelých do 1 roku života na 1 000 živě narozených v daném roce.

  Př. V roce 2007 ve věku 0 zemřelo v ČR 315 dětí, živě se narodilo 114 632 dětí, kvocient kojenecké úmrtnosti představoval tedy 2,75 zemřelých do 1 roku na 1 000 živě narozených.

  V podrobnějším členění lze v prvním roce života dále vymezit úmrtnost:

    - novorozeneckou, tj. v průběhu prvních 0 – 27 dní,
    - časnou 0 – 6 dní,
    - pozdní 7 – 27 dní,
    - ponovorozeneckou 28 – 364 dní.

- **Perinatální úmrtnost** – vyjadřuje počet mrtvě narozených a zemřelých v průběhu prvních šesti dnů života na 1 000 všech narozených (živě i mrtvě).

2.5 Úmrtnostní tabulky a jejich využití v pojišťovnictví

Tato kapitola 2.5 a podkapitola 2.5.1 jsou zpracovány dle [1], [7], [8], [15], [17] a [23].

Úmrtnostní tabulky představují nejčastější způsob analýzy úmrtnosti. Přesnými mírami vystihují vymírání populace v jednotlivých věkových skupinách. Tradičně se používají při odhadování úmrtnosti v pojišťovnách, i když údaje v nich se zřejmě nikdy v budoucnu nebudou úplně opakovat.

Aktuár životní pojišťovny na většinu výpočtů při produktech životního a důchodového pojištění nevyhnutelně potřebuje právě úmrtnostní tabulky.

Od té doby, kdy byla v civilizovaných zemích podstatně redukována porodní úmrtnost, vykazuje mužská populace prakticky ve všech věkových kategoriích větší sklon k úmrtnosti než populace ženská. Ženy žijí v průměru déle než muži a tento fakt pojišťovny většinou zohledňují ve svých sazebníkách. Přístupy při řešení těchto problémů jsou různé:

a) pojišťovna počítá sazby pojistného pro muže a ženy tak, že používá úmrtnostní tabulky zvlášť pro muže a zvlášť pro ženy,
b) pojišťovna počítá pojišťovací sazby pro muže s použitím mužských úmrtnostních tabulek a sazby pro ženy získá vhodným věkovým posunutím sazeb pro muže.

Pokud pojišťovna použije celostátní úmrtnostní tabulku, zkonstruovanou na základě demografických údajů obyvatelstva, musí počítat s rizikem, že úmrtnost pojištěnců bude do jisté míry odlišná. Pojistné smlouvy při životním pojištění uzavírají relativně zdraví jedinci s perspektivou dlouhého života, pro které je pravděpodobnost úmrtí nižší než v celé populaci. Na druhé straně pojistné smlouvy na úmrtí uzavírá taková část populace, pro kterou je vyšší pravděpodobnost úmrtí. Riziko se v takovémto případě nazývá někdy rizikem selekce. Na tuto skutečnost pojišťovny reagují určitými opatřeními. Velmi častým opatřením je věkový posun použité úmrtnostní tabulky ve prospěch pojišťovny. Přirozenou ochranou pojišťovny, která však často nepůsobí dobře hlavně z propagačního hlediska, je použití zastaralých úmrtnostních tabulek.

Základní dělení úmrtnostních tabulek:

1. generační úmrtnostní tabulky,
2. průřezové úmrtnostní tabulky.

- **Generační úmrtnostní tabulky**

Generační úmrtnostní tabulky představují záznam průběhu života konkrétní populace současně narozených jedinců od jejich narození až po smrt posledního z nich. Konstrukce těchto úmrtnostních tabulek je obtížná, předpokládá, kromě jiného, statistické sledování populace po dlouhou dobu (sledovaná populace se v dlouhém časovém úseku zmenšuje nejen vymíráním, ale i migrací). Generační tabulky mají častější použití při sledování délky života zvířat (např. sledování kmene bakterií, kde je průměrná délka života obvykle o dost kratší než u člověka. Generační úmrtnostní tabulky se také často nazývají kohortní úmrtnostní tabulky, protože pro generaci, tj. pro osoby narozené v tom samém období se v demografii používá název kohorta. Je to všeobecnější označení pro skupinu osob, při které došlo k určité události ve stejném čase, resp. období (skupina pacientů podstupující současně nějakou terapii).

- **Průřezové úmrtnostní tabulky**

  Průřezové úmrtnostní tabulky vycházejí z úmrtnostních událostí dané populace v průběhu krátkého časového období (většinou je obdobím rok). Na základě úmrtnostním měr podle jednotlivých věkových skupin se konstruuje obraz života hypotetické populace současně narozených jedinců. Průřezové tabulky jsou ve skutečnosti běžné úmrtnostní tabulky a používají se mnohem častěji než generační tabulky.

Dále můžeme úmrtnostní tabulky rozdělit podle délky sledovaného věkového intervalu takto:

- podrobné úmrtnostní tabulky, které pracují s věkovým intervalem jeden rok,
- zkrácené úmrtnostní tabulky, které pracují s věkovým intervalem delším než jeden rok (zpravidla se volí věkový interval 5 let)

V životním pojištění pracujeme s podrobnými průřezovými tabulkami.

2.5.1 Popis sloupců úmrtnostní tabulky

Úmrtnostní tabulky obsahují ve sloupcích nejčastěji tyto ukazatele úmrtnosti pro různé věky x.

$\boldsymbol{x}$ – vyjadřuje věk osoby,

pro $x \in \{0,1,\dots,\omega\}$, přičemž ω je předpokládaný nejvyšší věk, kterého může dosáhnout osoba sledovaného souboru.

$\boldsymbol{l_x}$ – vyjadřuje počet osob, dožívajících se věku x. Počáteční hodnota l_0 vyjadřuje počáteční počet osob modelovaného souboru a nazývá se *kořen úmrtnostní tabulky* (radix). Libovolná další hodnota l_x představuje počet jedinců z kořene l_0, kteří se dožijí věku x. Konečná posloupnost $\{l_x\}_0^\omega$ je nerostoucí, tj. $l_0 \geq l_1 \geq \cdots \geq l_\omega$ a nazývá se postoupnost žijících, neboli funkce přežití.

$\boldsymbol{q_x}$ – vyjadřuje pravděpodobnost toho, že x letá osoba se nedožije věku $(x+1)$ a nazývá se roční míra úmrtnosti.

$\boldsymbol{p_x}$ - vyjadřuje pravděpodobnost toho, že x letá osoba se dožije věku $(x+1)$ a nazývá se roční míra dožití. Platí:

$$q_x + p_x = 1$$

Z klasické definice pravděpodobnosti vyplývá platnost následujících vztahů:

$$p_x = \frac{l_{x+1}}{l_x}$$

$$q_x = \frac{l_x - l_{x+1}}{l_x}$$

$\boldsymbol{d_x}$ - vyjadřuje tabulkový počet jedinců, kteří zemřeli ve věku x. Platí zde vztahy:

$$d_x = l_x - l_{x+1}$$

$$q_x = \frac{d_x}{l_x},$$

$$l_\omega = d_\omega.$$

$\boldsymbol{a_x}$ – vyjadřuje délku části posledního roka života jedinců, kteří zemřeli ve věku x. Jde o průměrný počet let prožitých ve věkovém intervalu $\langle x, x+1)$ těmi d_x jedinci, kteří v tomto věkovém intervalu zemřeli. Hodnoty a_x se v některých úmrtnostních tabulkách neuvádějí, protože mají jen pomocný charakter. Za předpokladu rovnoměrného počtu zemřelých v časovém intervalu $\langle x|x+1)$ je $a_x = \frac{1}{2}$.

$\boldsymbol{L_x}$ - vyjadřuje očekávaný počet let, prožitých osobami ve věku x, tj. celkový počet člověkoroků, které ještě prožijí osoby ve věku x, tj. které ještě prožije počet osob l_x. Pro tuto hodnotu platí

$$L_x = l_{x+1} + a_x \cdot d_x,$$

protože každý z l_{x+1} dožívajících přispěje do L_x jedním rokem, pokud každý z d_x osob, které zemřeli ve věku x, přispěje v průměru hodnotou a_x. Pokud je například $a_x = 0{,}5$, potom dostaneme

$$L_x = l_{x+1} + 0{,}5d_x = \frac{(l_x + l_{x+1})}{2}.$$

$\boldsymbol{T_x}$ –vyjadřuje počet zbývajících roků života všech osob ve věku x, tedy

$$T_x = L_x + L_{x+1} + \cdots + L_\omega$$

$\boldsymbol{e_x^0}$ - vyjadřuje střední délku života osoby ve věku x, tj. průměrný počet roků, které se ještě dožije jedinec ve věku x. Platí:

$$e_x^0 = \frac{T_x}{l_x}$$

3 LOGISTICKÝ REGRESNÍ MODEL JAKO MODEL PŘEŽITÍ

Celá následující kapitola 3 je pracována dle [10] a [13].

Logistická regrese je metoda vhodná na modelování jednostranné závislosti mezi proměnnými. Jako závislá proměnná v modelu vystupuje kategoriální proměnná a jako vysvětlující mohou být proměnné spojitého nebo kategoriálního typu. Jde o modelování podmíněné pravděpodobnosti jedné obměny kategoriálně závislé proměnné v závislosti na jiných proměnných.

Využívá se na modelování závislosti proměnných, ve kterých zkoumaná (závislá, referenční) proměnná není spojitá. Se závislými proměnnými kategoriálního typu se setkáváme poměrně často v mnohých vědních disciplínách a také v praxi. Vyskytují se v nominálních (např. ano/ne, úspěch/neúspěch) nebo v ordinálních obměnách (např. dobrý/velmi dobrý/výborný) a mohou být dichotomické nebo i množné.

Logistická regrese zkoumá závislost mezi pravděpodobností konkrétní obměny závisle proměnné na nezávisle proměnných. Vysvětlující proměnné mohou být stejně tak kvalitativní jako kvantitativní. V případě jednoduchého logistického modelu, tj. v modelu jen s jednou vysvětlující proměnnou, která je kategoriální, se logistická regrese redukuje na analýzu kontingenční tabulky.

3.1 Relativní riziko a pravděpodobnostní poměr

Označme písmenem E libovolnou událost (event) a písmenem F libovolný faktor, který ovlivňuje nastání události E.

Použijme označení pro E:

- 1 (když událost nastane),
- 0 (nenastane E).

Pro rizikový faktor F použijme analogicky označení:

- 1 (výskyt faktoru F),
- 0 (nevyskytuje se).

Výsledky výskytu události E v závislosti na rizikovém faktoru F přehledně zobrazuje asociační tabulka (tabulka 1).

Tabulka 1: Asociační tabulka [13]

Rizikový faktor F	**Událost E**		**Sloupcové součty**
	Ano (E = 1)	**Ne (E = 0)**	
Ano (F = 1)	**a** n_{11}	**b** n_{10}	$\boldsymbol{a+b}$
Ne (F = 0)	**c** n_{01}	**d** n_{00}	$\boldsymbol{c+d}$
Řádkové součty	$\boldsymbol{a+c}$	$\boldsymbol{b+d}$	**n**

Symboly v tabulce 1 značí:

$\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$ – četnosti 2. stupně,

např. $\boldsymbol{d}$ – počet těch, u kterých nenastala událost a nevyskytl se ani rizikový faktor (např. neumřeli a jsou nekuřáci v průběhu sledování).

Řádkové četnosti 1. stupně:

$\boldsymbol{a+b}$ – vztahuje se pouze na výskyt faktoru F

$\boldsymbol{c+d}$ – počet těch, u kterých se rizikový faktor F nevyskytl

Sloupcové četnosti 1. stupně vztahující se k výskytu události E:

$\boldsymbol{a+c}$ – počet těch, u kterých se událost E vyskytla (např. počet zemřelých)

$\boldsymbol{b+d}$ – počet těch, u kterých se událost nevyskytla

Asociační tabulku 1 můžeme využít na ověření toho, či faktor F významně ovlivňuje nastání události E nebo nikoli a v případě jeho významnosti kvantifikovat intenzitu jeho vlivu. Ověřujeme tedy hypotézy:

H_0: E nezávisí na faktoru F

H_1: E významně závisí na faktoru F

K ověření těchto hypotéz používáme testovací charakteristiku chí kvadrát:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

kde O_i jsou skutečné zjištěné početnosti (2. stupně) – a, b, c, d a E_i jsou očekávané početnosti, když platí H_0, které dostaneme takto:

$$a_0 = \frac{(a+b)(a+c)}{n}$$

$$b_0 = \frac{(a+b)(b+d)}{n}$$

$$c_0 = \frac{(c+d)(a+c)}{n}$$

$$d_0 = \frac{(c+d)(b+d)}{n}$$

Pro podmíněné pravděpodobnosti platí:

$$P(E = 1/F = 1) = \frac{a}{a+b}$$

$$P(E = 1/F = 0) = \frac{c}{c+d}$$

Důležitým ukazatelem vlivu faktoru F na nastání události E je relativní riziko (RR):

$$RR = \frac{P(E = 1/F = 1)}{P(E = 1/F = 0)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a(c+d)}{c(a+b)}$$

Časté využití má i tzv. odds ratio (OR), nabývá pravděpodobnostní poměr:

$$OR = \frac{\frac{P(F = 1/E = 1)}{P(F = 0/E = 1)}}{\frac{P(F = 1/E = 0)}{P(F = 0/E = 0)}} = \frac{\frac{\frac{a}{a+c}}{\frac{c}{a+c}}}{\frac{\frac{b}{b+d}}{\frac{d}{b+d}}} = \frac{a \cdot d}{b \cdot c}$$

3.1.1 Ukázka aplikace

Data pro ukázku byla vybrána ze statistického programového balíku Statgraphics Centurion XV, konkrétně byl zvolen soubor *collisions.sf6*. Tento soubor obsahuje údaje umožňující měřit vliv vybraných rizikových faktorů na událost, kterou je smrtelná nehoda při srážce dvou automobilů. Sledovaným rizikovým faktorem je rychlost vozidla, které bylo příčinou srážky (km/h), upravená do tvaru dichotomické kategoriální proměnné.

Rizikový faktor F - **rychlost** tedy uvažujeme jako dichotomickou proměnnou se dvěma variantami:

0 < 100 km/h

1 > 100 km/h

Událost E:

1 úmrtí

0 přežití

Tabulka 2: Asociační tabulka pro příklad 1

Rizikový faktor F	**Událost E**		**Sloupcové součty**
	1	**0**	
1	a = 25	b = 14	a + b = 39
0	c = 9	d = 10	c + d = 19
Řádkové součty	a + c = 34	b + d = 24	n = 58

Z tabulky 2 je možné vypočítat relativní riziko a odds ratio:

Výpočet relativního rizika:

$$RR = \frac{P(E = 1/F = 1)}{P(E = 1/F = 0)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a(c+d)}{c(a+b)} = \frac{25 \cdot 19}{9 \cdot 39} = \frac{475}{351} = 1{,}3533$$

Pravděpodobnost, že událost E – smrtelný úraz nastane, je tedy 1,3533 násobně větší v případě, že je rychlost vozidla, které zapříčinilo srážku, nad 100 km/h (při výskytu faktoru F), než v případě, že je rychlost menší než 100 km/h, tedy faktor F se nevyskytl.

Výpočet odds ratio:

$$OR = \frac{\frac{P(F = 1/E = 1)}{P(F = 0/E = 1)}}{\frac{P(F = 1/E = 0)}{P(F = 0/E = 0)}} = \frac{a \cdot d}{b \cdot c} = \frac{25 \cdot 10}{14 \cdot 9} = 1{,}9841$$

Odds ratio nás informuje o tom, že podíl pravděpodobnosti výskytu faktoru F (rychlost vozidla byla nad 100 km/h) a pravděpodobnosti toho, že se faktor F nevyskytl, je 1,9841-krát vyšší v případě, pokud srážka automobilů měla smrtelné následky, než v případě, že srážka smrtelné následky neměla.

Výstupy ze systému Statgraphics Centurion XV

Frequency Table for rychlost_cat by zavažnost

	0	1	Row Total	
0	10	9	19	1)
	17,24%	15,52%	32,76%	2)
	7,86	11,14		3)
	0,58	0,41		4)
1	14	25	39	1)
	24,14%	43,10%	67,24%	2)
	16,14	22,86		3)
	0,28	0,20		4)
Column Total	24	34	58	
	41,38%	58,62%	100,00%	

Cell contents:
Observed frequency
Percentage of table
Expected frequency
Contribution to chi-squared

Obrázek 2: Výstup ze systému Statgraphics Centurion XV

Obrázek 2 obsahuje tyto údaje:

1) Observed frequency (skutečné zjištění) O_i

2) Percentage of table (procento z celkového počtu n)

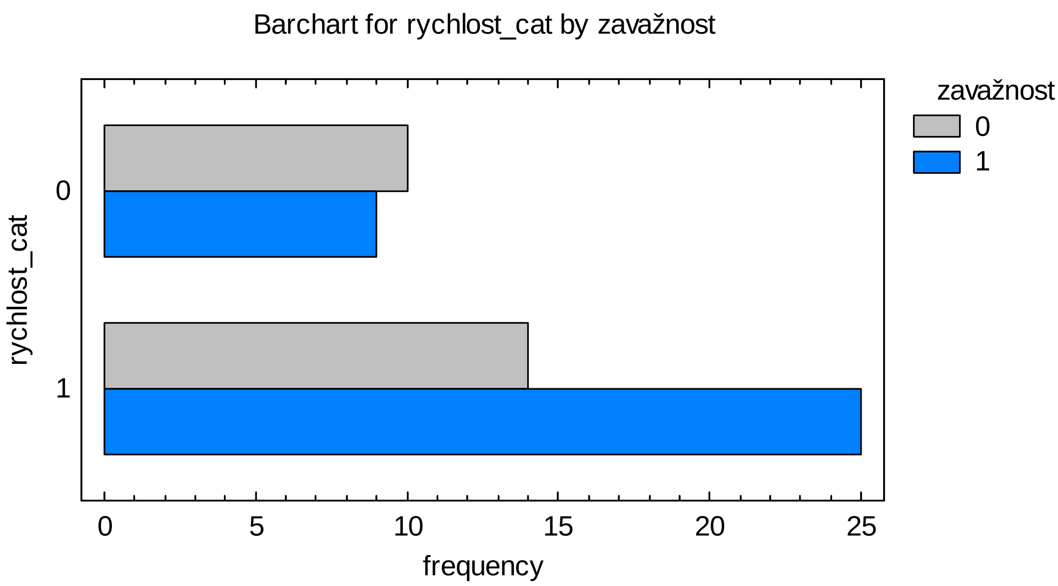
3) Expected frequency (očekávané početnosti) E_i, pokud faktor F není významný

4) Contribution to chi-squared (náležitost do chí-kvadrátu)

Chí-kvadrát test nezávislosti události E na faktoru F:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 0{,}58 + 0{,}41 + 0{,}28 + 0{,}20 = 1{,}475$$

Hodnota testovacího kritéria $\chi^2 = 1{,}475$ je nižší než kritická hodnota na hladině významnosti 0,05, která je 3,841 459, proto se nepotvrdila významná závislost smrtelných následků srážky vozidel na tom, či rychlost byla anebo nebyla nad 100 km/h.



Obrázek 3: Závislost úmrtí na rychlosti jízdy (výstup ze Statgraphics Centurion XV)

Z obrázku 3 je jasně viditelné, že je při překročení rychlostní hranice 100 km/h vyšší pravděpodobnost dopravní nehody a vyšší riziko úmrtí než při jízdě pod 100 km/h. Při rychlosti nad 100 km/h převažuje počet smrtelných úrazů v porovnání s těmi, kde úraz nebyl smrtelný, při rychlosti pod 100 km/h je podíl smrtelných úrazů menší než v prvním případě.

Tests of Independence

Test	*Statistic*	*Df*	*P-Value*
Chi-Squared	1,475	1	0,2246

Odds Ratios and Relative Risk

Relative Risk

Numerator	*Denominator*	*Relative Risk*	*95% LCL*	*95% UCL*
0	1	1,3533	0,80609	2,66675

Odds Ratios

Numerator	*Denominator*	*Odds Ratio*	*95% LCL*	*95% UCL*
0	1	1,98413	0,651852	6,03934

Obrázek 4: Výstup ze systému Statgraphics Centurion XV

Uvedené výpočty potvrzuje i výstup ze systému Statgraphics Centurion XV na obrázku 4. Intervaly spolehlivosti pro relativní riziko a odds ratio interpretujeme takto:

S pravděpodobností 0,95 je relativní riziko nastalé události E, pokud faktor F nastane anebo nenastane, z intervalu 0,80609 až 2,66675. Analogicky odds ratio je s pravděpodobností 0,95 z intervalu 0,651852 až 6,03934.

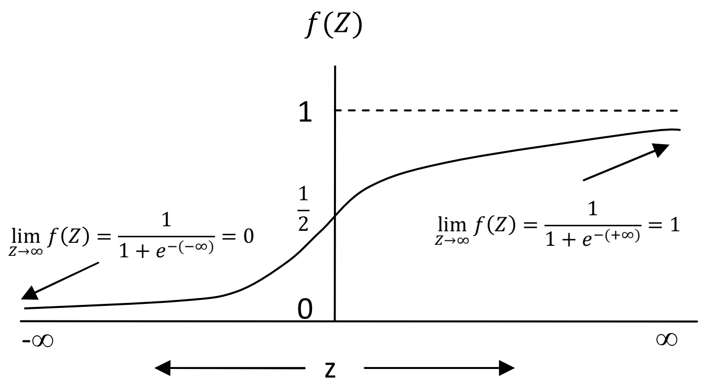
3.2 Logistický regresní model

Cílem aplikace logistického modelu v analýze přežití je odhad vztahů mezi rizikovými faktory (F) a proměnnou nastání události (E), u které když nabývá hodnoty 0, tak událost nenastala a pokud nabývá hodnoty 1, tak událost nastala.

Rizikový faktor F spolu s kontrolními proměnnými C_1, C_2, C_3, představují soubor nezávislých proměnných, které popisují závislou proměnnou E. Nezávisle proměnné označujeme všeobecně symboly $X_1, X_2 \dots X_k$, kde k vyjadřuje počet uvažovaných nezávislých proměnných. X může značit i kombinaci některých nezávislých proměnných ($C_1 \times C_2, E^2$).

V logistické regresní analýze přímo odhadujeme pravděpodobnost nastání události. Logistická funkce, která popisuje logistický model, má tvar

$$f(z) = \frac{1}{1 + e^{-Z}}$$



Obrázek 5: Logistická funkce [13]

Logistická křivka je prodloužená S-křivka, jejíž tvar se podobá distribuční funkci normálního rozdělení. Nabývá hodnoty z intervalu (0, 1). Logistická funkce se využívá prostřednictvím proměnné z, která kombinuje vlivy rizikových faktorů.

$$Z = \alpha + \beta_1 \cdot X_1 + \cdots + \beta_k \cdot X_k$$

kde X_i jsou nezávislé proměnné a α a β_i jsou konstanty, reprezentující neznámé parametry.

Na odhad parametrů logistického regresního modelu využíváme metodu maximální věrohodnosti. Protože logistický regresní model je nelineární, parametry modelu získáme iteračními metodami.

Logistický model souvisí s následujícím problémem: zjistili jsme hodnoty nezávislých proměnných $X_1, X_2 \dots X_k$ pro skupinu osob, u kterých jsme zjistili i stav události (0 – událost nenastala, 1 – událost nastala). Tyto informace chceme využít na popsání pravděpodobnosti vzniku události $P(E = 1|X_1, X_2 \dots X_k)$ po určité období (T_0 až T_1) s hodnotami nezávislých proměnných $X_1, X_2 \dots X_k$, naměřených v čase T_0. Pro zjednodušení tuto pravděpodobnost označme $P(\mathbf{X})$, kde $\mathbf{X}$ vyjadřuje sloupcový vektor proměnných $(X_1, X_2 \dots X_k)'$. Logistický model bude mít tvar

$$P(X) = P(E = 1|X_1, X_2 \dots X_k) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i \cdot X_i)}}$$

Neznámé parametry α a β_i odhadneme na základě získaných údajů proměnných $X_1, X_2 \dots X_k$ a E. Pokud známe parametry α a β_i a hodnoty $X_1, X_2 \dots X_k$ konkrétní osoby, můžeme tento model využít na odhad pravděpodobnosti, že u sledované osoby nastane během nějakého předem definovaného časového intervalu sledovaná událost.

Alternativním způsobem zápisu logistického modelu je *logit tvar*, který získáme transformováním modelu. Logit transformaci označujeme $logit\ P(\mathbf{X})$ a má tvar

$$logit\ P(X) = \ln\left[\frac{P(X)}{1 - P(X)}\right]$$

kde

$$P(X) = \frac{1}{1 - e^{-(\alpha + \sum \beta_i \cdot X_i)}}$$

Po úpravách

$$\ln OR = logit\ P(X) = \ln\left[\frac{P(X)}{1 - P(X)}\right] = \ln\left[e^{(\alpha + \sum \beta_i \cdot X_i)}\right] = \alpha + \sum \beta_i \cdot X_i$$

Pro zjednodušení se logistický model vyjadřuje většinou v logit tvaru.

3.2.1 Odds ratio

Efekt rizikového faktoru výstupní proměnné měří odds ratio (OR). Logistický koeficient β_i interpretujeme jako změnu v logaritmu OR spojenou s jednotkovou změnou nezávislé proměnné. Z toho e^{β_i} je hodnota, o kterou se změní OR při změně i-té nezávislé proměnné o jednu měrnou jednotku. Pokud je β_i kladné, potom e^{β_i} bude větší jak 1, tzn. OR se zvýší.

OR je poměr dvou šancí na nastání události

$$OR = \frac{šance_1}{šance_0}$$

ve kterých index vyjadřuje dvě osoby nebo dvě skupiny osob, které porovnáváme.

OR všeobecně porovnává dvě skupiny osob definované souborem proměnných $\mathbf{X} = (X_1, X_2, \dots, X_k)$. Nechť $\mathbf{X}_1 = (X_{11}, X_{12}, \dots, X_{1k})$ vyjadřuje soubor proměnných, specifikujících skupinu 1 a $\mathbf{X}_0 = (X_{01}, X_{02}, \dots, X_{0k})$ soubor proměnných, specifikujících skupinu 0.

$$OR_{x_1, x_0} = \frac{\dfrac{P(X_1)}{1 - P(X_1)}}{\dfrac{P(X_0)}{1 - P(X_0)}} = \frac{e^{(\alpha + \sum \beta_i \cdot X_{1i})}}{e^{(\alpha + \sum \beta_i \cdot X_{0i})}} = e^{\sum \beta_i (X_{1i} - X_{0i})}$$

Takto získáme všeobecný vztah pro OR logistického modelu porovnávajícího dvě skupiny osob, specifikované jako $\mathbf{X}_1$ a $\mathbf{X}_0$. Tento vztah obsahuje β_i, ale neobsahuje α.

S využitím matematického pravidla $e^{a+b} = e^a \times e^b$ můžeme tento vztah zapsat alternativním způsobem

$$OR_{x_1,x_0} = \prod_{i=1}^{k} e^{\beta_i(X_{1i}-X_{0i})}$$

tudíž vztah pro OR vyjadřuje, jakou částí přispívají jednotlivé proměnné do modelu. Jednotlivé proměnné přispívají do vztahu pro OR multiplikativním způsobem. Například pokud $e^{\beta_2(X_{12}-X_{02})} = 3$ a $e^{\beta_5(X_{15}-X_{05})} = 4$, potom společný vliv těchto dvou proměnných na OR bude $3 \times 4 = 12$.

3.2.2 Ukázka aplikace

Pro tuto ukázku byla zvolená stejná data jako v předešlé ukázce, tedy datový soubor *collisions.sf6*. Jsou zde údaje z 58 dopravních nehod (boční náraz), tedy *n = 58*. Hodnota proměnné E je binární, kvantifikující, zda dopravní nehoda byla smrtelná ($E = 1$) nebo ne ($E = 0$).

Jsou zde brány v potaz tyto faktory:

- věk řidiče,
- akcelerace,
- rychlost.

Následující tabulka 3 ukazuje část tohoto datového souboru:

Tabulka 3: Ukázka dat ze souboru collisions.sf6 ze Statgraphics Centurion XV

Věk	Akcelerace	Rychlost	Událost E
22	50	98	0
21	49	160	0
40	50	134	1
43	50	142	1
23	51	118	0
58	51	143	1
29	51	77	0
29	51	184	0
47	51	100	1
…	…	…	…

Závisle proměnná E se rovná 1 pokud událost nastala a 0 pokud událost nenastala. Nezávislé náhodné veličiny jsou: věk osoby, která zapříčinila nehodu, akcelerace a rychlost, při které se nehoda stala.

Logistický model má následující formu:

$$P(X) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i \cdot X_i)}}$$

Výstupy ze systému Statgraphics Centurion XV

Estimated Regression Model (Maximum Likelihood)

		Standard	*Estimated*
Parameter	*Estimate*	*Error*	*Odds Ratio*
CONSTANT	-15,7207	5,33566	
vek	0,164496	0,0417988	1,1788
zrychlení	0,131833	0,0564584	1,14092
rychlost_100=0	-0,653095	0,878665	0,520433

Analysis of Deviance

Source	*Deviance*	*Df*	*P-Value*
Model	33,905	3	0,0000
Residual	44,7673	54	0,8105
Total (corr.)	78,6723	57	

Percentage of deviance explained by model = 43,0965
Adjusted percentage = 32,9278

Likelihood Ratio Tests

Factor	*Chi-Squared*	*Df*	*P-Value*
vek	30,3302	1	0,0000
zrychlení	6,98415	1	0,0082
rychlost_100	0,56427	1	0,4525

Obrázek 6: Výstup ze systému Statgraphics Centurion XV

Obrázek 6 poskytuje tyto důležité informace:

- Odhadnutý logistický regresní model (Estimated Regression Model) ve tvaru:

$$\boldsymbol{P}(\boldsymbol{E}=\mathbf{1}) = \frac{1}{1+e^{-(\alpha+\sum \beta_i \cdot X_i)}}$$

$$= \frac{1}{1+e^{-(-15{,}7207+0{,}164496 \cdot věk+0{,}131833 \cdot zrychlení+(-0{,}653096) \cdot rychlost)}}$$

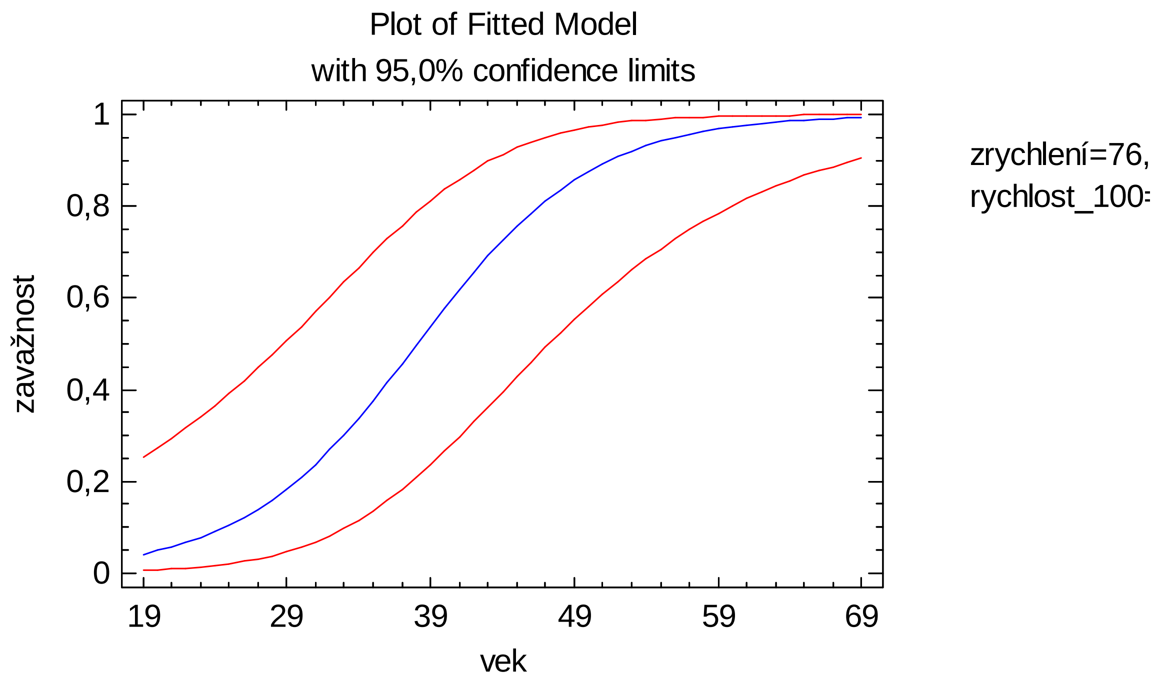
- Potvrzení významnosti tohoto modelu (p-Value 0,0000), který vysvětluje až 43,0965 variability hodnot závisle proměnné.

- Odds ratio pro nezávislé proměnné, vypočítaný z koeficientů $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ takto:

$$odds\ ratio = e^{(\widehat{\beta}_j)}.$$

- S každým rokem věku řidiče se odds ratio zvýší 1,1788 násobně.

- Test významnosti (Likelihood Ratio Tests) pro každý koeficient modelu potvrzuje významný rozdíl tohoto koeficientu od hodnoty 0, nebo významnou závislost na příslušné nezávislé proměnné. Malá P-hodnota dokazuje, že nastání události E významně ovlivňují faktory věk a akcelerace.

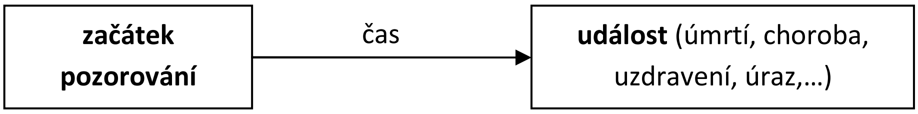


Obrázek 7: Logistická křivka (výstup ze Statgraphics Centurion XV)

4 ANALÝZA PŘEŽITÍ

Následující kapitola 4 včetně podkapitol 5.1 a 5.2 jsou zpracovány podle [10], [11], [12] a [14].

Analýza přežití je soubor statistických postupů, pro které je výslednou proměnnou čas po nastání události.



Obrázek 8: Analýza přežití [14]

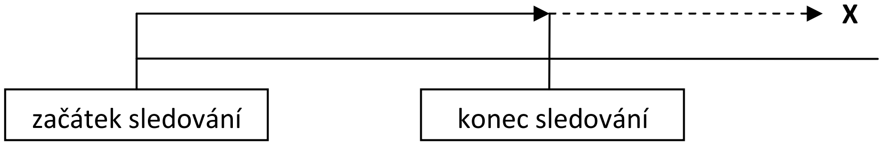
Hlavní cíle analýzy přežití jsou:

- odhad a interpretace funkce přežití a funkce rizika z údajů o přežívajících osobách,
- porovnání funkcí přežití mezi různými skupinami osob,
- odhad vztahů mezi vysvětlujícími proměnnými.

Při analýze dat se nejčastěji zaměřujeme na funkci přežití $S(t)$. Tato funkce vyjadřuje procento osob, které přežívá v čase t. Funkce přežití je nerostoucí funkce, pro kterou platí $S(0) = 1, S(\infty) = 0$.

4.1 Způsoby cenzorování

Nejčastějším problémem v analýze přežití je to, že ne vždy umíme zjistit informaci o délce přežití, protože událost může nastat až po ukončení sledování. V takových případech poznáme jen cenzorované (useknuté) pozorování. Cenzorovaná pozorování obsahují jen částečnou informaci o náhodné proměnné.



Obrázek 9: Cenzorované (useknuté) pozorování [14]

Necenzorovaná pozorování se obvykle označují jako T_i a cenzorovaná jako T_i^+. Nechť $T_1, T_2, \dots, T_n$ jsou nezávislé, identicky rozdělené náhodné proměnné.

- **I. typ cenzorování**

Nechť t_c je fixní číslo. Namísto pozorovaných $T_1, T_2, \dots, T_n$ pozorujeme náhodné veličiny $Y_1, Y_2, \dots, Y_n$, kde

$$Y_i = \begin{cases} T_i \; ak \; T_i \leq t_c \\ t_c \; ak \; t_c < T_i \end{cases}$$

- **II. typ cenzorování**

Nechť $r < n$ je fixní a nechť $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ jsou seřazené statistiky $T_1, T_2, \dots, T_n$. Pozorování ukončíme po r událostech, takže získáme $Y_{(1)}, Y_{(2)}, \dots, Y_{(r)}$:

$$Y_{(1)} = T_{(1)}$$
$$\vdots$$
$$Y_{(r)} = T_{(r)}$$
$$Y_{(r+1)} = T_{(r)}$$
$$\vdots$$
$$Y_{(n)} = T_{(r)}$$

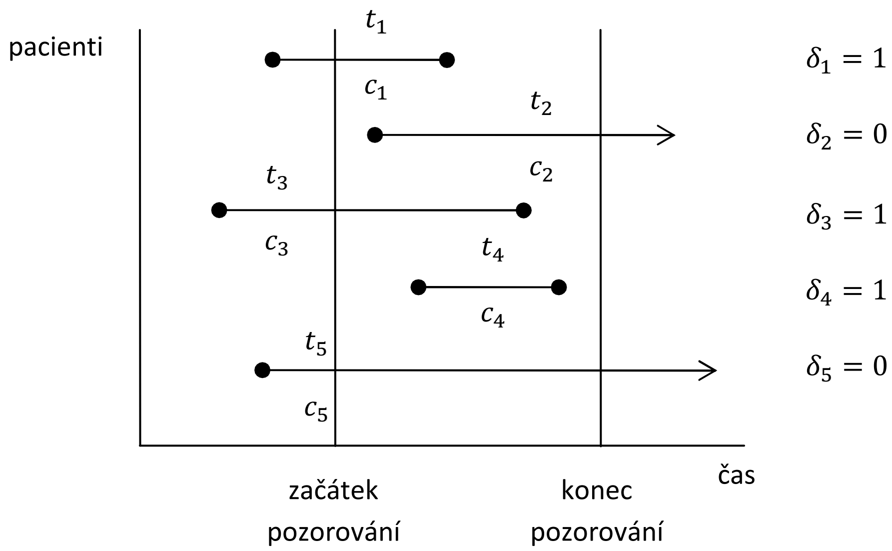
Cenzorování I. i II. typu vzniká v technických aplikacích. Například, když testujeme žárovky, začneme je sledovat v čase $t = 0$ a zapíšeme čas pro jejich vypálení. Některé žárovky mohou vydržet déle, než očekáváme. Proto v předem určeném čase t_c sledování ukončíme (cenzurování I. typu) anebo čekáme, dokud se nevypálí $(r/n) \cdot 100\%$ žárovek (cenzorování II. typu).

- **Náhodné cenzorování**

Nechť $C_1, C_2, \dots, C_n$ jsou nezávislé, identicky rozdělené náhodné proměnné, kde C_i je cenzorovaný čas přiřazený k T_i. Pozorujeme $(Y_1, \delta_1), \dots, (Y_n, \delta_n)$, kde

$$Y_i = \min(T_i, C_i)$$
$$\delta_i = \begin{cases} 1 \; ak \; T_i \leq C_i \; t.j. T_i \; je \; necenzorovné \\ 0 \; ak \; T_i > C_i \; t.j. T_i \; je \; cenzurovné \end{cases}$$

kde náhodné proměnné $\delta_1, \delta_2, \dots, \delta_n$ obsahují informaci o cenzorování, což je patrné z obrázku 10.



Obrázek 10: Ukázka náhodného cenzorování (upraveno z [11])

Náhodné cenzorování vzniká v lékařských aplikacích, kde pacienti vstupují do studie v různých časech. Cenzorování se vyskytuje v různých formách:

- úmrtí pacienta,
- ztráty na zkoumání – pacient se přestěhuje a nikdy ho už neuvidíme,
- vypadnutí ze studie – pacient odmítne pokračovat v léčbě, případně z důvodu vedlejších účinků musí být léčba pozastavena,
- omezení studie.

Důležitým předpokladem při náhodném cenzorování je to, že T_i nezávisí od C_i. Vypadnutí ze studie však často závisí na průběhu terapie, a proto může být závislost T_i od C_i oprávněná.

Předpokládejme, že ze souboru 100 sob bylo cenzorováno celkem 24 osob. Pokud bychom cenzory vyřadili od začátku sledování, vycházeli bychom pouze ze 76 pozorování. Tím by se ztratila část informace, kterou poskytují cenzory do doby, než „zmizí ze sledování“. Při stejných absolutních počtech zemřelých d_i je jejich relativní podíl vzhledem k nižší hodnotě n_0 vyšší, a tedy výsledná kumulativní pravděpodobnost přežití nižší.

Správné řešení problému cenzorování našli Kaplan a Meier. Počet cenzorovaných údajů q_i v čase t_i zohlednili tak, že počet n_{i+1} odpovídá počtu n_i, zmenšenému nejen o počet zemřelých d_i, ale i o počet cenzorovaných údajů w_i.

Tabulka 4: Cenzorování podle Kaplana a Meiera [14]

i	čas	cenzory se nevyskytly				cenzory byly vyřazené				cenzory byly respektované				
	t_i	n_i	d_i	p_i	$\prod \cdot p_i$	n_i	d_i	p_i	$\prod \cdot p_i$	n_i	d_i	w_i	p_i	$\prod \cdot p_i$
1	0	100	0	1,00	1,00	76	0	1,00	1,00	100	0	0	1,00	1,00
2	10	100	7	0,93	0,93	76	7	0,91	0,91	100	7	3	0,93	0,93
3	15	93	0	1,00	0,93	69	0	1,00	0,91	90	0	4	1,00	0,93
4	20	93	13	0,86	0,80	69	13	0,81	0,74	86	13	13	0,85	0,79
5	30	80	1	0,99	0,79	56	1	0,98	0,72	60	1	4	0,98	0,78
6	32	79	0	1,00	0,79	55	0	1,00	0,72	55	0	0	1,00	0,78

Kumulativní pravděpodobnost přežití Kaplan-Meierova přístupu je nižší než v prvním a vyšší než v druhém případě. V prvním případě zůstali v souboru všechny cenzorované pozorování až do konce sledování, v druhém se ztratila informace z jejich částečného trvání v souboru. Osud cenzorovaných pozorování po skončení sledování není známý, je však velmi málo pravděpodobné, že by všechny osoby okamžitě zemřely, proto je logické s nimi uvažovat.

4.2 Neparametrické jednorozměrné metody odhadu křivky přežití

Při tomto odhadu křivky přežití sledované období rozdělíme na menší časové intervaly. Na odhad kumulativní pravděpodobnosti přežití použijeme všechny částečné pravděpodobnosti z jednotlivých intervalů, čímž využijeme všechna dostupná data.

Abychom mohli tabulky přežití použít, předpokládáme, že podmínky přežití se v průběhu sledování nebudou měnit, tj. osoba sledovaná dnes se bude chovat stejně, jako by se chovala například před rokem. Stejně tak musíme předpokládat, že cenzorovaná pozorování se neodlišují od pozorování, při kterých událost nastala.

Údaje uspořádáme do matice, která má n řádků, kde každý řádek charakterizuje právě jednu osobu ve tvaru (#, t_j, δ_j, x_j, …, x_{jp}). První sloupec # obsahuje identifikační číslo a jiné informace, týkající se dané osoby. Druhý, t_j, obsahuje informace o čase přežití osoby, vyjadřující, kdy událost nastala, anebo kdy byl údaj cenzorovaný. Na odlišení osob, u kterých nastala událost, od osob cenzorovaných, využíváme dichotomickou proměnnou označující cenzorování s variantami 0 – údaj byl cenzorovaný, 1 – nastala událost. Hodnoty x_i představují vysvětlující proměnné, které využíváme na modelování délky přežití.

Celkový čas rozdělíme do intervalů I_1, I_2, …I_k, které jsou obvykle stejné délky. Použijeme označení:

n_i počet přežívajících na začátku období I_i,

d_i počet úmrtní během období I_i,

w_i počet cenzur během období I_i,

q_i pravděpodobnost úmrtí během intervalu I_i,

p_i podmíněná pravděpodobnost přežití období I_i za předpokladu, že osoba byla naživu na začátku období I_i.

4.2.1 Aktuárská metoda

Problémem je, že cenzorované osoby mohou ještě v daném intervalu, ve kterém byly cenzorované, zemřít. V aktuárské metodě předpokládáme, že cenzorované osoby byly v tomto intervalu jen polovinu času.

Při použití aktuárské metody dělíme celkovou délku sledování na menší intervaly. V každém intervalu zaznamenáme počet pozorovaných osob na začátku intervalu a počet událostí a cenzor v daném intervalu, ze kterých odhadneme pravděpodobnost nastání události během tohoto intervalu. Z odhadů pravděpodobností v jednotlivých intervalech odhadneme kumulativní pravděpodobnosti. Cenzorované údaje se tedy použijí ve všech intervalech, během kterých byly pozorované.

Při použití aktuárské metody se rozhodneme pro délku časového intervalu pro výpočet odhadu křivky přežití. Když počítáme například roční míry, data uspořádáme do 12 měsíčních intervalů.

Kumulativní pravděpodobnost přežití $S(t_k)$ můžeme rozložit na podmíněné pravděpodobnosti

$$S(t_k) = P(T > t_k) = P(T > t_1) \times P(T > t_2 | T > t_1) \times \ldots \times P(T > t_k | T > t_{k-1})$$
$$= p_1 \cdot p_2 \cdot \ldots \cdot p_k$$

kde

$$p_i = P(T > t_i | T > t_{i-1})$$

Aktuárskou metodou odhadujeme jednotlivé p_i samostatně a pomocí nich vypočítáme $S(t_k)$.

Za předpokladu, že interval I_i neobsahuje žádné cenzorované pozorování, můžeme na odhad P_i použít vztah $1 - \frac{d_i}{n_i}$. Pokud se v intervalu I_i vyskytují i cenzory, předpokládáme, že cenzurovaná pozorování byla vystavena riziku jen polovina intervalu.

Potom

$$n_i' = n_i - \tfrac{1}{2} \cdot w_i$$
$$\hat{q}_i = \frac{d_i}{n_i'} = \frac{d_i}{n_i - \frac{1}{2} \cdot w_i}$$
$$\hat{p}_i = 1 - \hat{q}_i$$

Aktuárský odhad funkce přežití je potom

$$\hat{S}(t_k) = \prod_{i=1}^{k} \hat{p}_i = \prod_{i=1}^{k} \left(1 - \frac{d_i}{n_i - \frac{1}{2} \cdot w_i}\right)$$

4.2.2 Kaplan – Meierova metoda

Pokud jsou údaje o přežívání zapsané detailněji, přesnost zachováme pouze při použití menších intervalů. Kaplan – Meierův odhad je založený na myšlence početnějších intervalů, které mohou být libovolně velké. Je podobný aktérskému odhadu až na to, že délky intervalů mohou být různé. Kaplan – Meierovou metodou odhadujeme křivku přežití v každém časovém okamžiku, ve kterém událost nastala, tj. pro každý necenzurovaný čas.

Kaplan – Meierovy křivky vycházejí z tabulky tříděných časů událostí pro každou skupinu. Nechť $t_{(1)} < t_{(2)} < \cdots < t_{(n)}$ jsou seřazené časy $t_1, t_2, \dots, t_n$ a nechť $\delta_{(i)}$ je hodnota δ, přiřatená k $t_{(i)}$, tj. $\delta_{(i)} = \delta_j$ když $t_{(i)} = t_j$.

Tabulka 5: Tabulka tříděných časů událostí

$t_{(j)}$	n_j	d_j	w_j
$t_{(0)} = 0$	n_0	$d_0 = 0$	w_0
$t_{(1)}$	n_1	d_1	w_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_{(k)}$	n_k	d_k	w_k

Tabulka 5 začíná časem přežití 0. Hodnota n_j vyjadřuje počet osob vystavených riziku na začátku j-tého období, d_j vyjadřuje počet událostí a w_j počet cenzor během j-tého období.

Křivku přežití odhadujeme v každém časovém okamžiku, ve kterém událost nastala a je konstantní v celém intervalu až po následující čas nastání události. Kumulativní pravděpodobnost přežití v konkrétním čase vypočítáme vynásobením všech individuálních pravděpodobností přežití až po tento čas.

Pokud soubor dat obsahuje cenzorované údaje (když je některé w_j různé od nuly), musíme použít alternativní způsob výpočtu pravděpodobnosti přežití – „Kaplan – Meierův přístup“, který dostaneme z podmíněných pravděpodobností $P(T > t_{(j)} \mid T \geq t_{(j)})$. Podstata Kaplan – Meierova přístupu spočívá v tom, že pozorování nemůžeme použít v čase $t_{(j)}$ tehdy, když před časem $t_{(j)}$ nastala událost anebo bylo pozorování cenzurované. Kaplan – Meierův vztah má tvar

$$\hat{S}(t_{(j)}) = \hat{p}(T > t_{(0)} \mid \mathrm{T} \geq \mathrm{t}_{(0)}) \cdot \hat{P}(T > t_{(1)} \mid \mathrm{T} \geq \mathrm{t}_{(1)}) \ldots \ldots \hat{P}(T > t_{(j)} \mid T \geq t_{(j)})$$

Pokud součin prvních *j-1* podmíněných pravděpodobností $\prod_{i=1}^{j-1} \hat{p}\left(T > t_{(i)} \middle| T \geq t_{(i)}\right)$ nahradíme kumulativní pravděpodobností $\hat{S}\left(t_{(j-1)}\right)$, můžeme uvedený vztah vyjádřit ve tvaru

$$\hat{S}\left(t_{(j)}\right) = \hat{S}\left(t_{(j-1)}\right) \cdot \hat{p}\left(T > t_{(j)} \middle| T \geq t_{(j)}\right)$$

Odhad pravděpodobnosti přežití v čase $t_{(0)}$ je vždy rovný hodnotě jedna. Kaplan-Meierovy pravděpodobnosti přežití udávají pravděpodobnost přežití na základě předchozí pravděpodobnosti přežití v čase $t_{(j-1)}$, násobeného podmíněnou pravděpodobností přežití $\hat{p}\left(T > t_{(j)} \middle| T \geq t_{(j)}\right)$.

Z odhadu

$$\hat{q}_i = \frac{d_i}{n_i}$$

dostaneme

$$\hat{p}_i = 1 - \hat{q}_i = \begin{cases} 1 - \dfrac{d_i}{n_i} \; pokud \; \delta_{(i)} = 1 \; (necenzurované) \\ \quad 1 \quad \; pokud \; \delta_{(i)} = 0 \; (cenzurované) \end{cases}$$

Z toho Kaplan – Meierův odhad

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \hat{p}_i = \prod_{t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right)^{\delta_{(i)}}$$

Při velkých souborech dostaneme pomocí Kaplan – Meierovy metody poměrně rozsáhlé tabulky přežití, proto se v takovýchto případech upřednostňuje aktuárská metoda.

4.3 Coxův regresní model úmrtnosti

Celá kapitola 4.3 je zpracována dle [10], [14] a [16].

4.3.1 Coxův model proporcionálního rizika

Přežívání není jen funkcí času, ale i různých rizikových faktorů. Coxův model dovoluje odhadnout kromě výchozí funkce přežití i hodnoty regresních koeficientů jednotlivých rizikových faktorů.

Coxův model se upřednostňuje před logistickým modelem tehdy, když máme k dispozici i cenzorované informace o čase přežití. Coxův model využívá časy přežití, ale i cenzorované údaje. Logistický model využívá pouze dichotomickou proměnnou, která vyjadřuje, jestli událost nastala a ignoruje časy přežití a cenzorované údaje.

Coxův model proporcionálního rizika vyjadřuje riziko v čase t pro osobu s určitými hodnotami vysvětlujících proměnných (rizikových faktorů), označených vektorem **X**.

$$h(t,\mathbf{X}) = h_0(t) \cdot e^{\sum_{i=1}^{p} \beta_i \cdot x_i}; \qquad \mathbf{X} = \left(X_1, X_2, \dots, X_p\right)$$

Riziko se v čase t skládá ze dvou částí:

První část $h_0(t)$, která obsahuje t, ale ne $\mathbf{X}$, se nazývá výchozí funkce přežití. Riziko $h_0(t)$ je neznámým parametrem, který se musí odhadnout.

Druhá část, která obsahuje vektor $\mathbf{X}$, ale ne t, je exponenciální vyjádření $e^{\sum_{i=1}^{p} \beta_i \cdot X_i}$. Proměnnou X může být například pohlaví, věk, váha anebo kuřák/nekuřák. Účinek vysvětlujících proměnných spočívá v násobení výchozího rizika faktorem $e^{\beta X}$.

Předchozí vztah můžeme přepsat do tvaru

$$h(t,X) = e^{\alpha(t) + \sum_{i=1}^{p} \beta_i \cdot X_i}$$

kde $\alpha(t) = \ln h_0(t)$. Za předpokladu specifikace funkce $\alpha(t)$ bychom dostali následující parametrické modely

$\alpha(t) = \alpha$ exponenciální model

$\alpha(t) = \alpha \cdot t$ Gompertzův model

$\alpha(t) = \alpha \cdot \ln t$ Weibullův model

Nejdůležitějším předpokladem modelu je to, že pokud má první osoba v čase t_1 riziko nastání události n-násobně vyšší v porovnání s druhou osobou, potom i v čase t_2 má první osoba n-násobně vyšší riziko úmrtí v porovnání s druhou osobou. Předpoklad proporcionality funkcí rizika nemusí platit například při proměnné charakterizující způsob terapie (medikamentózní a chirurgická léčba).

Coxův model má následující vlastnosti:

- pokud jsou všechny X_i nulové, exponenciální část Coxova modelu $e^{\sum_{i=1}^{p} \beta_i \cdot X_i}$ je rovna 1 a původní vztah se redukuje na výchozí funkci rizika $h_0(t)$,
- pro osobu s konstantními hodnotami vysvětlující proměnné X_i je funkce rizika $h_0(t) \times C(\mathbf{X})$, kde $C(\mathbf{X})$ je konstanta, závislá na hodnotách $\mathbf{X}$. Konstanta $C(\mathbf{X})$ je vždy kladná,
- pro jakékoliv dvě osoby je *poměr funkcí rizika*

$$\frac{h_0(t) \times C(\mathbf{X}_{(1)})}{h_0(t) \times C(\mathbf{X}_{(2)})} = \frac{C(\mathbf{X}_{(1)})}{C(\mathbf{X}_{(2)})}$$

Poměr dvou funkcí rizika nezávisí na čase t. Funkce rizika jsou tedy proporcionální v čase. Proto se takovýto model nazývá modelem proporcionálního rizika.

Východiskové riziko Coxova modelu $h_0(t)$ není specifikováno. Coxův model je neparametrický model. I když není výchozí riziko specifikované, můžeme odhadnout regresní koeficienty, poměr funkcí rizika (hazard ratio) a upravené křivky přežití.

Výsledkem Coxova modelu proporcionálního rizika je velmi blízká aproximace výsledků vhodného parametrického modelu. Použitím Coxova modelu dostaneme řešení porovnatelné s vhodným parametrickým modelem.

I když jsme si jistí konkrétním modelem, je lepší použít parametrický model. Pokud však nevíme, jaký model sestrojit, Coxův model je „bezpečnou“ volbou a nemusíme se obávat, že jsme zvolily špatné parametry modelu.

4.3.2 Hazard ratio (poměr rizika)

Hlavním cílem Coxova modelu je měření efektu rizikových faktorů. Na odměření tohoto efektu slouží hazard ratio (HR). Hazard ratio vysvětluje vztah mezi rizikovým faktorem a délkou přežití s ohledem na kontrolní proměnné.

Pokud HR = 1, potom určitý rizikový faktor nemá na čas přežití žádný efekt. Pokud HR > 1, potom skupina vystavená tomuto rizikovému faktoru má HR-násobně vyšší riziko v porovnání se skupinou, která není vystavená rizikovému faktoru. Pokud je HR vyšší pro skupinu vystavenou rizikovému faktoru v čase t, potom pravděpodobnost přežití skupiny vystavené rizikovému faktoru je nižší.

HR je definované jako riziko jedné osoby dělené rizikem pro jinou osobu

$$\widehat{HR} = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})}$$

kde $\mathbf{X}^*$ označuje soubor proměnných první osoby, $\mathbf{X}^* = \left(X_1^*, X_2^*, \dots, X_p^*\right)$

$\mathbf{X}$ označuje soubor proměnných druhé osoby, $\mathbf{X} = (X_1, X_2, \dots, X_p)$

Pokud platí, že $\widehat{HR} > 1 \Rightarrow \hat{h}(t, \mathbf{X}^*) > \hat{h}(t, \mathbf{X})$. Potom skupina se souborem proměnných $\mathbf{X}^*$ je vystavená vyššímu riziku v porovnání se skupinou $\mathbf{X}$.

Vztah $\widehat{HR} = \frac{\hat{h}(t,\mathbf{X}^*)}{\hat{h}(t,\mathbf{X})}$ pro HR upravíme do následujícího tvaru

$$\widehat{HR} = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})} = \frac{\hat{h}_0(t) \cdot e^{\sum_{i=1}^{p} \hat{\beta}_i \cdot X_i^*}}{\hat{h}_0(t) \cdot e^{\sum_{i=1}^{p} \hat{\beta}_i \cdot X_i}} = e^{\sum_{i=1}^{p} \hat{\beta}_i \cdot (X_i^* - X_i)}$$

a dostaneme upravené křivky přežití s použitím Coxova modelu proporcionálního rizika.

Další důležitou charakteristikou analýzy přežití jsou křivky přežití. Nejjednodušeji můžeme křivky přežití odhadnout s použitím Kaplan-Meierovy metody. Pokud použijeme na odhad dat přežití Coxův model, jednu z vysvětlujících proměnných budeme považovat za primární rizikový faktor a ostatní považujeme za kontrolní regresory. Hlavním cílem analýzy je vysvětlení vztahu mezi primárním rizikovým faktorem a časem do nastání události, s kontrolou ostatních regresorů. Takovéto křivky se potom nazývají rozšířené křivky přežití.

Odhad křivky přežití pro konkrétní osobu vypočítáme tak, že východiskovou křivku přežití umocníme hodnotou $e^{\beta X}$. Pokud hodnota tohoto exponentu je větší než 1, potom výsledný čas přežití bude kratší než východiskový čas přežití. Takže kladná hodnota příslušného koeficientu hovoří o nepříznivém vlivu daného rizikového faktoru ve smyslu zkracování doby přežívání, a naopak, záporná hodnota koeficientu svědčí o příznivém vlivu příslušného „rizikového“ (vlastně protirizikového) faktoru, tedy o prodlužování doby přežívání.

Funkce rizika Coxova modelu

$$\hat{h}(t, X) = \hat{h}_0(t) \cdot e^{\sum_{i=1}^{p} \hat{\beta}_i \cdot X_i},$$

může být konvertovaná na vztah funkce přežití, který je základem na výpočet rozšířených křivek přežití

$$\hat{S}(t, X) = \left[\hat{S}_0(t)\right]^{e^{\sum_{i=1}^{p} \hat{\beta}_i \cdot X_i}}.$$

Pokud chceme získat upravené křivky přežití, které jsou přizpůsobené pro všechny regresory, potom použijeme

$$\hat{S}(t, X) = \left[\hat{S}_0(t)\right]^{e^{\sum_{i=1}^{p} \hat{\beta}_i \cdot \bar{X}_i}}.$$

4.3.3 Ukázka aplikace

Pro tuto ukázku byla data vybrána opět z programového balíku Statgraphics Centurion XV, konkrétně soubor *coxph.sf6*. Soubor obsahuje data ze studie o 36 pacientech ($n = 36$) se zhoubným nádorem na ledvině. Pacienti byli rozděleni do třech věkových skupin. Někteří pacienty v každé skupině podstoupili nefrektomii (chirurgické odstranění ledviny), někteří nikoliv. Část souboru ukazuje následující tabulka 6:

Tabulka 6: Ukázka z části souboru coxph.sf6

Doba přežití [měsíce]	**Cenzory**[1)]	**Nefrektomie**[2)]	**Věková skupina**[3)]
9	0	0	1
6	0	0	1
21	0	0	1
15	0	0	2
8	0	0	2
17	0	0	2
12	0	0	3
104	1	1	1
9	0	1	1
56	0	1	1
35	0	1	1
35	0	1	1
52	0	1	1
68	0	1	1
77	1	1	1
84	0	1	1
8	0	1	1
38	0	1	1
…	…	…	…

[1)] Cenzory

1 žijí

0 zemřeli

[2)] Nefrektomie

1 podstoupili chirurgické odstranění ledviny

0 nepodstoupili chirurgické odstranění ledviny

[3)] Věková skupina

1 < 60 let

2 60 – 70 let

3 > 70 let

Coxův model proporcionálního rizika jsme našli použitím procedury *Cox proportional Hazards* statistického programového balíku Statgraphics Centurion XV. Výstup Coxovy regrese 1. stupně, kdy neuvažujeme interakci faktorů, obsahuje obrázek 11.

Výstupy ze Statgraphics Centurion XV:

Cox Proportional Hazards - prožití

Dependent variable: prožití
Censoring: cenzor
Factors:
nefrektomie
vek

Number of uncensored values: 32
Number of right-censored values: 4

Estimated Regression Model

		Standard	*Lower 95,0%*	*Upper 95,0%*
Parameter	*Estimate*	*Error*	*Conf. Limit*	*Conf. Limit*
nefrektomie=1	-1,41108	0,377288	-2,15056	-0,67161
vek=2	0,0124456	0,330352	-0,635033	0,659924
vek=3	1,34132	0,448089	0,463082	2,21956

Log likelihood = -82,7542

Likelihood Ratio Tests

Factor	*Chi-Squared*	*Df*	*P-Value*
nefrektomie	6,66386	1	0,0098
vek	4,73827	2	0,0936

Obrázek 11: Výstup ze Statgraphics Centurion XV

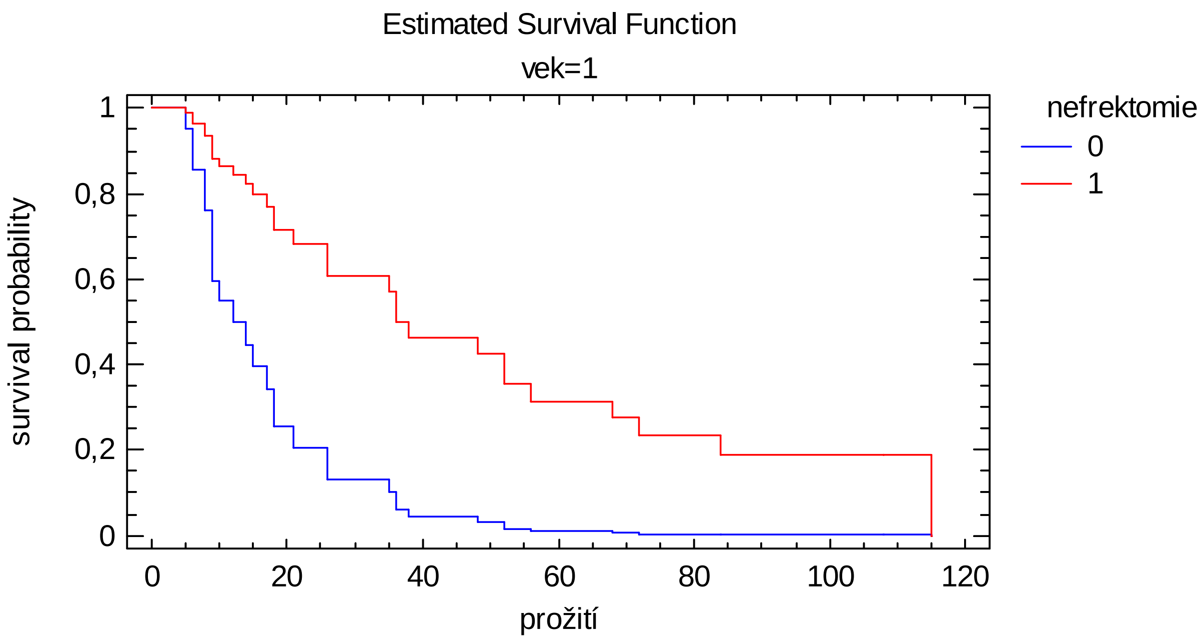
- Souhrn vstupních dat pro Coxův model proporcionálního rizika, zahrnujících počet pozorování $\boldsymbol{n}$, vhodných pro model.
- Část výstupu Estimated Regression Model (odhadnutý regresní model) obsahuje maximálně věrohodné odhady koeficientů $\widehat{\boldsymbol{\beta}}_i$ Coxova modelu, standardní chyby odhadů a 95% intervalové odhady.
- Likelihood Ratio Tests (test poměrů věrohodnosti) udává, že jsou koeficienty $\widehat{\boldsymbol{\beta}}_i$ významně různé od hodnoty 0. Je zde zobrazena oboustranná P-hodnota. Malá P-hodnota značí statistický významnou proměnnou.

Odhadnutý model na základě našich dat je následující:

$$h_x(t) = exp[-1.41108\ (Nephrectomy = 1) + 0.0124456(Age = 2) + 1.34132(Age = 3)]h_0(t)$$

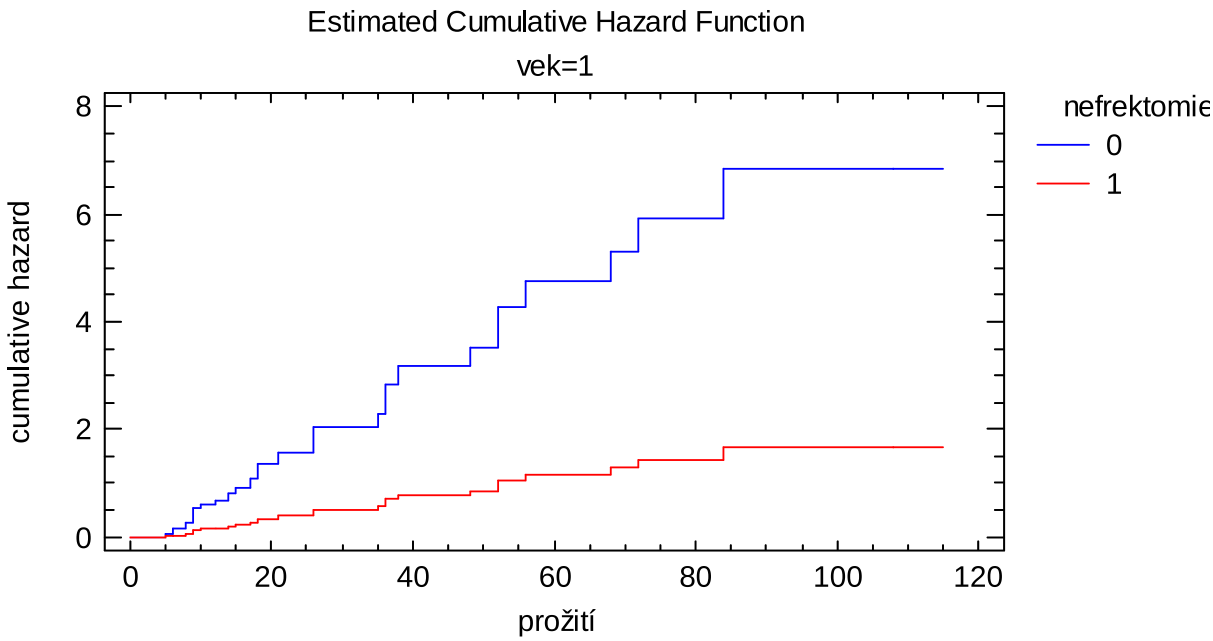
Základní funkce $h_0(t)$ je Coxova funkce rizika pro případ, že jsou všechny vysvětlující proměnné nulové.

$Nephrectomy = 1$, $Age = 2$ a $Age = 3$ jsou dummy proměnné, které nabývají hodnoty 0, při variantě ne a 1 při variantě ano. Test významnosti koeficientů ukazuje významné prodloužení života u těch, kteří podstoupili chirurgické odstranění ledviny, ale efekt ohledně věku zde není na 5% hladině významnosti významný. Nutno podotknout, že efekt nefrektomie je negativní, ukazuje tedy na úspěšnost redukce funkce rizika, a tudíž se zlepšuje přežití.



Obrázek 12: Funkce přežití pro dvě různé skupiny pacientů

Obrázek 12 zobrazuje odhadnutou funkci přežití pro pacienty ze skupiny, kteří podstoupili nefrektomii (1) a kteří ji nepodstoupili (0) a jsou zařazeni do první věkové skupiny, tedy $Age = 1$. Z grafického znázornění je vidět jasné zlepšení v přežití pacientů, kteří operaci podstoupili. Analogický závěr lze udělat i na základě grafického zobrazení funkce rizika na obrázku 13. Obdobný výsledek bychom dostali i pro další dvě věkové kategorie.



Obrázek 13: Coxova funkce proporcionálního rizika pro dvě různé skupiny pacientů

Pokud zvolíme Coxův model 2. řádu, tedy do modelu zařídíme i interakce faktorů nefrektomie a věk, dostaneme výstup na obrázku 14. Z výsledků testu významnosti koeficientů modelu vyplývá nevýznamnost interakce těchto faktorů.

Cox Proportional Hazards - prožití

Dependent variable: prožití
Censoring: cenzor
Factors:
nefrektomie
vek

Number of uncensored values: 32
Number of right-censored values: 4

Estimated Regression Model

		Standard	*Lower 95,0%*	*Upper 95,0%*
Parameter	*Estimate*	*Error*	*Conf. Limit*	*Conf. Limit*
nefrektomie=1	-1,94365	0,562645	-3,04642	-0,840886
vek=2	0,00491685	0,730185	-1,42622	1,43606
vek=3	0,0650905	1,03264	-1,95885	2,08903
nefrektomie=1*vek=2	-0,0504223	0,819766	-1,65714	1,55629
nefrektomie=1*vek=3	2,00278	1,14729	-0,245857	4,25143

Log likelihood = -81,2395

Likelihood Ratio Tests

Factor	*Chi-Squared*	*Df*	*P-Value*
nefrektomie	5,42807	1	0,0198
vek	0,00322947	2	0,9984
nefrektomie*vek	3,0294	2	0,2199

Obrázek 14: Výstup ze Statgraphics Centurion XV

ZÁVĚR

Zkušenosti životní pojišťovny se opírají o dlouhodobá pozorování úmrtnosti v závislosti na základních rizikových faktorech, kterými jsou věk a pohlaví. Tyto zkušenosti se dále porovnávají se standartními úmrtnostními tabulkami a jsou základem pro výpočty v životním a důchodovém pojištění.

Cílem této diplomové práce bylo kvantifikovat vliv nejen základních, ale i dalších rizikových faktorů v analýze přežití, popsat a aplikovat vhodné metody této kvantifikace. Vhodnými metodami pro stanovený cíl jsou kromě úmrtnostních tabulek také logistický regresní model jako model přežití, Kaplan Meierovy křivky přežití a Coxův regresní model úmrtnosti. Pro jejich aplikace je nutno využít vhodný statistický programový balík, kterým je v této diplomové práci statistický programový systém Statgraphics Centurion XV.

První kapitola této práce se zabývá významem pojištění a jeho vývojem od počátků po současnost.

V kapitole 2 je charakterizován význam a možnosti modelování úmrtnosti. V životním pojištění jsou kryta dvě základní rizika a to riziko úmrtí a riziko dožití. Délka života, chápaná jako spojitá náhodná veličina, umožní vytvořit pravděpodobnostní model úmrtnosti. Tento model lze následně využit pro stanovení různých charakteristik úmrtnosti, především intenzity úmrtnosti.

Značná část práce se zabývá konstrukcí úmrtnostních tabulek, které stále představují nejčastější způsob modelování úmrtnosti. Zde se doba života chápe jako diskrétní náhodná proměnná.

Zásadní kapitola 3 se věnuje logistickému regresnímu modelu jakožto modelu přežití. Je uveden stručný teoretický výklad logistické regrese, která je vhodná na modelování jednostranné závislosti pravděpodobnosti úmrtí či přežití na kvantitativních i kategoriálních proměnných – sledovaných faktorech. Na konkrétním příkladu byl ukázán výpočet relativního rizika a pravděpodobnostního poměru. Sledovala se zde závislost úmrtí či přežití na rizikovém faktoru, kterým byla rychlost jízdy automobilu v okamžiku dopravní nehody. Rychlost jízdy byla stanovena jako kategoriální proměnná, přičemž jako jedna varianta byla zvolena rychlost do 100 km/h a druhá rychlost nad 100 km/h. Rychlost jízdy lze považovat za rizikový faktor, což potvrzují i naše data, přesto v regresním modelu závislost smrtelných následků srážky vozidel na rychlosti nebyla významná. V ukázce aplikace logistické regrese jsme uvažovali tři rizikové faktory a to věk řidiče, akceleraci a rychlost automobilu v okamžiku dopravní nehody. Zkoumala se zde závislost úmrtí či přežití na těchto faktorech. Pomocí procedury logistické regrese programového balíku Statgraphics Centurion XV byla zjištěna významná závislost na dvou faktorech, a to na věku řidiče a na akceleraci.

V poslední kapitole je popsána analýza přežití, jsou zde definovány různé metody odhadu křivky přežití a dále Coxův regresní model úmrtnosti, jehož aplikace byla ukázána na konkrétním příkladu. Zde byl vstupem soubor pacientů se zhoubným nádorem na ledvině. Zkoumala se závislost úmrtí či přežití na tom, zda pacient podstoupil nefrektomii, přičemž jsme brali v úvahu i věk pacientů. Závislost se v tomto případě potvrdila, protože pacienti, kteří podstoupili chirurgické odstranění ledviny, žili déle.

Problematika kvantifikace příčin úmrtí tvoří nedílnou součást demografické analýzy i aktuárských analýz v životních pojišťovnách a důchodových fondech. Diplomová práce potvrdila skutečnost, že úmrtnost závisí na různých rizikových faktorech a tuto skutečnost je v komerční životní pojišťovně nutno zohlednit při výpočtu pojistného. Kvantifikace rizikových faktorů umožní také účinnější prevenci a aktivní péči o lidské zdraví.

SEZNAM LITERATURY

Odborná literatura

[1] Sivašová, D.: *Aktuárska demografia v prostredí konkurenčného poistného trhu*. Bratislava: EKONÓM, 2008. 98 stran. ISBN 978-80-225-2509-1

[2] Čejková, V., Martinovičová, D.: *Pojišťovnictví*. Brno: Vysoké účení technické, 2003. 133 stran. ISBN 80-214-2404-4

[3] Ducháčková, E.: *Principy pojištění a pojišťovnictví*. 2.vyd. Praha: Ekopress, 2005. 178 stran. ISBN 80-86119-92-0

[4] Hradec, M., Zárybnická, J., Křivohlávek, V.: *Pojištění a pojišťovnictví*. Vysoká škola finanční a správní, Praha, 2007. 216 stran. ISBN 80-86754-48-0

[5] Cipra, T.: *Pojistná matematika – teorie a praxe*. Praha: Ekopress, 2006. 411 stran. ISBN 80-86929-11-6

[6] Cipra, T.: *Finanční a pojistné vzorce*. Praha: Grada Publishing, 2006. 374 stran ISBN 80-247-1633-X

[7] Koschin, F.: *Aktuárská demografie*. Praha: VŠE, 1997. 123 stran. ISBN 80-7079-112-8

[8] Čejková, V., Čámský, F., Řezáč, F., Šedová, J.: *Pojišťovnictví*. Brno: Masarykova univerzita v Brně, 1999. 179 stran. ISBN 80-210-1637-X

[9] Kalibová K.: *Úvod do demografie*. Praha: Univerzita Karlova v Praze – Nakladatelství Karolinum, 2005. 52 stran. ISBN 80-246-0222-9

[10] Fiala, T.: *Výpočty aktuárské demografie v tabulkovém procesoru*. Praha: Vysoká škola ekonomická v Praze, 2005. 177 stran. ISBN 80-245-0821-4

[11] Hendl, J.: *Přehled statistických metod zpracování dat*. Praha: Portál, s. r. o., 2006. 583 stran. ISBN 80-7367-123-9

[12] Benjamin, B., Pollard, S. H.: *The Analysis of Mortality and other actuarial statistics*. Institute of Actuaries, Scotland, London, 1993

[13] Kleinbaum, D. G.: *Logistic Regression*: *A Self – Learning Text*. Springer-Verlag, New York, 1996

[14] Kleinbaum, D. G.: *Survival analysis*: *A Self – Learning Text*. Springer-Verlag, New York, 1996

[15] SEKERKA, B.: *Matematické a statistické metody ve financování, cenných papírech a pojišťovnictví*. Praha: Profess Consulting, 2002. 397 stran. ISBN 80-7259-031-6

[16] Cox, D. R., Oakes, D.: *Analysis of Survival Data.* Chapman and Hall, London, 1988

[17] Cipra, T.: *Praktický průvodce finanční a pojistnou matematikou*. Praha: HZ Praha, 1995. 320 stran. ISBN 80-901918-0-0

Internetové zdroje

[18] Česká asociace pojišťoven: *Česká asociace pojišťoven* [online]. c ČAP 2007-2008. Dostupný z WWW: <http://www.cap.cz/ZobrazFolder.aspx?folder=Lists%2fMenu+Verejneho+webu%2fPr%c5%afvodce+poji%c5%a1t%c4%9bn%c3%adm%2fPoji%c5%a1t%c4%9bn%c3%ad+dle+druhu+rizika%2fPoji%c5%a1t%c4%9bn%c3%ad+osob>

[19] Finanční vzdělávání: *Finanční vzdělávání* [online]. c FinančníVzdělávání 2007. Dostupný z WWW: <http://www.financnivzdelavani.cz/webmagazine/page.asp?idk=394>

[20] EDU server: *Pojišťovnictví - Ing. Vojtěch Jindra* [online]. Dostupný z WWW: <https://edu.uhk.cz/~jindrvo1/pojistovnictvi/historie.html>

[21] Demografický informační portál: *Demografický informační portál Obecně* [online]. c 2004 – 2008. Dostupný z WWW: <http://www.demografie.info/?cz_umrtnostobecne=&PHPSESSID=dccb9f08e8cd6c53567572514dcb43d1>

[22] Demografický informační portál: *Demografický informační portál Ukazatele* [online]. c 2004 – 2008. Dostupný z WWW: <http://www.demografie.info/?cz_umrtnostukazatele=&PHPSESSID=dccb9f08e8cd6c53567572514dcb43d1>

[23] Demografický informační portál: *Demografický informační portál Úmrtnostní tabulky* [online]. c 2004 – 2008. Dostupný z WWW: <http://www.demografie.info/?cz_umrtnosttabulky=&PHPSESSID=dccb9f08e8cd6c53567572514dcb43d1>

SEZNAM TABULEK

SEZNAM OBRÁZKŮ

SEZNAM POUŽITÝCH ZKRATEK

ČSFR	Československá federativní republika
ČR	Česká republika
MZ ČR	Ministerstvo zdravotnictví České republiky
OR	odds ratio
RR	relativní riziko
HR	hazard ratio

SEZNAM PŘÍLOH

PŘÍLOHY

Příloha 1: Podrobné úmrtnostní tabulky Česká republika, muži, 2007

2007	**Česká republika**						
			Muži *Males*				
věk *age*	qx	px	lx	dx	Lx	Tx	ex
0	0,003728	0,996272	100000	373	99667	7366630	73,67
1	0,000280	0,999720	99627	28	99613	7266964	72,94
2	0,000310	0,999690	99599	31	99584	7167351	71,96
3	0,000223	0,999777	99568	22	99557	7067767	70,98
4	0,000173	0,999827	99546	17	99538	6968210	70,00
5	0,000111	0,999889	99529	11	99523	6868672	69,01
6	0,000094	0,999906	99518	9	99513	6769148	68,02
7	0,000090	0,999910	99509	9	99504	6669635	67,03
8	0,000122	0,999878	99500	12	99494	6570131	66,03
9	0,000153	0,999847	99488	15	99480	6470638	65,04
10	0,000148	0,999852	99472	15	99465	6371158	64,05
11	0,000141	0,999859	99458	14	99451	6271693	63,06
12	0,000156	0,999844	99444	16	99436	6172242	62,07
13	0,000139	0,999861	99428	14	99421	6072806	61,08
14	0,000199	0,999801	99414	20	99404	5973385	60,09
15	0,000247	0,999753	99394	25	99382	5873981	59,10
16	0,000376	0,999624	99370	37	99351	5774599	58,11
17	0,000517	0,999483	99332	51	99307	5675248	57,13
18	0,000756	0,999244	99281	75	99244	5575941	56,16
19	0,000900	0,999100	99206	89	99161	5476697	55,21
20	0,000980	0,999020	99117	97	99068	5377536	54,25
21	0,000937	0,999063	99020	93	98973	5278468	53,31
22	0,000949	0,999051	98927	94	98880	5179494	52,36
23	0,000952	0,999048	98833	94	98786	5080614	51,41
24	0,001003	0,998997	98739	99	98689	4981829	50,45
25	0,001006	0,998994	98640	99	98590	4883139	49,50
26	0,000984	0,999016	98541	97	98492	4784549	48,55
27	0,000954	0,999046	98444	94	98397	4686057	47,60
28	0,000975	0,999025	98350	96	98302	4587660	46,65
29	0,001006	0,998994	98254	99	98204	4489358	45,69
30	0,000928	0,999072	98155	91	98109	4391154	44,74
31	0,000908	0,999092	98064	89	98019	4293044	43,78
32	0,000966	0,999034	97975	95	97928	4195025	42,82
33	0,001121	0,998879	97880	110	97825	4097097	41,86
34	0,001222	0,998778	97771	119	97711	3999272	40,90
35	0,001335	0,998665	97651	130	97586	3901561	39,95
36	0,001407	0,998593	97521	137	97452	3803975	39,01
37	0,001592	0,998408	97384	155	97306	3706523	38,06
38	0,001805	0,998195	97228	175	97141	3609217	37,12
39	0,001948	0,998052	97053	189	96958	3512076	36,19
40	0,001970	0,998030	96864	191	96768	3415118	35,26
41	0,002101	0,997899	96673	203	96572	3318349	34,33
42	0,002251	0,997749	96470	217	96361	3221778	33,40
43	0,002554	0,997446	96253	246	96130	3125416	32,47
44	0,002891	0,997109	96007	278	95868	3029287	31,55
45	0,003180	0,996820	95729	304	95577	2933418	30,64
46	0,003541	0,996459	95425	338	95256	2837841	29,74
47	0,003985	0,996015	95087	379	94898	2742585	28,84
48	0,004555	0,995445	94708	431	94492	2647687	27,96
49	0,005235	0,994765	94277	494	94030	2553195	27,08

věk *age*	qx	px	lx	dx	Lx	Tx	ex
50	0,005884	0,994116	93783	552	93507	2459165	26,22
51	0,006664	0,993336	93231	621	92921	2365658	25,37
52	0,007421	0,992579	92610	687	92266	2272737	24,54
53	0,008071	0,991929	91923	742	91552	2180470	23,72
54	0,009076	0,990924	91181	828	90767	2088918	22,91
55	0,010179	0,989821	90353	920	89894	1998151	22,11
56	0,011324	0,988676	89434	1013	88927	1908258	21,34
57	0,012602	0,987398	88421	1114	87864	1819330	20,58
58	0,013616	0,986384	87307	1189	86712	1731467	19,83
59	0,014675	0,985325	86118	1264	85486	1644754	19,10
60	0,016232	0,983768	84854	1377	84165	1559268	18,38
61	0,017935	0,982065	83477	1497	82728	1475103	17,67
62	0,019510	0,980490	81980	1599	81180	1392375	16,98
63	0,021221	0,978779	80380	1706	79527	1311195	16,31
64	0,022211	0,977789	78674	1747	77801	1231668	15,66
65	0,023587	0,976413	76927	1814	76020	1153867	15,00
66	0,025110	0,974890	75113	1886	74169	1077847	14,35
67	0,026854	0,973146	73226	1966	72243	1003678	13,71
68	0,028572	0,971428	71260	2036	70242	931434	13,07
69	0,031512	0,968488	69224	2181	68133	861192	12,44
70	0,034306	0,965694	67043	2300	65893	793059	11,83
71	0,038165	0,961835	64743	2471	63507	727167	11,23
72	0,042436	0,957564	62272	2643	60950	663659	10,66
73	0,045200	0,954800	59629	2695	58282	602709	10,11
74	0,048609	0,951391	56934	2767	55550	544427	9,56
75	0,053390	0,946610	54166	2892	52720	488877	9,03
76	0,058772	0,941228	51274	3013	49768	436157	8,51
77	0,064629	0,935371	48261	3119	46701	386389	8,01
78	0,070814	0,929186	45142	3197	43544	339688	7,52
79	0,077261	0,922739	41945	3241	40325	296144	7,06
80	0,085121	0,914879	38704	3295	37057	255819	6,61
81	0,093974	0,906026	35410	3328	33746	218762	6,18
82	0,103410	0,896590	32082	3318	30423	185016	5,77
83	0,113818	0,886182	28765	3274	27128	154593	5,37
84	0,125285	0,874715	25491	3194	23894	127465	5,00
85	0,137901	0,862099	22297	3075	20760	103571	4,65
86	0,151759	0,848241	19222	2917	17764	82811	4,31
87	0,166953	0,833047	16305	2722	14944	65047	3,99
88	0,183581	0,816419	13583	2494	12336	50103	3,69
89	0,201736	0,798264	11089	2237	9971	37767	3,41
90	0,221511	0,778489	8852	1961	7872	27796	3,14
91	0,242991	0,757009	6891	1675	6054	19925	2,89
92	0,266253	0,733747	5217	1389	4522	13870	2,66
93	0,291356	0,708644	3828	1115	3270	9348	2,44
94	0,318345	0,681655	2713	864	2281	6078	2,24
95	0,347239	0,652761	1849	642	1528	3797	2,05
96	0,378025	0,621975	1207	456	979	2269	1,88
97	0,410655	0,589345	751	308	597	1290	1,72
98	0,445035	0,554965	442	197	344	694	1,57
99	0,481023	0,518977	246	118	186	350	1,42
100	0,518419	0,481581	127	66	94	163	1,28
101	0,556961	0,443039	61	34	44	69	1,12
102	0,596324	0,403676	27	16	19	25	0,90
103	1,000000	0,000000	11	11	5	5	0,50

Příloha 2: Podrobné úmrtnostní tabulky Česká republika, ženy, 2007

2007	Česká republika						
Ženy *Females*							
věk *age*	qx	px	lx	dx	Lx	Tx	ex
0	0,002529	0,997471	100000	253	99783	7990074	79,90
1	0,000276	0,999724	99747	28	99733	7890292	79,10
2	0,000205	0,999795	99720	20	99709	7790558	78,12
3	0,000086	0,999914	99699	9	99695	7690849	77,14
4	0,000121	0,999879	99691	12	99685	7591154	76,15
5	0,000115	0,999885	99678	11	99673	7491470	75,16
6	0,000109	0,999891	99667	11	99662	7391797	74,16
7	0,000070	0,999930	99656	7	99653	7292135	73,17
8	0,000078	0,999922	99649	8	99645	7192483	72,18
9	0,000084	0,999916	99641	8	99637	7092837	71,18
10	0,000114	0,999886	99633	11	99627	6993200	70,19
11	0,000105	0,999895	99622	10	99617	6893573	69,20
12	0,000091	0,999909	99611	9	99607	6793956	68,20
13	0,000080	0,999920	99602	8	99598	6694349	67,21
14	0,000115	0,999885	99594	11	99589	6594751	66,22
15	0,000156	0,999844	99583	16	99575	6495163	65,22
16	0,000202	0,999798	99567	20	99557	6395588	64,23
17	0,000240	0,999760	99547	24	99535	6296030	63,25
18	0,000258	0,999742	99523	26	99510	6196495	62,26
19	0,000279	0,999721	99498	28	99484	6096985	61,28
20	0,000254	0,999746	99470	25	99457	5997501	60,29
21	0,000225	0,999775	99445	22	99433	5898044	59,31
22	0,000183	0,999817	99422	18	99413	5798611	58,32
23	0,000212	0,999788	99404	21	99393	5699198	57,33
24	0,000232	0,999768	99383	23	99371	5599804	56,35
25	0,000236	0,999764	99360	23	99348	5500433	55,36
26	0,000202	0,999798	99336	20	99326	5401085	54,37
27	0,000216	0,999784	99316	21	99306	5301758	53,38
28	0,000243	0,999757	99295	24	99283	5202453	52,39
29	0,000325	0,999675	99271	32	99255	5103170	51,41
30	0,000371	0,999629	99239	37	99220	5003915	50,42
31	0,000369	0,999631	99202	37	99183	4904695	49,44
32	0,000307	0,999693	99165	30	99150	4805512	48,46
33	0,000333	0,999667	99135	33	99118	4706362	47,47
34	0,000369	0,999631	99102	37	99083	4607244	46,49
35	0,000445	0,999555	99065	44	99043	4508160	45,51
36	0,000570	0,999430	99021	56	98993	4409117	44,53
37	0,000663	0,999337	98965	66	98932	4310124	43,55
38	0,000793	0,999207	98899	78	98860	4211193	42,58
39	0,000927	0,999073	98821	92	98775	4112333	41,61
40	0,001028	0,998972	98729	101	98678	4013558	40,65
41	0,001133	0,998867	98627	112	98572	3914880	39,69
42	0,001265	0,998735	98516	125	98453	3816308	38,74
43	0,001359	0,998641	98391	134	98324	3717855	37,79
44	0,001533	0,998467	98257	151	98182	3619531	36,84
45	0,001557	0,998443	98107	153	98030	3521349	35,89
46	0,001671	0,998329	97954	164	97872	3423318	34,95
47	0,001800	0,998200	97790	176	97702	3325446	34,01
48	0,002066	0,997934	97614	202	97513	3227744	33,07
49	0,002334	0,997666	97413	227	97299	3130230	32,13

věk *age*	qx	px	lx	dx	Lx	Tx	ex
50	0,002620	0,997380	97185	255	97058	3032932	31,21
51	0,002931	0,997069	96931	284	96789	2935874	30,29
52	0,003272	0,996728	96647	316	96488	2839085	29,38
53	0,003556	0,996444	96330	343	96159	2742597	28,47
54	0,003970	0,996030	95988	381	95797	2646437	27,57
55	0,004258	0,995742	95607	407	95403	2550640	26,68
56	0,004421	0,995579	95200	421	94989	2455237	25,79
57	0,005081	0,994919	94779	482	94538	2360248	24,90
58	0,005721	0,994279	94297	539	94027	2265710	24,03
59	0,006491	0,993509	93758	609	93453	2171683	23,16
60	0,007147	0,992853	93149	666	92816	2078229	22,31
61	0,007794	0,992206	92483	721	92123	1985413	21,47
62	0,008265	0,991735	91763	758	91383	1893290	20,63
63	0,008982	0,991018	91004	817	90595	1801907	19,80
64	0,009566	0,990434	90187	863	89755	1711311	18,98
65	0,010541	0,989459	89324	942	88853	1621556	18,15
66	0,011545	0,988455	88382	1020	87872	1532703	17,34
67	0,012641	0,987359	87362	1104	86810	1444830	16,54
68	0,014535	0,985465	86258	1254	85631	1358021	15,74
69	0,016392	0,983608	85004	1393	84307	1272390	14,97
70	0,018033	0,981967	83611	1508	82857	1188082	14,21
71	0,019749	0,980251	82103	1621	81292	1105226	13,46
72	0,021456	0,978544	80481	1727	79618	1023934	12,72
73	0,023730	0,976270	78755	1869	77820	944315	11,99
74	0,027802	0,972198	76886	2138	75817	866495	11,27
75	0,032092	0,967908	74748	2399	73549	790678	10,58
76	0,036728	0,963272	72349	2657	71021	717129	9,91
77	0,041286	0,958714	69692	2877	68254	646109	9,27
78	0,046216	0,953784	66815	3088	65271	577855	8,65
79	0,052329	0,947671	63727	3335	62060	512584	8,04
80	0,059555	0,940445	60392	3597	58594	450525	7,46
81	0,068036	0,931964	56796	3864	54863	391931	6,90
82	0,077462	0,922538	52931	4100	50881	337067	6,37
83	0,088234	0,911766	48831	4309	46677	286186	5,86
84	0,100483	0,899517	44523	4474	42286	239509	5,38
85	0,114396	0,885604	40049	4581	37758	197223	4,92
86	0,130167	0,869833	35467	4617	33159	159465	4,50
87	0,147996	0,852004	30851	4566	28568	126306	4,09
88	0,168093	0,831907	26285	4418	24076	97738	3,72
89	0,190671	0,809329	21867	4169	19782	73662	3,37
90	0,215934	0,784066	17697	3821	15787	53880	3,04
91	0,244077	0,755923	13876	3387	12182	38094	2,75
92	0,275264	0,724736	10489	2887	9045	25911	2,47
93	0,309618	0,690382	7602	2354	6425	16866	2,22
94	0,347200	0,652800	5248	1822	4337	10441	1,99
95	0,387986	0,612014	3426	1329	2761	6104	1,78
96	0,431846	0,568154	2097	905	1644	3342	1,59
97	0,478515	0,521485	1191	570	906	1698	1,43
98	0,527571	0,472429	621	328	457	792	1,27
99	0,578417	0,421583	293	170	209	335	1,14
100	0,630272	0,369728	124	78	85	126	1,02
101	0,682180	0,317820	46	31	30	41	0,90
102	0,733035	0,266965	15	11	9	11	0,77
103	1,000000	0,000000	4	4	2	2	0,50