

**Zpracování softwarové databáze
pro aplikační nástroj
Filemaker Pro 11 pro evidenci
identifikačních parametrů pacientů
s roztroušenou sklerózou
mozkomíšní.**

**Medical data processing for
a Filemaker Pro 11 software tool to
identify patients with cerebrospinal
sclerosis.**

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 6.5.2011

.....

Abstrakt

Náplní této práce je návrh a implementace databázového systému pro klinickou studii. Tato práce je členěna do čtyř hlavních částí. První část práce je věnována úvodu do problematiky roztroušené sklerózy, průběhu a charakteru tohoto onemocnění, a teoriích o jeho příčinách. Druhá část přibližuje problematiku návrhu databázového systému a obsahuje popis systému správy databáze. Jsou zde stručně prezentovány tři nejvýznamnější datové modely, které byly brány v úvahu při konstrukci databáze, se stručnou diskuzí jejich výhod a nevýhod. Třetí část je věnována problematice tvorby uživatelského rozhraní k této databázi. Prezentuje základní principy pro tvorbu uživatelského rozhraní. V této části jsou diskutovány designová rozhodnutí při tvorbě uživatelského rozhraní, které vzniklo v rámci této bakalářské práce, pro databázi pacientů klinické studie. Čtvrtá část je věnována nástinu problematiky statistického vyhodnocování dat, které budou po ukončení příslušné studie obsaženy v navržené databázi. Jsou prezentovány bootstrapové metody pro výpočet středních hodnot a rozptylů statistických veličin. Tato část také obsahuje příklady algoritmů pro výpočet střední hodnoty náhodné veličiny v jazyku Python.

Klíčová slova: roztroušená skleróza, bootstrap, databáze, uživatelské rozhraní

Abstract

The goal of this Bachelor thesis is to design and implement a database system for clinical study. This thesis is divided into four main parts. The first part introduces multiple sclerosis, the nature of this disease and theories about its causes. The second part is about database design issues and contains a description of a database management system. There is a brief presentation of three major data models that have been taken into consideration when designing the database, a brief discussion of their advantages and disadvantages is included. The third part is devoted to creation of the user interface for the database. Basic principles for creating user interfaces are presented. This section discusses design decisions made when creating user interface for the patient database for clinical studies. The fourth part is devoted to outlining the problem of statistical evaluation of data that will be realised after completion of the clinical study. Bootstrap methods are presented for calculating the mean and variance of statistical variables. This section also contains examples of algorithms for calculating the mean value of random variable in Python.

Keywords: multiple sclerosis, bootstrap, database, user interface

Seznam použitých zkratk a symbolů

FM11	– FileMaker 11
FNO	– Fakultní nemocnice ostrava
RS	– roztroušená skleróza
MR	– magnetická rezonance
VJI	– Vena Jugularis Interna (Vnitřní hrdelní žíla)
VAZY	– Vena Azygos (vena azygos je nepárová žíla, která běží zprava podél hrudní páteře.)
CCSVI	– Chronická Cerebrospinální venózní insuficience
PTA	– perkutánní transluminální angioplastika
EDSS	– rozšířená stupnice stavu invalidity
FSS	– fatigue severity scale (únavová škála)
MFSS	– modified fatigue severity scale (modifikovaná únavová škála)
MSIS-29	– multiple sclerosis impact scale (škála závažnosti dopadu roztroušené sklerózy)
QoL	– quality of life (dotazník kvality života)
DBMS	– DataBase Management System (systém správy databáze)
E(X)	– střední hodnota náhodné veličiny X
D(X)	– rozptyl náhodné veličiny X
1:1	– relace jedna k jedné
1:M	– relace jedna ku více
M:N	– relace více k více

Obsah

1	Úvod	2
2	Roztroušená skleróza	4
2.1	Obecné informace	4
2.2	Průběh onemocnění	4
2.2.1	Klasifikace RS podle průběhu	5
2.3	Možné příčiny onemocnění	6
2.4	Léčba	7
2.5	CCSVI (Chronic cerebrospinal venous insufficiency)	7
3	Klinická studie CCSVI FNO	9
3.1	Cíle studie	9
4	Výběr vhodného databázového modelu a jeho implementace v Database management systému FM11 Pro	11
4.1	Database management system	11
4.2	Metodologie návrhu databázového systému	11
4.2.1	Popis systému	12
4.2.2	Model pro databázi FNO	13
4.3	Databázové modely	14
4.3.1	Objektově orientovaný model	14
4.3.2	Relační model	15
4.3.3	Hierarchický model	15
4.3.4	Model pro databázi FNO	16
5	Design uživatelského rozhraní pro databázi	18
5.1	Uživatelské rozhraní pro studii FNO	18
5.2	Komentář k uživatelskému rozhraní databáze	20
6	Metody Bootstrap	33
6.1	Parametrický bootstrap	35
6.2	Střední kvadratická chyba odhadu	35
6.3	Bootstrap střední kvadratické chyby, rozptylu a směrodatné odchylky	37
6.4	Konstrukce bootstrapových intervalů spolehlivosti	38
6.4.1	Odhad střední hodnoty	38
6.4.2	Odhad rozptylu	42

6.4.3	Zobecnění	43
7	Reference	45

Přílohy

A - Seznam databázových polí se stručným popisem

B - EDSS škála

C - MSIS-29 (Multiple Sclerosis Impact Scale)

1 Úvod

Roztroušená skleróza (RS) je jedním z nejvážnějších onemocnění a to nejen samotným charakterem průběhu choroby, ale také masivností výskytu, zejména mezi populací v produktivním věku. Hlavní náplní této práce je návrh a implementace databázového systému pro konkrétní klinickou studii RS, včetně nástinu jedné z možností následného statistického vyhodnocení získaných dat. Na první pohled se může zdát že jde o jednoduchou problematiku. Přihlédneme-li však k množství hromaděných údajů, přibližně 2000 proměnných na jeden záznam, a počtu záznamů je zřejmé, že efektivní konstrukce databázového systému je nezbytně nutná.

Text je členěn do čtyř hlavních částí.

První část práce je věnována úvodu do problematiky roztroušené sklerózy, popisu průběhu a charakteru tohoto onemocnění a teoriích o jeho příčinách. První část rovněž poskytuje základní informace o klinické studii „Otevřená randomizovaná pilotní studie posuzující bezpečnost a efektivitu endovaskulární léčby CCSVI - Chronické Cerebrospinnální Venózní Insuficience u pacientů s RS - Roztroušenou Mozkomíšní Sklerózou“, která je postupně realizována ve Fakultní nemocnici Ostrava(FNO).

Druhá část přibližuje obecnou problematiku návrhu databázového systému a obsahuje popis database management systému. Teoretický základ byl čerpán z [6] a [8]. Jsou zde stručně prezentovány tři nejvýznamnější datové modely, které byly brány v úvahu při konstrukci databáze k uvedené studii, se stručnou diskuzí jejich výhod a nevýhod.

Třetí část je věnována problematice tvorby uživatelského rozhraní k této databázi. Jako podklady pro tuto část byly použity [9] a [10]. Prezentuje základní principy pro tvorbu uživatelského rozhraní. V této části jsou diskutovány designová rozhodnutí při tvorbě uživatelského rozhraní, které vzniklo v rámci této bakalářské práce, pro databázi pacientů klinické studie.

Protože autor předpokládá, že se bude podílet na statistickém vyhodnocení výsledků po ukončení studie, čtvrtá část je věnována problematice statistického vyhodnocení dat. Budou prezentovány výlučně bootstrapové metody (čerpáno hlavně z [12] a [13]). Jiné možnosti statistického vyhodnocení jsou prezentovány v bakalářské práci Michala Bělocha „Statistické prostředky pro podporu léčby pacientů s roztroušenou sklerózou mozkomíšní“. Velkou výhodou těchto metod je fakt že nekladou žádné předpoklady na rozdělení zkoumané náhodné veličiny. Jsou zde hlavně popsány algoritmy pro výpočet odhadů středních hodnot a rozptylů statistických veličin. Tato část rovněž obsahuje ukázkou implementace v jazyku Python.

Tato práce předpokládá základní povědomí o teorii databázových systémů, metodách tvorby uživatelských rozhraní a statistiky.

2 Roztroušená skleróza

2.1 Obecné informace

Roztroušená skleróza (dále jen RS) je zánětlivé autoimunitní onemocnění které způsobuje poškození centrální nervové soustavy. RS ovlivňuje schopnost buněk v mozku a míše navzájem komunikovat pomocí nervových vzruchů přes dlouhé neuronové výběžky (takzvané axony), které jsou obalené izolační látkou zvanou myelin¹. Při onemocnění RS dochází vlivem autoimunitní reakce k narušení a postupné ztrátě myelinu, což vede ke ztrátě funkčnosti buňky. Slovo skleróza (z řeckého skleros - tuhý) v názvu nemoci odkazuje na zjizvení v bílé mozkové hmotě a míše, které jsou převážně tvořeny myelinem. Místa po odeznění zánětu odborně nazýváme plaka².

I přestože o procesu choroby je známo mnoho, samotná příčina vzniku onemocnění je stále neznámá. Na tuto nemoc v dnešní době není znám lék, existují však přípravky pro zpomalení postupu nemoci či potlačení jejích symptomů, ty však mívají vedlejší účinky. Na každých 100 000 obyvatel připadá 50-180 pacientů (viz [3]) trpících RS v závislosti na geografické lokalitě. Vzhledem k charakteru nemoci je těžké určit prognózu pacienta. Střední délka života je u pacientů s RS o 5-10 let kratší než u zdravého člověka. V řadě případů dochází ke změnám kvality života. Onemocnění poprvé popsal francouzský neurolog Jean-Martin Charcot (1825,1893) až v roce 1868. Zvláštní případy průběhu nemoci popsali Eugene Devic (1858,1930), József Baló (1895,1979), Paul Ferdinand Schilder (1886,1940) a Otto Marburg (1874,1948).

2.2 Průběh onemocnění

Onemocnění RS je typické epizodickými stavy akutního zhoršení zdravotního stavu pacienta - relapsy neboli ataky. Ataky jsou zpravidla nepředvídatelné a bez známých příčin. Frekvence výskytu a závažnost atak je tedy velice individuální. RS má více forem s různou frekvencí ataků a rychlostí rozvoje nemoci. Mezi jednotlivými atakami mohou symptomy i na nějakou dobu vymizet. Jednoduše lze RS podle průběhu dělit na maligní³ a benigní⁴. Při maligním průběhu onemocnění jsou ataky časté, s závažnými následky a u pacienta postupně dochází k nárůstu invalidity. Naopak u benigního průběhu onemocnění jsou

¹ tvoří obal neuronů, pomáhá přenosu elektrických signálů

² syn. skléry, léze

³ zhoubnou

⁴ nezhoubnou

ataka lehká a období mezi ataky může být v individuálních případech až několik let. Opakem ataky je remise, tedy období kdy dochází k stabilizaci příznaků a v některých případech i vymizení příznaků RS.

2.2.1 Klasifikace RS podle průběhu

u onemocnění RS se mohou vyskytovat různé neurologické symptomy⁵, které často přechází ve fyzickou invaliditu a kognitivní potíže. Vzhledem k faktu že frekvence atak a charakter nových neurologických symptomů je vysoce individuální, je potřeba objektivní stupnice pro zhodnocení průběhu nemoci. Pro tento účel se používá EDSS stupnice⁶ (Expanded Disability Status Scale). EDSS klasifikuje pacienty podle závažnosti invalidity na stupnici 0-10 s krokem 0,5; 0 znamená normální neurologický nálezn a 10 smrt v důsledku RS. Existuje také klasifikace forem progresivní roztroušené sklerózy. Tyto formy se snaží předvídat průběhu nemoci na základě dřívějšího průběhu onemocnění. V roce 1996 americká National Multiple Sclerosis Society⁷ standardizovala definice čtyř forem RS. Dále popíšeme následující formy onemocnění RS:

- Remitentní - relabující (RR)
- Sekundárně progresivní (chronicko-progresivní)
- Primárně progresivní
- Relabující progresivní

Remitentní - relabující

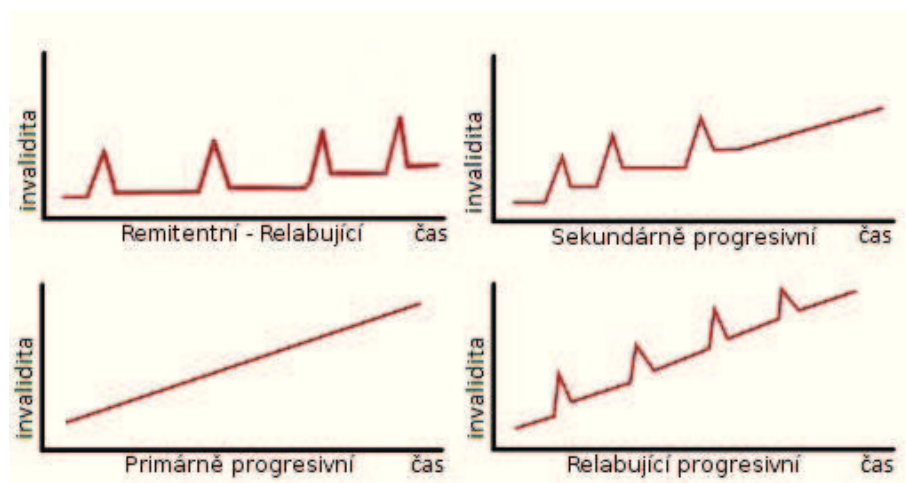
Jde o nejběžnější formu průběhu onemocnění RS, onemocnění RS také obvykle začíná v této formě. Dochází k výskytu atak trvajících několik týdnů až měsíců. Remise je v této formě částečná či úplná, organizmus je schopen kompenzovat ztrátu myelinu i bez pomoci léčiv. Obnova myelinu však nenastává v místech s probíhajícím autoimunitním zánětem, v tomto případě bývají předepsána imunosupresiva. Tato forma je, narozdíl od ostatních, dobře ovlivnitelná farmaceutickými přípravky.

Sekundárně progresivní

⁵např. únava, křeče, potíže s pamětí, emocionální nestabilita, deprese

⁶rozšířená stupnice stavu invalidity

⁷Národní společnost pro roztroušenou sklerózu



Obrázek 1: Ilustrace možných vztahů mezi elementy

Tato forma předpokládá vyčerpání schopnosti organismu nadále kompenzovat ztrátu myelinu. Dochází k mírnému nárůstu invalidity. Remitentní forma přechází na Sekundárně progresivní u 50% pacientů po 19-ti letech průběhu nemoci. Účinnost léků je u této formy nižší, velkou roli hraje životospráva jednotlivých pacientů.

Primárně progresivní

Tato forma se vyskytuje u pacientů mezi 40. a 50. rokem života. Při této formě dochází k progresivnímu nárůstu invalidity. Farmaceutické přípravky nemají na onemocnění primárně progresivní formy velký vliv.

Relabující progresivní

Jedná se o vzácnou formu průběhu onemocnění RS vyskytující se u pacientů nad 40 let. po atacích nedochází ke zlepšení zdravotního stavu pacienta a neurologické postižení pacienta se s každým relapsem stupňuje. V této fázi je život pacienta ohrožen a je potřeba využívat veškerých dostupných prostředků k zbrždění postupu nemoci, nehledě na možné nežádoucí účinky farmaceutických přípravků.

2.3 Možné příčiny onemocnění

RS je pravděpodobně způsobena kombinací genetických faktorů, vlivů prostředí, infekčních faktorů a vaskulárních potíží. Pravděpodobně se jedná o vliv CCSVI⁸, jež je předmětem klinické studie FNO. Rizikovou skupinu tvoří lidé ve věku od 20 do 40 let.

⁸Chronická Cerebrospinální venózní insuficience

Výskyt nemoci je častější u žen (v přibližném poměru 2:1). Primárně progresivní forma onemocnění je nejběžnější u jedinců nad 50 let. Nad věkovou hranicí 50 let se rovněž vyrovnává poměr počtu mužů a žen.

Genetické faktory

RS není považována za dědičnou chorobu, ale šance výskytu onemocnění je vyšší u příbuzných a potomků pacientů trpících RS. Onemocnění je nejčastější u euro-poidní rasy, dále pak negroidní, nejmenší výskyt je zaznamenán u mongoloidní rasy.

Faktory prostředí

Výskyt RS je četnější u populací vzdálenějších od rovníku, toto je pravděpodobně způsobeno sníženým vystavením slunečnímu záření. Snížená produkce vitamínu D je hlavní biologický mechanismus použitý k vysvětlení tohoto jevu. Kouření bylo popsáno jako nezávislý rizikový faktor.

Infekční faktory

Během výzkumu nemoci byly vypracovány různé hypotetické mechanismy vzniku tohoto onemocnění. Například tzv. hygienická hypotéza uvádí, že jedinci vystavení různým patogenům během dětství mají lepší ochranu proti onemocnění RS. RS by podle této teorie pak byla později vyvolána autoimunitní reakcí spuštěnou u náchylných jedinců četnými infekčními mikroorganismy, s rizikem rostoucím s věkem při infekci.

2.4 Léčba

v dnešní době není znám lék na RS, nicméně existuje řada farmaceutických přípravků jež pomáhají zmírnit četnost a závažnost atak a potlačují neurologické symptomy choroby. k užívaným přípravkům patří imunomodulační přípravky, imunosupresiva a výjimečně i cytostatika. Na regulaci zánětů se používají kortikoidy. Někteří pacienti také vyhledávají alternativní způsoby léčby jako například diety, bylinnou léčbu, drogy (marihuanu) nebo cvičení jako tai chi nebo jógu.

2.5 CCSVI (Chronic cerebrospinal venous insufficiency)

Vliv CCSVI na kvalitu života pacientů trpících RS je hlavním předmětem probíhající studie FNO. CCSVI je stav, kdy je omezen nebo blokován krevní oběh z mozku a páteře zpět do srdce. Tento stav je zapříčiněn zúžením žil zajišťujících odtok krve. Jedná se především

o krční žíly a venu azygos⁹. Narušená činnost zúžených žil časem způsobuje reflux¹⁰ krve zpět do mozku či páteře. V takovémto případě může docházet k úniku krvinek a dalších látek do tkání mozku a páteře. Při delším zdržování krve v mozku dochází k pomalé dodávce živin a pomalého odvádění neokysličené krve, což způsobuje nedostatek kyslíku. Tento fakt bývá spojován s únavovými stavy typickými pro pacienty trpící RS. Německý patolog Georg Eduard Rindfleisch jako první popsal rozložení zánětlivých lézí způsobených RS kolem žil. Vztah mezi CCSVI a RS poprvé popsal italský operatér Dr. Paolo Zamboni v r.2009. Dle jeho teorie CCSVI může způsobovat hromadění železa, jež může být příčinou autoimunitní reakce RS.

⁹nepárová žíla běžící zprava podél páteře

¹⁰zpětný tok

3 Klinická studie CCSVI FNO

Oficiální název studie je "Otevřená randomizovaná pilotní studie posuzující bezpečnost a efektivitu endovaskulární léčby CCSVI - Chronické Cerebrospinální Venózní Insuficience u pacientů s RS - Roztroušenou Mozkomíšni Sklerózou". Do studie bude postupně během doby 24 měsíců zařazeno 100 pacientů v léčené¹¹ a 100 pacientů v kontrolní skupině¹², s tím že po 12 měsících bude cca 75 pacientů při splnění kritérií studie zařazeno do druhé intervenční skupiny (to znamená že budou léčeni stejně jako pacienti první skupiny). Pacienti budou sledováni po dobu 24 měsíců. Obě skupiny pacientů budou pokračovat v zavedené léčbě RS beze změny medikace. Obě skupiny podstoupí neinvazivní vyšetření duplexní dopplerovskou sonografií a MR¹³ vyšetření, intervenční skupina podstoupí vyšetření cerebrocervikální venózní cirkulace a podle zařazení do skupiny zákrok PTA¹⁴ - perkutánní transluminální angioplastiky stenózujícího postižení krčních žil nebo veny azygos. v rámci studie bude dokumentováno procento výskytu CCSVI s měřením venózního tlaku u pacientů, vliv PTA na redukci klinického postižení, zabránění relapsu postižení, sledovány kvantitativní změny demyelinizačních ložisek. Pacienti kontrolní skupiny v případě splnění indikačních podmínek budou zařazeni do druhé intervenční skupiny. u obou skupin bude posuzován výsledný klinický stav vzhledem ke včasnosti indikace intervenčního výkonu.

3.1 Cíle studie

Vyhodnotit bezpečnost venózní PTA VJI¹⁵ a VAZY¹⁶ na základě:

- četnosti a povahy vedlejších účinků po zákroku
- sledování potenciálních nežádoucích účinků na zhoršení neurologického nálezu na EDSS stupnici
- srovnání průběhu RS před a po zákroku

Vyhodnotit účinnost PTA, PTA považováno za účinné jestliže:

- dojde k redukci stenózy na < 30% lumina¹⁷ žíly

¹¹pacientovi bude provedena perkutánní transluminální angioplastika

¹²pacienti budou léčeni klasicky

¹³magnetická rezonance

¹⁴tzv. liberační procedura - užitím balónkového katetru je uvolněn průtok krve žílou

¹⁵vena jugularis interna, párové krční žíly

¹⁶vena azygos - nepárová žíla běžící zprava podél hrudní páteře.

¹⁷medicínský termín pro průřez žíly

- dojde k efektivnímu ovlivnění stenóz¹⁸
- dojde ke zlepšení průtokových parametrů žil

po provedení PTA stenóz žilního systému je léčba CCSVI-RSM považována za účinnou pokud:

- dojde ke snížení klinického stavu EDSS, FSS, MFIS, MSIS-29¹⁹
- dojde k redukci relapsů
- dojde ke zlepšení QOL²⁰
- dojde ke stabilizaci MR nálezů

Pro aplikaci bootstrapové metody na data získaná během průběhu studie pro nás tedy budou zajímavé zejména hodnoty železa, počty plak a skóre dotazníků. Předmětem našeho zájmu budou především hodnoty vyšetření ECD²¹ vzhledem k charakteru zákroku PTA²².

¹⁸ zúžení

¹⁹škály pro hodnocení stavu pacienta trpícího RS

²⁰dotazník kvality života

²¹hodnoty průměrů a průtoků žil

²²užitím balónkového katetru je odstraněna obstrukce žíly

4 Výběr vhodného databázového modelu a jeho implementace v Database management systému FM11 Pro

Na základě specifikací účelu databáze a způsobu hromadění dat získávaných v rámci studie bylo třeba sestavit vhodný model organizace dat. Využití vhodného datového modelu je velmi důležité pro usnadnění pozdějšího přístupu k datům za účelem jejich dalšího zpracování pomocí statistických programů.

4.1 Database management system

Jedná se o aplikační prostředí jež zajišťuje vytváření tabulek, vkládání dat, správu a interakci s databází apod.. FNO má zakoupeno komerční řešení FileMaker Pro 11. Proto není v tomto případě potřeba řešit problematiku fyzické implementace databázových systémů. Fyzická integrita a bezpečnost dat je tímto systémem rovněž zajištěna. DBMS plní tyto funkce:

- zajišťuje přístup k datům
- zajišťuje a spravuje struktury nezbytné pro ukládání dat
- zajišťuje nezávislosti dat na jejich logické organizaci při jejich prezentování uživateli
- správa zabezpečení dat
- správa přístupových práv více uživatelů k databázi
- správa zálohování a obnovy dat
- zajišťuje integritu dat
- zajišťuje přístup k databázi pomocí dotazových jazyků (např. SQL)
- zajišťuje uživatelská rozhraní pro přístup k databázi (např. Filemaker má možnost Instant Web Publishing)

4.2 Metodologie návrhu databázového systému

Při konceptuálním návrhu databázového systému je třeba danou problematiku zhodnotit v kontextu následujících pojmů:

Atribut - charakteristika entity, vlastnost konkrétního objektu, odpovídá jednotlivým polím v databázi

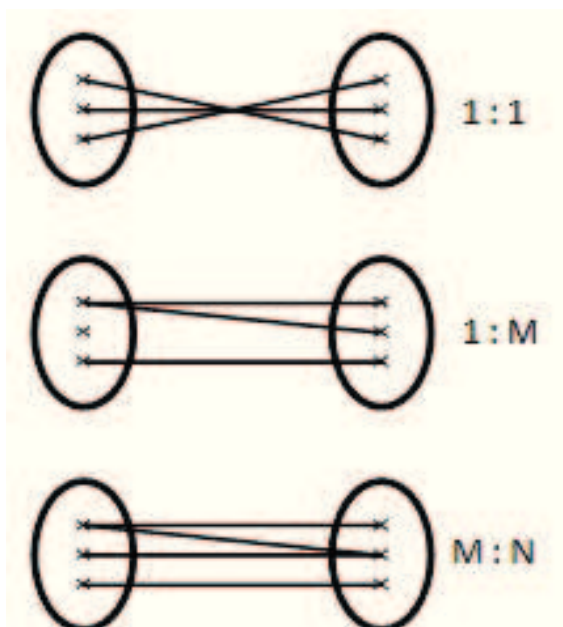
Záznam - seskupení atributů jedné entity, jeden řádek v tabulce

Tabulka - seskupení záznamů stejného typu

Transakce - jeden zvratitelný úkon provedený v databázi

Vztah - popisuje vztah mezi dvěma (případně více) objekty, existují tři typy vztahů viz obr. 2

- One-to-many - vztah jeden k více, surjektivní relace (1:M)
- Many-to-many - vztah více k více, (M:N)
- One-to-one - vztah jedna k jedné, injektivní funkce příklad: datum narození pacienta (1:1)



Obrázek 2: Ilustrace možných vztahů mezi elementy

4.2.1 Popis systému

Konstrukce databázového systému by se měla začít formulací jednoduchých pravidel jež v systému mezi daty budou platit. Na základě těchto pravidel je pak snadné sestavit funkční model databáze. Pravidla popisují prvky v systému, jejich atributy a vztahy mezi prvky systému. Zdrojem těchto pravidel jsou v našem případě specifikace studie

a požadavky lékařů. Pečlivá formulace takovýchto pravidel je velmi hodnotná, protože dává možnost odstranit chybné prekoncepce o fungování daného systému před jeho implementací. Dává rovněž možnost zhodnotit povahu, roli a rozsah systému. Na základě těchto pravidel by rovněž měl být zkonstruován vhodný systém pojmenovávání jednotlivých atributů v rámci databáze.

4.2.2 Model pro databázi FNO

Konstrukci databáze pro studii FNO jsme zahájili²³ formulací pravidel pro databázový systém. Tato pravidla přímo vycházejí z protokolu o studii a konzultací s lékaři.

- Studie bude mít 200 pacientů rozdělených do 3 skupin - A,B,C (kontrolní, intervenční, druhá intervenční)
- Skupina a (kontrolní) bude mít 100 pacientů
- Skupina B (intervenční) bude mít 100 pacientů
- Prvních 12 měsíců budou pacienti pouze ve skupinách A,B
- po 12 měsících bude určitý počet pacientů kontrolní skupiny zařazen do druhé intervenční skupiny (C)
- Skupina C (druhá intervenční) bude mít cca 75 pacientů
- Studie bude obsahovat celkem 10 kontrol - T0, T1, T2, T3, T6, T12, T13, T15, T 18, T24; navíc bude jedna kontrola provedena pro přeřazení
- Kontrola se skládá z více vyšetření a dotazníků - Celkové klinické vyšetření, EDC, MR, EDSS, dotazníky (FSS, MSIS-29, MFIS), vyšetření očního pozadí, likvor, venografie, dotazník QoL; jednotlivé kontroly zahrnují jen některé z vyšetření, dotazníky jsou součástí všech kontrol
- Skupina a bude mít kontroly T0, T3, T6, T12 ,T24
- Skupina B bude mít kontroly T0, T1, T2, T3, T6, T12, T 18, T24
- Skupina C bude mít kontroly T12, T13, T15, T18, T24
- Databáze by měla také obsahovat informace administrativního charakteru, jako data kontrol

²³Ve spolupráci s Michalem Bělochem

v kontextu těchto pravidel byla prozkoumána vhodnost vybraných datových modelů pro následnou implementaci databázového systému pro klinickou studii.

4.3 Databázové modely

V této části zmíníme tři typické databázové modely. Uvedeme je v pořadí v jakém byly brány v potaz při tvorbě databáze, tedy sestupně podle míry abstrakce, které lze jejich použitím dosáhnout.

4.3.1 Objektově orientovaný model

Reálné problémy se stále narůstající složitostí demonstrovaly potřebu konstrukce přesnějších modelů. První z takových to modelů byl SDM²⁴ v roce 1981. SDM prezentuje data a jejich vztahy zapomocí konstrukce zvané objekt, SDM tedy představuje OODM²⁵.

Princip: Objekt je abstrakce reálné entity. Atributy popisují vlastnosti objektu. Objekty s podobnou charakteristikou jsou seskupeny do tříd, u kterých jsou definovány metody interakce s atributy dané třídy. Třídy jsou řazené do hierarchie která dovoluje dědit společné vlastnosti.

Výhody:

- zahrnuje sémantický kontext
- dědičnost podporuje integritu databáze
- strukturální²⁶ a datová²⁷ nezávislost

Nevýhody:

- pomalý vývoj standardů
- složitý navigační přístup
- strmá křivka učení
- velká náročnost transakcí na fyzické zdroje

po zhodnocení rozsahu databáze co se týče různých entit²⁸ jež se budou v systému vyskytovat jsme rozhodli, že objektově orientovaný model by byl pro databázi FNO příliš

²⁴semantický datový model

²⁵object oriented data model

²⁶způsob přístupu k datům není závislý na jejich organizaci

²⁷fyzická organizace dat je maskována DBMS

²⁸pacient, kontrola, vyšetření krve apod.

složitý. Přednosti objektově orientovaného modelu vynikají u složitých²⁹ systémů s četnými transakcemi. Databáze FNO není z tohoto pohledu natolik složitá, aby vyžadovala konstrukci hierarchie tříd. Počet a charakter transakcí je předem znám. Proto by OODM nebyl nejvhodnější.

4.3.2 Relační model

Vyvinut E.F. Coddem (IBM) v roce 1970. Konceptuální jednoduchost tohoto modelu je vyvážena zvýšenou náročností na fyzické prostředky při implementaci.

Princip: relační tabulky³⁰

Výhody:

- strukturální nezávislost, snadno proveditelné zásahy do struktury databáze
- konceptuálně jednoduchý model, dovoluje úplně ignorovat fyzickou implementaci databáze
- jednoduchý design databáze, implementace a management
- možnost podpory ad hoc dotazů (SQL)

Nevýhody:

- velmi náročný na software a fyzické prostředky pro zajištění databáze
- konceptuální jednoduchost může vést ke špatným designům, implementacím a datovým anomáliím
- bývá pomalejší než ostatní modely

Systém relačních tabulek se na první pohled jeví jako velmi vhodný model pro databázi FNO. Nicméně tento systém by vyžadoval navržení číslovací konvence pro efektivní řetězení záznamů a tím by vznikly zbytečné nároky na proškolení personálu který bude vkládat data do systému.

4.3.3 Hierarchický model

Jedná se o jeden z prvních databázových modelů. Tento model byl vytvořen firmou IBM v roce 1960. Asi nejznámější databází s hierarchickým modelem je windows registry.

Principem tohoto modelu je logická organizace dat do víceúrovňové stromové struktury.

Výhody:

²⁹s velkým množstvím různých druhů entit a metod pro interakci mezi nimi

³⁰tabulky propojené pomocí navzájem společných hodnot (klíčů)

- velmi efektivně zachycuje systémy s mnoha úrovněmi 1:M vztahů
- konceptuální jednoduchost
- bezpečnost databáze je zajištěná DBMS
- datová nezávislost (aplikace přistupující k datům nemohou zasahovat do organizační struktury databáze)
- model podporuje datovou integritu v systému (jasně definované parent-child vztahy)

Nevýhody:

- komplikovaná implementace
- složitý navigační přístup (pro přístup k elementům hluboko ve struktuře je potřeba procházet nadřazenou strukturou od kořene)
- složitá správa databáze, zásahy do struktury databáze jsou velmi komplikované
- závislost na stromové struktuře
- limitováno na 1:1 a 1:M vztahy
- absence standardizované implementace

Jednotlivé kontroly ve studii lze obecně chápat jako jeden druh entity. S tímto předpokladem je možné je implementovat jako jeden druh záznamu, což přináší zjednodušení konstrukce databáze. Avšak s takovýmto rozdělením by vznikla také možnost inkonzistencí vlivem špatného přidružování jednotlivých kontrol k pacientům a v důsledku by takovýto systém mohl být nepřehledný.

4.3.4 Model pro databázi FNO

Na základě zhodnocení různých modelů se ukázalo že by bylo vhodné databázi koncipovat jednoduše jako jednu monolitickou tabulku. Tato koncepce přináší zvýšenou odolnost proti lidské chybě při vkládání údajů (odpadají možnosti uživatelských chyb v tvorbě vazeb mezi jednotlivými tabulkami) za cenu zvýšené náročnosti konstrukce uživatelského rozhraní. Hlavní výhodou tohoto uspořádání je že usnadní pozdější export dat pro zpracování v externích statistických programech. Data bude možno pohodlně exportovat jako jeden soubor ve formátech .csv, .xml nebo .xls (případně .xlsx), který pak bude možné snadno naimportovat do příslušného statistického softwaru. Další výhodou je také fakt že tento jednoduchý model nevyžaduje žádné proškolení personálu, který

bude data v nemocnici do systému vkládat. Při použití relačních nebo objektových modelů by mohlo vlivem špatného řetězení dat dojít k inkonzistencím a duplikacím vlivem vlastností těchto modelů. Jednoduchý model má v tomto případě významné výhody. Pro pojmenování jednotlivých atributů tabulky bylo vybráno následující schéma:

lab_amylaza_T00,

kde *lab* je třípísmenná zkratka která popisuje typ údaje, *amylaza* je název konkrétní sledované proměnné, *T00* popisuje z které kontroly daná hodnota pochází. Pro pojmenování proměnných získaných z dotazníků bylo zvoleno vzhledem k odlišnému charakteru dat pozměněné schéma:

T00_QOL_01,

kde *T00* popisuje z které kontroly daná hodnota pochází, *QOL* je název konkrétního dotazníku a *01* je číslo příslušné otázky v dotazníku. Seznam všech hodnot se stručným popisem jejich významu je v příloze *A - Seznam databázových polí se stručným popisem*.

5 Design uživatelského rozhraní pro databázi

Pro systém takového rozsahu je důležitá přehlednost a organizovanost uživatelského rozhraní. Toho lze dosáhnout konstrukcí uživatelského rozhraní na základě základních designových principů:

Konzistence

Při tvorbě uživatelského rozhraní by měla být dodržena jednotnost vzhledu a organizace jednotlivých oken.

Efektivita

Uživatelské rozhraní je třeba konstruovat s konkrétním záměrem ohledně uživatelského zážitku z práce s rozhraním tak aby proces byl co nejjednodušší.

Bezpečnost

Je třeba rozhraní koncipovat tak aby jeho konstrukce nebyla příčinou chyb ze strany uživatele. Rozhraní by tedy mělo být jednoduché, jasné a přehledné.

Jednoduchost použití

Obsluha daného uživatelského rozhraní by neměla klást na uživatele zvýšené nároky. Uživatelské rozhraní by mělo být intuitivní.

Jednoduchý na porozumění

Prvky v uživatelském rozhraní by měly organizačně odpovídat nebo alespoň být dostatečně podobné úkonům již známým uživateli.


Zmíněné zásady vedly k následující implementaci:

5.1 Uživatelské rozhraní pro studii FNO

Uživatelské rozhraní má jednoduchý tabulkový vzhled, organizace jednotlivých polí byla volena tak aby vizuálně odpovídala laboratorním zprávám generovaným vyšetřeními a tak usnadnila nemocničnickému personálu práci při zadávání dat do systému.

Na následujících stránkách bude představeno uživatelské rozhraní databáze a budou komentována designová rozhodnutí na jejichž základě vzniklo.

Otevřená randomizovaná pilotní studie posuzující bezpečnost a efektivitu endovaskulární léčby CCSVI



ID	Skupina	Jméno	Příjmení	Rodné číslo		Telefon				
26	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	T0	T Int	T1	T3	T6	T12	T13	T15	T18	T24
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
27	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	T0	T Int	T1	T3	T6	T12	T13	T15	T18	T24
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
28	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	T0	T Int	T1	T3	T6	T12	T13	T15	T18	T24
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
29	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	T0	T Int	T1	T3	T6	T12	T13	T15	T18	T24
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Obrázek 3: Ukázka uživatelského rozhraní databáze - administrativní prvky

5.2 Komentář k uživatelskému rozhraní databáze

Obrázek 3 ukazuje přehledovou stránku kontrol pacientů vzniklou na základě dodatečného požadavku lékařů na poskytnutí informativního přehledu o datech kontrol pacientů. Tato obrazovka obsahuje záznamy pacientů obsahující údaje které je v systému identifikují, konkrétně jméno, příjmení, identifikační číslo v databázovém systému a rodné číslo. Dále obsahuje telefonní číslo pro případné kontaktování pacienta ohledně kontrol a pole pro samotná data jednotlivých kontrol. Pole pro data kontrol jsou mírně odsazena aby vzniklo vizuální odlišení jednotlivých záznamů. Jednotlivé záznamy jsou také pro zvýšenou přehlednost separovány horizontální linkou.

Otevřená randomizovaná pilotní studie posuzující bezpečnost a efektivitu endovaskulární léčby CCSVI
Chronické Cerebrospinální Venózní Insuficience u pacientů s RS-Roztroušenou Mozkomíšni Sklerózou

ID 29 Rodné číslo /
 Skupina Příjmení Jméno

Anamnéza Administrace EDSS Dotazníky **Vyšetření/Testy** Medikace

ID 29 Skupina A B C

Rodné číslo / Pohlaví Kuřák Ano Ne
 Jméno Výška Těhotenství Ano Ne
 Příjmení Váha HLA2 Ano Ne

Doba trvání nemoci Roky Měsíce Klasifikace RS EDSS

Zde vložená data jsou konstantní a nekontrolována

22

Obrázek 4: Ukázka uživatelského rozhraní databáze - anamnéza

Obrázek 4 ukazuje výchozí záložku pacienta s anamnestickými údaji a zařazením do skupiny. Pravá horní část této obrazovky je společná pro všechny další a obsahuje údaje pro identifikaci pacienta v rámci studie. Jednotlivé záložky (Anamnéza, Dotazníky, Vyšetření) jsou barevně odlišeny pro podporu lepší orientace uživatele v systému. Pole této záložky jsou vizuálně rozdělena do tří skupin - identifikační číslo pacienta v rámci systému a skupina; údaje o pacientovi; obecné údaje o jeho onemocnění RS. Pohlaví, klasifikace formy RS jsou vybírány z drop-down listu. V pravé dolní části je check box pro potvrzení že tato část je kompletní, tento check box je využit pro funkčnost administrativní záložky. Tato záložka se vizuálně jeví prázdná, toto je kompromis pro udržení stejných velikostí s ostatními záložkami, které obsahují více prvků.

Anamnéza Administrace EDSS Dotazníky Vyšetření/Testy Medikace

Data návštěv		T A -> C											Poznámky
		Anamnéza	T0	T Int	T1	T3	T6	T12	T13	T15	T18	T24	
T0	<input type="checkbox"/>	<input type="checkbox"/>											
T Int	<input type="checkbox"/>	MSIS-29	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T1	<input type="checkbox"/>	FSS	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T3	<input type="checkbox"/>	MFIS	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T6	<input type="checkbox"/>	QOL	<input type="checkbox"/>					<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>
T12	<input type="checkbox"/>	MRI	<input type="checkbox"/>				<input type="checkbox"/>				<input type="checkbox"/>		
T13	<input type="checkbox"/>	Likvor	<input type="checkbox"/>					<input type="checkbox"/>					
T15	<input type="checkbox"/>	MEDS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T18	<input type="checkbox"/>	Lab	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T24	<input type="checkbox"/>	ECD	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Telefon	<input type="text"/>	Celkové Klinické			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Obrázek 5: Ukázka uživatelského rozhraní databáze - administrativní prvky

Obrázek 5 ukazuje administrativní záložku s přehledem zadaných dat a políček pro data návštěv. Tato záložka vznikla na požadavek lékařů o administrativní přehled. Záložka je vizuálně dělená na dvě části - data kontrol s check boxy a tabulku kontrol. Tabulka pomocí check boxů zobrazuje které kontroly a vyšetření jsou již v systému vloženy. Vpravo od této tabulky je poskytnut prostor pro administrativní poznámky k pacientovi.

Anamnéza	Administrace	EDSS	Dotazníky	Vyšetření/Testy	Medikace		
FSS	MFIS	MSIS-29	QOL				
T0 AB	T3 AB	T6 AB	T12 AB	T13 C	T15 C	T18 ABC	T24 ABC

1. Moje motivace je nižší když jsem unavený/á. 1 2 3 4 5 6 7
2. Cvičení mi přináší únavu 1 2 3 4 5 6 7
3. Jsem lehce unavený/á. 1 2 3 4 5 6 7
4. Únava se střetává s mým fyzickým fungováním. 1 2 3 4 5 6 7
5. Únava mi způsobuje časté problémy. 1 2 3 4 5 6 7
6. Moje únava mi znemožňuje trvalé fyzické fungování. 1 2 3 4 5 6 7
7. Únava se střetává s prováděním určitých povinností a odpovědností. 1 2 3 4 5 6 7
8. Únava patří mezi mé tři největší oslabující symptomy. 1 2 3 4 5 6 7
9. Únava se střetává s mou prací, rodinou, či společenským životem. 1 2 3 4 5 6 7

Celkové hodnocení

Zde vložená data jsou konkrétní a zkontrolována

Obrázek 6: Ukázka uživatelského rozhraní databáze - dotazník fatigue severity scale

Obrázek 6 ukazuje záložku jednoho z dotazníků používaných ve studii. Konkrétně se jedná o únavovou škálu (FSS). Tato škála je zde implementována jako pole radio buttonů. Toto pole je možné rychle a pohodlně vyplnit pomocí standardních klávesových zkratk (Tab, Enter, Mezerník, šipky). Dole pak je pole s jednoduchým scriptem pro výpočet hodnocení dotazníku. Záložky pro jednotlivé kontroly T0, T2. . . jsou barevně odlišeny pro usnadnění orientace uživatele.

Anamnéza Administrace EDSS Dotazníky **Vyšetření/Testy** Medikace

Likvor MRI ECD Lab Katetrizační venografie Oční pozadí Ostatní

T0 T6 T18

3D - SWI

Depozita FC Ano Ne

Uzávěr periventikulárních žil Ano Ne

Demineralizační ložiska - Flair

Počet plak

Počet GD+ Plak
 Pozitivní Nové Léze

3D - MR - venografie

Stenóza, obstrukce

VJI Ano Ne

TB Ano Ne

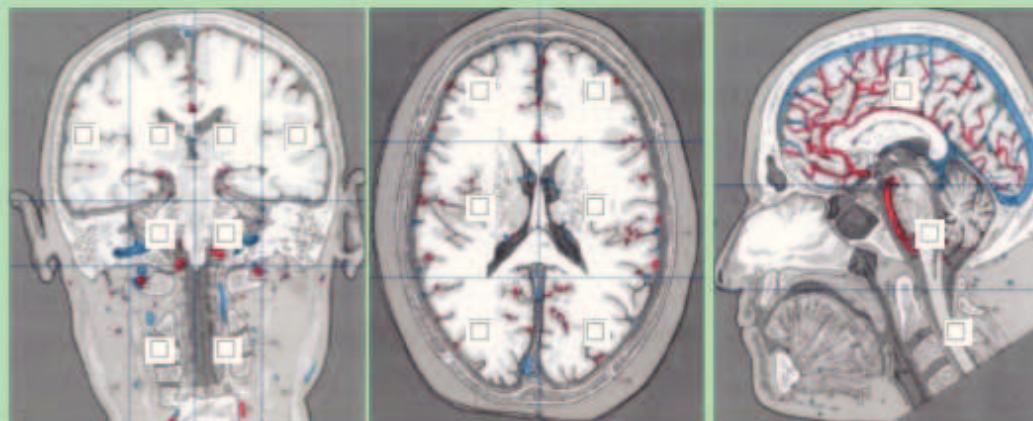
VAZ Ano Ne

VCS Ano Ne

Kompenzační oběh

Krční páteř Ano Ne

IK Ano Ne



Zde vložené data jsou konkrétní a zkontrolovány

Obrázek 7: Ukázka uživatelského rozhraní databáze - Magnetická rezonance

Obrázek 7 ukazuje záložku magnetické rezonance, kde jsou sbírána data získaná vyšetřením pomocí magnetické rezonance. Tato záložka sbírá data o obstrukcích žil, počtu a přibližné lokalizaci plak. Data získané v této záložce budou vhodná pro zpracování kontingenčních tabulek a případné vyšetřování závislostí mezi lokalizací plak a obstrukcemi žil nebo skóre dotazníku kvality života.

Anamnéza	Administrace	EDSS	Dotazníky	Vyšetření/Testy	Medikace					
Likvor	MRI	ECD	Lab	Katetrizační venografie	Oční pozadí	Ostatní				
T0 AB	T1 AB	T3 AB	T8 AB	T12 AB	A->C	T13 C	T15 C	T18 ABC	T24 ABC	
Datum						Základní biochemie				
Krevní obraz		Krevní diferenciál		Koagulace		Nátrium v séru (mmol/l) _____ Kálium v séru (mmol/l) _____ Chloridy v séru (mmol/l) _____ Urea-serum (μmol/l) _____ Kreatinin serum (μmol/l) _____ Bilirubin celkový (μmol/l) _____ Amyláza sérum (μkat/l) _____ Alanintransferáza _____ Aspartáttransferáza _____ Alkalická fosfatáza (μkat/l) _____ Glutamyltransferáza (μkat/l) _____ Glukóza-sérum (mmol/l) _____ Cholesterol (mmol/l) _____ Triacylglycerol (mmol/l) _____ C reak. prot. (mg/l) _____ Kreatinínáza (μkat/l) _____ CK-MB frakce (μkat/l) _____ železo v séru (μmol/l) _____ Lipáza (μkat/l) _____ Glyk. Hemoglobin (%) _____ ferritin (μg/l) _____ transferrin (g/l) _____ vit. D2 (μg/l) _____ volné transferrinové receptory (mg/l) _____ volné transferrinové receptory - index sTfR/Ferr _____ Anti-Cardiolipin-Ab IgG _____ Anti-Cardiolipin-Ab IgM _____ Anti-b2-Glycoprotein 1 Ab IgG a IgM _____				
Leukocyty	_____	Erytrocyty	_____							
Dif. segmenty (%)	_____	Hemoglobin (g/l)	_____							
Neutrofilly	_____	Hematokrit	_____							
Dif. lymfocyty (%)	_____	Str. obj. ERY (fl)	_____							
APLy - Lymfocyty	_____	Hemogl. ERY (pg)	_____							
Dif. monocyty (%)	_____	Hemogl. koncentrace (g/dl)	_____							
APMo - Monocyty	_____	RDW (%)	_____							
Dif. eozinofily (%)	_____	Trombocyty	_____							
APeO - Eozinofily	_____	Str. obj. tromb. (fl)	_____							
Dif. bazofily (%)	_____									
APBa - Bazofily	_____									
Dif. komentář	_____									
Zde vložené data jsou konstantní a zkontrolované <input type="checkbox"/>										

30

Obrázek 8: Ukázka uživatelského rozhraní databáze - laboratorní výsledky

Obrázek 8 ukazuje záložku laboratorních výsledků. Konkrétně se jedná o hodnoty základní biochemie, krevního obrazu, krevního diferenciálu a koagulace. Záložky pro jednotlivé kontroly T0, T2... jsou rovněž barevně odlišeny pro usnadnění orientace uživatele.

Anamnéza		Administrace		EDSS		Dotazníky		Vyšetření/Testy		Medikace	
T0 AB	T Int B	T1 B	T3 AB	T6 AB	T12 AB	T13 C	T15 C	T18 ABC	T24 ABC		
Antidepressivum			Nesteroidní antirevmatikum			Antiulcerozum			Centrální myorelaxans		Jiné
Citaloprami hydrobromidum <input type="checkbox"/>			Ibuprofenum <input type="checkbox"/>			Famotidinum <input type="checkbox"/>			Tizanidini hydrochloridum <input type="checkbox"/>		
Fluoxetini hydrochloridum <input type="checkbox"/>			Nimesulidum <input type="checkbox"/>			Omeprazolom <input type="checkbox"/>			Baclofenom <input type="checkbox"/>		
Sertralini hydrochloridum <input type="checkbox"/>			Piroxicamum betadexum <input type="checkbox"/>			Ranitidini hydrochloridum <input type="checkbox"/>					
Venlafaxini hydrochloridum <input type="checkbox"/>			Indometacidum <input type="checkbox"/>								
Glukokortikoid			Imunomodulans			Soli a ionty			Vitamin D		
Prednisonum <input type="checkbox"/>			Interferonum beta 1a <input type="checkbox"/>			Mg lactici <input type="checkbox"/>			Colecalciferolum <input type="checkbox"/>		
Metylprednisolonum <input type="checkbox"/>			Glatirameri acetat <input type="checkbox"/>			KCl <input type="checkbox"/>			Ergocalciferolum <input type="checkbox"/>		
Anxiolytika			Imunosupresivum			Vazodilatans			Venofarmakum		
Hydroxyzini dihydrochloridum <input type="checkbox"/>			Azathioprinum <input type="checkbox"/>			Naftidrofuryli oxalat <input type="checkbox"/>			Diosminum <input type="checkbox"/>		
Antikoagulancia											
Warfarinum natrium <input type="checkbox"/>											

Zde vložené data jsou kontrolovány a aktualizovány

Obrázek 9: Ukázka uživatelského rozhraní databáze - sledování léků

Obrázek 9 ukazuje záložku medikace užívané pacientem. Léčiva jsou rozdělena do skupin podle jejich funkce. Na základě konzultací s lékaři bylo rozhodnuto že u léčiva se budou zaznamenávat podle účinné látky, protože na základě jmen jednotlivých přípravků by nebylo možno vytvořit v obecné rovině dostatečně vyčerpávající seznam. Implementovaný seznam zahrnuje nejběžnější přípravky užívané pacienty trpící RS, nicméně není úplně vyčerpávající a proto je na pravé straně poskytnut prostor pro zápis jiných přípravků nebo poznámek.

6 Metody Bootstrap

Základní princip metody bootstrap poprvé představil ve své práci Bradley Efron v roce 1979 [15]. Jedná se o metodu patřící do širší třídy metod převzorkování statistického souboru, které mají své uplatnění při vyhodnocování statistických vzorků s malou četností. Tato metoda je typická svou vyšší výpočetní náročností. Používá se hlavně pro sestavování intervalů spolehlivosti a hypotézové testy. Principem metody je výpočet hledaného parametru na základě četných převzorkování původního datového vzorku. Například jednoduchý intervalový odhad střední hodnoty lze realizovat následujícím postupem:

1. Náhodně vygenerujeme celkem B převzorkování původního datového vzorku. Jinými slovy, z n realizací (x_1, x_2, \dots, x_n) náhodné veličiny X pomocí náhodných výběrů s opakováním vytvoříme celkem B serií n „fiktivních“ realizací.
2. Pro každé z výše uvedených bootstrapových převzorkování spočítáme výběrový průměr.
3. Hodnoty výběrových průměrů seřadíme podle velikosti.
4. Jako odhad dolní meze vybereme $n\alpha$ tou hodnotu³¹, jako odhad horní meze vybereme $n(1 - \alpha)$ tou hodnotu. V podstatě tedy určíme α a $1 - \alpha$ kvantily

Uvedený postup lze snadno implementovat ve skriptovacím jazyce Python³² pouze s využitím základního balíčku `random`, jež je součástí distribuce Pythonu. Balíček `random` generuje náhodná čísla pomocí kvalitního algoritmu Mersenne Twister.

Jako příklad uvedme vyšetřování jedné z důležitých hodnot měřených u pacientů s CCSVI a to konkrétně střední hodnotu průměru [cm] žíly jugularis interna v sedě (jedná se o jednu z hodnot vyšetření ECD):

0.69	1.09	0.43	0.75
0.38	0.52	0.3	0.46
0.24	0.18	0.7	0.22
0.46	0.22	0.26	0.13
0.13	0.08	0.29	1
0.36	0.25	0.92	0.4

³¹ přesněji hodnotu nejbližší k $n\alpha$

³² příklady v této práci byly zpracovány s verzí Python 2.7.1

```

import random

def bootstrap(sample, nsamples):
    n = len(sample)
    X = []
    for i in range(nsamples):
        # bootstrapove prevzorkovani
        resample = [random.choice(sample) for j in range(0,n-1)]
        # vypocet prumeru
        x = sum(resample)/ float(len(resample))
        X.append(x)
    return X

def confinterval(b, alpha):
    n = len(b)
    # serazeni hodnot prumeru jednotlivych vzorku
    b.sort()
    # vypocty hranic konfidencniho intervalu
    lb = b[int(alpha * n)]
    ub = b[int((1 - alpha ) * n)]
    return [lb,ub]

# seed PRNG pro reprodukovatelnost vysledku
random.seed(1337)

# datovy vzorek
s = (0.69,1.09,0.43,0.75,0.38,0.52,0.3,0.46,0.24,0.18,0.7,0.22,0.46,0.22,0.26,
0.13,0.13,0.08,0.29,1,0.36,0.25,0.92,0.4)

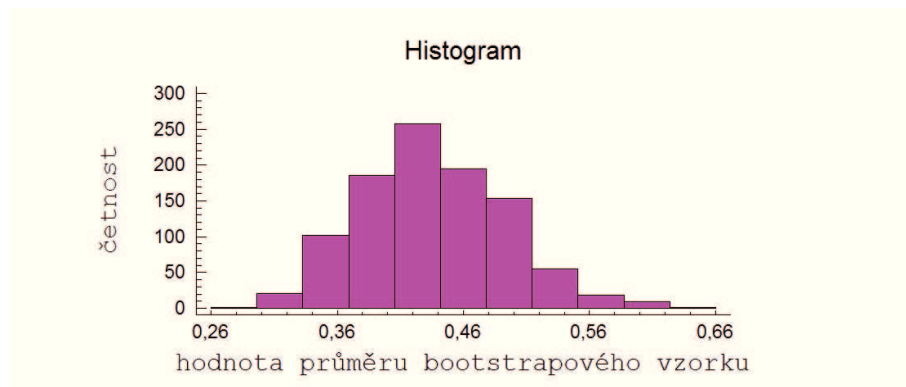
# generujeme 1000 bootstrapovych vzorku
x = bootstrap(s,1000)
# vybereme hranice konfidencniho intervalu  $s \sim \alpha = 0.05$ 
ci = confinterval(x,0.05)
# vypis hranic intervalu

```

```
print ci
```

```
# [0.34347826086956523, 0.5369565217391304]
```

Pro uvedená data tedy po zaokrouhlení získáme bootstrapový odhad $\langle 0,34; 0,54 \rangle$ pro střední hodnotu průměru [cm] žíly jugularis interna v sedě. Průměry jednotlivých bootstrapových vzorků jsou přehledně znázorněny v přiloženém histogramu



Obrázek 10: Histogram průměrů bootstrapových vzorků

Poznamenejme, že opakování³³ výše uvedeného postupu experimentálně potvrdilo očekávanou stabilitu metody.

6.1 Parametrický bootstrap

Metodu bootstrap lze také aplikovat na daný problém parametricky. Při aplikaci metody na náhodnou veličinu X pak vycházíme z předpokladu, že známe rozdělení pravděpodobnosti pozorované náhodné veličiny X . Z hodnot původního výběru (x_1, \dots, x_n) vypočítáme odhady potřebných parametrů distribuční funkce této veličiny. Bootstrapové vzorky dále generujeme na základě takto odhadnuté distribuční funkce. Pro řadu sledovaných parametrů studie však zřejmě bude obtížné věrohodně odhadnout typ rozdělení příslušných náhodných veličin.

6.2 Střední kvadratická chyba odhadu

Nechť θ je parametr rozdělení pravděpodobnosti náhodné veličiny X . Při náhodném výběru (X_1, \dots, X_n) z této náhodné veličiny o rozsahu n , na základě hodnot generovaných výběrem, vypočteme odhad neznámého parametru θ a označme jej $\hat{\theta}$. $\hat{\theta}$ tedy chápeme

³³se zakomentovaným řádkem „random.seed(1337)“

jako náhodnou veličinu proměnných (X_1, \dots, X_n) . Pak střední kvadratickou chybou tohoto odhadu $\hat{\theta}$ rozumíme hodnotu

$$MSE = E(\hat{\theta} - \theta)^2 \quad (1)$$

Dále předpokládejme, že $\hat{\theta}$ je nevychýleným odhadem θ , to znamená, že $E(\hat{\theta}) = \theta$, potom $(MSE) = E(\hat{\theta} - E(\hat{\theta}))^2$, takže střední kvadratická chyba je vlastně rozptylem $\hat{\theta}$. To pak umožňuje použití Čebyševovy nerovnosti:

$$P\left(|\hat{\theta} - \theta| \leq k\sqrt{MSE}\right) \geq 1 - \frac{1}{k^2} \text{ pro libovolné } k \in \mathbb{N} \quad (2)$$

Speciálně pro $k = 2$ pak dostaneme odhad

$$P\left(|\hat{\theta} - \theta| \leq 2\sqrt{MSE}\right) \geq 0,75 \quad (3)$$

Číslo $2\sqrt{MSE}$ nazýváme tolerancí chyby odhadu a používáme ho jako hrubou míru přesnosti odhadu v případech, kdy žádnou vhodnější míru nemáme k dispozici. Uvedme ilustrační příklad. Při odhadu střední hodnoty rozdělení $E(X)$ výběrovým průměrem \bar{X} můžeme využít faktu, že rozptyl výběrového průměru je

$$D(\bar{X}) = \frac{D(X)}{n}. \quad (4)$$

Odhadneme-li dále rozptyl rozdělení $D(X)$ výběrovým rozptylem S^2 , pak s využitím centrální limitní věty dostáváme asymptotický konfidenční interval se spolehlivostí 0,95:

$$E(X) \in \left(\bar{X} - 2\frac{S}{\sqrt{n}}; \bar{X} + 2\frac{S}{\sqrt{n}}\right) \quad (5)$$

Toleranci chyby odhadu \bar{X} pak můžeme odhadnout jako $2Sn^{-1/2}$. Není-li n dostatečně velké nebo v případech odhadů jiných parametrů často nemáme k dispozici analytickou metodu výpočtu střední kvadratické chyby a tolerance chyby odhadu. V takových případech hledáme vhodný odhad MSE. Při opakování náhodných výběrů o rozsahu n z náhodné veličiny X můžeme odhad parametru θ získaný z hodnot výběru při i -tém opakování označit $\hat{\theta}_i$ s chybou MSE_i . S dostatečně velkým počtem opakování lze střední kvadratickou chybu odhadu $\hat{\theta}$ přibližně vyjádřit jako

$$\widehat{MSE} = \frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i - \theta)^2 \text{ kde } B \text{ je počet provedených výběrů.} \quad (6)$$

Rovněž přímá metoda výpočtu \widehat{MSE} je často nerealizovatelná, zejména v situacích kdy není možné získat dostatečný počet náhodných výběrů. V následující části uvedeme jednu z metod jak aproximovat \widehat{MSE} .

6.3 Bootstrap střední kvadratické chyby, rozptylu a směrodatné odchylky

V případě že rozdělení pravděpodobnosti pozorované náhodné veličiny X není známo nebo nemáme k dispozici intervalový odhad jejího parametru θ , lze postupovat tak, že nahradíme pozorované hodnoty (x_1, \dots, x_n) výběru (X_1, \dots, X_n) novými, získanými z (x_1, \dots, x_n) náhodným výběrem (s opakováním). Nově získaný náhodný výběr (x_{b1}, \dots, x_{bn}) nazýváme bootstrapový vzorek.

Pro odhad chyby, rozptylu a směrodatné odchylky $\hat{\theta}$ můžeme uvést následující algoritmus:

1. Z hodnot (x_1, \dots, x_n) výběru (X_1, \dots, X_n) vypočítáme odhad $\hat{\theta}$ parametru θ .
2. Realizujeme B bootstrapových výběrů o rozsahu n z pozorovaných hodnot (x_1, \dots, x_n) . Je obvyklé zvolit $B > 1000$.
3. Pro každý výběr vypočítáme $\hat{\theta}_{b,i}$ kde $i = 1, 2, \dots, B$ tedy odhad parametru θ na základě i -tého bootstrapového vzorku.
4. Bootstrapový odhad hledaného parametru θ dostaneme jako aritmetický průměr jednotlivých odhadů, tedy:

$$\hat{\theta}_b = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_{b,i}. \quad (7)$$

Dále pak získané odhady $\hat{\theta}_{b,i}$ můžeme například použít

- k odhadu střední kvadratické chyby MSE odhadu $\hat{\theta}$:

$$\widehat{\text{MSE}}_b = \frac{1}{B} \sum_{i=1}^B (\hat{\theta}_{b,i} - \hat{\theta})^2 \quad (8)$$

- k odhadu rozptylu $D(\hat{\theta})$:

$$\widehat{D}(\hat{\theta})_b = \frac{1}{B-1} \sum_{i=1}^B \left(\hat{\theta}_{b,i} - \frac{1}{B} \sum_{i=1}^B \hat{\theta}_{b,i} \right)^2 \quad (9)$$

- k odhadu směrodatné odchylky $\sigma(\hat{\theta})$:

$$\widehat{\sigma}(\hat{\theta})_b = \sqrt{\widehat{D}(\hat{\theta})_b} \quad (10)$$

V případě, že $\hat{\sigma}$ je nestranným odhadem pak můžeme aplikovat výš uvedenou Čebyševovu nerovnost.

6.4 Konstrukce bootstrapových intervalů spolehlivosti

Uvedmě ještě alternativní modifikaci klasických metod určování konfidenčního intervalu pomocí techniky bootstrapu. Necht' Z je spojitá náhodná veličina se střední hodnotou $E(Z) = 0$, rozptylem $D(Z) = 1$ a hustotou pravděpodobnosti $f(z)$. Necht' X je spojitá náhodná veličina daná vztahem

$$X = \mu + \sigma Z, \text{ kde } \sigma > 0. \quad (11)$$

X má tedy hustotu pravděpodobnosti

$$g(x) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right). \quad (12)$$

Potom střední hodnota $E(X) = \mu$, rozptyl $D(X) = \sigma^2$ a směrodatná odchylka $\sigma(X) = \sigma$. V této podkapitole ukážeme, jak získat metodou bootstrap odhad konfidenčního intervalu pro odhady střední hodnoty μ , rozptylu σ^2 a směrodatné odchylky σ . Přitom budeme používat následující bodové odhady parametrů: μ odhadneme výběrovým průměrem \bar{X} a σ odhadneme výběrovou směrodatnou odchylkou S . Dále budeme prezentovat algoritmy pro sestavení oboustranných konfidenčních intervalů se spolehlivostí $1 - 2\alpha$ získané pomocí $(1 - \alpha)$ -kvantilu a α -kvantilu příslušných rozdělení pravděpodobnosti.

6.4.1 Odhad střední hodnoty

Je známo že pokud Z pochází z normovaného normálního rozdělení pak statistika

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (13)$$

má Studentovo rozdělení pravděpodobnosti s $n - 1$ stupni volnosti a platí

$$P\left(-t_{1-\alpha} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{1-\alpha}\right) = 1 - 2\alpha, \quad (14)$$

kde $t_{1-\alpha}$ je $(1 - \alpha)$ -kvantil Studentova rozdělení s $n - 1$ stupni volnosti (ze symetrie rozdělení plyne $-t_{1-\alpha} = t_{\alpha}$). Z tohoto dostaneme vztah pro dolní a horní mez konfidenčního intervalu pro μ :

$$\mu \in \left(\bar{X} - t_{1-\alpha} \frac{S}{\sqrt{n}}; \bar{X} + t_{1-\alpha} \frac{S}{\sqrt{n}}\right). \quad (15)$$

Nyní předpokládejme, že Z nemá normální rozdělení pravděpodobnosti. Potom rozdělení pravděpodobnosti statistiky t je stále nezávislé na μ i σ , ale už se obecně nejedná o Studentovo rozdělení. I bez přesné znalosti patřičného rozdělení statisticky stále platí:

$$P\left(t_\alpha < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{1-\alpha}\right) = 1 - 2\alpha, \quad (16)$$

takže meze konfidenčního intervalu splňují:

$$\mu \in \left(\bar{X} - t_{1-\alpha} \frac{S}{\sqrt{n}}; \bar{X} - t_\alpha \frac{S}{\sqrt{n}}\right). \quad (17)$$

Pro konkrétní určení příslušných mezí ovšem musíme odhadnout $t_{1-\alpha}, t_\alpha$. Hodnoty kvantilů rozdělení neznámé pravděpodobnostní statistiky t odhadneme použitím bootstrapu. Algoritmus pro bootstrap $\mu = E(X)$:

1. Z hodnot (x_1, \dots, x_n) náhodného výběru (X_1, \dots, X_n) vypočítáme výběrový průměr \bar{X} a výběrovou směrodatnou odchylku S .
2. Realizujeme B bootstrapových výběrů o rozsahu n . Obvykle volíme $B < 1000$.
3. Pro každý bootstrapový vzorek vypočítáme pozorovanou hodnotu výběrového průměru $\bar{X}_{b,i}$ a výběrové směrodatné odchylky $S_{b,i}$ a hodnotu statistiky t

$$t_{b,i} = \frac{\bar{X}_{b,i} - \bar{X}}{S_{b,i}/\sqrt{n}} \text{ kde } i = 1, 2, \dots, B \quad (18)$$

4. α -kvantil a $(1 - \alpha)$ -kvantil rozdělení pravděpodobnosti statistiky t_b odhadneme tak že bootstrapové odhady seřadíme podle velikosti a vybereme hodnotu s pořadím i aby odpovídalo $i/B = 1 - \alpha$ pro dolní mez (označíme $t_{b,1-\alpha}$) a hodnotu $i/B = \alpha$ pro horní mez (označíme $t_{b,\alpha}$).
5. Bootstrapovým konfidenčním odhadem pro střední hodnotu μ pak rozumíme vztah:

$$\mu \in \left(\bar{X} - t_{b,1-\alpha} \frac{s}{\sqrt{n}}; \bar{X} - t_{b,\alpha} \frac{s}{\sqrt{n}}\right). \quad (19)$$

Pro ilustraci uvedme implementaci tohoto algoritmu v jazyce Python. Prozkoumejme znovu příklad (viz. 6) vyšetřování střední hodnoty průměru žíly jugularis interna v sedě:

```
import random
from math import sqrt

def meanstdv(x):# funkce vypoctu prumeru a smerodatne odchylky
    n, mean, std = len(x), 0, 0
    for a in x:
        mean = mean + a
    mean = mean / float(n)# vypočet prumeru
    for a in x:
        std = std + (a - mean)**2
    std = sqrt(std / float(n-1))# vypočet smerodatne odchylky
    return mean, std

def bootstrap(sample, nb):
    n = len(sample)
    # vypočet prumeru a smerodatne odchylky puvodniho souboru
    xp, xd = meanstdv(sample)
    T = []
    for i in range(nb):
        # bootstrapove prevzorkovani
        resample = [random.choice(sample) for j in range(0,n-1)]
        x, s= meanstdv(resample)
        # vypočet t statistky neznameho rozdeleni
        t = (x - xp)/(s/sqrt(n))
        T.append(t)
    return [T, n, xp, xd]

def confinterval(T, ns, xp, xd, alpha):
    n = len(T)
    T.sort()# serazeni hodnot prumeru jednotlivych vzorku
    # vypocty hranic konfidencniho intervalu
    ub = xp - T[int(alpha * n)]*(xd/sqrt(ns))
```

```

lb = xp - T[int((1 -alpha ) *n)]*(xd/sqrt(ns))
return [lb,ub]

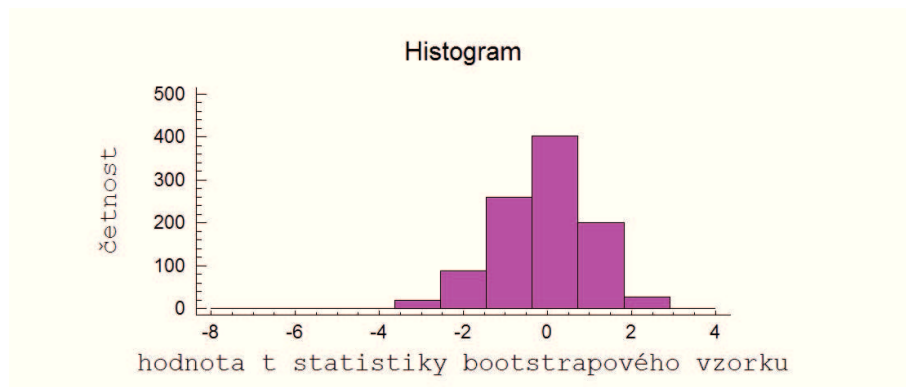
# seed PRNG pro reprodukovatelnost vysledku
random.seed(1337)

# datovy vzorek
s = (0.69,1.09,0.43,0.75,0.38,0.52,0.3,0.46,0.24,0.18,0.7,0.22,0.46,0.22,0.26,
0.13,0.13,0.08,0.29,1,0.36,0.25,0.92,0.4)

# generujeme 1000 bootstrapovych vzorku
x = bootstrap(s,1000)
# vypoctame hranice konfidencniho intervalu s~alpha = 0.05
ci = confinterval(x[0],x[1], x[2], x[3], 0.05)
print ci
# [0.3448360459347557, 0.5545824058473234]

```

Při uvedeném postupu jsme tedy na hladině významnosti $\alpha = 0,05$ odhadli konfidenční interval pro střední hodnotu průměru [cm] žíly jugularis interna v sedě takto: $0 < 0,34; 0,55 >$.



Obrázek 11: Histogram hodnot t statistiky bootstrapových vzorků

6.4.2 Odhad rozptylu

Budeme postupovat podobně jako v předchozím odstavci. Pokud Z pochází z normovaného normálního rozdělení, pak statistika

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \quad (20)$$

má Pearsonovo χ^2 rozdělení pravděpodobnosti s $(n-1)$ stupni volnosti a platí

$$P\left(\chi_\alpha^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\alpha}^2\right) = 1 - 2\alpha, \quad (21)$$

kde χ_α^2 a $\chi_{1-\alpha}^2$ jsou α -kvantil a $1-\alpha$ -kvantil Pearsonova rozdělení s $n-1$ stupni volnosti. Odtud můžeme určit meze konfidenčního intervalu pro rozptyl σ^2 :

$$\sigma^2 \in \left(\frac{(n-1)S^2}{\chi_{1-\alpha}^2}; \frac{(n-1)S^2}{\chi_\alpha^2}\right). \quad (22)$$

Pokud Z nemá normální rozdělení pravděpodobnosti, rozdělení pravděpodobnosti statistiky χ^2 je stále nezávislé na μ i σ , obecně se však už nejedná o Pearsonovo rozdělení. Nicméně, pokud bychom mohli nějakým způsobem zjistit hodnoty kvantilu tohoto neznámého rozdělení, výše uvedené vztahy by stále platily. Hodnoty kvantilu rozdělení pravděpodobnosti neznámé statistiky χ^2 odhadneme pomocí bootstrapu.

Algoritmický postup pro získání konfidenčního intervalu pro $\sigma^2 = D(X)$ a $\sigma = \sigma(X)$ bude tedy následující:

1. Z hodnot (x_1, \dots, x_n) výběru (X_1, \dots, X_n) vypočítáme hodnotu výběrového průměru \bar{X} a výběrové směrodatné odchylky S .
2. Vygenerujeme B bootstrapových vzorků o rozsahu n . Obvykle volíme $B > 1000$.
3. Pro každý bootstrapový vzorek určíme hodnotu výběrového rozptylu $S_{b,i}^2$ a hodnotu příslušné statistiky:

$$\chi_{b,i}^2 = \frac{(n-1)S_{b,i}^2}{S^2} \text{ kde } i = 1, 2, \dots, B. \quad (23)$$

4. α -kvantil a $(1-\alpha)$ -kvantil rozdělení pravděpodobnosti statistiky χ_b^2 dostaneme, tak že bootstrapové odhady seřadíme podle velikosti a vybereme hodnotu s pořadím i tak aby odpovídalo $i/B = 1-\alpha$ pro dolní mez (označíme $\chi_{b,1-\alpha}^2$) a hodnotu $i/B = \alpha$ pro horní mez (označíme $\chi_{b,\alpha}^2$).
5. Bootstrapový konfidenční interval pro σ^2 je pak určen vztahem:

$$\sigma^2 \in \left(\frac{(n-1)S^2}{\chi_{b,1-\alpha}^2}; \frac{(n-1)S^2}{\chi_{b,\alpha}^2}\right). \quad (24)$$

6. Meze bootstrapového konfidenčního intervalu potom lze zapsat takto:

$$\sigma \in \left(\sqrt{\frac{(n-1)S^2}{\chi_{b,1-\alpha}^2}}; \sqrt{\frac{(n-1)S^2}{\chi_{b,\alpha}^2}} \right). \quad (25)$$

6.4.3 Zobecnění

Obecné konfidenční intervaly pro libovolné parametry rozdělení pravděpodobnosti odhadované libovolnými statistikami uvádí Efron [13]. Necht' θ je libovolný parametr rozdělení pravděpodobnosti náhodné veličiny X a necht' $\hat{\theta}$ je nějakým jeho odhadem. Necht' $\hat{\sigma}(\hat{\theta})$ je bootstrapovým odhadem směrodatné odchylky odhadu $\hat{\theta}$. Pokud u náhodné veličiny $\hat{\theta}_b$, tj. odhadu parametru θ na základě bootstrapových náhodných výběrů můžeme předpokládat normální rozdělení pravděpodobnosti, pak statistika

$$t = \frac{\hat{\theta} - \theta}{\hat{\sigma}(\hat{\theta})_b} \quad (26)$$

má Studentovo rozdělení pravděpodobnosti s $(n-1)$ stupni volnosti a platí

$$P \left(-t_{1-\alpha} < \frac{\hat{\theta} - \theta}{\hat{\sigma}(\hat{\theta})_b} < t_{1-\alpha} \right) = 1 - 2\alpha, \quad (27)$$

kde $t_{1-\alpha}$ je $(1-\alpha)$ -kvantil Studentova rozdělení s $n-1$ stupni volnosti. Odtud lze bezprostředně odvodit bootstrapový konfidenční interval se spolehlivostí $1-2\alpha$ pro odhad parametru θ :

$$\theta \in (\hat{\theta} - t_{1-\alpha}\hat{\sigma}(\hat{\theta})_b; \hat{\theta} + t_{1-\alpha}\hat{\sigma}(\hat{\theta})_b). \quad (28)$$

V případě, kdy u náhodné veličiny $\hat{\theta}_b$ nelze předpokládat normální rozdělení, musíme hodnoty kvantilu rozdělení pravděpodobnosti neznámé statistiky t opět odhadnout metodou bootstrap. Uvedme algoritmus pro získání konfidenčního intervalu parametru θ :

1. Z hodnot (x_1, \dots, x_n) výběru (X_1, \dots, X_n) určíme hodnotu vybraného parametru $\hat{\theta}$.
2. Vygenerujeme B bootstrapových vzorků o rozsahu n . Obvykle volíme $B > 1000$.
3. Pro každý bootstrapový vzorek určíme odhad $\hat{\theta}_{b,i}$ parametru θ a jeho směrodatnou odchylku $\sigma(\hat{\theta}_{b,i})$. Pokud pro ni neznáme žádné analytické vyjádření, odhadneme ji opakováním techniky bootstrap (viz (25)). Přitom obvykle volíme $B > 100$.
4. Pro každý bootstrapový vzorek určíme hodnotu statistiky

$$t_{b,i} = \frac{\hat{\theta}_{b,i} - \hat{\theta}}{\sigma(\hat{\theta}_{b,i})} \text{ kde } i = 1, 2, \dots, B \quad (29)$$

5. Vypočítáme odhad směrodatné odchylky $\widehat{\sigma}(\widehat{\theta})_b$ odhadu $\widehat{\theta}$ podle (??)
6. α -kvantil a $(1 - \alpha)$ -kvantil rozdělení pravděpodobnosti statistiky t_b dostaneme tak že bootstrapové odhady seřadíme podle velikosti a vybereme vhodnou hodnotu s pořadím i aby odpovídalo $i/B = 1 - \alpha$ pro dolní mez (označíme $t_{b,1-\alpha}$) a hodnotu $i/B = \alpha$ pro horní mez (označíme $t_{b,\alpha}$)
7. Meze hledaného konfidenčního intervalu parametru θ pak vyhovují vztahu:

$$\theta \in \left(\widehat{\theta} - t_{b,1-\alpha} \widehat{\sigma}(\widehat{\theta})_b; \widehat{\theta} - t_{b,\alpha} \widehat{\sigma}(\widehat{\theta})_b \right). \quad (30)$$

7 Reference

- [1] LENSKÝ, Petr. *Roztroušená skleróza mozkomíšni : nemoc, nemocný a jeho problémy*. Praha : Unie Roska, 1996.
- [2] JEDLIČKA, P. *Léčba roztroušené mozkomíšni sklerózy*. Vyd. 1. Praha 1 : Avicenum, 1991. 144 s. ISBN 80-201-0121-7.
- [3] *Atlas of MS Database* [online]. 2009 [cit. 2011-04-30]. Atlas of MS Database. Dostupné z WWW: <http://www.atlasofms.org/index.aspx>.
- [4] *Multiple Sclerosis*. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, 2001 [cit. 2011-05-01]. Dostupné z WWW: http://en.wikipedia.org/wiki/Multiple_sclerosis.
- [5] LUCCHINETTI, Claudia F.; HOHLFELD, Reinhard. *Multiple Sclerosis*. 1st ed. [s.l.] : Elsevier Saunders, 2010. 496 s. ISBN 978-1-4160-6068-0.
- [6] ROB, Peter; CORONEL, Carlos. *Database Systems: Design, Implementation, and Management*. United States : Course Technology, 2009. 795 s. ISBN 978-1-4239-0201-0, 1-4239-0201-7.
- [7] *Database*. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, 2001 [cit. 2011-05-01]. Dostupné z WWW: <http://en.wikipedia.org/wiki/Database>.
- [8] CHURCHER, Clare. *Beginning Database Design: From Novice to Professional*. United States : Apress, 2007. 240 s. ISBN 978-1-59059-769-9, 978-1-4302-0366-7.
- [9] SEARS, Andrew; JACKO, Julie A. . *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Second Edition. United States of America : Lawrence Erlbaum, 2008. 1358 s. ISBN 978-0-8058-5870-9
- [10] ZAPHIRIS, Panayiotis; SIANG ANG, Chee. *Human Computer Interaction: Concepts, Methodologies, Tools and Applications*. United States of America, : Information Science Reference, 2009. 3024 s. ISBN 1605660523.
- [11] *Human,computer interaction*. In *Wikipedia : the free encyclopedia* [online]. St. Peterburg (Florida) : Wikipedia Foundation, 2001, last modified on 08:08, 22 April 2011 [cit. 2011-05-01]. Dostupné z WWW: http://en.wikipedia.org/wiki/Human-computer_interaction

- [12] PAVLÍČKOVÁ, Lucie. *Metoda bootstrap a její aplikace*. Brno, 2009. 64 s. Diplomová práce. Vysoké učení technické v Brně.
- [13] EFRON, Bradley; TIBSHIRANI, Rob. *An introduction to the bootstrap*. United States of America : Chapman & Hall, 1993. 436 s.
- [14] EFRON, Bradley. *The Jackknife, the Bootstrap and Other Resampling Plans*. Vermont : Capital City Press, 1982. 92 s. ISBN 0-89871-179-7.
- [15] EFRON, Bradley. *The 1977 Rietz Lecture : Bootstrap Methods: Another Look at the Jackknife*. In *The Annals of Statistics*. Vol. 7, No.1. [s.l.] : Institute of Mathematical Statistics, 1979. s. 235. ISSN 00905364.
- [16] JANSSEN, Arnold; PAULS, Thorsten. *How Do Bootstrap and Permutation Tests Work?*. In *The Annals of Statistics*. Vol. 31, No.3. [s.l.] : Institute of Mathematical Statistics, 2003. s. 705-1032. ISSN 00905364.
- [17] CHERNICK, Michael R. *Bootstrap methods : a guide for Practitioners and Researchers*. Second Edition. New Jersey : Wiley, 2008. 369 s. ISBN 978-0-471-75621-7.
- [18] Briš R., Litschmannová M., *Statistika I*, elektronické skriptum VŠB TUO, FEI, 2004
- [19] Briš R., Litschmannová M., *STATISTIKA II.*, E-learningový prvek pro podporu výuky odborných a technických předmětů, v rámci projektu CZ.O4.01.3/3.2.15.2/0326, VŠB TU Ostrava, 2007, ISBN 978-80-248-1482-7.