

Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra informatiky a výpočetní techniky

**Diplomová práce**

**Statistické zpracování  
lékařských dat**

ZÁPADOČESKÁ UNIVERZITA V PLZNI  
Fakulta aplikovaných věd  
Akademický rok: 2010/2011

## ZADÁNÍ DIPLOMOVÉ PRÁCE (PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Miroslav HEJNA**  
Osobní číslo: **A09N0164P**  
Studijní program: **N3918 Aplikované vědy a informatika**  
Studijní obor: **Finanční informatika a statistika**  
Název tématu: **Statistické zpracování lékařských dat**  
Zadávací katedra: **Katedra informatiky a výpočetní techniky**

### Z á s a d y p r o v y p r a c o v á n í :

1. Seznamte se s aplikacemi různých statistických metod v oblasti zpracování lékařských dat.
2. Proveďte analýzu vybrané studie a navrhňte ve spolupráci s lékaři některé metody statistického vyhodnocení.
3. Návrh realizujte a ověřte na vhodném vzorku dat.
4. Získané výsledky vyhodnoťte, případně navrhňte další možnosti statistického zpracování.

Rozsah grafických prací: **dle potřeby**  
Rozsah pracovní zprávy: **min. 40 stran původního textu**  
Forma zpracování diplomové práce: **tištěná**  
Seznam odborné literatury:  
**dodá vedoucí diplomové práce**

Vedoucí diplomové práce: **Doc. Dr. Ing. Jana Klečková**  
Katedra informatiky a výpočetní techniky

Datum zadání diplomové práce: **30. září 2010**  
Termín odevzdání diplomové práce: **19. května 2011**

  
Doc. Ing. František Vávra, CSc.  
děkan



  
Prof. Ing. Jiří Šafařík, CSc.  
vedoucí katedry

V Plzni dne 1. listopadu 2010

## **Čestné prohlášení**

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 27. června 2012

.....

Miroslav Hejna

## **Poděkování**

Tímto bych chtěl poděkovat paní Doc. Dr. Ing. Janě Klečkové za vedení diplomové práce a panu Ing. Petru Včelákovi za odbornou pomoc. Poděkování také patří mé rodině a blízkým za jejich trpělivost a podporu po celou dobu studia.

## **Abstrakt**

### **Statistické zpracování lékařských dat**

Tato diplomová práce se zaměřuje na statistické zpracování dat pacientů, kteří prodělali cévní mozkovou příhodu. Data pocházejí ze znalostní báze experimentálního lékařského systému, jehož technologické pozadí je ve stručnosti představeno. Byly navrženy statistické metody s ohledem na charakter dat a požadované typy statistických analýz. Sérií příkazů ve statistickém prostředí R byla navržena statistická aplikace, která umožňuje budoucí statistické zpracování nových dat ze znalostní báze systému. Aplikace byla otestována na aktuálním souboru dat, kde jsou data také testována dalšími metodami.

## **Abstract**

### **Statistical processing of medical data**

The aim of the diploma thesis was to statistical processing of data on patients who have had stroke. Data come from the knowledge base of medical experimental system, the technological background is briefly introduced. Statistical methods have been designed with regard to the nature of the types of data and statistical analyzes. Series of commands in environment for statistical computing R was designed statistical application that enables the future statistical processing of new data from knowledge base. The application was tested on the current data file where the data is also tested by other method.

# Obsah

|  |           |
|--|-----------|
| <b>1. Úvod .....</b>   | <b>4</b>  |
| <b>2. Experimentální lékařský systém.....</b>                | <b>5</b>  |
| 2.1 Registr SITS .....                                       | 5         |
| 2.2 Experimentální systém .....                              | 5         |
| 2.3 RDF .....  | 6         |
| 2.4 Datové úložiště.....                                     | 8         |
| 2.4.1 Data mining.....                                       | 8         |
| 2.5 Statistická aplikace.....                                | 9         |
| <b>3. Statistické testování .....</b>                        | <b>10</b> |
| 3.1 Statistické hypotézy .....                               | 10        |
| 3.2 Statistické metody pro porovnání skupin.....             | 10        |
| 3.2.1 Kruskal – Wallisův test.....                           | 11        |
| 3.2.2 Simultánní porovnávání.....                            | 12        |
| 3.2.3 $\chi^2$ test dobré shody .....                        | 12        |
| 3.2.4 Randomizační test dobré shody .....                    | 14        |
| 3.2.5 Testy v kontingenčních tabulkách .....                 | 15        |
| <b>4. Statistická analýza trombolyzovaných pacientů.....</b> | <b>18</b> |
| 4.1 Základní popisná statistika .....                        | 18        |
| 4.2 Rozdělení pacientů podle ICD .....                       | 20        |
| 4.2.1 Věk.....   | 21        |
| 4.2.2 Pohlaví .....  | 23        |
| 4.2.3 NIHSS .....  | 25        |
| 4.2.4 Klinické změny .....                                   | 31        |
| 4.2.5 Rankin score .....                                     | 34        |
| 4.2.6 Průkaz zavřené tepny .....                             | 36        |

|           |   |           |
|-----------|---|-----------|
| 4.2.7     | Průkaz rekanalizace .....                         | 37        |
| 4.3       | Analýza pacientů s krvácením .....                | 39        |
| 4.3.1     | Diabetes mellitus.....                            | 40        |
| 4.3.2     | Arteriální hypertenze .....                       | 40        |
| 4.3.3     | Nikotinismus .....                                | 41        |
| 4.3.4     | Antiagregační léčba .....                         | 41        |
| 4.3.5     | Čas do zahájení léčby .....                       | 42        |
| 4.3.6     | Známka ischemie na CT .....                       | 44        |
| 4.3.7     | Klinická změna .....                              | 45        |
| 4.3.8     | Spektrum ICD .....                                | 45        |
| 4.3.9     | Rankin score .....                                | 45        |
| 4.4       | Analýza pacientů s kompletním CT vyšetřením ..... | 46        |
| 4.4.1     | Spektrum ICD .....                                | 47        |
| 4.4.2     | NIHSS .....                                       | 48        |
| 4.4.3     | Klinická změna .....                              | 49        |
| 4.4.4     | Rankin score .....                                | 52        |
| 4.4.5     | Čas do zahájení léčby .....                       | 53        |
| 4.4.6     | Známka ischemie na CT .....                       | 54        |
| <b>5.</b> | <b>Závěr .....</b>                                | <b>57</b> |
|           | <b>Seznam tabulek .....</b>                       | <b>58</b> |
|           | <b>Seznam obrázků .....</b>                       | <b>60</b> |
|           | <b>Seznam použité literatury .....</b>            | <b>61</b> |
|           | <b>Seznam příloh.....</b>                         | <b>63</b> |



## 1. Úvod

Cévní mozkové příhody (nebo též mozková mrtvice) jsou celosvětově druhou příčinou úmrtí a nejčastější příčinou invalidity u lidí středního a vyššího věku. Je to v současné době aktuální a celosvětově rozšířený problém. Česká republika patří dokonce k zemím s nejvyšší úmrtností (dvoj až trojnásobně vyšší oproti vyspělým státům Evropy).

Z iniciativy lékařů proto vznikl globální registr SITS, jehož cílem je shromažďovat informace o průbězích a výsledcích léčby pacientů s prodělanou mozkovou mrtvicí z celého světa. Následnou analýzou či výzkumem nad těmito daty by měla být zkvalitněna léčba a prevence.

Do registru SITS přispívá také Fakultní nemocnice v Plzni, s jejíž spoluprací je na Západočeské univerzitě v Plzni vyvíjen experimentální lékařský systém. Tento systém si klade za cíl vytvořit komplexní znalostní bázi nad daty z registru SITS a propojit tyto údaje s dalšími lékařskými záznamy.

Významnou roli zde hraje statistická analýza dat, která jednak popisuje základní charakteristiky souboru dat, ale také může dávat odpovědi na některé domněnky (např. postihuje mozková mrtvice častěji kuřáky/diabetiky?) či objevit nové souvislosti. A právě statistická analýza dat je předmětem této práce.

V úvodu práce je stručně popsán experimentální lékařský systém a jeho technologické pozadí. Další část se věnuje statistickým metodám. Po stručném úvodu do statistického testování je následně navrženo několik statistických metod, u nichž jsou popsány jejich principy. Metody jsou vybrány v závislosti na typu dat a požadovaných analýz.

V poslední části aplikujeme vybrané metody na reálná data z registru SITS. V případě potřeby zde také data podrobujeme dalším analýzám a testům.

## 2. Experimentální lékařský systém

### 2.1 Registr SITS

Informace o registru SITS jsou čerpány z [1].

SITS (Safe Implementation of Treatment in Stroke) je mezinárodní nezisková organizace se základnou ve Švédsku. Jejím úkolem je propojit znalostní bázi lékařů s akademickou půdou ohledně případů cévních mozkových příhod. Vznikla na základě iniciativ lékařů s cílem vylepšit klinické pokusy a také zajistit lepší prevenci cévních mozkových příhod. SITS zahrnuje širokou škálu nemocnic a zdravotních středisek z celého světa.

Organizace založila registr SITS, který je v současné době největší databází pacientů s prodělanou mozkovou mrtvicí. Nad tímto registrem spolupracují lékaři s akademiky s cílem zkvalitnit léčbu cévních mozkových příhod. Lékaři zde mohou porovnávat postupy a výsledky léčby, zatímco vědečtí pracovníci získávají pro své výzkumy mnoho reálných dat. V současné době (květen 2012) je v registru přes 71 000 pacientů z více než 1 300 zdravotních zařízení z celého světa. V České republice je evidováno téměř 7 000 pacientů.

Do roku 2010 obsahoval registr SITS pouze pacienty léčené trombolýzou. „Trombolýza je proces rozpouštění krevní sraženiny (trombu). Je přirozeným procesem organismu nebo může být dosažena podáním léků.“ [2] Od roku 2010 je nasazena nová databáze, která umožňuje vkládat pacienty i s jiným typem léčby, např. mechanickou rekanalizací. [3]

Fakultní nemocnice v Plzni je jedním ze zdravotních zařízení v České republice, která v registru SITS sdružuje své pacienty s prodělanou mozkovou mrtvicí. Aktuálně (únor 2012) je zde 435 pacientů z období 2006 – 2011.

### 2.2 Experimentální systém

Registr SITS je založen na webovém přístupu, kde lékaři vyplňují informace o pacientech přes webové rozhraní. To sebou ale nese i své nevýhody. Registr existuje zcela nezávisle na lékařských informačních systémech a data získaná z registru jsou většinou textového charakteru. Údaje z registru tak není možné propojit s dalšími

potřebnými lékařskými záznamy o pacientovi v informačních systémech. Data proto nejsou příliš vhodná pro automatizované zpracování či vědecké výzkumy.

Výzkumy cévních mozkových příhod jsou založeny na heterogenních lékařských datech. Jsou to lékařské záznamy o pacientovi a různé biomedicínské obrazové záznamy. Pro tyto data existují standardizované formáty. Pro lékařské záznamy je používán formát DASTA (Data Standard), který je národním standardem v České republice, a HL7 (Health Level Seven), který je používán celosvětově. Pro obrazové záznamy (např. snímky z CT, EEG) je používán formát DICOM (Digital Imaging and Communications in Medicine). [4]

Experimentální lékařský systém, vyvíjený ve spolupráci Fakultní nemocnice v Plzni a Západočeské univerzity v Plzni, se snaží jít dále. Klade si za cíl propojit tyto heterogenní data s údaji z registru SITS do jednoho systému a vytvořit tak komplexní znalostní bázi.

### 2.3 RDF

Při studiu tohoto tématu bylo čerpáno z [5].

Základem znalostní báze jsou dobře strukturovaná data a efektivní nástroje k získávání a vyhledávání dat. K tomuto účelu byl vybrán sémantický model dat postavený na moderním formátu RDF (Resource Description Framework).

Resource Description Framework nabízí systém pro uložení popisu informací na webu a vztahů mezi nimi, zejména pak pro reprezentaci metadat o webových zdrojích, jako je název, jméno autora či aktualizace webové stránky. I přesto, že byl RDF primárně navržen jako základ pro sémantický web, je možné ho také využít pro popis informací o věcech, které nemusí být na webu přímo dosažitelné. RDF je tak často používán např. pro reprezentaci osobních informací nebo vztahů v sociálních sítích.

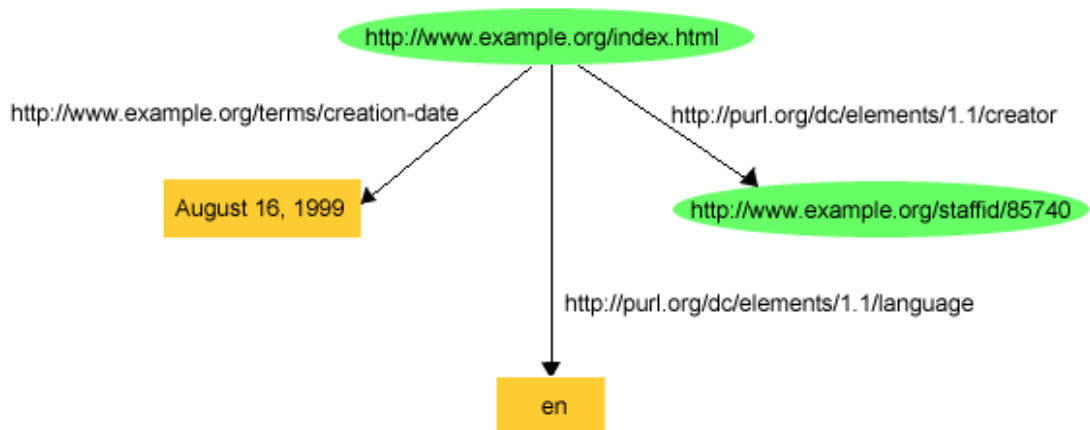
RDF je určeno především pro zpracování informací aplikacemi, ne pro jejich zobrazování uživateli. Nabízí tak efektivní způsob pro výměnu informací mezi aplikacemi bez ztráty jejich významu. Základní myšlenkou je identifikace věcí pomocí webových identifikátorů (URI<sup>1</sup>) a popis zdrojů jako vlastností a jejich hodnot. RDF tak umožňuje zobrazení informace jako grafu s uzly představující zdroje a jejich vlastnosti.

---

<sup>1</sup> Uniform Resource Identifiers - <http://www.ietf.org/rfc/rfc2396.txt>

Vlastnosti jsou popisovány pomocí tzv. RDF trojic, které se skládají z předmětu (subject), predikátu (predicate) a objektu (object). Předmět představuje zdroj, ke kterému se informace vztahuje, predikát popisuje jeho vlastnost a objekt je pak hodnota této vlastnosti. Objektem může být konečná hodnota (literál) nebo další URI reference.

Uvedeme si příklad na popisu zdroje, kterým je webová stránka. Vycházíme z Obr. 1, který zobrazuje ukázkou RDF grafu takového zdroje. Popisujeme webovou stránku `http://www.example.org/index.html` (předmět), kterou vytvořil (predikát `http://purl.org/dc/elements/1.1/creator`) uživatel s id 85740 (objekt `http://www.example.org/staffid/85740`). Další vlastností je datum vytvoření webové stránky (predikát `http://www.example.org/terms/creation-date`) s hodnotou 16. srpna 1999 (objekt August 16, 1999). Poslední vlastností je pak jazyk (predikát `http://purl.org/dc/elements/1.1/language`) webové stránky s hodnotou angličtina (objekt en).



Obr. 1: RDF graf

RDF graf z Obr. 1 zobrazuje tři RDF trojice, jež jsou shrnuty v Tab. 1.

| Předmět  | Predikát  | Objekt  |
|--|---|---|
| <code>http://www.example.org/index.html</code> | <code>http://purl.org/dc/elements/1.1/creator</code>    | <code>http://www.example.org/staffid/85740</code> |
| <code>http://www.example.org/index.html</code> | <code>http://www.example.org/terms/creation-date</code> | August 16, 1999                                   |
| <code>http://www.example.org/index.html</code> | <code>http://purl.org/dc/elements/1.1/language</code>   | en  |

Tab. 1: RDF trojice

## 2.4 Datové úložiště

Experimentální lékařský systém jako své datové úložiště používá databázi **Oracle 11g**, která umožňuje efektivní ukládání sémantických dat. Nabízí také nativní podporu pro RDF standard umožňující ukládání RDF trojic a DML<sup>2</sup> operace s těmito daty. [6]

### 2.4.1 Data mining

Pod pojmem „data mining“ rozumíme proces získávání (dolování) dat pro účely další analýzy. V našem případě se jedná o získávání dat ze znalostní báze experimentálního systému, tedy ze sémantické databáze Oracle.

Pro práci s RDF daty byl speciálně navržen dotazovací jazyk SPARQL<sup>3</sup>, který je určen k manipulaci s daty a k vytváření dotazů nad RDF grafy. SPARQL je standardem W3C<sup>4</sup> syntaxí podobný klasickému SQL<sup>5</sup>. Oracle neumožňuje přímé dotazování v SPARQL jazyce, je ale možné využít prostředníka, tzv. SPARQL endpoint<sup>6</sup>. Nicméně Oracle nabízí podporu SPARQL přímo v SQL jazyce díky zabudované tabulkové funkci SEM\_MATCH. [7]

Uvedeme si nyní jednoduché ukázky v obou jazycích. Pokračujme v ukázkovém RDF modelu a jeho grafu na Obr. 1. Chceme získat datum vytvoření webové stránky.

Zápis ve SPARQL:

```
SELECT ?date WHERE
{
  <http://www.example.org/index.html>
  <http://www.example.org/terms/creation-date>
  ?date .
};
```

Zápis v SQL s použitím funkce SEM\_MATCH:

```
SELECT date FROM TABLE (SEM_MATCH('{
  <http://www.example.org/index.html>
  <http://www.example.org/terms/creation-date>
  ?date .
}', SEM_MODELS('web')));
```

<sup>2</sup> Data Manipulation Language - [http://databases.about.com/od/sql/a/sqlfundamentals\\_3.htm](http://databases.about.com/od/sql/a/sqlfundamentals_3.htm)

<sup>3</sup> Simple Protocol and RDF Query Language - <http://www.w3.org/TR/rdf-sparql-query>

<sup>4</sup> World Wide Web Consortium - <http://www.w3.org/>

<sup>5</sup> Structured Query Language - <http://www.sql.org/>

<sup>6</sup> [http://semanticweb.org/wiki/SPARQL\\_endpoint](http://semanticweb.org/wiki/SPARQL_endpoint)

Pro spouštění SPARQL dotazů v databázi Oracle je možné použít open source nástroj MetaMed<sup>7</sup>, který kromě jiného může sloužit právě jako jednoduchý SPARQL dotazovací nástroj.

## 2.5 Statistická aplikace

Jedním z našich cílů je navrhnout statistickou aplikaci, která by se mohla stát součástí experimentálního systému a samostatně provádět statistické analýzy nad daty ze znalostní báze. Je požadováno zpracování ve **statistickém výpočetním prostředí R**<sup>8</sup>.

Prostředí R je open source výpočetní nástroj, který je v základu konzolovou aplikací, nicméně existují grafické nadstavby, které nabízejí komfortnější ovládání (např. RStudio<sup>9</sup>). V prostředí R můžeme nalézt podporu pro širokou škálu statistických metod a také pro grafické výstupy, jako jsou nejrůznější grafy.

Statistická aplikace se sestává ze sady příkazů, které mají za úkol načíst data, zpracovat požadované analýzy a jejich výstupy uložit do samostatných souborů, ať už v podobě tabulek či grafů. Ideou je spouštět aplikaci skrze experimentální systém. Aplikace by měla při spouštění získávat data automaticky skrze nástroj MetaMed přímo ze znalostní báze a tak provádět analýzy na stále nových datech.

Prostředí R bylo použito ve verzi 2.15.0.

---

<sup>7</sup> Meta Data Extraction and Manipulation Project - <http://medical.kiv.zcu.cz/projects/metamed/>

<sup>8</sup> <http://www.r-project.org/>

<sup>9</sup> <http://rstudio.org/>

### 3. Statistické testování

#### 3.1 Statistické hypotézy

„Statistická hypotéza je určitý předpoklad (výrok) o vlastnostech rozdělení zkoumané náhodné veličiny, který chceme prošetřit a určit, zda je platný či nikoliv.“ [8] Hypotéza, která vyjadřuje náš předpoklad a kterou chceme testovat, se nazývá nulová hypotéza ( $H_0$ ). Hypotéza popírající nulovou se pak nazývá alternativní hypotéza ( $H_1$ ).

Většina statistického software vrací jako výsledek statistických testů tzv. p-hodnotu. Ta vyjadřuje pravděpodobnost, s jakou bychom za platnosti nulové hypotézy dostali stejné či ještě více extrémnější hodnoty odporující nulové hypotéze. Pokud je p-hodnota menší než požadovaná hladina významnosti, nulovou hypotézu zamítáme. P-hodnota může být také interpretována jako nejmenší hladina významnosti, při které bychom nulovou hypotézu ještě zamítli. [8]

#### 3.2 Statistické metody pro porovnání skupin

Jedním z požadavků lékařů na analýzu dat bylo porovnávání různých skupin pacientů podle vybraných kritérií (znaků). V analyzovaném souboru dat se setkáváme s kvantitativní i kvalitativní znaky. U jednotlivých statistických metod rozlišujeme, na který typ dat je vhodné je použít.

Statistické metody lze obecně rozdělit do 2 základních skupin:

- parametrické metody – u těchto metod se předpokládá, že známe rozdělení zkoumaných dat
- neparametrické metody – nejsou zde předpoklady na rozdělení dat, metody tak mají obecnější platnost

Vzhledem k malému rozsahu analyzovaného vzorku dat (jen u zlomku pacientů jsou vyplněny všechny potřebné údaje) a nepředvídatelnému vývoji budoucích dat volíme neparametrické metody. U některých těchto metod je doporučováno určité minimální množství dat, aby testy byly vypovídající. Bohužel ani tohoto minimálního množství dat nebylo v mnohých případech dosaženo, proto zde také uvedeme postupy, jak se vypořádat s velmi malým vzorkem dat.

### 3.2.1 Kruskal – Wallisův test

Touto metodou testujeme, zda dva nebo více nezávislých náhodných vzorků dat pochází ze stejného pravděpodobnostního rozdělení. Jedná se o neparametrickou verzi metody ANOVA<sup>10</sup>. Kruskal – Wallisův test (dále jen KW) tedy používáme v případě, kdy nemůžeme vycházet z předpokladu, že měřená data pocházejí z normálního rozdělení. Předpokládáme ovšem, že data pocházejí z rozdělení se stejným tvarem a jednotlivé skupiny jsou na sobě nezávislé.

Testujeme, zda naměřená data ve skupinách mají stejné mediány. Nulovou a alternativní hypotézu tedy můžeme zapsat jako [9]:

$$H_0: \tilde{\mu}_1 = \tilde{\mu}_2 = \dots = \tilde{\mu}_m$$

$$H_1: \text{Alespoň pro jednu dvojici } i, j \text{ platí, že } \tilde{\mu}_i \neq \tilde{\mu}_j$$

Test se řadí do skupiny tzv. rankových statistik, kde vycházíme z pořadí jednotlivých měření. Nejdříve tedy uspořádáme měření ze všech skupin podle velikosti. V případě shody některých dat jim přiřadíme průměrné pořadí, např. máme-li shodná měření na pozicích 5, 6, 7, 8, budou tyto data mít pořadí 6,5. Naměřená data dále nahradíme jejich pořadím a vypočítáme součty pořadí v jednotlivých skupinách. Součet pořadí měření ve skupině  $i$  označme  $R_i$ . Testovací statistika vychází z měření rozdílnosti průměrů pořadí v jednotlivých skupinách [9]:

$$KW = \left[ \frac{12}{n(n-1)} \sum_{i=1}^m \left( \frac{R_i}{n_i} \right)^2 \right] - 3(n+1). \quad (3.2.1.1)$$

Za platnosti nulové hypotézy má testovací statistika přibližně  $\chi^2$ -rozdělení se stupni volnosti  $df^{11} = m - 1$ . Pro dobrou aproximaci je doporučováno mít alespoň 4 skupiny a v každé skupině alespoň 5 hodnot [10]. V případě malého počtu naměřených hodnot bychom měli použít tabelované kritické hodnoty pro KW.

Při větším počtu shodných měření upravujeme statistiku KW tak, že ji podělíme výslednou hodnotou ze vzorce:

$$1 - \sum_{i=1}^r (t_i^3 - t_i)/(n^3 - n), \quad (3.2.1.2)$$

<sup>10</sup> Analysis of variance (analýza rozptylu)

<sup>11</sup> Degrees of freedom (stupně volnosti)



kde  $r$  je počet skupin, kde se vyskytují shodná měření,  $t_i$  je počet shodných měření ve skupině  $i$  a  $n$  je počet všech měření [9].

Zamítneme-li na základě KW testu nulovou hypotézu, znamená to, že minimálně jedna skupina se liší alespoň od jedné jiné skupiny. Z KW testu ovšem nezjistíme, které skupiny jsou odlišné, ani kolik takových skupin je. Skupiny můžeme porovnat buď vizuálně např. z krabicových grafů nebo použít tzv. post hoc analýzu. Jedná se o simultánní porovnávání jednotlivých skupin mezi sebou.

### 3.2.2 Simultánní porovnávání

V literatuře se můžeme setkat s několika způsoby simultánního porovnávání po zamítnutí nulové hypotézy KW testem. My se zde zaměříme na variantu „Siegel & Castellan“ [11], která je použita v prostředí R ve funkci *kruskalmc*<sup>12</sup> (kruskal multiple comparison).

Při porovnávání skupin  $i$  a  $j$  testujeme hypotézu  $H_0: \tilde{\mu}_i = \tilde{\mu}_j$  proti  $H_1: \tilde{\mu}_i \neq \tilde{\mu}_j$ , kde  $\tilde{\mu}_i$  je medián  $i$ -té skupiny. Chceme-li porovnat všechny skupiny navzájem, musíme těchto porovnání provést  $m(m-1)/2$ , kde  $m$  je počet skupin. Jako testovací statistiku zde uvažujeme rozdíl průměrných pořadí  $|\bar{R}_i - \bar{R}_j|$ . Tyto rozdíly mají při dostatečném množství dat přibližně normální rozdělení. Testovací statistiku pak porovnááme s upravenou kritickou hodnotou standardizovaného normálního rozdělení:

$$|\bar{R}_i - \bar{R}_j| \geq z_{\alpha/m(m-1)} \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad (3.2.2.1)$$

kde  $\bar{R}_i$  je průměrné pořadí ve skupině  $i$ ,  $\alpha$  je hladina významnosti,  $m$  je počet skupin,  $n$  je celkový počet dat a  $n_i$  je počet dat ve skupině  $i$ . Pokud je splněna nerovnost (3.2.2.1), zamítáme nulovou hypotézu na hladině významnosti  $\alpha$ .

### 3.2.3 $\chi^2$ test dobré shody

Testujeme hypotézu ( $H_0$ ), zda naše data pocházejí z očekávaného (teoretického) pravděpodobnostního rozdělení proti alternativní hypotéze ( $H_1$ ), že data z tohoto rozdělení nepocházejí. Máme-li naměřená data rozdělena do několika kategorií (v případě kategoriálního typu dat) nebo do několika intervalů (v případě spojitého typu

<sup>12</sup> Součást externí knihovny „pgirmess“ - <http://cran.r-project.org/web/packages/pgirmess/index.html>

dat), testujeme, zda četnosti v jednotlivých kategoriích (intervalech) se shodují s očekávanými četnostmi.

Tento test je založen na porovnávání rozdílů naměřených a očekávaných (teoretických) četností dat v jednotlivých kategoriích. Jedná se o neparametrický test, proto zde nejsou předpoklady na rozdělení dat. Podmínkou ale je, aby data v jednotlivých kategoriích byla na sobě nezávislá.

Rozdíly mezi naměřenými a očekávanými četnostmi zachycuje testovací statistika

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi_{k-1}^2, \quad (3.2.3.1)$$

kde  $O_i$  jsou naměřené a  $E_i$  očekávané četnosti v  $i$ -té kategorii,  $k$  je počet kategorií. Testovací statistika má za platnosti nulové hypotézy asymptoticky  $\chi^2$ -rozdělení se stupni volnosti  $df = k - 1$ . Očekávané četnosti můžeme také určit jako  $E_i = np_i$ , kde  $n$  je rozsah výběru a  $p_i$  je teoretická pravděpodobnost výskytu jevu v kategorii  $i$ .

Protože je tento test asymptoticky platný (při zvětšujícím se rozsahu výběru dostáváme přesnější výsledky), musí být počet dat dostatečně velký. Obecně doporučované podmínky pro dobrou aproximaci jsou [10]:

- pro  $df = 1$ , žádná z očekávaných četností nesmí být menší než 5
- pro  $df > 1$ , žádná z očekávaných četností nesmí být menší než 1 a nesmí být více než 20% z očekávaných četností menších než 5.

Pokud tyto podmínky nejsou splněny, test může dávat nepřesné výsledky. Nejjednodušší možností jak zvýšit malý počet očekávaných četností je sloučit dvě či více sousedních kategorií. Ne vždy je ale tato možnost vhodná. Jinou možností je použití tzv. randomizačních testů, kde dostáváme zejména při malém počtu dat přesnější výsledky.

**Příklad 2.1** [10]: Kostka byla vržena 40 krát s těmito četnostmi čísel (od 1 do 6) 8, 6, 3, 10, 7, 6. Testujeme hypotézu, zda je rozdělení četností pro jednotlivá čísla rovnoměrné. V R dostáváme tyto výsledky:

```
> chisq.test(c(8,6,3,10,7,6))
  Chi-squared test for given probabilities
data:  c(8, 6, 3, 10, 7, 6)
X-squared = 4.1, df = 5, p-value = 0.5351
```

Můžeme si také zobrazit, jaké očekávané četnosti byly vypočteny. Máme 6 kategorií, v každé kategorii očekáváme stejný počet četností, tedy  $40/6 = 6,67$ .

```
> data.frame(namerene=chi$observed, ocekavane=chi$expected)
  namerene ocekavane
1         8  6.666667
2         6  6.666667
3         3  6.666667
4        10  6.666667
5         7  6.666667
6         6  6.666667
```

Na základě p-hodnoty ( $p > 0.05$ ) nezamítáme nulovou hypotézu (na 5% hladině významnosti).

### 3.2.4 Randomizační test dobré shody

Na rozdíl od  $\chi^2$  testu dobré shody (3.2.3) se zde nevyužívá aproximace  $\chi^2$  – rozdělením. Není proto nutné mít velký počet dat. Pro malé očekávané četnosti tento test vrací přesnější výsledky než  $\chi^2$  test dobré shody.

Princip je následující. Náhodně generujeme data stejného rozsahu jako naše naměřená data s očekávaným rozdělením četností v jednotlivých kategoriích. Takto generujeme co nejvíce náhodných vzorků dat (pro velmi přesnou p-hodnotu je doporučeno generovat alespoň 1 000 000 dat). Pro každý vygenerovaný soubor dat spočteme  $\chi^2$  statistiku a zjišťujeme, v kolika případech je tato hodnota rovna nebo větší než  $\chi^2$  hodnota z původních naměřených dat. Je-li těchto případů více než 5% z celkového počtu náhodných vzorků, znamená to, že pokud by platila nulová hypotéza, dostali bychom naše naměřená data nebo horší (s většími rozdíly mezi četnostmi) ve více než 5% případů. Nemůžeme tedy zamítnout nulovou hypotézu (na 5% hladině významnosti). Pokud těchto případů bude méně než 5%, nulovou hypotézu zamítáme. [12]

Vzhledem k velkému počtu dat a tím spojené výpočetní náročnosti je nutné tento test realizovat pomocí vhodného softwaru. Tento software pak vrací p-hodnotu testu. Protože randomizační test je založen na generování náhodného vzorku dat ze všech možných kombinací, vrací při každém průběhu mírně odlišné p-hodnoty. Proto čím více dat vygenerujeme, tím bude odhad p-hodnoty přesnější.

**Příklad 2.2:** Pro srovnání s  $\chi^2$  testem dobré shody (3.2.3) otestujeme stejný příklad 2.1 pomocí randomizačního testu. Budeme generovat 100 000 náhodných vzorků:

```
> chisq.test(c(8,6,3,10,7,6),simulate.p.value=T,B=100000)

Chi-squared test for given probabilities with simulated
p-value (based on 1e+05 replicates)

data:  c(8, 6, 3, 10, 7, 6)
X-squared = 4.1, df = NA, p-value = 0.5666
```

Výsledná p-hodnota je mírně odlišná oproti příkladu 2.1, ale končíme se stejným rozhodnutím – nezamítáme nulovou hypotézu.

### 3.2.5 Testy v kontingenčních tabulkách

V  $\chi^2$  testu dobré shody (3.2.3) porovnáváme data podle jednoho znaku. Chceme-li porovnávat data podle dvou znaků ( $A$  a  $B$ ), uspořádáme je do kontingenční tabulky. Kategorie jednoho znaku určují řádky a kategorie druhého znaku určují sloupce. Má-li znak určující řádky  $r$  kategorií a znak určující sloupce  $c$  kategorií, dostáváme kontingenční tabulku typu  $r \times c$ .

| úroveň | $B_1$    | $B_2$    | ... | $B_c$    |       |
|--------|----------|----------|-----|----------|-------|
| $A_1$  | $O_{11}$ | $O_{12}$ | ... | $O_{1c}$ | $n_1$ |
| $A_2$  | $O_{21}$ | $O_{22}$ | ... | $O_{2c}$ | $n_2$ |
| ...    | ...      | ...      | ... | ...      | ...   |
| $A_r$  | $O_{r1}$ | $O_{r2}$ | ... | $O_{rc}$ | $n_r$ |
|        | $c_1$    | $c_2$    | ... | $c_c$    | $n$   |

Obr. 2: Kontingenční tabulka typu  $r \times c$

Na Obr. 2 vidíme příklad takovéto kontingenční tabulky, kde  $n_i$  jsou řádkové součty,  $c_j$  jsou sloupcové součty a  $O_{ij}$  jsou pozorované četnosti v daných kategoriích  $A_i, B_j$ .

### Hypotéza homogenity

Označíme-li  $P_{ij}$  relativní četnost prvků v kategorii  $A_i$ , které jsou v  $j$ -té kategorii znaku  $B$ , můžeme hypotézu homogenity vyjádřit rovnicí  $H_0: P_{1j} = P_{2j} = \dots = P_{ij}$  pro všechna  $j = 1, 2, \dots, c$  [9]. To znamená, že pro každou kategorii v  $B$  má být relativní četnost prvků stejná pro všechny úrovně  $A$ .

### Hypotéza nezávislosti

Uvažujeme náhodný výběr, kde každý prvek z toho výběru je klasifikován podle dvou kategoriálních proměnných  $A$  a  $B$ . Hypotézou nezávislosti zkoumáme, zda hodnoty proměnné  $A$  neovlivňují rozdělení proměnné  $B$  a naopak. Nulová hypotéza tedy zní, že obě proměnné jsou na sobě stochasticky nezávislé. [9]

Označme  $p_{ij}$  pravděpodobnost, že prvek bude náležet do kategorie  $i$  v proměnné  $A$  a zároveň do kategorie  $j$  v proměnné  $B$ . Pravděpodobnost, že prvek padne do kategorie  $A_i$ , resp. padne do kategorie  $B_j$ , označme  $p_{i.}$ , resp.  $p_{.j}$ . Hypotézu nezávislosti obou proměnných pak můžeme vyjádřit rovnicemi [9]:

$$p_{ij} = p_{i.}p_{.j}, \quad \text{kde } p_{i.} = \sum_{j=1}^c p_{ij}, \quad p_{.j} = \sum_{i=1}^r p_{ij}. \quad (3.2.5.1)$$

Hypotézy homogenity i nezávislosti testujeme stejným způsobem. Nejprve musíme vypočítat očekávané četnosti  $O_{ij}$  v kontingenční tabulce. Pravděpodobnost, že prvek bude náležet do kategorií  $A_i$  a  $B_j$ , se za předpokladu platnosti nulové hypotézy spočte jako  $p_{ij} = p_{i.}p_{.j}$ . Hodnoty obou pravděpodobností odhadneme jako  $p_{i.} = n_i/n$  a  $p_{.j} = c_j/n$ . Odhad pravděpodobnosti  $p_{ij}$  pak bude  $p_{ij} = n_i c_j / n^2$ . Protože očekávanou četnost můžeme určit jako  $O_{ij} = p_{ij}n$ , tak po dosazení  $p_{ij}$  dostáváme výraz:

$$O_{ij} = \frac{n_i c_j}{n} \quad (3.2.5.2)$$

pro  $i = 1, 2, \dots, r$  a  $j = 1, 2, \dots, c$ . Testovací statistiku pak spočteme podle vzorce

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - O_{ij})^2}{O_{ij}}. \quad (3.2.5.3)$$

Statistika  $\chi^2$  má za platnosti nulové hypotézy asymptoticky  $\chi^2$ -rozdělení se stupni volnosti  $df = (r - 1)(c - 1)$ .

Pro dobrou aproximaci platí podobné podmínky jako v případě  $\chi^2$  testu dobré shody [10]:

- pro  $df = 1$ , tedy kontingenční tabulky 2x2, žádná z očekávaných četností nesmí být menší než 5
- pro  $df > 1$ , žádná z očekávaných četností nesmí být menší než 1 a nesmí být více než 20% z očekávaných četností menších než 5

V případě, že očekávané četnosti nesplňují tyto podmínky, je možné použít **Yatesovu korekci kontinuity** [13]. Tato korekce se využívá zejména pro  $df = 1$ , tedy pro kontingenční tabulky typu 2x2. Upravená hodnota testovací statistiky má tvar:

$$\chi_{Yates}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(|n_{ij} - O_{ij}| - 0,5)^2}{O_{ij}}. \quad (3.2.5.4)$$

Tato korekce zlepšuje aproximaci  $\chi^2$  rozdělení. Zmenšuje hodnotu testovací statistiky a zvyšuje tak p-hodnotu, tedy snižuje pravděpodobnost zamítnutí nulové hypotézy. Tuto korekci zde zmiňujeme v souvislosti s tím, že výpočet  $\chi^2$  testu v prostředí R pro kontingenční tabulky typu 2x2 probíhá automaticky s touto korekcí. Použití korekce je nicméně možné vypnout parametrem *correct=F*.

## 4. Statistická analýza trombolyzovaných pacientů

Analyzovaný vzorek dat pochází z registru SITS Fakultní nemocnice v Plzni. Data jsou získávána ale již přímo ze znalostní báze experimentálního systému, kam jsou data z registru SITS transformována. Data jsou uložena v sémantickém modelu SITS. Pokud zde navážeme na terminologii RDF trojic (kap. 2.3), tak jako zdroj (předmět) zde vystupuje pacient, jako predikát (vlastnost) je námi analyzovaný parametr a jako objekt konkrétní hodnota tohoto parametru (vlastnosti). Na začátku každé analyzované části proto uvádíme také název vlastnosti (predikátu) či podmínky vlastností, za kterých byla data získána.

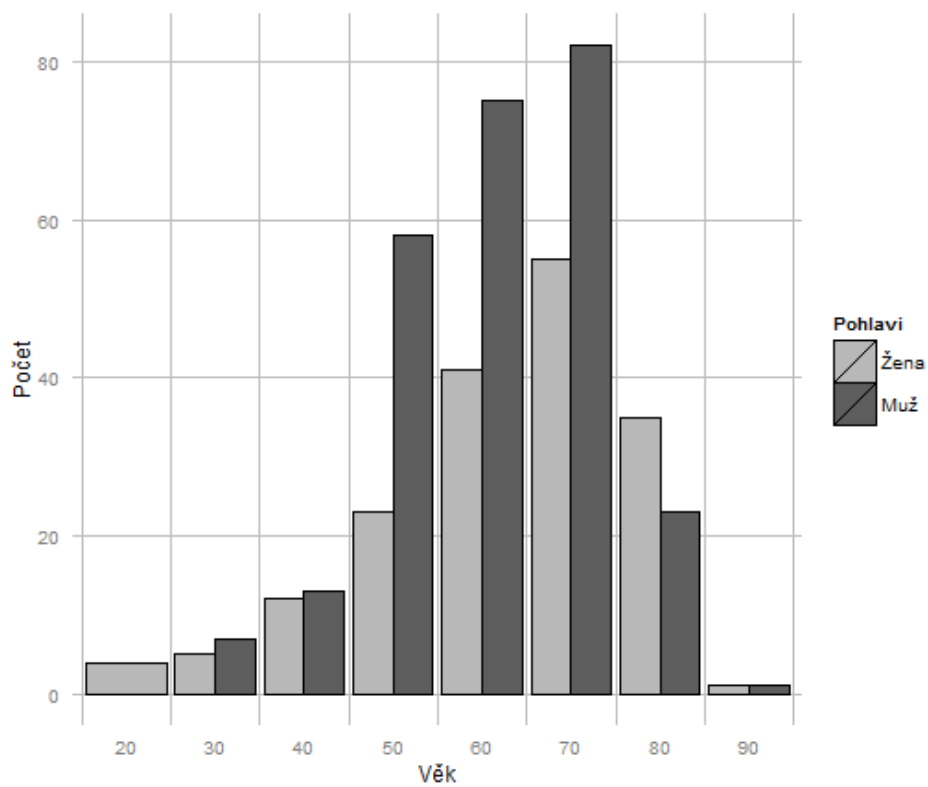
Data obsahují údaje o pacientech, kteří prodělali cévní mozkovou mrtvici a byli léčeni trombolýzou. Pro zachování soukromí pacientů jsou veškerá data ve znalostní bázi anonymizována, obsahují tak jen obecné údaje o pacientech (věk, pohlaví, atd.), průběhu a výsledku léčby.

### 4.1 Základní popisná statistika

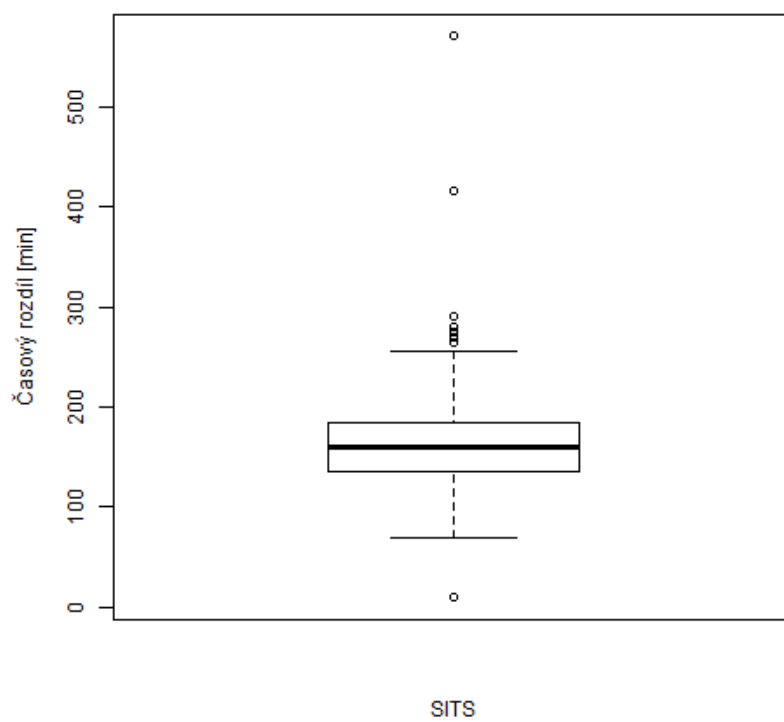
Charakteristika aktuálních dat (únor 2012):

- k mozkové příhodě došlo v letech 2006 – 2011
- počet unikátních pacientů: 435
- průměrný věk pacienta: 67
- počet mužů/žen: 259 / 176
- průměrný věk mužů/žen: 65,9 / 67,4

Počty pacientů podle pohlaví v jednotlivých věkových kategoriích ukazuje graf na Obr. 3. Důležitou charakteristikou je průměrný čas do zahájení léčby, tedy průměrný časový rozdíl mezi vznikem mozkové příhody a začátkem léčby. Průměrný časový rozdíl je 163,72 minut a rozložení jednotlivých rozdílů můžeme vidět v krabicovém grafu na Obr. 4.



Obr. 3: Četnosti pacientů podle pohlaví a věkových kategorií



Obr. 4: Rozložení časových rozdílů pro celý soubor dat ze SITS



## 4.2 Rozdělení pacientů podle ICD

ICD<sup>13</sup> je mezinárodním standardem pro klasifikaci a diagnostiku nemocí pro lékařské účely pod záštitou světové zdravotnické organizace (WHO<sup>14</sup>). Pro diagnostiku cévních mozkových příhod je vymezena kategorie I63, jejíž podkategorie shrnuje Obr. 5. V této části se budeme věnovat statistické analýze pacientů rozdělených do těchto kategorií.

V SITS modelu odpovídá vlastnosti:

*After7D\_\_ICDCode*

| <b>I63</b>   | <b>Cerebral infarction</b>  |
|--------------|---|
|              | <i>Incl.:</i> occlusion and stenosis of cerebral and precerebral arteries, resulting in cerebral infarction |
|              | <i>Excl.:</i> sequelae of cerebral infarction ( <a href="#">I69.3</a> )                                     |
| <b>I63.0</b> | <b>Cerebral infarction due to thrombosis of precerebral arteries</b>  |
| <b>I63.1</b> | <b>Cerebral infarction due to embolism of precerebral arteries</b>  |
| <b>I63.2</b> | <b>Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries</b>                 |
| <b>I63.3</b> | <b>Cerebral infarction due to thrombosis of cerebral arteries</b>   |
| <b>I63.4</b> | <b>Cerebral infarction due to embolism of cerebral arteries</b>   |
| <b>I63.5</b> | <b>Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries</b>                    |
| <b>I63.6</b> | <b>Cerebral infarction due to cerebral venous thrombosis, nonpyogenic</b>                                   |
| <b>I63.8</b> | <b>Other cerebral infarction</b>  |
| <b>I63.9</b> | <b>Cerebral infarction, unspecified</b>   |

Obr. 5: ICD kódování pro diagnostiku cévních mozkových příhod

Četnosti pacientů v jednotlivých kategoriích ukazuje Tab. 2. Pacienti, kteří nejsou zařazeni do žádné kategorie (151 pacientů), nejsou v této části analýzy uvažováni.

| ICD               | I63.0   | I63.3   | I63.4   | I63.5  | I63.8  | I63.9  |
|-------------------|---------|---------|---------|--------|--------|--------|
| Absolutní četnost | 81      | 88      | 76      | 22     | 10     | 7      |
| Relativní četnost | 28,52 % | 30,99 % | 26,76 % | 7,75 % | 3,52 % | 2,46 % |

Tab. 2: Četnosti pacientů ve skupinách ICD

<sup>13</sup> International Classification of Diseases - <http://www.who.int/classifications/icd>

<sup>14</sup> World Health Organization

Tyto skupiny budeme následně porovnávat podle vybraných parametrů (dle požadavků lékařů).

#### 4.2.1 Věk

V SITS modelu odpovídá vlastnosti:

*Baseline\_\_Age*

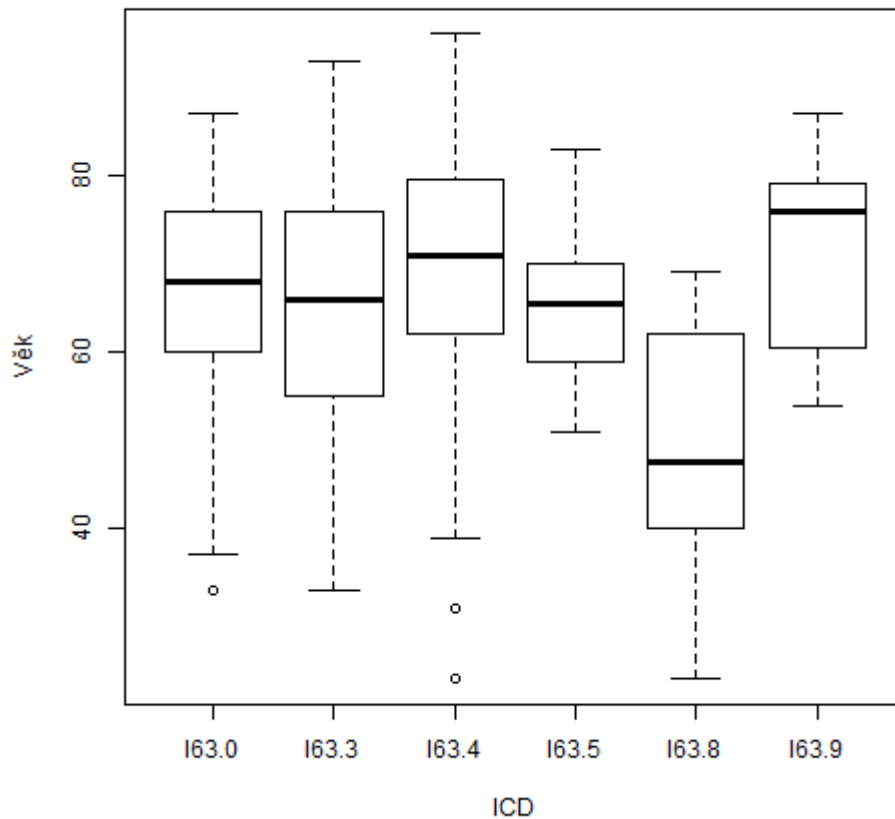
Věk pacientů je spojitá náhodná veličina. Průměr a medián v jednotlivých kategoriích zobrazuje Tab. 3. Pro porovnání rozdělení věků v jednotlivých skupinách použijeme Kruskal – Wallisův test.

| ICD    | I63.0 | I63.3 | I63.4 | I63.5 | I63.8 | I63.9 |
|--------|-------|-------|-------|-------|-------|-------|
| Průměr | 66,85 | 65,67 | 69,47 | 65,00 | 48,50 | 70,86 |
| Medián | 68    | 66    | 71    | 65,5  | 47,5  | 76    |

Tab. 3: Průměr a medián věkového rozložení podle ICD

Pro grafické znázornění rozložení naměřených hodnot se typicky používá tzv. krabicový diagram (boxplot). Výhodou tohoto diagramu je, že přehledně zobrazuje několik charakteristik souboru: nejmenší hodnota, dolní kvartil, medián, horní kvartil, největší hodnotu pozorování a příp. také odlehlá pozorování. Můžeme si tak udělat základní představu o variabilitě souboru, příp. rychlé porovnání skupin.

Na Obr. 6 jsou znázorněny krabicové diagramy pro věková rozdělení v jednotlivých skupinách ICD.



Obr. 6: Graf věkových rozložení skupin podle ICD

Testujeme nulovou hypotézu, zda je věkové rozložení v jednotlivých skupinách stejné. Provedením KW testu v R dostáváme tento výsledek:

```
> kruskal.test(AGE, fICD)

Kruskal-wallis rank sum test

data: AGE and fICD
Kruskal-wallis chi-squared = 20.8968, df = 5, p-value = 0.0008472
```

Na hladině významnosti  $\alpha = 5\%$  tedy zamítáme nulovou hypotézu ( $p < \alpha$ ). Nulovou hypotézu bychom mohli dokonce zamítnout i na úrovni  $\alpha = 1\%$ .

Protože zde zamítáme nulovou hypotézu, může nás také zajímat, které nebo kolik skupin je odlišných. K tomu vyžijeme simultánní porovnávání (3.2.2), kde porovnáváme všechny skupiny navzájem mezi sebou. ICD skupin máme 6, proto musíme provést  $6 \cdot (6-1)/2 = 15$  těchto porovnaní.

V R dostáváme tento výstup:

```
> kruskalmc(AGE ~ ICD,data=data)
Multiple comparison test after Kruskal-wallis
p.value: 0.05
Comparisons
      obs.dif critical.dif difference
I63.0-I63.3  9.866021    37.11830    FALSE
I63.0-I63.4 19.379792    38.49720    FALSE
I63.0-I63.5 19.309203    57.95533    FALSE
I63.0-I63.8 96.372840    80.79915     TRUE
I63.0-I63.9 23.827160    94.96824    FALSE
I63.3-I63.4 29.245813    37.74873    FALSE
I63.3-I63.5  9.443182    57.46088    FALSE
I63.3-I63.8 86.506818    80.44523     TRUE
I63.3-I63.9 33.693182    94.66731    FALSE
I63.4-I63.5 38.688995    58.36110    FALSE
I63.4-I63.8 115.752632   81.09070     TRUE
I63.4-I63.9  4.447368    95.21642    FALSE
I63.5-I63.8 77.063636    91.93741    FALSE
I63.5-I63.9 43.136364   104.60856    FALSE
I63.8-I63.9 120.200000   118.79661     TRUE
```

Vidíme, že ve všech případech, kdy byla indikována odlišnost na hladině významnosti 5%, vystupuje skupina I63.8. Tato skupina se odlišuje od většiny ostatních skupin. Při vizuálním porovnání z Obr. 6 můžeme také vidět, že pacienti ve skupině I63.8 dosahují v průměru nižšího věku než pacienti v ostatních skupinách.

#### 4.2.2 Pohlaví

V SITS modelu odpovídá vlastnosti:

*Baseline\_\_Gender*

Pozorované četnosti v jednotlivých kategoriích zobrazuje Tab. 4.

| ICD   | Žena | Muž |
|-------|------|-----|
| I63.0 | 20   | 61  |
| I63.3 | 32   | 56  |

|       |    |    |
|-------|----|----|
| I63.4 | 42 | 34 |
| I63.5 | 4  | 18 |
| I63.8 | 7  | 3  |
| I63.9 | 4  | 3  |

Tab. 4: Pozorované četnosti pohlaví ve skupinách ICD

Po výpočtu očekávaných četností podle vzorce (3.2.5.2) dostáváme tabulku Tab. 5, kde si můžeme všimnout, že ve třech buňkách je očekávaná četnost menší než 5 (to je 25% buněk). Tím naše data porušují podmínky pro dobrou aproximaci, jak bylo uvedeno v kap. 3.2.5.

| ICD   | Žena  | Muž   |
|-------|-------|-------|
| I63.0 | 31,09 | 49,91 |
| I63.3 | 33,77 | 54,23 |
| I63.4 | 29,17 | 46,83 |
| I63.5 | 8,44  | 13,56 |
| I63.8 | 3,84  | 6,16  |
| I63.9 | 2,69  | 4,31  |

Tab. 5: Očekávané četnosti pohlaví ve skupinách ICD

Chceme testovat, zda četnosti pohlaví jsou v jednotlivých skupinách proporcčně stejně rozděleny, resp. zda je rozdělení nezávislé na skupině ICD. Provedeme-li  $\chi^2$  test v kontingenční tabulce, dostáváme z R tento výstup:

```
> chisq.test(fICD, GENDER)

Pearson's Chi-squared test

data:  fICD and GENDER
X-squared = 24.7936, df = 5, p-value = 0.0001527

warning message:
In chisq.test(fICD, GENDER) : Chi-squared approximation may be
incorrect
```

Dostáváme varování, že aproximace nemusí být korektní. Jak bylo řečeno v kap. 3.2.3, řešením může být sloučení skupin, kde očekávané četnosti porušují podmínky. V našem případě bychom mohli sloučit skupiny „I63.8“ a „I63.9“. Ovšem skupiny mají různé diagnostiky cévních mozkových příhod, proto nedává příliš smysl skupiny slučovat.

Druhým řešením je použít randomizační test, kde nejsou kladeny podmínky na počet očekávaných četností. Provedeme tedy tento test s 1 000 000 vygenerovanými náhodnými vzorky:

```
> chisq.test(fICD,GENDER,simulate.p.value=T,B=1000000)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data:  fICD and GENDER
X-squared = 24.7936, df = NA, p-value = 9.7e-05
```

Na hladině významnosti 5% tedy zamítáme hypotézu, že rozložení pohlaví je ve všech skupinách stejné.

### 4.2.3 NIHSS

NIHSS<sup>15</sup> je standardizovaný systém bodového hodnocení neurologického vyšetření u pacientů s cévní mozkovou příhodou [14]. Hodnotí se např. úroveň vědomí, zorné pole, motorika těla, řeč atd. NIHSS se u pacientů zjišťuje v několika časových odstupech:

- vstupní NIHSS
- NIHSS po 24 hodinách
- NIHSS po 7 dnech

Zajímá nás, zda je rozložení NIHSS ve vybraných časových fázích stejné ve všech skupinách ICD. Zajímá nás také rozložení rozdílů v hodnocení mezi těmito fázemi. Těmto analýzám se budeme věnovat v následující části. Počty pacientů s vyplněným NIHSS v jednotlivých skupinách ICD ukazuje Tab. 6.

---

<sup>15</sup> National Institutes of Health Stroke Scale

| ICD   | I63.0 | I63.3 | I63.4 | I63.5 | I63.8 | I63.9 |
|-------|-------|-------|-------|-------|-------|-------|
| Počet | 12    | 15    | 18    | 1     | 2     | NA    |

Tab. 6: Počty pacientů s vyplněným NIHSS

### Vstupní NIHSS

V SITS modelu odpovídá vlastnosti:

```
Baseline_NIHscore
```

Krabicové grafy rozložení vstupního NIHSS ve skupinách ukazuje Obr. 7. Pro test hypotézy, zda je rozložení NIHSS ve skupinách stejné, použijeme KW test:

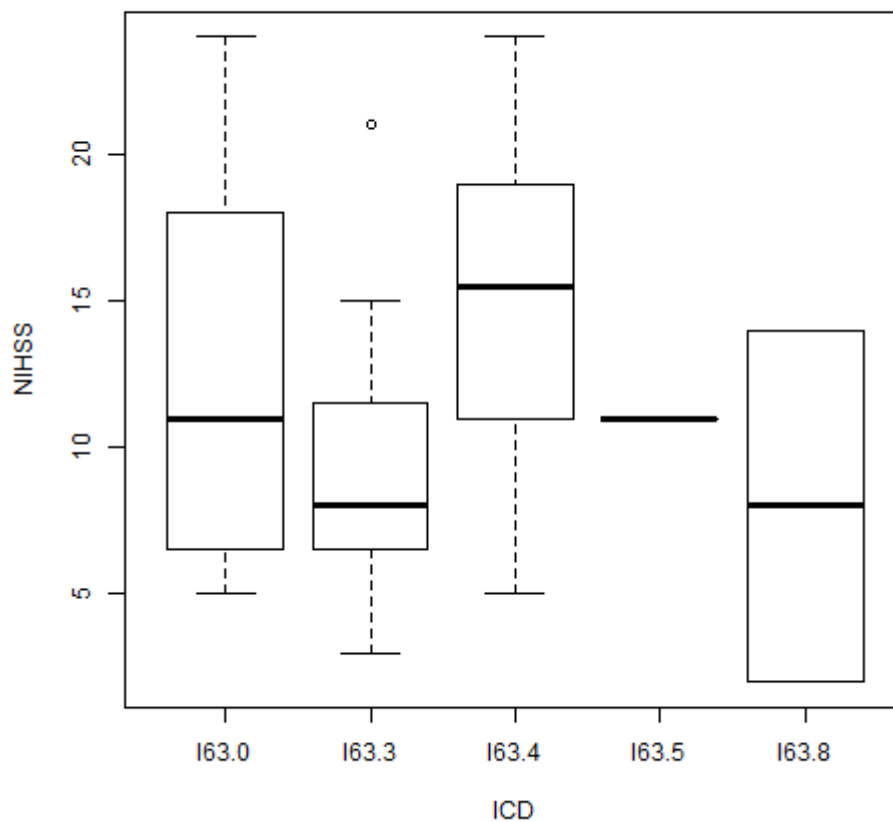
```
> kruskal.test(NIH, fICD)
```

```
    Kruskal-wallis rank sum test
```

```
data:  NIH and fICD
```

```
Kruskal-wallis chi-squared = 8.8263, df = 4, p-value = 0.06559
```

Na hladině významnosti 5% tedy nezamítáme hypotézu, že rozložení NIHSS je ve všech skupinách stejné.



Obr. 7: Graf rozložení NIHSS

### Trend NIHSS po 24 hodinách

V SITS modelu odpovídá rozdílu mezi vlastnostmi:

```
Baseline_NIHScore a After24H_NIHScore
```

Zkoumáme rozložení trendu, resp. rozdílu vstupního NIHSS a NIHSS po 24 hodinách. Krabicový graf na Obr. 8 zachycuje tato rozložení pro jednotlivé skupiny ICD.

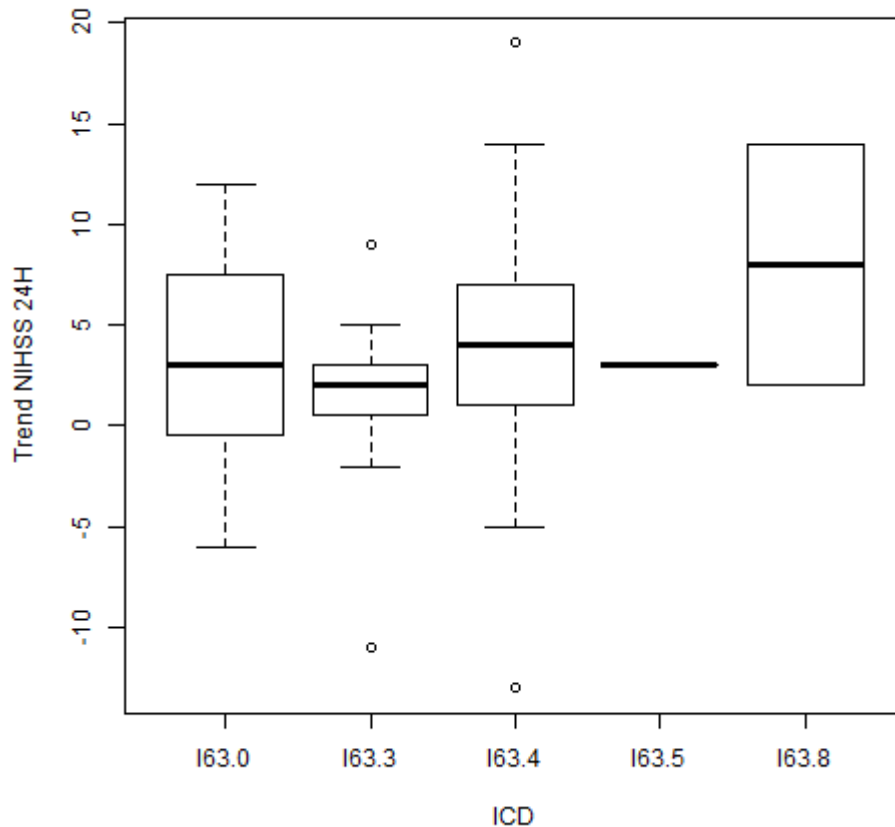
Testujeme nulovou hypotézu, zda rozložení rozdílů je ve všech skupinách stejné. K testování použijeme KW test:

```
> kruskal.test(trend, fICD)
      kruskal-wallis rank sum test
data: trend and fICD
```



Kruskal-wallis chi-squared = 2.9867, df = 4, p-value = 0.56

Na hladině významnosti 5% tedy nezamítáme hypotézu, že rozložení rozdílů NIHSS a NIHSS po 24 hodinách je ve všech skupinách stejné.



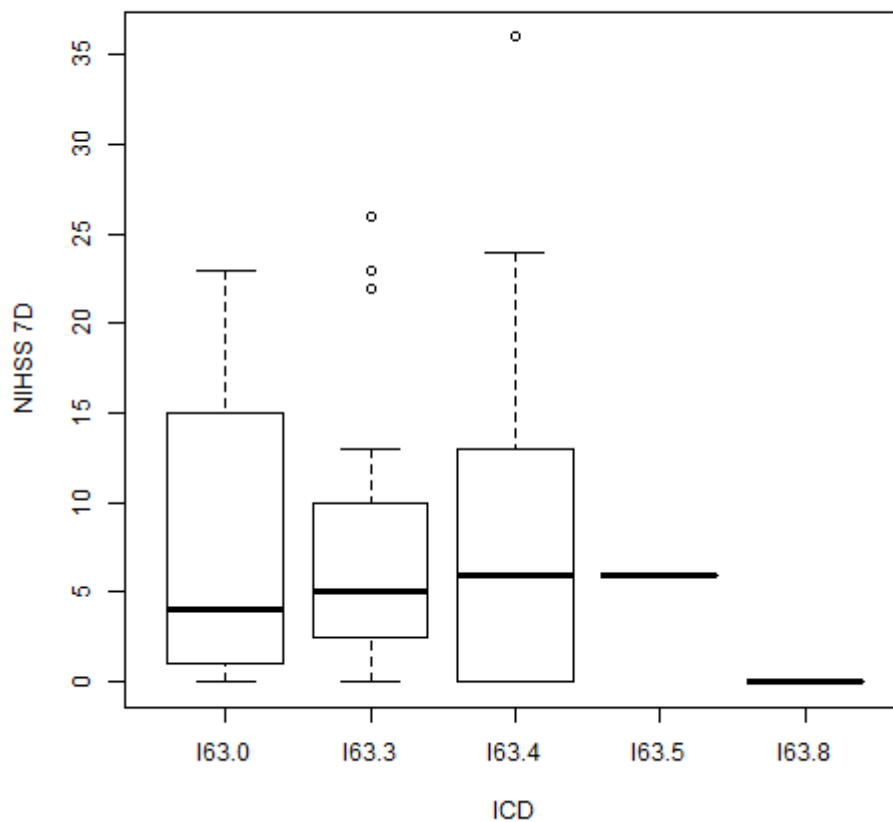
Obr. 8: Graf rozložení rozdílů vstupního NIHSS a NIHSS po 24h

### NIHSS po 7 dnech

V SITS modelu odpovídá vlastnosti:

*After7D\_\_NIHScore*

V této části zkoumáme hodnocení NIHSS po 7 dnech od začátku léčby. Opět nás zajímá, zda je rozložení NIHSS ve všech skupinách stejné. Na Obr. 9 jsou tyto rozložení NIHSS vykresleny pro jednotlivé skupiny ICD.



Obr. 9: Graf rozložení NIHSS po 7 dnech

Po provedení KW testu dostáváme tento výsledek:

```
> kruskal.test(AFTER7DNIHSS, fICD)

Kruskal-wallis rank sum test

data: AFTER7DNIHSS and fICD
Kruskal-wallis chi-squared = 4.1254, df = 4, p-value = 0.3893
```

Na hladině významnosti 5% nezamítáme hypotézu, že rozložení NIHSS po 7 dnech je ve všech skupinách stejné.

### Trend NIHSS po 7 dnech

V SITS modelu odpovídá rozdílu mezi vlastnostmi:

*Baseline\_NIHScore a After7D\_NIHScore*

Graf na Obr. 10 zachycuje rozložení rozdílů vstupního NIHSS a NIHSS po 7 dnech v jednotlivých skupinách ICD. Opět testujeme, zda je rozložení rozdílů ve všech skupinách stejné.

Stejně jako v předchozích případech, i zde použijeme KW test:

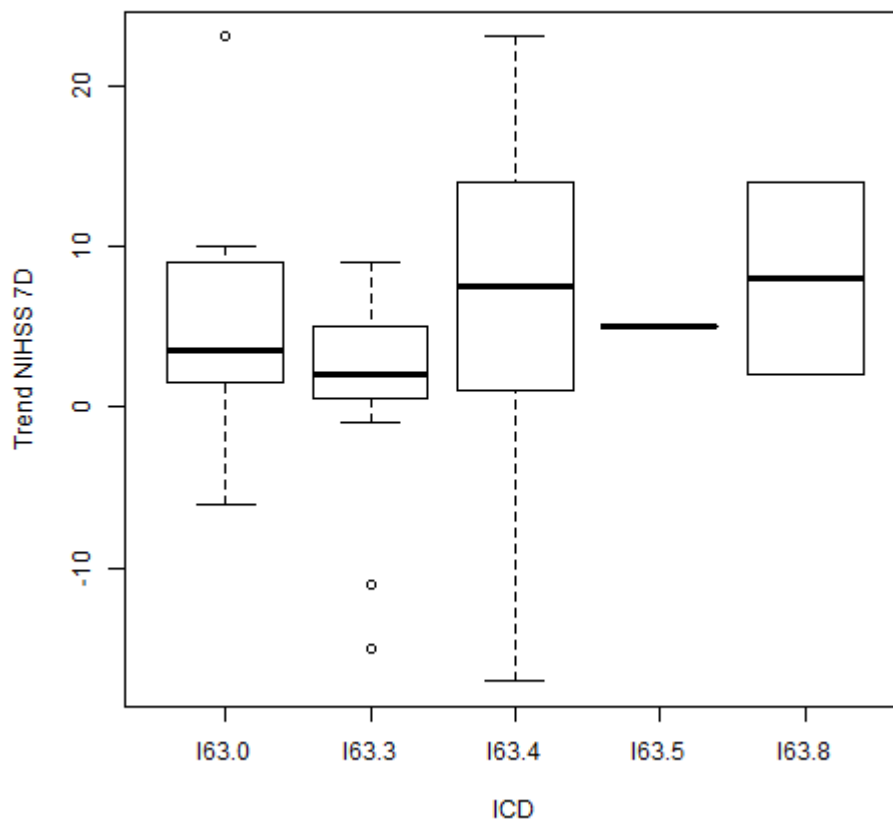
```
> kruskal.test(trend, fICD)
```

```
    Kruskal-wallis rank sum test
```

```
data: trend and fICD
```

```
Kruskal-wallis chi-squared = 4.7503, df = 4, p-value = 0.313
```

Na hladině významnosti 5% nezamítáme hypotézu, že rozložení rozdílů NIHSS a NIHSS po 7 dnech je ve všech skupinách stejné.



Obr. 10: Graf rozložení rozdílu vstupního NIHSS a NIHSS po 7 dnech

#### 4.2.4 Klinické změny

V SITS modelu odpovídá vlastnostem:

*After24H\_GlobalOutcome*  
*After7D\_GlobalOutcome*

U pacientů byla zjišťována klinická změna po 24 hodinách a po 7 dnech od zahájení léčby. Pozorované četnosti v jednotlivých skupinách jsou zobrazeny v Tab. 7 a Tab. 8. Chceme testovat hypotézu, zda četnosti v kategoriích klinické změny jsou proporcčně stejně rozděleny v jednotlivých skupinách ICD, resp. zda je rozdělení četností nezávislé na skupinách ICD.

| ICD   | Better | Dead | Much better | Much worse | Unchanged | Worse |
|-------|--------|------|-------------|------------|-----------|-------|
| I63.0 | 37     | 0    | 19          | 0          | 17        | 8     |
| I63.3 | 44     | 0    | 13          | 4          | 13        | 14    |
| I63.4 | 24     | 0    | 20          | 3          | 16        | 13    |
| I63.5 | 6      | 0    | 7           | 0          | 5         | 4     |
| I63.8 | 5      | 0    | 2           | 0          | 2         | 1     |
| I63.9 | 2      | 0    | 3           | 0          | 2         | 0     |

Tab. 7: Klinická změna po 24 hodinách – pozorované četnosti

| ICD   | Better | Dead | Much better | Much worse | Unchanged | Worse |
|-------|--------|------|-------------|------------|-----------|-------|
| I63.0 | 32     | 1    | 31          | 2          | 11        | 4     |
| I63.3 | 41     | 3    | 21          | 4          | 12        | 7     |
| I63.4 | 23     | 2    | 25          | 2          | 19        | 5     |
| I63.5 | 8      | 0    | 8           | 0          | 4         | 2     |
| I63.8 | 4      | 0    | 4           | 0          | 2         | 0     |
| I63.9 | 2      | 0    | 4           | 0          | 1         | 0     |

Tab. 8: Klinická změna po 7 dnech - pozorované četnosti

Pro testování použijeme  $\chi^2$  test v kontingenční tabulce. Vypočtené očekávané četnosti jsou zobrazeny v Tab. 9 a Tab. 10. V prvním případě z testování automaticky vyřazujeme kategorii „Dead“, která má ve všech skupinách ICD nulové četnosti. V obou případech vidíme porušení podmínek pro dobrou aproximaci, tedy že více než 20% z očekávaných četností je menší než 5 a některé z očekávaných četností jsou menší než 1.

| ICD   | Better | Much better | Much worse | Unchanged | Worse |
|-------|--------|-------------|------------|-----------|-------|
| I63.0 | 33,65  | 18,25       | 2,00       | 15,69     | 11,41 |
| I63.3 | 36,56  | 19,83       | 2,17       | 17,04     | 12,39 |
| I63.4 | 31,58  | 17,13       | 1,87       | 14,72     | 10,70 |
| I63.5 | 9,14   | 4,96        | 0,54       | 4,26      | 3,10  |
| I63.8 | 4,15   | 2,25        | 0,25       | 1,94      | 1,41  |
| I63.9 | 2,91   | 1,58        | 0,17       | 1,36      | 0,99  |

Tab. 9: Klinická změna po 24 hodinách - očekávané četnosti

| ICD   | Better | Dead | Much better | Much worse | Unchanged | Worse |
|-------|--------|------|-------------|------------|-----------|-------|
| I63.0 | 31,37  | 1,71 | 26,52       | 2,28       | 13,98     | 5,13  |
| I63.3 | 34,08  | 1,86 | 28,82       | 2,48       | 15,18     | 5,58  |
| I63.4 | 29,44  | 1,61 | 24,89       | 2,14       | 13,11     | 4,82  |
| I63.5 | 8,52   | 0,46 | 7,20        | 0,62       | 3,80      | 1,39  |
| I63.8 | 3,87   | 0,21 | 3,27        | 0,28       | 1,73      | 0,63  |
| I63.9 | 2,71   | 0,15 | 2,29        | 0,20       | 1,21      | 0,44  |

Tab. 10: Klinická změna po 7 dnech - očekávané četnosti

Po provedení  $\chi^2$  testů dostáváme opět varovné hlášení o nekorektnosti aproximace:

```
> chisq.test(ICD,AFTER24H)
```

```
Pearson's Chi-squared test
```

```

data: ICD and AFTER24H
X-squared = 20.0998, df = 20, p-value = 0.4517

Warning message:
In chisq.test(ICD, AFTER24H) : Chi-squared approximation may be
incorrect

*****

> chisq.test(fICD,AFTER7D)

      Pearson's Chi-squared test

data: fICD and AFTER7D
X-squared = 17.4303, df = 25, p-value = 0.8656

Warning message:
In chisq.test(fICD, AFTER7D) : Chi-squared approximation may be
incorrect

```

Provedeme-li přesnější randomizační test pro 1 000 000 náhodných vzorků, dostáváme podobné p-hodnoty:

```

> chisq.test(ICD,AFTER24H,simulate.p.value=T,B=1000000)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data: ICD and AFTER24H
X-squared = 20.0998, df = NA, p-value = 0.4444

*****

> chisq.test(fICD,AFTER7D,simulate.p.value=T,B=1000000)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data: fICD and AFTER7D
X-squared = 17.4303, df = NA, p-value = 0.8599

```

Na hladině významnosti 5% tedy nezamítáme hypotézy, že rozložení četností klinických změn po 24 hodinách a po 7 dnech je stejné pro všechny skupiny ICD.

#### 4.2.5 Rankin score

V SITS modelu odpovídá vlastnosti:

*After3M\_\_RankinScore*

Modifikované Rankin score (mRS<sup>16</sup>) je škála hodnocení stupně zdravotní neschopnosti či nesamostatnosti pacientů, kteří prodělali cévní mozkovou příhodu. Obsahuje hodnoty 0 – 6, od naprosto zdravého pacienta bez dalších následků, až po smrt pacienta. Pro představu uvádíme popis této škály (v originálním znění) v Tab. 11. [15]

| Score | Description   |
|-------|---|
| 0     | No symptoms at all  |
| 1     | No significant disability despite symptoms; able to carry out all usual duties and activities                               |
| 2     | Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance       |
| 3     | Moderate disability; requiring some help, but able to walk without assistance   |
| 4     | Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance |
| 5     | Severe disability; bedridden, incontinent and requiring constant nursing care and attention                                 |
| 6     | Dead  |

Tab. 11: Rankin score – popis škály

V této části analyzujeme hodnocení Rankin score po 3 měsících od začátku léčby. Škálu Rankin score dělíme dle požadavku lékařů do dvou kategorií:

- dobrý výsledek: hodnoty 0-2
- suboptimální výsledek: hodnoty 3-6.

Pozorované a očekávané četnosti v těchto kategoriích ukazuje Tab. 12.

<sup>16</sup> Modified Rankin Scale - <http://www.rankinscale.org/>

| ICD   | pozorované četnosti |           | očekávané četnosti |           |
|-------|---------------------|-----------|--------------------|-----------|
|       | score 0-2           | score 3-6 | score 0-2          | score 3-6 |
| I63.0 | 44                  | 31        | 41,13              | 33,87     |
| I63.3 | 43                  | 31        | 40,58              | 33,42     |
| I63.4 | 26                  | 38        | 35,10              | 28,90     |
| I63.5 | 14                  | 7         | 11,52              | 9,48      |
| I63.8 | 4                   | 3         | 3,84               | 3,16      |
| I63.9 | 5                   | 2         | 3,84               | 3,16      |

Tab. 12: Rankin score - pozorované a očekávané četnosti

Očekávané četnosti opět nesplňují podmínky pro dobrou aproximaci  $\chi^2$  testu v kontingenční tabulce. Spolu s  $\chi^2$  testem proto také provedeme randomizační test:

```
> chisq.test(fICD,rankinCategory)

Pearson's Chi-squared test

data: fICD and rankinCategory
X-squared = 7.9632, df = 5, p-value = 0.1583

warning message:
In chisq.test(fICD, rankinCategory) :
Chi-squared approximation may be incorrect

*****

> chisq.test(fICD,rankinCategory,simulate.p.value=T,B=1000000)

Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data: fICD and rankinCategory
X-squared = 7.9632, df = NA, p-value = 0.1574
```

Výsledné p-hodnoty jsou si velice blízké. Na jejich základě nezamítáme nulovou hypotézu (na hladině významnosti 5%), že je rozložení četností v jednotlivých skupinách ICD stejné.



#### 4.2.6 Průkaz zavřené tepny

V SITS modelu odpovídá složené podmínce:

*Baseline\_\_CTAOcclusion == 'Yes'*  
nebo (*Baseline\_\_CTAOcclusion == 'Not done'* a současně  
*Baseline\_\_CTDenseArterySign == 'Yes'*)

U pacientů byl zjišťován průkaz zavřené tepny na vstupním CT. Chceme testovat hypotézu, zda je poměr pacientů s tímto příznakem stejný ve všech skupinách ICD. Pacienty rozdělíme do dvou kategorií: se zjištěným průkazem zavřené tepny a ostatní. Pozorované četnosti v jednotlivých kategoriích podle skupin ICD ukazuje Tab. 13.

| ICD   | Ostatní | Průkaz zavřené tepny |
|-------|---------|----------------------|
| I63.0 | 29      | 52                   |
| I63.3 | 42      | 46                   |
| I63.4 | 24      | 52                   |
| I63.5 | 22      | 0                    |
| I63.8 | 7       | 3                    |
| I63.9 | 4       | 3                    |

Tab. 13: Průkaz zavřené tepny na CT - pozorované četnosti

Vypočtené očekávané četnosti můžeme vidět v Tab. 14. Hypotézu otestujeme pomocí  $\chi^2$  testu v kontingenční tabulce. Ovšem očekávané četnosti nesplňují podmínky pro dobrou aproximaci (více než 20% očekávaných četností je menších než 5), proto zkusíme také provést randomizační test.

| ICD   | Ostatní | Průkaz zavřené tepny |
|-------|---------|----------------------|
| I63.0 | 36,51   | 44,49                |
| I63.3 | 39,66   | 48,34                |
| I63.4 | 34,25   | 41,75                |

|       |      |       |
|-------|------|-------|
| I63.5 | 9,92 | 12,08 |
| I63.8 | 4,51 | 5,49  |
| I63.9 | 3,15 | 3,85  |

Tab. 14: Průkaz zavřené tepny na CT - očekávané četnosti

Výstup testů z R:

```
> chisq.test(ICD,CTAOCC)

      Pearson's Chi-squared test

data:  ICD and CTAOCC
X-squared = 38.3839, df = 5, p-value = 3.159e-07

Warning message:
In chisq.test(ICD, CTAOCC) :
  Chi-squared approximation may be incorrect

*****

> chisq.test(ICD, CTAOCC,simulate.p.value=T,B=1000000)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data:  ICD and CTAOCC
X-squared = 38.3839, df = NA, p-value = 1e-06
```

Na hladině významnosti 5% tedy zamítáme hypotézu, že by poměr pacientů s průkazem zavřené tepny na vstupním CT byl ve všech skupinách stejný.

#### 4.2.7 Průkaz rekanalizace

V SITS modelu odpovídá složené podmínce:

```
After24H_CTAOcclusion == 'Yes'
nebo (Baseline_CTAOcclusion == 'Not Done' a současně
Baseline_CTDenseArterySign == 'Yes')
```

U pacientů byl také zjišťován průkaz rekanalizace. Zajímá nás, zda je poměr pacientů s tímto příznakem stejný ve všech skupinách ICD. Pacienty rozdělíme do dvou skupin:

se zjištěným průkazem rekanalizace a ostatní. Tab. 15 ukazuje pozorované četnosti v jednotlivých kategoriích.

| ICD   | Ostatní | Průkaz rekanalizace |
|-------|---------|---------------------|
| I63.0 | 47      | 34                  |
| I63.3 | 69      | 19                  |
| I63.4 | 41      | 35                  |
| I63.5 | 15      | 7                   |
| I63.8 | 6       | 4                   |
| I63.9 | 6       | 1                   |

Tab. 15: Průkaz rekanalizace - pozorované četnosti

Očekávané četnosti dostáváme v Tab. 16. Ani zde očekávané četnosti nesplňují podmínky pro dobrou aproximaci  $\chi^2$ -rozdělením v testu v kontingenční tabulce.

| ICD   | Ostatní | Průkaz rekanalizace |
|-------|---------|---------------------|
| I63.0 | 52,48   | 28,52               |
| I63.3 | 57,01   | 30,99               |
| I63.4 | 49,24   | 26,76               |
| I63.5 | 14,25   | 7,75                |
| I63.8 | 6,48    | 3,52                |
| I63.9 | 4,54    | 2,46                |

Tab. 16: Průkaz rekanalizace - očekávané četnosti

Spolu s  $\chi^2$  testem v kontingenční tabulce provedeme také randomizační test:

```
> chisq.test(ICD,CTAOCCAFTER24)

Pearson's Chi-squared test

data:  ICD and CTAOCCAFTER24
X-squared = 14.2514, df = 5, p-value = 0.01409
```

```
Warning message:
In chisq.test(ICD, CTAOCCAFTER24) :
  Chi-squared approximation may be incorrect

*****

> chisq.test(ICD, CTAOCCAFTER24, simulate.p.value=T, B=1000000)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data:  ICD and CTAOCCAFTER24
X-squared = 14.2514, df = NA, p-value = 0.01244
```

Zamítáme tedy hypotézu (na hladině významnosti 5%), že je poměr pacientů s průkazem rekanalizace stejný ve všech skupinách ICD.

### 4.3 Analýza pacientů s krvácením

V SITS modelu odpovídá vlastnosti:

```
pacienti s krvácením:
  After24H_CTLocalHaemorrhage == {'HI1', 'HI2', 'PH1', 'PH2'} nebo
  After24H_CTRemoteHaemorrhage == {'PHr1', 'PHr2'}
```

V této kapitole budeme analyzovat pacienty s krvácením po trombolýze. Pacienty s krvácením rozdělujeme do dvou skupin dle symptomatičnosti. Ta je dána rozdílem hodnocení NIHSS po 24 hodinách a vstupního NIHSS. Skupiny jsou následující:

- symptomatické krvácení (symptomatičnost > 3)
- asymptomatické krvácení (symptomatičnost ≤ 3)

Tyto skupiny budeme dále porovnávat podle vybraných parametrů (dle požadavků lékařů). V analyzovaném vzorku dat je 42 pacientů s krvácením. Počty pacientů, u kterých bylo zároveň hodnoceno NIHSS a které je tak možné rozdělit do skupin, zobrazuje Tab. 17. Protože zde máme velmi malý vzorek dat, nemá zde ve většině případů zřejmě smysl testovat hypotézy o nezávislosti. V takových případech uvedeme

jen četnosti v jednotlivých kategoriích bez dalšího testování. Nicméně předpokládáme budoucí nárůst těchto dat, proto jsou testy ve statistické aplikaci připraveny.

| symptomatické krvácení | asymptomatické krvácení |
|------------------------|-------------------------|
| 2                      | 2                       |

Tab. 17: Četnost pacientů s krvácením a hodnocením NIHSS

#### 4.3.1 Diabetes mellitus

V SITS modelu odpovídá vlastnosti:

*Baseline\_DiabetesDiagnosisOfDiabetes*

U pacientů byla zjišťována diagnóza diabetes. Chceme testovat hypotézu, zda rozložení četností pacientů s touto diagnózou jsou stejné pro obě skupiny pacientů s krvácením, resp. zda diagnóza diabetes je závislá na skupině pacientů s krvácením.

Četnosti těchto pacientů zobrazuje Tab. 18. V analyzovaném souboru dat se nevyskytli žádní pacienti s diagnózou diabetes, u kterých by zároveň bylo hodnoceno NIHSS a bylo je tak možné zařadit do některé ze skupin dle symptomatickosti. Sloupec s nulovými četnostmi vyřazujeme, proto zde nyní nemůžeme provést  $\chi^2$  test v kontingenční tabulce.

| Diabetes            | Ne | Ano |
|---------------------|----|-----|
| asymptomatické krv. | 2  | 0   |
| symptomatické krv.  | 2  | 0   |

Tab. 18: Četnost pacientů s diagnózou diabetes

#### 4.3.2 Arteriální hypertenze

V SITS modelu odpovídá vlastnosti:

*Baseline\_HypertensionDiagnosisOfHypertension*

U pacientů byla také zjišťována diagnóza arteriální hypertenze (vysoký krevní tlak). Četnosti pacientů s touto diagnózou ve skupinách podle symptomatickosti ukazuje Tab. 19. U všech pacientů, u kterých bylo hodnoceno NIHSS, byla také zjištěna diagnóza art. hypertenze, proto i zde dostáváme sloupec s nulovými četnostmi. Ani zde tedy nebudeme provádět  $\chi^2$  test v kontingenční tabulce.

| Art. hypertenze     | Ne | Ano |
|---------------------|----|-----|
| asymptomatické krv. | 0  | 2   |
| symptomatické krv.  | 0  | 2   |

Tab. 19: Četnost pacientů s diagnózou art. hypertenze

### 4.3.3 Nikotinismus

V SITS modelu odpovídá vlastnosti:

*Baseline\_\_CurrentSmokerCurrentSmokerAtStrokeOnset == 'Yes'* nebo  
*Baseline\_\_PreviousSmokerPreviousSmokerButStoppedBeforeStrokeOnset == 'Yes'*

V této části nás bude zajímat, zda byl nebo je pacient kuřák. Jak vidíme z Tab. 20, všichni pacienti, které v této části analyzujeme, nejsou kuřáci. Dostáváme opět sloupec s nulovými četnostmi, nebudeme tedy ani zde provádět  $\chi^2$  test.

| Kuřák               | Ne | Ano |
|---------------------|----|-----|
| asymptomatické krv. | 2  | 0   |
| symptomatické krv.  | 2  | 0   |

Tab. 20: Četnost pacientů kuřáků

### 4.3.4 Antiagregační léčba

V SITS modelu odpovídá vlastnostem:

*Baseline\_\_AspirinLessThan75MgOverDay*

*Baseline\_\_Aspirin75MgOverDay*  
*Baseline\_\_AspirinGreaterThan75MgLessThen200Mg*  
*Baseline\_\_AspirinGreaterThan200Mg*  
*Baseline\_\_DipyridamoleSlowRelease*  
*Baseline\_\_Clopidogrel*  
*Baseline\_\_OtherAntiplatelet*

Jedná se o léčbu snižující krevní srážlivost. Používá se jako prevence trombózy či infarktu myokardu<sup>17</sup>. Jak vidíme z Tab. 21, nemáme v tomto případě v analyzovaném souboru dat žádné pacienty, u kterých by byla nasazena antiagregační léčba.

| Antiagregační léčba | asymptomatické krv. | symptomatické krv. |
|---------------------|---------------------|--------------------|
| Aspirin <75 Mg      | 0                   | 0                  |
| Aspirin 75 Mg       | 0                   | 0                  |
| Aspirin 75-200 Mg   | 0                   | 0                  |
| Aspirin >200 Mg     | 0                   | 0                  |
| Dipyridamole        | 0                   | 0                  |
| Clopidogrel         | 0                   | 0                  |
| Jiné antiplatelet   | 0                   | 0                  |

Tab. 21: Četnost pacientů s nasazenou antiagregační léčbou

#### 4.3.5 Čas do zahájení léčby

V SITS modelu odpovídá rozdílu mezi vlastnostmi:

*Treatment\_\_DateAndTimeForActivateTreatment* a  
*Baseline\_\_DateAndTimeStrokeOnset*

Zajímá nás čas do zahájení léčby, tzn. časové rozdíly mezi časem mozkové příhody a časem zahájení léčby. V tomto případě nás samozřejmě zajímají časy v rámci pacientů s krvácením, které v této kapitole analyzujeme. Průměrné časové rozdíly zobrazuje Tab.

<sup>17</sup> <http://lekarske.slovníky.cz/pojem/antiagregacni-lecba>

22. Krabicové grafy rozložení těchto časových rozdílů v jednotlivých skupinách pak ukazuje Obr. 11.

| Krvácení            | Průměrný čas do zahájení léčby [min] |
|---------------------|--------------------------------------|
| asymptomatické krv. | 155,0                                |
| symptomatické krv.  | 177,5                                |

Tab. 22: Průměrný čas do zahájení léčby ve skupinách dle symptomatičnosti

Hypotézu, zda je rozložení časových rozdílů ve všech skupinách stejné, otestujeme KW testem:

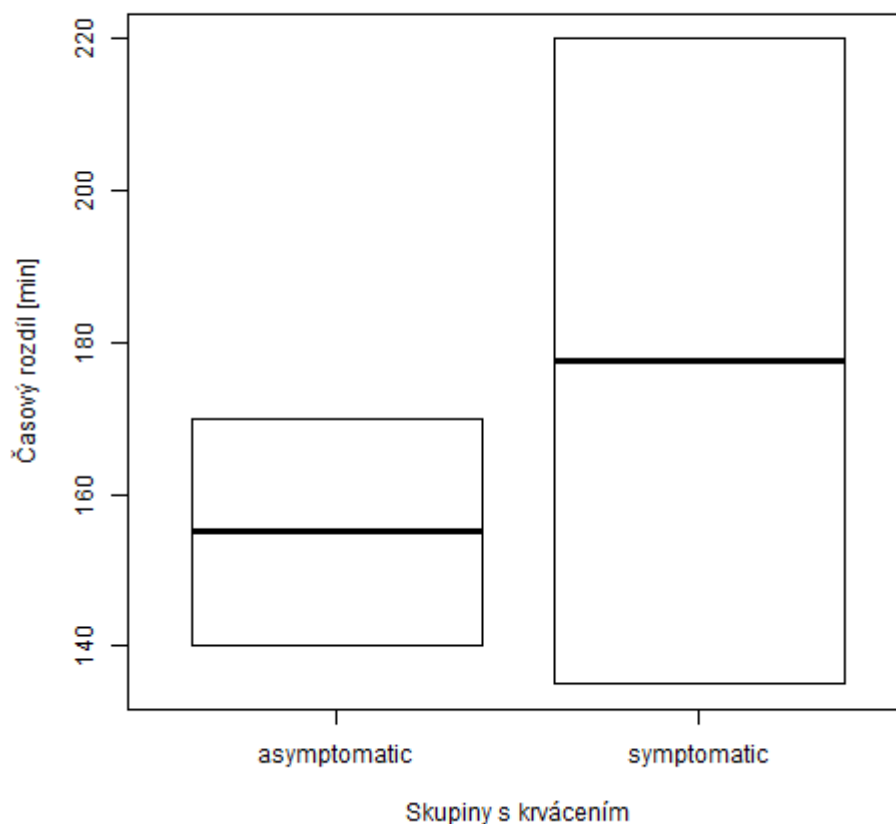
```
> kruskal.test(diff_time_min, fBleed)

Kruskal-wallis rank sum test

data: diff_time_min and fBleed
kruskal-wallis chi-squared = 0, df = 1, p-value = 1
```

Jak bylo uvedeno v kap. 3.2.1, pro dobrou aproximaci  $\chi^2$  – rozdělením je doporučováno mít alespoň 4 skupiny po 5 hodnotách. V tomto případě tyto podmínky splněny nejsou, proto výsledek KW testu nemusí být přesný. Nicméně zde na 5% hladině významnosti nezamítáme nulovou hypotézu, že je rozložení rozdílů časů ve všech skupinách stejné.





Obr. 11: Graf rozložení časů do zahájení léčby

### 4.3.6 Znamka ischemie na CT

V SITS modelu odpovídá vlastnosti:

*Baseline\_\_CTCurrentInfarct == 'Yes'*

Ischémie je obecně nedokrevnost určitého orgánu či tkáně. Znamky ischemie je možné rozeznat z CT snímku. Pozorované četnosti ukazuje Tab. 23.

| Krvácení            | Ne | Ano |
|---------------------|----|-----|
| asymptomatické krv. | 2  | 0   |
| symptomatické krv.  | 2  | 0   |

Tab. 23: Znamka ischemie na CT – pozorované četnosti

### 4.3.7 Klinická změna

V SITS modelu odpovídá vlastnosti:

*After7D\_GlobalOutcome*

V tomto případě nás zajímá klinická změna po 7 dnech od zahájení léčby v rámci skupin dle symptomatickosti. Pozorované četnosti v jednotlivých kategoriích ukazuje Tab. 24.

| Krvácení            | Better | Dead | Much better | Much worse | Unchanged | Worse |
|---------------------|--------|------|-------------|------------|-----------|-------|
| asymptomatické krv. | 0      | 0    | 0           | 0          | 2         | 0     |
| symptomatické krv.  | 0      | 0    | 0           | 1          | 1         | 0     |

Tab. 24: Klinická změna po 7 dnech – pozorované četnosti

### 4.3.8 Spektrum ICD

V SITS modelu odpovídá vlastnosti:

*After7D\_ICDCode*

Zajímá nás rozložení četností pacientů ve skupinách ICD. Tyto četnosti jsou zobrazeny v Tab. 25.

| Krvácení            | I63.0 | I63.3 | I63.4 | I63.5 | I63.8 | I63.9 |
|---------------------|-------|-------|-------|-------|-------|-------|
| asymptomatické krv. | 0     | 0     | 2     | 0     | 0     | 0     |
| symptomatické krv.  | 0     | 1     | 1     | 0     | 0     | 0     |

Tab. 25: Pozorované četnost pacientů ve skupinách ICD

### 4.3.9 Rankin score

V SITS modelu odpovídá vlastnosti:

*After3M\_\_RankinScore*

Rankin score dělíme podle kap. 4.2.5 do dvou kategorií podle hodnocení: dobrý a suboptimální výsledek. Pozorované četnosti v těchto kategoriích ukazuje Tab. 26

| Krvácení            | score 0-2 | score 3-6 |
|---------------------|-----------|-----------|
| asymptomatické krv. | 0         | 2         |
| symptomatické krv.  | 0         | 2         |

Tab. 26: Rankin score - pozorované četnosti

#### 4.4 Analýza pacientů s kompletním CT vyšetřením

V SITS modelu odpovídá vlastnostem:

kompletní CT vyšetření a rekanalizace

*Baseline\_\_CTDone* == 'Yes' a současně

*Baseline\_\_CTAOcclusion* == 'Yes' a současně

*Baseline\_\_CTPerfusionDeficit* == 'Yes' a současně

*After24H\_\_CTDone* == 'Yes' a současně

*After24H\_\_CTAOcclusion* == 'No'

Kompletní CT vyšetření a perzistující uzávěr

*Baseline\_\_CTDone* == 'Yes' a současně

*Baseline\_\_CTAOcclusion* == 'Yes' a současně

*Baseline\_\_CTPerfusionDeficit* == 'Yes' a současně

*After24H\_\_CTDone* == 'Yes' a současně

*After24H\_\_CTAOcclusion* == 'Yes'

V této části analyzujeme pacienty s kompletním CT vyšetřením, u kterých byl prokázán CTA<sup>18</sup> uzávěr mozkové tepny na vstupním CT. Tyto pacienty rozdělujeme do dvou skupin:

- rekanalizovaní
- s perzistujícím uzávěrem (nerekanalizovaní)

<sup>18</sup> Computed Tomography Angiography

Četnosti pacientů v těchto skupinách ukazuje Tab. 27. Tyto skupiny budeme v dalších částech porovnávat podle vybraných parametrů.

| CTA uzávěr        | rekanalizovaní | nerekanalizovaní |
|-------------------|----------------|------------------|
| Absolutní četnost | 56             | 44               |
| Relativní četnost | 56,0%          | 44,0%            |

Tab. 27: Četnost pacientů ve skupinách s CT vyšetřením

#### 4.4.1 Spektrum ICD

V SITS modelu odpovídá vlastnosti:

*After7D\_\_ICDCode*

Analyzujeme rozložení pacientů ve skupinách ICD. Chceme testovat hypotézu, zda je toto rozložení stejné pro obě skupiny pacientů. Pozorované četnosti pacientů v těchto kategoriích ukazuje Tab. 28. Očekávané četnosti pak Tab. 29.

| CTA uzávěr       | I63.0 | I63.3 | I63.4 | I63.5 | I63.8 | I63.9 |
|------------------|-------|-------|-------|-------|-------|-------|
| rekanalizovaní   | 20    | 9     | 26    | 0     | 1     | 0     |
| nerekanalizovaní | 15    | 14    | 13    | 0     | 1     | 1     |

Tab. 28: Spektrum ICD - pozorované četnosti

| CTA uzávěr       | I63.0 | I63.3 | I63.4 | I63.8 | I63.9 |
|------------------|-------|-------|-------|-------|-------|
| rekanalizovaní   | 19,6  | 12,88 | 21,84 | 1,12  | 0,56  |
| nerekanalizovaní | 15,4  | 10,12 | 17,16 | 0,88  | 0,44  |

Tab. 29: Spektrum ICD - očekávané četnosti

Očekávané četnosti nesplňují podmínky dobré aproximace, proto zde spolu s  $\chi^2$  testem v kontingenčních tabulkách provedeme také randomizační test:

```
> chisq.test(ctcomplet$CTA,ctcomplet$ICD)

      Pearson's Chi-squared test

data:  ctcomplet$CTA and ctcomplet$ICD
X-squared = 5.7778, df = 4, p-value = 0.2164

Warning message:
In chisq.test(ctcomplet$CTA, ctcomplet$ICD) :
  Chi-squared approximation may be incorrect

*****

> chisq.test(ctcomplet$CTA,ctcomplet$ICD,simulate.p.value=T,B=1000000)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data:  ctcomplet$CTA and ctcomplet$ICD
X-squared = 5.7778, df = NA, p-value = 0.1807
```

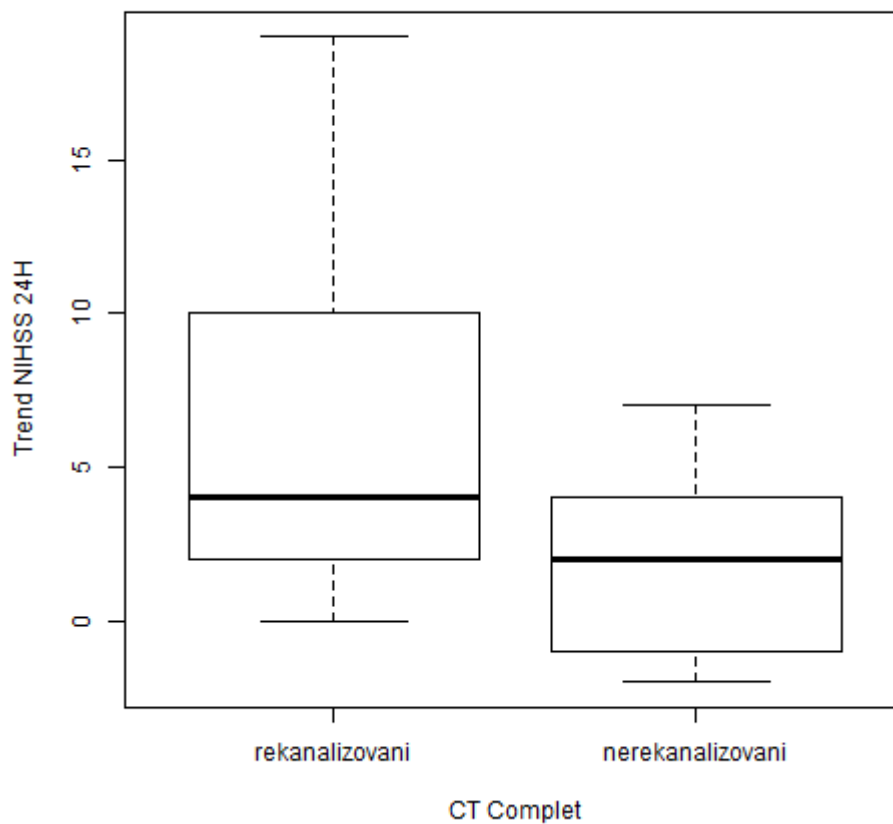
Na hladině významnosti 5% tedy nezamítáme hypotézu, že je rozložení četností v obou skupinách stejné.

#### 4.4.2 NIHSS

V SITS modelu odpovídá rozdílu mezi vlastnostmi:

```
Baseline__NIHScore a After24H__NIHScore
```

Z hodnocení NIHSS nás zajímá trend po 24 hodinách, resp. rozdíl vstupního NIHSS a NIHSS po 24 hodinách. Chceme testovat hypotézu, zda je rozložení těchto rozdílů pro obě skupiny stejné. Rozložení rozdílů zachycují krabicové grafy na Obr. 12.



Obr. 12: Graf rozložení rozdílů vstupního NIHSS a NIHSS po 24h

Naší hypotézu otestujeme KW testem:

```
> kruskal.test(trend,ctcomplet$CTA)

Kruskal-wallis rank sum test

data: trend and ctcomplet$CTA
Kruskal-wallis chi-squared = 4.1136, df = 1, p-value = 0.04254
```

Zamítáme tedy hypotézu (na hladině významnosti 5%), že by rozložení rozdílů vstupního NIHSS a NIHSS po 24 hodinách bylo stejné v obou skupinách.

#### 4.4.3 Klinická změna

V SITS modelu odpovídá vlastnostem:

```
After24H_GlobalOutcome
After7D_GlobalOutcome
```

Chceme testovat hypotézu, zda rozložení četností klinických změn je v obou skupinách stejné. Zajímají nás klinické změny po 24 hodinách a po 7 dnech od začátku léčby. Pozorované četnosti v těchto kategoriích ukazují Tab. 30 a Tab. 31.

| CTA uzávěr       | Better | Dead | Much better | Much worse | Unchanged | Worse |
|------------------|--------|------|-------------|------------|-----------|-------|
| rekanalizovaní   | 27     | 0    | 19          | 0          | 7         | 3     |
| nerekanalizovaní | 18     | 0    | 6           | 0          | 12        | 8     |

Tab. 30: Klinická změna po 24 hodinách - pozorované četnosti

| CTA uzávěr       | Better | Dead | Much better | Much worse | Unchanged | Worse |
|------------------|--------|------|-------------|------------|-----------|-------|
| rekanalizovaní   | 18     | 1    | 32          | 1          | 3         | 1     |
| nerekanalizovaní | 18     | 1    | 7           | 0          | 15        | 3     |

Tab. 31: Klinická změna po 7 dnech - pozorované četnosti

Očekávané četnosti jsou vypočteny v Tab. 32 a Tab. 33. Vyřazujeme sloupce s nulovými četnostmi, tedy v případě kategorií klinické změny po 24 hodinách sloupce „Dead“ a „Much worse“. Hypotézu budeme testovat  $\chi^2$  testem v kontingenčních tabulkách. Protože očekávané četnosti nesplňují podmínky dobré aproximace, provedeme také randomizační test.

| CTA uzávěr       | Better | Much better | Unchanged | Worse |
|------------------|--------|-------------|-----------|-------|
| rekanalizovaní   | 25,2   | 14          | 10,64     | 6,16  |
| nerekanalizovaní | 19,8   | 11          | 8,36      | 4,84  |

Tab. 32: Klinická změna po 24 hodinách - očekávané četnosti

Výsledky testů pro klinickou změnu po 24 hodinách:

```
> chisq.test(ctcomplet$CTA, ctcomplet$AFTER24H)

Pearson's Chi-squared test

data: ctcomplet$CTA and ctcomplet$AFTER24H
```

```
X-squared = 10.865, df = 3, p-value = 0.01248

Warning message:
In chisq.test(ctcomplet$CTA, ctcomplet$AFTER24H) :
  Chi-squared approximation may be incorrect

*****

>
chisq.test(ctcomplet$CTA,ctcomplet$AFTER24H,simulate.p.value=T,B=100000)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data:  ctcomplet$CTA and ctcomplet$AFTER24H
X-squared = 10.865, df = NA, p-value = 0.01157
```

Na základě těchto p-hodnot zamítáme hypotézu (na hladině významnosti 5%), že by rozložení četností klinické změny po 24 hodinách bylo stejné ve skupinách s prokázaným CTA uzávěrem.

| CTA uzávěr       | Better | Dead | Much better | Much worse | Unchanged | Worse |
|------------------|--------|------|-------------|------------|-----------|-------|
| rekanalizovaní   | 20,16  | 1,12 | 21,84       | 0,56       | 10,8      | 2,24  |
| nerekanalizovaní | 15,84  | 0,88 | 17,16       | 0,44       | 7,92      | 1,76  |

Tab. 33: Klinická změna po 7 dnech - očekávané četnosti

Výsledky testů pro klinickou změnu po 7 dnech:

```
> chisq.test(ctcomplet$CTA,ctcomplet$AFTER7D)

      Pearson's Chi-squared test

data:  ctcomplet$CTA and ctcomplet$AFTER7D
X-squared = 24.9448, df = 5, p-value = 0.0001428

Warning message:
In chisq.test(ctcomplet$CTA, ctcomplet$AFTER7D) :
  Chi-squared approximation may be incorrect

*****
```



```
>
chisq.test(ctcomplet$CTA,ctcomplet$AFTER7D,simulate.p.value=T,B=100000
0)

      Pearson's Chi-squared test with simulated p-value (based on
1e+06 replicates)

data:  ctcomplet$CTA and ctcomplet$AFTER7D
X-squared = 24.9448, df = NA, p-value = 8e-06
```

Také v tomto případě zamítáme hypotézu (na 5% hladině významnosti), že je rozložení četností klinické změny po 7 dnech stejné ve skupinách s prokázaným CTA uzávěrem.

#### 4.4.4 Rankin score

V SITS modelu odpovídá vlastnosti:

*After3M\_\_RankinScore*

Jak bylo uvedeno v kap. 4.2.5, Rankin score dělíme do dvou kategorií podle hodnocení: na dobrý a suboptimální výsledek. Četnosti v těchto kategoriích jsou zobrazeny v Tab. 34. Chceme testovat hypotézu, zda rozložení těchto četností je stejné v obou skupinách pacientů s prokázaným uzávěrem na CTA.

| CTA uzávěr       | score 0-2 | score 3-6 |
|------------------|-----------|-----------|
| rekanalizovaní   | 34        | 13        |
| nerekanalizovaní | 15        | 25        |

Tab. 34: Rankin score - pozorované četnosti

Hypotézu budeme testovat  $\chi^2$  testem v kontingenční tabulce, v tomto případě v kontingenční tabulce 2x2. Očekávané četnosti jsou vypočteny v Tab. 35.

| CTA uzávěr       | score 0-2 | score 3-6 |
|------------------|-----------|-----------|
| rekanalizovaní   | 26,47     | 20,53     |
| nerekanalizovaní | 22,53     | 17,47     |

Tab. 35: Rankin score - očekávané četnosti

Provedením testu v R dostáváme tento výstup:

```
> chisq.test(ctcomplet$CTA,rankinCategory,correct=F)

Pearson's Chi-squared test

data: ctcomplet$CTA and rankinCategory
X-squared = 10.6626, df = 1, p-value = 0.001093
```

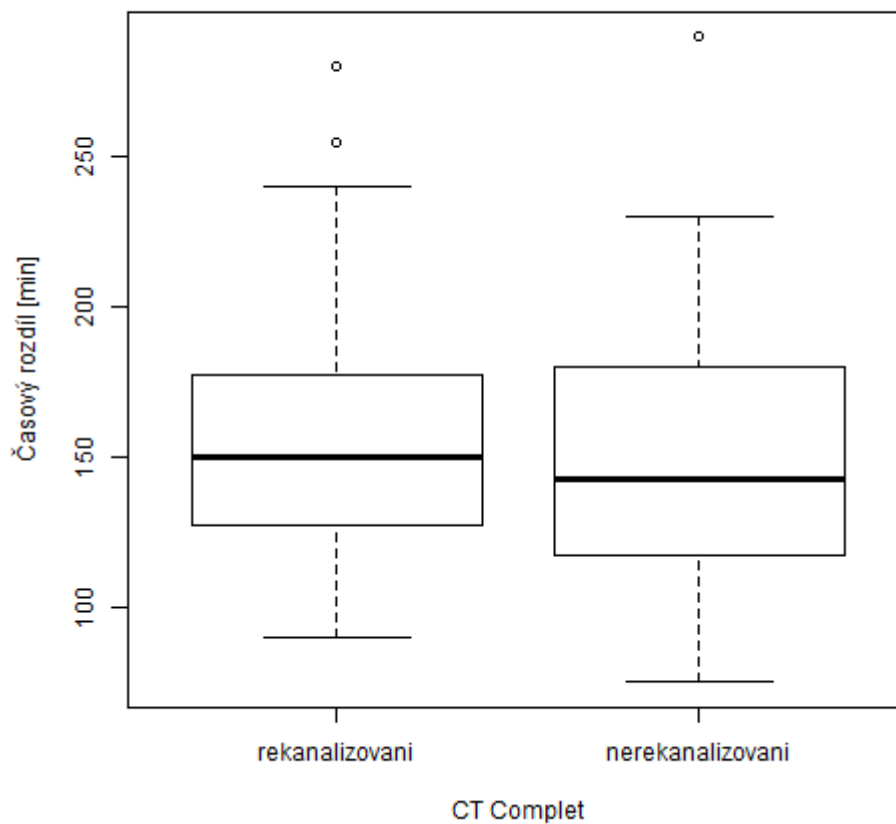
Zamítáme tedy hypotézu (na 5% hladině významnosti), že je rozložení četností stejné v obou skupinách. Můžeme si všimnout, že ve skupině pacientů s rekanalizací převažují pacienti s dobrým výsledkem hodnocení Rankin score, zatímco ve skupině pacientů s perzistujícím uzávěrem převažují pacienti se suboptimálním výsledkem.

#### 4.4.5 Čas do zahájení léčby

V SITS modelu odpovídá rozdílu mezi vlastnostmi:

```
Treatment__DateAndTimeForActilyseTreatment a  
Baseline__DateAndTimeStrokeOnset
```

V této části budeme analyzovat čas do zahájení léčby, resp. jaká doba uběhla od začátku mozkové příhody do zahájení léčby. Rozložení těchto časů je znázorněno v krabicových grafech na Obr. 13. Zajímá nás hypotéza, zda je toto rozložení časů stejné v obou skupinách pacientů, u kterých byl prokázán uzávěr na CTA.



Obr. 13: Graf rozložení časů do zahájení léčby

Hypotézu testujeme KW testem:

```
> kruskal.test(diff_time_min,ctcomplet$CTAOCCAFTER24)

Kruskal-wallis rank sum test

data: diff_time_min and ctcomplet$CTAOCCAFTER24
Kruskal-wallis chi-squared = 0.5276, df = 1, p-value = 0.4676
```

Nezamítáme tedy hypotézu (na hladině významnosti 5%), že by rozložení časů do zahájení léčby bylo stejné pro obě skupiny.

#### 4.4.6 Znamka ischemie na CT

V SITS modelu odpovídá vlastnostem:

```
Baseline__CTCurrentInfarct == 'Yes'
After24H__CTCurrentInfarct == 'Yes'
```

V této části nás bude zajímat známka ischemie na vstupním a na kontrolním CT snímku. Kontrolní snímek je pořizován po 24 hodinách. V Tab. 36 jsou zobrazeny pozorované a očekávané četnosti pacientů, u kterých byla zjišťována známka ischemie na vstupním CT. Zajímá nás, zda jsou poměry těchto pacientů stejné pro obě skupiny pacientů s prokázaným uzávěrem na CTA.

| Krvácení         | pozorované četnosti |     | očekávané četnosti |     |
|------------------|---------------------|-----|--------------------|-----|
|                  | Ne                  | Ano | Ne                 | Ano |
| rekanalizovaní   | 37                  | 9   | 36,8               | 9,2 |
| nerekanalizovaní | 31                  | 8   | 31,2               | 7,8 |

Tab. 36: Znamka ischemie na vstupním CT

Očekávané četnosti splňují podmínky pro dobrou aproximaci, proto by  $\chi^2$  test v kontingenční tabulce měl být korektní. Není zde proto nutné provádět i randomizační test. Provedením  $\chi^2$  testu dostáváme tento výstup:

```
> chisq.test(ctcomplet$CTAOCCAFTER24,ctcomplet$INFARCT,correct=F)

Pearson's Chi-squared test

data: ctcomplet$CTAOCCAFTER24 and ctcomplet$INFARCT
X-squared = 0.0118, df = 1, p-value = 0.9133
```

Nezamítáme tedy hypotézu (na hladině významnosti 5%), že jsou poměry zjištěných a nezjištěných známek ischemie na vstupním CT stejné v obou skupinách, resp. že zjištěná známka ischemie nezávisí na tom, do které skupiny pacient patří.

V Tab. 37 máme četnosti druhého případu, tedy zjištěné známky ischemie na kontrolním CT. I zde očekávané četnosti splňují podmínky pro dobrou aproximaci, proto provedeme pouze  $\chi^2$  test v kontingenční tabulce:

```
>
chisq.test(ctcomplet$CTAOCCAFTER24,ctcomplet$INFARCTAFTER24,correct=F)

Pearson's Chi-squared test

data: ctcomplet$CTAOCCAFTER24 and ctcomplet$INFARCTAFTER24
```

$\chi^2 = 6.1376$ ,  $df = 1$ ,  $p\text{-value} = 0.01323$

| Krvácení         | pozorované četnosti |     | očekávané četnosti |       |
|------------------|---------------------|-----|--------------------|-------|
|                  | Ne                  | Ano | Ne                 | Ano   |
| rekanalizovaní   | 17                  | 35  | 11,92              | 40,08 |
| nerekanalizovaní | 5                   | 39  | 10,08              | 33,92 |

Tab. 37: Znamka ischemie na kontrolním CT

Zde tedy na hladině významnosti 5% zamítáme hypotézu, že jsou poměry stejné v obou skupinách.

## 5. Závěr

Úvodem jsme se seznámili s technologickým pozadím experimentálního lékařského systému, který je vyvíjen ve spolupráci Fakultní nemocnice v Plzni a Západočeské univerzity v Plzni. Systém si klade za cíl vytvořit komplexní znalostní bázi pacientů s prodělanou cévní mozkovou příhodou.

Cílem bylo navrhnout statistické metody pro lékaři požadované analýzy pacientů s prodělanou mozkovou mrtvicí z registru SITS. Většina těchto analýz byla typu porovnání skupin podle určitých kritérií. Porovnávali jsme pacienty s rozdílnou diagnostikou mozkové mrtvice, pacienty, kteří po léčbě trombolýzou krváceli a pacienty s kompletním CT vyšetřením, u kterých byl na CT prokázán uzávěr mozkové tepny.

Metody byly vybrány s ohledem na charakter dat a požadavek budoucího samostatného provádění analýz v rámci experimentálního systému. Z těchto důvodů bylo nutné navrhnout metody jednoduché a s obecnější platností. Ty jsme hledali v oblasti neparametrických metod, které mají nižší nároky na data a mají tak univerzálnější použití. Jako hlavní metody jsme navrhli Kruskal - Wallisův test a  $\chi^2$  testy. V mnohých případech jsme se potýkali s velmi malým množstvím dat, proto jsme se také zabývali randomizačními testy.

Vytvořili jsme statistickou aplikaci, která se skládá ze série příkazů ve statistickém prostředí R. Aplikace měla získávat data automaticky přímo ze znalostní báze experimentálního systému skrze nástroj MetaMed. Bohužel se nástroj MetaMed ukázal při složitějších dotazech jako nespolehlivý. Tento problém se dosud nepodařilo vyřešit, může zde tedy být prostor k dalšímu vývoji. Aplikace načítá data ze souboru, provádí navrhované metody a výsledky (tabulky, grafy) ukládá do samostatných souborů.

Aplikace byla v poslední části otestována na aktuálních datech z registru SITS. Byly zde také představeny jednotlivé příkazy provádějící testy v prostředí R.

## Seznam tabulek

|  |    |
|--|----|
| Tab. 1: RDF trojice .....  | 7  |
| Tab. 2: Četnosti pacientů ve skupinách ICD .....                               | 20 |
| Tab. 3: Průměr a medián věkového rozložení podle ICD.....                      | 21 |
| Tab. 4: Pozorované četnosti pohlaví ve skupinách ICD .....                     | 24 |
| Tab. 5: Očekávané četnosti pohlaví ve skupinách ICD .....                      | 24 |
| Tab. 6: Počty pacientů s vyplněným NIHSS .....                                 | 26 |
| Tab. 7: Klinická změna po 24 hodinách – pozorované četnosti .....              | 31 |
| Tab. 8: Klinická změna po 7 dnech - pozorované četnosti.....                   | 31 |
| Tab. 9: Klinická změna po 24 hodinách - očekávané četnosti.....                | 32 |
| Tab. 10: Klinická změna po 7 dnech - očekávané četnosti.....                   | 32 |
| Tab. 11: Rankin score – popis škály .....                                      | 34 |
| Tab. 12: Rankin score - pozorované a očekávané četnosti .....                  | 35 |
| Tab. 13: Průkaz zavřené tepny na CT - pozorované četnosti .....                | 36 |
| Tab. 14: Průkaz zavřené tepny na CT - očekávané četnosti .....                 | 37 |
| Tab. 15: Průkaz rekanalizace - pozorované četnosti.....                        | 38 |
| Tab. 16: Průkaz rekanalizace - očekávané četnosti .....                        | 38 |
| Tab. 17: Četnost pacientů s krvácením a hodnocením NIHSS.....                  | 40 |
| Tab. 18: Četnost pacientů s diagnózou diabetes .....                           | 40 |
| Tab. 19: Četnost pacientů s diagnózou art. hypertenze .....                    | 41 |
| Tab. 20: Četnost pacientů kuřáků .....   | 41 |
| Tab. 21: Četnost pacientů s nasazenou antiagregační léčbou.....                | 42 |
| Tab. 22: Průměrný čas do zahájení léčby ve skupinách dle symptomatičnosti..... | 43 |
| Tab. 23: Znamka ischemie na CT – pozorované četnosti.....                      | 44 |
| Tab. 24: Klinická změna po 7 dnech – pozorované četnosti .....                 | 45 |
| Tab. 25: Pozorované četnost pacientů ve skupinách ICD .....                    | 45 |

---

|   |    |
|---|----|
| Tab. 26: Rankin score - pozorované četnosti.....                  | 46 |
| Tab. 27: Četnost pacientů ve skupinách s CT vyšetřením.....       | 47 |
| Tab. 28: Spektrum ICD - pozorované četnosti .....                 | 47 |
| Tab. 29: Spektrum ICD - očekávané četnosti .....                  | 47 |
| Tab. 30: Klinická změna po 24 hodinách - pozorované četnosti..... | 50 |
| Tab. 31: Klinická změna po 7 dnech - pozorované četnosti.....     | 50 |
| Tab. 32: Klinická změna po 24 hodinách - očekávané četnosti.....  | 50 |
| Tab. 33: Klinická změna po 7 dnech - očekávané četnosti.....      | 51 |
| Tab. 34: Rankin score - pozorované četnosti.....                  | 52 |
| Tab. 35: Rankin score - očekávané četnosti .....                  | 53 |
| Tab. 36: Znamka ischemie na vstupním CT .....                     | 55 |
| Tab. 37: Znamka ischemie na kontrolním CT .....                   | 56 |



## Seznam obrázků

|  |    |
|--|----|
| Obr. 1: RDF graf.....  | 7  |
| Obr. 2: Kontingenční tabulka typu r x c .....                            | 15 |
| Obr. 3: Četnosti pacientů podle pohlaví a věkových kategorií.....        | 19 |
| Obr. 4: Rozložení časových rozdílů pro celý soubor dat ze SITS .....     | 19 |
| Obr. 5: ICD kódování pro diagnostiku cévních mozkových příhod.....       | 20 |
| Obr. 6: Graf věkových rozložení skupin podle ICD.....                    | 22 |
| Obr. 7: Graf rozložení NIHSS .....                                       | 27 |
| Obr. 8: Graf rozložení rozdílů vstupního NIHSS a NIHSS po 24h.....       | 28 |
| Obr. 9: Graf rozložení NIHSS po 7 dnech.....                             | 29 |
| Obr. 10: Graf rozložení rozdílu vstupního NIHSS a NIHSS po 7 dnech ..... | 30 |
| Obr. 11: Graf rozložení časů do zahájení léčby.....                      | 44 |
| Obr. 12: Graf rozložení rozdílů vstupního NIHSS a NIHSS po 24h.....      | 49 |
| Obr. 13: Graf rozložení časů do zahájení léčby.....                      | 54 |

## Seznam použité literatury

- [1] *Safe Implementation of Treatments in Stroke*. [Online] 2012. [cit. 10. 6. 2012.]  
Dostupný z www: <https://sitsinternational.org/>.
- [2] **MLČOCH, Zbyněk**. Cévní mozková příhoda, mrtvice - komplexní článek, trombolýza (fibrinolýza). *MUDr. Zbyněk Mlčoch*. [Online] 2008. [cit. 21. 5. 2012.]  
Dostupný z www: <http://www.zbynekmlcoch.cz/informace/medicina/neurologie-nemoci-vysetreni/cevni-mozkova-prihoda-mrtvice-komplexni-clanek-trombolyza-fibrinolyza>.
- [3] **MIKULÍK, Robert**. Registr SITS. [Online] 2010. [cit. 21. 6. 2012.] Dostupný z www: [http://www.cmp.cz/jnp/cz/pece\\_o\\_cmp\\_v\\_cr/registr\\_sits.html](http://www.cmp.cz/jnp/cz/pece_o_cmp_v_cr/registr_sits.html).
- [4] **VČELÁK, Petr, KLEČKOVÁ, Jana a ROHAN, Vladimír**. *Experimental database implementation for the cerebrovascular diseases research integrates together different kinds of medical data*. International Multi-Conference on Innovative Development in ICT, 2010.
- [5] **World Wide Web Consortium**. RDF Primer. [Online] 2004. [cit. 20. 6. 2012.]  
Dostupný z www: <http://www.w3.org/TR/rdf-primer/>.
- [6] **ORACLE**. Oracle Database Semantic Technologies. [Online] 2009. [cit. 20. 6. 2012.] Dostupný z www:  
<http://www.oracle.com/technetwork/database/options/semantic-tech/semtech11gr2-featover-131765.pdf>.
- [7] **ORACLE**. Oracle Database Semantic Technologies Overview. [Online] 2010. [cit. 20. 6. 2012.] Dostupný z www:  
[http://docs.oracle.com/cd/E18283\\_01/appdev.112/e11828/sdo\\_rdf\\_concepts.htm](http://docs.oracle.com/cd/E18283_01/appdev.112/e11828/sdo_rdf_concepts.htm).
- [8] **HEJNA, Miroslav**. *Přehled statistických testů a softwarová podpora těchto testů na ZČU*. Plzeň, 2009. Bakalářská práce. Západočeská univerzita. Fakulta aplikovaných věd. Katedra matematiky. Vedoucí práce Blanka Šedivá.
- [9] **HENDL, Jan**. *Přehled statistických metod*. Praha : Portál, s.r.o, 2009. ISBN 978-80-7367-482-3.
- [10] **MARQUES de Sá, Joaquim**. *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*. 2. vyd. Berlin : Springer, 2007. ISBN 978-3-540-71971-7.

- [11] **SIEGEL Sidney, CASTELLAN John N.** *Nonparametric statistics for the behavioral sciences*. 2. vyd. Columbus : McGraw-Hill, 1988. ISBN 978-0070573574.
- [12] **MCDONALD, John H.** Handbook of biological statistics. [Online] 2009. [cit. 30. 5. 2012.] Dostupný z www: <http://udel.edu/~mcdonald/statintro.html>.
- [13] **STEFANESCU, C., BERGER, V. a HERSHBERGER, S.** Yates's continuity correction. [Online] 2004. [cit. 21. 6. 2012.] Dostupný z www: <http://faculty.london.edu/cstefanescu/Yates.pdf>.
- [14] **MIKULÍK, DUFEK, GOLDEMUND, REIF.** Pokyny k vyplňování NIHSS škály. *Cerebrovaskulární sekce České neurologické společnosti ČLS JEP*. [Online] 2003. [cit. 30. 5. 2012.] Dostupný z www: [http://www.cmp.cz/public/a2/28/a5/4011\\_16402\\_pokyny\\_k\\_vyplnovani\\_nihss.pdf](http://www.cmp.cz/public/a2/28/a5/4011_16402_pokyny_k_vyplnovani_nihss.pdf).
- [15] Modified Rankin Scale. *Internet Stroke Center*. [Online] [cit. 6. 6. 2012.] Dostupný z www: [http://www.strokecenter.org/wp-content/uploads/2011/08/modified\\_rankin.pdf](http://www.strokecenter.org/wp-content/uploads/2011/08/modified_rankin.pdf).
- [16] **ZVÁROVÁ, Jana.** *Biomedicínská statistika I*. Praha : Karolinum, 2002. ISBN 80-7184-786-0.
- [17] An Introduction to R. *The R Project*. [Online] Dostupný z www: <http://cran.r-project.org/doc/manuals/R-intro.html>.
- [18] **MZČR.** Věstník č. 2/2010: Péče o pacienty s cerebrovaskulárním onemocněním České republiky. [Online] 2010. [cit. 20. 6. 2012.] Dostupný z www: <http://www.mzcr.cz/Legislativa/Soubor.ashx?souborID=9372>.

## **Seznam příloh**

**Příloha A:** Kódy v dotazovacích jazycích pro získání dat

## Příloha A: Kódy v dotazovacích jazycích pro získání dat

SQL příkaz pro získání potřebných dat z modelu SITS:

```
SELECT s, o, age, gender, nih, after24hnih, after7dnh, after24h,
after7d, after3m, ctaocc, ctddense, ctaoccafter24, ctddenseafter24,
dattreatment, datonset, lhaemorrhage, rhaemorrhage, diabetes,
hypertension, currsmoker, prevsmoker, AspirinL75, Aspirin75,
Aspiring75, Aspiring200, Dipyridamole, Clopidogrel, Antiplatelet,
Infarct, Infarctafter24, ctPerfusion, CTDone, CTDoneafter24, icd FROM
table(SEM_MATCH(
  {?s sits:Baseline__TreatmentFile ?o
  . OPTIONAL {?s sits:Baseline__Age ?age}
  . OPTIONAL {?s sits:Baseline__Gender ?gender}
  . OPTIONAL {?s sits:Baseline__NIHScore ?nih}
  . OPTIONAL {?s sits:After24H__NIHScore ?after24hnh}
  . OPTIONAL {?s sits:After7D__NIHScore ?after7dnh}
  . OPTIONAL {?s sits:After24H__GlobalOutcome ?after24h}
  . OPTIONAL {?s sits:After7D__GlobalOutcome ?after7d}
  . OPTIONAL {?s sits:After3M__RankinScore ?after3m}
  . OPTIONAL {?s sits:Baseline__CTAOcclusion ?ctaocc}
  . OPTIONAL {?s sits:Baseline__CTDenseArterySign ?ctddense}
  . OPTIONAL {?s sits:After24H__CTAOcclusion ?ctaoccafter24}
  . OPTIONAL {?s sits:After24H__CTDenseArterySign ?ctddenseafter24}
  . OPTIONAL {?s sits:Treatment__DateAndTimeForActilyseTreatment
?dattreatment}
  . OPTIONAL {?s sits:Baseline__DateAndTimeStrokeOnset ?datonset}
  . OPTIONAL {?s sits:After24H__CTLocalHaemorrhage ?lhaemorrhage}
  . OPTIONAL {?s sits:After24H__CTRemoteHaemorrhage ?rhaemorrhage}
  . OPTIONAL {?s sits:Baseline__DiabetesDiagnosisOfDiabetes ?diabetes}
  . OPTIONAL {?s sits:Baseline__HypertensionDiagnosisOfHypertension
?hypertension}
  . OPTIONAL {?s sits:Baseline__CurrentSmokerCurrentSmokerAtStrokeOnset
?currsmoker}
  . OPTIONAL {?s
sits:Baseline__PreviousSmokerPreviousSmokerButStoppedBeforeStrokeOnset
?prevsmoker}
  . OPTIONAL {?s sits:Baseline__AspirinLessThen75MgOverDay ?AspirinL75}
  . OPTIONAL {?s sits:Baseline__Aspirin75MgOverDay ?Aspirin75}
  . OPTIONAL {?s sits:Baseline__AspirinGreaterThen75MgLessThen200Mg
?Aspiring75}
  . OPTIONAL {?s sits:Baseline__AspirinGreaterThen200Mg ?Aspiring200}
  . OPTIONAL {?s sits:Baseline__DipyridamoleSlowRelease ?Dipyridamole}
  . OPTIONAL {?s sits:Baseline__Clopidogrel ?Clopidogrel}
  . OPTIONAL {?s sits:Baseline__OtherAntiplatelet ?Antiplatelet}
  . OPTIONAL {?s sits:Baseline__CTCurrentInfarct ?Infarct}
  . OPTIONAL {?s sits:After24H__CTCurrentInfarct ?Infarctafter24}
  . OPTIONAL {?s sits:Baseline__CTPerfusionDeficit ?ctPerfusion}
```

```
. OPTIONAL {?s sits:Baseline__CTDone ?CTDone}
. OPTIONAL {?s sits:After24H__CTDone ?CTDoneafter24}
. OPTIONAL {?s sits:After7D__ICDCode ?icd}}',
SEM_Models('SITS'),
null,
SEM_aliases(SEM_ALIAS('sits',
'http://mre.kiv.zcu.cz/ontology/2012/01/sits.owl#')),
null));
```

SPARQL příkaz pro získání potřebných dat z modelu SITS:

```
PREFIX sits: <http://mre.kiv.zcu.cz/ontology/2012/01/sits.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT
  ?Baseline__TreatmentFile
  ?Baseline__Age
  ?Baseline__Gender
  ?Baseline__NIHScore
  ?After24H__NIHScore
  ?After7D__NIHScore
  ?After24H__GlobalOutcome
  ?After7D__GlobalOutcome
  ?After3M__RankinScore
  ?Baseline__CTAOcclusion
  ?Baseline__CTDenseArterySign
  ?After24H__CTAOcclusion
  ?After24H__CTDenseArterySign
  ?Treatment__DateAndTimeForActilyseTreatment
  ?Baseline__DateAndTimeStrokeOnset
  ?After24H__CTLocalHaemorrhage
  ?After24H__CTRemoteHaemorrhage
  ?Baseline__DiabetesDiagnosisOfDiabetes
  ?Baseline__HypertensionDiagnosisOfHypertension
  ?Baseline__CurrentSmokerCurrentSmokerAtStrokeOnset
  ?Baseline__PreviousSmokerPreviousSmokerButStoppedBeforeStrokeOnset
  ?Baseline__AspirinLessThen75MgOverDay
  ?Baseline__Aspirin75MgOverDay
  ?Baseline__AspirinGreaterThen75MgLessThen200Mg
  ?Baseline__AspirinGreaterThen200Mg
  ?Baseline__DipyridamoleSlowRelease
  ?Baseline__Clopidogrel
  ?Baseline__OtherAntiplatelet
  ?Baseline__CTCurrentInfarct
  ?After24H__CTCurrentInfarct
  ?Baseline__CTPerfusionDeficit
  ?Baseline__CTDone
```

```

?After24H__CTDone
?After7D__ICDCode
WHERE
{
  ?report rdf:type sits:SITSReport .
  ?report sits:Baseline__TreatmentFile ?Baseline__TreatmentFile .
  OPTIONAL {?report sits:Baseline__Age ?Baseline__Age .}
  OPTIONAL {?report sits:Baseline__Gender ?Baseline__Gender .}
  OPTIONAL {?report sits:Baseline__NIHScore ?Baseline__NIHScore .}
  OPTIONAL {?report sits:After24H__NIHScore ?After24H__NIHScore .}
  OPTIONAL {?report sits:After7D__NIHScore ?After7D__NIHScore .}
  OPTIONAL {?report sits:After24H__GlobalOutcome
?After24H__GlobalOutcome .}
  OPTIONAL {?report sits:After7D__GlobalOutcome
?After7D__GlobalOutcome .}
  OPTIONAL {?report sits:After3M__RankinScore ?After3M__RankinScore .}
  OPTIONAL {?report sits:Baseline__CTAOcclusion
?Baseline__CTAOcclusion .}
  OPTIONAL {?report sits:Baseline__CTDenseArterySign
?Baseline__CTDenseArterySign .}
  OPTIONAL {?report sits:After24H__CTAOcclusion
?After24H__CTAOcclusion .}
  OPTIONAL {?report sits:After24H__CTDenseArterySign
?After24H__CTDenseArterySign .}
  OPTIONAL {?report sits:Treatment__DateAndTimeForActilyseTreatment
?Treatment__DateAndTimeForActilyseTreatment .}
  OPTIONAL {?report sits:Baseline__DateAndTimeStrokeOnset
?Baseline__DateAndTimeStrokeOnset .}
  OPTIONAL {?report sits:After24H__CTLocalHaemorrhage
?After24H__CTLocalHaemorrhage .}
  OPTIONAL {?report sits:After24H__CTRemoteHaemorrhage
?After24H__CTRemoteHaemorrhage .}
  OPTIONAL {?report sits:Baseline__DiabetesDiagnosisOfDiabetes
?Baseline__DiabetesDiagnosisOfDiabetes .}
  OPTIONAL {?report sits:Baseline__HypertensionDiagnosisOfHypertension
?Baseline__HypertensionDiagnosisOfHypertension .}
  OPTIONAL {?report
sits:Baseline__CurrentSmokerCurrentSmokerAtStrokeOnset
?Baseline__CurrentSmokerCurrentSmokerAtStrokeOnset .}
  OPTIONAL {?report
sits:Baseline__PreviousSmokerPreviousSmokerButStoppedBeforeStrokeOnset
?Baseline__PreviousSmokerPreviousSmokerButStoppedBeforeStrokeOnset .}
  OPTIONAL {?report sits:Baseline__AspirinLessThan75MgOverDay
?Baseline__AspirinLessThan75MgOverDay .}
  OPTIONAL {?report sits:Baseline__Aspirin75MgOverDay
?Baseline__Aspirin75MgOverDay .}
  OPTIONAL {?report sits:Baseline__AspirinGreaterThan75MgLessThan200Mg
?Baseline__AspirinGreaterThan75MgLessThan200Mg .}
}

```

```
OPTIONAL {?report sits:Baseline__AspirinGreaterThan200Mg
?Baseline__AspirinGreaterThan200Mg .}
OPTIONAL {?report sits:Baseline__DipyridamoleSlowRelease
?Baseline__DipyridamoleSlowRelease .}
OPTIONAL {?report sits:Baseline__Clopidogrel ?Baseline__Clopidogrel
.}
OPTIONAL {?report sits:Baseline__OtherAntiplatelet
?Baseline__OtherAntiplatelet .}
OPTIONAL {?report sits:Baseline__CTCurrentInfarct
?Baseline__CTCurrentInfarct .}
OPTIONAL {?report sits:After24H__CTCurrentInfarct
?After24H__CTCurrentInfarct .}
OPTIONAL {?report sits:Baseline__CTPerfusionDeficit
?Baseline__CTPerfusionDeficit .}
OPTIONAL {?report sits:Baseline__CTDone ?Baseline__CTDone .}
OPTIONAL {?report sits:After24H__CTDone ?After24H__CTDone .}
OPTIONAL {?report sits:After7D__ICDCode ?After7D__ICDCode .}
}
```