Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra informatiky a výpočetní techniky

Bakalářská práce

Vizualizace výsledků statistického medicínského šetření

Plzeň 2015 Iva Ptáčková

Prohlášení

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

V Plzni dne 25. června 2015

Iva Ptáčková

Abstract

Visualization of statistical medical research

This bachelor thesis solves visualization of statistical medical research, that is based on the data from patients, who suffered a stroke. The goal is to find optimal means of visualizing chosen statistical methods. The thesis contains basic theory about the phases of statistical prosessing of data, about the register from which it originates and the principle of statistical methods. Solution includes recommendations on graphs with certain methods, on implementation and subsequent implementation of the assignment.

Vizualizace výsledků statistického medicínského šetření

Tato bakalářská práce řeší vizualizaci výsledků statistického medicínského šetření, které vychází z údajů o pacientech s prodělanou mozkovou příhodou. Jejím cílem je najít optimální způsob vizualizace vybraných statistických metod. Práce obsahuje základní teorii, zabývající se fázemi statistického zpracování dat dat, registrem, z kterého data pochází, a principem statistických metod. Náplň cíle zahrnuje doporučení ohledně grafů u daných metod, implementaci a následnou realizaci zadání.

Obsah

1 Úvod

Cílem této bakalářské práce je navrhnout, a následně implementovat a otestovat vhodné řešení pro vizualizaci výsledků statistického šetření medicínských dat, které bude usnadňovat práci při vyhodnocení získaný dat.

První tři kapitoly jsou čistě seznámení se s teorií. Obsahují informace o problematice vizualizace dat, kterou se tato práce, v okrajové části medicínského výzkumu, snaží zmírnit. Dále seznamuje čtenáře s fázemi statistického zpracování medicínských dat, které vedou k výsledným statistickým souborům dat, s možnostmi grafického vyhodnocení těchto dat, s ohledem na povahu dat. A také objasňuje původ dat, s kterými v dalších kapitolách pak pracuje. Tedy seznamuje čtenáře s registrem SITS.

V dalších kapitolách se pak práce začíná zabývat řešením. Je objasněn princip obvykle používaných metod v oblasti zpracování medicínských dat a následně navrhnuto grafické řešení. Vybraná možná řešení jsou pak v dále implementována ve zvoleném programu s spolu se statistickými metodami a následně otestována na dostupných datech.

2 Problematika vizualizace

Data, která se získávají z medicínských přístrojů, se již dají považovat za jistou interpretaci, ale bez příslušných znalostí nejsou tak čitelná a jednoznačná. V některých případech se jejich hodnota zvýší až v kombinaci s jiným údajem, či výstupem jiné analytické metody.

Základní data, s kterými se pracuje při vizualizaci, jsou data získaná z papírových schématických dat. Tato data jsou vhodně přizpůsobena a přeformulována do elektronické podoby. Jedná se například o výstup při provádění počítačové tomografie známé také jako CT (Computer Tomography). Data jsou sice už v elektronické podobě, ale ne ve formě, ze které by se dala dále zpracovat komplexněji. [18]

Přeformulování a zadání dat do systému je prací lékařského pracovníka. Protože nejsou vyhodnocována žádným programem a je zde lidský faktor, jsou vkládaná data ovlivněna vnímáním, zkušenostmi a také i interakcí zadavatele s počítačem. I tak je tu snaha omezit tento faktor, a to definováním standardního postupu a podpory získávání znalostí. [18]

Z toho důvodu jsou faktory jako vnímání, poznávání, komunikace člověka s počítačem a podpora znalostí podstatnými prvky v procesu, který vede k vizualizaci. [18]

2.1 Výhody

S růstem možných vyšetření a medicínských záznamů roste i množství informací o jednotlivých pacientech. Na jednoho pacienta jsou pak k dispozici velká kvanta různorodých dat - textové, numerické, diagramové, apod. Jejich vizualizace je možným

řešením pro uvědomění si komplexnosti anamnézy a využití celého jejího potenciálu. Při kreativním řešení je šance dosáhnout výsledků, jako oproštění se od tradičního smýšlení a nalezení nových východisek při dané léčbě. [18]

2.2 Kompletace dat do grafického 3D objektu

Jednou z myšlenek jak data vizualizovat pro urychlení celkových diagnostik je nechat projít data speciálním programem, který je zpracuje do specifických bloků. Tyto bloky jsou pak součástí jednoho grafického 3D objektu, který obsahuje vrstvy. Výsledkem je tedy virtuální variace pacienta. [19]

Tato myšlenka vznikla na základě velkého nárůstu počtu informací, které jsme schopni zpracovat. Tento nárůst je se pak projevuje i např. v množství snímků, které během vyšetření jsou schopny přístroje vyprodukovat. Pro srovnání jak gigantický je to obrat oproti minulosti: 100 obrazů o zhruba 50MB v minulosti a 24000 obrazů o 20GB v dnes. Další důvod proč je tento software vyvíjen je časová náročnost prohlédnutí a zanalyzování všech těchto snímků. [19]

Program má jednu nevýhodu, a to jsou velké požadavky na grafiku přístroje. Ale spolu sním je vyvíjen i způsob práce s výsledným objektem. Pro jejich prohlížení existují dotykové pracovní plochy různých velikostí. Využití je možné, jak ve vzdělávacích nebo výzkumných institutech, tak i do budoucna v nemocnicích [19]

2.3 Data produkovaná lékařským systémem

Hlavní problematikou je tedy prezentace data, které medicínský systém vyprodukuje. Typy dat jsou děleny do skupin: textová-vypovídající o něčem, numerické hodnoty např. z laboratorních výsledků, signálové např. ECG a obrazové. Struktura dat je

pak závislá na způsobu vyjádření speciálním kódováním, protože standardní slovní zásoba, na kterou jsme zvyklí, nedokáže vyjádřit efektivněji význam těchto hodnot. [20]

3 Statistické zpracování medicínských dat

V medicíně mají statistické metody velký význam v oblasti rozhodování, kde jsou její výsledky využívány k zajištění nejlepší možné péče, alokaci nákazy v případě epidemie apod. Je to neodmyslitelná část výzkumu v oblasti medicíny, vedoucí k pokroku a vývoji např. léčebných metod. [15]

Statistika v lékařství je rozdělena do 3 fází [15]:

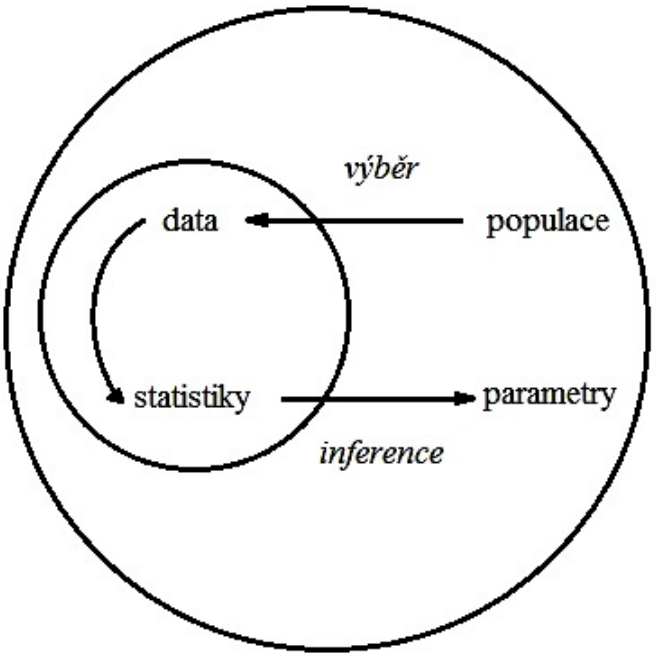
- sběr dat,
- analýza dat,
- statistické usuzování.

Jako první krok, kterým se začíná toto statistické šetření, je vytyčení otázek, na které hledáme odpovědi, a ustanovení cílové skupiny. Sběr dat se točí kolem sledovaného vzorku, protože od jeho kvality se pak odvíjí i kvalita výzkumu a přesnost jeho výsledků. Absence kvality je tedy nejčastějším problémem při aplikaci statistické metody. Hlavním měřítkem při sběru vzorku je jeho náhodnost a reprezentativnost. K náhodnosti vzorku nám pomáhá software, který pomocí pseudonáhodných čísel vygeneruje jeden náhodný vzorek z celkového datového souboru.[15]

Ve druhé fázi, tedy v analýze dat, jde o vystihnutí podstaty sesbíraných dat, která dostatečně shrne vlastnosti dat. K tomu napomáhá vhodný statistický software. Nejčastěji se využívají hodnoty deskriptivní statistiky, jako je medián, průměr, modus či směrodatná odchylka. Každé ze zde zmiňovaného má své klady a zápory, a je vhodné pouze za jistých okolností. Průměr je například žádoucí u velkého množství nashromážděných dat.[15]

Ke zorientování se v datech nám slouží různá grafická vyjádření. Nejčastěji se při analýze dat používají grafy typu histogram, a pak krabicový nebo bodový graf. Jejich výběr je odvíjen od používaných dat.[15]

Poslední fází je tedy statistické usuzování (viz obr.3.1). Jedná se o odhad skutečné pravděpodobnosti výskytu události. Hlavní úlohu zde hraje především už zmiňovaná náhodnost vzorku, protože se z jeho zpracování vyvozují závěry týkající se původce vzorku, nějaké množiny pozorovaných. Vesměs nahodilost vzorků by měla být taková, že při novém zpracování člověk dojde opět k podobným výsledkům s minimálním rozdílem či ke konstantě. Úskalím poslední fáze je schopnost objektivního zhodnocení, ke které je potřeba rozumět metodám zkoumání a znát jejich předpoklady. [15]



Obrázek 3.1: Proces statistického usuzování

S první fází je spojen i častý jev v rámci dat jedné cílové skupiny, a to různorodost nasbíraných dat. Shromažďovaná data mohou být kombinací kvalitativních

a kvantitativních dat. To je důsledek stylizace otázek, která je v rámci medicínských dat pestrá.

Kvalitativní data pak mohou být trojího druhu: binární, nominální a ordinární. V případě binárních dat jsou data reprezentována většinou *true/false* hodnotou, která může být dále vyjádřena např. *ano/ne* nebo *1 a 0*. Obecně jsou data schopna nabývat pouze dvou hodnot. Nominálními daty jsou označena data, v kterých je obsaženo více kategorií bez možnosti seřazení dle významu nebo hodnoty. U ordinálních dat je seřazení možné i přes kategoriální obsah. Jedná se většinou o data, která vyjadřují nějakou určitou stupnici nebo velikost.

Kvantitativní data je možné dělit na spojité a diskrétní. U spojitých se objevují různá čísla z určitého intervalu. Interval může být nějaká množina čísel, která je logicky ohraničena. Např. věkové rozpětí, teplota, atd. Diskrétní data jsou vyjádřením libovolné četnosti celočíselnou hodnotou, nejsou omezeny intervalem.

3.1 Grafická interpretace statistických dat

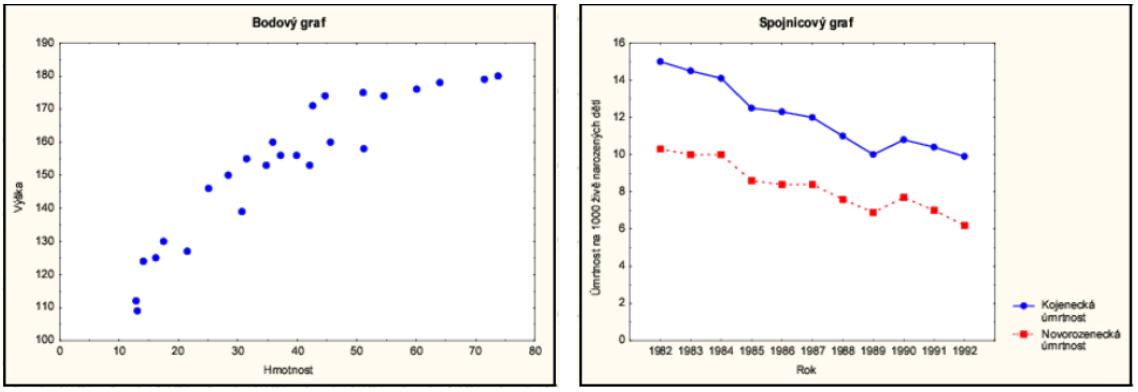
Ve statistice jsou k dispozici různé způsoby a možnosti jak vyjádřit nasbíraná data, a to tabulkou či grafem. Kombinací tabulky a grafu lze dosáhnou přesné představy, vyplývajících souvislostí a dispozicí daných dat. [8]

3.1.1 Bodový a spojnicový graf

Bodový graf (obr.3.2 vlevo)je používán převážně v případě, že je nutné vyjádřit závislost znaků. Naměřené hodnoty jsou vyjádřeny pomocí souřadnic. Graf je vhodný pouze pro číselné hodnoty o velké četnosti. Přesnost měření je ovlivněna objemem dat. V případě, že je zkoumáno více hodnot jiných skupin, jsou použity jiné symboly

při vykreslení. V tomto případě je pak možné, že velké množství zobrazených dat může být matoucí, proto se doporučuje použití spojnicového grafu. [8]

Spojnicový graf (obr.3.2 vpravo) je určen převážně pro data, která jsou závislá na čase. Lze jej použít jako polygon četností, pokud jím je znázorněno rozdělení relativních a absolutních četností spojitého znaku. [8]



Obrázek 3.2: Prezentace rozdílu bodového a spojového grafu [8]

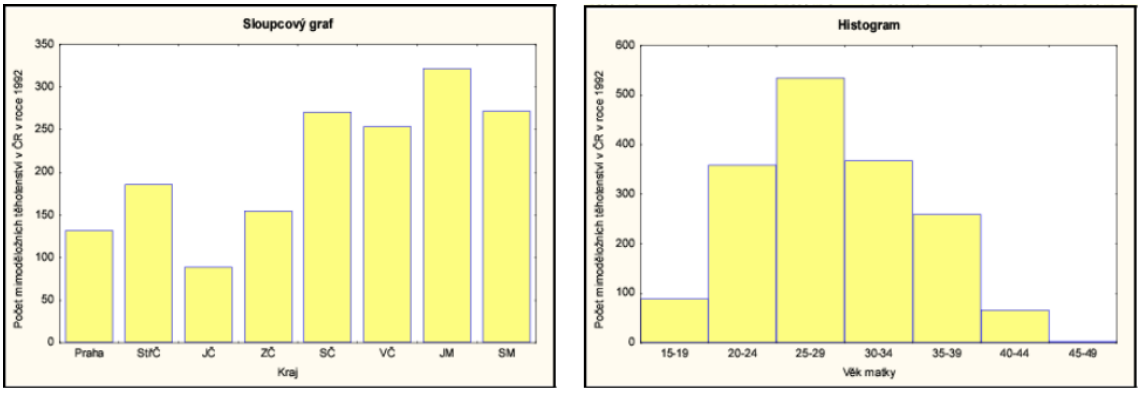
3.1.2 Sloupcový graf a histogram

Sloupcový graf (obr.3.3 vlevo)je určen pro porovnávání četností v rámci několikanásobného rozdělení. Používá se tedy pokud zkoumáme jeden faktor, ale u více skupin. U těchto skupin se jedná spíše o kvalitativní veličiny.[8]

Jeho stavba je bez hlubšího významu, pokud tedy nejsou různé velikosti tříd, kdy je pak třeba zachytit přímou úměrnost šířky sloupce na velikosti třídy. V tomto směru se pak už jedná spíše o histogram. [8]

Histogram (obr.3.3 vpravo) je používán pro vyjádření relativních nebo absolutních četností spojitého znaku. Význam stavby grafu není založen pouze na výšce sloupce, ale také na jeho ploše, protože četnost třídy je shodná s násobkem výšky

a šířky sloupce. [8]



Obrázek 3.3: Prezentace rozdílu sloupcového grafu a histogramu [8]

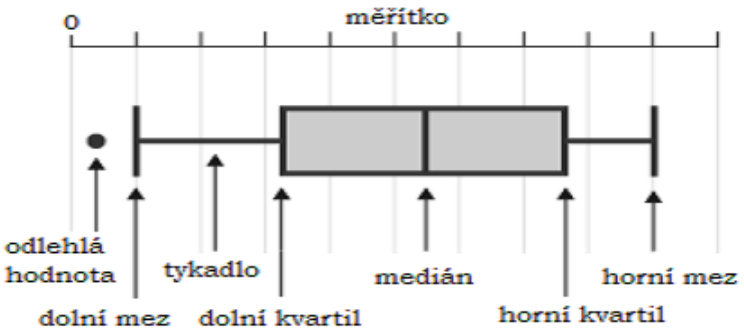
3.1.3 Kruhový graf

Někdy označován také jako *koláčkový* nebo *výsečový graf*. Je používán v kombinaci s procenty, kdy celá jeho plocha je grafické vyjádření celému souboru a četnosti a procentuální zastoupení je vyjádřeno kruhovými výsečemi. Jednotlivé části tohoto kruhu, které tvoří výseče, jsou pak typicky barevně rozlišeny pro jednoznačnost. [8]

3.1.4 Krabicový diagram

Je používán k vizualizaci číselných dat, a to pomocí jejich kvartilů. U krabicového diagramu (viz. 3.4) se nejčastěji vyobrazuje medián, horní a dolní kvartil, odlehlé hodnoty a extrémy, a také pak i tykadla (tzv. vousky), která jsou v prostoru mezi jedním z extrémů (např. nejméně) a kvartilem (např. dolním). Je vhodný pro neparametrická měření.[8]

V některých případech je možné jej použít i na jiné typy dat. V tom případě je



Obrázek 3.4: Popis krabicového diagramu

pak medián chápán jako průměr a místo kvartilů jsou uvedeny násobky směrodatné chyby a extrémy pak uvádějí násobek směrodatné odchylky. [8]

4 Registr SITS

Safe Implementation of Treatments in Stroke dále jen SITS je registr užívaný k sběru dat ohledně pacientů s mrtvicí. Tyto data jsou dále zpracovávána a použita k výzkumu a analýze pro zlepšení následné péče a léčby pacientů. Jedná se o mezinárodní registr a sběr dat je zcela anonymní.

Formulář pro sběr dat se skládá z více tzv. kapitol s podbody. Začíná identifikací pacienta a shrnutím informací o něm a jeho hospitalizaci. Ostatní otázky k vyplnění jsou již zaměřeny na průběh a následky mozkové příhody a reakce ošetřujícího lékaře na ni. Tj. informace o podaných lécích a zjištěné informace z vyšetření.

Každá položka ve formuláři představuje statisticky hodnotnou informaci. Vyplnění formuláře je součástí statistického šetření. Přesněji fáze analýzy dat. Formulář je sestaven tak, aby vyplňování nestálo ošetřujícího lékaře zbytečně příliš mnoho času. Téměř vše je v bodech s možností zaškrtnutí jedné varianty. Pro statistické vyhodnocování ideální, většinou číselně zhodnoceno nebo jsou zde kladné/záporné varianty. Některé hodnoty jsou pro porovnávání opět měřeny i s odstupem času. U dat jako jsou *modified rakin scale* a *national institute of health stroke scale* je možné přesně vidět, jak může být průběh sběru dat někdy složitý, a proto je u některých dat zjednodušen výběr anamnézy pouhou stupnicí. Ovšem i to někdy může být zavádějící.

4.1 Modified Rankin Scale (mRS)

Modifikovaná Rankinova stupnice je používána k zhodnocení celkového stavu pacienta po mrtvici. Stupnice je v rozmezí 1-6, kde ke každému číslu je přiřazen stav, ve kterém se pacient může nacházet [12]:

0 Bez symptomů
1 Bez výraznějšího omezení, schopen vykonávat všechny obvyklé denní potřeby a aktivity
2 Lehká invalidita: neschopnost vykonávat všechny dříve obvyklé aktivity, schopen vykonávat všechny své potřeby bez dopomoci
3 Mírná invalidita: vyžaduje pomoc, ale je schopen chůze bez pomoci
4 Středně těžká invalidita: neschopnost chůze bez dopomoci, neschopnost vykonávat tělesné potřeby bez dopomoci
5 Těžká invalidita: upoután na lůžko, inkontinentní, vyžaduje nepřetržitou péči
6 Smrt

Při vyhodnocování lze postupovat systematicky (graficky viz. Příloha A). První položenou otázkou je, zda pacient je plně soběstačný a schopen žít sám. V případě že ano, následuje otázka, zda je pacient schopen vykonávat všechny aktivity jako před příhodou. Pokud ne, je mu přiřazeno v stupnici číslo 2, tedy lehká invalidita. Jestliže je však schopen těchto aktivit, následuje poslední otázka, zda je zcela bez deficitu. Ano – je označen jako zcela bez symptomů, tedy hodnotou 0. Ne - je označen jako bez výraznějšího omezení, ve stupnici hodnota 1.[13]

Pokud na první otázku zní odpověď ne, je položena otázka, zda je schopen chůze bez pomoci jiné osoby. Z kladné odpovědi dostáváme, že pacientovy následky se dají označit jako mírná invalidita, čili na stupnici hodnota 3. Jestliže je odpověď ne, je dále na místě položení další otázky, zda je, či není upoután na lůžko. V přídě záporné odpovědi je osoba po mrtvici označena jako středně invalidní, tedy stupeň 4. A pokud je odpověď kladná, znamená to pro pacienta těžkou invaliditu, tedy 5. stupeň. Dalším, 6. stupněm je smrt.[13]

4.2 National Institute of Health Stroke Scale (NIHSS)

NIH Stroke Scale je standardizované neurologické vyšetření sloužící k zjištění deficitu u pacienta s mozkovou příhodou. Byl vytvořen za účelem homogenního vyhodnocení stavu pacienta, a to z důvodu dalšího zpracování dat (porovnávání dat pacientů).[12, 13]

Vyšetření zahrnuje těchto 15 bodů:

1a. Level of Consciusness
1b. LOC Questions
1c. LOC Commands
2. Best Gaze
3. Visual
4. Facial Palsy
5a. Motor Right Arm
5b. Motor Left Arm
6a. Motor Right Leg
6b. Motor Left Leg
7. Limb Ataxia
8. Sensory
9. Best Language
10. Dysarthria
11. Extinction and Inattention - Neglect

Postup je jednotný. Pokládají se postupně otázky, na které pacient odpovídá bez vměšování se vyšetřujícího. Otázka se vyhodnotí buď jako správná či nesprávná. První odpověď je ta, s kterou se pracuje, pokud se pacient později opraví, nebere se na to zřetel. Vyhodnocuje se pouze to, čeho je v danou chvíli pacient schopen.[12, 13]

První bod *Level of Consciusness* se nikdy nepřeskakuje, jiné, které se nevyskytují, lze vynechat. Stupnice hodnocení je pak u každého bodu rozdílná (viz. Přiloha B).

4.3 Imaging-CT

4.3.1 CT ASPECTS Score

Jedná se o 10 bodové kvantitativní topografické hodnocení a je podotázkou v zobrazování CT (Imaging-CT). Alberta Stroke Program Early CT Score se používá jako nástroj k přesnějšímu hodnocení častých známek ischemie nebo-li infarktu, hlavním důvodem používání je sjednocení a větší spolehlivost výsledných závěrů. Je vyhodnocován z původních snímků z CT.[12, 13]

5 Časté statistické metody v oblasti medicíny

5.1 Analýza dat

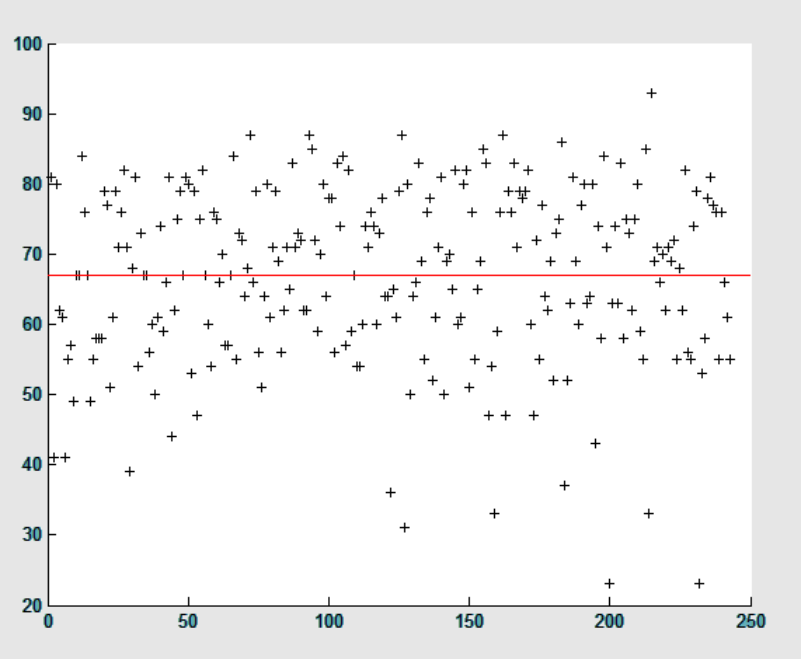
Analýza dat je oblast statistiky, která je známá také jako popisná statistika. Je to způsob charakterizace či prezentace dat. Analyzovaná data jsou zpracována většinou za účelem porozumění řešené problematiky. Tohoto cíle lze dosáhnou především je-li znám i kontext nasbíraných dat. Nemalou součástí procesu dospění k pravdivé odpovědi je pak i schopnost porozumět grafickým i výpočetním výstupům pocházejících z statistického zpracování dat. Tato skutečnost se ovšem týká celé oblasti statistiky.[5, 8]

Pro zobrazování dat jsou využívána tabulková či grafová vyjádření. Tabulkové řešení je vhodné při předpokladu, že bude docházet ještě k dalšímu zpracování naměřených hodnot, nebo pokud je nutné zachovat výsledek statistického šetření v korektním tvaru. Grafické vyjádření je upřednostňováno hlavně v případě, když jsou hlavním zkoumaným faktorem dat kvalitativní vlastnosti. Obecně je grafické zpracování výsledných hodnot vždy přínosem pro pochopení globálního hlediska získaných výstupů.[5, 8]

V téhle oblasti zpracování dat se jedná o naměřené veličiny vycházející z většího souboru dat. Jednou skupinou těchto veličin jsou střední hodnoty, resp. míry centrální tendence. Typicky je řeč o aritmetickém průměru, nebo jeho rozšíření jako váženého aritmetického průměru. A dále pak modus a medián, který je oproti aritmetickému průměru více nevšímavý k odlehlým hodnotám. [5]

Pro většinu z nich je vhodné použití bodového grafu (ukázka viz obr. 5.1) s vy-

značenou výslednou střední hodnotou. Při tomto grafickém vyjádření je pak možné zachytit i rozptyl kolem osy, která je samotným vyjádřením střední hodnoty.



Obrázek 5.1: Vizualizace aritmetického průměru ($\bar{6}6.893$)

5.2 Neparametrické testy

Neparametrické testy jsou vhodné porovnávání souborů dat, která nemají definované své rozdělení. Jedná se o metody univerzální povahy se sníženou statistickou efektivitou. Výstup těchto metod je pak nutné chápat z obecného hlediska.

5.2.1 Testování hypotéz

Při vyhodnocování statistických testů je nutné stanovení nulové a alternativní hypotézy. Nulová hypotéza (H_o) je tvrzení, které může být formulováno jak s potřebou kladné, tak záporné odpovědi. Záleží, zda chceme dosáhnout menšího okruhu možností, nebo se chceme dozvědět, zda toto určité tvrzení je správné. V konečném důsledku to ale má stejný význam. Alternativní hypotéza jsou pak všechna zbylá tvrzení. Tedy jen nám při svém potvrzení říká, že je H_0 zamítnuta. Nutnost formulace obou hypotéz je tedy bezpředmětná. [8]

Výpočtem statistické metody je pak zjištěno s jakou pravděpodobností bychom mohli dostat pozorovaná data, která by ještě více odporovala H_0, za předpokladu, že je H_0 pravdivá. Vypočtená pravděpodobnost je pak označena jako *dosažená hladina významnosti* p. Důvěryhodnost H_0 je pak závislá na velikosti p, s větší hodnotou p roste důvěryhodnost. [8]

Pro rozhodnutí zda se H_0 zamítne či nikoliv je nutné určit *hladinu významnosti*. Obvykle je to 0,1 nebo 0,05. Pokud je p pak menší než je určená hodnota, je H_0 zamítnuta. Tedy nulovou hypotézu zamítneme, pokud p překročí určitou mezní hodnotu. [8]

5.2.2 Kruskal-Wallisův test

Kruskal-Wallisův test je neparametrickou verzí metody analýzy rozptylu jednoduchého třídění. Tento způsob testování dat je využíván pokud jsou výběry z rozdělení, které je značně odlišné od normálního rozdělení. Je aplikován při testování shody zvoleného pravděpodobnostního rozdělení srovnávaných skupin. Data, s kterými pracuje, nevycházejí z normálního rozdělení, a jsou na sobě nezávislé. Jeden z předpokladů použití této metody je přítomnost dat, které obsahují dva a více naměřených

údajů. [2, 5, 1]

V principu jsou data rozdělena do skupin (např. žena, muž). Je zjištěn stupeň volnosti a zvolena kritická hodnota (χ^2-rozdělení). Data skupin jsou seřazena dle velikosti napříč skupinami, a následně je jim přiřazena hodnota pořadí (dále jen rank). V případě shodných naměřených hodnot se přechází k přiřazení průměru z pořadí. Data jsou nadále zpět rozřazena do svých skupin, ale reprezentována svojí rank hodnotou. Skupiny jsou pak sumarizovány a je určena četnost jejich dat. Po dosazení do vzorce (viz. Příloha C) je výsledek porovnán s hladinou významnosti. H_0 je pak zamítnuta nebo přijata na základě tohoto porovnání.

Kruskal-Wallisův test je jedním z testů, které nelze efektivně vizualizovat jinak než s pomocí krabicového diagramu. Hlavním důvodem je jeho skupinové rozdělení, při kterém se pracuje vždy na dvou a více rozdělení. V diagramu se pak nejeví tak skromně a údaje z něj vyplývající jsou zřejmé.

5.2.3 Simultánní porovnávání

Toto porovnávání je zároveň také *post hoc* analýzou, která se používá v případě zamítnutí H_o u předešlé metody. Analýzu je možné provádět, aniž by tomu předcházela specifikace srovnání dat. Princip metody je postaven na porovnávání mediánů statisticky usuzovaných skupin. Pro výslednou hodnotu je potřeba porovnat navzájem všechny skupiny (vzorec viz Příloha C). Pro výslednou vizualizaci je samozřejmostí krabicový graf, protože obsahuje v sobě medián.[5]

5.2.4 χ^2 test dobré shody

Tento test je neparametrickou metodou, která je používána v případě na sobě nezávislých dat. Základ této metody je v ověření shody usuzovaných četností s četnostmi,

které byly vypozorovány (vzorec viz. Příloha C). Data je možné rozdělit do kategorií nebo na intervaly. Záleží na typu dat, jestli jsou kategoriálního typu či intervalového typu.[3, 4, 1]

V praxi je porovnávána nominální proměnná s dvěma a více hodnotami. Porovnávány jsou pak pozorované hodnoty s očekávanými hodnotami, které je možné vypočítat prostřednictvím nějakého teoretického očekávání (Např. 1:1, kdyby šlo o pohlaví).[7]

Pro přesnější výsledky se u této metody doporučuje větší množství dat. V opačném případě mohou být výsledky nepřesné. Test je aplikovatelný na již zmíněné kategoriální údaje. Tj. například pohlaví či typ údaje, který posouvá jedince do jisté kategorie.[7]

V testu jsou hlavní zkoumanou veličinou četnosti, proto je logické zobrazování těchto dat pomocí např. dvou histogramů. Podobnost očekávaných a pozorovaných četností je možné tímto způsobem vyhodnotit i bez přesného měření.

5.2.5 Randomizační test dobré shody

Je používán, pokud je nominální proměnná se třemi a více hodnotami a pro χ^2 test dobré shody je vzorek dat příliš malý. Test je prováděn v případě, že z jednoho testu dobré shody není možné pro malého množství očekávaných četností dojít správného výsledku. Aproximační vztah tak malého vzorku dat není přesný.[6, 1]

Základem randomizační verze tohoto testu je pak opakované měření při ještě menším vzorku dat, kdy počítáme vždy jen s náhodně vybraným vzorkem dat z celého vzorku. Přitom je vždy dodržen poměr naměřených dat. Řešení grafické zpracování tohoto testu pak bude obdobné jako u χ^2 testu dobré shody. [6]

6 Implementace v Matlabu

Původní návrh zpracování praktické části této práce v Excelu byl zamítnut z důvodu nemožnosti využití vytvořeného nástroje i na jiných operačních systémech než je Windows. To je nepřehlédnutelný nedostatek, který elegantně vyřešilo rozhodnutí pracovat s Matlabem a udělat nástroj s grafickým uživatelským rozhraním (dále jen GUI) s výstupem v podobě grafů a statistických hodnot. Systémové požadavky na Matlab jsou navíc rozsáhlejší jak u Excelu. Je podporování u Windows (32-bit, 64-bit), Mac OS X (64-bit) a Linux (64-bit). Implementace je provedena ve verzi Matlab R2014a.

Matlab (Matrix Laboratory) je vyšší programovací jazyk a zároveň interaktivní programové prostředí pro numerické výpočty, grafické znázornění a programování, který má mnoho využití, ale primárně slouží k analýze dat, sestavování algoritmů a k vytváření modelů a aplikací. Pracovní prostředí je ovládáno příkazovou řádkou, která umožňuje okamžité zpracování, nebo je možné prostřednictvím skriptů s příponou *.m vykonávat více příkazů naráz. Jeho základní datovou strukturou jsou matice. [16]

Jako pracovní prostředí umožňuje, v rámci grafiky, vykreslování či zobrazování dvou a tří dimenzionálních grafů, obrázků a animací, a dále pak také vizualizaci dat a webový přístup. Pro práci s externími datovými zdroji je možný export i import textových souborů, tabulkových procesů i o velkém objemu dat.[16]

Při pokročilém vývoji softwaru Matlab podporuje objektově-orientované programování a externí rozhraní $Java, C/C++, .NET, ...$[16]

6.1 Princip práce s Matlabem

Základem pro porozumění Matlabu je minimální znalost matic a vektorů. Výhoda maticové datové struktury je hlavně možnost selekce jednotlivých hodnot pomocí indexů, které také určují pozici hodnoty v datovém bloku. Každá hodnota je opatřena specifickým identifikátorem a lze ji snadno oddělit od zbytku dat.

6.2 GUIDE

V pracovní prostředí Matlab je možnost vytvoření si vlastní aplikace nebo GUI, které při používání vlastních či knihovních funkcí, usnadňuje práci širšímu spektru uživatelů. Vytvoření GUI je snadné a přístupné ze základního pracovního prostředí. Při otevření nového souboru v návrhovém prostředí (dále jen GUIDE) se objeví možnosti čisté nebo již přednastavené plochy (základní ovládací prvky, graf s menu, přednastavený dialog, viz příloha C). Souboru s GUI pak náleží přípona *.fig.

Okno pro nastavení GUI obsahuje panel s nástroji pro rychlejší práci. Panel obsahuje základní prvky (také viz. příloha C):

- Push Button
- Slider
- Radio Button
- Check Box
- Edit Text
- Static Text
- Pop-up Menu
- Listbox

- Tooggle Button
- Table
- Axes
- Panel
- Button Group
- ActiveX Control

Všechny prvky jsou velice variabilní a jejich využití může být standardním i nestandardním způsobem. Záleží na schopnostech a potřebě uživatele. Při sázení těchto prvků na čistou plochu se automaticky generuje základní kód do skriptového souboru s příponou *.m.

Pro aplikaci tak bude využito *push button* pro potvrzování, *edit text* a *table* jako textový výstup pro výsledné hodnoty plynoucí z aplikování jedné z metod. Pro výčet metod je pak ideální *list box*, který bude také využit pro výběr souboru dat a jejich případné rozdělení u neparametrických metod. Původní myšlenka byla použít *pop-up menu*, které by funkčně splňovalo veškeré potřeby. Bylo od ní upuštěno z důvodu nepřehlednosti. Vlastnosti této komponenty (není prostorově náročná) byly využity při upřesňování výběru. Pro grafický výstup bylo zvoleno standardní řešení a to prvek *axes*.

6.3 Import dat

Předpokladem importu dat v této práci je určitý formát souboru, který obsahuje datový blok. Soubor je ve formátu *.xls případně *.xlsx. Nosičem dat je tedy tabulkový procesor (tzv. spreadsheet). Pro tabulkové procesory jsou v Matlabu speciální

příkazy. U verze R2014a je tento způsob importu dat však zbytečně složitý. Tedy jsou zde jednoduší způsoby importu dat.

R2014a je vybavena možností importu dat přes GUI Matlabu. Součástí této funkce je zároveň i generování skriptu pro import dat, který je možné pak dále použít ve vlastním GUI. Při generování je možné nastavit i datový typ hodnot v tabulce. V případě že tabulka je naplněna nejen číselnými hodnotami, je vhodné nastavit ji přímo *cell array*.

Různorodé formáty dat je nutné převádět téměř pro každou funkci. Styl *cell array* slouží jen pro zachování všech údajů. K tomuto procesu je využíváno jak převodních tak detekčních funkcí. U převodních se jedná o funkce *cell2mat()* a *num2str()*, které například v tomto pořadí převedou numerická data typu *cell array* na data typu *char*. V případě detekčních funkcí jsou tu funkce *is*()* jako třeba *iscellstr()*, která zjistí, zda jsou data v *cell array* typu *string* nebo nikoliv.

6.4 Statistické funkce

6.4.1 Kruskal-Wallis

Pro Kruskal-Wallisův test je připravena funkce *kruskalwallis()*, která pracuje s vloženými daty ve formě matice, skupinou a také je schopna rovnou údaje graficky zpracovat do podoby krabicového diagramu. Hodnoty, které jsou pak navráceny, jsou p-hodnota, ANOVA (*Analysis of variance*) tabulka (*cell array*) a nebo v struktuře, na kterou je možné navázat v dalších testech.[9]

Metoda pracuje v první řadě s jedním nebo s více kvantitativními vektory (soubory). Ale data, která je nutno zpracovat, jsou v cca devadesáti procentech nenumerická (kvalitativní). Některá z nich mají ordinální charakter, proto je smysluplný

jejich převod.

Tato data jsou tedy náležitě zpracována funkcí *unique*, jež primárně využívána k vyhledávání unikátních záznamů. Funkce obsahuje vícenásobné výstupy. V kombinaci s nastavením *sorted* je schopná přiřadit číselnou hodnotu, která plní substituční funkci pro původní data. Hodnota odpovídá pořadí po seřazení původních nenumerických dat.

Hodnoty na výstupu jsou vypisovány pomocí komponenty *static text*. Tento blok dat obsahuje informace o výsledku Kruskal-Wallisova testu a pravděpodobnost, na základě které uživatel sám rozhoduje zamítnutí/nezamítnutí H_0.

6.4.2 Simultánní porovnávání

U *post hoc* analýzy se používá funkce *multcompare()*, funkce pro vícenásobné porovnávání. Na jejím vstupu jsou nutná data z předešlé metody, v tomto případě se jedná o data pocházející z *kuskalwallis()*. Funkce jako taková ještě není přesným vyjádřením simultánního porovnávání, předpokladem je nastavení *CType* na *scheffe*. Pro přesnější výsledek byla přidána ještě možnost nastavení *post hoc* analýzy na *Turkeyho metodu*. Její použití se omezuje na data se symetrickým tříděním.

Data, která jsou během post hoc analýzy zpracována, jsou jedním z výstupů Kruskal-Wallisova testu. Jde o výstup *stats*, který obsahuje strukturovaná data (informace o porovnávaných skupinách - mediány, jejich četnosti atd.) daného testu. Tento výstup metody je primárně určen k dalšímu zpracování v případě neuspokojivého výsledku.

Výčet získaných hodnot je zprostředkován *uitable*. Obsahuje jednotlivé pravděpodobnosti vycházející z porovnávání jednotlivých skupin, rozdíly jejich průměrů a jejich .

6.4.3 χ^2 test dobré shody

Při použití χ^2 testu dobré shody je tu možná funkce *chi2gof()*, která pracuje s daty a se specifikací typu dat a nastavením jejich hodnoty. Na výstupu jsou pak hodnoty jako výsledná hypotéza, p-hodnota a struktura.

Data, která se u této funkce zpracovávají jsou nominálního typu. Proto je třeba před vložením do funkce jejich úprava. Pro modifikaci zvolených dat posloužila funkce *crosstable()*. Funkce na výstupu poskytuje dva parametry s daty, která jsou dále zpracována jako vstupní data pro hlavní funkci testování. První data jsou obsahují vzájemné četnosti položek, další jsou jejich popisky tzv. labely. Data s četnostmi jsou použita k zjištění pozorovaných (označení pro matlab je *Frequency*) a očekávaných (v matlabu *Expected*) četností, které se následně vkládají do hlavní funkce.

Data oznamující výsledek metody jsou informace o výsledku hypotézy, zda je zamítnuta či nikoliv. Pravděpodobnost a hladina významnosti, v rámci kterých se došlo předešlého statusu.

6.4.4 Randomizační test dobré shody

U této části se opakují všechny funkce z předchozího testování. Funkce náhodného výběru je zastoupena funkcí *randi()*. Počet opakování je nastaven na 1000 a provádí ji cyklus *for*. Výsledem je soubor se stejným počtem pravděpodobností kolik je opakování. Tato data jsou následně porovnána s původní pravděpodobnostní hodnotou.

Na výstupu je standardně informace o zamítnutí/nezamítnutí H_0 a jaké soubory byly využity. Celý test je prováděn s hladinou významnosti 0,5.

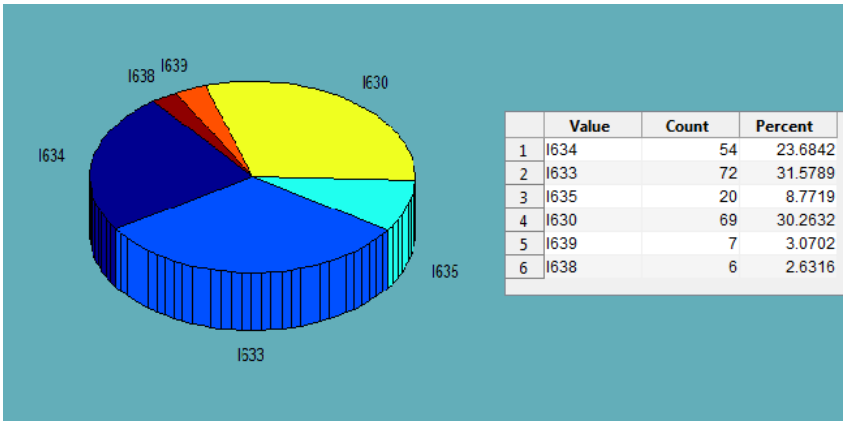
6.5 Statistické grafy

Matlab je vybaven funkcemi s grafickým výstupem. Je tak možné svá data vyjádřit i v jiné než textové/číselné formě. V prostředí jsou obsaženy základní grafická vyjádření.[17]

6.5.1 Kruhový graf

Kruhový graf je obsažen ve dvou formách, a to 2D a 3D. Jejich použití je naprosto totožné proto nemá výběr vliv na výsledný dojem. Samotné funkce jsou *pie()* a *pie3()*. Na vstupu jsou buď kvantitativní data nebo kvantitativní a kvalitativní data(ty pak slouží primárně jen k popisu grafu).

Výsledek úpravy v matlabu je na obrázku 6.1 za použití procentuálního zpracování dat. Zobrazuje zastoupení jednotlivých ICD v rámci celého souboru dat, kde I634 a I633 dohromady udávají nadpoloviční většinu. Tento fakt odpovídá i naměřeným údajům v tabulce.



	Value	Count	Percent
1	I634	54	23.6842
2	I633	72	31.5789
3	I635	20	8.7719
4	I630	69	30.2632
5	I639	7	3.0702
6	I638	6	2.6316

Obrázek 6.1: Ukázka 3D grafu výstupu při výpočtu procentuálního zastoupení hodnot v souboru dat ICD

ICD a jeho značení je globálním standardem k diagnostice mozkových příhod. I63.* je specifický kód pro mozkový infarkt a jeho příčiny (viz. 6.2). Proto lze usuzovat, že nejčastějšími příčinami mozkového infarktu jsou trombóza nebo embólie mozkových tepen.

. 0	Mozkový infarkt způsobený trombózou přívodných mozkových tepen
. 1	Mozkový infarkt způsobený embolií přívodných mozkových tepen
. 2	Mozkový infarkt způsobený neurčenou okluzí nebo stenózou přívodných mozkových tepen
. 3	Mozkový infarkt způsobený trombózou mozkových tepen
. 4	Mozkový infarkt způsobený embolií mozkových tepen
. 5	Mozkový infarkt způsobený neurčenou okluzí nebo stenózou mozkových tepen
. 6	Mozkový infarkt způsobený mozkovou žilní trombózou, nehnisavou
. 8	Jiný mozkový infarkt
. 9	Mozkový infarkt NS

Obrázek 6.2: Kód I63.* vyjadřující příčinu mozkového infarktu.

6.5.2 Bodový graf

Bodový graf, který je znám také jako korelační diagram, je v Matlabu jako funkce *scatter*. Pro další práci s ním je třeba jistá úprava. Nabízí možnosti vykreslování daných bodů od koleček, po znaménka matematických operací nebo diamanty. Pro vyznačení určitého mezníku (průměr, medián) byla použita funkce *refline()*, do které je nutné zadat jen souřadnice pro vykreslení linky (výsledek viz. 5.1).

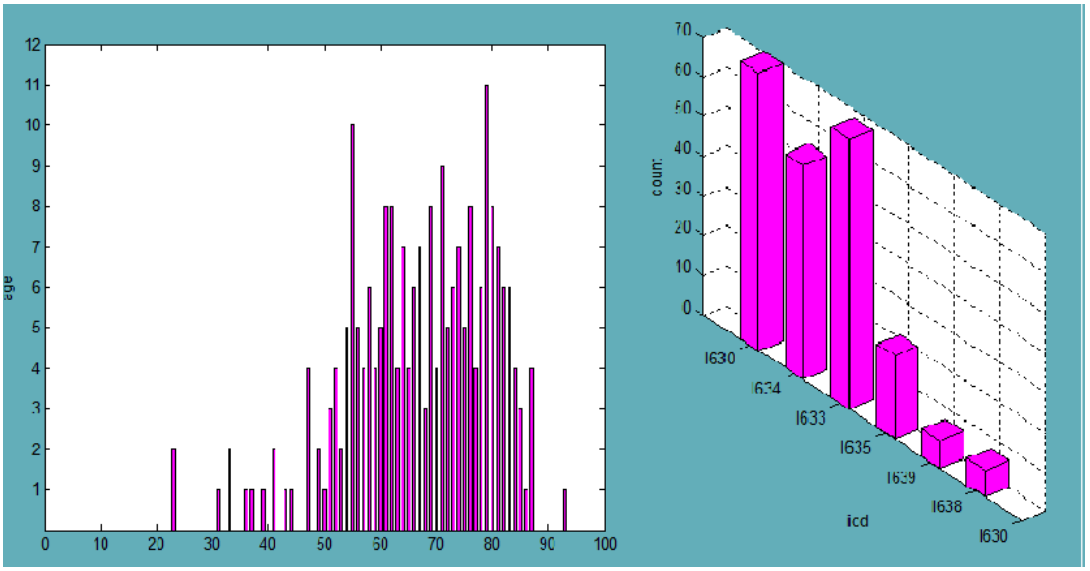
6.5.3 Histogram

Histogram je další velice variabilní možností grafického zpracování dat. V Matlabu je pod funkcí *hist()*. Je v něm možné nastavení, v kolika sloupcích budou daná data vykreslena. To je vlastnost, se kterou je docíleno přesnějšího vyjádření dat.[11].

Pro vhodnější zpracování byla i přesto zvolena funkce *bar()*, která je schopna

pracovat s kategoriemi. Data na vstupu jsou pouze kvantitativního typu. Pro zpracování dat na vhodný vstupní formát byla použita funkce *tabular()*, která ze souboru dat sumarizuje procentuální zastoupení jednotlivých položek v souboru. Je to taky hlavní funkce analytického zpracování dat. Způsob převodu dat se v případě χ^2 liší pouze v počtu zpracovávaných kategorií. Hlavní kategorie, se kterou se v tomto případě porovnává, jsou celková data, a to se zvolenou kategorií. Tuto volbu provádí uživatel.

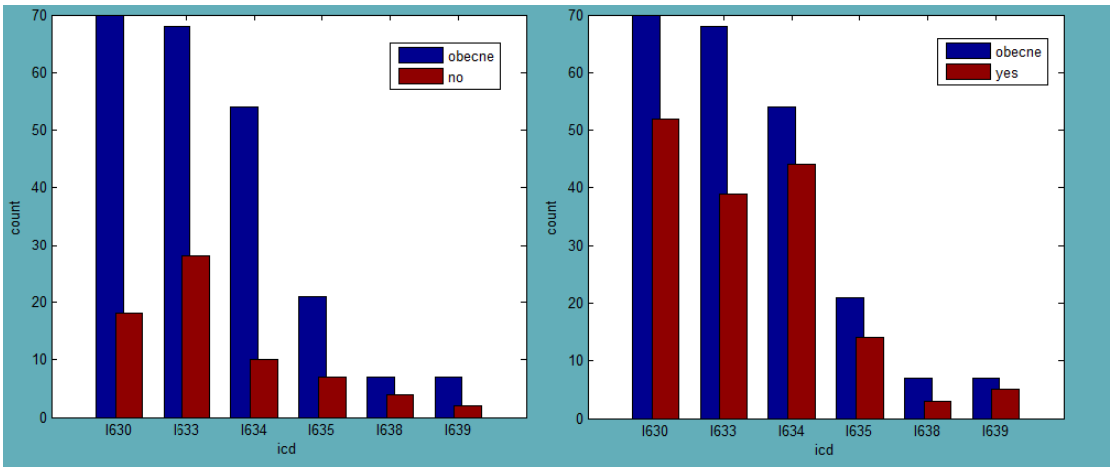
U analýzy dat byla dokonce pro vizuální efektivitu zvolena 3D verze funkce *bar3()*. Příčinou byla vágnost grafu vzhledem k jednomu pozorovanému datovému souboru ve většině případů kvalitativního charakteru. Pro případ mnohočetného kvantitativního souboru byla aplikace nastavena pro přehlednost množství sloupců zpět na 2D verzi. Tohle řešení se zdálo nejlepší vzhledem k možnosti analyzovat i data obsahující informace o věku pacienta (porovnání obou verzí viz obr. 6.3).



Obrázek 6.3: Porovnání 2D grafu (věk) a 3D grafu (ICD z předešlého příkladu)

V případě χ^2 testu byla zvolena 2D verze (viz obr.6.4). Je to výstup testu, kdy zjišťujeme, zda má hypertenze (vysoký tlak) vliv na rozložení četností ICD skupin.

První graf vyjadřující data, která na hladině významnosti 0,5 není zamítnuta, ale pouze na ní. Při hladině významnosti 0.1 už by byla zamítnuta. Totéž naznačuje graf vlevo, na kterém je větší rozdíl u prvního sloupce. Pokud by jsme si představily vrcholy daných četností jako křivku, tento menší rozdíl by byl očividnější.



Obrázek 6.4: Porovnání 2D grafu (věk) a 3D grafu (ICD z předešlého příkladu)

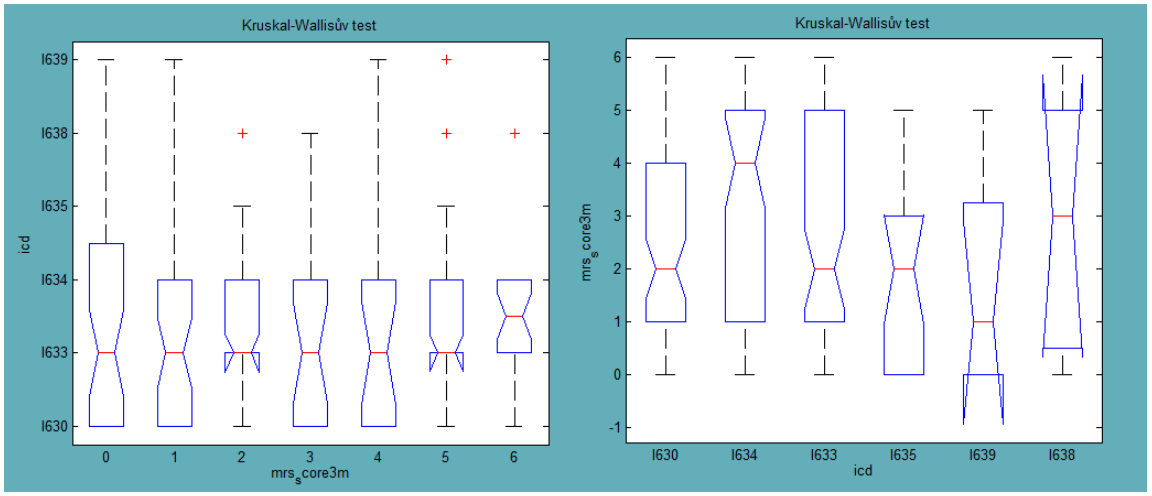
U druhé skupiny na obr. 6.4 vpravo je větší podobnost mezi četnostmi, pokud se pro srovnání použijí obecné nekategorizované četnosti souboru dat. Ty naznačují distribuční křivku, a tedy i jak by měli vypadat očekávané četnosti. Data (kategorie lidí s hypertenzí) jsou z pohledu na graf více podobna jejich předpokladu. To potvrzuje i fakt, že nulová hypotéza nebyla zamítnuta na hladině 0.01. Ovšem, vezme-li se v potaz skutečnost, že u některých četností se suma minima z celku dostává sotva na 10, bylo by vhodnější užití i jiné srovnávací metody. Data mohou být zkreslena menší objemností dat.

6.5.4 Krabicový diagram

Krabicový diagram je zde pod funkcí *boxplot)*. Výstup je rozdílný na základě vstupních dat (ne jen v případě různých hodnot nacházejících se v souboru dat), a to

v množství výstupních diagramů. V případě souboru dat, který je maticí, je na výstupu tolik krabicových diagramů, kolik je sloupců v matici. Při datech o jedné sledované proměnné tedy o jednom vektoru je na výstupu jen jeden diagram. [10]

V případě prvního grafu (obr. 6.5) nelze odmítnout nulovou hypotézu, že je rozložení ICD u mRS skóre stejné, a to na hladině významnosti 0,5. Tuto skutečnost lze usuzovat i z levého grafu. Z druhého grafu je evidentní, že rozložení mRS skóre v rámci ICD skupin je více různorodé, a proto je možné zamítnou H_0 na hladině významnosti 0,1. Ovšem na hladině významnosti 0,01 ji nezamítáme. V případě přesnějšího výsledku, tedy které skupiny se liší natolik, že snižují celkovou pravděpodobnost, by bylo vhodné použít *post hoc* analýzu.



Obrázek 6.5: Ukázka dvou krabicových diagramu u Kruskal-Wallisova testu

7 Závěr

V úvodu práce byla prezentována problematika vizualizace medicínských dat, jež objasnila důvody, které vedly k zadání této práce. Zadání bylo: jednoduše vytvořit aplikaci, která by usnadnila práci analytikům a zároveň na ni mohli navazovat dalšími podpůrnými programy.

Během této práce byla stěžejní manipulace s daty. Pro základní znalosti obecného získávání dat pro statistické účely bylo nutné se seznámit s jednotlivými fázemi statistického šetření. Poskytnutá data byla původem z registru SITS. Z toho důvodu byl popsán význam tohoto registru, způsob zadávání těchto dat do systému, průběh získávání dat a samotný formulář. Formulář byl základním zdrojem informací, protože bylo možné na základě jeho obsahu lépe identifikovat poskytnutá data.

Součástí aplikace měla být možnost výpočtu a nejen grafického výstupu, proto bylo nezbytné se alespoň v principu seznámit s vhodnými metodami. Byly vybrány na základě předchozí práce, jež se zabývala vhodnými metodami v oblasti zpracování lékařských dat. V rámci implementace bylo náročnější částí pochopení těchto metod, a to celý proces zpracování dat do podoby, kterou by daná funkce byla schopna otestovat. A to i přesto, že předpokladem jsou čistá data. Hlavním úskalím proto byl obecně Matlab, který sice poskytuje velké množství funkcí, ale ne všechny při práci lze objevit ve vhodném časovém úseku. Pro začátečníka velice nevhodné prostředí pro větší časově vymezený projekt.

Program obecně splňuje všechny základní požadavky, které byly zadány v rámci této práce. V případě osvědčení jeho funkčnosti v rámci cílové skupiny je další vývoj s větší ovladatelností uživatele určitě na místě. Např. volba vlastního grafu v případě pokročilejšího uživatele (v programu pouze u analýzy dat), možnost jisté modifikace samotných metod, která by ve výsledku znamenala více prostoru pro statistické

testování dat. To jsou ale poznatky, kterých se mi dostalo v rámci nynějších znalostí, proto nebylo na začátku práce možné je definovat či určit jako cíle.

Seznam obrázků

Literatura

[1] HEJNA, Miroslav. *Statistické zpracování lékařských dat.* Plzeň, 2012. Dostupné z: `https://otik.uk.zcu.cz/bitstream/handle/11025/3044/Hejna_DP.pdf`. Diplomová práce. Západočeská univerzita, Fakulta aplikovaných věd, Katedra informatiky a výpočetní techniky.

[2] ANDĚL, Jiří. Statistické metody. 4., upr. vyd. Praha: Matfyzpress, 2007, 299 s. ISBN 978-80-7378-003-6.

[3] HEBÁK, Petr, Jiří HUSTOPECKÝ, Eva JAROŠOVÁ a Ivana MALÁ. *Vícerozměrné statistické metody.* Vyd. 1. Praha: Informatorium, 2004-2005, 3 sv. ISBN 80-7333-025-3.

[4] ANTOCH, Jaromír a Dana VORLÍČKOVÁ. *Vybrané metody statistické analýzy dat.* 1. vyd. Praha: Academia, 1992, 279 s. ISBN 8020002049.

[5] HENDL, Jan. *Přehled statistických metod zpracování dat: analýza a metaanalýza dat.* Vyd. 1. Praha: Portál, 2004, 583 s. ISBN 8071788201.

[6] Haandbook of Biological Statistics: Randomization test of goodness-of-fit [online]. 2009. [cit. 2015-06-11]. Dostupné z: `http://udel.edu/~mcdonald/statrand.html`

[7] Chi-square test of goodness-of-fit: Haandbook of Biological Statistics [online]. 2014. [cit. 2015-06-09]. Dostupné z: http://www.biostathandbook.com/chigof.html

[8] ZVÁROVÁ, J. *Základy statistiky pro biomedicínské obory.* Vyd. 1. Praha: Karolinum, 2002, 218 s. ISBN 80-718-4786-0. Dostupné z: http://new.euromise.org/czech/tajne/ucebnice/html/html/statist.html

[9] Kruskal-Wallis test. *MathWotks: MATLAB and Simulink for Technical Commputing* [online]. 1994, 2015 [cit. 2015-06-09]. Dostupné z: http://www.mathworks.com/help/stats/kruskalwallis.html

[10] Box plot. *MathWotks: MATLAB and Simulink for Technical Commputing* [online]. 1994, 2015 [cit. 2015-06-09]. Dostupné z: http://www.mathworks.com/help/stats/boxplot.html

[11] Histogram plot. *MathWorks: MATLAB and Simulink for Technical Commputing* [online]. 1994, 2015 [cit. 2015-06-10]. Dostupné z: http://www.mathworks.com/help/matlab/ref/hist.html

[12] MUDr: Lékařské klasifikace • Online kalkulačky • Skóre • Tabulky • MKN v.2 [online]. 2008-2009 [cit. 2014-11-16]. Dostupné z: http://www.mudr.org

[13] Diagnostický a terapeutický manuál cévních onemocnění mozku [online]. 2009, 2014 [cit. 2014-11-23]. Dostupné z: http://cmp-manual.wbs.cz/

[14] ABZ.cz: slovník cizích slov - on-line hledání [online]. 2005, 2014 [cit. 2014-11-27]. Dostupné z: http://slovnik-cizich-slov.abz.cz/

[15] BĚLÁŠKOVÁ, Silvie a Lenka BLAŽKOVÁ. *Moderní analýza medicínských dat.* [online]. 2010 [cit. 2014-11-07]. Dostupné z: http://www.systemonline.cz/it-pro-verejny-sektor-a-zdravotnictvi/moderni-analyza-medicinskych-dat.htm

[16] MATLAB documentation *MathWorks: MATLAB and Simulink for Technical Commputing* [online]. 1994, 2015 [cit. 2015-03-20]. Dostupné z: http://www.mathworks.com/help/matlab/index.html

[17] Graphics *MathWorks: MATLAB and Simulink for Technical Commputing* [online]. 1994, 2015 [cit. 2015-03-24]. Dostupné z: http://www.mathworks.com/help/matlab/graphics.html

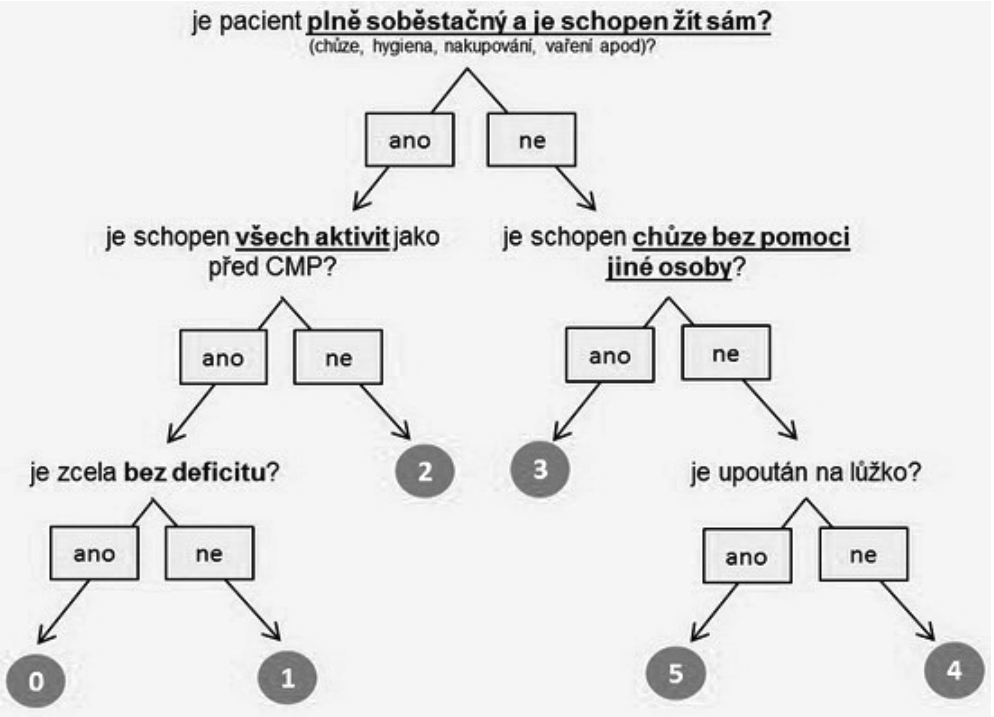
[18] BUI, Alex A.T. a William HSU. *Medical imaging informatics*. New York: Springer, 2010. ISBN 1441903852-. Dostupné z: http://www.mii.ucla.edu/~willhsu/pubs/bui.mii.ch4.pdf

[19] Anders Ynnerman: Visualizing the medical data explosion *TED: Ideas worth spreading*[online]. 2010 [cit. 2015-06-16]. Dostupné z: http://www.ted.com/talks/anders_ynnerman_visualizing_the_medical_data_explosion/transcript?language=en

[20] SZOLOVITS, Peter. *Medical Informatics Computer: Computer Applications in Health Care* [online]. 1997 [cit. 2015-06-16]. Dostupné z: http://groups.csail.mit.edu/medg/courses/6872/96/notes/Tsien/

Přílohy

A Vyhodnocování mRS



B Stupnice vyšetřovaných bodů NIHSS

Level of Consciusness	0	plně při vědomí, spolupracující
	1	spavý, po mírné stimulaci poslechne, odpoví
	2	opakovaná stimulace k pozornosti, sopor
	3	koma (reflexní či žádná odpověď)
LOC Questions	0	obě odpovědi zcela správně
	1	jedna správně, těžká dysarthrie či jiná bariéra (OTI)
	2	obě špatně, afázie, kóma
LOC Commands	0	oba úkoly správně
	1	jeden úkol správně
	2	žádný správně, kóma
Best Gaze	0	bez patologie
	1	izol. paresa okohybného nervu, deviace či pohledová paresa potlačitelná OC manévry
	2	nepotlačitelná deviace či pohledová paresa
Visual	0	bez postižení
	1	částečná hemianopsie, fenomén extinkce
	2	kompletní hemianopsie
	3	oboustranná hemianopsie (slepota,včetně kortikální slepoty)
Facial Palsy	0	symetrický pohyb, bez postižení
	1	lehká paresa (např. asymetrie NL rýhy)
	2	úplná nebo částečná paréza dolní větve (centrální paresa)
	3	kompletní (perif.) paréza uni- či bilaterální, koma
Motor Arm/Leg	0	bez kolísání
	1	kolísání nebo pokles, bez úplného pádu na podložku

	2	určitý pohyb proti gravitaci, neudrží nad podložkou
	3	pohyb po podložce
	4	plegie, bez pohybu, koma (pro všechny konč.)
	9	amputace, ankylóza aj. příčiny patolog. nálezu nesouvisející s příhodou.
Limb Ataxia	0	nepřítomna, nebo jen důsledek paresy. Koma.
	1	na jedné končetině
	2	přítomna na více končetinách
	9	amputace, ankylóza aj.
Sensory	0	bez poruchy čití
	1	lehká a střední porucha sense (hypestezie,hypalgezie)
	2	těžká porucha sense až anestezie uni, či bilat. Koma.
Best Language	0	bez afázie
	1	lehčí fatická porucha, lze porozumět
	2	těžká fatická porucha
	3	globální afázie, mutismus, kóma
Dysarthria	0	nepřítomna
	1	setřelá řeč, je mu rozumět
	2	výrazně setřelá výslovnost, není rozumět, mutismus, kóma
	9	intubace, jiná bariéra
Extinction and Inattention - Neglect	0	nepřítomen
	1	neglektuje 1 kvalitu, anosognoze
	2	neglektuje více jak 1 kvalitu, kóma.

C Vzorce statistických metod

Krukal-Walllisuv test

$$H_{KW} = \frac{12}{n(n-1)} \sum_{i=1}^{m} \frac{R_i^2}{n_i} - 3n(n+1)$$

n_i	četnost skupiny
R_i	suma hodnot skupiny

χ^2-test dobré shody

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)}{E_i} \sim \chi^2_{k-1}$$

O_i	pozorované četnost
R_i	očekávané četnosti

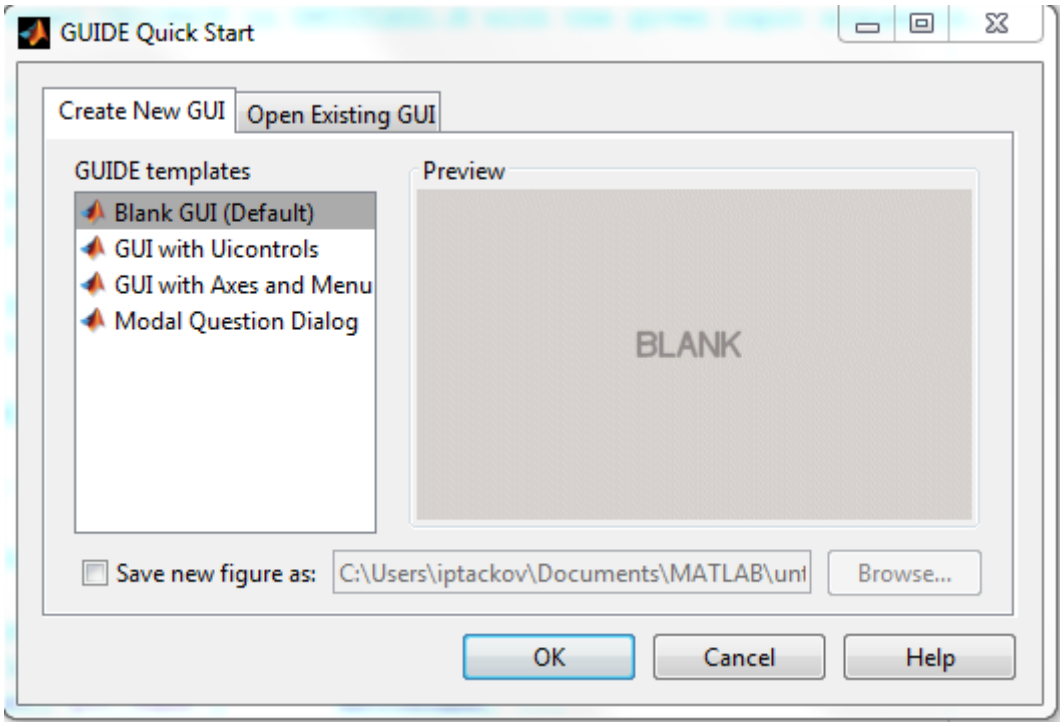
Simultánní porovnání

Počet porovnávání, které musíme provézt: $\frac{m(m-1)}{2}$

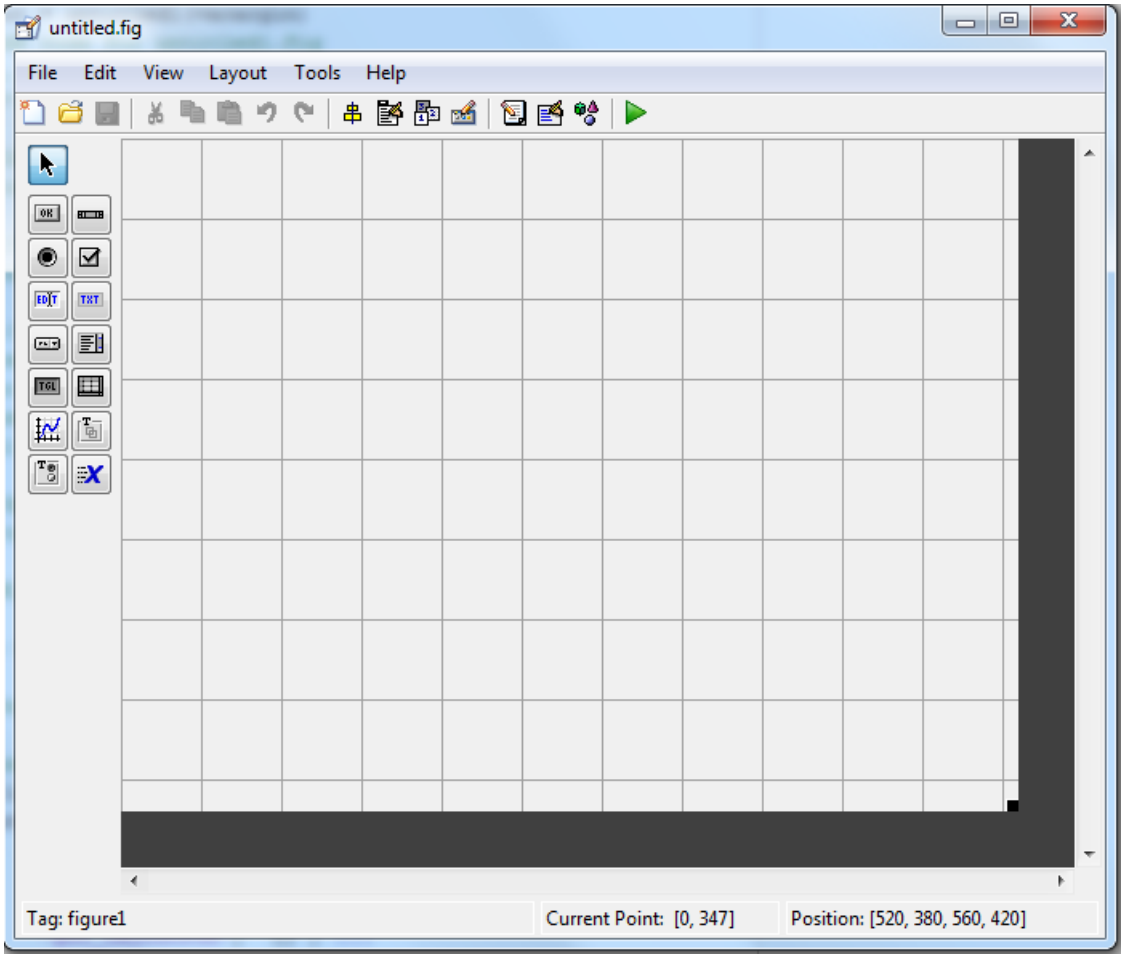
$$|\overline{R_i} - \overline{R_j}| \geq z_{\alpha/m(m-1)} \sqrt{\frac{n(n+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

$\overline{R_i}$	průměrná pořadí ve skupině (i,j)
$\overline{R_j}$	
m	počet skupin
n	suma počtu dat
n_i	počet dat v rámci skupiny i

D GUIDE



Obrázek D.1: Úvodni okno



Obrázek D.2: Pracovní prostředí

E Uživatelská příručka

Import data

Program pracuje s údaji, které mají minimální nárok na strukturu svých dat. To znamená, že údaje obsahují popisek těchto dat (pohlaví,věk,..). V případě, že se jedná o data bez označení, program nijak nereaguje a bere první řádek jako popisek naměřených dat. Ve výsledku nebudou následně do testování nijak zapojována.

Samotný import dat: *Open File*$\longrightarrow$ vyberte soubor (např. *data.xls*)$\longrightarrow$ *Otevřít.*

Výběr metody

Samotný výběr metody je na uživateli: *Analýza dat* a *Neparametrické testy.*

Analýza dat se týká základních statistických hodnot a grafů.

- **Procenta:** *Použití* pouze pro data, která v sobě mají kategorie nebo se aspoň opakují. *Graf* je možné zvolit typu koláč nebo histogram.
- **Aritmetický průměr a Medián:** *Použití* pouze u kvantitativních dat. *Graficky* zpracováno bodovým grafem s vyznačením průměru (mediánu).
- **Medián a Modus:** *Použití* pro všechny typy dat. U mediánu, v případě nekvantitativních dat, může docházet k nerozhodnému výsledku, který zapříčiní vypsaní položky, která je v seznamu dále. Tento fakt může zapříčinit nepatrné zkreslení výsledku. Medián je doplněn bodovým *grafem*, modus histogramem.

Kruskal-Wallisův test

Kruskal-Wallisův test je neparametrickou verzí metody analýzy rozptylu jednoduchého třídění. Tento způsob testování dat je využíván pokud jsou výběry z rozdělení, které je značně odlišné od normálního rozdělení. Je aplikován při testování shody zvoleného pravděpodobnostního rozdělení srovnávaných skupin. Data, s kterými pracuje, nevycházejí z normálního rozdělení, a jsou na sobě nezávislé. Jeden z předpokladů použití této metody je přítomnost dat, které obsahují dva a více naměřených údajů.

V principu jsou data rozdělena do skupin (např. žena, muž). Je zjištěn stupeň volnosti a zvolena kritická hodnota (χ^2-rozdělení). Data skupin jsou seřazena dle velikosti napříč skupinami, a následně je jim přiřazena hodnota pořadí (dále jen rank). V případě shodných naměřených hodnot se přechází k přiřazení průměru z pořadí. Data jsou nadále zpět rozřazena do svých skupin, ale reprezentována svojí rank hodnotou. Skupiny jsou pak sumarizovány a je určena četnost jejich dat. Po dosazení do vzorce je výsledek porovnán s hladinou významnosti. H_0 je pak zamítnuta nebo přijata na základě tohoto porovnání.

- **Použití:** Spojitá data v kombinaci s kategoriálními.
- **Hodnoty:** KW je výsledek daného testu. P-value je pravděpodobnost, že se jedná o stejnou distribuci. Zamítnutí nulové hypotézy závisí na pozorovateli.
- **Graf:** Krabicový diagram.

Simultánní porovnávání

Toto porovnávání je zároveň také post hoc analýzou, která se používá v případě zamítnutí H_o u předešlé metody. Analýzu je možné provádět, aniž by tomu předcházela

specifikace srovnání dat. Princip metody je postaven na porovnávání mediánů statisticky usuzovaných skupin. Pro výslednou hodnotu je potřeba porovnat navzájem všechny skupiny.

- **Použití:** Další krok v případě zamítnutí nulové hypotézy u KW testu. Je možné nastavení dvou metod. Scheffe a HSD, která je ale jen pro symetricky setříděná data.
- **Hodnoty:** Výsledné hodnoty jsou navzájem porovnávané skupiny: jejich pravděpodobnost, že jsou ze stejné distribuce a také rozdíly jejich průměrů. Vše v rámci výstupní tabulky.
- **Graf:** Krabicový diagram.

χ^2-test dobré shody

Tento test je neparametrickou metodou, která je používána v případě na sobě nezávislých dat. Základ této metody je v ověření shody usuzovaných četností s četnostmi, které byly vypozorovány. Data je možné rozdělit do kategorií nebo na intervaly. Záleží na typu dat, jestli jsou kategoriálního typu či intervalového typu.

V praxi je porovnávána nominální proměnná s dvěma a více hodnotami. Porovnávány jsou pak pozorované hodnoty s očekávanými hodnotami, které je možné vypočítat prostřednictvím nějakého teoretického očekávání (Např. 1:1, kdyby šlo o pohlaví). Pro přesnější výsledky se u této metody doporučuje větší množství dat. V opačném případě mohou být výsledky nepřesné. Test je aplikovatelný na již zmíněné kategoriální údaje. Tj. například pohlaví či typ údaje, který posouvá jedince do jisté kategorie.

- **Použití:** Data nominální, ordinální a diskrétní. Větší počet dat.

- **Hodnoty:** Rozhodnutí o hypotéze, pravděpodobnost, hladina významnosti a použité datové soubory.
- **Graf:** Sloupcový graf (histogram).

Randomizační test dobré shody

Je používán, pokud je nominální proměnná se třemi a více hodnotami a pro χ^2 test dobré shody je vzorek dat příliš malý. Test je prováděn v případě, že z jednoho testu dobré shody není možné pro malého množství očekávaných četností dojít správného výsledku. Aproximační vztah tak malého vzorku dat není přesný. Základem randomizační verze tohoto testu je pak opakované měření při ještě menším vzorku dat, kdy počítáme vždy jen s náhodně vybraným vzorkem dat z celého vzorku. Přitom je vždy dodržen poměr naměřených dat.

- **Použití:** Data nominální, ordinální a diskrétní. V případě malého vzorku dat.
- **Hodnoty:** Rozhodnutí o hypotéze, hladina významnosti (zde vždy 0,05) a použité datové soubory.
- **Graf:** Sloupcový graf (histogram).