



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická

Úloha zpracování dat

Projekt pro předmět Lékařská informatika

Studijní program: Biomedicínské inženýrství a informatika
Studijní obor: Biomedicínské inženýrství

Martin Koretz, Jan Vlasák

Praha, březen '13

Obsah

1. Zadání.....	3
2. Zpracování dat.....	4
2.1 Analýza dat jako celku a rozdělení do skupin	4
2.2 Rozdělení do skupin	5
2.2.1 Věk	5
2.2.2 Krevní tlak	5
2.2.3 BMI	5
2.3 Statistiky	6
2.4 Rizika a nepřesnosti.....	6
2.5 Analýza v závislosti na věku	6
2.6 Analýza v závislosti na krevním tlaku.....	8
2.7 Analýza v závislosti na body mass indexu (BMI).....	10
2.8 Analýza v závislosti na jiné medikaci	12
2.9 Navržení rozhodovacího stromu.....	14
3. Závěr	16
3.1 Možná vylepšení a návaznost	17
4. Použitá literatura	18

1. Zadání

Na zadaných datech proveďte analýzu, která bude obsahovat rozřídění dat do skupin v závislosti na typu podávaného léku, na výsledku léčby primární choroby (primární pro stanovení účinnosti léku), přítomnosti jiné medikace, věku, průměrného krevního tlaku [1] a BMI [2] (nazvěme tyto 4 proměnné faktory) a bude zohledňovat výskyt sekundárních chorob před a po léčbě. Analýza bude obsahovat sledované pacienty, kteří se jedním či druhým lékem vyléčili, tak ale i ty, kteří se nevléčili. Zároveň by analýza měla ve všech skupinách (rozdělení podle faktorů) obsahovat statistické testování pro dokázání (vyvrácení), zda je některý z léků účinnější, nebo zda mohly rozdíly mezi léky vzniknout náhodou (vždy na hladině významnosti 0.05). Diskutujte případný vliv věku, průměrného tlaku, BMI či přítomnosti jiné medikace na výsledek léčby (ve všech variantách).

Data jsou přiložena k zadání.

2. Zpracování dat

Data byla zpracována jako celek a poté rozdělena do skupin podle proměnných faktorů a každá skupina byla analyzována zvlášť.

Sekundární choroby jsou pro přehlednost nadále označovány jako alfa (první sekundární choroba) a beta (druhá sekundární choroba).

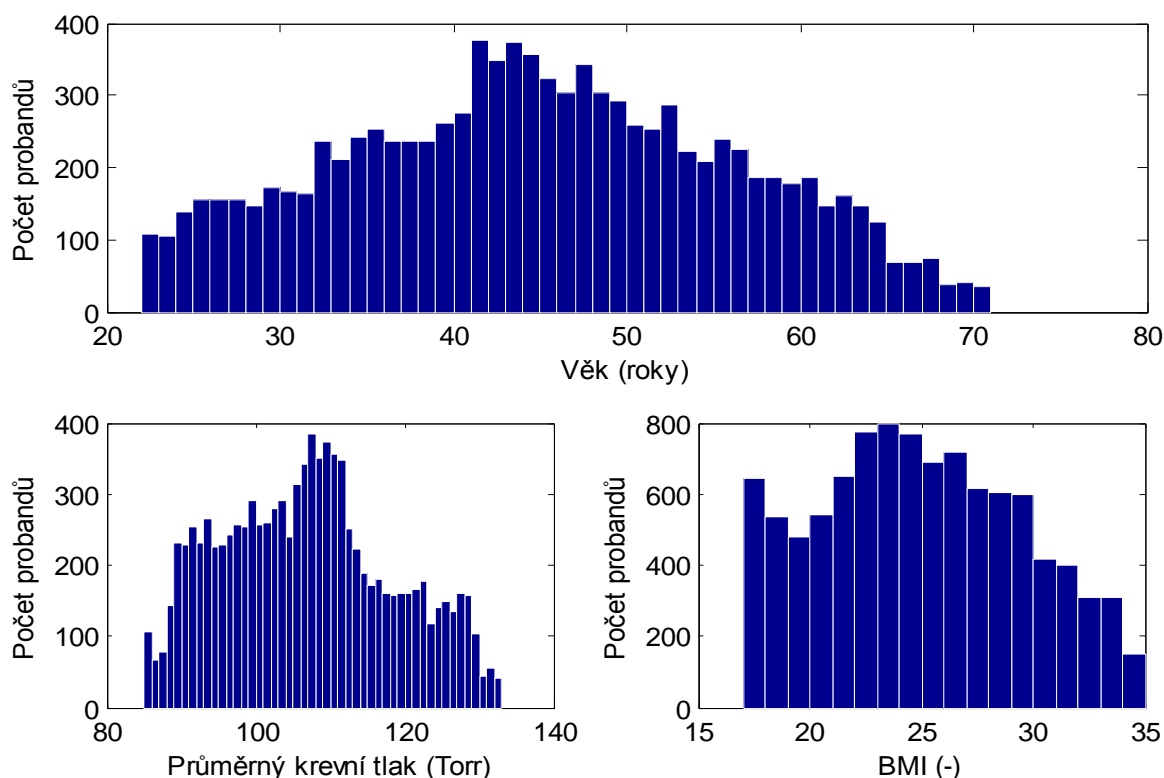
2.1 Analýza dat jako celku a rozdělení do skupin

Data byla v první fázi analyzována jako celek bez dalšího členění. Tato analýza má informativní charakter a jejím cílem je analyzovat strukturu a složení zadaných dat pro určení dalšího postupu. Celkový počet probandů byl 10 000.

První analýzou bylo vyšetření některých charakteristik vyjádřitelných jednoduchými prohlášeními platnými pro realizovaný výběr:

- Celkový počet probandů byl 10 000.
- Primární chorobou před léčbou trpělo 10 000 probandů.
- Jinou medikaci užívalo 3 315 probandů.
- Lékem 1 bylo léčeno 4834 probandů.
- Lékem 2 bylo léčeno 5166 probandů.
- Vyléčeno bylo 5 978 probandů. Účinnost léčby je 0,5978 %.
- Účinnost léčby lékem 1 u primární choroby je 0,46276 %.
- Účinnost léčby lékem 2 u primární choroby je 0,72416 %.
- Účinnost léčby alfa choroby je 0,068512 %.
- Účinnost léčby beta choroby je 0,2114 %.
- Účinnost léčby alfa choroby lekem 1 je 0,3244 %.
- Účinnost léčby alfa choroby lekem 2 je -0,3538 %.
- Účinnost léčby beta choroby lekem 1 je 0,20343 %.
- Účinnost léčby beta choroby lekem 2 je 0,21985 %.

Jako součást další analýzy bylo zobrazeno zastoupení jednotlivých proměnných faktorů u probandů formou histogramů (Obr. 1).



Obr. 1: Zastoupení jednotlivých proměnných faktorů u probandů

2.2 Rozdělení do skupin

Na základě histogramů (Obr. 1) bylo s přihlédnutím k zaběhnutým klasifikačním systémům zvoleno rozdělení do jednotlivých skupin.

2.2.1 Věk

Skupiny věku byly zvoleny po desítkách let (do 30 let, 31–40 let, ...), jelikož v těchto rozmezích jsou počty probandů srovnatelné, jejich celkový počet v každé ze skupin je poměrně konzistentní až na poslední a předposlední skupinu, které byly sloučeny do skupiny nad 60 let.

2.2.2 Krevní tlak

Skupiny krevního tlaku byly zvoleny s ohledem na úzus zavedený British hypertension society, WHO a ESH [1], [2]. Daná rozmezí (107,5–114 Torr, 114–129 Torr a extrémní) byla dopočítána pro průměrný tlak (nikoliv střední tlak) a nejsou v rozporu s rozložením hodnot tlaku ve zkoumané skupině.

2.2.3 BMI

Skupiny BMI byly zvoleny dle norem WHO [3], které vyhovují zadaným datům. Skupiny jsou do 18,5; 18,5–25; 25–30; nad 30. Skupina BMI do 18,5 má oproti ostatním skupinám malé zastoupení probandů a proto hrozí, že bude nutné tuto nedostatečnost kompenzovat statistickými metodami.

2.3 Statistiky

Statistiky byly prováděny testem chí-kvadrát nezávislosti v kontingenční tabulce [4]. Jako nulová hypotéza H_0 byla stanovena nezávislost znaků a a b . Tyto znaky byly v průběhu měněny dle zadání a jejich znění je k vidění v příloženém m-souboru. Např. v prvním případě jsme použili znaky *lidé* a *použitá medicína*. Z tohoto případu vzešli všechny hypotézy pravdivé, a proto nemůžeme zamítnout nezávislost znaků a tedy oba léky odpovídají normálnímu rozdělení a nelze určit, zda je některý z nich lepší. Když vyjde H_0 záporná, předpokládáme, že jeden z léků je účinnější a na základě účinnosti určíme který z nich.

Hladina významnosti byla při všech testech 5 %.

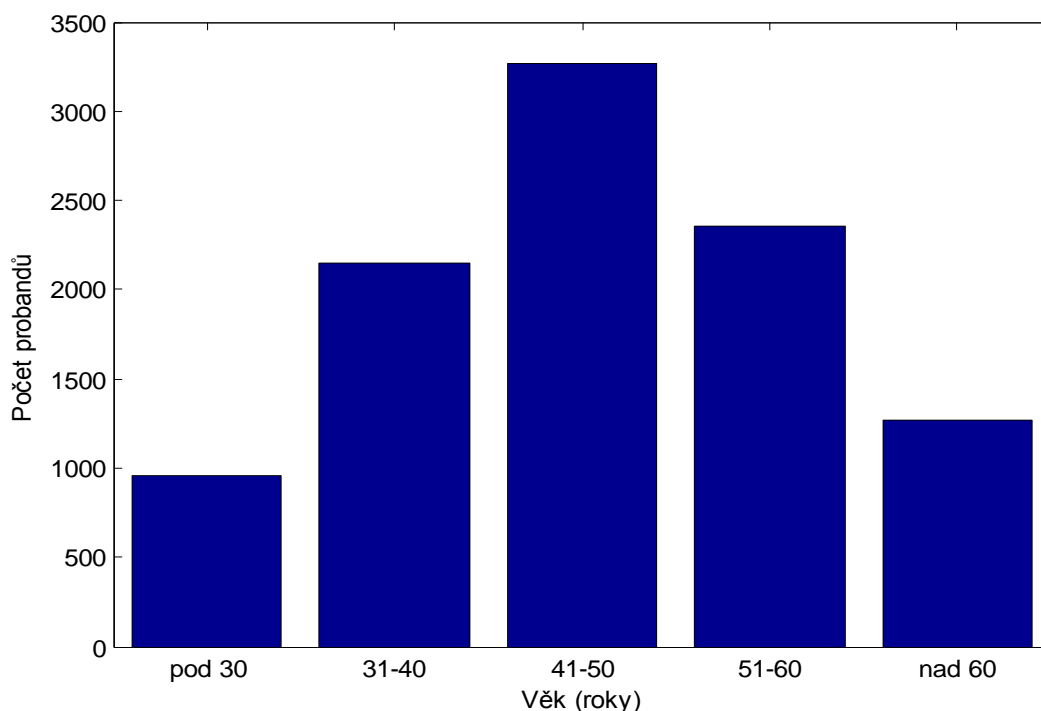
2.4 Rizika a nepřesnosti

Pro zjednodušení budu dále uvádět, že lék 1 (nebo 2) je účinnější pro danou chorobu. Toto označení je nepřesné a proto na něj upozorňuji a žádám čtenáře, aby tento fakt přijali a zařídili se dle vlastní potřeby.

Dále vědomě použiji pouze grafické znázornění dat, jelikož přesná číselná reprezentace může být nalezena v příloženém m-souboru a díky použití statistického testu jsou závěry z daných dat jasné. Tabulkové uvedení dat by bylo redundantní.

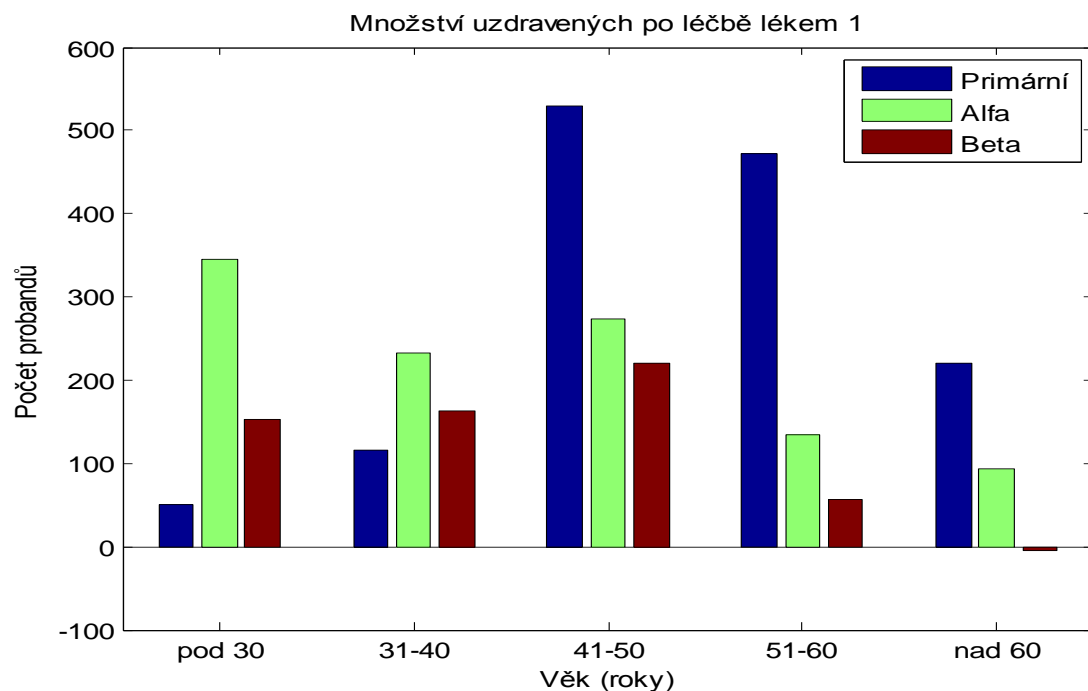
2.5 Analýza v závislosti na věku

Rozdělení pro danou skupinu je patrné na Obr. 2. Rozdělení je dle našich očekávání normální.

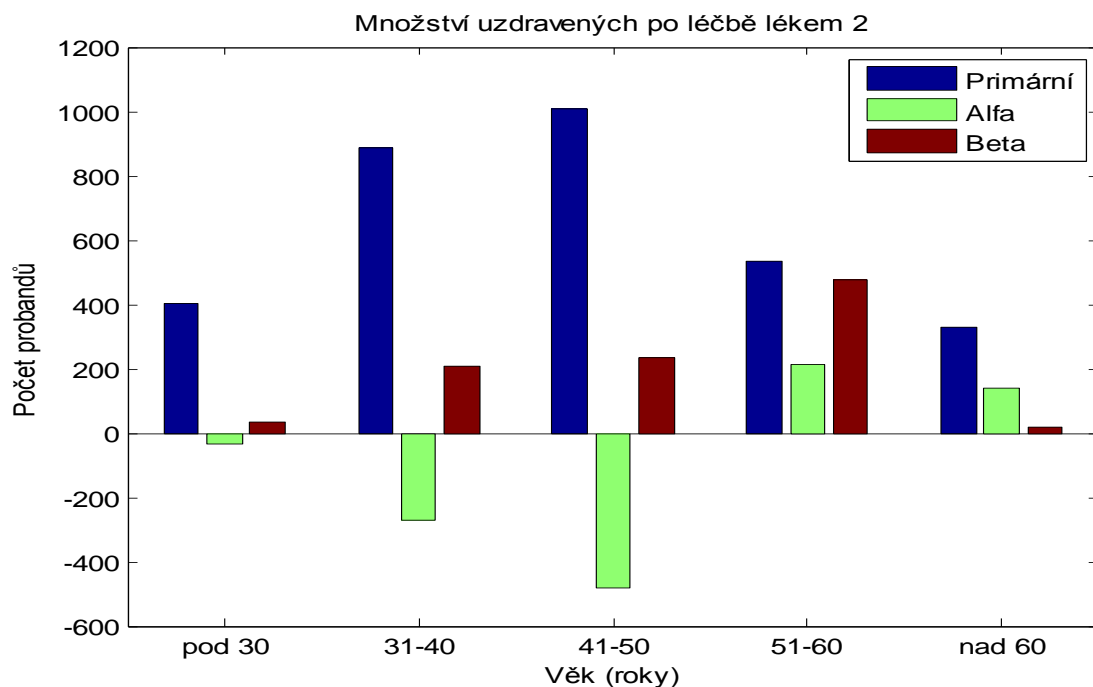


Obr. 2: Zastoupení jednotlivých věkových skupin dle zvoleného rozdělení

Množství probandů uzdravených po léčbě lékem 1 a těch, kteří byli uzdravení po léčbě lékem 2, je zobrazeno na Obr. 3 a Obr. 4.



Obr. 3: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 1 v závislosti na věkové skupině



Obr. 4: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 2 v závislosti na věkové skupině

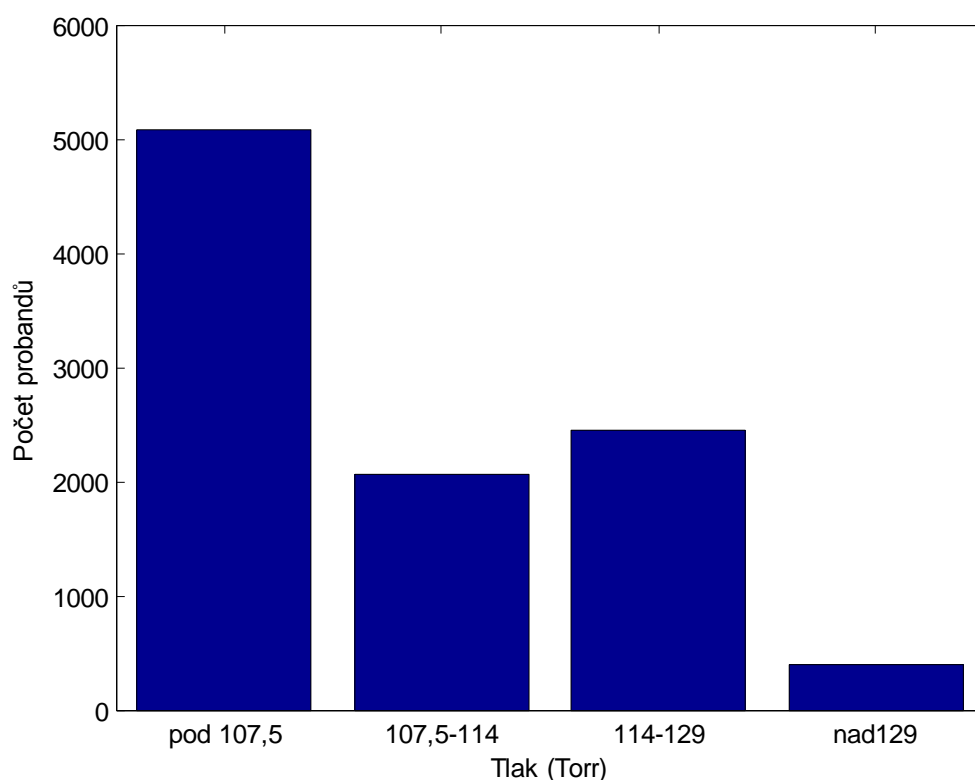
Tabulka vhodného léku (Tab. 1) na základě statistického testu uvádí, zda je pro léčbu dané nemoci vhodnější lék 1, lék 2, nebo zda jsou oba léky srovnatelně úspěšné.

Tab. 1: Vhodnost léku na danou chorobu s ohledem na věkovou skupinu

	Věková skupina				
	pod 30	31-40	41-50	51-60	nad 60
primární nemoc	1, 2	2	1, 2	1, 2	2
alfa	1	1	1, 2	1, 2	1, 2
beta	1, 2	1, 2	1, 2	2	1, 2

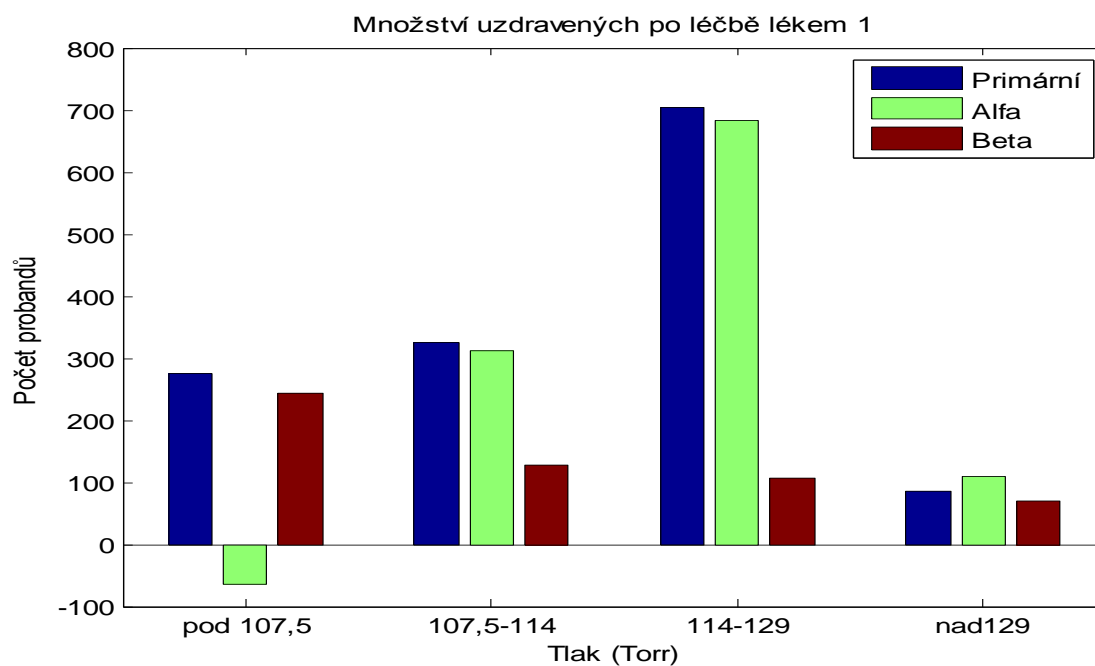
2.6 Analýza v závislosti na krevním tlaku

Rozdělení pro danou skupinu je patrné na Obr. 5. Rozdělení je velmi nerovnoměrné, což nám zřejmě ovlivní statistické testy.

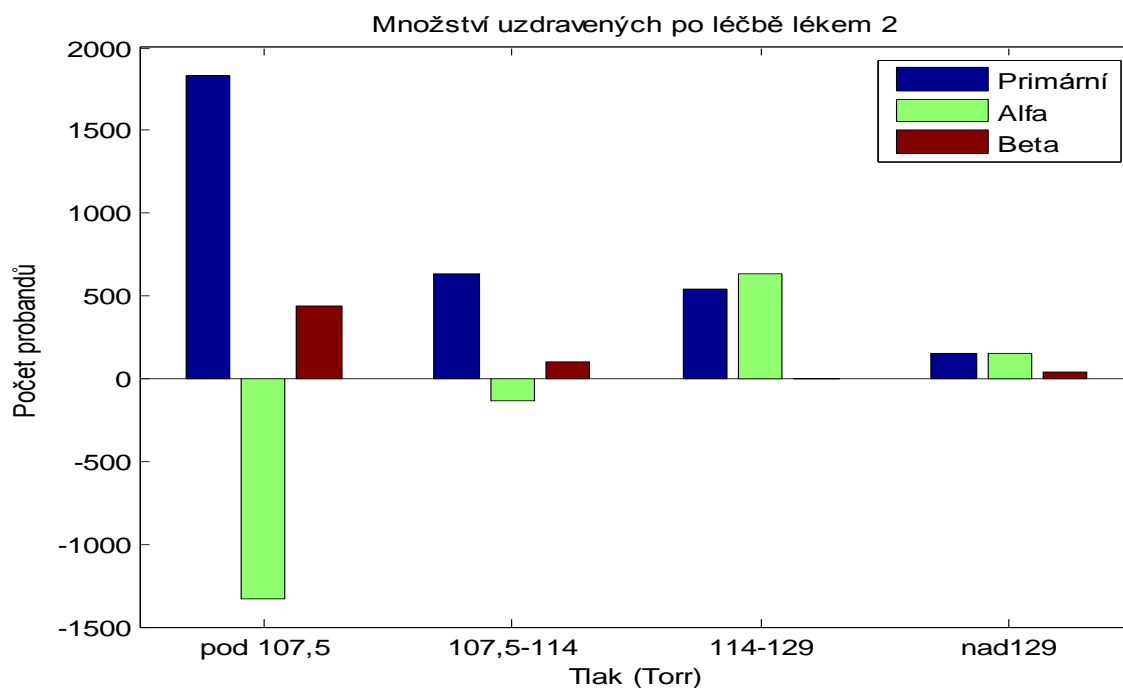


Obr. 5: Zastoupení jednotlivých skupin krevního tlaku dle zvoleného rozdělení

Množství probandů uzdravených po léčbě lékem 1 a těch, kteří byli uzdraveni po léčbě lékem 2 vzhledem k zařazení do skupin podle průměrného krevního tlaku je zobrazeno na Obr. 6 a Obr. 7.



Obr. 6: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 1 v závislosti na s. krevního tlaku



Obr. 7: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 2 v závislosti na s. krevního tlaku

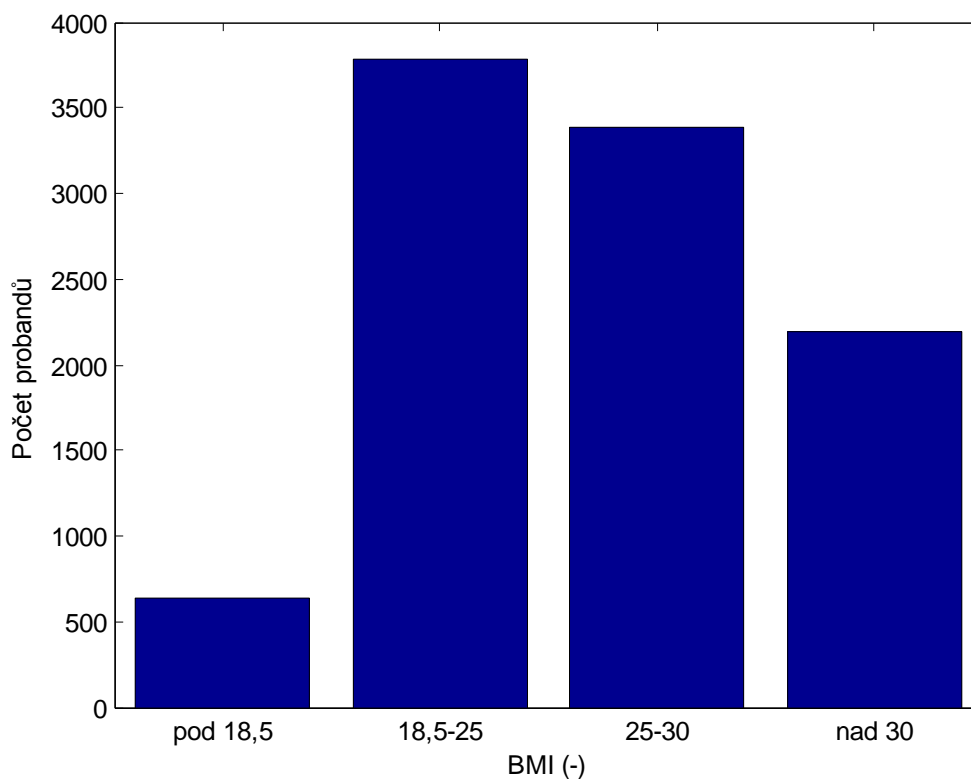
Tabulka vhodného léku (Tab. 2) na základě statistického testu uvádí, zda je pro léčbu dané nemoci vhodnější lék 1, lék 2, nebo zda jsou oba léky srovnatelně úspěšné.

Tab. 2: Vhodnost léku na danou chorobu s ohledem na sk. krevního tlaku

	Skupina krevního tlaku			
	pod 107,5	107,5-114	114-129	nad 129
primární nemoc	2	2	1, 2	1, 2
alfa	1	1	1, 2	1, 2
beta	1, 2	1, 2	1	1, 2

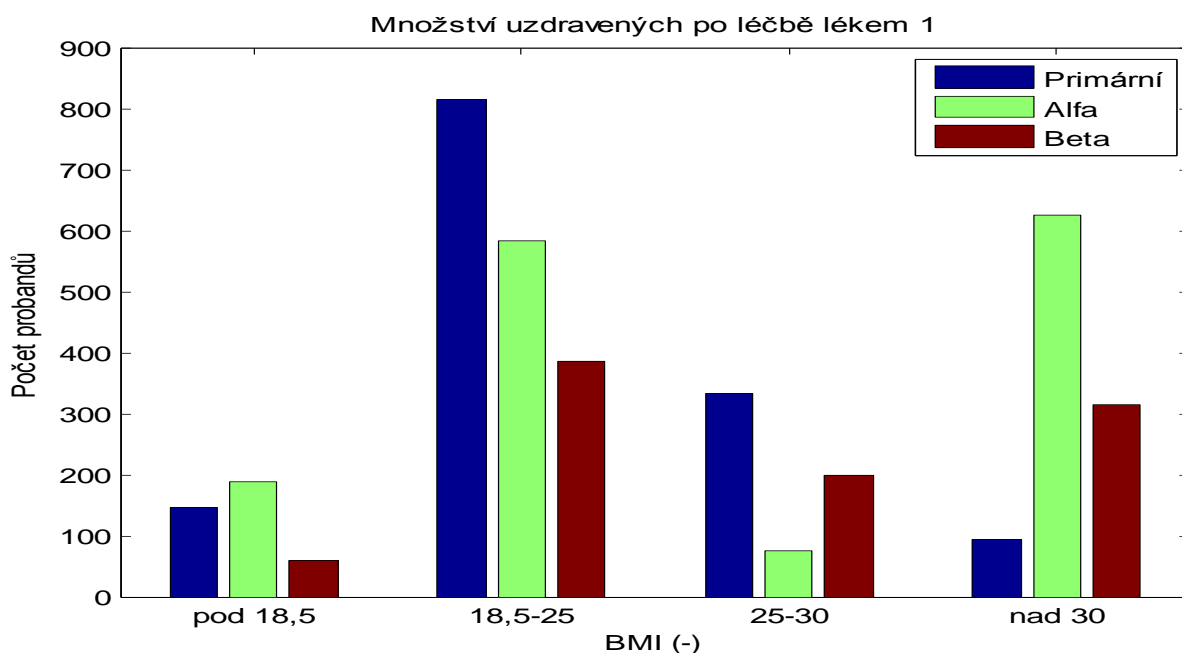
2.7 Analýza v závislosti na body mass indexu (BMI)

Rozdělení pro danou skupinu je patrné na Obr. 8. Rozdělení může způsobit problém v testech pro první skupinu.

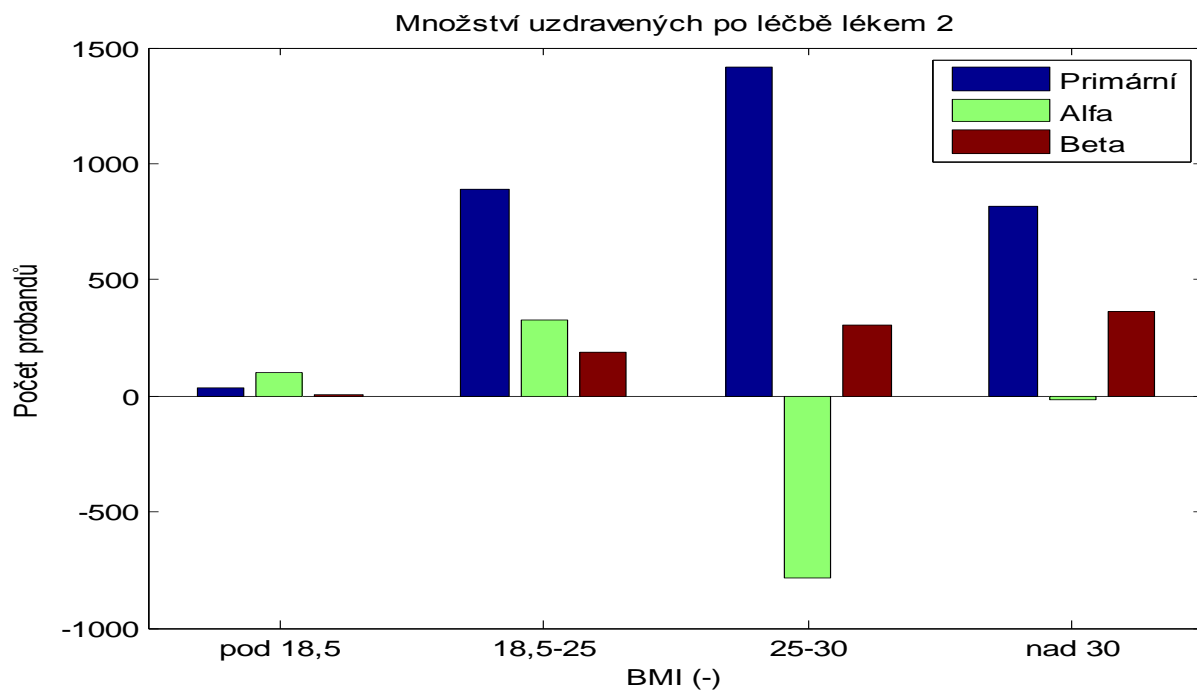


Obr. 8: Zastoupení jednotlivých skupin BMI dle zvoleného rozdělení

Množství probandů uzdravených po léčbě lékem 1 a těch, kteří byli uzdraveni po léčbě lékem 2 vzhledem k zařazení do skupin podle body mass indexu je zobrazeno na Obr. 9 a Obr. 10.



Obr. 9: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 1 v závislosti na skupině BMI



Obr. 10: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 2 v závislosti na skupině BMI

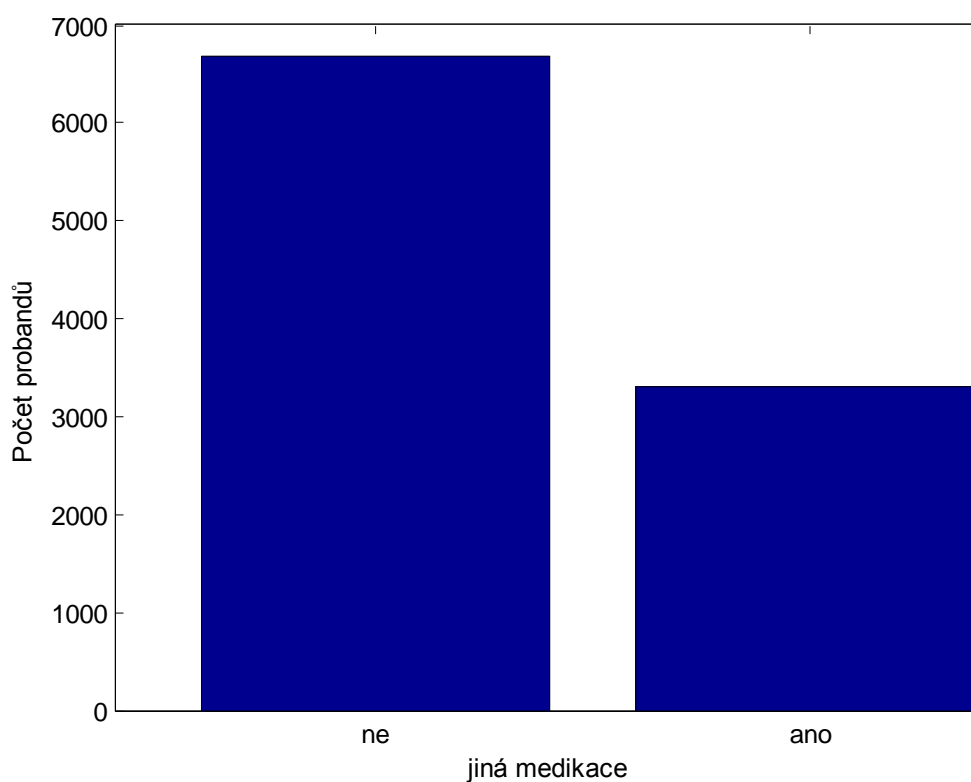
Tabulka vhodného léku (Tab. 3) na základě statistického testu uvádí, zda je pro léčbu dané nemoci vhodnější lék 1, lék 2, nebo zda jsou oba léky srovnatelně úspěšné.

Tab. 3: Vhodnost léku na danou chorobu s ohledem na skupinu BMI

	Skupina BMI			
	pod 18,5	18,5-25	25-30	nad 30
primární nemoc	1	1, 2	2	2
alfa	1, 2	1, 2	1	1
beta	1, 2	1, 2	1, 2	1, 2

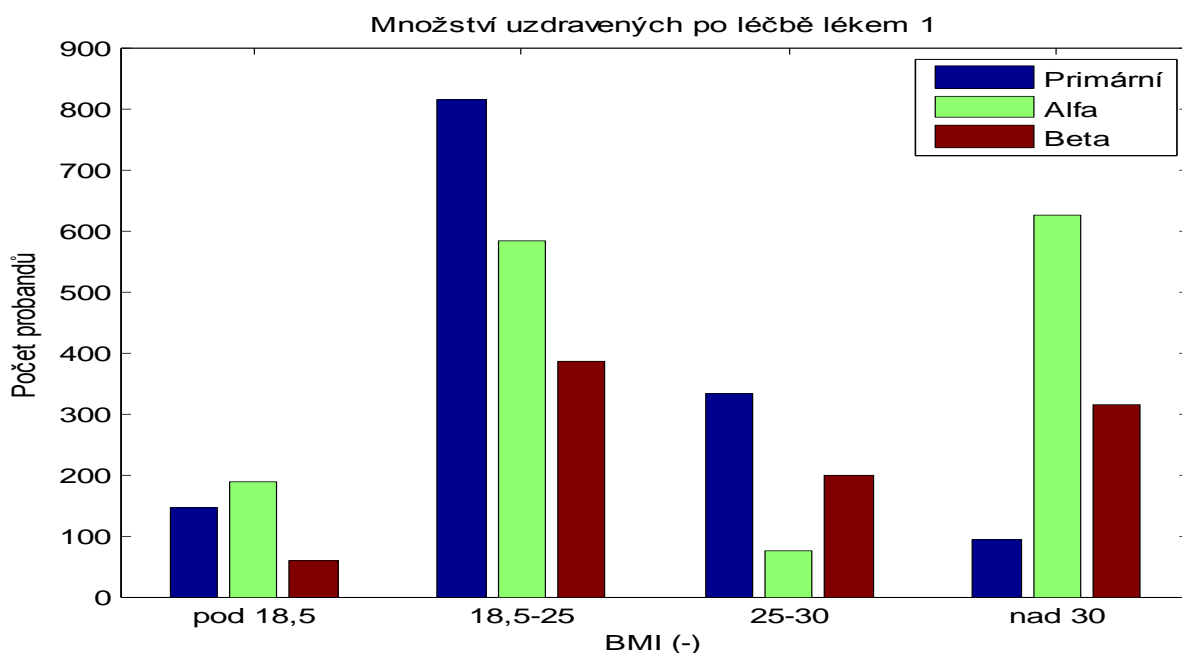
2.8 Analýza v závislosti na jiné medikaci

Rozdělení pro danou skupinu je patrné na Obr. 8. Rozdělení může způsobit problém v testech pro první skupinu.

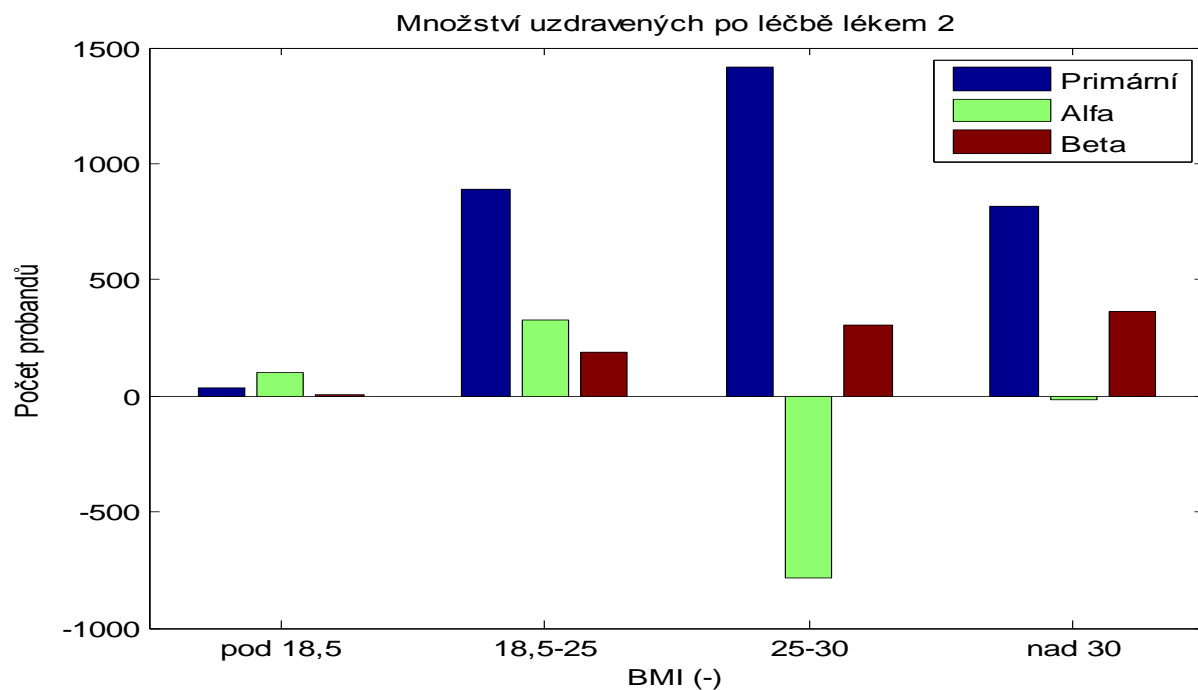


Obr. 11: Zastoupení jednotlivých skupin BMI dle zvoleného rozdělení

Množství probandů uzdravených po léčbě lékem 1 a těch, kteří byli uzdravení po léčbě lékem 2 vzhledem k zařazení do skupin podle jiné medikace je zobrazeno na Obr. 12 a Obr. 13.



Obr. 12: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 1 v závislosti na skupině BMI



Obr. 13: Množství probandů uzdravených z jednotlivých nemocí po léčbě lékem 2 v závislosti na skupině BMI

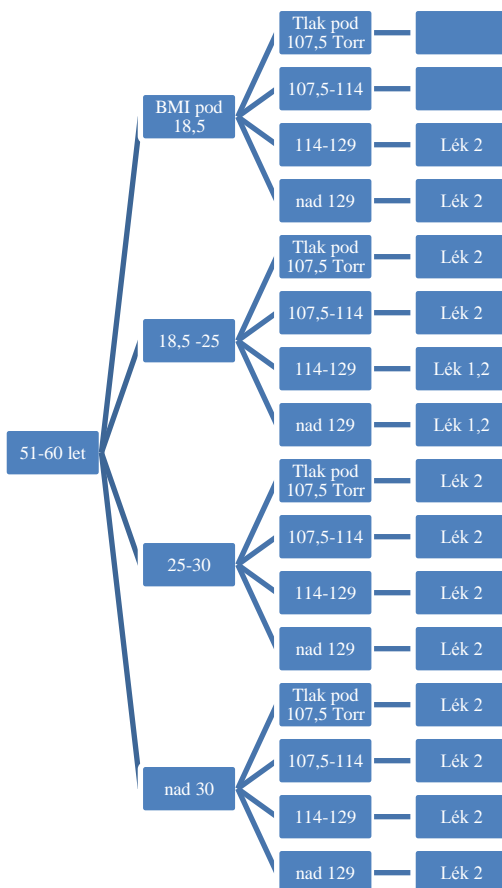
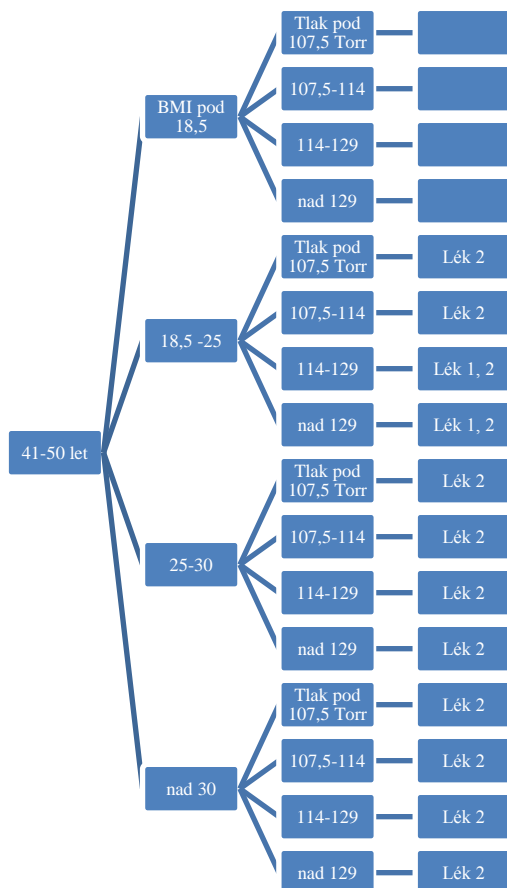
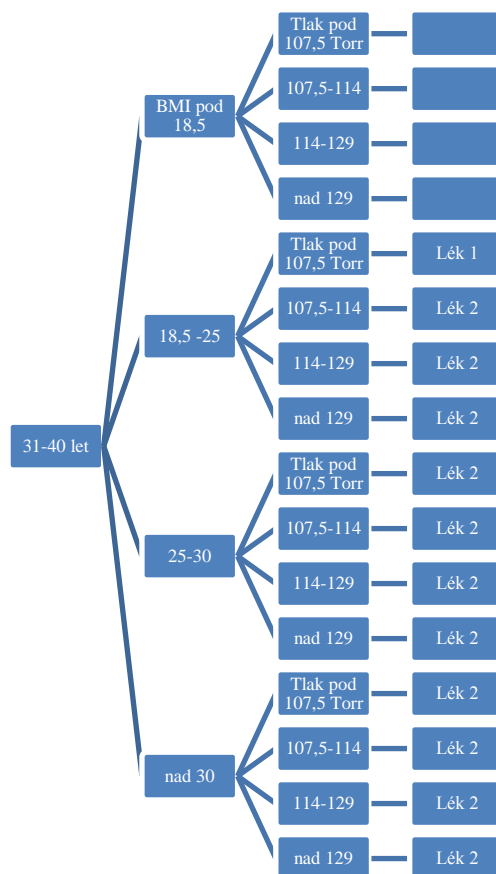
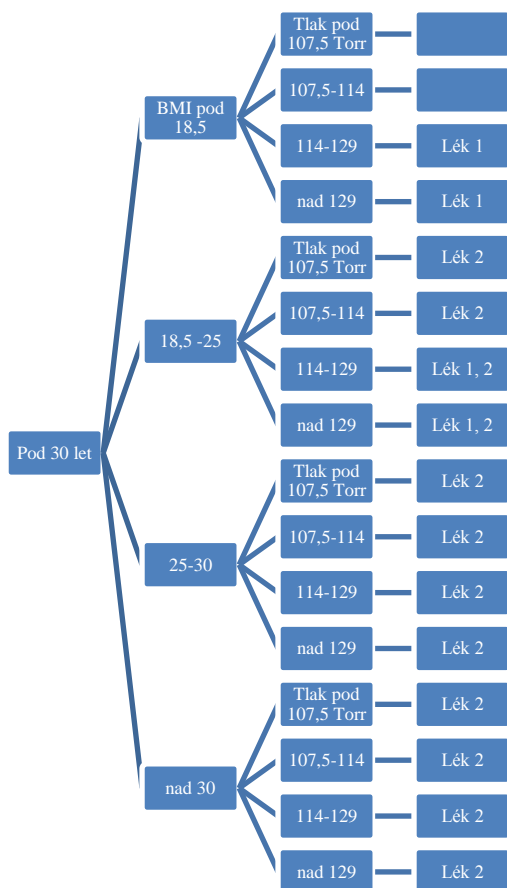
Tabulka vhodného léku (Tab. 3) na základě statistického testu uvádí, zda je pro léčbu dané nemoci vhodnější lék 1, lék 2, nebo zda jsou oba léky srovnatelně úspěšné.

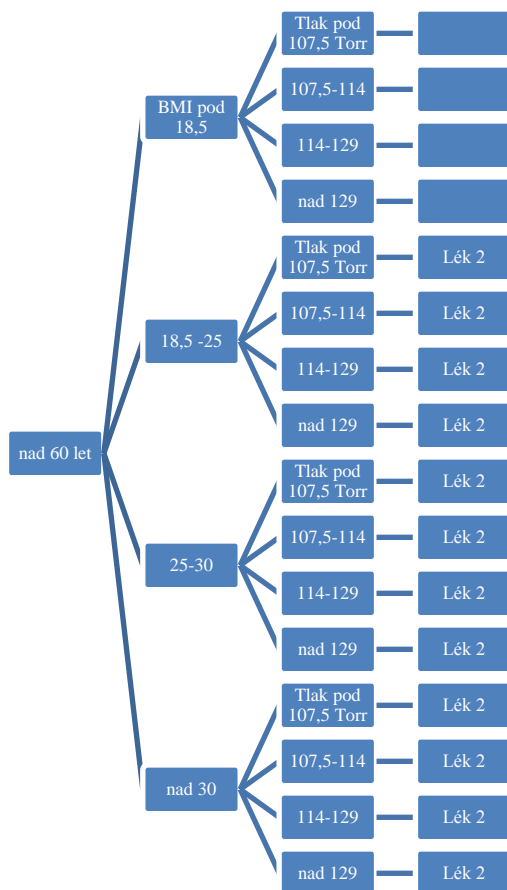
Tab. 3: Vhodnost léku na danou chorobu s ohledem na jinou medikaci

	Skupina BMI			
	pod 107,5	107,5-114	114-129	nad 129
primární nemoc	1	1, 2	2	2
alfa	1, 2	1, 2	1	1
beta	1, 2	1, 2	1, 2	1, 2

2.9 Navržení rozhodovacího stromu

Rozhodovací strom má sloužit k navržení vhodného léku pro pacienta, u nějž známe zařazení do jednotlivých skupin. Rozhodovací strom realizujeme na základě vhodnosti léků, které vzešli z testování jednotlivých skupin pro primární chorobu. Sekundární choroby předpokládáme nezávažné a tak nakažení prostě riskneme.





3. Závěr

Klasifikovali jsme jednotlivé skupiny i celý soubor dat, udělali jsme statistické testy, abychom věděli, jaký lék je v dané kategorii vhodný a poté jsme vytvořili rozhodovací strom.

Některá data jsou těžko klasifikovatelná pro nevhodný počet probandů ve skupině. Tento jev je kompenzován kompatibilitou s běžně užívanými klasifikačními systémy. Pro vyšší vypovídací hodnotu můžeme skupiny upravit, ale ztratíme kompatibilitu a porovnatelnost s případnými dalšími studiemi, které se pravděpodobně budou řídit zaběhnutou klasifikací.

V klasifikaci jsme u tlaku zanedbali symetricky extrémní hodnoty systolického a zároveň diastolického tlaku, které se na střední hodnotě neprojeví, čímž se zkreslí výsledky.

Chyby jsou způsobené (většinou řízeným) zanedbáním a generalizací dat. Variabilita mezi výsledky stejného zkoumaného balíku dat různými výzkumnými týmy je velmi pravděpodobná.

3.1 Možná vylepšení a návaznost

Bylo by ještě možné vylepšit výpovědní hodnotu zohledněním faktu, kolik lidí onemocnělo sekundárními chorobami a kolik se jich vyléčilo. Ne za použití pouhého rozdílu těchto hodnot na začátku a na konci studie.

Dalším možným vylepšením je zmíněná změna hranic jednotlivých skupin, která však nese již jmenovaná rizika. Neblahý jev s nevhodným počtem probandů v jednotlivých skupinách může být kompenzován doplněním zkoumaného vzorku o další jedince nebo pomocí komparační studie.

4. Použitá literatura

- [1] *British hypertension society* [online]. [vid. 19. 3. 2013]. Dostupné z:
http://www.medicalcriteria.com/site/index.php?option=com_content&view=article&id=134%3Acarbhsiv&catid=46%3Acardiology&Itemid=80&lang=en
- [2] *World Health Organisation & ISH* [online]. [vid. 19. 3. 2013]. Dostupné z:
<http://www.medscape.com/viewarticle/471863>
- [3] *World Health Organisation* [online]. [vid. 19. 3. 2013]. Dostupné z:
http://apps.who.int/bmi/index.jsp?introPage=intro_3.htmlNAGAYA, Ken et al. Airway Humidification With a Heated Wire Humidifier During High-Frequency Ventilation Using Babylog 8000 Plus® in Neonates. *Pediatric Pulmonology* [PDF]. 2009, 44: 260–266. [vid. 11. 12. 2011].
- [4] KÁBRT, M. *Test chí-kvadrát nezávislosti v kontingenční tabulce*, Univerzita Hradec Králové, 2011. [online]. [vid. 19. 3. 2013]. Dostupné z:
<http://www.milankabrt.cz/testNezavislosti/>