

# Lékařská informatika - Úloha zpracování a analýzy dat

Peter Schmiedt, Eduard Füzesséry

10. novembra 2015

# 1 Úvod

Cieľom tejto úlohy je spracovanie a analýza dát z testovania účinnosti liečby dvoch rôznych liekov na primárnu chorobu a aký majú na to vplyv vek, BMI, sekundárne choroby, prítomnosť inej medikácie, atď.

Ďalšou úlohou je nájsť zaujímavé fakty a návaznosti medzi veličinami štatistickej významnosti.

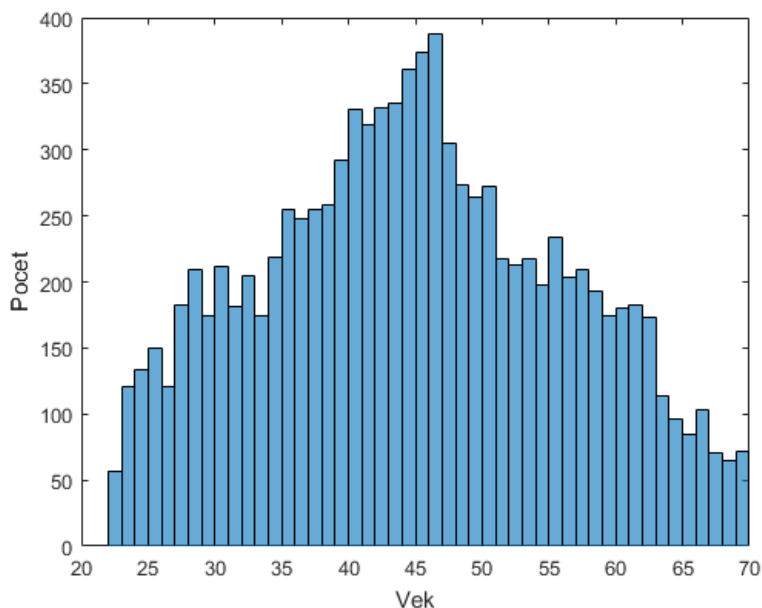
## 2 Spracovávané Dáta

Budeme spracovávať štatistické dáta pacientov, ktorý užívali dva rôzne lieky (nikdy nie naraz) proti primárnej chorobe. Máme k dispozícii dáta o vzorke 10 000 pacientov.

K dispozícii máme, aký liek bol podávaný pacientovi a aký účinok to malo na jeho primárne ochorenie. Ďalej máme k dispozícii základné fyzické proporcie pacienta ako: vek, index telesnej hmotnosti a priemerný krvný tlak. Je nám známa taktiež anamnéza pacienta (výskyt sekundárnych chorôb pred a po liečbe a iná nešpecifikovaná medikácia).

### 2.1 Vekové spektrum

Máme k dispozícii dáta pacientov so širokým vekovým spektrom, ktoré môžeme vidieť v histograme 1. Histogram udáva absolútny počet pacientov v jednotlivých vekových kategóriách. Pomocou t-testu sme si overili normálne rozdelenie týchto veličín na hladine štatistickej významnosti (0.05). V tabuľke 1 je rozdelenie pacientov do vekových podskupín.



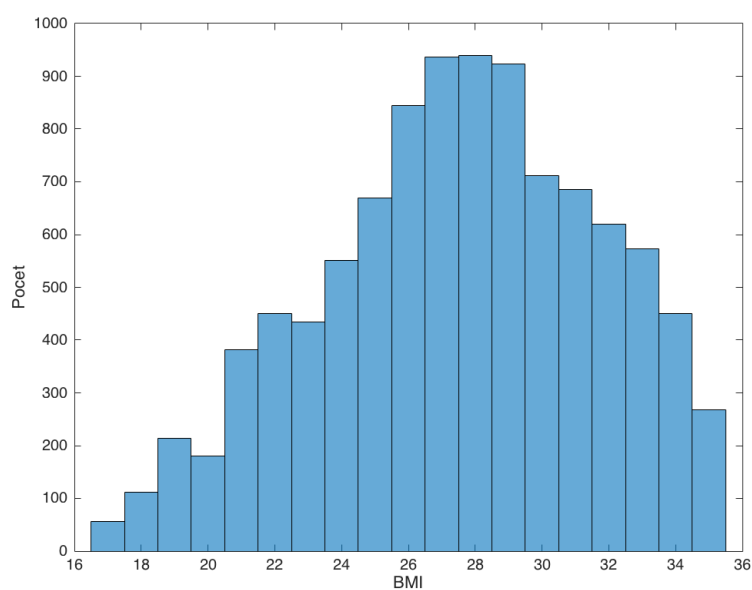
Obr. 1: Histogram Vekov

Vek	<30	30-39	40-49	50-59	60<
Počet Pacientov	1148	2299	3283	2130	1140

Tabuľka 1: Vek pacientov

## 2.2 BMI spektrum

BMI (ang. Body Mass Index) je index telesnej hmotnosti. Je to pomer medzi aktuálnou váhou a výškou<sup>2</sup>. V histograme 2 je rozloženie BMI pacientov. V tabuľke 1 sú dáta zobrazené v jednotlivých podskupinách. Za pomoci t-testom sme si overili normálne rozdelenie veličín na hladine štatistickej významnosti (0.05).



Obr. 2: Histogram BMI

BMI	Podvýživa	Zdravá Váha	Mierna Nadváha	Obezita
Počet Pacientov	167	2212	4314	3307

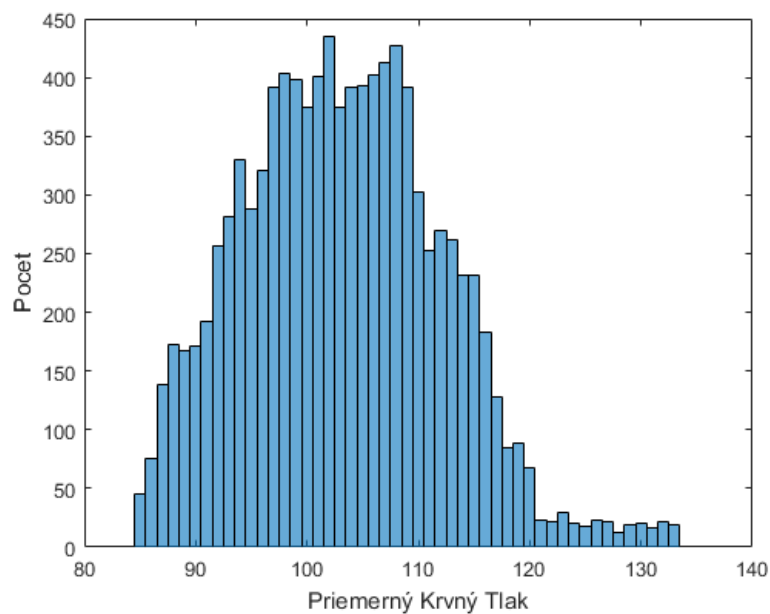
Tabuľka 2: BMI Pacientov

## 2.3 MAP spektrum

MAP (ang. Mean Arterial Pressure) je priemerný krvný tlak udávaný v mmHg. Rozdelenie medzi pacientami môžeme vidieť v histograme 3 a rozdelenie do podskupín v tabuľke 3.

MAP	<70	70-92	93-105	106-119	120<
Počet Pacientov	0	1218	4785	3665	332

Tabuľka 3: MAP Pacientov



Obr. 3: Histogram MAP

## 2.4 Lieky a Medikácia

V tabuľke 4 môžeme vidieť počet vyliečených a nevyliečených pacientov (primárna choroba) po podaní lieku 1 a lieku 2.

	Liek 1	Liek 2	Celkom
<b>Vyliečený</b>	1625	3630	5255
<b>Nevyliečený</b>	3837	908	4745
<b>Celkom</b>	5462	4538	10 000

Tabuľka 4: Vyliečený a nevyliečený pacienti s rôznymi liekmi

V tabuľke 5 vidíme účinnosť jednotlivých liekov na našej vzorke. Účinnosť myslíme ako pomer vyliečených k celkovému počtu pacientov liečených daným liekom.

	Liek 1	Liek 2
<b>Účinnosť</b>	29,75%	80%

Tabuľka 5: Účinnosť jednotlivých liekov

V tabuľke 6 pozorujeme pridružené sekundárne choroby pred a po liečbe daným typom lieku. Niektoré dáta sa prekrývajú s tým, že niektorí pacienti môžu mať výskyt prvej a aj výskyt druhej sekundárnej choroby pred a po liečbe.

	<b>1. sek. choroba</b>	<b>2. sek. choroba</b>
<b>Výskyt pred liečbou</b>	4717	6387
<b>Výskyt po liečbe</b>	4505	3888

Tabuľka 6: Výskyt sekundárnych chorôb pred a po liečbe

### 3 Rozhodovací systém

Prieniky jednotlivých kategórií si budeme značiť systémom ID, kde ID bude 4-ciferné číslo a každá cifra znamená nejakú konkrétnu kategóriu. Prvá cifra udáva vek, druhá cifra udáva MAP, tretia BMI a štvrtá cifra nám udáva prítomnosť inej medicíny počas liečby.

Vek		MAP		BMI		Iná medicína	
1000	<30	100	<70	10	Podvýživa	Prítomná	1
2000	30-39	200	70-92	20	Normál	Neprítomná	0
3000	40-49	300	93-105	30	Nadváha		
4000	50-59	400	106-119	40	Obezita		
5000	60<	500	120<				

Tabuľka 7: Rozhodovací systém

Napríklad: **ID-2321** znamená že ide o pacientov vo veku 30-39 rokov, s MAP v rozsahu 93-105mmHg, BMI normálne a s prítomnosťou inej medicíny.

Podskupiny s počtom pacientov menších ako 25 budeme zanedbávať z dôvodu štatistickej nevýznamnosti.

Tabuľku s výsledkami je možné vidieť v prílohe A.

## 4 Testovanie jednotlivých liekov

Každá podskupina (so svojím vlastným ID) je testovaná na účinnosť lieku 1 alebo 2. Resp. ktorý liek je účinnejší.

Budeme testovať nasledujúcu hypotézu: “Medzi liekom 1 a 2 nie je žiadny rozdiel účinnosti.” Hypotézu budeme testovať za pomoci štatistického testu dobrej zhody (chí-kvadrát test) [1] [2]. Výsledok týchto testov môžeme vidieť v tabuľke v prílohe A.

V niektorých podskupinách nie je potvrdená účinnosť ani jedného z liekov a to už z dôvodu malej vzorky (menej ako 25 pacientov) alebo z neúčinnosti oboch liekov.

### 4.1 Vplyv liekov na pridružené choroby

Testovali sme aj vplyv liekov na pridružené sekundárne choroby.

	Liek 1	Liek 2
1. sek. choroba+		
1. sek choroba–		
2. sek choroba+		
2. sek choroba–		

Tabuľka 8: Vplyv liekov na pridružené ochorenia



## 5 Záver

Zo štatistickej analýzy dát nám vyšlo na hladine významnosti 5% , lepšia účinnosť druhého lieku. Dáta sme najprv zanalyzovali a rozdelili do skupín podľa veku, tlaku, bmi a samozrejme, ktorý liek užívali. Tieto dáta sme otestovali t-testom, na overenie správnosti nášho rozdelenia. Potom sme si určili systém, na základe ktorého sme urobili štatistickú analýzu dát, z ktorých sme odstránili štatisticky nevýznamné skupiny (pod 25 pacientov v skupine). Tieto štatisticky významné dáta sme následne otestovali chí-kvadrát testom, všetky tieto údaje sú v prílohe A.

## Literatúra

- [1] Wikipedia. Chi-squared test, 2015. [Online; pristupované 6-November-2015].
- [2] Jana Zvárová. *Základy statistiky pro biomedicínské obory*. Karolinum, 2007.

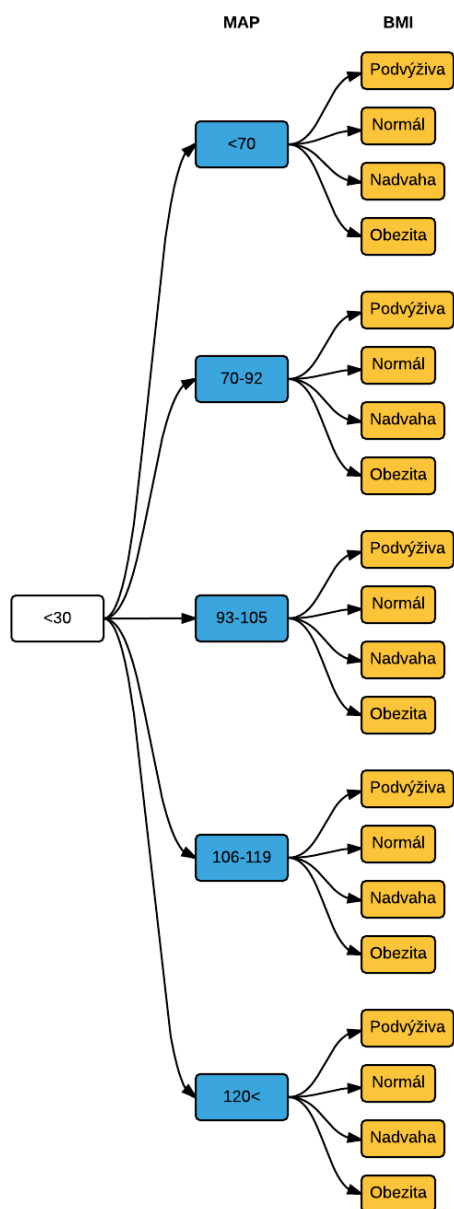
## A Tabuľka účinnosti liekov

ID	Celkový počet	Liek 1 Vyliečený	Liek 1 Nevyliečený	Liek 2 Vyliečený	Liek 2 Nevyliečený	Chi-kvadrát			
1230	109	34	45	3	27	23.064	0	0	0
1231	31	26	3	1	1	35.207	0	0	0
1320	45	10	8	2	25	21.814	0	0	0
1330	329	162	94	13	60	132.04	0	0	0
1331	134	113	10	6	5	160.43	0	0	0
1420	49	16	7	0	26	33.056	0	0	0
1430	242	148	39	11	44	178.11	0	0	0
1431	100	85	9	1	5	119.33	0	0	0
2220	27	10	6	3	8	9.5911	0	0	0
2230	201	50	67	12	72	51.974	0	0	0
2231	65	52	12	1	0	57.107	0	0	0
2320	175	40	39	5	91	78.706	0	0	0
2330	597	269	142	31	155	269.33	0	0	0
2331	244	216	13	8	7	322.29	0	0	0
2341	31	30	1	0	0	NaN	0	0	0
2420	177	51	46	5	75	71.381	0	0	0
2430	400	235	74	11	80	275.13	0	0	0
2431	170	145	16	1	8	202.19	0	0	0
2530	41	25	9	1	6	25.683	0	0	0
3220	38	11	8	2	17	16.522	0	0	0
3230	236	51	34	19	132	111.93	0	0	0
3231	82	53	6	18	5	71.595	0	0	0
3320	321	121	64	11	125	166.97	0	0	0
3321	65	6	6	1	52	42.078	0	0	0
3330	780	253	122	35	370	437.09	0	0	0
3331	333	171	24	60	78	243.07	0	0	0
3410	29	7	16	0	6	4.8036	0	0	0
3420	319	96	100	14	109	96.754	0	0	0
3421	57	10	6	1	40	36.599	0	0	0
3430	586	231	94	27	234	344.87	0	0	0
3431	205	110	17	16	62	172.54	0	0	0
3520	26	9	7	2	8	8.2768	0	0	0
3530	62	20	7	3	32	40.152	0	0	0
4230	134	20	8	7	99	87.727	0	0	0

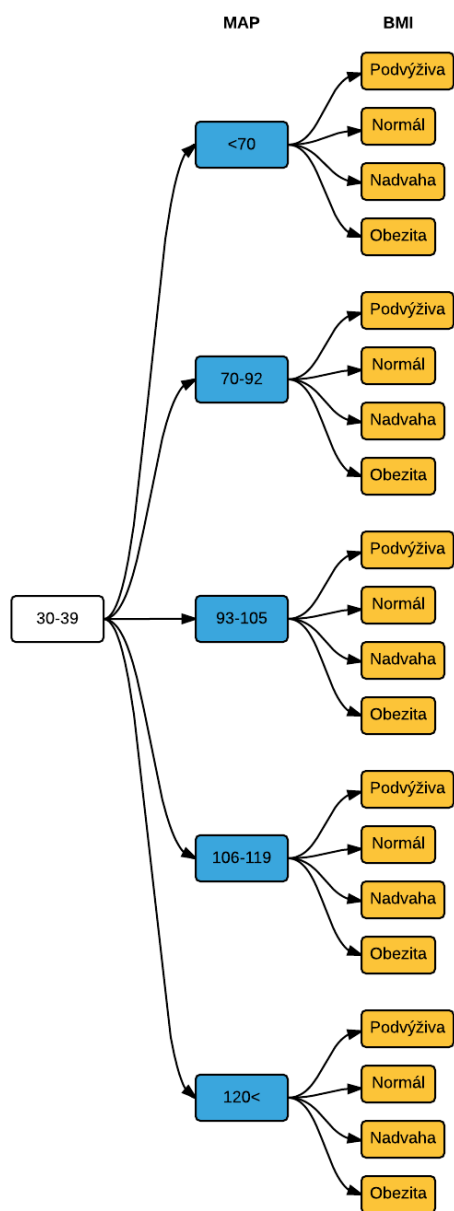
4231	66	7	4	47	8	4.4503	0	0	0
4320	213	86	58	13	56	75.744	0	0	0
4321	65	3	5	7	50	32.836	0	0	0
4330	511	115	57	27	312	299.51	0	0	0
4331	197	19	13	99	66	17.69	0	0	0
4420	221	89	75	2	55	74.827	0	0	0
4421	54	2	4	3	45	32.66	0	0	0
4430	403	131	51	16	205	255.34	0	0	0
4431	120	16	6	38	60	34.477	0	0	0
4530	32	6	4	1	21	18.733	0	0	0
5230	67	8	5	3	51	41.824	0	0	0
5231	56	7	1	43	5	9.5193	0	0	0
5320	74	21	10	3	40	43.702	0	0	0
5321	36	1	2	8	25	12.837	0	0	0
5330	281	35	20	20	206	164.24	0	0	0
5331	144	9	9	97	29	2.6243	0	0	0
5420	69	18	8	2	41	44.139	0	0	0
5430	216	48	16	15	137	137.5	0	0	0
5431	80	12	13	33	22	5.5488	0	0	0

---

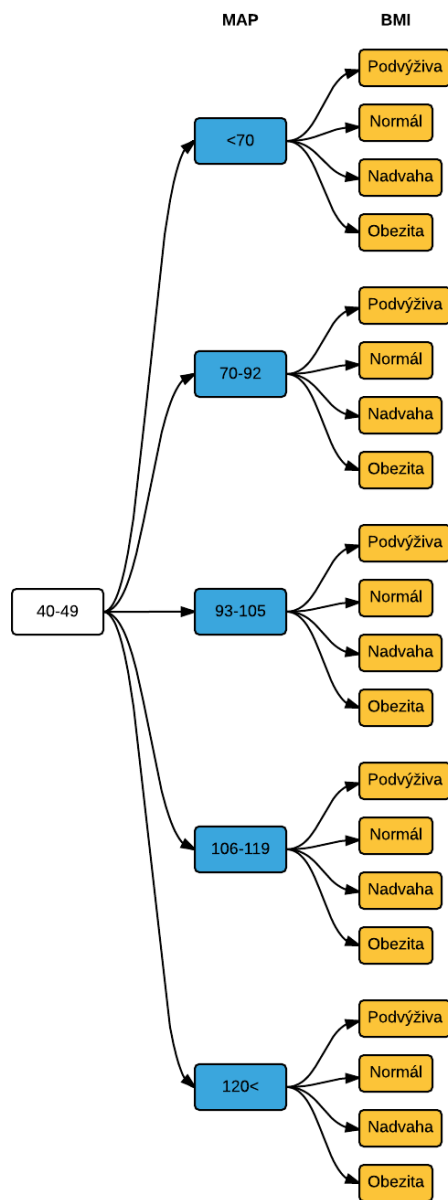
## B Rozhodovací strom



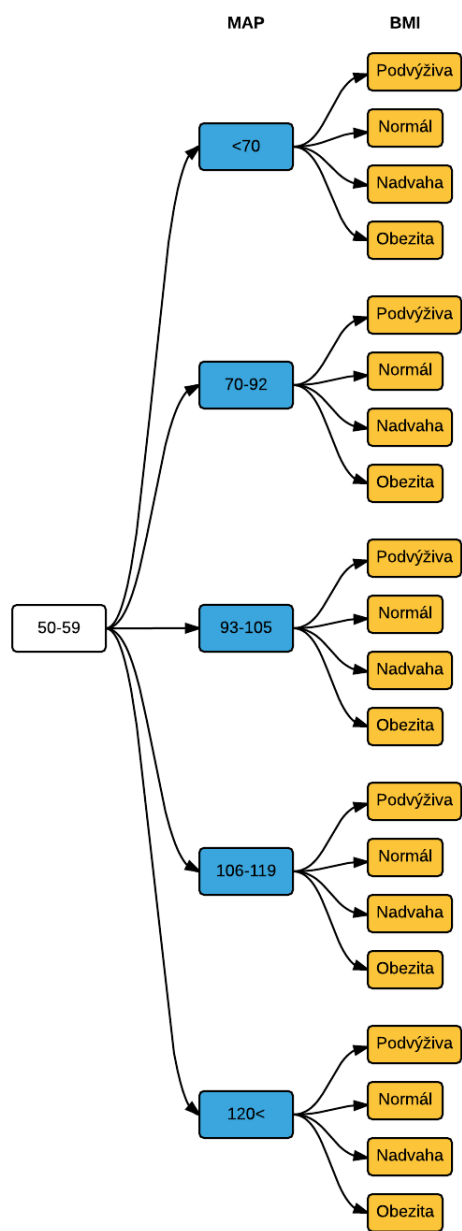
Obr. 4: Rozhodovací strom pre pacientov pod 30 rokov



Obr. 5: Rozhodovací strom pre pacientov od 30 do 39 rokov

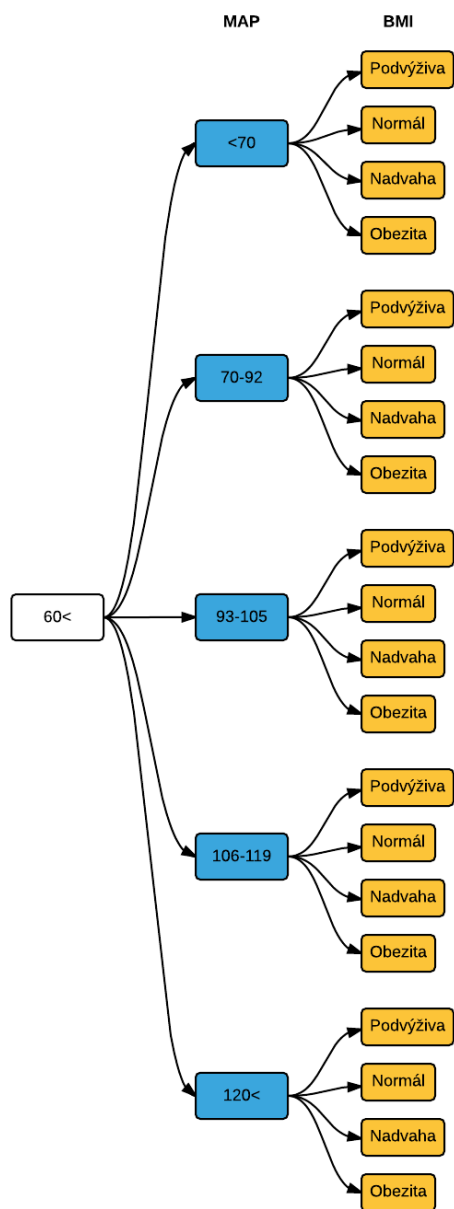


Obr. 6: Rozhodovací strom pre pacientov od 40 do 49 rokov



Obr. 7: Rozhodovací strom pre pacientov od 50 do 59 rokov





Obr. 8: Rozhodovací strom pre pacientov nad 60 rokov