

Lékařská informatika - Úloha zpracování a analýzy dat

Peter Schmiedt, Eduard Füzesséry

8. 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.

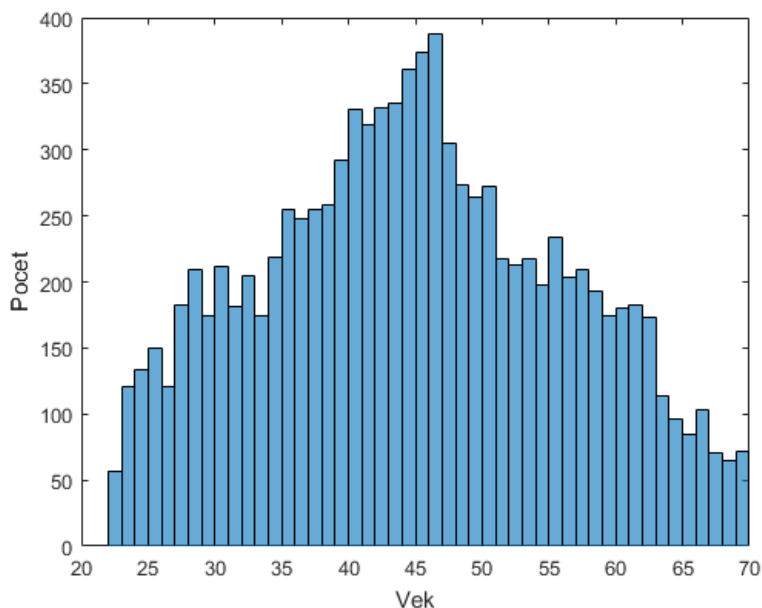
2 Spracovávané Dáta

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.



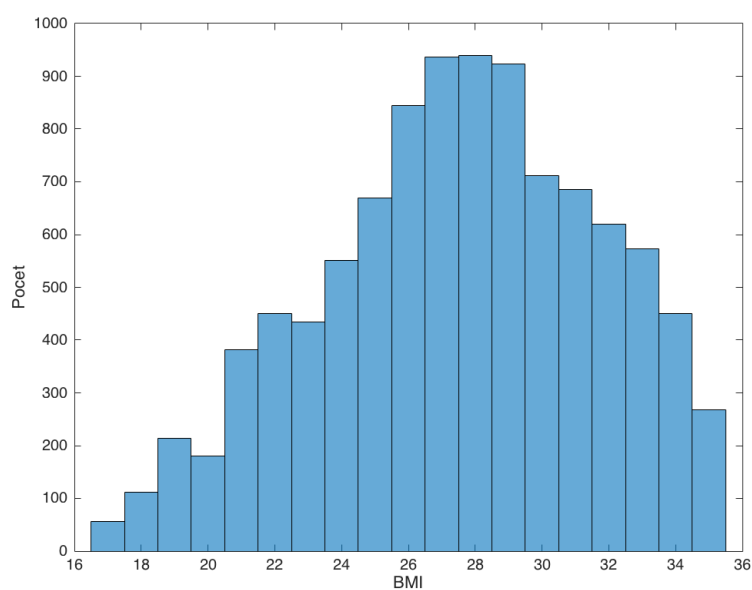
Obr. 1: Histogram Vekov

| 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 výškou². 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 |

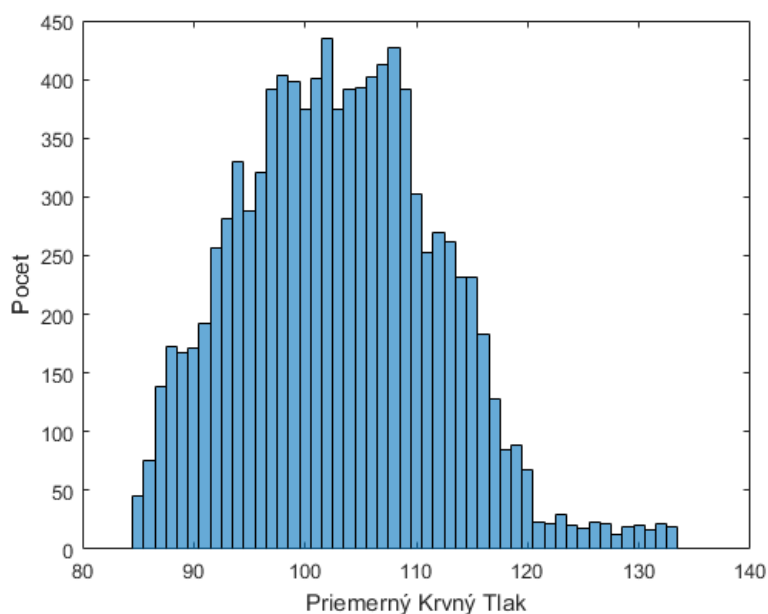
Tabuľka 2: BMI Pacientov

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 |
|--------------------|--------|--------|--------|
| Vyliečený | 1625 | 3630 | 5255 |
| Nevyliečený | 3837 | 908 | 4745 |
| Celkom | 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 |
|-----------------|--------|--------|
| Účinnosť | 29,75% | 80% |

Tabuľka 5: Účinnosť jednotlivých liekov

V tabuľke 6 pozorujeme pridružené sekundárne choroby pred a po liečbe daným typom lieku.

| | Prvá sekundárna choroba | Druhá sekundárna choroba |
|---------------------|-------------------------|--------------------------|
| Výskyt pred liečbou | 4717 | 6387 |
| Výskyt po liečbe | 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 | | Prítomnosť inej medicíny | |
|------|-------|-----|---------|-----|-----------|--------------------------|---|
| 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.

4 Testovanie jednotlivých liekov

Literatúra