

## 1 $\chi^2$ -test

V nemocnici sledovali pacientov prijatých na psychiatrické oddelenie – vzorok *A* a pacientov prijatých na interné oddelenie s problémami s trávením – vzorok *B*. U pacientov sledovali príjmové skupiny označené I, II, III, IV, V.

Príjmová skupina	Vzorok <i>A</i>	Vzorok <i>B</i>
I	17	5
II	25	21
III	39	34
IV	42	49
V	32	25
Celkom	155	134

Majú vzorky *A* a *B* rovnaké rozdelenie? Otestujte to  $\chi^2$  testom na hladine významnosti 95%.

## 2 Rozhodovacie stromy

### 2.1 Príklad A

Máme dáta o pacientoch po infarkte s nasledujúcimi položkami:

Vitamín pravidelne užívaný typ vitamínu.

Veľkosť rodiny veľkosť domácnosti, v ktorej pacient žil.

Cvičil či pravidelne cvičil.

Mal bypass či mal bypass.

Prežil či prežil aspoň 5 rokov od prvého infarktu.

Vitamín	Veľkosť rodiny	Cvičil	Mal bypass	Prežil
B complex	veľká	pravidelne	nie	–
B complex	veľká	pravidelne	áno	–
C	veľká	pravidelne	nie	+
D	stredná	pravidelne	nie	+
D	malá	málo	nie	+
D	malá	málo	áno	–
C	malá	málo	áno	+
B complex	stredná	pravidelne	nie	–
B complex	malá	málo	nie	+
D	stredná	málo	nie	+
B complex	stredná	málo	áno	+
C	stredná	pravidelne	áno	+
C	veľká	málo	nie	+
D	stredná	pravidelne	áno	–

Pomocou algoritmu TDIDT skonštruujte rozhodovací strom s cieľovým atribútom Prežil.

## 2.2 Príklad B

Máme dáta o vozidlách s nasledujúcimi položkami:

Veľkosť veľkosť vozidla: M – malé, S – stredné a V – veľké.

Cena cena vozidla: M – malá, S – stredná a V – vysoká.

Výkon výkon motora: V – veľký, M – malý.

Farba farba vozidla: S – svetlá a T – tmavá.

Kúpi či si zákazník takéto vozidlo kúpi: A – áno, N – nie.

Značka	Veľkosť	Cena	Výkon	Farba	Kúpi
Lada	S	M	V	S	N
Octavia	S	S	M	S	A
Smart	M	V	M	T	N
BMW	S	V	V	T	A
traktor	V	V	V	T	A
Babeta	M	M	M	S	A
Felicie	S	M	M	T	N
Ferrari	M	V	V	S	N

Pomocou algoritmu TDIDT skonštruujte rozhodovací strom s cieľovým atribútom Kúpi.