

Vzorová úloha č. 2 z Vícerozměrných metod - cvičení

Ve studii byl u 6 osob zjišťován systolický tlak a hladina celkového cholesterolu v krvi. Naměřené hodnoty jsou uvedeny v přiložené tabulce.

Pacient	Systolický tlak (mmHg)	Celkový cholesterol (mmol/l)
A	165	4,5
B	125	4,7
C	160	7,5
D	170	7,0
E	130	4,0
F	165	6,5

Vztahy mezi pacienty jsou vyjádřeny následujícími asociačními maticemi:

A) Asociační matice počítaná na původních datech

	A	B	C	D	E	F
A	0	40,00	5,83	5,59	35,00	2,00
B	40,00	0	35,11	45,06	5,05	40,04
C	5,83	35,11	0	10,01	30,20	5,10
D	5,59	45,06	10,01	0	40,11	5,02
E	35,00	5,05	30,20	40,11	0	35,09
F	2,00	40,04	5,10	5,02	35,09	0

B) Asociační matice počítaná na standardizovaných datech

	A	B	C	D	E	F
A	0	2,04	2,05	1,71	1,81	1,35
B	2,04	0	2,60	2,77	0,54	2,37
C	2,05	2,60	0	0,61	2,82	0,72
D	1,71	2,77	0,61	0	2,87	0,42
E	1,81	0,54	2,82	2,87	0	2,46
F	1,35	2,37	0,72	0,42	2,46	0

Určete shluky pacientů pomocí aglomerativního hierarchického shlukování, získejte následující výstupy a odpovězte na otázky:

1. Výše uvedené asociační matice vyjadřují podobnost nebo vzdálenost? Zdůvodněte. (1 bod)
Vzdálenost, protože na diagonále jsou nuly
2. Jakou z uvedených asociačních matic byste pro shlukování použili a proč? (2 body)
matici B, protože parametry mají jiné jednotky a velmi odlišný rozsah hodnot (tzn., mají různou variabilitu, jiné číselné hodnoty), proto je vhodné data standardizovat
3. Vytvořte dendrogram pomocí algoritmu nejvzdálenějšího souseda. Rozepište jednotlivé kroky výpočtu. (2 body)

Krok 1:

	A	B	C	D	E	F
A	0	2.04	2.05	1.71	1.81	1.35
B	2.04	0	2.60	2.77	0.54	2.37
C	2.05	2.60	0	0.61	2.82	0.72
D	1.71	2.77	0.61	0	2.87	0.42
E	1.81	0.54	2.82	2.87	0	2.46
F	1.35	2.37	0.72	0.42	2.46	0

Shluk: D-F

Krok 2:

	A	B	C	D+F	E
A	0	2.04	2.05	1.71	1.81
B	2.04	0	2.60	2.77	0.54
C	2.05	2.60	0	0.72	2.82
D+F	1.71	2.77	0.72	0	2.87
E	1.81	0.54	2.82	2.87	0

Shluk: B-E

Krok 3:

	A	B+E	C	D+F
A	0	2.04	2.05	1.71
B+E	2.04	0	2.82	2.87
C	2.05	2.82	0	0.72
D+F	1.71	2.87	0.72	0

Shluk: (D-F)-C

Krok 4:

	A	B+E	D+F+C
A	0	2.04	2.05
B+E	2.04	0	2.87
D+F+C	2.05	2.87	0

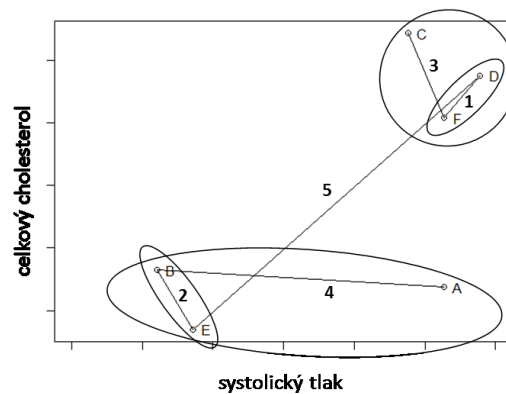
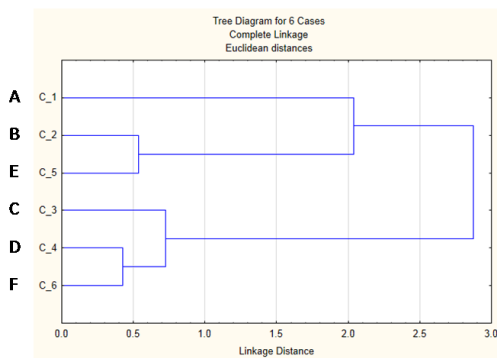
Shluk: (B-E)-A

Krok 5:

	B+E+A	D+F+C
B+E+A	0	2.87
D+F+C	2.87	0

Shluk: ((D-F)-C)-((B-E)-A)

Dendrogram



4. Pokud bychom v dendrogramu provedli řez na podobnosti/vzdálenosti 15 (v případě výpočtu pomocí asociační matice A) resp. 1,5 (v případě výpočtu pomocí asociační matice B), kolik dostaneme shluků? Kteří pacienti budou v jednotlivých shlucích? Výsledek interpretujte. (3 body)

V případě správného výpočtu pomocí matice B dostaneme 3 shluky: (C+D+F), (B+E) a (A) – v prvním shluku jsou pacienti s vysokým tlakem i vysokým cholesterolem, v druhém shluku pacienti s nízkým tlakem i nízkým cholesterolem, ve třetím shluku pacient s vysokým tlakem a nízkým cholesterolem.
5. Pokud bychom chtěli rozdělit osoby do několika shluků jednoho řádu, jakou metodu bychom použili? (1 bod) **k-průměrů nebo x-průměrů nebo k-medoidů**
6. Pokud bychom chtěli rozdělit osoby do několika skupin s podskupinami nižších řádů tak, že vzdálenost mezi shluky bude minimální vzdáleností mezi zástupci, jakou metodu bychom použili? (1 bod) **metodu nejbližšího sousedá**