

**UNIVERZITA PARDUBICE**  
**FAKULTA CHEMICKO-TECHNOLOGICKÁ**  
**KATEDRA ANALYTICKÉ CHEMIE**

**LICENČNÍ STUDIUM - STATISTICKÉ ZPRACOVÁNÍ DAT**

## **SEMESTRÁLNÍ PRÁCE**

**3.5 Klasifikace analýzou vícerozměrných dat**

Ing. Věra Fialová

BIOPHARM  
VÝZKUMNÝ ÚSTAV BIOFARMACIE A VETERINÁRNÍCH LÉČIV a.s.  
Jílové u Prahy

2006

## Obsah

<b>Úloha 1. Proved'te klasifikaci diskriminační analýzou DA. ....</b>	<b>3</b>
1.1. Popisné statistiky .....	6
1.2. Kovarianční matice .....	6
1.3. Korelační matice.....	6
1.4. Exploratorní analýza vícerozměrných dat EDA .....	7
1.5. Diskriminační analýza DA .....	10
1.5.1. Vyčíslení diskriminační funkce.....	10
1.5.2. Klasifikace kosatců.....	11
1.6. Klasifikace objektů kanonickou korelační analýzou.....	13
1.7. Zařazení neznámých objektů do tříd.....	15
1.8. Analýza shluků .....	16
1.9. Závěr .....	17
<b>Úloha 2. V úloze aplikujte buď logistickou regresi LR nebo kanonickou korelační analýzu CCA. ....</b>	<b>18</b>
2.1. Popisné statistiky .....	20
2.2. Kovarianční matice .....	20
2.3. Korelační matice.....	21
2.4. Exploratorní analýza vícerozměrných dat EDA .....	22
2.5. Zadání kanonických proměnných .....	24
2.6. Test významnosti kanonických kořenů .....	25
2.7. Struktura kanonických faktorů a redundance.....	25
2.8. Kanonické skóre .....	27
2.9. Grafy kanonických skóre .....	27
2.10. Analýza shluků .....	28
2.11. Závěr .....	28
<b>Úloha 3. Aplikujte metodu vícerozměrného škálování MDS a korespondenční analýzu CA kategoričkých dat. ....</b>	<b>29</b>
3.1. Popisné statistiky .....	29
3.2. Korelační matice.....	30
3.3. Exploratorní analýza vícerozměrných dat EDA .....	31
3.4. Mapování objektů vícerozměrným škálováním MDS.....	32
3.5. Korespondenční analýza CA.....	35
3.6. Analýza shluků .....	38
3.7. Závěr .....	38

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

**Úloha 1. Proved'te klasifikaci diskriminační analýzou DA.**

**Zadání:** U 124 kuřat byla stanovena celková živá hmotnost a hmotnosti jater, žaludku, srdce, prsních svalů a tkáně v oblasti krku. Hmotnosti byly stanovovány ve třech různých věkových kategoriích. Lze podle hmotností orgánů určit do jaké věkové kategorie patří? Zařad'te zvířata BB podle vytvořené funkce. (STATISTICA – data 3.5\_1.sta).

*Tab. 1: Hematologická data*

	Sex	Věk. skupina	Číslo zvířete	Hmotnost	Hm. jater	Hm. žaludek	Hm. srdce	Hm. prs. sv.	krk	Hm. ledvina
1	F	xx3	BB1	3430	36	46	12	775	85	13
2	F	xx1	BB2	1675	31	30	6	170	57	7
3	F	xx2	BB3	2235	36	41	7	370	75	8
4	F	xx2	BB4	2280	33	41	7	390	69	8
5	F	xx1	BB5	2320	34	46	7	382	55	10
6	F	xx3	BB6	3290	42	55	10	735	86	11
7	F	xx3	BB7	4080	53	38	12	825	100	10
8	M	xx3	BB8	3090	39	38	11	415	120	4
9	M	xx3	BB9	3645	59	53	13	500	185	8
10	M	3	129	4230	47	64	15	765	134	9
11	M	1	130	1980	30	55	8	250	67	6
12	M	3	133	5150	44	70	20	960	133	8
13	M	1	137	2325	31	40	9	334	79	6
14	M	2	138	2640	35	52	8	368	111	8
15	M	3	141	2600	28	41	8	410	85	5
16	M	3	143	4250	44	104	15	720	120	10
17	M	1	144	2280	33	46	7	327	79	7
18	M	3	149	3425	38	43	10	585	128	10
19	M	1	151	2180	34	45	7	280	100	9
20	M	1	153	1845	25	33	6	260	76	8
21	M	1	155	2725	37	58	10	388	110	9
22	M	1	156	2890	30	47	12	498	81	7
23	M	2	158	2760	37	49	14	396	99	7
24	M	1	160	2310	27	52	8	344	87	7
25	M	1	162	1820	26	45	7	225	71	8
26	M	1	164	1620	25	36	5	200	58	6
27	M	3	166	4470	52	53	17	845	141	13
28	M	2	170	2215	32	45	8	350	87	4
29	M	2	172	2975	40	48	9	430	120	11
30	M	3	176	4480	44	52	16	815	135	8
31	M	3	180	4830	47	52	18	920	140	9
32	M	3	184	5165	40	42	24	960	130	12
33	M	2	186	2765	41	52	9	365	110	6
34	M	2	188	2670	35	31	8	423	117	10
35	M	2	190	2620	32	61	9	380	75	6
36	M	3	192	3165	26	37	15	430	110	7
37	M	2	193	2430	38	44	10	387	101	7
38	M	2	196	2460	30	49	8	320	89	7
39	M	2	199	2060	29	53	8	250	67	6
40	F	3	201	3700	60	55	10	720	90	13
41	F	3	203	3470	41	37	11	785	80	10
42	F	3	206	4085	43	53	13	800	91	11
43	F	3	207	4125	52	44	15	935	98	12
44	F	1	209	1915	31	40	7	298	75	8

## SEMESTRÁLNÍ PRÁCE

### 3.5 Klasifikace analýzou vícerozměrných dat

	Sex	Věk. skupina	Číslo zvířete	Hmotnost	Hm. jater	Hm. žaludek	Hm. srdce	Hm. prs. sv.	krk	Hm. ledvina
45	F	1	212	2075	31	42	6	370	62	7
46	F	1	213	2000	33	45	7	280	59	8
47	F	3	216	3335	37	50	16	710	92	10
48	F	2	218	2975	31	48	12	542	88	8
49	F	1	220	1855	24	42	6	324	56	6
50	F	2	221	2815	33	48	9	477	92	8
51	F	1	224	2010	26	50	6	319	72	8
52	F	3	227	3325	37	59	12	615	84	10
53	F	1	229	1890	32	55	5	270	62	7
54	F	2	231	2520	30	49	9	454	83	5
55	F	1	232	1830	20	37	7	265	53	7
56	F	2	235	2250	27	34	7	444	63	7
57	F	3	237	3200	32	50	10	625	82	8
58	F	1	239	2320	33	49	11	404	65	7
59	F	3	241	3370	49	63	18	695	100	7
60	F	1	243	1685	34	41	5	215	60	6
61	F	2	249	2200	26	44	9	390	66	7
62	F	3	279	4275	41	58	18	930	105	8
63	F	1	281	2175	26	35	10	388	62	6
64	F	1	283	2020	29	41	7	343	60	7
65	F	1	284	1760	29	41	7	220	57	6
66	F	1	289	2260	35	40	7	334	73	9
67	F	3	292	4015	40	43	15	625	105	17
68	F	3	294	3485	32	39	15	1015	90	8
69	F	1	296	1970	30	42	7	288	59	7
70	F	1	298	1500	31	36	6	205	61	4
71	F	3	325	4475	35	46	13	910	105	8
72	M	1	330	2390	30	56	9	306	83	8
73	M	2	332	3400	39	62	11	498	110	8
74	M	3	335	4785	60	48	18	1010	135	11
75	M	2	336	2490	29	38	9	349	94	9
76	M	3	339	4505	68	57	15	815	145	10
77	F	2	341	2350	31	39	8	458	82	7
78	M	1	343	2030	32	41	7	256	72	6
79	M	3	347	5160	55	41	16	970	220	12
80	M	1	349	1840	29	37	6	236	65	7
81	F	3	352	4210	46	44	13	835	110	11
82	F	2	353	2820	43	55	10	570	82	9
83	F	2	357	1865	33	25	6	320	68	6
84	F	2	359	2900	38	42	10	545	83	8
85	F	2	362	2275	43	36	7	470	84	9
86	F	3	364	3915	47	44	12	800	80	10
87	F	2	368	2070	28	42	8	370	68	4
88	M	1	372	1910	34	45	7	25	69	6
89	F	3	374	4190	53	45	14	855	95	10
90	M	1	379	1530	19	36	6	135	66	8
91	M	1	380	1700	25	42	5	205	63	5
92	M	1	382	2150	32	56	6	275	66	6
93	M	1	384	2415	30	53	10	368	85	7
94	M	1	386	1846	30	39	7	135	53	7
95	M	1	390	1500	27	49	7	180	61	5
96	M	1	393	1665	24	38	5	160	63	6
97	F	2	398	2325	29	40	7	423	92	9
98	F	1	753	1600	25	36	6	215	55	5
99	F	2	754	2970	38	42	11	575	91	7
100	F	1	756	1885	32	32	6	215	65	8
101	F	3	764	3375	36	44	13	760	81	7

**SEMESTRÁLNÍ PRÁCE**  
**3.5 Klasifikace analýzou vícerozměrných dat**

	Sex	Věk. skupina	Číslo zvířete	Hmotnost	Hm. jater	Hm. žaludek	Hm. srdce	Hm. prs. sv.	krk	Hm. ledvina
102	M	3	768	4380	40	52	18	705	132	9
103	M	2	776	2945	46	49	11	387	94	11
104	F	3	778	2920	22	52	18	520	78	8
105	M	2	780	2450	19	60	9	350	79	6
106	M	3	783	5020	46	50	23	920	156	11
107	M	2	787	3025	33	47	10	486	107	7
108	M	2	792	2540	34	50	7	370	98	7
109	M	2	794	2470	36	43	8	280	79	5
110	M	2	795	2160	30	50	8	390	90	8
111	M	2	798	2310	26	33	9	330	72	7
112	M	3	800	3195	40	47	10	495	118	9
113	M	2	852	2800	37	47	9	444	88	6
114	F	2	852	2800	40	40	9	475	73	5
115	F	2	854	2940	36	42	9	588	88	6
116	F	2	856	1950	25	46	6	330	65	6
117	F	3	860	4130	54	43	13	870	87	10
118	F	3	861	3095	36	40	15	580	97	10
119	F	2	863	2570	33	65	10	450	77	7
120	F	3	867	3400	39	40	11	730	79	9
121	F	3	870	3280	37	41	10	745	75	9
122	F	2	872	1980	21	38	7	310	74	5
123	F	3	957	3355	38	35	9	755	77	8

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

**1.1. Popisné statistiky**

*Tab. 2: Popisné statistiky*

Skupina	Hm. jater	Hm. žaludek	Hm. srdce	Hm. prs. sv.	krk	Hm. ledvina	N
1	29	43	7	272	69	7	38
2	33	46	9	414	87	7	38
3	43	49	14	767	109	10	38
Celkově	35	46	10	484	88	8	114

*Tab. 3: Směrodatné odchytky*

Skupina	Hm. jater	Hm. žaludek	Hm. srdce	Hm. prs. sv.	krk	Hm. ledvina	N
1	4.05	6.99	1.63	88.98	12.53	1.22	38
2	6.02	8.52	1.70	83.27	14.86	1.71	38
3	9.55	12.10	3.67	157.75	29.77	2.11	38
Celkově	8.88	9.67	3.98	238.00	26.20	2.15	114

**1.2. Kovarianční matice**

*Tab. 4: Kovarianční matice*

	Hm. jater	Hm. žaludek	Hm. srdce	Hm. prs. sv.	krk	Hm. ledvina
Hm. jater	48	14	4	375	64	4.8
Hm. žaludek	14	89	7	119	45	0.2
Hm. srdce	4	7	6	147	25	0.8
Hm. prs. sv.	375	119	147	13245	829	52.3
krk	64	45	25	829	421	12.7
Hm. ledvina	5	0	1	52	13	3.0

**1.3. Korelační matice**

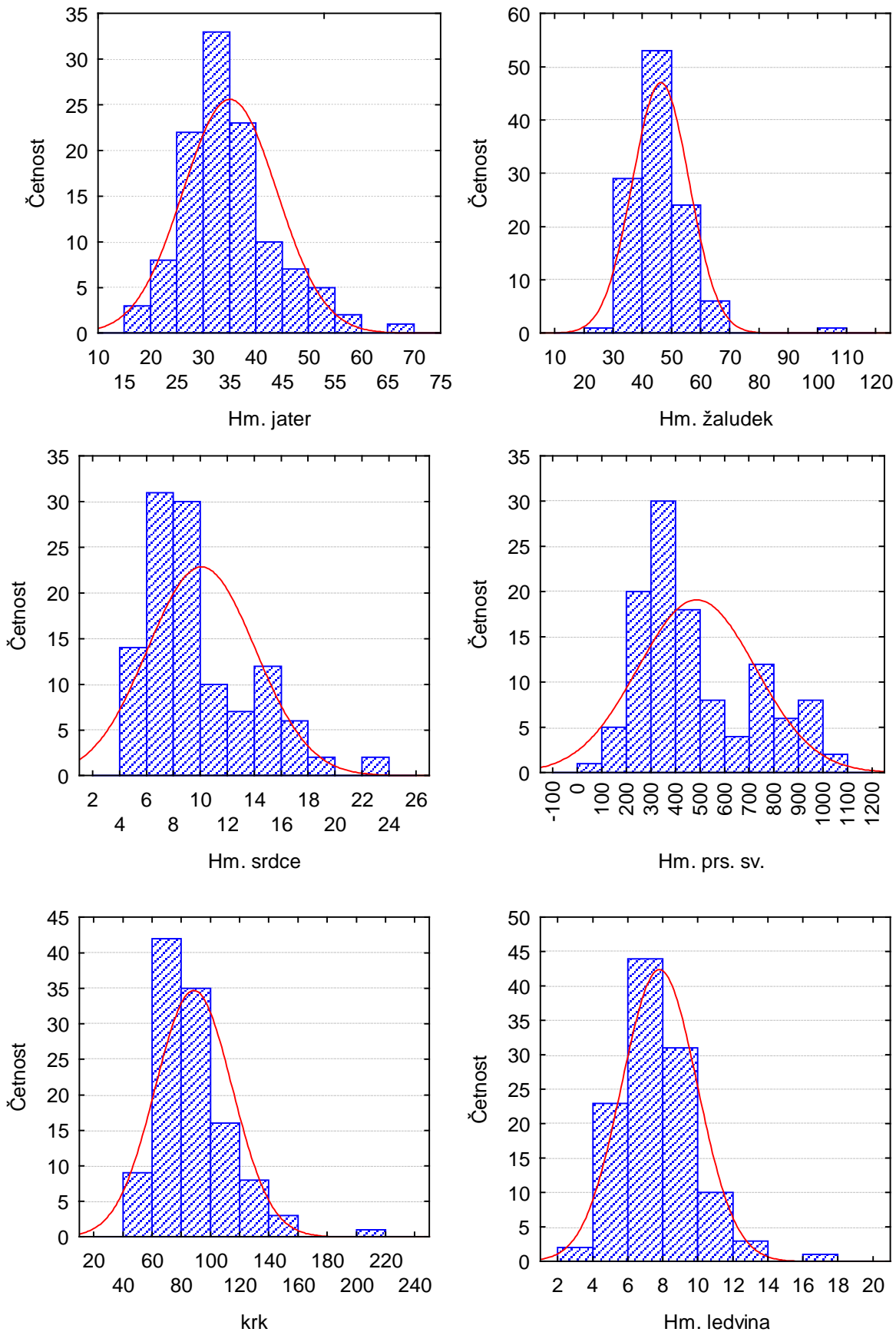
*Tab. 5: Korelační matice*

	Hm. jater	Hm. žaludek	Hm. srdce	Hm. prs. sv.	krk	Hm. ledvina
Hm. jater	1.00	0.21	0.22	0.47	0.45	0.40
Hm. žaludek	0.21	1.00	0.30	0.11	0.23	0.01
Hm. srdce	0.22	0.30	1.00	0.51	0.49	0.18
Hm. prs. sv.	0.47	0.11	0.51	1.00	0.35	0.26
krk	0.45	0.23	0.49	0.35	1.00	0.36
Hm. ledvina	0.40	0.01	0.18	0.26	0.36	1.00

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

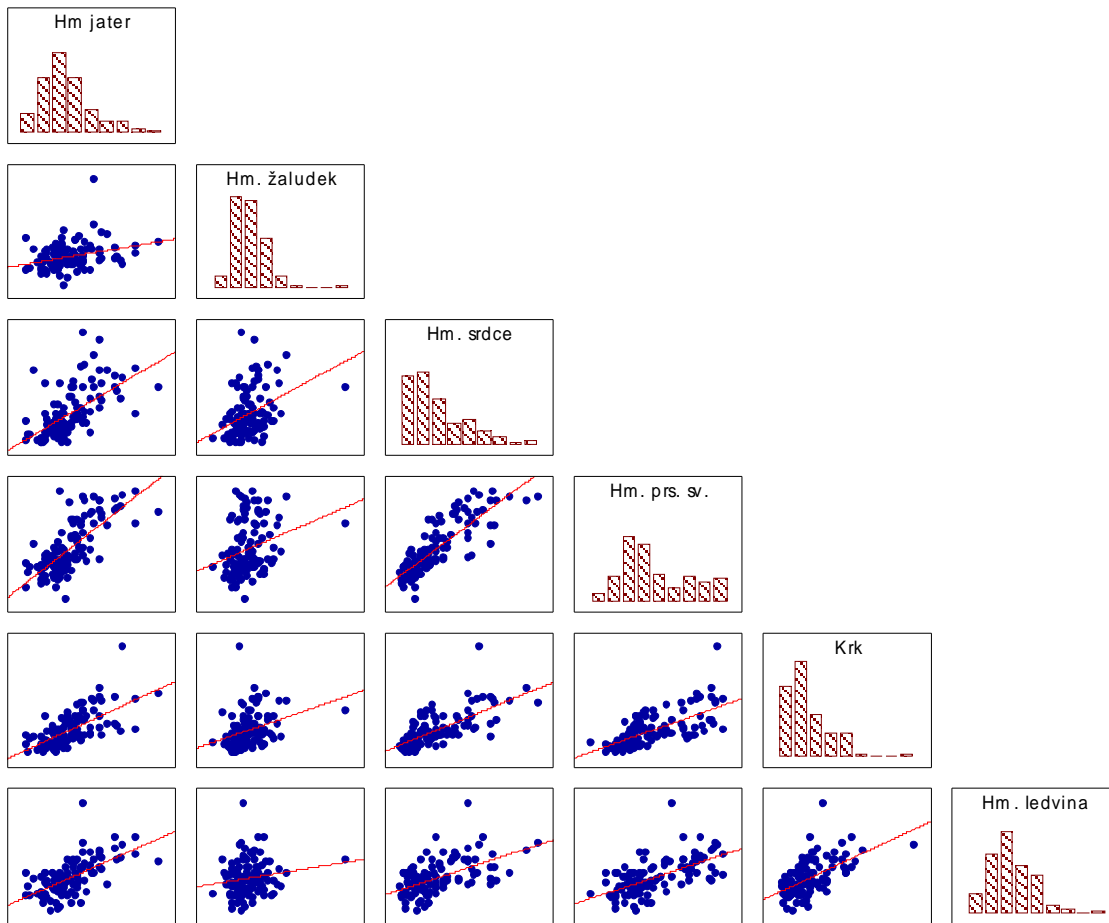
**1.4. Exploratorní analýza vícerozměrných dat EDA**

*Obr. 1: Histogram a jádrový odhad hustoty*



**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 2: Rozptylový diagram pro 6 znaků a 114 zvířatech

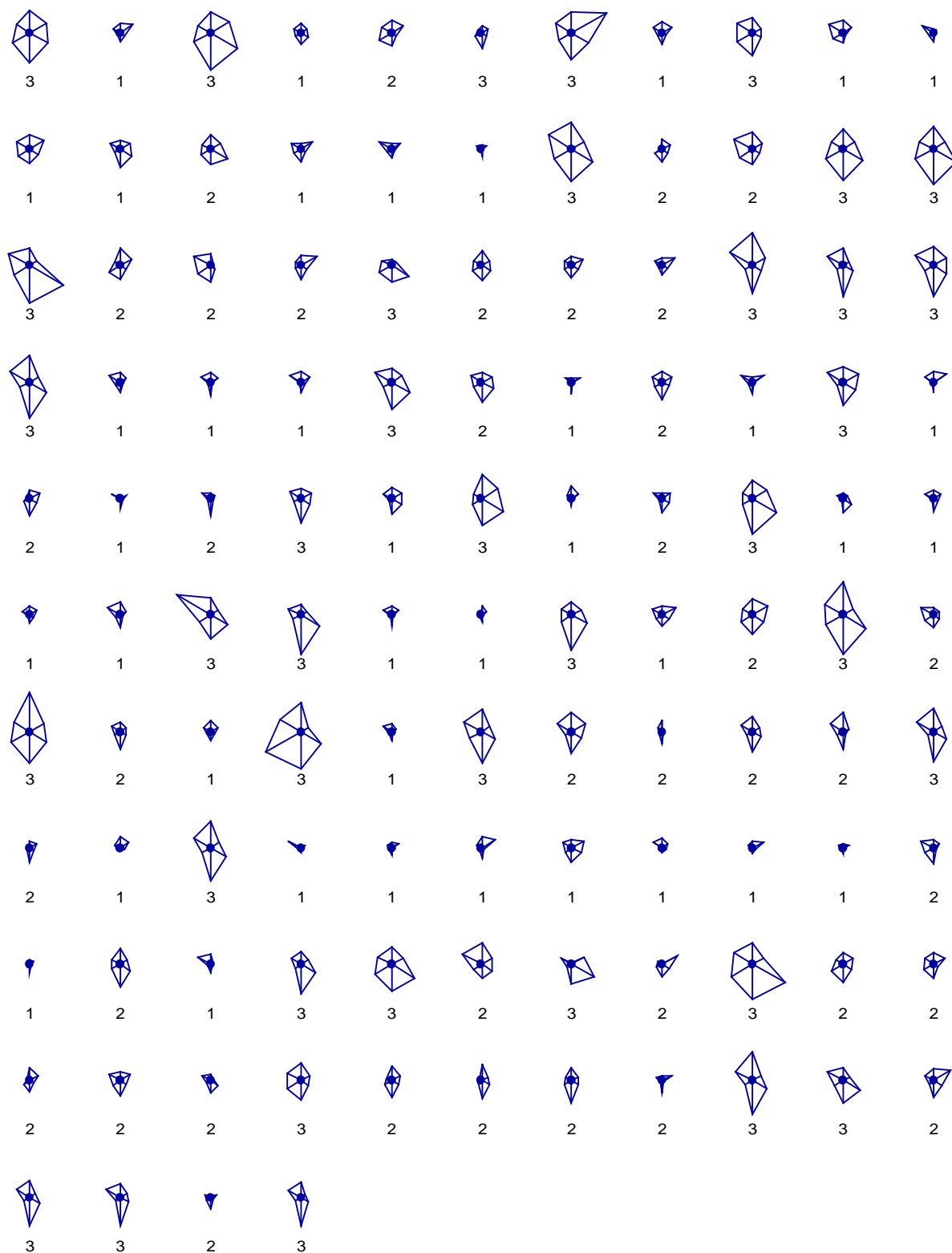


Rozptylový diagram vyjadřuje graficky závislost mezi jednotlivými proměnnými.



**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 3: Hvězdicový graf



\*Pravotočivě: Hm.jater, Hm.Žaludek, Hm.srdce, Hm.prs.sv., Krk, Hm.ledvina

Z hvězdicového grafu lze nejlépe odlišit většinu zvířat z 3 věkové skupiny – největší hvězdička.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

### 1.5. Diskriminační analýza DA

Analýzou hlavních komponent se snažíme vysvětlit závislost proměnných pomocí hlavních komponent.

Tab. 6: Vyšetření vlivu diskriminátorů

	Wilkovo $\lambda$	Parcial. $\lambda$	F-test	p-hladina	Toler.	1-Toler. - $R^2$
Hm. jater	0.1963	0.9965	0.19	0.829	0.6025	0.3975
Hm. žaludek	0.1957	0.9997	0.02	0.984	0.8562	0.1438
Hm. srdce	0.2055	0.9519	2.68	0.073	0.5727	0.4273
Hm. prs. sv.	0.3011	0.6497	28.58	0.000	0.5960	0.4040
krk	0.2093	0.9346	3.71	0.028	0.6106	0.3894
Hm. ledvina	0.2098	0.9324	3.84	0.024	0.7844	0.2156

Wilkovo  $\lambda$  testuje důsledek odstranění dotyčného diskriminátoru a vyjadřuje diskriminační sílu navrženého modelu. Parciální  $\lambda$  vyjadřuje příspěvek dané proměnné k diskriminační síle modelu. F test a p-hladina vyjadřuje hodnotu F kritéria a jeho pravděpodobnosti k testování statistické významnosti Wilkovo  $\lambda$ . V této úloze jsou statisticky významné pouze poslední tři proměnných. První dvě proměnné nemají vliv na diskriminační sílu modelu a proto by měly být z modelu vypuštěny.

Tab. 7: Vyšetření vlivu diskriminátorů – automatický výběr diskriminátorů

	Wilkovo $\lambda$	Partial $\lambda$	F-test	p-hladina	Toler.	1-Toler. - $R^2$
Hm. prs. sv.	0.3207	0.6123	34.1954	0.0000	0.7081	0.2919
Hm. ledvina	0.2119	0.9267	4.2715	0.0164	0.8448	0.1552
krk	0.2111	0.9301	4.0560	0.0200	0.6799	0.3201
Hm. srdce	0.2073	0.9475	2.9943	0.0542	0.6302	0.3698

Do modelu zahrneme 4 proměnné hmotnost prsního svalstva, hmotnost ledvin, krk, hmotnost srdce.

#### 1.5.1. Vyčíslení diskriminační funkce

Tab. 8: Vyčíslení diskriminační funkce

	1	2	3
Hm. prs. sv.	0.0030	0.0136	0.0352
Hm. ledvin	1.8002	1.5197	2.0727
krk	0.0698	0.1063	0.0729
Hm. srdce	0.5520	0.4552	0.8970
Abs. člen	-11.9848	-15.7833	-34.9436

Klasifikační funkce mají tvar

Pro 1 sk.:  $0.0030 * \text{Hm. prs. sv.} + 1.8002 * \text{Hm. Ledvin} + 0.0698 * \text{krk} + 0.5520 * \text{Hm. srdce} - 11.9848$

Pro 2 sk.:  $0.0136 * \text{Hm. prs. sv.} + 1.5197 * \text{Hm. Ledvin} + 0.1063 * \text{krk} + 0.4552 * \text{Hm. srdce} - 15.7833$

Pro 3 sk.:  $0.0352 * \text{Hm. prs. sv.} + 2.0727 * \text{Hm. Ledvin} + 0.0729 * \text{krk} + 0.8970 * \text{Hm. srdce} - 34.9436$

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

**1.5.2. Klasifikace kosatců**

Tab. 9: Klasifikační matice

Třída	Procenta	1	2	3
1	86.84	33	5	0
2	81.58	6	31	1
3	89.47	0	4	34
Total	85.96	39	40	35

Z klasifikační matice je patrné, že nejvíce nesprávně zařazených případů je u druhé skupiny, nejlépe jsou podle modelu zařazeni členové třetí skupiny.

Tab. 10: Zařazení zvířat do skupin

Číslo zvířete	Daná třída	Nalezená třída			Mahalanobisovy vzdálenosti			Procento zařazení do třídy		
		1	2	3	1	2	3	1	2	3
129	3	3	2	1	23.6	13.0	4.5	0.0	1.4	98.6
130	1	1	2	3	1.8	4.3	22.9	77.3	22.7	0.0
133	3	3	2	1	45.4	31.2	9.9	0.0	0.0	100.0
*137	1	2	1	3	2.0	2.0	16.3	50.0	50.0	0.0
138	2	2	1	3	5.8	3.1	17.6	20.7	79.2	0.1
*141	3	2	1	3	5.6	2.3	16.0	16.8	83.1	0.1
143	3	3	2	1	53.6	45.2	37.2	0.0	1.8	98.2
144	1	1	2	3	0.6	1.1	16.1	55.9	44.1	0.0
*149	3	2	3	1	15.5	8.4	10.2	2.0	69.5	28.5
151	1	1	2	3	4.2	5.0	21.6	59.4	40.6	0.0
153	1	1	2	3	3.1	5.0	22.6	72.4	27.6	0.0
*155	1	2	1	3	5.6	3.5	14.6	26.2	73.6	0.3
*156	1	2	1	3	5.6	2.9	6.5	18.1	70.4	11.5
158	2	2	1	3	9.8	8.7	16.3	36.8	61.8	1.4
*160	1	2	1	3	3.2	2.3	16.5	39.6	60.4	0.1
162	1	1	2	3	1.9	5.3	23.3	84.7	15.3	0.0
164	1	1	2	3	1.4	4.6	27.2	83.6	16.4	0.0
166	3	3	2	1	32.6	22.4	4.8	0.0	0.0	100.0
170	2	2	1	3	5.8	3.6	20.8	25.3	74.7	0.0
172	2	2	1	3	9.6	7.0	15.2	21.5	77.2	1.3
176	3	3	2	1	26.4	14.5	3.3	0.0	0.4	99.6
180	3	3	2	1	35.9	22.0	4.4	0.0	0.0	100.0
184	3	3	2	1	62.1	50.9	21.3	0.0	0.0	100.0
186	2	2	1	3	7.9	5.1	21.3	19.4	80.6	0.0
188	2	2	1	3	11.3	8.4	17.6	18.8	80.4	0.8
190	2	2	1	3	4.6	3.5	16.7	36.7	63.3	0.1
*192	3	2	1	3	15.4	12.5	17.8	18.0	76.6	5.5
193	2	2	1	3	4.4	2.2	14.7	25.1	74.8	0.1
196	2	2	1	3	1.6	1.1	16.8	43.4	56.6	0.0
*199	2	1	2	3	1.3	3.7	22.1	77.3	22.7	0.0
201	3	3	2	1	30.7	26.2	17.3	0.1	1.1	98.7
203	3	3	2	1	23.0	15.4	5.0	0.0	0.5	99.4
206	3	3	2	1	24.5	16.9	3.5	0.0	0.1	99.9
207	3	3	2	1	36.8	26.9	6.0	0.0	0.0	100.0
209	1	1	2	3	0.7	2.2	17.9	68.1	31.9	0.0
212	1	1	2	3	1.7	2.2	15.2	55.6	44.3	0.1
213	1	1	2	3	1.6	5.0	20.3	84.4	15.6	0.0
216	3	3	2	1	19.9	14.1	2.9	0.0	0.4	99.6

## SEMESTRÁLNÍ PRÁCE

### 3.5 Klasifikace analýzou vícerozměrných dat

Číslo zvířete	Daná třída	Nalezená třída			Mahalanobisovy vzdálenosti			Procento zařazení do třídy		
		1	2	3	1	2	3	1	2	3
218	2	2	3	1	7.3	3.3	4.8	8.7	61.9	29.4
220	1	1	2	3	2.8	3.5	20.0	58.7	41.3	0.0
221	2	2	1	3	4.6	1.0	8.9	14.0	84.4	1.6
224	1	1	2	3	5.0	5.9	20.9	61.4	38.6	0.0
227	3	3	2	1	13.3	9.4	5.3	1.6	11.4	87.0
229	1	1	2	3	3.2	5.4	24.7	74.2	25.8	0.0
231	2	2	1	3	6.6	2.7	14.0	12.3	87.4	0.3
232	1	1	2	3	2.6	5.4	22.3	80.0	20.0	0.0
235	2	2	1	3	5.5	4.0	13.5	31.7	67.7	0.6
237	3	3	2	1	11.8	6.2	5.0	2.1	35.4	62.5
239	1	1	2	3	4.3	4.5	13.0	52.8	46.5	0.7
241	3	3	2	1	26.5	19.8	12.4	0.1	2.5	97.5
243	1	1	2	3	2.9	6.2	28.7	83.5	16.5	0.0
*249	2	1	2	3	2.5	2.6	12.2	51.4	48.2	0.4
279	3	3	2	1	38.6	26.3	6.9	0.0	0.0	100.0
*281	1	2	1	3	4.6	4.5	14.8	49.3	50.4	0.3
283	1	1	2	3	0.9	2.0	15.6	63.5	36.5	0.0
284	1	1	2	3	1.0	4.8	24.5	87.1	12.9	0.0
289	1	1	2	3	1.6	3.0	16.5	66.8	33.2	0.0
292	3	3	2	1	44.2	42.9	29.1	0.1	0.1	99.8
294	3	3	2	1	51.4	37.8	14.8	0.0	0.0	100.0
296	1	1	2	3	0.4	2.8	19.3	76.8	23.2	0.0
298	1	1	2	3	5.0	7.5	31.9	77.0	23.0	0.0
325	3	3	2	1	35.9	22.8	8.2	0.0	0.1	99.9
330	1	1	2	3	2.3	3.3	17.7	61.4	38.6	0.0
332	2	2	1	3	8.1	3.4	9.8	8.4	87.9	3.7
335	3	3	2	1	45.6	31.7	8.3	0.0	0.0	100.0
336	2	2	1	3	3.4	2.9	15.1	44.0	55.9	0.1
339	3	3	2	1	38.3	27.2	17.4	0.0	0.7	99.3
341	2	2	1	3	3.8	1.0	10.0	19.7	79.4	0.9
343	1	1	2	3	0.7	2.8	21.5	74.1	25.9	0.0
347	3	3	2	1	75.8	56.2	39.5	0.0	0.0	100.0
349	1	1	2	3	0.7	3.6	23.1	81.4	18.6	0.0
352	3	3	2	1	25.1	15.2	1.8	0.0	0.1	99.9
353	2	2	3	1	9.1	6.0	6.1	9.7	46.3	43.9
*357	2	1	2	3	6.2	6.6	24.3	55.3	44.7	0.0
359	2	2	3	1	5.9	2.4	5.1	12.2	69.5	18.3
362	2	2	1	3	7.0	5.0	13.3	26.8	72.0	1.1
364	3	3	2	1	24.4	17.4	4.6	0.0	0.2	99.8
368	2	2	1	3	5.7	3.9	20.2	29.2	70.8	0.0
372	1	1	2	3	10.6	17.9	47.5	97.5	2.5	0.0
374	3	3	2	1	28.7	19.9	4.7	0.0	0.1	99.9
379	1	1	2	3	5.4	10.8	33.0	93.4	6.6	0.0
380	1	1	2	3	1.3	3.9	27.0	78.2	21.8	0.0
382	1	1	2	3	3.2	4.5	24.2	65.9	34.1	0.0
*384	1	2	1	3	2.2	1.6	12.9	42.5	57.2	0.2
386	1	1	2	3	4.7	11.3	33.4	96.4	3.6	0.0
390	1	1	2	3	1.6	5.3	28.3	86.2	13.8	0.0
393	1	1	2	3	1.6	5.3	30.0	86.6	13.4	0.0
398	2	2	1	3	6.0	3.6	13.8	23.8	75.8	0.5
753	1	1	2	3	2.2	5.0	27.4	80.0	20.0	0.0
754	2	2	3	1	8.0	2.9	5.0	5.5	69.9	24.6
756	1	1	2	3	3.1	7.4	26.3	89.6	10.4	0.0
764	3	3	2	1	20.2	12.2	3.3	0.0	1.1	98.9
768	3	3	2	1	23.4	14.8	5.3	0.0	0.9	99.1
*776	2	1	2	3	10.4	11.7	18.3	65.2	33.6	1.3
778	3	3	2	1	24.1	22.1	18.7	5.5	14.6	80.0
780	2	2	1	3	9.8	8.3	22.5	32.8	67.2	0.1
783	3	3	2	1	49.2	36.4	12.8	0.0	0.0	100.0
787	2	2	1	3	7.0	1.9	10.1	7.1	91.4	1.5
792	2	2	1	3	3.8	1.8	16.4	26.1	73.9	0.0

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Číslo zvířete	Daná třída	Nalezená třída			Mahalanobisovy vzdálenosti			Procento zařazení do třídy		
		1	2	3	1	2	3	1	2	3
*794	2	1	2	3	5.0	5.4	25.5	54.6	45.4	0.0
795	2	2	1	3	3.0	1.6	12.8	32.8	67.0	0.2
*798	2	1	2	3	2.8	3.7	16.4	61.3	38.7	0.1
*800	3	2	1	3	7.8	3.2	9.6	8.7	87.7	3.6
852	2	2	1	3	4.9	1.7	12.9	16.7	83.0	0.3
852	2	2	1	3	9.6	6.5	16.7	17.4	82.1	0.5
854	2	2	3	1	10.6	4.3	8.9	3.7	87.5	8.9
*856	2	1	2	3	2.7	2.8	19.1	51.2	48.8	0.0
860	3	3	2	1	31.8	23.3	7.8	0.0	0.0	100.0
861	3	3	2	1	13.2	9.6	5.5	1.9	11.2	86.9
863	2	2	1	3	7.5	5.6	13.6	26.9	71.8	1.3
867	3	3	2	1	17.3	10.6	3.3	0.1	2.5	97.4
870	3	3	2	1	20.5	13.5	5.8	0.1	2.1	97.8
872	2	2	1	3	4.3	3.9	21.4	44.4	55.6	0.0
957	3	3	2	1	21.6	13.7	7.2	0.1	3.7	96.2

### 1.6. Klasifikace objektů kanonickou korelační analýzou

Tab. 11: Tabulka vlastních vektorů pro 4 hlavní komponenty

	Vlastní číslo	Kan. korelace	Wilkovo $\lambda$	$\chi^2$	sv	Úroveň p
0	3.55	0.883	0.196	178.23	8	0.000
1	0.12	0.327	0.893	12.35	3	0.006

Oba kanonické kořeny jsou statisticky významné.

Tab. 12: Odhady koeficientů znaků u kanonických proměnných

	1. Kan. proměnná	2. Kan. proměnná
Hm. prs. sv.	0.0072	-0.0029
Hm. ledvin	0.0856	0.4696
krk	-0.0017	-0.0476
Hm. srdce	0.0911	0.2499
Abs. člen	-4.9208	-0.5712
Vlastní	3.5492	0.1194
KumPodíl	0.9675	1.0000

Tab. 13: Standardizované kanonické koeficienty znaků u kanonických proměnných

	1. Kan. Proměnná	2. Kan. Proměnná
Hm. prs. sv.	0.8287	-0.3310
Hm. ledvin	0.1475	0.8090
krk	-0.0346	-0.9772
Hm. srdce	0.2295	0.6295
Vlastní	3.5492	0.1194
KumPodíl	0.9675	1.0000

Standardizované koeficienty se používají k porovnávání objektů. První diskriminační funkce je nejvíce vážena proměnnou hmotnosti prsní svaloviny, druhá krkem.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Tab. 14: Koeficienty faktorové struktury

	1. Kan. Proměnná	2. Kan. Proměnná
Hm. prs. sv.	0.9717	-0.1417
Hm. ledvin	0.3943	0.4815
krk	0.4214	-0.4945
Hm. srdce	0.6583	0.1276

Tab. 15: Průměry kanonických proměnných

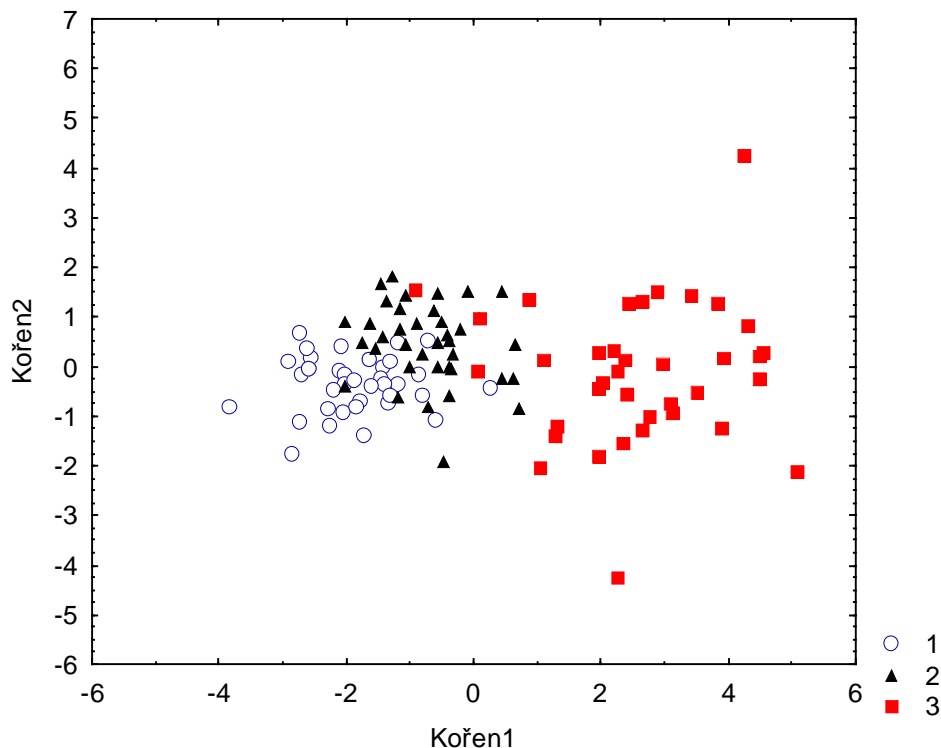
Hm. třída	1. Kan. Proměnná	2. Kan. Proměnná
1	-1.8485	0.3428
2	-0.6947	-0.4650
3	2.5432	0.1222

Tab. 16: Kanonické scóre všech objektů

Číslo zvířete	Daná třída	Scóre třídy G1	Scóre třídy G2	Číslo zvířete	Daná třída	Scóre třídy G1	Scóre třídy G2	Číslo zvířete	Daná třída	Scóre třídy G1	Scóre třídy G2
129	3	2.487	-1.270	218	2	0.567	0.311	364	3	2.706	1.193
130	1	-1.992	0.385	220	1	-1.684	-0.052	368	2	-1.378	-1.320
133	3	4.308	-0.731	221	2	-0.224	-0.677	372	1	-3.695	0.651
*137	1	-1.288	-0.152	224	1	-1.516	0.408	374	3	3.194	0.593
138	2	-1.087	-1.393	227	3	1.303	1.285	379	1	-2.830	1.225
*141	3	-0.967	-1.516	229	1	-2.043	0.102	380	1	-2.613	-0.369
143	3	2.237	-0.045	231	2	-0.585	-1.443	382	1	-2.016	-0.382
144	1	-1.410	0.001	232	1	-1.936	0.909	*384	1	-0.887	0.216
*149	3	0.795	-1.292	235	2	-0.638	0.064	386	1	-2.774	1.650
151	1	-1.657	-0.103	237	3	1.037	-0.022	390	1	-2.663	0.184
153	1	-1.945	0.319	239	1	-0.554	1.025	393	1	-2.913	-0.086
*155	1	-0.692	-0.493	241	3	2.099	0.292	398	2	-0.663	-0.403
*156	1	0.211	0.508	243	1	-2.502	-0.062	753	1	-2.543	-0.199
158	2	-0.324	0.486	*249	2	-0.772	0.873	754	2	0.669	-0.498
*160	1	-1.253	-0.355	279	3	3.910	-0.145	756	1	-2.225	1.108
162	1	-2.090	1.008	*281	1	-0.832	0.578	764	3	2.230	0.091
164	1	-2.616	0.037	283	1	-1.304	0.576	768	3	2.349	-0.131
166	3	3.538	0.485	284	1	-2.247	0.796	*776	2	-0.347	1.778
170	2	-1.437	-1.715	289	1	-1.300	0.682	778	3	0.943	2.190
172	2	-0.312	-0.246	292	3	2.193	4.309	780	2	-1.254	-0.525
176	3	2.867	-1.494	294	3	4.328	-0.103	783	3	4.443	0.219
180	3	3.832	-1.293	296	1	-1.755	0.684	787	2	-0.133	-1.488
184	3	4.984	2.119	298	1	-2.685	-0.799	792	2	-1.152	-1.127
186	2	-1.190	-1.917	325	3	3.297	-1.332	*794	2	-1.945	-1.053
188	2	-0.486	-0.639	330	1	-1.397	0.419	795	2	-0.831	-0.162
190	2	-1.027	-0.374	332	2	0.125	-0.946	*798	2	-1.228	0.625
*192	3	-0.026	0.084	335	3	4.649	-0.479	*800	3	0.089	-0.970
193	2	-0.836	-0.942	336	2	-1.064	0.065	852	2	-0.580	-1.182
196	2	-1.437	-0.561	339	3	2.867	-1.579	852	2	-0.418	-1.051
*199	2	-1.986	0.441	341	2	-0.433	-0.506	854	2	0.453	-1.510
201	3	2.126	1.639	343	1	-1.977	0.110	*856	2	-1.595	-0.285
203	3	2.379	0.513	347	3	4.166	-4.272	860	3	3.219	0.769
206	3	2.766	1.075	349	1	-2.175	0.404	861	3	1.246	1.349
207	3	3.954	1.121	352	3	2.972	-0.081	863	2	-0.320	0.146
209	1	-1.606	0.400	353	2	0.790	0.844	867	3	1.942	0.398
212	1	-1.179	0.348	*357	2	-1.719	-0.621	870	3	1.966	0.381
213	1	-1.662	1.348	359	2	0.461	0.189	872	2	-1.746	-0.772
216	3	2.305	1.576	362	2	-0.315	-0.132	957	3	1.912	-0.281

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 4: Klasifikační graf kanonických kořenů



Z grafu je patrné, že první kanonický kořen rozděluje skupinu poměrně jednoznačně. Druhá skupin leží mezi 1 a 3. Třetí skupina má největší variabilitu. Druhý kanonický kořen spíše rozděluje 1 a 2 skupinu.

### 1.7. Zařazení neznámých objektů do tříd

Tab. 17: Zařazení neznámých zvířat do věkových skupin

Číslo zvířete	Daná třída	Nalezená třída			Mahalanobisovy vzdálenosti			Procento zařazení do třídy		
		1	2	3	1	2	3	1	2	3
BB1	xx3	3	2	1	27.5	21.9	7.6	0.0	0.1	99.9
BB2	xx1	1	2	3	1.0	5.7	27.8	91.5	8.5	0.0
BB3*	xx2	1	2	3	1.8	2.1	14.2	53.2	46.7	0.1
BB4*	xx2	1	2	3	2.1	2.5	12.9	54.5	45.3	0.2
BB5	xx1	1	2	3	7.2	9.6	18.0	77.2	22.5	0.3
BB6	xx3	3	2	1	19.9	13.5	5.4	0.1	1.8	98.2
BB7	xx3	3	2	1	23.4	14.0	1.6	0.0	0.2	99.8
BB8	xx2	2	1	3	14.5	8.8	23.2	5.6	94.3	0.1
BB9	xx2	2	3	1	34.6	25.1	33.6	0.8	97.8	1.4

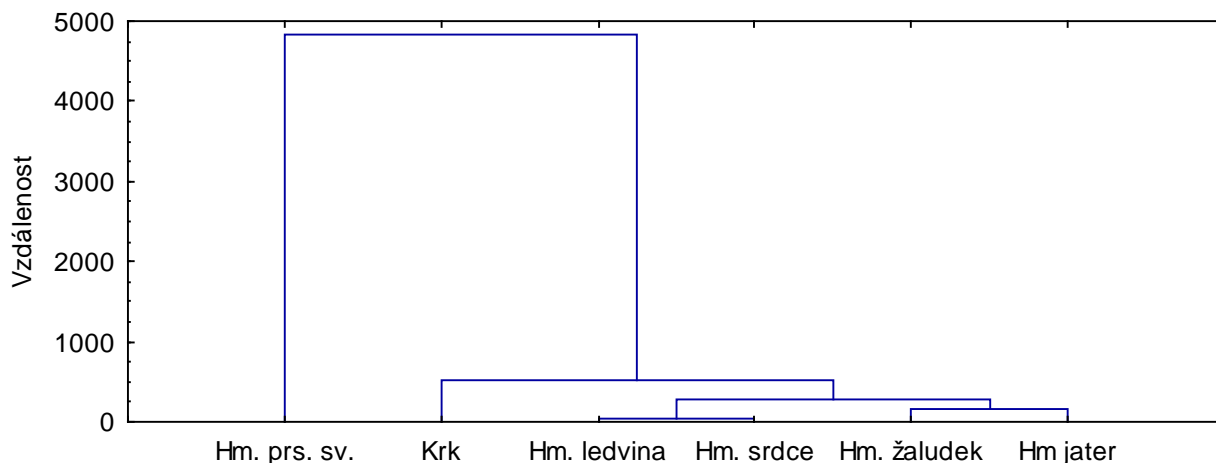
Dva objekty z druhé skupiny byly nesprávně zařazený do první věkové kategorie.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

**1.8. Analýza shluků**

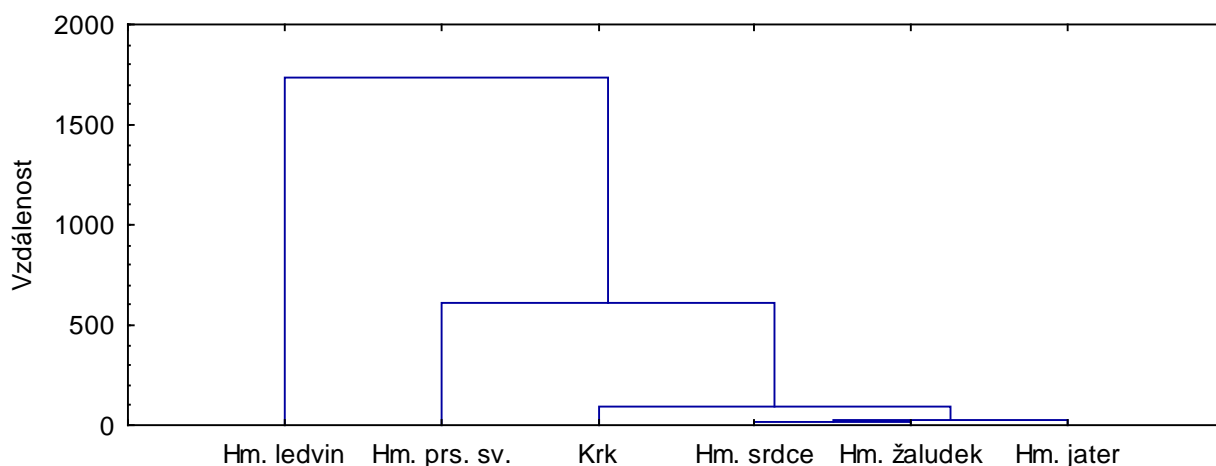
Analýza shluků se zabývá klasifikací dat pomocí shluků. Byly vytvořeny dendrogramy metodou nejbližšího souseda s použitím eukleidovské vzdálenosti.

*Obr. 5: Dendrogram – nstandardizovaná data*



Dendrogram z nstandardizovaných dat vyjadřuje, které stanovované hmotnosti jsou si nejpodobnější. Hmotnost ledvin je podobná hmotnosti srdce, hmotnost žaludku je podobná hmotnosti jater, nejvyšší hmotnost mají prsní svaly drůbeže.

*Obr. 6: Dendrogram – standardizace sloupcovým centrováním*



Dendrogram z dat standardizovaných sloupcovým centrováním vyjadřuje mezi rozptylem hmotností (velikost rozdílu hmotností v celé skupině). Nejmenší rozdíly v hmotnostech jsou u srdce, žaludku a jater. Největší rozptyl je pozorován u hmotnosti ledvin.



**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

**1.9. Závěr**

Diskriminační analýzou bylo stanoveno, že lze zvířata dělit do věkových kategorií podle hmotnosti prsního svalstva, ledvin, krku a srdce. Klasifikace podle diskriminační funkce zařadí správně 86 % zvířat. Největší vzdálenost od ostatních skupin má skupina 3 s největší pravděpodobností správného zařazení 89,5 %. Naopak nejmenší pravděpodobnost 81,6 % je u skupiny 2, která leží mezi skupinami 1 a 3. Dále lze skupinu charakterizovat 2 kanonickými proměnnými. Pro první proměnnou je hlavní hmotnost prsního svalstva a pro druhou jsou určující hmotnosti krku a ledvin.

Z 9 neznámých vzorků bylo správně zařazeno 7 zvířat. 2 zvířata z druhé skupiny byly nesprávně zařazeny do skupiny 1.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Úloha 2. V úloze aplikujte buď logistickou regresi LR nebo kanonickou korelační analýzu CCA.

**Zadání:** Před začátkem studie bylo 95 kuřatům provedeno stanovení základních biochemických parametrů, stanoveny hmotnosti, spotřeby krmiva a vody u jednotlivých kuřat. Určete jaký je vztah mezi základními biochemickými parametry a hmotností, spotřebou krmiva a vody. (STATISTICA – data3.5\_2.sta).

Tab. 18: Naměřené teploty

	AST (μkat/L)	TP (G/L)	ALB (G/L)	UA (μmol/L)	hmotnost (kg)	Spotřeba krmiva (kg)	Spotřeba vody (mL)
1	3.0	29	25	161	0.80	0.103	139
2	3.7	42	24	371	0.56	0.048	53
3	3.4	33	28	431	0.64	0.057	150
4	3.2	28	23	212	0.68	0.063	97
5	3.3	24	20	587	0.74	0.084	70
6	3.5	30	23	442	0.59	0.053	59
7	3.8	31	26	905	0.79	0.103	209
8	3.5	25	21	332	0.58	0.059	150
9	3.8	29	25	489	0.59	0.049	113
10	3.6	29	25	84	0.69	0.058	95
11	2.9	30	26	293	0.74	0.053	72
12	3.0	29	24	316	0.69	0.070	120
13	5.2	28	25	446	0.77	0.113	136
14	3.5	33	25	444	0.65	0.035	58
15	3.9	30	24	208	0.66	0.079	167
16	3.4	31	24	233	0.71	0.051	127
17	3.5	34	29	393	0.57	0.063	169
18	3.4	26	23	348	0.67	0.072	165
19	3.8	28	24	333	0.79	0.101	160
20	3.2	29	24	85	0.55	0.050	165
21	2.7	23	20	712	0.65	0.093	80
22	3.3	28	24	271	0.64	0.100	187
23	3.4	28	24	309	0.50	0.049	200
24	3.8	33	28	199	0.72	0.064	140
25	3.5	33	27	577	0.77	0.087	170
26	3.8	31	25	282	0.61	0.047	96
27	3.2	30	27	371	0.77	0.044	87
28	3.4	28	25	186	0.57	0.064	114
29	3.0	28	24	562	0.64	0.088	105
30	3.2	35	24	131	0.64	0.071	197
31	3.4	32	26	614	0.73	0.077	205
32	3.4	30	25	535	0.65	0.060	125
33	3.4	29	24	182	0.72	0.084	76
34	3.5	27	23	641	0.71	0.057	55
35	3.3	35	30	239	0.70	0.064	70
36	3.3	28	24	442	0.72	0.077	170
37	3.7	32	28	346	0.73	0.078	192
38	3.7	30	23	241	0.70	0.049	50
39	3.2	28	24	858	0.63	0.068	164
40	2.6	31	24	378	0.62	0.091	148
41	3.1	26	23	328	0.56	0.088	110
42	3.1	33	26	469	0.67	0.080	200
43	3.6	30	26	361	0.74	0.065	130
44	3.6	26	22	213	0.59	0.066	133
45	3.6	29	22	153	0.72	0.080	204
46	3.2	28	24	451	0.60	0.090	140
47	3.8	32	25	182	0.57	0.036	80
48	3.2	30	26	228	0.70	0.062	184
49	2.7	28	23	190	1.28	0.147	247
50	3.1	48	21	207	0.76	0.087	164
51	3.4	27	24	176	0.97	0.085	140
52	3.2	29	25	267	1.03	0.132	218

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

	AST (μkat/L)	TP (G/L)	ALB (G/L)	UA (μmol/L)	hmotnost (kg)	Spotřeba krmiva (kg)	Spotřeba vody (mL)
53	2.9	21	20	169	1.08	0.142	218
54	3.2	29	24	141	0.94	0.121	229
55	3.6	29	22	274	1.12	0.183	284
56	3.2	27	23	130	0.91	0.106	155
57	3.4	28	23	188	0.96	0.101	155
58	3.3	30	25	143	1.03	0.126	252
59	2.8	30	26	288	1.05	0.136	209
60	3.2	30	26	209	0.95	0.143	180
61	3.8	29	26	250	1.27	0.151	223
62	4.2	30	26	163	0.78	0.082	110
63	2.9	29	25	185	1.04	0.132	192
64	3.3	33	26	207	0.96	0.117	177
65	3.1	32	26	195	0.96	0.144	267
66	3.2	26	24	249	1.03	0.140	203
67	3.9	28	23	151	1.14	0.128	162
68	3.1	30	26	172	0.91	0.124	200
69	2.8	24	21	251	1.00	0.118	199
70	3.3	27	25	222	1.02	0.112	162
71	3.4	30	22	328	0.81	0.137	191
72	3.5	29	25	189	1.28	0.120	153
73	2.9	27	24	213	1.22	0.143	215
74	3.6	32	26	186	0.79	0.171	110
75	3.0	28	26	218	1.04	0.160	160
76	4.4	28	25	142	0.99	0.147	322
77	3.1	33	25	127	0.95	0.099	211
78	3.0	29	23	255	1.06	0.114	166
79	3.0	33	27	136	1.14	0.119	189
80	3.4	29	24	232	0.91	0.101	116
81	3.6	29	24	115	1.01	0.133	161
82	3.3	30	25	140	1.02	0.097	136
83	3.3	33	28	163	0.81	0.109	155
84	3.2	30	26	164	1.17	0.156	234
85	3.2	28	25	165	1.18	0.141	170
86	3.7	30	24	215	0.87	0.127	161
87	2.9	25	23	121	1.03	0.129	149
88	2.6	29	23	220	1.00	0.111	186
89	2.6	26	24	75	0.91	0.131	209
90	3.0	28	24	251	1.09	0.145	173
91	3.2	30	26	230	1.15	0.140	172
92	3.4	29	25	183	0.87	0.119	139
93	3.3	33	27	171	1.02	0.099	175
94	3.5	30	25	99	1.01	0.132	188
95	3.3	32	26	106	0.88	0.104	138

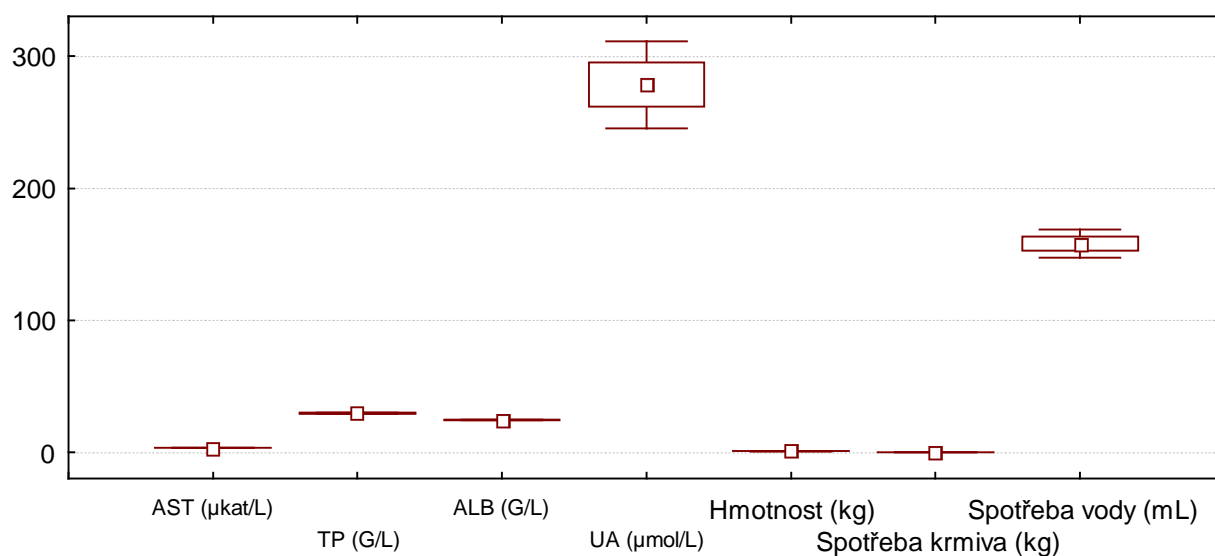
**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

## 2.1. Popisné statistiky

Tab. 19: Popisné statistiky

	AST ( $\mu\text{kat/L}$ )	TP (G/L)	ALB (G/L)	UA ( $\mu\text{mol/L}$ )	Hmotnost (kg)	Spotřeba krmiva (kg)	Spotřeba vody (mL)
Means	3.3	30	25	278	0.84	0.098	158
Std.Dev.	0.39	3.42	1.87	162.87	0.20	0.035	53.36

Obr. 7: Krabicový graf



## 2.2. Kovarianční matice

Tab. 20: Kovarianční matice

	AST ( $\mu\text{kat/L}$ )	TP (G/L)	ALB (G/L)	UA ( $\mu\text{mol/L}$ )	hmotnost (kg)	Spotřeba krmiva (kg)	Spotřeba vody (mL)
AST ( $\mu\text{kat/L}$ )	1.00	0.12	0.17	0.09	-0.21	-0.19	-0.12
TP (G/L)	0.12	1.00	0.44	-0.04	-0.18	-0.22	-0.07
ALB (G/L)	0.17	0.44	1.00	-0.05	0.02	-0.06	0.04
UA ( $\mu\text{mol/L}$ )	0.09	-0.04	-0.05	1.00	-0.42	-0.34	-0.24
Hmotnost (kg)	-0.21	-0.18	0.02	-0.42	1.00	0.82	0.55
Spotřeba krmiva (kg)	-0.19	-0.22	-0.06	-0.34	0.82	1.00	0.65
Spotřeba vody (mL)	-0.12	-0.07	0.04	-0.24	0.55	0.65	1.00

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

### 2.3. Korelační matice

Tab. 21: Korelační matice

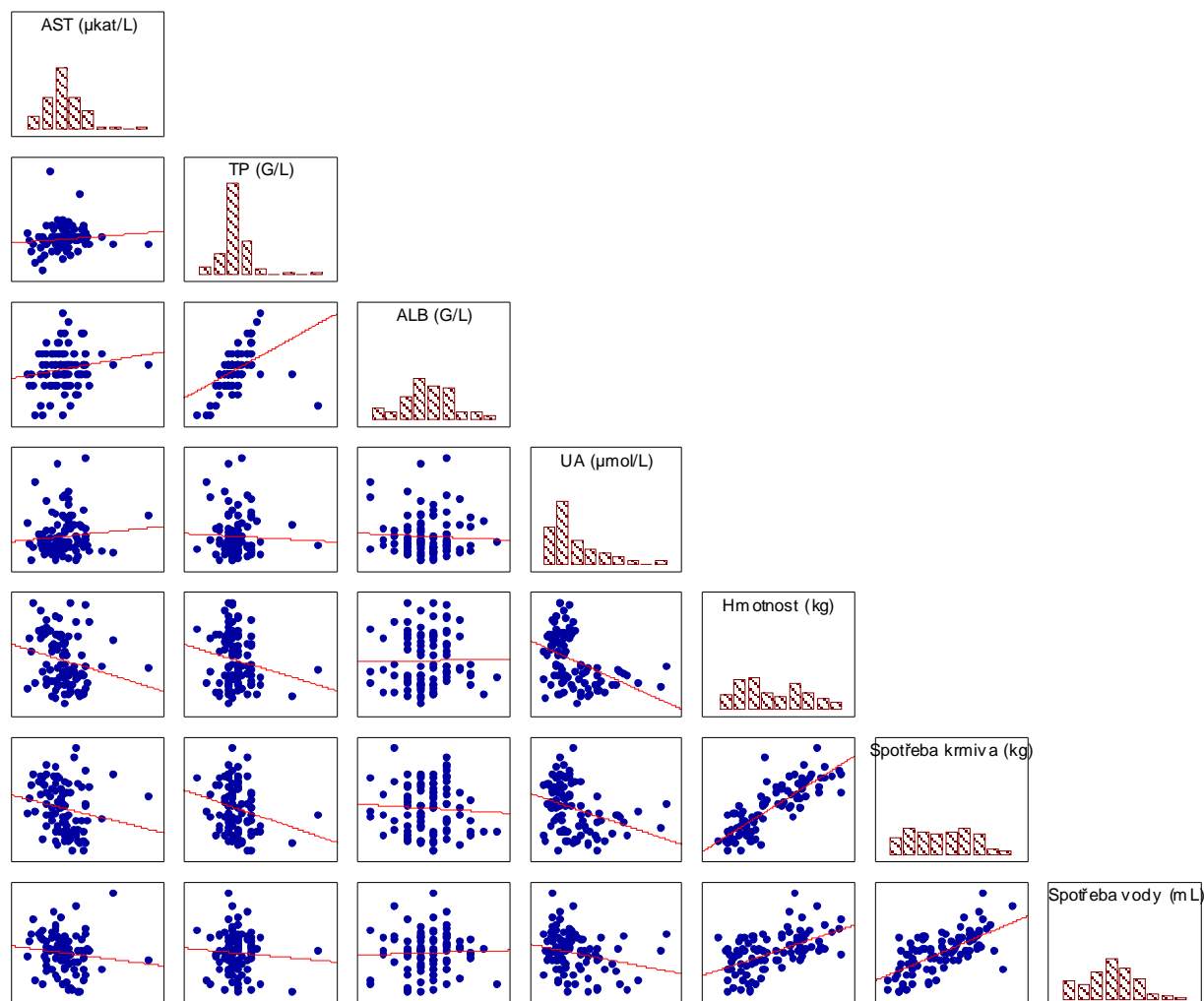
	AST ( $\mu$ kat/L)	TP (G/L)	ALB (G/L)	UA ( $\mu$ mol/L)	Hmotnost (kg)	Spotřeba krmiva (kg)	Spotřeba vody (mL)
AST ( $\mu$ kat/L)	1.00	0.12	0.17	0.09	-0.21	-0.19	-0.12
TP (G/L)	0.12	1.00	0.44	-0.04	-0.18	-0.22	-0.07
ALB (G/L)	0.17	0.44	1.00	-0.05	0.02	-0.06	0.04
UA ( $\mu$ mol/L)	0.09	-0.04	-0.05	1.00	-0.42	-0.34	-0.24
Hmotnost (kg)	-0.21	-0.18	0.02	-0.42	1.00	0.82	0.55
Spotřeba krmiva (kg)	-0.19	-0.22	-0.06	-0.34	0.82	1.00	0.65
Spotřeba vody (mL)	-0.12	-0.07	0.04	-0.24	0.55	0.65	1.00

Z korelační matice je patrná silná závislost mezi hmotností, spotřebou krmiva a vody.

SEMESTRÁLNÍ PRÁCE  
3.5 Klasifikace analýzou vícerozměrných dat

2.4. Exploratorní analýza vícerozměrných dat EDA

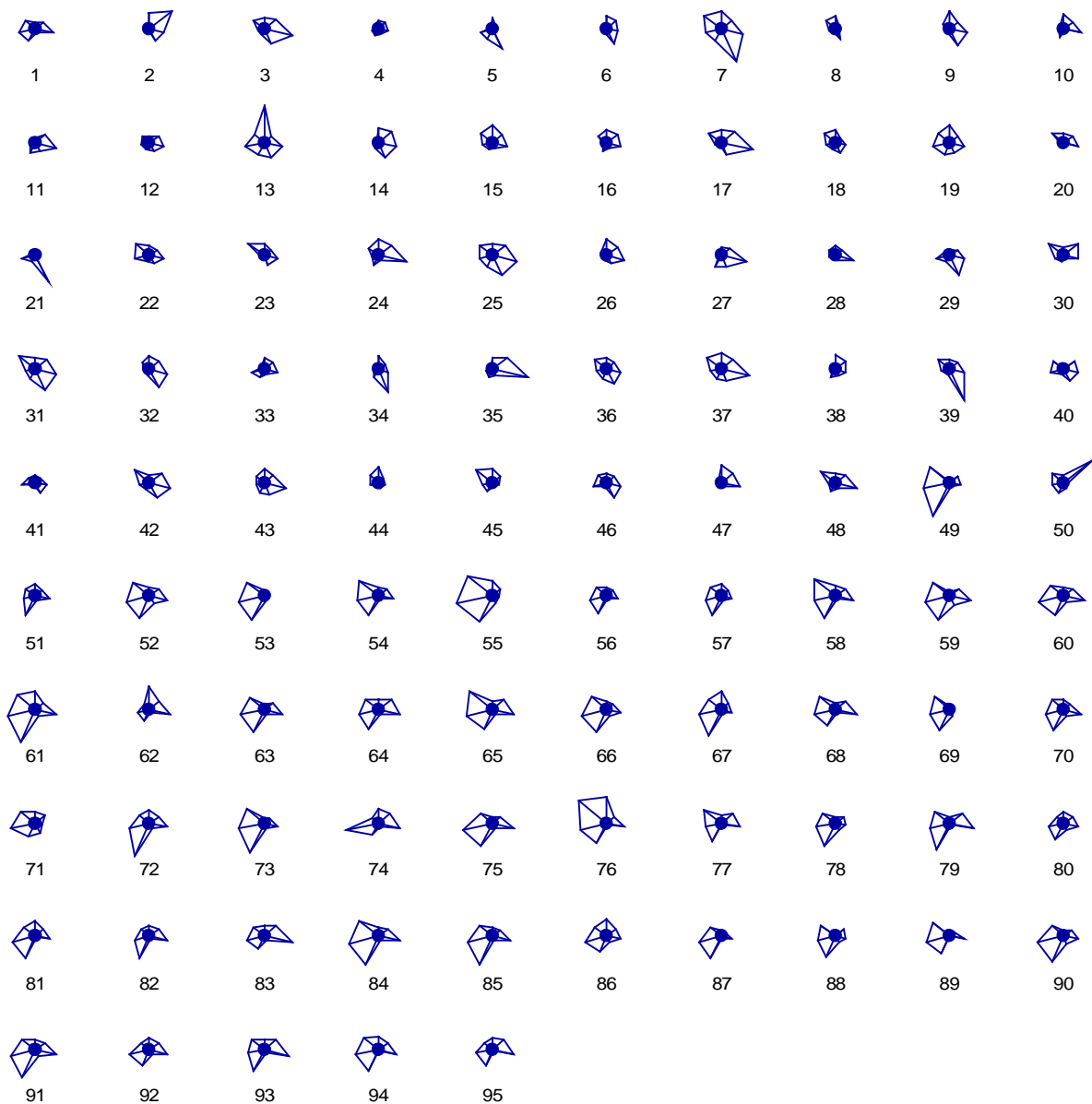
Obr. 8: Rozptylový diagram pro 95 objektů a 7 proměnných



Rozptylový diagram vyjadřuje graficky korelaci mezi jednotlivými proměnnými. Z tohoto grafu je patrná korelace mezi hmotností, spotřebou krmiva a vody.

SEMESTRÁLNÍ PRÁCE  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 9: Hvězdicový graf

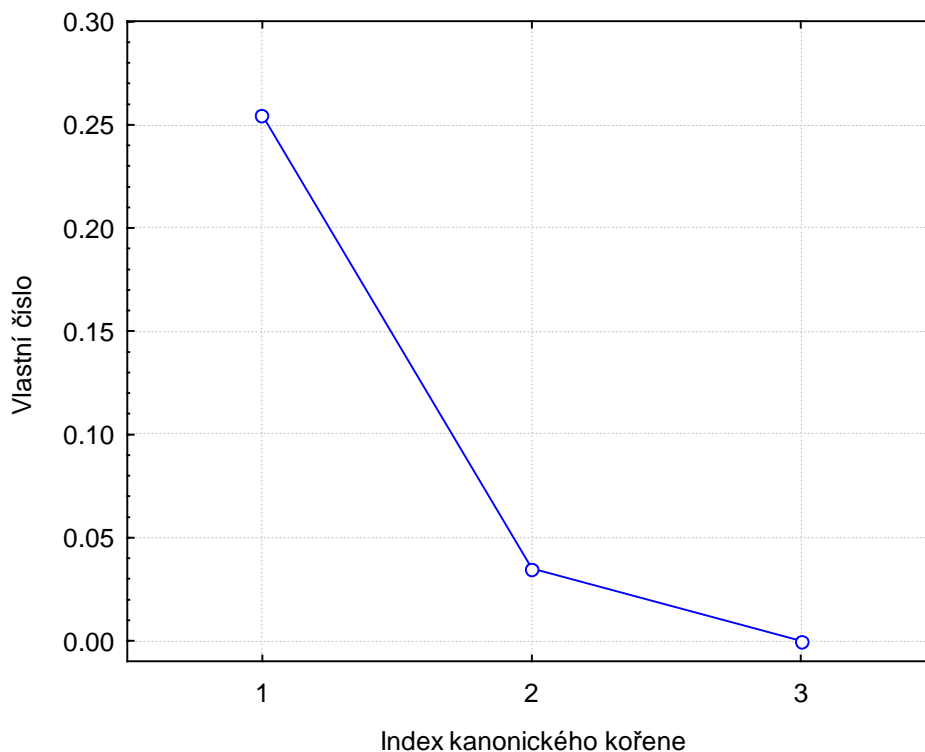


\*Pravotočivě: AST (μkat/L), TP (G/L), ALB (G/L), UA (μmol/L), Hmotnost (kg), Spotřeba krmiva (kg), Spotřeba vody (mL)

Z hvězdicovitého grafu je patrné zvláštní zvíře č. 13,50 a jiná.

## 2.5. Zadání kanonických proměnných

Obr. 10: Cattellův indexový graf úpatí vlastních čísel



Z grafu úpatí vlastních čísel lze určit kolik dvojic kanonických proměnných stačí k popisu dat. V tomto případě bude postačovat jedna dvojice.

Tab. 22: tabulka vlastních čísel

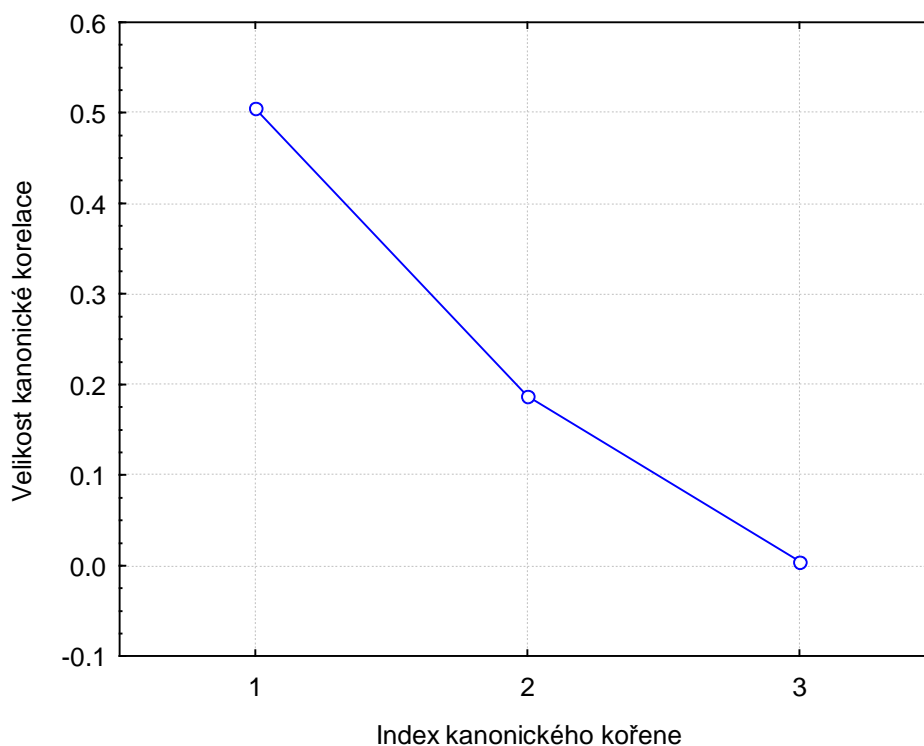
	Levý soubor závislých kanonických proměnných U	Pravý soubor nezávislých kanonických proměnných V
Počet znaků	4	3
Rozptyl extrahovaný	84.3 %	100.0 %
Celková redundance	7.7 %	17.4 %
Znaky:		
1	AST ( $\mu\text{kat/L}$ )	Hmotnost (kg)
2	TP (G/L)	Spotřeba krmiva (kg)
3	ALB (G/L)	Spotřeba vody (mL)
4	UA ( $\mu\text{mol/L}$ )	

Z extrahovaného rozptylu lze stanovit, že 3 dvojicemi kanonických proměnných lze vysvětlit 84 % dat levého souboru a 100 % pravého souboru. Celková redundance určuje celkovou korelaci v souborech pravých a levých proměnných. Nízká redundance ukazuje na nízkou korelaci v jednotlivých souborech



## 2.6. Test významnosti kanonických kořenů

Obr. 11: Graf kanonických korelací



Tab. 23: Test významnosti kanonických kořenů

Vypuštěný kořen	Kanonické R	Kanonické R <sup>2</sup>	$\chi^2$	df	p	Lambda
žádný	0.505	0.255	29.66	12	0.003	0.719
První kořen	0.187	0.035	3.21	6	0.782	0.965
První a druhý kořen	0.005	0.000	0.00	2	0.999	1.000

Kanonická korelace vyjádřená hodnotou kanonického R = 0,505 je statisticky významná, protože hodnota  $p < 0.05$ . Protože pro ostatní kořeny nejsou testy statisticky významné, lze určit, že statisticky významný je pouze první kanonický kořen.

## 2.7. Struktura kanonických faktorů a redundance

Tab. 24: Tabulka zátěží, korelací s kanonickou proměnnou U

	Kořen 1, U <sub>1</sub>	Kořen 2, U <sub>2</sub>	Kořen 3, U <sub>3</sub>
AST (μkat/L)	-0.424	-0.051	0.898
TP (G/L)	-0.362	-0.798	-0.032
ALB (G/L)	0.025	-0.864	0.090
UA (μmol/L)	-0.838	0.218	-0.328

Jednotlivé znaky korelují s různými kanonickými kořeny – AST s kořenem 3, TP a ALB s kořenem 2, UA s kořenem 1.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

*Tab. 25: Extrahovaný rozptyl a redundance u kanonické proměnné U*

	Rozptyl	Redundance.
Kořen 1, $U_1$	0.253	0.065
Kořen 2, $U_2$	0.358	0.013
Kořen 3, $U_3$	0.231	0.000

Kanonické kořeny vysvětlí každý od 23 do 36 % rozptylu všech položek biochemických parametrů.

*Tab. 26: Tabulka zátěží, korelací s kanonickou proměnnou V*

	Kořen 1, $V_1$	Kořen 2, $V_2$	Kořen 3, $V_3$
Hmotnost (kg)	0.999	-0.026	0.020
Spotřeba krmiva (kg)	0.840	0.429	-0.331
Spotřeba vody (mL)	0.557	-0.202	-0.805

Znaky hmotnosti a spotřeby krmiva korelují s prvním kořenem, spotřeba vody koreluje se třetím kořenem.

*Tab. 27: Extrahovaný rozptyl a redundance u kanonické proměnné V*

	Rozptyl	Redundance.
Kořen 1, $V_1$	0.672	0.171
Kořen 2, $V_2$	0.075	0.003
Kořen 3, $V_3$	0.253	0.000

První kanonický kořen vyčíslí 67 % rozptylu znaků na pravé straně a 17 % rozptylů ostatních proměnných.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

**2.8. Kanonické skóre**

Tab. 28: Tabulka vah znaků k výpočtu skóre u kanonické proměnné U

	Kořen 1, $U_1$	Kořen 2, $U_2$	Kořen 3, $U_3$
AST ( $\mu\text{kat/L}$ )	-0.338	0.102	0.957
TP (G/L)	-0.462	-0.521	-0.148
ALB (G/L)	0.242	-0.644	-0.025
UA ( $\mu\text{mol/L}$ )	-0.815	0.154	-0.420

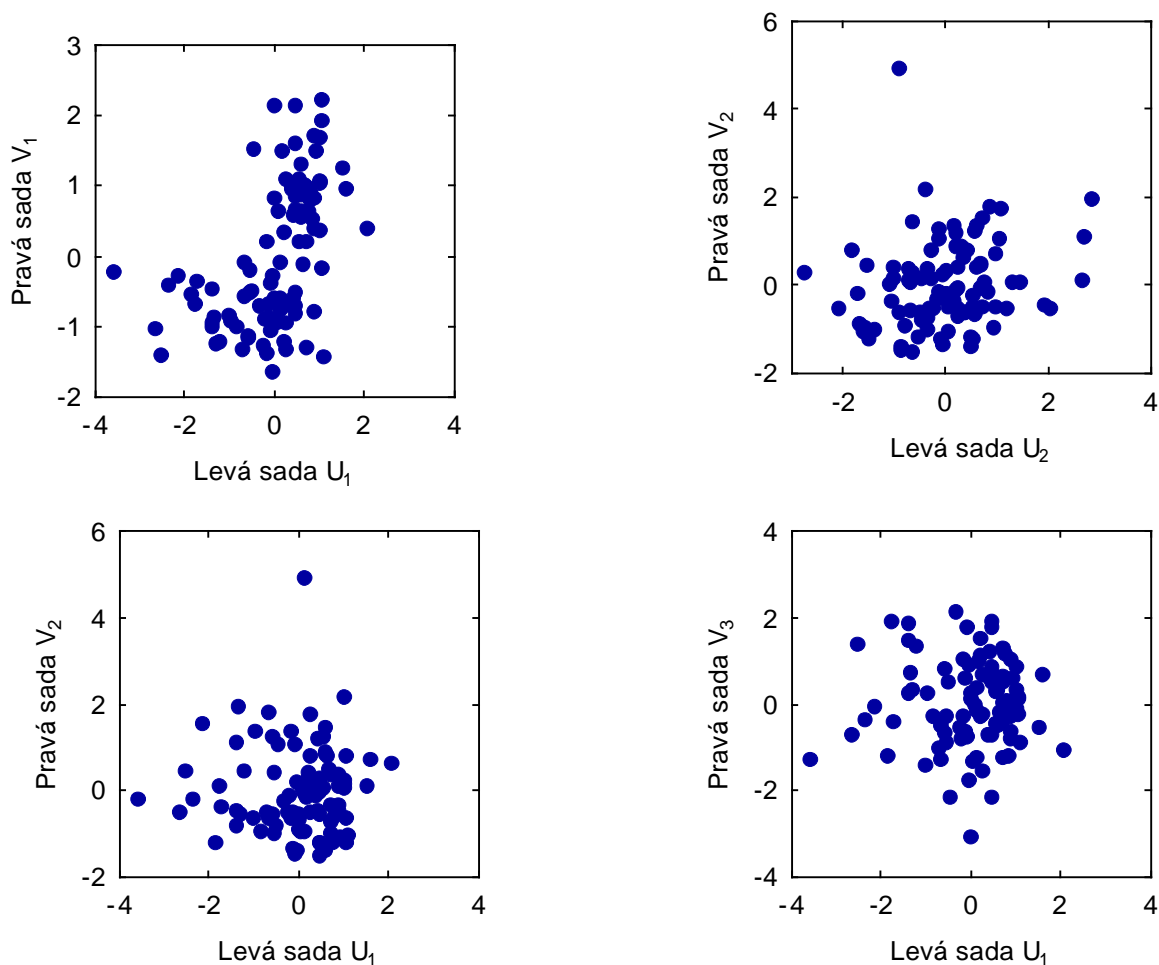
Tab. 29: Tabulka vah znaků k výpočtu skóre u kanonické proměnné V

	Kořen 1, $V_1$	Kořen 2, $V_2$	Kořen 3, $V_3$
Hmotnost (kg)	0.952	-1.136	0.944
Spotřeba krmiva (kg)	0.057	1.884	-0.433
Spotřeba vody (mL)	0.001	-0.803	-1.040

$-0.338 * \text{AST} - 0.462 * \text{TP} + 0.242 * \text{UA} = 0.952 * \text{hmotnost} + 0.057 * \text{spotřeba krmiva} + 0.001 * \text{spotřeba vody}$ ;  $R = 0.505$

**2.9. Grafy kanonických skóre**

Obr. 12: Rozptylové diagramy kanonických korelací

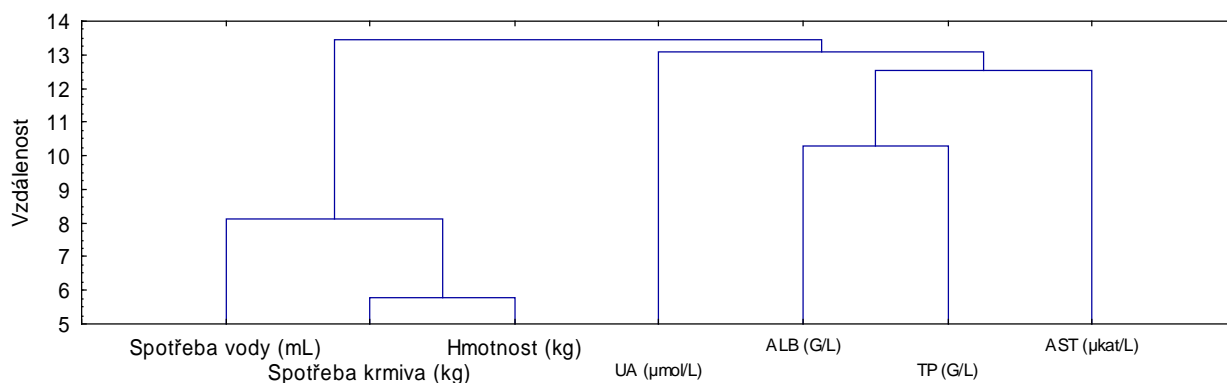


**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

### 2.10. Analýza shluků

Před tvorbou dendrogramů byla data standardizována.

Obr. 13: Dendrogram – metoda nejbližšího souseda – eukleidovská vzdálenost



Z tohoto dendrogramu standardizovaných dat lze určit, větší podobnost je mezi proměnnými na levé straně kanonické rovnice. A dále, že vzdálenosti uvnitř skupin jsou menší než mezi skupinami.

### 2.11. Závěr

Závěrem této úlohy lze říci, že kanonický korelační koeficient kořene první kanonické proměnné  $U_1$  a  $V_1$  je statisticky významný. Tento kořen vysvětluje 25 % rozptylu na levé straně a 67 % rozptylu na pravé straně.

Z kanonické analýzy lze vyvodit, že hmotnost je ovlivněna obsahem močové kyseliny v krvi a naopak na ní nemá vliv celkový obsah proteinů a aspartátaminotransferázy. Spotřeba krmiva a vody není ovlivněna žádným biochemickým parametrem.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Úloha 3. Aplikujte metodu vícerozměrného škálování MDS a korespondenční analýzu CA kategorických dat.

**Zadání:** Byla provedena studie na 32 prasatech. Zvířata byla rozdělena do 4 skupin podle podání látky: A bez podání léčiva, B podání standardní množství léčiva po dobu 5 dnů, C podání trojnásobnou dávku po dobu 5 dnů, D podání standardní dávky léčiva po dobu 15 dní. Po ukončení pokusu na jednotlivých zvířatech byly vyhodnocovány histopatologické nálezy – nálezy byly zařazeny k jednotlivým parametrům - 2 parametry u jater, tři u ledvin a jeden parametr pro plíce. Na základě více rozměrného škálování určete jestli jaká je podobnost mezi jednotlivými histopatologickými vyšetřeními bez ohledu na dávku podání léčiva. Korespondenční analýzou určete, pro kterou skupinu je významný, který histopatologický nálezy (jaká dávka má největší vliv na jednotlivý histopatologický nálezy) (STATISTICA – data3.5\_3a.sta, data3.5\_3b.smx).

Tab. 30: Tabulka histopatologických nálezy (počet nálezy pro jednotlivé skupiny)

	Játra 1	Játra 2	Ledviny 1	Ledviny 2	Ledviny 3	Plíce
A	3	2	6	1	2	1
B	0	0	3	0	0	0
C	4	3	2	0	1	0
D	3	3	4	0	0	0

Tab. 31: Tabulka podobnosti histopatologických nálezy (podle množství nálezy ve všech skupinách)

	Játra 1	Játra 2	Ledviny 1	Ledviny 2	Ledviny 3	Plíce
Játra 1	0	2	5	9	7	9
Játra 2	2	0	7	7	5	7
Ledviny 1	5	7	0	14	12	14
Ledviny 2	9	7	14	0	2	0
Ledviny 3	7	5	12	2	0	2
Plíce	9	7	14	0	2	0

### 3.1. Popisné statistiky

Tab. 32: Popisné statistiky – histopatologické nálezy

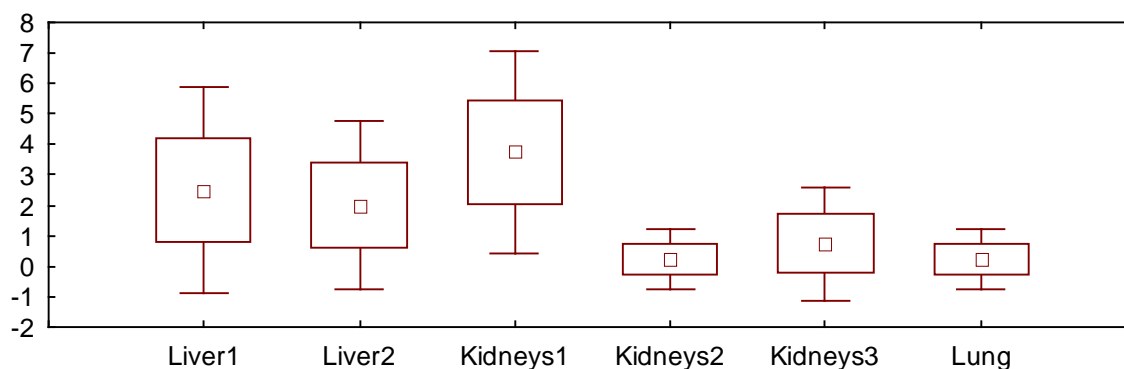
	Játra 1	Játra 2	Ledviny 1	Ledviny 2	Ledviny 3	Plíce
N	4	4	4	4	4	4
Průměr	2.50	2.00	3.75	0.25	0.75	0.25
Sm. odch.	1.73	1.41	1.71	0.50	0.96	0.50

Tab. 33: Popisné statistiky – skupiny

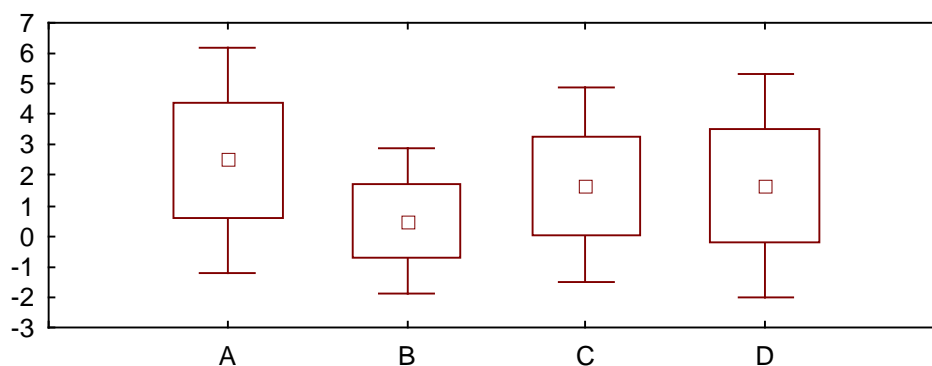
	A	B	C	D
N	6	6	6	6
Průměr	2.50	0.50	1.67	1.67
Sm. odch.	1.87	1.22	1.63	1.86

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 14: Krabicový graf – histopatologické nálezy



Obr. 15: Krabicový graf – skupiny



Z popisných statistik lze usuzovat na posobnost histopatologických vyšetření ledvin 2 a plic. Skupina B se odlišuje od ostatních skupin.

### 3.2. Korelační matice

Tab. 34: Korelační matice

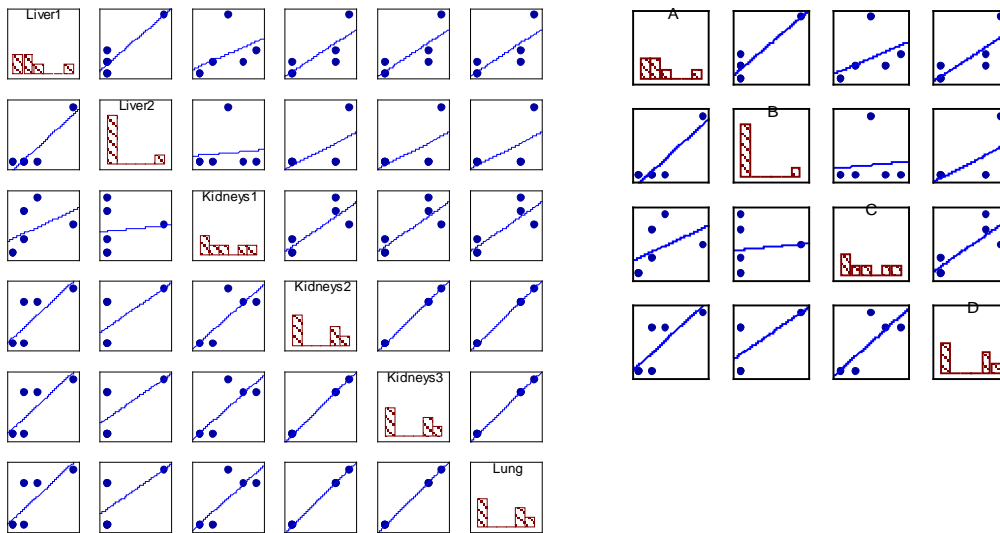
	Játra 1	Játra 2	Ledviny 1	Ledviny 2	Ledviny 3	Plíce
Játra 1	1.00	0.95	0.06	0.19	0.50	0.19
Játra 2	0.95	1.00	0.00	0.00	0.25	0.00
Ledviny 1	0.06	0.00	1.00	0.88	0.56	0.88
Ledviny 2	0.19	0.00	0.88	1.00	0.87	1.00
Ledviny 3	0.50	0.25	0.56	0.87	1.00	0.87
Plíce	0.19	0.00	0.88	1.00	0.87	1.00

Z korelační matice je vidět značnou míru podobnosti mezi histopatologickým vyšetřením jater a mezi vyšetřením ledvin 1, 2 a plic.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

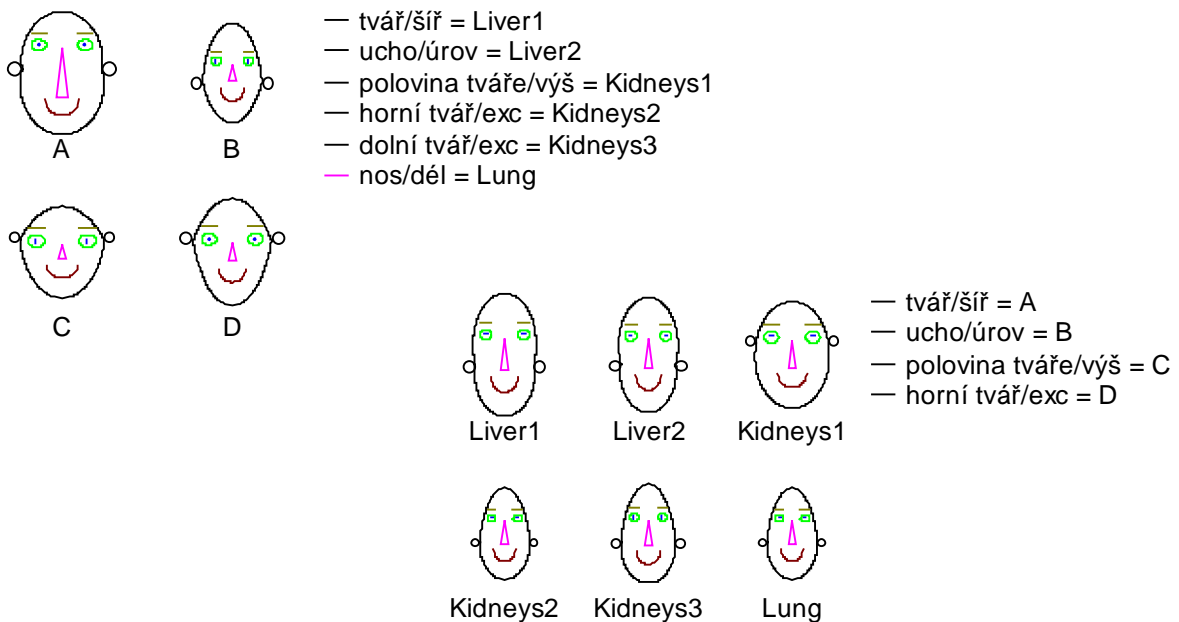
**3.3. Exploratorní analýza vícerozměrných dat EDA**

*Obr. 16: Rozptylový diagram pro histologické nálezy a pro skupiny*



Rozptylový diagram vyjadřuje graficky korelaci mezi jednotlivými proměnnými. Vzhledem k malému počtu kategoričkých dat není možno určit odlehle body ani přesnou závislost mezi proměnnými.

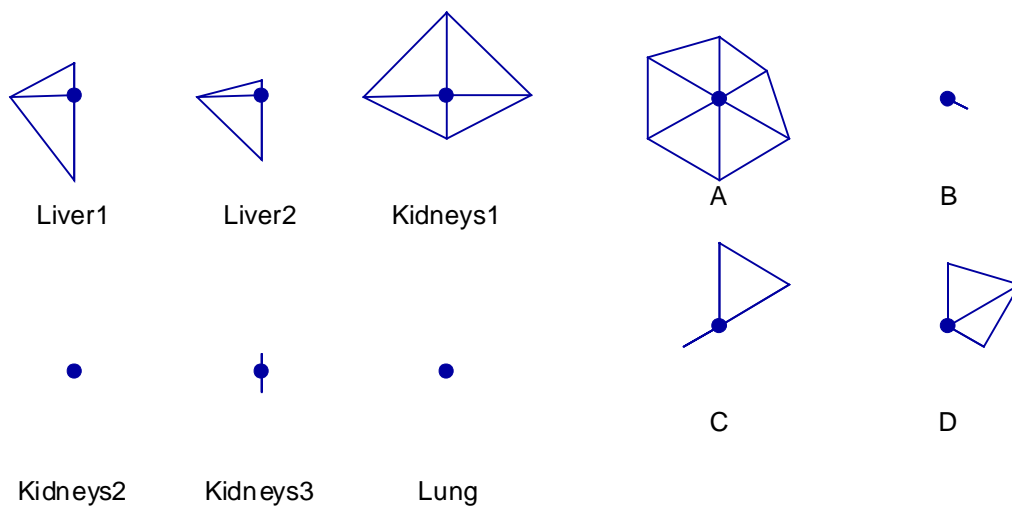
*Obr. 17: Symbolový graf – tváře*



Z grafu tváří je patrná podobnost vyšetření jater a podobnost mezi vyšetřeními ledvin 2, 3 a plic.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 18: Hvězdicový graf



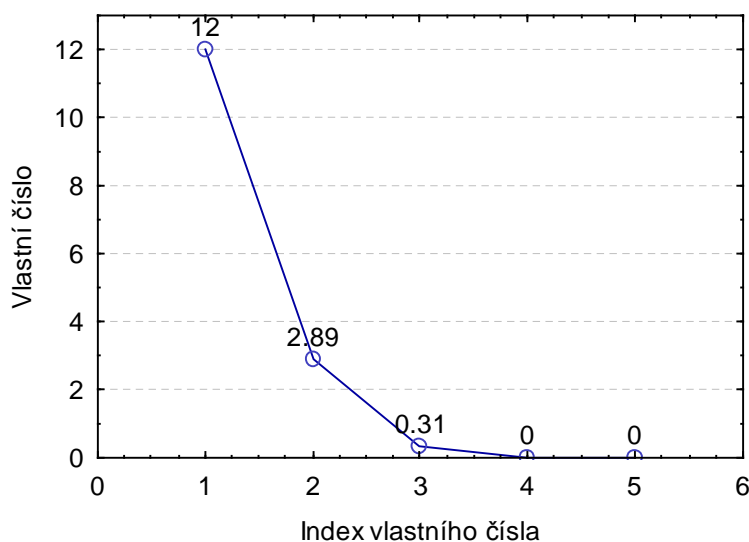
Z hvězdicového grafu lze určit podobnost vyšetření jater a podobnost mezi vyšetřeními ledvin 2, 3 a plic.

### 3.4. Mapování objektů vícerozměrným škálováním MDS

Tab. 35: tabulka vlastních čísel

Index	Vlastní číslo	Procento	Kumulativní vlastní číslo	Kumulativní procento
1	12,00	78,95	12,00	78,95
2	2,89	19,01	14,89	97,96
3	0,31	2,04	15,20	100,00
4	0,00	0,00	15,20	100,00
5	0,00	0,00	15,20	100,00

Obr. 19: Cattelův graf úpatí vlastních čísel



Z tabulky 12 a z Cattelova grafu lze stanovit první zlom u indexu 2 a pokles pod 1 u indexu 3. První dvě komponenty objasňují 97,96 % celkového rozptylu. Použijeme 2 faktory.



**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Tab. 36: Tabulka těsnosti proložení statistickou analýzou

Sloupec	Řádek	Skutečná vzdálenost.	Vypočtená vzdálenost (metrická m.)	Vypočtená vzdálenost (nemetrická m.)
3 – Ledviny 1	6 - Plíce	0.743	0.743	0.743
3 – Ledviny 1	4 - Ledviny 2	1.678	0.975	1.382
3 – Ledviny 1	5 - Ledviny 3	1.676	0.981	1.382
1 – Játra 1	4 - Ledviny 2	1.697	1.057	1.382
1 – Játra 1	6 - Plíce	1.520	1.076	1.382
2 – Játra 2	3 - Ledviny 1	0.975	1.081	1.382
1 – Játra 1	5 - Ledviny 3	1.076	1.520	1.382
2 – Játra 2	4 - Ledviny 2	1.081	1.529	1.382
2 – Játra 2	6 - Plíce	1.529	1.618	1.382
2 – Játra 2	5 - Ledviny 3	1.693	1.676	1.382
1 – Játra 1	3 - Ledviny 1	0.981	1.678	1.382
1 – Játra 1	2 – Játra 2	1.618	1.693	1.382
4 - Ledviny 2	5 - Ledviny 3	1.057	1.697	1.382
5 - Ledviny 3	6 - Plíce	2.389	2.389	2.389
4 - Ledviny 2	6 - Plíce	2.397	2.397	2.397
Stres – kritérium těsnosti		0.173		
Alienace		0.281		
Čistý stress (nemetrická m.)		1.082		
Čistý stress (metrická m.)		2.901		

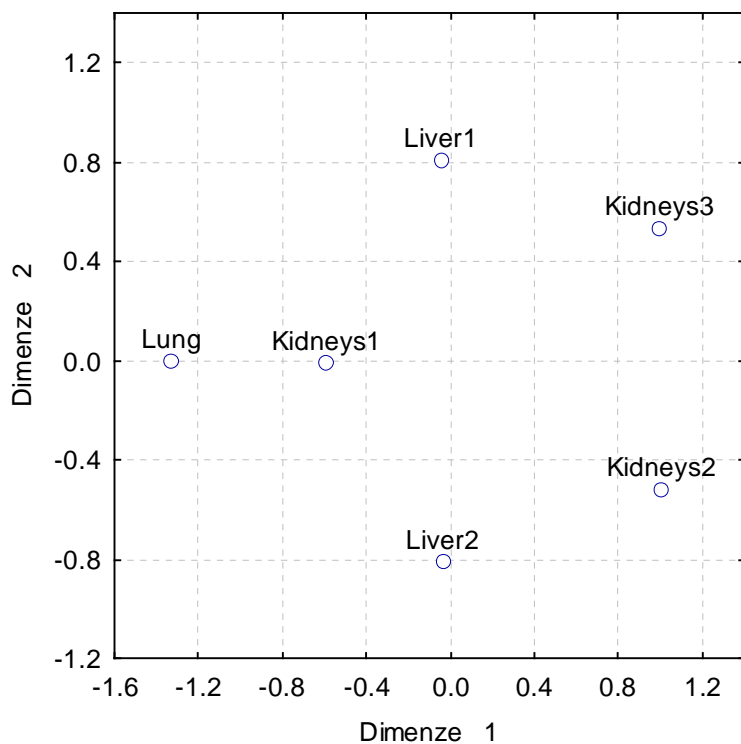
Hodnota stress vyjadřuje míru těsnosti proložení. Pro tuto úlohu není proložení dobré (> 0,05)

Tab. 37: Mapa objektů

	Souřadnice 1	Souřadnice 2
Liver1	-0,0443	0,8082
Liver2	-0,0360	-0,8096
Kidneys1	-0,5900	-0,0076
Kidneys2	1,0061	-0,5245
Kidneys3	0,9963	0,5328
Lung	-1,3333	0,0006

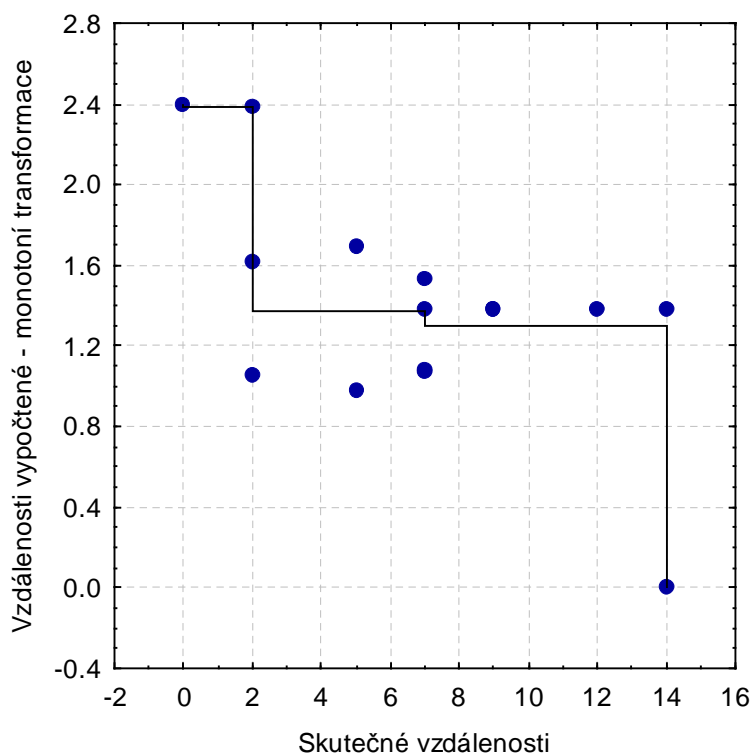
**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 20: Škálovací diagram



Škálovací diagram vyjadřuje graficky vzdálenosti mezi jednotlivými proměnnými. Z tohoto grafu je patrné, že podle první souřadnice, která vystihuje 80 % variability, si jsou nejvíce podobné vyšetření ledvin 2 a 3, vyšetření jater 1 a 2. Podle druhé souřadnice (19 % variability) podobnost vyšetření ledvin 1 a plic.

Obr. 21: Shepardův diagram

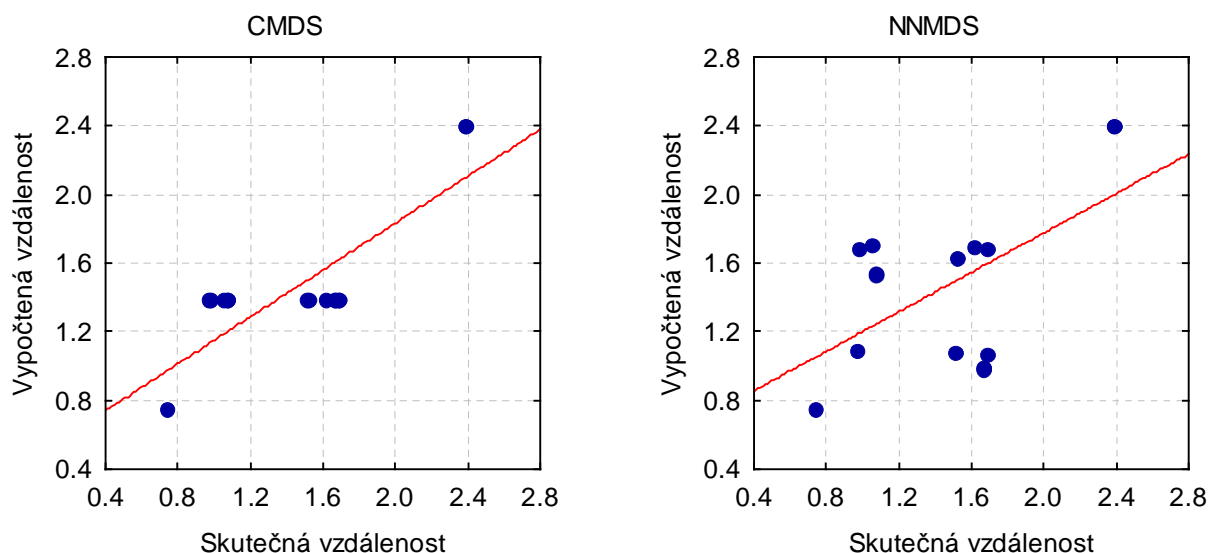


## SEMESTRÁLNÍ PRÁCE

### 3.5 Klasifikace analýzou vícerozměrných dat

Shepardův rozptylový diagram zobrazuje závislost mezi skutečnými a vypočtenými vzdálenostmi. Křivka představuje spojitou monotónní transformaci. Body, které jsou v blízkosti křivky představují dobrý model a body vzdálené představují špatné proložení.

Obr. 22: Těsnost proložení ve škálovacím disagramu CMDS a NNMDS



Těsnost proložení vyjadřuje vzdálenost od přímky. V tomto případě je vidět špatné proložení dat.

### 3.5. Korespondenční analýza CA

Tab. 38: Řádkové profily v procentech

	A	B	C	D	Celkem
Liver1	30.00	0.00	40.00	30.00	100.00
Liver2	25.00	0.00	37.50	37.50	100.00
Kidneys1	40.00	20.00	13.33	26.67	100.00
Kidneys2	100.00	0.00	0.00	0.00	100.00
Kidneys3	66.67	0.00	33.33	0.00	100.00
Lung	100.00	0.00	0.00	0.00	100.00
Průměr	60.28	3.33	20.69	15.69	100.00

Tab. 39: Sloupcové profily v procentech

	A	B	C	D	Průměr
Liver1	20.00	0.00	40.00	30.00	22.50
Liver2	13.33	0.00	30.00	30.00	18.33
Kidneys1	40.00	100.00	20.00	40.00	50.00
Kidneys2	6.67	0.00	0.00	0.00	1.67
Kidneys3	13.33	0.00	10.00	0.00	5.83
Lung	6.67	0.00	0.00	0.00	1.67
Celkem	100.00	100.00	100.00	100.00	100.00

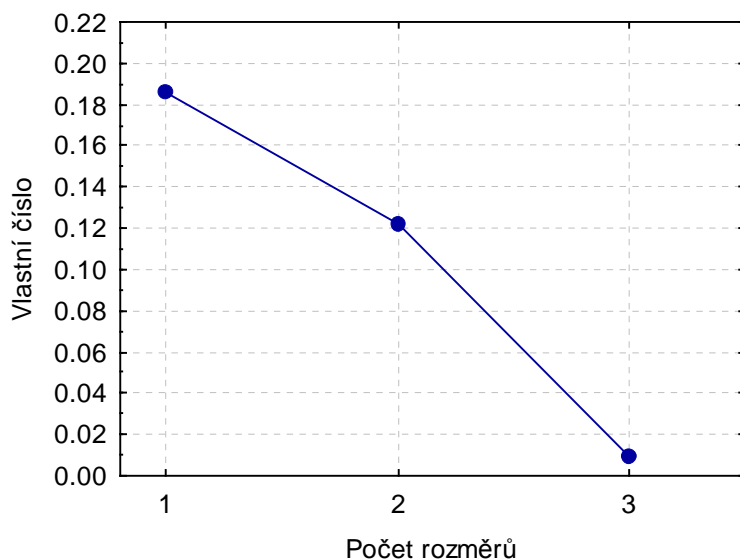
Sloupcové a řádkové profily slouží k vyšetření rozdílů mezi proměnnými a celkovým profilem.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Tab. 40: Tabulka vlastních čísel

m	Vlastní číslo	Individuální %	Kumulativní %	$\chi^2$
1	0.19	58.65	58.65	7.07
2	0.12	38.36	97.01	4.62
3	0.01	2.99	100.00	0.36
Součet	0.32			

Obr. 23: Cattelův indexový graf vlastních čísel



Z tabulky 41 a z Cattelova grafu není přesně vidět zlom ale lze usuzovat, že 97 % vysvětlené variability bude postačovat k vysvětlení podobnosti objektů.

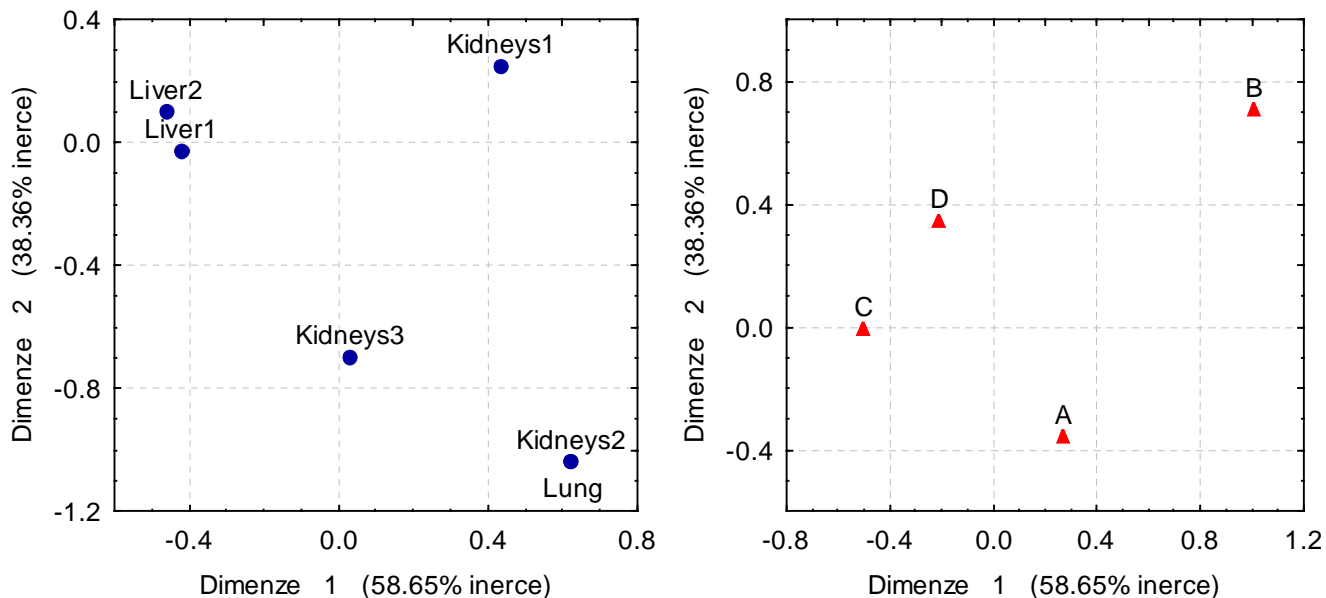
Tab. 41: Tabulka řádkových profilů a příspěvků do inercie

Kategorie	Souřadnice komp. 1	Souřadnice komp. 2	Mass	Kvalita	Relativní inercie	inercie komp. 1	$\text{Cos}^2$ komp. 1	inercie komp. 2	$\text{Cos}^2$ komp. 2
Liver1	-0.420	-0.028	0.263	0.996	0.148	0.250	0.991	0.002	0.005
Liver2	-0.458	0.098	0.211	0.968	0.151	0.238	0.926	0.016	0.042
Kidneys1	0.435	0.246	0.395	0.999	0.311	0.401	0.758	0.196	0.242
Kidneys2	0.625	-1.039	0.026	0.959	0.127	0.055	0.255	0.233	0.704
Kidneys3	0.031	-0.701	0.079	0.899	0.136	0.000	0.002	0.319	0.897
Lung	0.625	-1.039	0.026	0.959	0.127	0.055	0.255	0.233	0.704

Kvalita vyjadřuje míru zkreslení dat snížením počtu os. V tomto případě je pro všechny proměnné vyšší než 90%, data nejsou zkreslená. Mass představuje váhu, která připadá na danou proměnnou. Inercie představuje sumo všech četností v prvků tabulky pro danou proměnnou.  $\text{Cos}^2$  představuje příspěvek inercie do dané osy.

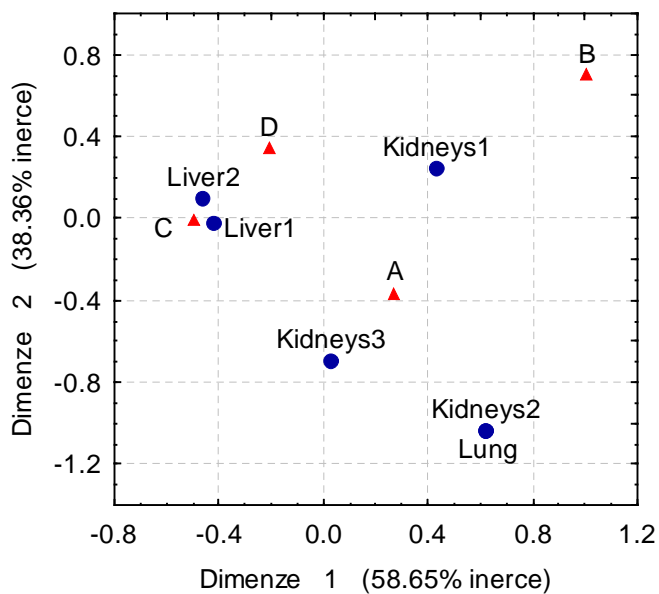
**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

Obr. 24: Grafy sloupcových a řádkových profilů



Podle obrázku 8, se vyšetření rozdělí na vyšetření jater a vyšetření ledvin 2 a plic. Podobné skupiny jsou C a D. Skupina B je velmi odlišná od ostatních skupin.

Obr. 25: Souhrnný graf sloupcových a řádkových profilů



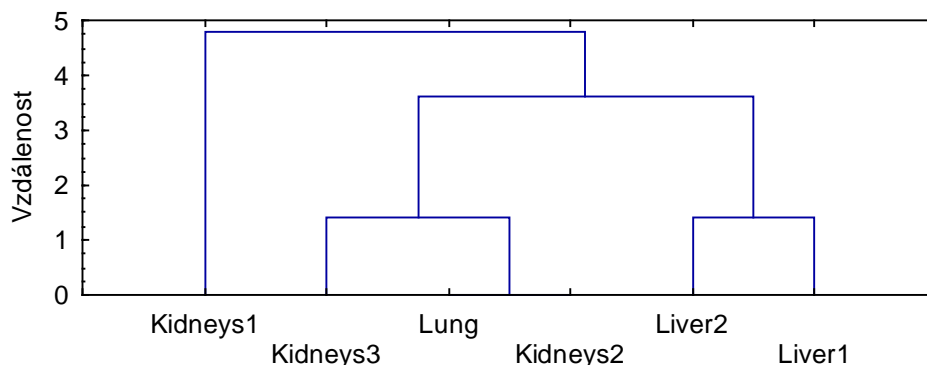
Ze souhrnného grafu lze určit, že nejvíce histopatologických nálezů jater bylo stanoveno pro skupinu C a také D. Průměrný počet všech druhů nálezů bylo určeno pro kontrolní skupinu A. U skupiny B nebyly hodnoceny žádné výrazné množství histopatologických nálezů.

**SEMESTRÁLNÍ PRÁCE**  
3.5 Klasifikace analýzou vícerozměrných dat

### 3.6. Analýza shluků

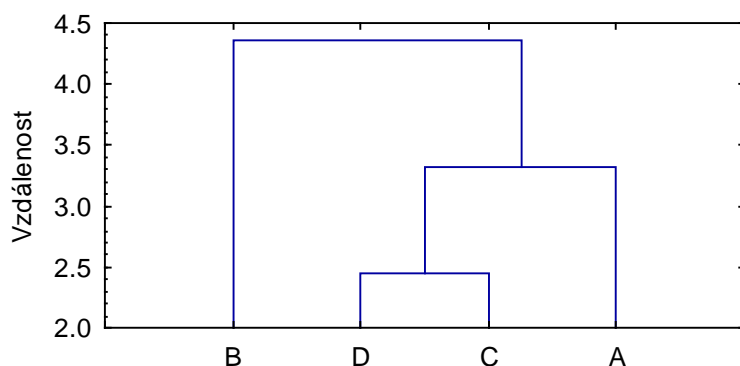
Analýza shluků se zabývá klasifikací dat pomocí shluků. Tato metoda používá pro rozdělení proměnných do shluků

Obr. 26: Dendrogram – metoda nejbližšího souseda – eukleidovská vzdálenost – nálezy



Z tohoto dendrogramu lze určit, že podobné množství histopatologických nálezů bylo stanoveno pro vyšetření ledvin 2 a plic a také ledvin 3, dále podobné množství bylo stanoveno pro obě vyšetření jater. Zvláštní skupinu tvoří vyšetření ledvin 1.

Obr. 27: Dendrogram – metoda nejbližšího souseda – eukleidovská vzdálenost – skupiny



Z tohoto dendrogramu lze určit, že podobné histopatologické nálezy byly stanoveny pro skupinu C a D – vyšší dávka a delší doba podání.

### 3.7. Závěr

Při hodnocení stanovení histopatologických nálezů pro 4 vyšetřované skupiny se ukázalo, že pro všechna zvířata jsou v závislosti oba typy nálezů jater a také je závislost u vyšetření ledvin 2 a plic. Při vyšetření ledvin 2 a plic bylo stanoveno nejméně nálezů. Nejvíce nálezů pro většinu skupin bylo stanoveno při vyšetření ledvin 1.

Nejčtenější nálezy v oblasti jater byly stanoveny pro skupinu C – podání trojnásobné dávky. Nejméně histopatologických nálezů bylo pozorováno u skupiny B s podáním standardního množství látky ve standardním čase. U kontrolní skupiny A byly nalezeny histopatologické změny u všech typů vyšetření.

Podaná látka je antibiotikum, které se podává hlavně při infekcích dýchacích cest nebo močového ústrojí.