

Semestrální práce

3.1 Matematické principy analýzy vícerozměrných dat

RNDr.Raimund HEDBÁVNÝ

Bioveta, a.s. Komenského 212, 683 23 Ivanovice na Hané

2007

3.1 Matematické principy analýzy vícerozměrných dat

Úloha 1

Najděte vlastní (charakteristická) čísla a vlastní vektory, determinant, stopu a odmocninu od této jednoduché matice

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 8 & 2 \\ 3 & 2 & 10 \end{pmatrix}$$

Determinant

- Determinant je definován pouze pro čtvercové matice.
- Pokud je determinant roven nule je matice singulární a neexistuje inverzní matice.
- Pokud je determinant nenulový je matice nesingulární a inverzní matice existuje.
- Pokud je determinant větší než nula má soustava n rovnic o n neznámých reprezentovaná maticí právě jedno řešení.

Determinant matice 2 x 2 se vypočítá jako rozdíl součinů prvků na hlavní a vedlejší diagonále

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$$

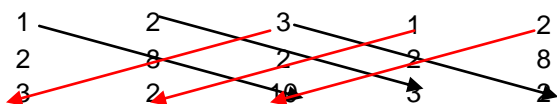
Pro výpočet determinantu matice 3 x 3 existuje několik způsobů (Gaussova eliminace, kofaktorová metoda, apod).

K výpočtu determinantu matice 3 x 3 lze použít Sarrusova pravidla, jak ukazuje následující postup

K matici

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 8 & 2 \\ 3 & 2 & 10 \end{vmatrix}$$

připíšeme první dva sloupce. Od součtu součinů trojic čísel ve směru černých šipek odečteme součet součinů ve směru šipek červených.



Výsledek:

$$\det(\mathbf{A}) = (80 + 12 + 12) - (72 + 4 + 40) = \underline{\underline{-12}}$$

Stopa matice

Stopa matice je součet diagonálních prvků matice

V případě zadané matice

$$\text{tr}(\mathbf{A}) = 1 + 8 + 10 = \underline{\underline{19}}$$

Vlastní čísla

Pro výpočet vlastních čísel matice je třeba řešit rovnici $\det(\mathbf{A} - \lambda\mathbf{E}) = 0$.

Pro zadanou matici

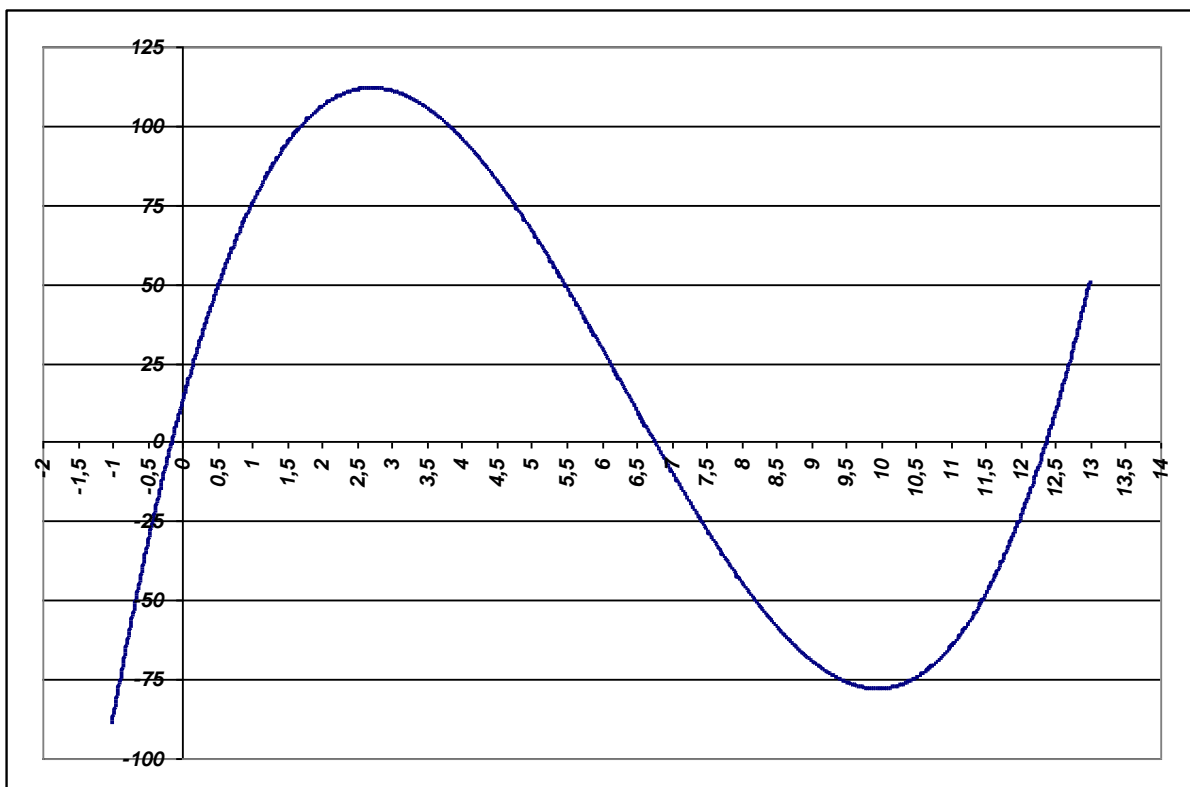
$$\det(\mathbf{A} - \lambda\mathbf{E}) = \det \begin{vmatrix} 1-\lambda & 2 & 3 \\ 2 & 8-\lambda & 2 \\ 3 & 2 & 10-\lambda \end{vmatrix} = 0$$

Výsledkem je následující kubická rovnice

$$x^3 - 19x^2 + 81x + 12 = 0$$

Kořeny této rovnice, které jsou vlastními čísly matice, byly určeny graficky. Průběh funkce

$y = x^3 - 19x^2 + 81x + 12$ ukazuje graf



Kořeny rovnice jsou

$$\lambda_1 = 12,3776, \quad \lambda_2 = 6,7657, \quad \lambda_3 = -0,1433$$

Pro kontrolu

$$\Sigma\lambda_i = \text{tr}(\mathbf{A}) = 19, \quad \Pi\lambda_i = \det(\mathbf{A}) = -12$$

Výpočet vlastních čísel a vlastních vektorů matice byl proveden i v programu OPgm. Byla nalezena stejná vlastní čísla a k nim odpovídající vlastní vektory

$$\mathbf{v}_1 \begin{vmatrix} 0,30189 \\ 0,50689 \\ 0,8074 \end{vmatrix} \quad \mathbf{v}_2 \begin{vmatrix} 0,014312 \\ 0,84444 \\ -0,53546 \end{vmatrix} \quad \mathbf{v}_3 \begin{vmatrix} 0,9532 \\ -0,17326 \\ -0,24776 \end{vmatrix}$$

Dosazením vypočítaných vlastních čísel λ_i a vlastních vektorů \mathbf{v}_i do vztahu

$$\mathbf{A} = \Sigma\lambda_i\mathbf{v}_i\mathbf{v}_i^T \quad \text{dostaneme}$$

$$\begin{vmatrix} 1,1281 & 1,8941 & 3,0170 \\ 1,8941 & 3,1803 & 5,0657 \\ 3,0170 & 5,0657 & 8,0689 \end{vmatrix} + \begin{vmatrix} 0,0014 & 0,0818 & -0,0518 \\ 0,0818 & 4,8245 & -3,0592 \\ -0,0518 & -3,0592 & 1,9398 \end{vmatrix} + \begin{vmatrix} -0,1302 & 0,0237 & 0,0338 \\ 0,0237 & -0,0043 & -0,0062 \\ 0,0338 & -0,0062 & -0,0088 \end{vmatrix}$$

což po sečtení a zaokrouhlení dává původní matici

$$\begin{vmatrix} 0,9992 & 1,9995 & 2,9990 \\ 1,9995 & 8,0004 & 2,0003 \\ 2,9990 & 2,0003 & 9,9999 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 8 & 2 \\ 3 & 2 & 10 \end{vmatrix}$$

Odmocnina z matice

Odmocnina z matice \mathbf{A} je definována vztahem

$$\mathbf{A}^{1/2} = \Sigma\lambda_i^{1/2}\mathbf{v}_i\mathbf{v}_i^T$$

Vzhledem ke skutečnosti, že $\lambda_3 < 0$ není odmocnina z matice v oboru reálných čísel definována.

Úloha 2

Pro typická data z vašeho pracoviště, minimálně 4 rozměrná, určete projekci do prvních dvou komponent, dvojný graf a diskutujte jeho význam.

V průběhu klinického ověřování polyvalentní vakcíny pro psy, která jako jednu ze složek obsahuje i antigeny navozující tvorbu protilátek proti třem nejčastěji se vyskytujícím sérovarům leptospir, bylo vakcinováno a následně revakcinováno 30 štěňat. Dalších 30 štěňat bez protilátek proti leptospirám sloužilo jako kontroly. Čtyři týdny po revakcinaci byla provedena čelenž virulentními sérovary *Leptospira grippotyphosa*, *Leptospira icterohaemorrhagiae* a *Leptospira canicola* (vždy 10 vakcinovaných a 10 kontrolních). Zvířata byla pozorována 4 týdny po infekci. V tomto období byla štěňatům odebírána krev a testována moč. Následně byla zvířata utracena a byla provedena pitva.

Testy

Biochemická a hematologická vyšetření krve: jaterní testy (ALT,ALP), bilirubin, podíl bílých krvinek (WBC count)

Mikrobiologické testy: přítomnost leptospir v krvi a moči (blood cult. a urine cult.)

Klinické příznaky: icterus (žloutenka), conjunktivitis (zánět spojivek)

Patologické vyšetření

Mezi sledované parametry byla zařazena i hmotnost v době čelenže. Znak vacc. je rozlišení mezi vakcinovanými a kontrolními skupinami.

Získaná data ukazuje následující tabulka 1

Tabulka 1 – Výsledky klinického testování leptospirové složky vakcíny

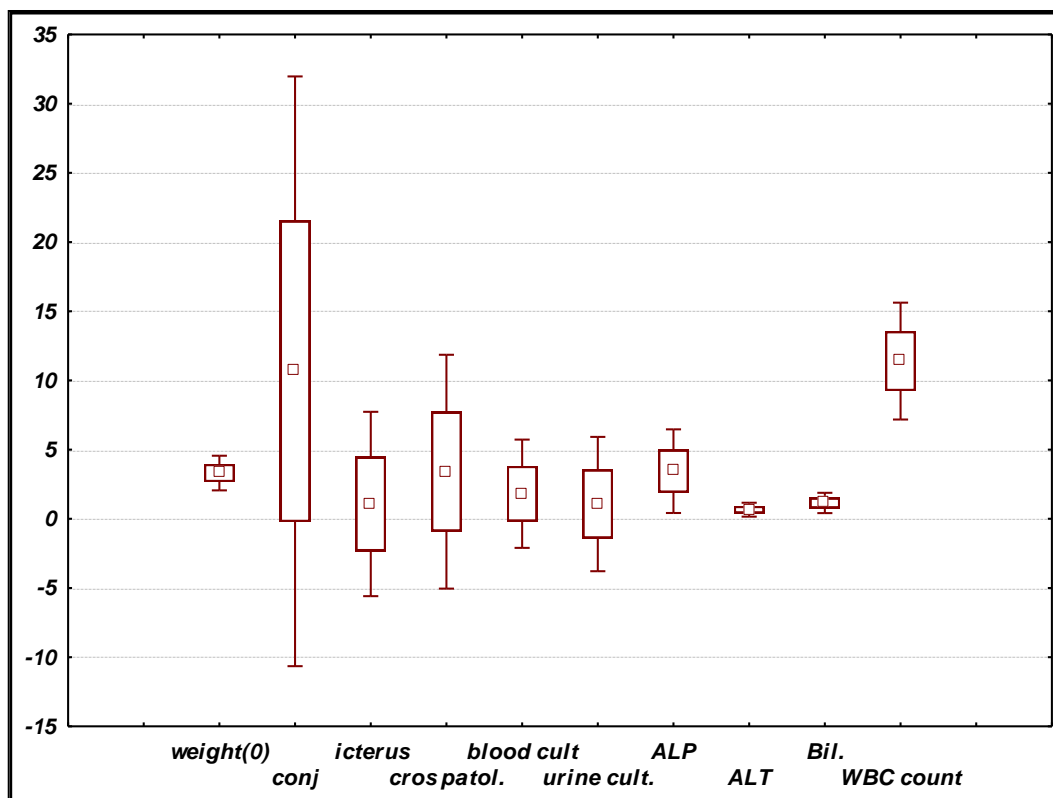
ID	vaccin.	weight (kg)	conjunctivitis	icterus	cross pathology	blood cult.	urine cult.	ALP (ukat/l)	ALT (ukat/l)	Bilirubin (umol/l)	WBC count (g/l)
grippe_v_253	1	3,7	0	0	0	0	0	3,655	0,6875	0,575	10,35
grippe_v_254	1	3,7	0	0	0	0	0	3,3425	0,73	0,75	10,1
grippe_v_255	1	3,8	0	0	0	0	0	2,6275	0,6425	0,6	11,525
grippe_v_256	1	3,8	11	0	0	1	0	3,24	0,7925	0,6	11,625
grippe_v_257	1	3,6	0	0	0	0	0	3,085	0,5625	0,425	9,975
grippe_v_258	1	3,1	0	0	0	1	0	2,81	0,68	0,825	11,65
grippe_v_259	1	3,6	0	0	0	0	0	3,455	0,525	1,025	11,5
grippe_v_260	1	4,1	6	0	0	0	0	2,4325	0,62	0,85	11,275
grippe_v_239	1	3	0	0	0	0	0	3,125	0,5475	0,825	12,025
grippe_v_240	1	3,1	0	0	0	0	0	2,63	0,5325	1,15	8,15
grippe_c_223	0	3,6	9	0	5	3	0	3,36	0,605	0,725	9,475
grippe_c_224	0	3,5	17	0	5	4	0	12,4525	2,0175	1,45	10,425
grippe_c_225	0	3,2	6	1	3	4	0	9,19	1,5875	1,225	12,275
grippe_c_226	0	3	17	4	5	4	6	4,3425	0,4225	1,25	11
grippe_c_227	0	4	7	0	0	3	3	3,1475	0,505	0,6	8,7
grippe_c_229	0	3,2	15	0	10	4	0	3,6775	0,4875	0,625	10,6
grippe_c_230	0	3,5	17	0	10	4	3	3,6675	0,5675	0,9	12,05
grippe_c_231	0	2,8	19	1	10	4	0	3,9125	0,5675	0,625	10,275
grippe_c_232	0	3,1	20	0	8	4	0	3,9075	0,5475	0,775	9,725
can_v_317	1	3,8	0	0	0	3	0	2,776	0,47	1,26	11,28
can_v_318	1	5	8	0	0	0	0	2,692	0,528	0,96	13,8
can_v_319	1	4,5	0	0	0	0	0	2,924	0,648	1,08	12,48
can_v_320	1	3,9	10	0	0	0	0	3,03	0,562	1,06	9,7
can_v_321	1	3,5	0	0	0	3	0	2,856	0,746	0,98	14,24
can_v_322	1	3,7	5	0	0	0	0	3,112	0,7	0,82	10,76
can_v_323	1	3,4	9	0	0	0	0	3,02	0,54	1,2	9,36
can_v_324	1	3,5	0	0	0	0	0	2,67	0,658	1,52	12,96
can_v_325	1	3,2	10	0	0	0	0	3,092	0,564	1,1	10,44
can_v_326	1	3,6	0	0	0	0	0	2,736	0,514	1,44	12
can_c_327	0	3,1	24	0	8	3	0	3,404	0,656	0,72	10,48
can_c_328	0	2,8	22	0	8	0	0	2,744	0,862	0,96	9,84
can_c_329	0	3,1	40	1	8	4	9	3,45	0,576	1,24	8,52
can_c_330	0	4	17	17	0	3	0	2,61	0,682	1,06	7,36
can_c_331	0	4,3	17	0	10	3	6	2,152	0,824	1,26	9,04
can_c_332	0	4,1	17	17	8	3	0	2,578	0,724	1,12	10,28
can_c_333	0	4,3	40	8	8	2	0	2,734	0,614	0,9	6,32
can_c_334	0	4,2	25	0	8	6	6	2,458	0,47	0,76	9,72
can_c_335	0	3,5	24	0	0	0	0	1,932	0,724	0,96	7,96
can_c_336	0	3	40	8	8	6	6	2,794	0,586	0,82	10,56
ictero_v_337	1	2,6	9	0	0	1	0	3,364	0,406	1	14,6
ictero_v_338	1	3,1	0	0	0	0	0	4,096	0,456	1,64	16,4
ictero_v_339	1	3,4	0	0	0	0	0	3,956	0,624	1,58	10,16
ictero_v_340	1	3,7	7	0	0	0	0	4,1825	0,4325	1,45	11,48
ictero_v_341	1	3,1	0	0	0	0	0	3,166	0,592	1,54	14,52
ictero_v_342	1	3	0	0	0	0	0	2,99	1,058	1,4	15,52
ictero_v_343	1	3,4	10	0	0	3	0	3,484	0,498	1,82	13,12
ictero_v_344	1	3	0	0	0	0	0	3,72	0,792	1,42	12,92
ictero_v_345	1	3,1	0	0	0	0	0	3,978	0,68	1,78	9,32

ictero_v_346	1	2,8	0	0	0	0	0	2,706	0,634	1,74	13,68
ictero_c_347	0	3	24	0	0	3	0	4,436	0,506	1,68	10,04
ictero_c_348	0	2,3	9	0	5	2	0	4,952	0,626	1,4	11,48
ictero_c_349	0	2,4	1	1	7	0	0	3,176	0,604	1,58	14,36
ictero_c_350	0	3,4	21	0	10	3	9	2,628	0,78	1,26	13,92
ictero_c_351	0	2	18	0	9	2	0	2,582	0,504	1,42	15,08
ictero_c_352	0	2,3	7	0	6	0	0	2,758	0,528	1,84	12,68
ictero_c_353	0	2,5	26	3	7	6	3	3,452	0,518	1,68	15,48
ictero_c_354	0	1,9	20	0	11	5	0	3,334	0,614	1,14	11,96
ictero_c_355	0	2,2	12	1	9	5	3	2,86	0,67	1,625	13,6
ictero_c_356	0	2,1	13	1	15	5	9	3,4	1,036	1,5	12,4

Ze skupiny grippo_c byl vyřazen pes č. 228, u kterého došlo pravděpodobně vlivem jiného onemocnění ke zkreslení výsledků.

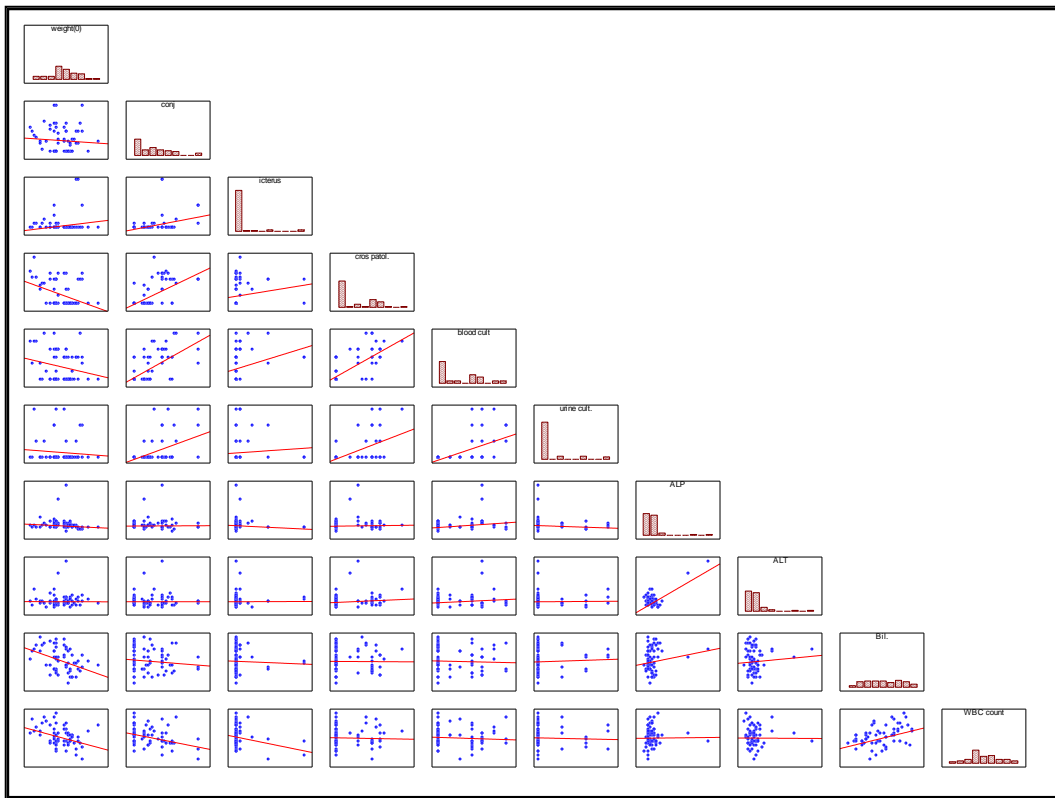
Variabilita znaků je vidět na krabicovém grafu na obr. 1

Obr.1 – Krabicový graf (STATISTICA)



U znaků weight, ALT a Bil. pozorujeme malou variabilitu. Data jsou v různých jednotkách, proto byla provedena standardizace. Na standardizované zdrojové matici byla prověřena přítomnost významných korelací mezi znaky. Výsledek ukazuje maticový graf na obr.2.

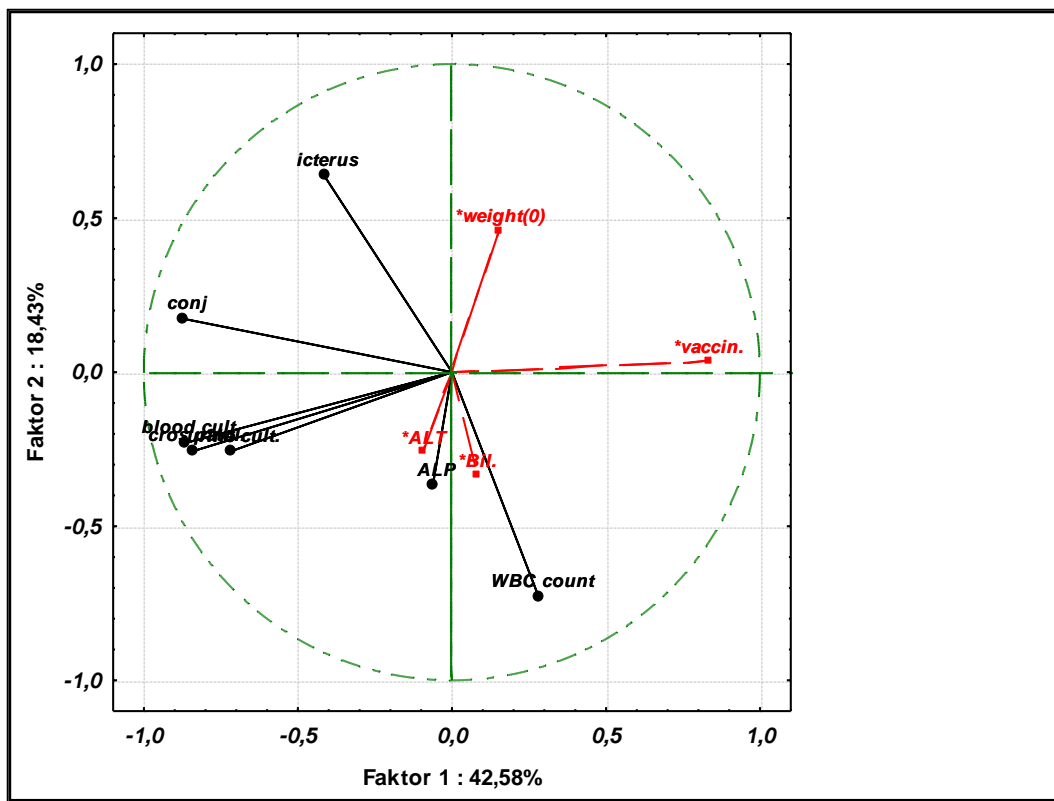
Obr.2 – Korelace znaků (STATISTICA)



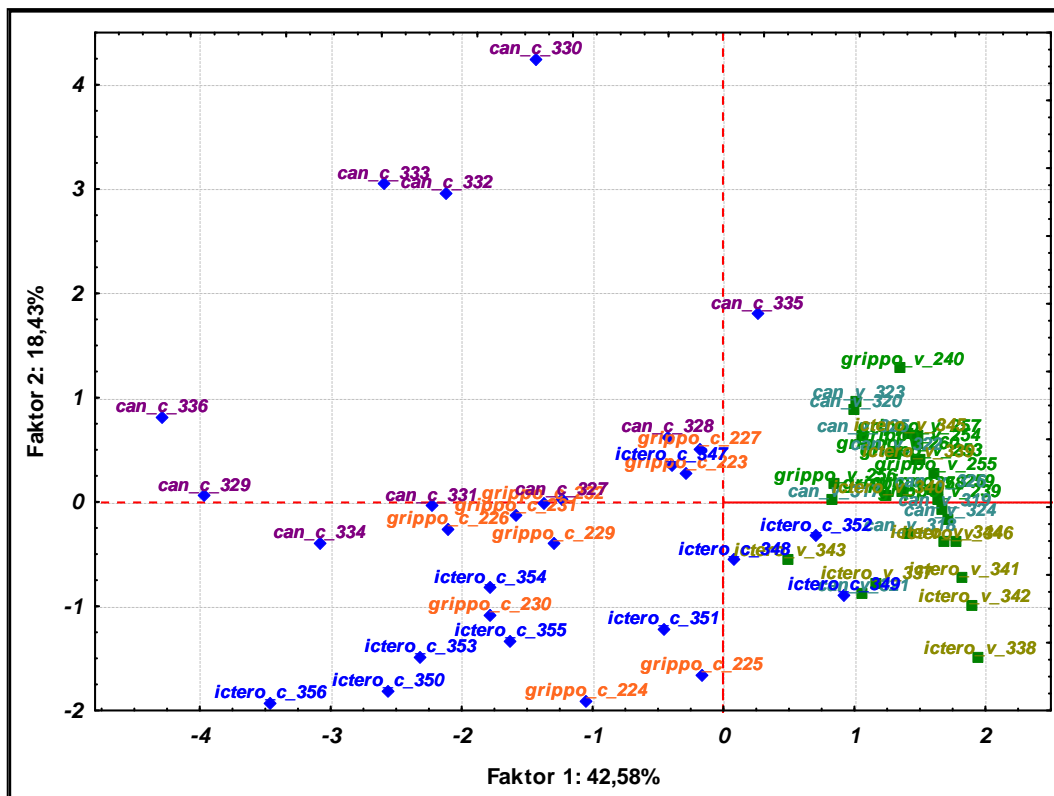
Mezi znaky se vyskytují významné korelace

Metodou PCA pro aktivní proměnné conj., icterus, cross patol., blood cult., urine cult., ALP a WBC count; pomocné proměnné vacc., weight, ALT a Bil. byla v programu Statistica získána následující projekce znaků (obr. 3a) a objektů (obr. 3b) do prvních dvou latentních proměnných.

Obr 3a – Graf komponentních vah (STATISTICA)

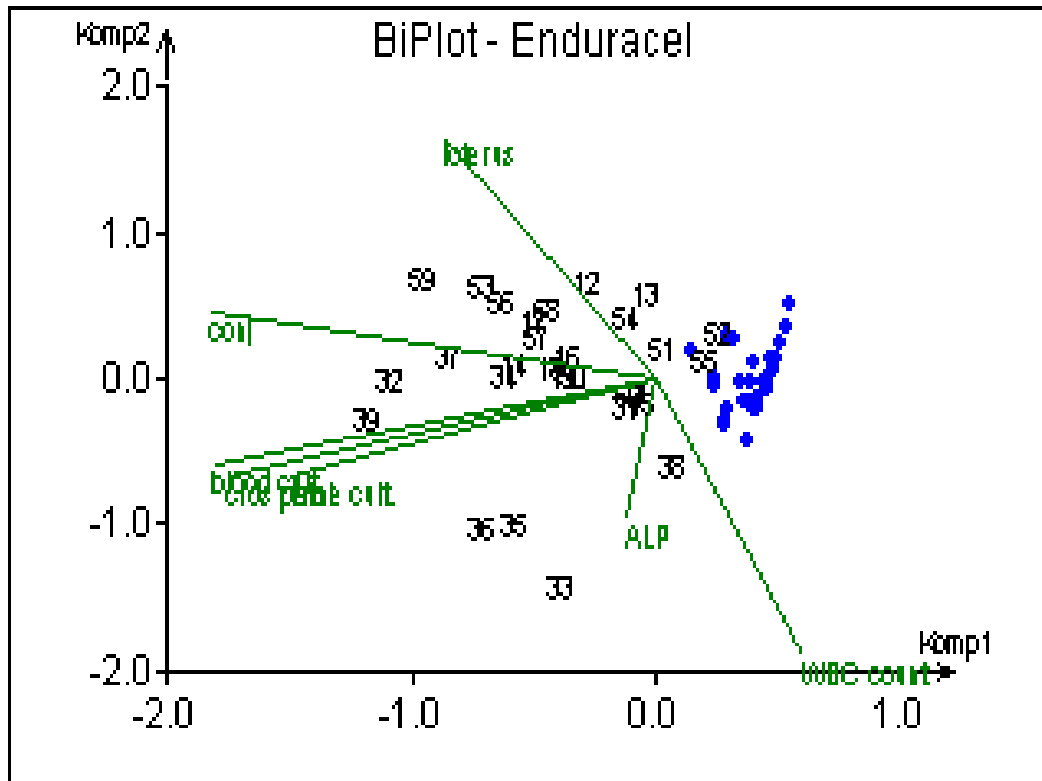


Obr 3b – Rozptylový diagram komponentního skóre (STATISTICA)



Sloučením výše uvedených grafů do jednoho dostáváme dvojný graf (Biplot). Na obr.4 je uveden dvojný graf vytvořený v QC Expertu 2.7. Ke konstrukci byly použity pouze proměnné uvedené výše jako aktivní.

Obr.4 – Biplot (QC Expert 2.7)



V grafu jsou čísla označena zvířata z kontrolních skupin. Modré body představují vakcinovaná štěňata. Poloha objektů koresponduje se směry znaků. Většina kontrol bez protilátek je rozptýlena kolem polopřímek reprezentujících příznaky infekce leptospirou, zatímco vakcinovaná zvířata tvoří kompaktní shluk mimo průvodiče znaků.

Závěr

Protilátková odezva vakcinovaných zvířat je dostatečná ke zvládnutí leptospirové infekce.