



SEMESTRÁLNÍ PRÁCE

Klasifikace analýzou vícerozměrných dat

2005/2006

Ing. Petr Eliáš



1. KLASIFIKACE DAT DISKRIMINAČNÍ ANALÝZOU

1.1 Zadání

V podniku Precheza a.s. se červený pigment TP 303 vyrábí rozkladem zelené skalice. Důležitým kvalitativním ukazatelem tohoto produktu je barevný odstín vyjádřený v barevném prostoru CIE L*a*b*. Šarže produktu, které splňují určité požadavky jsou zařazeny do tří skupin, jež pro jednoduchost označíme písmeny A – C. Na předložených datech ověřte s využitím DA výše uvedenou klasifikaci šarží.

1.2 Data

| Šarže | L* | a* | b* | DE _{CMC} | Třída |
|-------|-------|-------|------|-------------------|-------|
| 1 | 49,05 | 16,63 | 6,88 | 0,64 | C |
| 2 | 48,95 | 16,55 | 6,76 | 0,62 | B |
| 3 | 49,11 | 16,63 | 6,73 | 0,48 | B |
| 4 | 48,97 | 16,46 | 6,72 | 0,61 | B |
| 5 | 49,05 | 16,58 | 6,77 | 0,57 | B |
| 6 | 49,11 | 16,33 | 6,50 | 0,46 | B |
| 7 | 49,10 | 16,57 | 6,68 | 0,47 | B |
| 8 | 48,91 | 16,25 | 6,36 | 0,58 | B |
| 9 | 49,02 | 16,51 | 6,59 | 0,48 | B |
| 10 | 48,86 | 16,58 | 6,78 | 0,69 | B |
| 11 | 48,86 | 16,58 | 6,78 | 0,69 | B |
| 12 | 49,13 | 16,92 | 7,07 | 0,68 | C |
| 13 | 49,00 | 16,93 | 7,00 | 0,63 | C |
| 14 | 49,06 | 16,73 | 6,81 | 0,47 | B |
| 15 | 49,13 | 16,70 | 6,65 | 0,32 | B |
| 16 | 49,10 | 16,65 | 6,54 | 0,24 | B |
| 17 | 49,42 | 16,12 | 5,93 | 0,54 | A |
| 18 | 49,44 | 16,27 | 6,02 | 0,42 | B |
| 19 | 49,29 | 15,87 | 5,54 | 0,90 | A |
| 20 | 49,43 | 16,01 | 5,91 | 0,59 | A |
| 21 | 49,35 | 16,18 | 6,05 | 0,44 | B |
| 22 | 49,28 | 16,14 | 6,04 | 0,47 | B |
| 23 | 49,29 | 16,16 | 5,96 | 0,51 | B |
| 24 | 49,36 | 15,86 | 5,35 | 1,06 | A |
| 25 | 49,32 | 16,15 | 5,77 | 0,64 | A |
| 26 | 49,29 | 16,32 | 6,21 | 0,31 | B |
| 27 | 49,36 | 16,56 | 6,59 | 0,29 | B |
| 28 | 49,39 | 16,68 | 6,67 | 0,31 | B |
| 29 | 49,30 | 16,57 | 6,49 | 0,21 | B |
| 30 | 49,37 | 16,72 | 6,86 | 0,49 | B |
| 31 | 49,45 | 16,79 | 6,80 | 0,41 | B |
| 32 | 49,32 | 16,76 | 6,69 | 0,32 | B |
| 33 | 49,22 | 16,70 | 6,60 | 0,26 | B |
| 34 | 49,30 | 16,92 | 6,88 | 0,48 | B |
| 35 | 49,51 | 16,78 | 6,76 | 0,38 | B |
| 36 | 49,25 | 16,79 | 6,70 | 0,32 | B |
| 37 | 49,31 | 16,77 | 6,84 | 0,47 | B |
| 38 | 49,30 | 16,86 | 6,93 | 0,54 | C |
| 39 | 49,32 | 16,96 | 6,90 | 0,50 | B |
| 40 | 49,18 | 16,95 | 6,77 | 0,39 | B |
| 41 | 49,34 | 16,92 | 6,69 | 0,29 | B |
| 42 | 49,34 | 16,98 | 7,00 | 0,59 | C |
| 43 | 49,32 | 16,96 | 7,00 | 0,60 | C |
| 44 | 49,30 | 16,79 | 6,82 | 0,44 | B |
| 45 | 49,35 | 16,74 | 6,80 | 0,43 | B |
| 46 | 49,32 | 16,87 | 7,00 | 0,61 | C |
| 47 | 49,31 | 16,81 | 6,87 | 0,49 | B |
| 48 | 49,38 | 16,69 | 6,80 | 0,44 | B |
| 49 | 49,35 | 16,70 | 6,78 | 0,42 | B |
| 50 | 49,35 | 16,81 | 6,93 | 0,54 | B |
| 51 | 49,40 | 16,82 | 6,98 | 0,60 | C |
| 52 | 49,28 | 16,77 | 6,90 | 0,53 | B |
| 53 | 49,20 | 16,53 | 6,89 | 0,60 | C |
| 54 | 49,20 | 16,59 | 6,95 | 0,64 | C |

| Šarže | L* | a* | b* | DE _{CMC} | Třída |
|-------|-------|-------|------|-------------------|-------|
| 55 | 49,29 | 16,25 | 6,57 | 0,47 | B |
| 56 | 49,21 | 16,32 | 6,66 | 0,50 | B |
| 57 | 49,26 | 16,38 | 6,69 | 0,48 | B |
| 58 | 49,41 | 16,30 | 6,49 | 0,38 | B |
| 59 | 49,27 | 16,33 | 6,63 | 0,46 | B |
| 60 | 49,32 | 16,43 | 6,83 | 0,58 | C |
| 61 | 49,26 | 16,63 | 6,99 | 0,66 | C |
| 62 | 49,21 | 16,62 | 6,90 | 0,58 | C |
| 63 | 49,32 | 16,60 | 6,89 | 0,57 | C |
| 64 | 49,48 | 16,57 | 6,94 | 0,63 | C |
| 65 | 49,39 | 16,59 | 6,88 | 0,57 | C |
| 66 | 49,29 | 16,53 | 6,83 | 0,54 | B |
| 67 | 49,31 | 16,51 | 6,82 | 0,53 | B |
| 68 | 49,36 | 16,53 | 6,89 | 0,59 | C |
| 69 | 49,22 | 16,65 | 7,03 | 0,69 | C |
| 70 | 49,32 | 16,52 | 6,84 | 0,55 | B |
| 71 | 49,21 | 16,45 | 6,80 | 0,55 | B |

Soubor s daty: TP_303_sarze.xls
Použitý software: STATISTICA v7.1

1.3 Řešení

1.3.1 Vybrané úseky textového výstupu programu STATISTICA

Tab I Výsledky diskriminační funkční analýzy

| | Wilkovo λ | Parciální λ | F | Úroveň p | Tolerance | Tolerance 1-R ² |
|-------------------|-------------------|---------------------|--------|----------|-----------|----------------------------|
| b* | 0,279 | 0,517 | 30,397 | 0,000 | 0,431 | 0,569 |
| DE _{CMC} | 0,306 | 0,470 | 36,634 | 0,000 | 0,709 | 0,291 |
| L* | 0,176 | 0,820 | 7,151 | 0,002 | 0,802 | 0,198 |
| a* | 0,154 | 0,933 | 2,338 | 0,105 | 0,413 | 0,587 |

Tab II Test χ^2 významnosti kořenů

| Odstraněný kořen | Vlastní číslo | Kanonicke R | Wilkovo λ | χ^2 | Úroveň p |
|------------------|---------------|-------------|-------------------|----------|----------|
| 0 | 2,626 | 0,851 | 0,144 | 128,915 | 0,000 |
| 1 | 0,917 | 0,692 | 0,522 | 43,259 | 0,000 |

Tab. III Standardizované koeficienty pro kanonické proměnné

| Proměnná | Kořen 1 | Kořen 2 |
|-------------------|---------|---------|
| b* | -1,184 | -0,469 |
| DE _{CMC} | 0,663 | -0,947 |
| L* | 0,327 | -0,555 |
| a* | 0,473 | -0,022 |
| Vlastní | 2,626 | 0,917 |
| Kumul. podíl | 0,741 | 1,000 |

Tab IV Klasifikační funkce

| | C (p = 0,253) | B (p = 0,676) | A (p = 0,0704) |
|-------------------|---------------|---------------|----------------|
| b* | 2,573 | 0,219 | -11,368 |
| DE _{CMC} | 1,374 | -1,074 | 5,365 |
| L* | 0,659 | -0,462 | 2,060 |
| Konstanta | -0,352 | -0,230 | 3,473 |

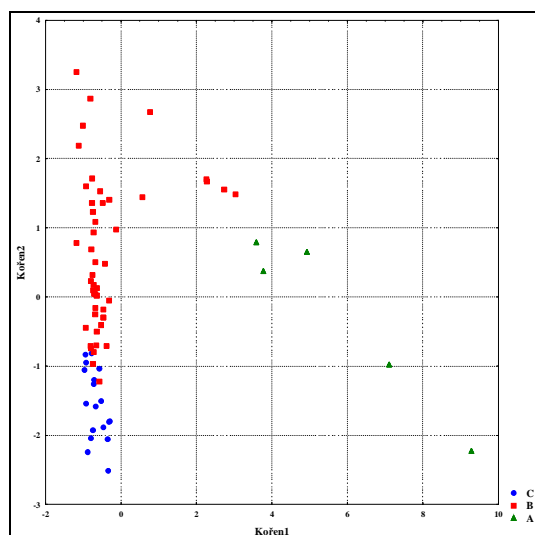
Tab. V Klasifikační matice

| | % správných | C (p = 0,253) | B (p = 0,676) | A (0,0704) |
|--------|-------------|---------------|---------------|------------|
| C | 88,889 | 16 | 2 | 0 |
| B | 95,833 | 2 | 46 | 0 |
| A | 100,000 | 0 | 0 | 5 |
| Celkem | 94,366 | 18 | 48 | 5 |

1.3.2 Komentář k vybraným úsekům textového výstupu programu STATISTICA

- Výsledky diskriminační funkční analýzy (tab. I)*
 Znaky, které nejlépe a nejvíce přispívají k rozdělení objektů do tříd jsou b^* a DE_{CMC} (mají nejvyšší hodnotu F a parciální Wilkovo λ nejbližší 0).
 Tolerance $1 - R^2$ je mírou nadbytečnosti znaku.
- Test χ^2 významnosti kořenů (tab II)*
 V prvním řádku se testuje stat. významnost všech (zde 2) kořenů. Druhý řádek se týká zbývajících kořenů bez prvního. V našem případě jsou obě diskriminační funkce (kanonické kořeny) významné.
- Standardizované koeficienty pro kanonické proměnné (tab. III)*
 Standardizované koeficienty jsou v normovaných vzájemně porovnatelných měřících. V první diskriminační funkci se nejsilněji projevují znaky b^* a DE_{CMC} , ve druhé potom znaky DE_{CMC} a srovnatelně L^* a b^* .
- Klasifikační funkce (tab. IV)*
 Šarže bude zařazena do takové třídy, ve které bude dosahovat nejvyššího klasifikačního skóre.
- Klasifikační matice (tab. V)*
 Z klasifikační matice jsou patrné počty správně a chybně zařazených šarží.

1.3.3 Vybrané grafy grafického výstupu programu STATISTICA



Obr. 1 Klasifikační graf kanonických kořenů

1.3.4 Komentář k vybraným grafům grafického výstupu programu STATISTICA

- *Klasifikační graf kanonických kořenů*
Kořen 1 (první diskriminační funkce) diferencuje šarže skupiny A od ostatních. Kořen 2 potom umožňuje od sebe separovat šarže skupin B a C, ve druhém případě není diference tak silná jako v případě prvním.

1.4 Závěr

Z předložených dat byla vypočtena klasifikační funkce umožňující zařadit „nové“ šarže výrobku TP 303 do definovaných tříd. 94,4 % šarží ve výše uvedených datech bylo zařazeno do správné kvalitativní skupiny. Drobný problém nastává při separaci materiálů skupin B a C, které nejsou ostře odděleny.

2. KLASIFIKACE DAT KANONICKOU KORELAČNÍ ANALÝZOU

2.1 Zadání

U vybraných šarží titanové běloby typu RGU se pokuste ověřit, zda existuje vztah mezi chemickým složením a dalšími sledovanými parametry.

2.2 Data

| Šarže | Spotřeba oleje | L* | b* | Al ₂ O ₃ | SiO ₂ | C organický |
|-------|----------------|-------|------|--------------------------------|------------------|-------------|
| 1 | 21,8 | 97,49 | 1,57 | 2,783 | 0,771 | 0,280 |
| 2 | 22,0 | 97,39 | 1,37 | 2,741 | 0,835 | 0,260 |
| 3 | 21,8 | 97,59 | 1,27 | 2,831 | 0,939 | 0,230 |
| 4 | 21,8 | 97,69 | 1,17 | 3,052 | 1,007 | 0,220 |
| 5 | 22,0 | 97,59 | 1,07 | 3,130 | 0,997 | 0,250 |
| 6 | 22,2 | 98,09 | 1,37 | 3,212 | 1,006 | 0,230 |
| 7 | 22,3 | 97,79 | 1,37 | 3,287 | 1,009 | 0,214 |
| 8 | 22,2 | 97,89 | 1,37 | 3,348 | 1,006 | 0,216 |
| 9 | 23,0 | 97,19 | 1,27 | 3,332 | 1,000 | 0,244 |
| 10 | 23,1 | 97,19 | 1,37 | 2,969 | 0,818 | 0,222 |
| 11 | 23,6 | 97,29 | 1,37 | 3,083 | 0,941 | 0,189 |
| 12 | 24,0 | 97,09 | 1,37 | 3,146 | 0,991 | 0,201 |
| 13 | 23,2 | 97,09 | 1,17 | 3,258 | 1,068 | 0,203 |
| 14 | 23,3 | 96,99 | 1,17 | 3,232 | 1,096 | 0,219 |
| 15 | 23,5 | 97,09 | 1,27 | 3,213 | 1,110 | 0,216 |
| 16 | 23,6 | 97,09 | 1,17 | 3,114 | 1,110 | 0,251 |
| 17 | 23,2 | 97,09 | 1,17 | 3,042 | 1,078 | 0,281 |
| 18 | 23,0 | 97,19 | 1,17 | 3,016 | 1,105 | 0,238 |
| 19 | 23,4 | 96,79 | 1,27 | 3,005 | 0,877 | 0,422 |
| 20 | 23,1 | 97,59 | 1,37 | 3,011 | 1,012 | 0,230 |
| 21 | 23,2 | 97,29 | 1,07 | 2,915 | 1,013 | 0,289 |
| 22 | 22,4 | 97,39 | 1,17 | 2,857 | 0,994 | 0,231 |
| 23 | 23,2 | 97,49 | 1,17 | 2,939 | 1,043 | 0,216 |
| 24 | 22,9 | 97,39 | 1,07 | 2,918 | 1,048 | 0,243 |
| 25 | 23,5 | 97,49 | 1,07 | 2,919 | 1,042 | 0,231 |
| 26 | 23,6 | 97,59 | 1,27 | 2,945 | 1,036 | 0,248 |
| 27 | 23,9 | 97,29 | 1,47 | 2,958 | 1,023 | 0,253 |
| 28 | 23,5 | 97,59 | 1,57 | 2,993 | 1,027 | 0,228 |
| 29 | 21,9 | 97,39 | 1,07 | 2,662 | 0,656 | 0,235 |
| 30 | 21,6 | 97,39 | 1,17 | 2,862 | 0,729 | 0,230 |
| 31 | 21,2 | 97,49 | 1,17 | 2,738 | 0,717 | 0,186 |
| 32 | 20,1 | 97,29 | 1,07 | 2,545 | 0,748 | 0,241 |
| 33 | 20,2 | 97,39 | 1,17 | 2,557 | 0,735 | 0,233 |
| 34 | 22,0 | 97,39 | 1,07 | 2,727 | 0,634 | 0,281 |
| 35 | 21,5 | 97,29 | 1,17 | 2,735 | 0,534 | 0,269 |
| 36 | 21,8 | 97,39 | 1,17 | 2,757 | 0,530 | 0,282 |
| 37 | 21,7 | 97,39 | 1,27 | 2,863 | 0,592 | 0,274 |
| 38 | 21,5 | 97,59 | 1,17 | 2,823 | 0,597 | 0,314 |
| 39 | 21,7 | 97,49 | 1,07 | 2,740 | 0,611 | 0,341 |
| 40 | 21,1 | 97,49 | 1,07 | 2,716 | 0,643 | 0,288 |
| 41 | 21,0 | 97,39 | 0,97 | 2,726 | 0,675 | 0,258 |
| 42 | 19,7 | 97,39 | 1,07 | 2,698 | 0,697 | 0,265 |
| 43 | 21,4 | 97,39 | 1,17 | 2,762 | 0,726 | 0,247 |
| 44 | 21,4 | 97,39 | 1,27 | 2,744 | 0,725 | 0,279 |
| 45 | 20,7 | 97,09 | 0,97 | 2,980 | 0,534 | 0,306 |
| 46 | 21,1 | 97,39 | 1,17 | 2,885 | 0,620 | 0,295 |
| 47 | 21,4 | 96,99 | 1,07 | 2,822 | 0,658 | 0,231 |
| 48 | 20,9 | 97,39 | 1,07 | 2,918 | 0,731 | 0,352 |
| 49 | 21,0 | 97,39 | 1,07 | 2,868 | 0,755 | 0,315 |
| 50 | 21,0 | 97,19 | 1,17 | 2,806 | 0,750 | 0,304 |
| 51 | 20,6 | 97,39 | 1,17 | 2,770 | 0,764 | 0,300 |
| 52 | 21,3 | 97,59 | 1,17 | 2,853 | 0,726 | 0,204 |
| 53 | 20,8 | 97,79 | 1,17 | 2,701 | 0,752 | 0,281 |
| 54 | 21,1 | 96,99 | 0,87 | 2,845 | 0,750 | 0,420 |
| 55 | 21,0 | 97,39 | 1,07 | 2,732 | 0,763 | 0,261 |
| 56 | 21,0 | 97,19 | 1,07 | 2,781 | 0,773 | 0,282 |
| 57 | 21,4 | 96,89 | 0,87 | 2,909 | 0,758 | 0,344 |
| 58 | 20,6 | 97,59 | 1,17 | 2,694 | 0,777 | 0,309 |

| Sarže | Spotřeba oleje | L* | b* | Al ₂ O ₃ | SiO ₂ | C organický |
|-------|----------------|-------|------|--------------------------------|------------------|-------------|
| 59 | 21,2 | 96,99 | 0,97 | 2,840 | 0,767 | 0,310 |
| 60 | 21,2 | 97,39 | 1,17 | 2,849 | 0,776 | 0,292 |
| 61 | 21,5 | 97,09 | 0,87 | 2,913 | 0,777 | 0,331 |
| 62 | 21,3 | 97,29 | 1,07 | 3,087 | 0,769 | 0,310 |
| 63 | 21,2 | 96,79 | 0,87 | 2,900 | 0,783 | 0,323 |
| 64 | 21,1 | 97,09 | 1,07 | 2,857 | 0,771 | 0,303 |
| 65 | 22,0 | 97,19 | 1,27 | 3,103 | 1,128 | 0,224 |
| 66 | 22,0 | 97,09 | 1,17 | 2,901 | 1,047 | 0,227 |
| 67 | 22,0 | 97,09 | 1,17 | 2,953 | 0,777 | 0,255 |
| 68 | 22,0 | 97,19 | 1,17 | 2,874 | 0,922 | 0,342 |
| 69 | 21,7 | 97,59 | 1,27 | 2,932 | 0,766 | 0,311 |
| 70 | 21,7 | 97,29 | 1,27 | 2,882 | 0,775 | 0,315 |
| 71 | 21,7 | 97,19 | 1,17 | 2,962 | 0,885 | 0,321 |
| 72 | 21,6 | 97,09 | 1,07 | 2,816 | 0,764 | 0,306 |
| 73 | 21,8 | 97,49 | 1,17 | 2,901 | 0,837 | 0,274 |
| 74 | 20,4 | 97,19 | 1,27 | 2,729 | 0,738 | 0,307 |
| 75 | 20,8 | 96,99 | 1,17 | 2,804 | 0,850 | 0,275 |
| 76 | 20,7 | 97,19 | 1,37 | 2,698 | 0,798 | 0,257 |
| 77 | 20,7 | 96,69 | 1,37 | 2,673 | 0,743 | 0,287 |
| 78 | 20,7 | 96,69 | 1,37 | 2,561 | 0,757 | 0,303 |
| 79 | 20,5 | 97,19 | 1,37 | 2,513 | 0,689 | 0,257 |
| 80 | 20,8 | 96,49 | 1,37 | 2,507 | 0,703 | 0,330 |
| 81 | 20,5 | 96,89 | 1,37 | 2,463 | 0,679 | 0,266 |
| 82 | 20,4 | 96,89 | 1,37 | 2,400 | 0,698 | 0,297 |
| 83 | 20,6 | 97,09 | 1,37 | 2,489 | 0,739 | 0,319 |
| 84 | 20,8 | 96,99 | 1,27 | 2,452 | 0,778 | 0,330 |
| 85 | 21,0 | 97,09 | 1,17 | 2,518 | 0,557 | 0,269 |
| 86 | 21,5 | 97,29 | 1,17 | 2,724 | 0,664 | 0,333 |
| 87 | 21,2 | 97,09 | 1,07 | 2,739 | 0,671 | 0,345 |
| 88 | 21,7 | 96,99 | 1,17 | 2,734 | 0,705 | 0,344 |
| 89 | 21,5 | 96,99 | 1,07 | 2,797 | 0,695 | 0,345 |
| 90 | 21,5 | 96,99 | 1,07 | 2,797 | 0,695 | 0,345 |
| 91 | 21,3 | 97,29 | 1,17 | 2,788 | 0,667 | 0,302 |
| 92 | 21,5 | 97,19 | 1,07 | 2,848 | 0,682 | 0,386 |
| 93 | 22,0 | 97,39 | 1,27 | 2,851 | 0,696 | 0,252 |
| 94 | 21,4 | 97,49 | 1,27 | 2,761 | 0,686 | 0,253 |
| 95 | 21,3 | 97,49 | 1,17 | 2,804 | 0,752 | 0,299 |
| 96 | 21,0 | 97,49 | 1,17 | 2,796 | 0,788 | 0,298 |
| 97 | 21,1 | 97,39 | 1,17 | 2,849 | 0,825 | 0,332 |
| 98 | 21,8 | 97,09 | 1,07 | 2,892 | 0,999 | 0,298 |
| 99 | 21,3 | 97,29 | 1,07 | 2,878 | 0,829 | 0,340 |
| 100 | 21,3 | 97,59 | 1,07 | 2,838 | 0,856 | 0,313 |

Soubor s daty: RGU_sarze.xls
Použitý software: STATISTICA v7.1

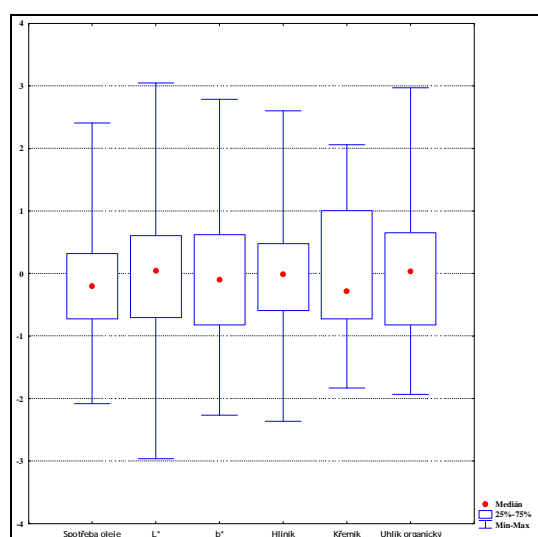
2.3 EDA

2.3.1 Vybrané úseky textového výstupu programu STATISTICA

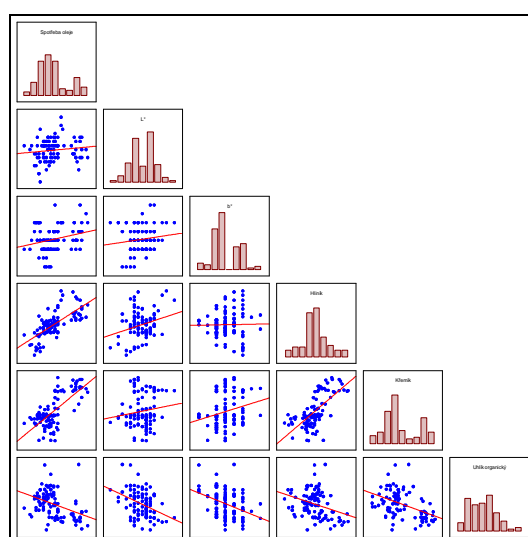
Tab. VI Korelace

| | Spotřeba oleje | L* | b* | Hliník | Křemík | Uhlík org. |
|----------------|----------------|--------|--------|--------|--------|------------|
| Spotřeba oleje | 1,000 | 0,123 | 0,255 | 0,703 | 0,713 | -0,389 |
| L* | 0,123 | 1,000 | 0,133 | 0,273 | 0,134 | -0,378 |
| b* | 0,255 | 0,133 | 1,000 | 0,018 | 0,234 | -0,372 |
| Hliník | 0,703 | 0,273 | 0,018 | 1,000 | 0,675 | -0,320 |
| Křemík | 0,713 | 0,134 | 0,234 | 0,675 | 1,000 | -0,438 |
| Uhlík org. | -0,389 | -0,378 | -0,372 | -0,320 | -0,438 | 1,000 |

2.3.2 Vybrané grafy grafického výstupu programu STATISTICA



Obr. 2 Krabic. graf pro stand. vstupní data



Obr. 3 Graf korelace

2.3.3 Komentář k vybraným grafům grafického výstupu programu STATISTICA

- *Krabicový graf (obr. 2)*
Znázorňuje proměnlivost jednotlivých znaků. Aby bylo možno takto jednotlivé znaky porovnat, musela být vzhledem k odlišnému měřítku data standardizována. Odlehle hodnoty nebyly indikovány.
- *Graf korelace (obr. 3)*
Poskytuje stejné informace, jako tabulka VI. Nejvýraznější jsou lineární závislosti mezi parametry Spotřeba oleje a Hliník, Spotřeba oleje a Křemík, Hliník a Křemík.

2.4 CCA

2.4.1 Vybrané úseky textového výstupu programu STATISTICA

Tab. VII Souhrn kanonické analýzy

| | | Levá sada kanonických prom. U | Pravá sada kanonických prom. V |
|----------------------------|-----------|-------------------------------|--------------------------------|
| Počet proměnných | | 3 | 3 |
| Získaný rozptyl | | 100,000% | 100,000% |
| Celková redundance | | 32,973% | 46,751% |
| Proměnné: | 1. | Spotřeba oleje | Hliník |
| | 2. | L* | Křemík |
| | 3. | B* | Uhlík organický |
| Kanonické R | | 0,798 | |
| p | | 0,000 | |
| χ^2 | | 127,300 | |

Tab. VIII Testy významnosti kanonických kořenů

| Odstraněný kořen | Kanonické R | Kanonické R ² | χ^2 | p | První λ |
|------------------|-------------|--------------------------|----------|-------|-----------------|
| 0 | 0,799 | 0,638 | 127,304 | 0,000 | 0,264 |
| 1 | 0,429 | 0,184 | 30,233 | 0,000 | 0,729 |
| 2 | 0,327 | 0,107 | 10,823 | 0,001 | 0,893 |

Tab. IX Zátěže (korelace) s kanonickou proměnnou U (struktura faktorů – levý soubor)

| | Kořen 1 (U ₁) | Kořen 2 (U ₂) | Kořen 3 (U ₃) |
|-----------------------|---------------------------|---------------------------|---------------------------|
| Spotřeba oleje | 0,968 | 0,092 | -0,235 |
| L* | 0,363 | -0,540 | 0,759 |
| b* | 0,217 | -0,828 | -0,517 |

Tab. X Extrahovaný (získaný) rozptyl a redundance (levá strana)

| | Získaný rozptyl | Redundance |
|--------------------------------|-----------------|------------|
| Kořen 1 (U₁) | 0,372 | 0,237 |
| Kořen 2 (U₂) | 0,329 | 0,060 |
| Kořen 3 (U₃) | 0,300 | 0,032 |

Tab. XI Zátěže (korelace) s kanonickou proměnnou V (struktura faktorů – pravý soubor)

| | Kořen 1 (V ₁) | Kořen 2 (V ₂) | Kořen 3 (V ₃) |
|------------------------|---------------------------|---------------------------|---------------------------|
| Hliník | 0,923 | 0,269 | 0,274 |
| Křemík | 0,875 | -0,002 | -0,484 |
| Uhlík organický | -0,556 | 0,825 | -0,103 |

Tab. XII Extrahovaný (získaný) rozptyl a redundance (pravá strana)

| | Získaný rozptyl | Redundance |
|--------------------------------|-----------------|------------|
| Kořen 1 (V₁) | 0,642 | 0,410 |
| Kořen 2 (V₂) | 0,251 | 0,046 |
| Kořen 3 (V₃) | 0,107 | 0,011 |

Tab. XIII Kanonické váhy (levá strana)

| | Kořen 1 | Kořen 2 | Kořen 3 |
|-----------------------|---------|---------|---------|
| Spotřeba oleje | 0,952 | 0,369 | -0,192 |
| L* | 0,253 | -0,471 | 0,861 |
| b* | -0,060 | -0,859 | -0,583 |

Tab. XIV Kanonické váhy (pravá strana)

| | Kořen 1 | Kořen 2 | Kořen 3 |
|-----------------|---------|---------|---------|
| Hliník | 0,603 | 0,542 | 1,087 |
| Křemík | 0,383 | 0,086 | -1,374 |
| Uhlík organický | -0,196 | 1,036 | -0,358 |

2.4.2 Komentář k vybraným úsekům textového výstupu programu STATISTICA

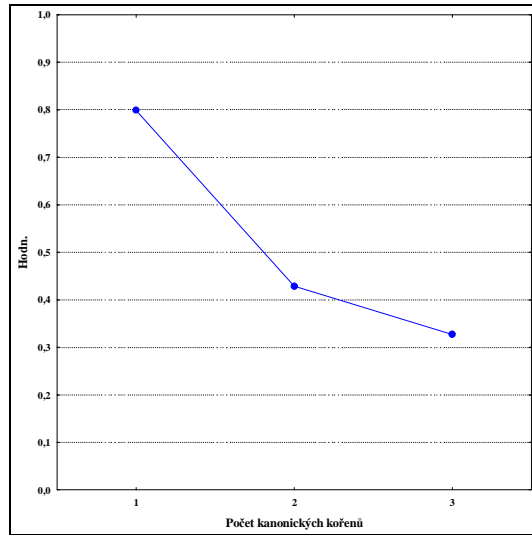
- Souhrn kanonické analýzy (tab. VII)*
 Kanonický korelační koeficient $R = 0,798$ je zde významný (vypočtená hladina významnosti $p < \alpha = 0,05$). R odpovídá nejvýznamnější kanonické proměnné U_1 a V_1 . Získaný (extrahovaný) rozptyl označuje průměrné množství spočteného rozptylu ze znaků v obou souborech pomocí kanonických proměnných. 3 kanonické proměnné dávají na levé i pravé straně rovnice 100 % rozptylu. Celková redundance ukazuje na velikost celkové korelace mezi znaky na levé straně rovnice 32,97 % a na pravé straně 46,75 %
- Testy významnosti kanonických kořenů (tab. VIII)*
 V prvním řádku nebyl odstraněn žádný kořen, testy jsou statisticky významné (hodnota $p < \alpha = 0,05$). Ve druhém řádku byl odstraněn nejvýznamnější kořen U_1 a V_1 , testy jsou nadále statisticky významné. V posledním řádku jsme odstranili navíc i kořen U_2 a V_2 , hodnota p napovídá, že testy jsou i v tomto případě statisticky významné. Všechny 3 kanonické kořeny jsou tedy statisticky významné.
- Struktura faktorů – levá strana (tab. IX)*
 Znak Spotřeba oleje vykazuje nejvyšší korelace u prvního kanonického faktoru. Znak L^* nejvíce koreluje s kanonickými faktory 2 a 3 a znak b^* s kanonickou proměnnou U_2 .
- Extrahovaný rozptyl – levá strana (tab. X)*
 První kanonický kořen vyčísluje 37,2 % rozptylu znaků na levé straně a 23,7 % rozptylu ostatních znaků založených na 1. kanonickém kořenu. V případě druhého kanonického kořene je to 32,9 a 6 % a v případě třetího kanonického kořene 30 a 3,2 %.
- Struktura faktorů – pravá strana (tab. XI)*
 Znak Hliník vykazuje nejvyšší korelace u prvního kanonického faktoru. Znak Křemík nejvíce koreluje s kanonickými faktory 1 a 3 a znak Uhlík organický s kanonickou proměnnou V_2 a V_1 .
- Extrahovaný rozptyl – pravá strana (tab. XII)*
 První kanonický kořen vyčísluje 64,2 % rozptylu znaků na pravé straně a 41,0 % rozptylu ostatních znaků založených na 1. kanonickém kořenu. V případě druhého kanonického kořene je to 25,1 a 4,6 % a v případě třetího kanonického kořene 10,7 a 1,1 %.
- Kanonické váhy – levá a pravá strana (tab. XIII, XIV)*
 Kanonické proměnné U a V jsou váženými součty znaků v každém souboru. V našem případě

$$U_1 = a_1 y_1 + a_2 y_2 + a_3 y_3$$

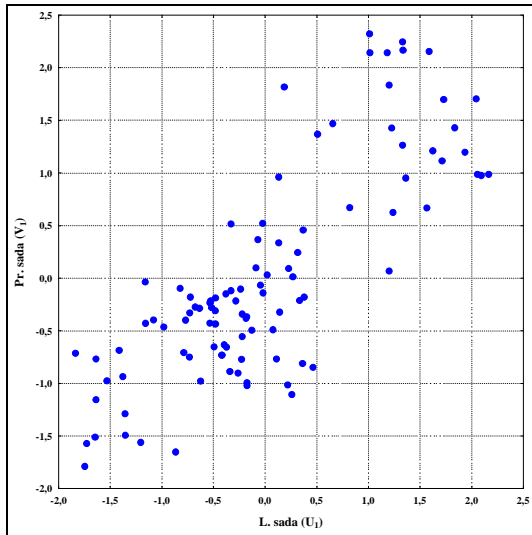
$$V_1 = b_1 y_1 + b_2 y_2 + b_3 y_3$$

Analogické jsou vztahy i pro proměnné U_2 a V_2 , resp. U_3 a V_3 .

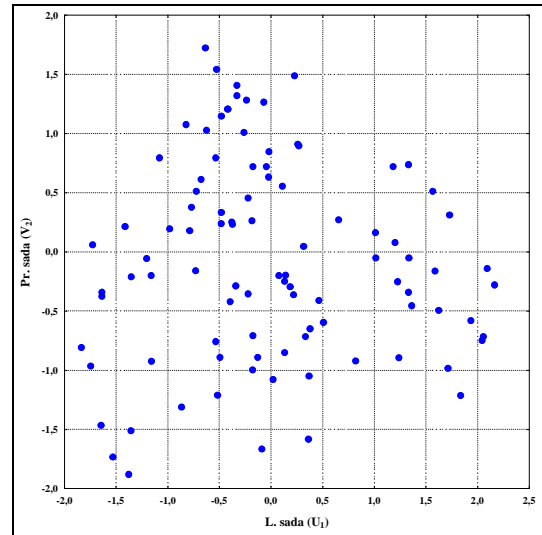
2.4.3 Vybrané grafy grafického výstupu programu STATISTICA



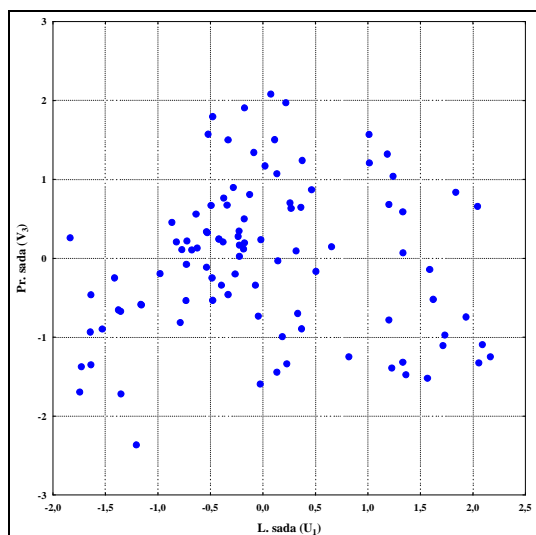
Obr. 4 Graf kanonických korelací



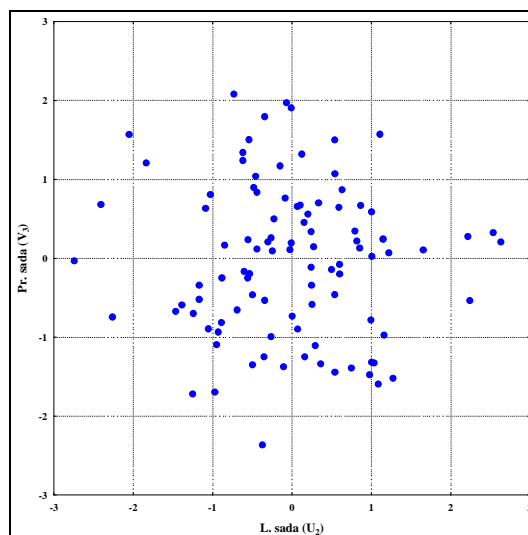
Obr. 5 Rozptyl. diagram kanonických korelací
 $V_1 = f(U_1)$



Obr. 6 Rozptyl. diagram kanon. korelací
 $V_2 = f(U_1)$



Obr. 7 Rozptyl. diagram kanonických korelací
 $V_3 = f(U_1)$



Obr. 8 Rozptyl. diagram kanon. korelací
 $V_3 = f(U_2)$

2.4.4 Komentář k vybraným grafům grafického výstupu programu STATISTICA

- *Grafy kanonických skóre (obr. 5, 6, 7, 8)*
 Objekty leží na přímce, lze indikovat shluky objektů, heterogenita je způsobena tím, že vybrané šarže nepochází z jedné výrobní kampaně.

2.5 Závěr

Provedením kanonické korelační analýzy bylo potvrzeno vzájemné ovlivnění chemického složení a dalších hodnocených parametrů u výrobku RGU.

3. KLASIFIKACE DAT KORESPONDENČNÍ ANALÝZOU A MDS

3.1 Zadání

Cílem ankety, které se zúčastnilo celkem 6371 respondentů, bylo objasnit vztah mezi jejich věkem a zdravotním stavem. Zdravotní stav byl hodnocen stupnicí velmi dobrý – dobrý – průměrný – špatný – velmi špatný. Každý z účastníků ankety zvolil kategorii, která nejlépe vystihovala jeho stav a tu označil jedničkou, ostatním kategoriím příslušely nuly. Získaná data jsou uvedena níže.

Pokuste se danou problematiku vyhodnotit pomocí metody korespondenční analýzy a vícenásobného škálování.

3.2 Data

| | Velmi dobrý | Dobrý | Průměrný | Špatný | Velmi špatný | Suma |
|-------------|-------------|-------|----------|--------|--------------|------|
| 16 – 24 | 243 | 789 | 167 | 18 | 6 | 1223 |
| 25 – 34 | 220 | 809 | 164 | 35 | 6 | 1234 |
| 35 – 44 | 147 | 658 | 181 | 41 | 8 | 1035 |
| 45 – 54 | 90 | 469 | 236 | 50 | 16 | 861 |
| 55 – 64 | 53 | 414 | 306 | 106 | 30 | 909 |
| 65 – 74 | 44 | 267 | 284 | 98 | 20 | 713 |
| Více než 75 | 20 | 136 | 157 | 66 | 17 | 396 |
| Suma | 817 | 3542 | 1495 | 414 | 103 | 6371 |

Soubor s daty: Zdravotni_stav.xls

Použitý software: STATISTICA v7.1

3.3 Řešení

3.3.1 Vybrané tabulky textového výstupu programu STATISTICA

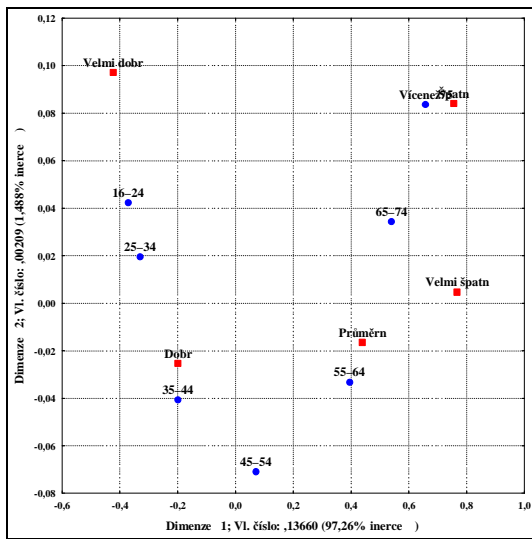
Tab. I Korelační matice

| | 16 – 24 | 25 – 34 | 35 – 44 | 45 – 54 | 55 – 64 | 65 – 74 |
|-----------|---------|---------|---------|---------|---------|---------|
| 16 – 24 | 1,000 | 0,999 | 0,990 | 0,931 | 0,785 | 0,602 |
| 25 – 34 | 0,999 | 1,000 | 0,994 | 0,936 | 0,797 | 0,613 |
| 35 – 44 | 0,990 | 0,994 | 1,000 | 0,967 | 0,856 | 0,689 |
| 45 – 54 | 0,931 | 0,936 | 0,967 | 1,000 | 0,956 | 0,848 |
| 55 – 64 | 0,785 | 0,797 | 0,856 | 0,956 | 1,000 | 0,960 |
| 65 – 74 | 0,602 | 0,613 | 0,689 | 0,848 | 0,960 | 1,000 |
| 75 a více | 0,515 | 0,529 | 0,613 | 0,788 | 0,931 | 0,993 |

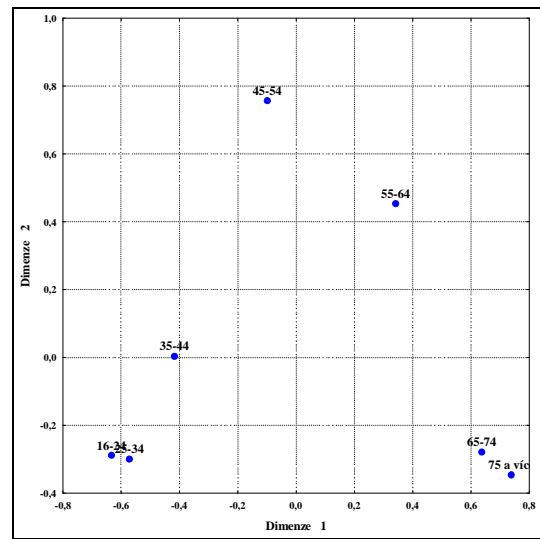
Tab. II Matice vzdáleností

| | 16 – 24 | 25 – 34 | 35 – 44 | 45 – 54 | 55 – 64 | 65 – 74 |
|-----------|---------|---------|---------|---------|---------|---------|
| 16 – 24 | 0,000 | 0,261 | 0,519 | 1,135 | 1,860 | 2,339 |
| 25 – 34 | 0,261 | 0,000 | 0,308 | 0,918 | 1,708 | 2,228 |
| 35 – 44 | 0,519 | 0,308 | 0,000 | 0,618 | 1,405 | 1,940 |
| 45 – 54 | 1,135 | 0,918 | 0,618 | 0,000 | 0,903 | 1,504 |
| 55 – 64 | 1,860 | 1,708 | 1,405 | 0,903 | 0,000 | 0,618 |
| 65 – 74 | 2,339 | 2,228 | 1,940 | 1,504 | 0,618 | 0,000 |
| 75 a více | 2,402 | 2,344 | 2,088 | 1,778 | 1,002 | 0,511 |

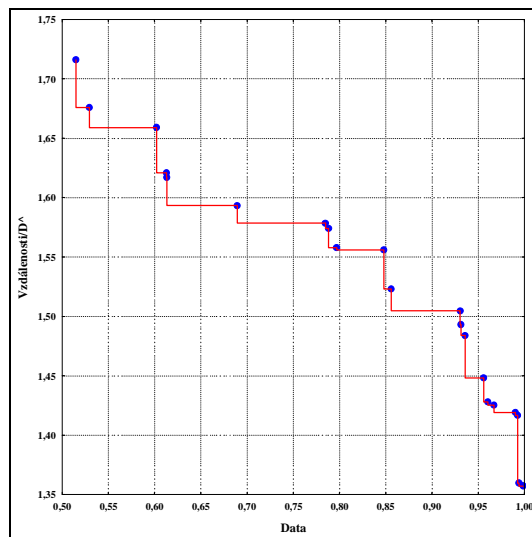
3.3.2 Vybrané grafy grafického výstupu programu STATISTICA



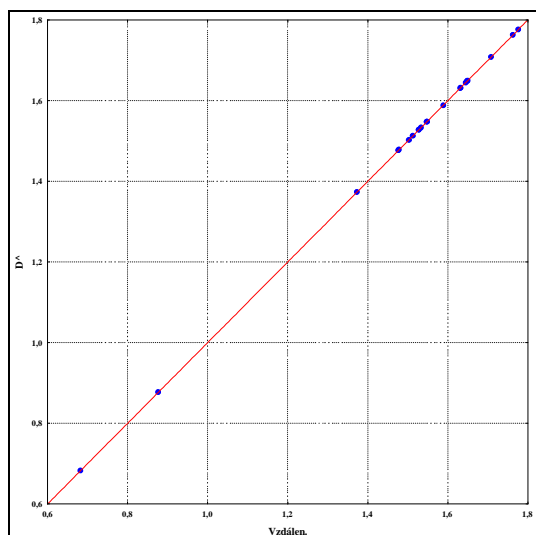
Obr. 9 Sloupcový a řádkový graf



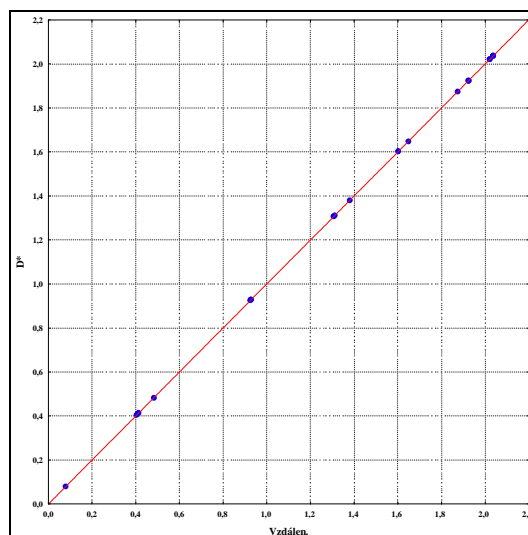
Obr. 10 Dvourozm. škál. diagram CMDS



Obr. 11 Shepardův diagram MDS



Obr. 12 Těsnost proložení (metricky)



Obr. 13 Těsnost proložení (nemetricky)

3.3.3 Komentář k vybraným grafům grafického výstupu programu STATISTICA

- *Sloupcový a řádkový graf (obr.9)*
Z grafu je patrné, že skupiny respondentů ve věku do 45 let vnímají svůj zdravotní stav jako velmi dobrý, případně dobrý, respondenti nad 64 let potom jako špatný, případně velmi špatný.
- *Dvourozměrný škálovací diagram (obr. 10)*
Z grafu je patrná podobnost skupin 16 - 24 a 25 - 34 let a rovněž 65 - 74 a 74 a více let.
- *Shepardův diagram (obr. 11)*
Všechny body leží těsně u křivky, což značí dobrý model.
- *Těsnost proložení (obr. 12 a 13)*
Těsnost je velmi dobrá, hodnoty vzdáleností d_{ij} vypočtené metricky a nemetricky jsou v dobré shodě.

3.4 Závěr

Pomocí korespondenční analýzy a MDS se podařilo osvětlit názorovou podobnost mezi respondenty, v podstatě byly indikovány 3 hlavní skupiny, a to respondentů ve věku 16 - 44 let, 45 - 64 let a více než 64 let.