AnalizaWW PD1

Maciej Nasinski November 16, 2015

FOTOGRAFIA SUKCESU - ANALIZA KANONICZNA

Za pomoca klasycznych współczynników korelacji oraz analizy kanonicznej zabadono zaleznosc pomiedzy zmiennymi opisujacymi ocene oraz cechy zdjec zamieszczonych na serwisie internetowym. Wnioskowanie pozwoliło na odnalezienie kilku interesujacych zaleznosci.

```
library(moments)
library(CCA)
library(ggplot2)
```

Badanie zaczynamy od wgrania oraz obróbki danych. Poczatkowo zakladamy iz wszystkie zmienne sa typu "numeric" co pozwoli na strorzenie macierzy. Nastepnie tworzymy nowy data frame w którym konkretne zmienne formatujemy na typ "factor" z odpowiednio przypisanymi "levels".

W programie R CRAN nie istnieje funkcja rozbijajacy zmienna typu factor na osobne kolumny o kodowaniu 0-1. W przypadku zmiennej 2 poziomowej wystarczy odjac od wszystkich wartosci w kolumnie wartosc 1. Dla zmiennych wielopoziomowych procedura nie jest tak prosta. Autor stworzyl nowa procedure pozwalajaca na rozbicie zmiennej factor na konkretne levele w osobnych kolumnach. Tworzymy trzeci data frame w którym rozbijamy zmienne typu factor.

```
zdjecia02<-zdjecia0
sap1<-sapply(c("miejsce","odbicia","kolor"),function(x) zdjecia02[,x]<<-as.factor(zdjecia02[,x]))
for(i in c("miejsce","odbicia","kolor")){
sap2<-sapply(2:length(levels(zdjecia02[,i])),function(x) zdjecia02[paste0(i,x)]<<-as.numeric(zdjecia02[,
}
zdjecia02<-zdjecia02[,!names(zdjecia02) %in% c("miejsce","odbicia","kolor")]</pre>
```

Warto zaprezentowac wyglad analizowanych data frame-ow.

```
head(zdjecia0,3)
```

```
koty dzien osoby komentarze artyzm miejsce odbicia ostrosc ocena kolor
##
## 1
               8
                      7
                                 42
                                          6
                                                   1
                                                            0
                                                                      3
                                                                                   1
                                                                      2
## 2
         4
                                 26
                                                                            3
              20
                      5
                                           4
                                                   0
                                                            1
                                                                                   4
## 3
         4
              18
                      2
                                 30
                                                   0
                                                                            5
                                                                                   3
```

```
head(zdjecia,3)
```

```
##
     koty dzien osoby komentarze
                                         artyzm
                                                   miejsce
                                                                 odbicia
## 1
        2
              8
                     7
                                42 rewelacyjnie inny kraj
                                                                    brak
        4
                                26
## 2
             20
                     5
                                         dobrze
                                                    Polska wys. odbicia
## 3
        4
             18
                                30
                                         dobrze
                     2
                                                    Polska wys. odbicia
##
         ostrosc
                        ocena
                                   kolor
                     b.dobrze fioletowy
## 1 przecietnie
## 2
           slabo przecietnie
                                    inny
## 3
          dobrze
                     b.dobrze niebieski
```

head(zdjecia02,3)

```
##
     koty dzien osoby komentarze artyzm ostrosc ocena miejsce2 odbicia2
## 1
         2
               8
                      7
                                  42
                                           6
                                                    3
                                                          5
                                                                     1
                                                                               0
## 2
         4
              20
                      5
                                  26
                                                    2
                                                          3
                                                                     0
                                                                               1
                                           4
## 3
         4
              18
                      2
                                  30
                                           4
                                                    4
                                                          5
                                                                     0
                                                                               1
##
     kolor2 kolor3 kolor4
## 1
           0
                   0
## 2
           0
                   0
                           1
## 3
           0
                   1
                           0
```

Czy istnieja istotne róznice w liczbie komentarzy zdjec zrobionych zagranica i w Polsce?

```
komPolska<-zdjecia[which(zdjecia$miejsce=="Polska"),4]
kominne<-zdjecia[which(zdjecia$miejsce=="inny kraj"),4]
t.test(komPolska,kominne)</pre>
```

```
##
## Welch Two Sample t-test
##
## data: komPolska and kominne
## t = -14.624, df = 2782.4, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -4.137852 -3.159418
## sample estimates:
## mean of x mean of y
## 22.88379 26.53243</pre>
```

Odrzucamy hipoteze zerowa - "true difference in means is equal to 0". T-test wskazuje na istotna roznice wzgledem sredniej.

Czy wsród zdjec o dominujacym kolorze niebieskim srednia liczba osób jest równa sredniej liczbie kotów na zdjeciu ?

Przystepujemy do wydzielenia z data frame wierszy dla których zdjecie posiada kolor niebieski. Nastepnie w latwy sposób obliczamy statystyke t dla róznicy liczby osob oraz kotów w nowym data frame.

```
zdjecianie<-zdjecia[which(zdjecia[,"kolor"]=="niebieski"),]
zdjecianieosoby<-zdjecianie$osoby
zdjecianiekoty<-zdjecianie$koty
t.test(zdjecianieosoby,zdjecianiekoty)</pre>
```

```
##
## Welch Two Sample t-test
##
## data: zdjecianieosoby and zdjecianiekoty
## t = -0.81289, df = 1457.5, p-value = 0.4164
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.19544816    0.08092006
## sample estimates:
## mean of x mean of y
## 2.959703    3.016967
```

W tym przypadku nie mozemy odrzucie hipotezy zerowej-"true difference in means is equal to 0".

Czy wyzsze oceny za artyzm sa skorelowane z nizszymi ocenami za ostrosc zdjecia?

```
tableao <- as.matrix(table((zdjecia$artyzm), (zdjecia$ostrosc), deparse.level = 2))</pre>
tableao
                   (zdjecia$ostrosc)
##
   (zdjecia$artyzm) b.slabo slabo przecietnie dobrze b.dobrze rewelacyjnie
##
       b.slabo
                          0
                                2
                                            4
                                                   8
                                                            9
                                           29
                                                           70
##
       slabo
                          5
                                9
                                                  73
                                                                        41
##
      przecietnie
                          9
                               21
                                           68
                                                 124
                                                          125
                                                                        62
##
       dobrze
                         89
                              167
                                          378
                                                 645
                                                          433
                                                                       140
##
       b.dobrze
                         40
                               66
                                           90
                                                 124
                                                           67
                                                                        24
##
      rewelacyjnie
                         10
                                6
                                            8
                                                   6
                                                            3
                                                                         0
sumr <- apply(tableao, 1, sum)</pre>
sumc <- apply(tableao, 2, sum)</pre>
tableao <- rbind(tableao, sumc)</pre>
tableao <- cbind(tableao, sumr = c(sumr, sum(sumr)))</pre>
sapply(1:6, function(x) tableao[, x]/tableao[7, x])
##
                      [,1]
                                  [,2]
                                              [,3]
                                                          [,4]
                                                                       [,5]
                0.00000000 0.007380074 0.006932409 0.008163265 0.012729844
## b.slabo
## slabo
                0.03267974 0.033210332 0.050259965 0.074489796 0.099009901
## przecietnie 0.05882353 0.077490775 0.117850953 0.126530612 0.176803395
## dobrze
               0.58169935 0.616236162 0.655112652 0.658163265 0.612446959
                0.26143791\ 0.243542435\ 0.155979203\ 0.126530612\ 0.094766620
## b.dobrze
## rewelacyjnie 0.06535948 0.022140221 0.013864818 0.006122449 0.004243281
                ## sumc
##
                      [,6]
## b.slabo
               0.02197802
## slabo
               0.15018315
## przecietnie 0.22710623
## dobrze
               0.51282051
## b.dobrze
                0.08791209
## rewelacyjnie 0.00000000
               1.00000000
## sumc
# correalation and significance of it
cor.test(as.numeric(zdjecia0$artyzm), as.numeric(zdjecia0$ostrosc), method = c("spearman"))
## Warning in cor.test.default(as.numeric(zdjecia0$artyzm),
## as.numeric(zdjeciaO$ostrosc), : Cannot compute exact p-value with ties
##
##
   Spearman's rank correlation rho
##
## data: as.numeric(zdjecia0$artyzm) and as.numeric(zdjecia0$ostrosc)
## S = 5284400000, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##
          rho
## -0.2213333
```

Canonical Correlation

Przeprowadzono losowanie 781 liczb ze zbioru liczb naturalnych od 1 do 2961 (liczba wierszy bazy danych). W celu umozliwienia latwego powtórzenia badania z identycznymi wynikami, ustawiono generator liczb losowych na konkretnym poziomie.

```
set.seed(308914)
randomrows<-sort(sample(1:nrow(zdjecia0), 781, replace = FALSE, prob = NULL))
zdjecia.random<-zdjecia0[randomrows,]
summary(zdjecia.random)</pre>
```

```
##
                         dzien
                                          osoby
                                                         komentarze
         koty
##
    Min.
           :0.000
                     Min.
                            : 3.00
                                      Min.
                                             :0.000
                                                       Min.
                                                              : 5.00
##
    1st Qu.:2.000
                     1st Qu.:14.00
                                      1st Qu.:2.000
                                                       1st Qu.:21.00
                     Median :17.00
                                      Median :3.000
                                                       Median :25.00
##
   Median :3.000
    Mean
           :3.044
                            :17.27
                                             :3.093
                                                              :25.06
##
                     Mean
                                      Mean
                                                       Mean
                     3rd Qu.:20.00
##
    3rd Qu.:4.000
                                      3rd Qu.:4.000
                                                       3rd Qu.:29.00
           :7.000
                     Max.
                            :30.00
##
    Max.
                                      Max.
                                              :9.000
                                                       Max.
                                                               :45.00
##
        artyzm
                       miejsce
                                         odbicia
                                                           ostrosc
           :1.00
                           :0.0000
                                              :0.0000
                                                                :1.000
##
    Min.
                    Min.
                                      Min.
                                                        Min.
    1st Qu.:4.00
                    1st Qu.:0.0000
                                      1st Qu.:0.0000
                                                        1st Qu.:3.000
##
##
   Median:4.00
                    Median :1.0000
                                      Median :0.0000
                                                        Median :4.000
                           :0.5826
                                                                :3.963
##
    Mean
           :3.85
                    Mean
                                      Mean
                                              :0.2971
                                                        Mean
##
    3rd Qu.:4.00
                    3rd Qu.:1.0000
                                      3rd Qu.:1.0000
                                                        3rd Qu.:5.000
                           :1.0000
                                              :1.0000
                                                                :6.000
##
    Max.
           :6.00
                    Max.
                                      Max.
                                                        Max.
##
                        kolor
        ocena
##
   Min.
           :1.00
                    Min.
                           :1.00
    1st Qu.:3.00
                    1st Qu.:3.00
##
##
    Median:4.00
                    Median:3.00
           :3.72
##
    Mean
                    Mean
                           :3.12
    3rd Qu.:5.00
                    3rd Qu.:4.00
##
           :6.00
##
   Max.
                    Max.
                           :4.00
```

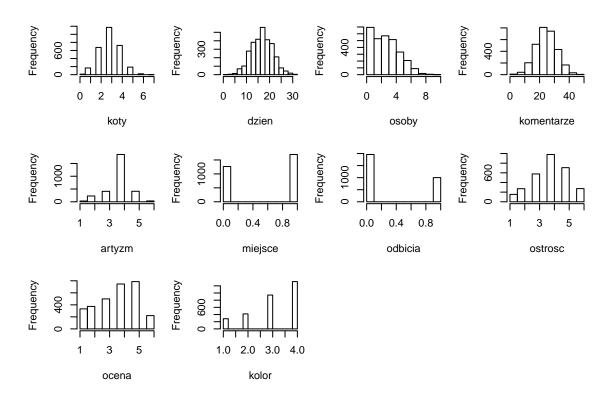
Przy pomocy pakietu "moments" zbudowano tabele z podstawowymi charakterystykami zmiennych w modelu.

```
kurtosis<-sapply(1:ncol(zdjecia.random),function(x) kurtosis(zdjecia.random[,x]))
skewness<-sapply(1:ncol(zdjecia.random),function(x) skewness(zdjecia.random[,x]))
sd<-sapply(1:ncol(zdjecia.random),function(x) sd(zdjecia.random[,x]))
mean<-sapply(1:ncol(zdjecia.random),function(x) mean(zdjecia.random[,x]))
bound<-mean +3*sd
jarque.stat<-sapply(1:ncol(zdjecia.random),function(x) jarque.test(zdjecia.random[,x])$statistic)
stats<-matrix(1:(ncol(zdjecia0)*6),ncol=ncol(zdjecia0),nrow=6)
stats<-rbind(kurtosis,skewness,jarque.stat,sd,mean,bound)
colnames(stats)<- colnames(zdjecia0)
stats</pre>
```

```
##
                koty
                         dzien
                                  osoby komentarze
                                                    artyzm
## kurtosis
           3.23745291
                     2.91991493
                               2.6165719 3.1418756
                                                 4.8621288
## skewness
           ## jarque.stat 1.91156766 0.68422148 12.3027748 2.1248894 219.0633588
## sd
            1.03435837 4.55853726 1.8201778 6.6820087
                                                 0.8018188
## mean
           3.04353393 17.26632522 3.0934699 25.0576184
                                                 3.8501921
## bound
           6.14660904 30.94193702 8.5540032 45.1036445
                                                 6.2556485
##
               miejsce
                        odbicia
                                 ostrosc
                                            ocena
                                                    kolor
## kurtosis
             1.1121890
                       1.7889658 2.6325425 2.1929912 2.6723447
## skewness
            ## jarque.stat 130.5762485 150.4227830 22.5144641 44.2531658 95.3557914
## sd
             0.5825864 0.2970551 3.9628681 3.7195903 3.1203585
## mean
## bound
             2.0629314    1.6688165    7.7968104    8.1323213    6.0040387
```

W celu potwierdzenia rozkladu zmiennych, na jedym wykresie przedstawiono hiostogramy dla wszystkich zmiennych. Opcjonalnie mozna wykorzystac funkcje qplot z pakietu ggplot2 przedstawiajac ten sam wykres za pomoca jednej funkcji.

```
par(mfrow=c(3,4), mar=c(4,4,2,1), oma=rep(2,4))
for(i in 1:ncol(zdjecia0)){
   hist(zdjecia0[,i],xlab = colnames(zdjecia0)[i],main = "")
}
```



Wiekszosc zmiennych posiada charakterystyki nie pozwalajace odrzucic hipotezy zerowej zakladajacej normalnosc rozkladu. Zmienne typu factor których poziomy nie reprezentuja "rozwoju" danej cechy, posiadaja chrakterystyki nie pozwalajace na przyjecie hipotezy zerowej. Nalezy wyróznic zmienna osoby która posiada charakterystyki rozkladu wykladniczego, a koty juz nie.

Korelacja spearmana pomiedzy wszystkimi zmiennymi.

round(cor(zdjecia0, method = c("spearman")), 2)

```
##
              koty dzien osoby komentarze artyzm miejsce odbicia ostrosc
## koty
              1.00 0.02 -0.27
                                    -0.19 -0.17
                                                   0.18
                                                           0.14
                                                                  -0.16
## dzien
              0.02 1.00 0.24
                                    -0.19 -0.18
                                                  -0.25
                                                           0.11
                                                                   0.30
## osoby
             -0.27 0.24 1.00
                                    0.32
                                           0.28
                                                   0.29
                                                          -0.19
                                                                   0.08
## komentarze -0.19 -0.19 0.32
                                    1.00
                                           0.86
                                                   0.26
                                                          -0.30
                                                                  -0.24
## artyzm
             -0.17 -0.18 0.28
                                    0.86
                                           1.00
                                                   0.23
                                                          -0.28
                                                                  -0.22
## miejsce
              0.18 -0.25 0.29
                                    0.26
                                           0.23
                                                   1.00
                                                          -0.17
                                                                  -0.02
## odbicia
              0.14 0.11 -0.19
                                    -0.30 -0.28
                                                  -0.17
                                                          1.00
                                                                  -0.15
             -0.16 0.30 0.08
                                    -0.24 -0.22
## ostrosc
                                                  -0.02
                                                          -0.15
                                                                   1.00
                                                          -0.07
## ocena
              0.35 -0.41 -0.06
                                    0.16
                                          0.16
                                                   0.22
                                                                  -0.13
                                   -0.36 -0.31
                                                                   0.30
## kolor
              0.08 0.10 0.01
                                                   0.04
                                                          -0.02
##
             ocena kolor
## koty
              0.35 0.08
## dzien
             -0.41 0.10
## osoby
             -0.06 0.01
## komentarze 0.16 -0.36
## artyzm
              0.16 - 0.31
## miejsce
              0.22 0.04
## odbicia
             -0.07 -0.02
## ostrosc
             -0.13 0.30
## ocena
              1.00 -0.18
## kolor
             -0.18 1.00
```

Korelacja Pearsona pomiedzy wszystkimi zmiennymi.

```
round(cor(zdjecia0, method = c("pearson")),2)
```

```
##
               koty dzien osoby komentarze artyzm miejsce odbicia ostrosc
## koty
               1.00
                     0.03 -0.28
                                      -0.20
                                             -0.18
                                                       0.18
                                                               0.14
                                                                      -0.16
                                             -0.20
                                                     -0.25
                                                               0.12
## dzien
               0.03
                     1.00 0.24
                                      -0.21
                                                                       0.31
## osoby
              -0.28
                     0.24
                           1.00
                                       0.32
                                              0.28
                                                       0.29
                                                              -0.20
                                                                       0.08
## komentarze -0.20 -0.21
                                              0.90
                                                      0.26
                                                              -0.30
                                                                      -0.25
                           0.32
                                       1.00
              -0.18 -0.20
                                       0.90
                                              1.00
                                                      0.23
                                                              -0.28
                                                                      -0.22
## artyzm
                           0.28
## miejsce
               0.18 -0.25
                           0.29
                                       0.26
                                              0.23
                                                      1.00
                                                              -0.17
                                                                      -0.03
                                      -0.30
                                                                      -0.15
## odbicia
               0.14
                     0.12 -0.20
                                             -0.28
                                                     -0.17
                                                               1.00
## ostrosc
              -0.16
                     0.31
                           0.08
                                      -0.25
                                             -0.22
                                                     -0.03
                                                              -0.15
                                                                       1.00
## ocena
               0.35 -0.41 -0.06
                                       0.16
                                              0.15
                                                      0.22
                                                              -0.08
                                                                      -0.13
## kolor
               0.09 0.11 0.01
                                      -0.36 -0.31
                                                      0.04
                                                              -0.01
                                                                       0.30
              ocena kolor
##
## koty
               0.35
                     0.09
## dzien
              -0.41
                     0.11
## osoby
              -0.06 0.01
## komentarze
              0.16 - 0.36
## artyzm
               0.15 - 0.31
## miejsce
               0.22 0.04
## odbicia
              -0.08 -0.01
## ostrosc
              -0.13 0.30
## ocena
               1.00 -0.18
## kolor
              -0.18 1.00
```

Nalezy wyróznic korelacje pomiedzy zmiennymi komentarz oraz artyzm, która jest najwieksza co do wartosci bezwzglednej - 0.8980848. Relacja ta wydaje sie racjonalna. Druga korelacja warta wyróznienia jest relacja pomiedzy dniem miesiaca publikacji oraz ocena. Interpretacja tego wspólczynnika ukazuje potencjal publikowania zdjec na poczatku miesiaca. Nalezy zaznaczyc ze zmienna dla dnia publikacji ma charakterystyki wskazujące na rozkład istotnie normalny.

Testy na istotnosc pozwalaja na odrzucenie hipotezy zerowej zakladajacej korelacje równa zeru.

*TESTY NA ISTOTNOSC DWÓCH NAJBARDZIEJ INTERESUJACYCH KORELACJI ZAPREZENTOWANO NA DRUGIEJ STRONIE

```
cor.test(as.numeric(zdjecia0$artyzm),as.numeric(zdjecia0$komentarze),method = c("pearson"))
##
##
   Pearson's product-moment correlation
##
## data: as.numeric(zdjecia0$artyzm) and as.numeric(zdjecia0$komentarze)
## t = 111.07, df = 2959, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.8908837 0.9048346
## sample estimates:
##
         cor
## 0.8980848
cor.test(as.numeric(zdjecia0$ocena), as.numeric(zdjecia0$dzien), method = c("spearman"))
## Warning in cor.test.default(as.numeric(zdjecia0$ocena),
## as.numeric(zdjecia0$dzien), : Cannot compute exact p-value with ties
##
   Spearman's rank correlation rho
##
## data: as.numeric(zdjecia0$ocena) and as.numeric(zdjecia0$dzien)
## S = 6081800000, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##
          rho
## -0.4056195
```

Canonical Correlations

Stworzono zmienne konieczne do przeprowadzenia analizy. Kryterium doboru zmiennych było narzucone z góry.

```
zbior1<-c("artyzm", "ostrosc", "ocena")
zbior2<-c("kolor2", "kolor3", "kolor4", "koty", "miejsce2", "odbicia2", "dzien", "osoby")
matY<-zdjecia02[,zbior1]
matX<-zdjecia02[,zbior2]
corrall<-matcor(matX,matY)
corrall$XYcor</pre>
```

```
##
                 kolor2
                             kolor3
                                        kolor4
                                                              miejsce2
                                                     koty
## kolor2
            1.0000000000 -0.27676090 -0.36261720 -0.01568453 -0.007554501
## kolor3
           -0.2767609033 1.00000000 -0.61225761 -0.01520725 -0.017777133
## kolor4
           -0.3626171977 -0.61225761 1.00000000 0.07232030 0.037196226
## koty
           -0.0156845282 -0.01520725 0.07232030 1.00000000
                                                           0.182786069
## miejsce2 -0.0075545010 -0.01777713 0.03719623
                                               0.18278607
                                                           1.000000000
## odbicia2 -0.0225184220 0.02943472 -0.02235972 0.14026794 -0.165036439
## dzien
           -0.0448402407 -0.01317768 0.09018178 0.02574509 -0.250060550
## osoby
           -0.0003863498 -0.02022860 0.01873075 -0.28194470 0.286915235
## artyzm
            0.1163630200 0.08690286 -0.28245211 -0.18057792 0.231579762
## ostrosc -0.1198801629 -0.06119485 0.26504809 -0.16037719 -0.025652641
## ocena
            0.224193457
##
              odbicia2
                            dzien
                                         osoby
                                                    artyzm
## kolor2
           -0.02251842 -0.04484024 -0.0003863498
                                               0.11636302 -0.11988016
## kolor3
            0.02943472 -0.01317768 -0.0202285993 0.08690286 -0.06119485
## kolor4
           -0.02235972 0.09018178 0.0187307472 -0.28245211 0.26504809
## koty
            ## miejsce2 -0.16503644 -0.25006055 0.2869152345 0.23157976 -0.02565264
## odbicia2 1.00000000 0.11917154 -0.1993564115 -0.27839121 -0.14752440
## dzien
            0.11917154 \quad 1.00000000 \quad 0.2376557400 \quad -0.19518259 \quad 0.31168792
## osoby
           -0.19935641 0.23765574 1.000000000 0.28336508
                                                           0.08392537
## artyzm
           -0.27839121 -0.19518259 0.2833650768 1.00000000 -0.22461926
## ostrosc -0.14752440 0.31168792 0.0839253673 -0.22461926 1.00000000
           -0.07538635 -0.41102732 -0.0603780705 0.15013748 -0.12761502
## ocena
##
                ocena
## kolor2
            0.08193121
## kolor3
            0.03668024
## kolor4
           -0.15883778
## koty
            0.35400384
## miejsce2 0.22419346
## odbicia2 -0.07538635
           -0.41102732
## dzien
## osoby
           -0.06037807
## artyzm
            0.15013748
## ostrosc -0.12761502
## ocena
            1.00000000
```

W macierzy nie mozna wyróznic istotnej ilosci duzych co do wartosci bezwzglednej korelacji.

```
cc1 <- cc(matX,matY)
# sklad oraz rodzaje wynikow dla funkcji cc
str(cc1)</pre>
```

```
## List of 5
## $ cor : num [1:3] 0.699 0.515 0.256
## $ names :List of 3
    ..$ Xnames : chr [1:8] "kolor2" "kolor3" "kolor4" "koty" ...
    ..$ Ynames : chr [1:3] "artyzm" "ostrosc" "ocena"
##
     ..$ ind.names: chr [1:2961] "1" "2" "3" "4" ...
## $ xcoef : num [1:8, 1:3] 0.49 0.765 1.352 -0.613 0.162 ...
    ..- attr(*, "dimnames")=List of 2
    ....$ : chr [1:8] "kolor2" "kolor3" "kolor4" "koty" ...
##
    ....$ : NULL
   $ ycoef : num [1:3, 1:3] -0.128 0.416 -0.511 -1.235 -0.259 ...
##
##
    ..- attr(*, "dimnames")=List of 2
    ....$ : chr [1:3] "artyzm" "ostrosc" "ocena"
##
     .. ..$ : NULL
##
## $ scores:List of 6
##
    ..$ xscores
                     : num [1:2961, 1:3] -2.389 -0.137 -0.575 1.469 -1.414 ...
##
     ...- attr(*, "dimnames")=List of 2
     .. .. ..$ : chr [1:2961] "1" "2" "3" "4" ...
##
##
     .. .. ..$ : NULL
##
     ..$ yscores
                      : num [1:2961, 1:3] -1.334 -0.472 -0.662 1.8 -2.166 ...
##
     ...- attr(*, "dimnames")=List of 2
##
     ....$ : chr [1:2961] "1" "2" "3" "4" ...
     .. .. ..$ : NULL
##
     ..$ corr.X.xscores: num [1:8, 1:3] -0.1961 -0.0988 0.4136 -0.4723 -0.2941 ...
##
     ... - attr(*, "dimnames")=List of 2
     .....$ : chr [1:8] "kolor2" "kolor3" "kolor4" "koty" ...
##
     .. .. ..$ : NULL
##
##
     ..$ corr.Y.xscores: num [1:3, 1:3] -0.2359 0.4539 -0.5804 -0.4782 -0.0617 ...
     ... - attr(*, "dimnames")=List of 2
##
     .. .. ..$ : chr [1:3] "artyzm" "ostrosc" "ocena"
##
     .. .. ..$ : NULL
##
##
     ..$ corr.X.yscores: num [1:8, 1:3] -0.1371 -0.0691 0.2891 -0.3301 -0.2056 ...
##
     ... - attr(*, "dimnames")=List of 2
     .....$ : chr [1:8] "kolor2" "kolor3" "kolor4" "koty" ...
##
     .. .. ..$ : NULL
##
##
     ..$ corr.Y.yscores: num [1:3, 1:3] -0.338 0.649 -0.83 -0.928 -0.12 ...
     ... - attr(*, "dimnames")=List of 2
##
     .....$ : chr [1:3] "artyzm" "ostrosc" "ocena"
##
     .. .. ..$ : NULL
##
```

Zinterpretuj powiazania miedzy zbiorami zmiennych, analizujac wielkosci standaryzowanych wag kanonicznych i ladunków czynnikowych. Czy wystepuja róznice? Jakie? Interpretacja powinna byc przeprowadzona jedynie dla istotnych korelacji kanonicznych.

Warto ukazac najwazniejsze wyniki otrzymane z funkcji cc.

```
cc1$cor
```

[1] 0.6990544 0.5150777 0.2562271

```
cc1$ycoef
```

```
## [,1] [,2] [,3]

## artyzm -0.1279292 -1.2347593 0.08647124

## ostrosc 0.4162045 -0.2591983 -0.64425872

## ocena -0.5112507 0.1089434 -0.45933313
```

cc1\$xcoef

```
[,1]
                       [,2]
                                [,3]
##
## kolor2
         ## kolor3
         ## kolor4
         1.35169665 1.053796889 -1.64884308
## koty
        ## miejsce2 0.16208561 -0.630496519 -0.55895285
## odbicia2 -0.06930292 0.974533662 1.30897499
         0.16929266 -0.009423395 -0.01496501
## dzien
## osoby
        -0.16269409 -0.171015591 0.03798971
```

Pierwszy wspólczynnik korelacj kanonicznej jest równy 0.6990544 - i.e. najwiekszy mozliwy do osiagniecia wspólczynnik otrzymany z liniowyej kombinacji zmiennych z obu badanych zbiorow. Zgodnie z oczekiwaniami policzony wspólczynnik jest wiekszy niz dowolny wspólczynnik z macierzy korelacji miedzy zmienna ze zbioru X, a zmienna ze zbioru Y.

Poniewaz zmienne ukazano na róznych skalach oraz maja rózna wariancje powinnismy analizowac "wystandaryzowane wagi kanoniczne". Róznica w interpretacji polega jedynie na zmianie miar zmian danej zmiennej.

Standardized Canonical Coefficients

```
wyniki<-list()
wyniki[[1]] <- diag(sqrt(diag(cov(matY)))) %*% cc1$ycoef
rownames(wyniki[[1]])<-rownames(cc1$ycoef)
wyniki[[2]] <- diag(sqrt(diag(cov(matX)))) %*% cc1$xcoef
rownames(wyniki[[2]])<-rownames(cc1$xcoef)
wyniki</pre>
```

```
## [[1]]
##
               [,1]
                         [,2]
                                   [,3]
## artyzm
         -0.1063451 -1.0264315 0.07188187
## ostrosc 0.5300802 -0.3301163 -0.82053126
         -0.7466721 0.1591099 -0.67084741
## ocena
##
## [[2]]
##
                [,1]
                           [,2]
                                      [,3]
           0.17059743
                     0.13264550 -0.22264605
## kolor2
           ## kolor3
## kolor4
           ## koty
          -0.63026500 0.44082126 -0.42936781
## miejsce2 0.08019321 -0.31194341 -0.27654659
## odbicia2 -0.03278925 0.46108061 0.61931467
           0.77085433 -0.04290833 -0.06814142
## dzien
## osoby
          -0.31072910 -0.32662232 0.07255647
```

Analizujemy kilka losowych zmiennych o najwyzszych co do wartosci bezwglednej wartosciach wspólczynników. Nelezy zaznaczyc ze wyniki mozna analizowac dwukierunkowo. Przemnozenie stron przez minus jeden powoduje odwrócenie wnioskowania przy zachowaniu wiarygodności interpretacji.

Pierwsza zmienna kanoniczna – wysoka ostrosc(0.5300802) oraz slaba ocena(-0.7466721) - kolor – kolor inny niz zielony niebieksi i filetowy powinien wpływac ujemnie na ocene lub/i dodatnio na ostrosc - koty – mniej kotów to gorsza ocena lub/i lepsza ostrosc - dzien – im wczesniej w miesiacu zrobimy zdjecie tym lepiej dla oceny i/lub otrzymamy lepsza ostrosc (moze jestesmy bardziej wypoczeci;p)

Druga zmienna kanoniczna – slabe walory artystyczne (-1.0264315) - kolor – kolor inny niz zielony niebieksi i filetowy powinien wpływac ujemnie na walory artystyczne zdjecia - koty – wiecej kotów to gorsza wartosc artystyczna

 $\label{thm:condition} \begin{tabular}{ll} Trzecia zmienna kanoniczna - slaba ostrosc oraz slaba ocena - kolor - kolor inny niz zielony niebieksi i filetowy powinien wpływac dodatnio na ostrosc i/lub ocene - odbicia - wystepowanie odbicia powinien wpływac ujemnie na ostrosc i/lub ocene \\ \end{tabular}$

Ile par zmiennych kanonicznych wybrac?

Procedura Liczenia lambdy WILKSA powinna ulatwic analize wiekszych zbiorów danych - i.e. wyboru istotnych zmiennych kanonicznych do dalszej analizy.

```
WILKSL<-function(matX,matY,cc1){</pre>
ev <- (1 - cc1$cor^2)
n <- dim(matX)[1]</pre>
p <- length(matX)</pre>
q <- length(matY)
k \leftarrow min(p, q)
m < -n - 3/2 - (p + q)/2
w <- rev(cumprod(rev(ev)))</pre>
# initialize
d1 <- d2 <- f <- vector("numeric", k)</pre>
for (i in 1:k) {
    s \leftarrow sqrt((p^2 * q^2 - 4)/(p^2 + q^2 - 5))
    si <- 1/s
    d1[i] \leftarrow p * q
    d2[i] <- m * s - p * q/2 + 1
    r \leftarrow (1 - w[i]^si)/w[i]^si
    f[i] <- r * d2[i]/d1[i]
    p < -p - 1
    q < -q - 1
pv <- pf(f, d1, d2, lower.tail = FALSE)</pre>
dmat <- cbind(WilksL = w, F = f, df1 = d1, df2 = d2, p = pv)</pre>
return(dmat)
## source: http://www.ats.ucla.edu/stat/r/dae/canonical.htm
```

FUNKCJA DO SZYBKIEGO POLICZENIA LAMBDY WILKSA

```
WILKS2<-function(cc1){
   dmat2<-matrix(0,nrow=ncol(matY),ncol=2)
   sapply(1:ncol(matY),function(i) dmat2[i,1]<<-pre>col((1-cc1$cor^2)[i:(ncol(matY))]))
   return(dmat2)
}
```

Testowanie istotnosci koreacji kanonicznych dla analizowanych danych.

WILKSL(matX,matY,cc1)

```
## WilksL F df1 df2 p
## [1,] 0.3510030 154.99349 24 8556.499 0.000000e+00
## [2,] 0.6864606 87.24763 14 5902.000 2.001233e-228
## [3,] 0.9343477 34.57059 6 2952.000 1.417238e-40
```

WILKS2(cc1)

```
## [,1] [,2]
## [1,] 0.3510030 0
## [2,] 0.6864606 0
## [3,] 0.9343477 0
```

Wspólczynnik redundancji

Redundacja mowi nam ile przecietnie wariancji w jednym zbiorze jest wyjasnione przez dana zmienna kanoniczna przy danym innym zbiorze zmiennych. Wartosc redundacji calkowitej moze byc wazna analityczna informacja o naszym modelu. Mowi o procencie calkowitej wariancji jednego zbioru wyjasniona w ramach modelu.

```
REDUNT<-function(matX,matY,cc1){</pre>
    eigenmatY<-cc1$cor
    vector1<-vector(,length(matY))</pre>
    sapply(1:ncol(matY),function(i) vector1[i]<<-sqrt(eigenmatY[i]))</pre>
    names1<-c("opposite variance", "own variance")</pre>
    names2<-c("prop stdvar v","prop stdvar u")</pre>
    matim<<-list()</pre>
    for(i in (1:ncol(matY))){
        a<-round(sum((cc1$scores$corr.Y.xscores[,i])^2)/ncol(matY),3)
        b<-round(a/vector1[i],3)
        c<-round(sum((cc1$scores$corr.X.yscores[,i])^2)/ncol(matX),3)</pre>
        d<-round(c/vector1[i],3)</pre>
        assign(paste0("amat",i),matrix(c(c,d,a,b),byrow=TRUE,nrow=2,ncol=2,dimnames=list(names2,names1)
        matim[[i]]<<-get(paste0("amat",i))</pre>
    }
    return(matim)
}
```

Zastosotanie funkcji REDUNT

```
REDUNT(matX,matY,cc1)
```

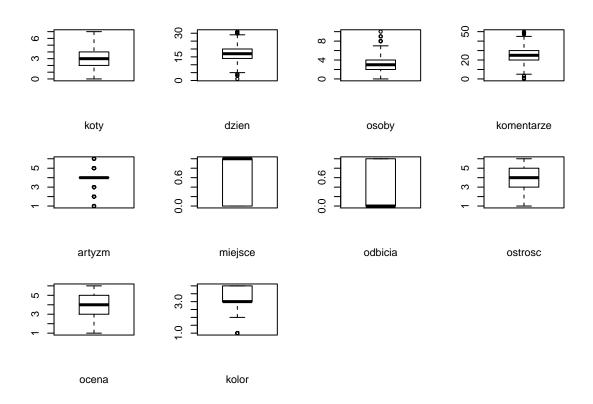
```
## [[1]]
                  opposite variance own variance
                                            0.075
## prop stdvar v
                              0.063
## prop stdvar u
                              0.200
                                            0.239
##
## [[2]]
##
                  opposite variance own variance
## prop stdvar v
                              0.047
                                            0.065
## prop stdvar u
                              0.078
                                            0.109
##
## [[3]]
##
                  opposite variance own variance
## prop stdvar v
                              0.009
                                            0.018
                              0.020
                                            0.040
## prop stdvar u
```

DANE W PROCENTACH:

- 1 zmienna kanoniczna wyjasnia przecietnie 7.5 zmienności w zbiorze Y w oparciu o X. Redundancja wynosi 6.3
- 2 zmienna kanoniczna wyjasnia przecietnie 6.5 zmienności w zbiorze Y w oparciu o X. Redundancja wynosi 4.7
- 2 zmienna kanoniczna wyjasnia przecietnie 1.8 zmienności w zbiorze Y w oparciu o X. Redundancja wynosi 4.7

Sprawdz, czy w zbiorze danych wystepuja obserwacje odstajace i okresl, czy budza one podejrzenia. Jako etykiety mozesz uzyc zmiennej nr zdjecia.

```
par(mfrow=c(3,4), mar=c(4,4,2,1), oma=rep(2,4))
for(i in 1:ncol(zdjecia0)){
   boxplot(zdjecia0[,i],xlab = colnames(zdjecia0)[i],main = "")
}
```



```
outliers<-list()
for(i in 1:ncol(zdjecia0)){
   tt<-zdjecia0[which(zdjecia0[,i]>stats[6,i]),]
   outliers[[i]]<-tt
}
   names(outliers)<-c(colnames(zdjecia0))</pre>
```

OUTLIERS

outliers[1:4]

```
## $koty
       koty dzien osoby komentarze artyzm miejsce odbicia ostrosc ocena
## 2586 7 21
                    0
                           18
                                       3 1
                                                       1
       kolor
## 2586
##
## $dzien
       koty dzien osoby komentarze artyzm miejsce odbicia ostrosc ocena
## 478
                31
                      5
                                24
                                         4
                                                0
                                                        0
## 539
           2
                31
                       6
                                22
                                         4
                                                 0
                                                        0
                                                                 4
                                                                      1
## 2249
                      5
                                                                 6
           3
               31
                                 18
                                         3
                                                 0
                                                         0
                                                                      1
           3
               31
                      5
                                15
                                         2
                                                 0
                                                                5
## 2378
                                                         1
                                                                      1
##
       kolor
## 478
           3
## 539
## 2249
            3
## 2378
            4
##
## $osobv
       koty dzien osoby komentarze artyzm miejsce odbicia ostrosc ocena
## 201
          1
               15
                      9
                                32
                                         4
                                                 1
                                                        0
## 270
           3
               18
                      10
                                27
                                         4
                                                 1
                                                        0
                                                                 4
                                                                      3
## 1025
           3
               16
                      9
                                34
                                         5
                                                 1
                                                        0
                                                                2
                                                                      5
## 1048
        1
             19
                      9
                                31
                                         4
                                                1
                                                        0
                                                                      3
## 2287
           2
             20
                      9
                                38
                                         5
                                                        0
                                                                      3
                                                1
                                27
                                                                      2
## 2336
              19
                      9
                                                                5
           0
                                         4
                                                 1
                                                        0
## 2934
           1
               12
                      9
                                39
                                         5
                                                 1
                                                        0
                                                                2
                                                                      3
##
       kolor
## 201
            3
## 270
            4
## 1025
            4
## 1048
## 2287
            3
## 2336
            3
## 2934
##
## $komentarze
       koty dzien osoby komentarze artyzm miejsce odbicia ostrosc ocena
## 5
                21
                                         6
           4
                      7
                                49
                                                1
                                                        0
                                                                1
## 459
           1
                11
                       4
                                49
                                         6
                                                 1
                                                        0
                                                                3
                                                                      5
## 740
                9
                       6
                                47
                                                 1
                                                                4
                                                                      5
           1
                                         6
                                                        0
## 851
           2
                9
                       4
                                50
                                         6
                                                 1
                                                        0
                                                                2
                                                                      4
## 2635
           4
               15
                      3
                                48
                                         6
                                                 1
                                                        1
                                                                1
       kolor
##
## 5
            1
## 459
            1
## 740
## 851
            1
## 2635
```

Jakich wskazówek udzielisz Aminie na podstawie przeprowadzonej analizy? Jakie cechy powinny miec publikowane zdjecia, zeby uzyskiwaly wyzsze oceny? Uzasadnij.

Klasyczne wnioskowanie skazuje na potencjal publikacji zdjec na poczatku tygodnia. W przypadku komentowania zdjecia badz artyzmu ciezko wyznaczyc racjonalny zwiazek przycznowo skutkowy. Zacheta do komentowania moze stanowic poprawe oceny za artyzm. Zacheta moze byc np. kontrawersyjnosc zdjecia. Analiza kanoniczna pozwala na bardziej szczególowe zbadanie zaleznosci pomiedzy zmiennymi. Duzo kotów na zdjeciu powinno prowadzic do wyzszych ocen lecz koty prawdop. ciezo sfotografowac co powoduje mniejsze wyniki za ostrosc. Rozwiazaniem moze byc zakup lepszego aparatu. Wieksza ilosc kotów na zdjieciu moze takze prowadzic do nizszej oceny za walory artystyczne. Brak odbicia na zdjeciu powinien wpływac dodatnio na wyniki.