348HWw6

Shay Lebovitz

10/18/2020

1) *a)*

```
flour_data <- read.table('/Users/shaylebovitz/R/flour.txt', header = TRUE)
(cor_mat <- cor(flour_data))
```

```
##                  ash     protein    glutenin      moisture     lactic
## ash       1.00000000 -0.06238327 -0.10350563 -0.0581970690 0.13212860
## protein  -0.06238327  1.00000000  0.86945983 -0.2859916887 0.12514129
## glutenin -0.10350563  0.86945983  1.00000000 -0.0270066490 0.17808384
## moisture -0.05819707 -0.28599169 -0.02700665  1.0000000000 0.07759354
## lactic    0.13212860  0.12514129  0.17808384  0.0775935407 1.00000000
## weight    0.09789416  0.40215965  0.26181042 -0.0005899613 0.25261627
##                 weight
## ash       0.0978941603
## protein   0.4021596469
## glutenin  0.2618104210
## moisture -0.0005899613
## lactic    0.2526162701
## weight    1.0000000000
```

`protein` and `glutenin` are highly correlated, which makes sense as glutenin is a type of protein. The only other correlation of any significance is `weight` and `protein`, which makes sense as protein is likely more dense than carbohydrates. `weight` and `glutenin` is also decently correlated at 0.26, as well as `moisture` and `protein` at -.29.

b)

```
flour <- flour_data[, 1:3]
wheat <- flour_data[, 4:6]
flour_cor <- cancor(flour, wheat)
print(flour_cor)
```

```
## $cor
## [1] 0.67047004 0.28143464 0.06792907
##
## $xcoef
##                [,1]        [,2]         [,3]
## ash       0.1426455  0.63102862  1.036393524
## protein   0.2858834 -0.09319866 -0.002521618
## glutenin -0.5197908  0.52133076 -0.157208041
##
## $ycoef
##                 [,1]       [,2]        [,3]
## moisture -0.20968339 0.09273076 -0.14978483
## lactic   -0.15596027 1.20093511  1.21369813
```

```
## weight    0.06403856 0.04054605 -0.06845426
##
## $xcenter
##       ash   protein  glutenin
## 0.5578125 9.9959375 3.4962500
##
## $ycenter
## moisture   lactic   weight
## 10.94688  0.33500 59.18125
```

```
14*flour_cor$xcoef
```

```
##                [,1]      [,2]        [,3]
## ash       1.997037  8.834401 14.50950933
## protein   4.002368 -1.304781 -0.03530266
## glutenin -7.277071  7.298631 -2.20091257
```

```
20*flour_cor$ycoef
```

```
##                [,1]      [,2]      [,3]
## moisture -4.193668  1.854615 -2.995697
## lactic   -3.119205 24.018702 24.273963
## weight    1.280771  0.810921 -1.369085
```

The first canonical correlation is 0.6705, the second is 0.2814, and the third is 0.0680. For flour, the first canonical variables are roughly -7glutenin + 4protein + 2ash. For wheat, the first canonical variables are roughly -4moisture - 3lactic + weight

```
flour.cc1<-as.matrix(flour)%*%flour_cor$xcoef[,1]
wheat.cc1<-as.matrix(wheat)%*%flour_cor$ycoef[,1]

cor(flour.cc1, wheat.cc1)
```

```
##         [,1]
## [1,] 0.67047
```

We see that the first canonical correlation is 0.67.

```
cor(flour.cc1, flour)
```

```
##            ash   protein  glutenin
## [1,] 0.1427265 0.6988342 0.2589947
```

All the correlations of the first canonical variable with the data are positive, with protein being the greatest. Even though the coefficient of glutenin is negative, it is positively correlated with the first canonical variable.

```
cor(wheat.cc1, wheat)
```

```
##        moisture      lactic    weight
## [1,] -0.7758317 0.009820858 0.6248059
```

Moisture is negatively correlated, whereas weight and lactic are positively correlated with the first canonical variable, although the lactic correlation is very small. The lactic coefficient is negative, but its correlation to the first canonical variable is positive.

2)

```
sales_data <- read.table('/Users/shaylebovitz/R/sales.txt', header = TRUE)
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## number of items read is not a multiple of the number of columns
```

```r
sales_data <- sales_data[1:50, ]
perf <- sales_data[, 1:3]
perf <- as.data.frame(sapply(perf, as.numeric))
tests <- sales_data[, 4:7]
sales_cor <- cancor(perf, tests)
print(sales_cor)
```

```
## $cor
## [1] 0.9944827 0.8781065 0.3836057
##
## $xcoef
##                [,1]        [,2]        [,3]
## Growth 0.008911125  0.02486719 -0.05387899
## Prof   0.002989377 -0.03459487  0.01478786
## New    0.011179739  0.03404200  0.05477358
##
## $ycoef
##               [,1]         [,2]         [,3]         [,4]
## Creat 0.009964020  0.027484475  0.035222369 -0.002672241
## Mech  0.004391186 -0.028796340 -0.020270754 -0.047626303
## Abs   0.012794882  0.070823323 -0.040032008 -0.007549523
## Math  0.008975711 -0.009759438  0.001618942  0.013413863
##
## $xcenter
##  Growth    Prof     New
##  98.836 106.622 102.810
##
## $ycenter
## Creat  Mech   Abs  Math
## 11.22 14.18 10.56 29.76
```

```r
100*sales_cor$xcoef
```

```
##             [,1]      [,2]      [,3]
## Growth 0.8911125  2.486719 -5.387899
## Prof   0.2989377 -3.459487  1.478786
## New    1.1179739  3.404200  5.477358
```

```r
100*sales_cor$ycoef
```

```
##            [,1]       [,2]       [,3]       [,4]
## Creat 0.9964020  2.7484475  3.5222369 -0.2672241
## Mech  0.4391186 -2.8796340 -2.0270754 -4.7626303
## Abs   1.2794882  7.0823323 -4.0032008 -0.7549523
## Math  0.8975711 -0.9759438  0.1618942  1.3413863
```

The first canonical correlation is 0.99, the second is 0.88, and the third is 0.38. This suggests that the two canonical variables are highly correlated. For perf, the canonical variables are roughly growth + new. For tests, the canonical variables are roughly creat + abs + math

```r
perf_cc1<-as.matrix(perf)%*%sales_cor$xcoef[,1]
tests_cc1<-as.matrix(tests)%*%sales_cor$ycoef[,1]

cor(perf_cc1, tests_cc1)
```

```
##           [,1]
```

```
## [1,] 0.9944827
```

We see that the first canonical correlation is 0.99.

```r
cor(perf_cc1, perf)
```

```
##         Growth      Prof      New
## [1,] 0.9798776 0.9464085 0.951862
```

All three variables correlate highly with the canoncical variable. These variables are highly correlated with each other, as someone with a lot of sales growth will likely aslo have sales profitability and new account sales.

```r
cor(perf)
```

```
##           Growth      Prof       New
## Growth 1.0000000 0.9260758 0.8840023
## Prof   0.9260758 1.0000000 0.8425232
## New    0.8840023 0.8425232 1.0000000
```

```r
cor(tests_cc1, tests)
```

```
##          Creat      Mech       Abs      Math
## [1,] 0.6383313 0.7211626 0.6472493 0.9440859
```

All tests are positively correlated with the first canonical variable, with Math being the highest. This reveals taht doing well on the math test likely has the strongest influence on sales performance. Because all four correlations are greater than 0.6, they all are decently important for predicted sales.