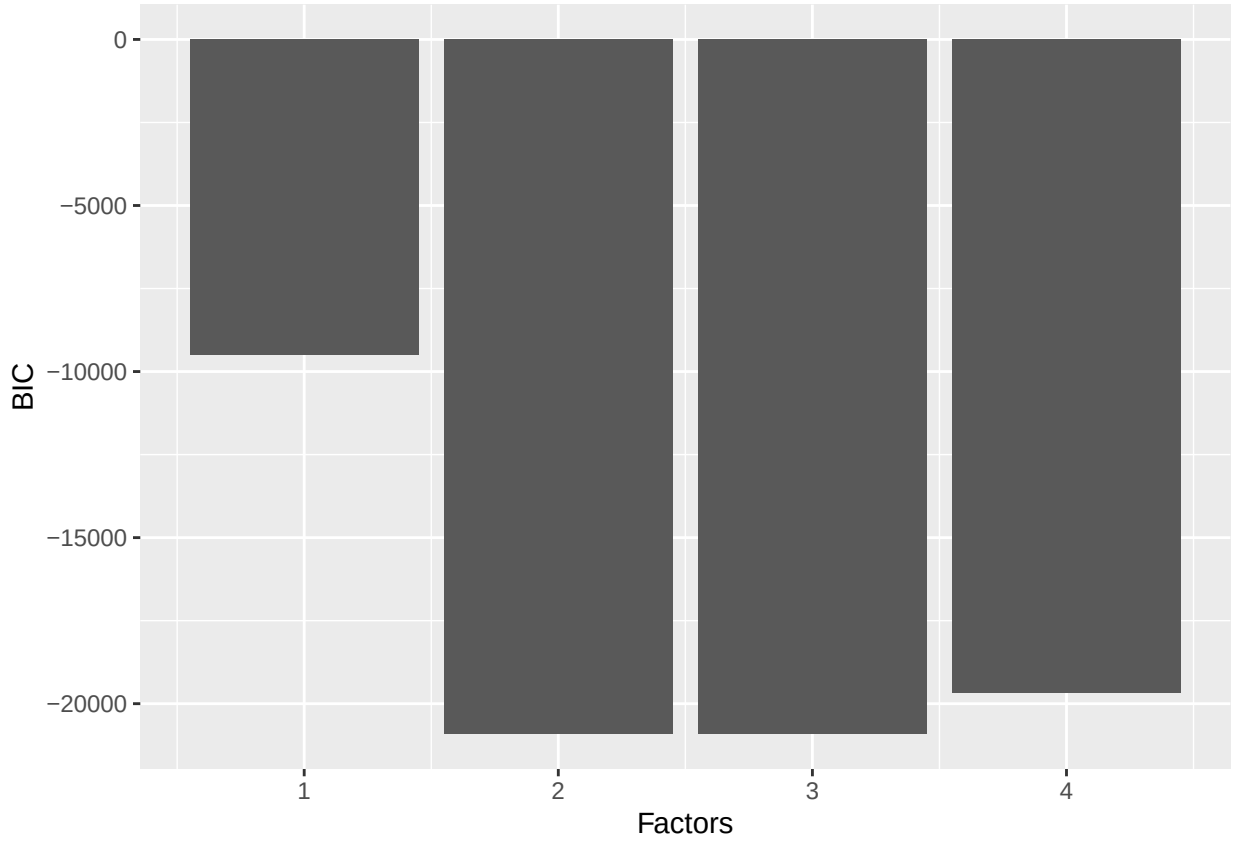
Stat501 Homework 7

Kelby Kies

4/16/2021

Question 1:

Part A.) Provide a factor analysis of the dataset for q = 1, 2, 3, 4 factors. Use BIC to determine the number of optimal factors. [10 points]




```
## [1] "Q=3 Factors are Optimal!"
```

```
## [1] "Here is the full analysis:"
```

```
##
## Call:
```

```
## fad(x = as.matrix(GRB[, -1]), factors = 3, scores = "regression",     rotation = "promax", method =
##
## Uniquenesses:
##   T50   T90    F1    F2    F3    F4   P64  P256 P1024
## 0.036 0.029 0.053 0.005 0.081 0.488 0.033 0.005 0.040
##
## Loadings:
##       [,1]   [,2]   [,3]
##  [1,]  1.018 -0.112
##  [2,]  1.018
##  [3,]  0.606         0.580
##  [4,]  0.577         0.625
##  [5,]  0.561  0.224  0.392
##  [6,]  0.419  0.435
##  [7,] -0.235  1.063
##  [8,]         1.068
##  [9,]  0.200  0.840
##
##                  [,1]  [,2]  [,3]
## SS loadings    3.366 3.239 0.908
## Proportion Var 0.374 0.360 0.101
## Cumulative Var 0.374 0.734 0.835
##
## Factor Correlations:
##         Factor1 Factor2 Factor3
## Factor1   1.000  -0.219  -0.806
## Factor2  -0.219   1.000   0.464
## Factor3  -0.806   0.464   1.000
##
## The BIC is:  -20920.06
```

In order to estimate the number of optimal factors, we want to choose q (#factors) such that our log likelihood function is maximized and thus our BIC is minimized. Here I used the fad() function to run a factor analysis on the GRB data, but I didn't include the class variable. When trying different rotation methods, I found that promax provided me results where I could interpret the lowest # of factors. By looking at the BIC's, we can see that q=3 factors produces the smallest BIC value therefore 3 factors is optimal.

Part B.) Describe the uniquenesses. How do they compare with the original variances? [5 points]

```
## [1] "Here are the uniqueness values from the factor analysis with 3 factors:"

##        T50        T90         F1         F2         F3         F4        P64
## 0.03552007 0.02873954 0.05262395 0.00500000 0.08065347 0.48754868 0.03255673
##       P256      P1024
## 0.00500000 0.03972816

## [1] "Here are the original Variances:"

##       T50       T90        F1        F2        F3        F4       P64      P256
## 0.8476781 0.8406582 0.8434463 0.7254612 0.6031329 0.7374003 0.1899601 0.2034551
##     P1024
## 0.2395857
```

Here we can see that relatively most of the variance is explained by common factors for the majority of our variables. We can see this when looking at the variable T50, T90, F1, F2, F3, P64 and P256. These uniqueness values are all small! The uniqueness value for the variable F4 accounts for a big portion of the original variance as well.Here the uniqueness value is: 0.48754868 and the original variance of F4 is: 0.7374003.

Part C.) Describe the factor loadings at the optimal number of factors that are obtained by BIC. [10 points]

```
## [1] "Here are the factor loadings for 3 factors:"
```

```
##
## Loadings:
##       [,1]   [,2]   [,3]
##  [1,]  1.018 -0.112
##  [2,]  1.018
##  [3,]  0.606         0.580
##  [4,]  0.577         0.625
##  [5,]  0.561  0.224  0.392
##  [6,]  0.419  0.435
##  [7,] -0.235  1.063
##  [8,]         1.068
##  [9,]  0.200  0.840
##
##                  [,1]  [,2]  [,3]
## SS loadings    3.366 3.239 0.908
## Proportion Var 0.374 0.360 0.101
## Cumulative Var 0.374 0.734 0.835
```

Loadings close to -1 or 1 indicate that the factor strongly influences the variable. Loadings close to 0 indicate that the factor has a weak influence on the variable. Here we can see that the 1st factor has the most influence on T50, T90, F1, F2 and F3. The 2nd factor has the most influence on the variables P64, P256 and P1024.

The 3 factors explain about 93.9% of the total variance.

Question 2: We will now characterize the variability of handwriting in the context of digits. Specifically, we will focus on the zip code dataset and only on the digit 5. (For this reduced dataset, n = 168 $<$ 256 = p so that the fad package will need to be used.)

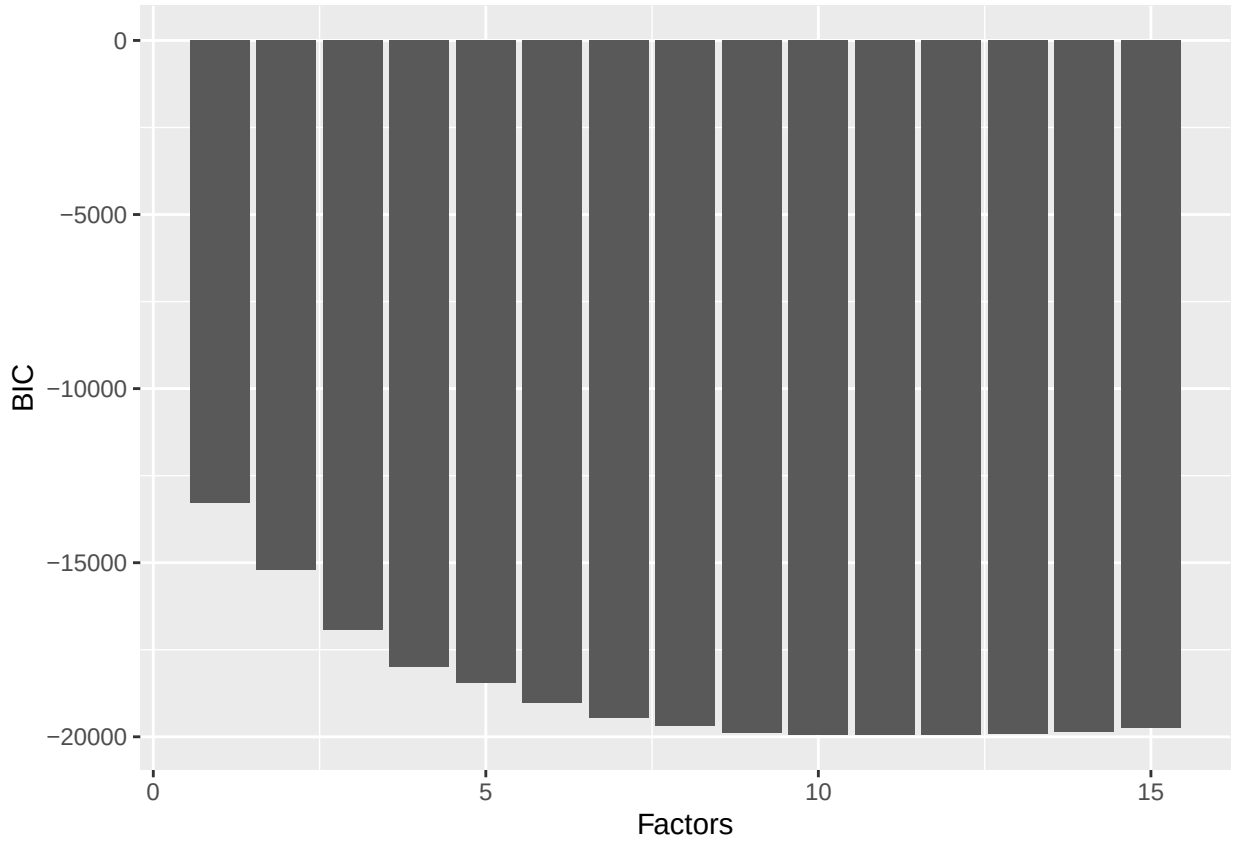
Part A.) Use factor analysis with q=1,2,. . . ,15 factors and BIC to decide on the number of factors. Display the factor loadings obtained and the uniquenesses by means of images. [15 points]

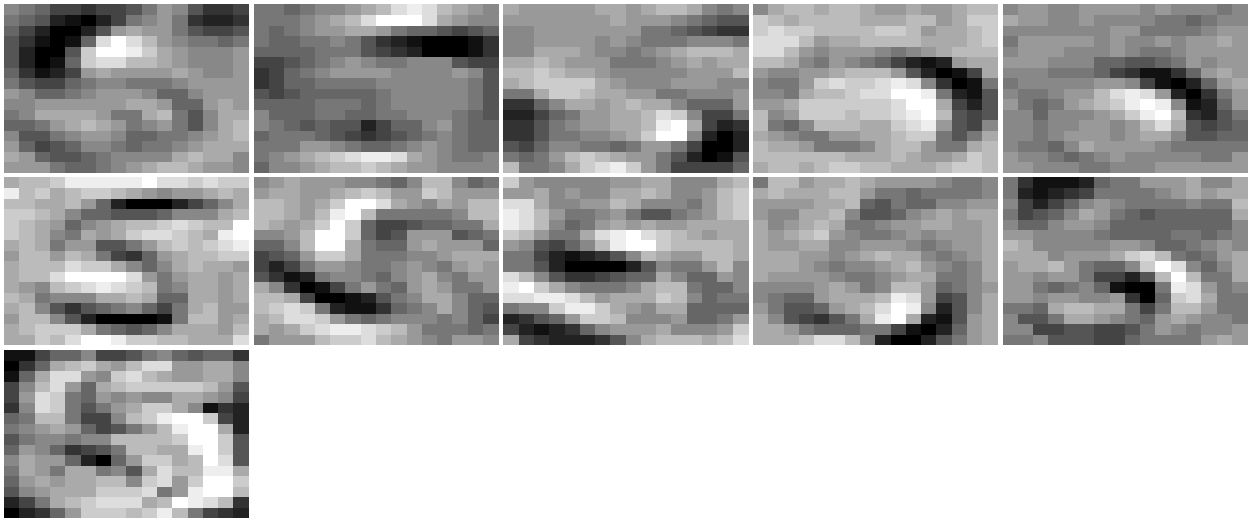
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## [1] 256  10
```





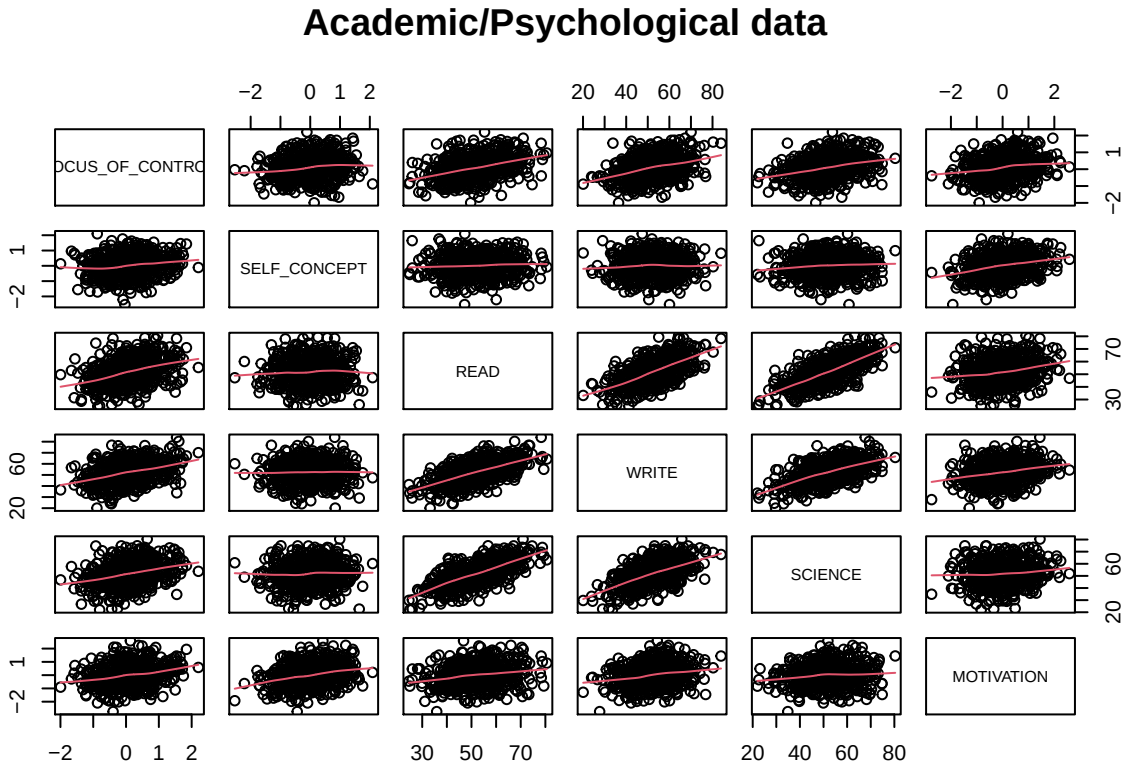
Part B.) Intepret the results from (a) above. [10 points]

After running a fad() on the zip data only for the digit 5, I determined that 10 factors is optimal due to the analysis producing the lowest "BIC" value as shown in the plot. The BIC is -19864.71.

Next I used images to display the results of the anlaysis i.e. the factor loadings (first image) and the uniqueness values (second image).The factor loadings are in a matrix that is 256 row by 10 columns. TO create this image, I took each column and made it into a 16 by 16 matrix. The image resulted in 10 very blurry images of the number 5. Also, I think the images get a little harder to distinguish '5'. However, these images are not as blurry as the image created with the uniqueness values.

Question 3: We go back to the psychological dataset of HW5.

Part A.) Perform a canonical correlation analysis between the academic variables and the psychological response variables. Comment. [10 points]




```
## $cor
## [1] 0.4432105 0.1533331 0.0188313
##
## $xcoef
##                   [,1]          [,2]         [,3]
## READ    -2.046639e-03 -0.0001046692  0.005735423
## WRITE   -2.482373e-03 -0.0037343450 -0.003293940
## SCIENCE -5.697164e-05  0.0050432025 -0.003243506
##
## $ycoef
##                            [,1]        [,2]        [,3]
## LOCUS_OF_CONTROL -0.051173272  0.03122170 -0.02016879
## SELF_CONCEPT      0.009807288  0.03449017  0.04915559
## MOTIVATION       -0.021168635 -0.04487452  0.01875833
##
## $xcenter
##     READ    WRITE  SCIENCE
## 51.90183 52.38483 51.76333
##
## $ycenter
## LOCUS_OF_CONTROL     SELF_CONCEPT       MOTIVATION
##      0.096533337      0.004916681      0.003897900
```

The first canoncial correlation between the academic(READ,WRITE,SCIENCE) and the psychological(MOTIVATION, SELF CONCEPT, LOCUS_OF_CONTROL) variables is 0.4432105. The second canoncial correlation between the variable sets is 0.1533331 and the 3rd canonical correlation is 0.0188313. I would say that the two sets of variables do not have a strong correlation because the first correlation value is not that close to 1 or -1. However the variable sets are postively correlated.

Part B.) Perform individual canonical correlation analyses between the academic and the psychological response sets of variables for each of the three program groups. Comment. [15 points]

```
## $cor
## [1] 0.43000305 0.21720225 0.07193644
##
## $xcoef
##                          [,1]        [,2]        [,3]
## LOCUS_OF_CONTROL -0.09408129 -0.07949994 -0.03124793
## SELF_CONCEPT      0.05797444 -0.08647038  0.05569847
## MOTIVATION       -0.05218647  0.05137459  0.08114058
##
## $ycoef
##                 [,1]         [,2]         [,3]
## READ    -0.001353831 -0.006488376  0.011856035
## WRITE   -0.009065860  0.008334064 -0.002118237
## SCIENCE  0.002057689 -0.007365294 -0.010963890
##
## $xcenter
## LOCUS_OF_CONTROL     SELF_CONCEPT       MOTIVATION
##      -0.09852908      -0.25822566      -0.39539753
##
## $ycenter
##     READ    WRITE  SCIENCE
## 49.44327 50.82720 50.45020

## $cor
## [1] 0.4918117 0.2135887 0.0143087
##
## $xcoef
##                  [,1]         [,2]         [,3]
## READ    -0.0031678566 -0.003457637  0.007826057
## WRITE   -0.0030157490 -0.003522361 -0.006528835
## SCIENCE -0.0004861281  0.008377870 -0.002123223
##
## $ycoef
##                           [,1]        [,2]        [,3]
## LOCUS_OF_CONTROL -0.073640129  0.04807287 -0.03415424
## SELF_CONCEPT      0.002310551  0.03550069  0.08734306
## MOTIVATION       -0.033526762 -0.07479059  0.01220857
##
## $xcenter
##     READ    WRITE  SCIENCE
## 52.42997 52.05612 51.87864
##
## $ycenter
```

```
## LOCUS_OF_CONTROL     SELF_CONCEPT       MOTIVATION
##      0.089768777      0.024466135      0.002515469

## $cor
## [1] 0.36199732 0.10202825 0.02834979
##
## $xcoef
##                   [,1]         [,2]         [,3]
## READ    -0.0047279133  0.008607613  0.003527045
## WRITE   -0.0040990823 -0.008472264  0.003622463
## SCIENCE  0.0003119441 -0.001945928 -0.010572288
##
## $ycoef
##                           [,1]        [,2]        [,3]
## LOCUS_OF_CONTROL -0.10298836  0.03924442 -0.04403748
## SELF_CONCEPT      0.03952920 -0.02598603 -0.10139096
## MOTIVATION       -0.02731847 -0.08682507  0.03389487
##
## $xcenter
##     READ    WRITE  SCIENCE
## 52.92883 53.97664 52.54848
##
## $ycenter
## LOCUS_OF_CONTROL     SELF_CONCEPT       MOTIVATION
##        0.2470664        0.1673028        0.2943555
```

For the students in PROG1, the first canoncial correlation between the academic(READ,WRITE,SCIENCE) and the psychological(MOTIVATION, SELF CONCEPT, LOCUS_OF_CONTROL) variables is 0.43000305. The second canoncial correlation between the variable sets is 0.21720225 and the 3rd canonical correlation is 0.07193644. I would say that the two sets of variables do not have a strong correlation because all of the canonical correlations are not that close to 1 or -1. However the variable sets do have a slight positive correlation.

For the students in PROG2, the first canoncial correlation between the academic(READ,WRITE,SCIENCE) and the psychological(MOTIVATION, SELF CONCEPT, LOCUS_OF_CONTROL) variables is 0.4918117. The second canoncial correlation between the variable sets is 0.2135887 and the 3rd canonical correlation is 0.0143087. I would say that the two sets of variables do not have a strong correlation because all of the canonical correlations are not that close to 1 or -1. The variable sets do have a slight positive correlation.are postively correlated.

For the students in PROG2, the first canoncial correlation between the academic(READ,WRITE,SCIENCE) and the psychological(MOTIVATION, SELF CONCEPT, LOCUS_OF_CONTROL) variables is 0.36199732. The second canoncial correlation between the variable sets is 0.10202825 and the 3rd canonical correlation is 0.02834979. I would say that the two sets of variables do not have a strong correlation because all of the canonical correlations are not that close to 1 or -1. The variable sets do have a slight positive correlation.are postively correlated.

We can see that between the 3 programs, Prog 2 is showing the strongest correlation between the academic and psychological variables with a canoncial correlation values: 0.4918117 0.2135887 0.0143087. PROG 3 seemed to have the weakest correlation between acadekic and psychological variables. However there isnn't that much difference btween the 3 programs.