PCA analysis

Jieqi Tu, jt3098

11/16/2019

```
# Import datasets
ABC_total_raw = readxl::read_excel("./ABC_Cord Blood_Metabolomics_new.xlsx") %>% as.data.frame()

# Tidy dataset
ABC_total =
  ABC_total_raw %>%
  janitor::clean_names()

# Divide the dataset to confounding information and experimental data
ABC_information = ABC_total[1:10]
ABC_data = ABC_total[11:930]

# Make a copy of ABC_data
ABC_data_orginal = ABC_data
```

```
# Convert 0 to NA and add 1 to columns that contains 0
ABC_data[ABC_data == 0] = NA
NA_value = sapply(ABC_data[1:920], function(x) sum(which(is.na(x))))

for(i in 1:920) {
  if(NA_value[i]>0) {
    ABC_data_orginal[i] = ABC_data_orginal[i] + 1
  }
}


# Base 10 log transformation of original data
ABC_data = sapply(ABC_data_orginal[1:920], function(x) log10(x))
ABC_logtrans = cbind.data.frame(ABC_information, ABC_data)

# Extract the subset of all controls
subset_control =
  ABC_logtrans %>%
  group_by(strata) %>%
  filter(asd == 0)

# Calculate the mean and standard deviation of the control group
mean_control = sapply(subset_control[11:930], function(x) mean(x))
sd_control = sapply(subset_control[11:930], function(x) sd(x))

# Write a for-loop to conduct scaling of data for both control and ASD groups
for(i in 11:930) {
  ABC_logtrans[i] = (ABC_logtrans[i] - mean_control[i-10])/sd_control[i-10]
}
```

```
# Principal component analysis
ABC_pca = prcomp(ABC_logtrans[c(11:930)], center = F, scale. = F)
summary(ABC_pca)
```

```
## Importance of components:
##                            PC1     PC2     PC3     PC4    PC5     PC6
## Standard deviation     9.81118 9.33956 7.58852 7.11932 6.2273 5.73083
## Proportion of Variance 0.09655 0.08749 0.05776 0.05084 0.0389 0.03294
## Cumulative Proportion  0.09655 0.18404 0.24180 0.29263 0.3315 0.36447
##                            PC7    PC8     PC9    PC10    PC11    PC12
## Standard deviation     5.56231 5.0915 5.04799 4.70662 4.55151 4.40485
## Proportion of Variance 0.03103 0.0260 0.02556 0.02222 0.02078 0.01946
## Cumulative Proportion  0.39550 0.4215 0.44706 0.46928 0.49006 0.50952
##                           PC13    PC14    PC15    PC16    PC17   PC18
## Standard deviation     4.39485 4.09946 3.97601 3.77288 3.66647 3.6271
## Proportion of Variance 0.01937 0.01686 0.01586 0.01428 0.01348 0.0132
## Cumulative Proportion  0.52890 0.54575 0.56161 0.57588 0.58937 0.6026
##                           PC19    PC20    PC21    PC22    PC23    PC24
## Standard deviation     3.43106 3.32388 3.17586 3.12792 3.08699 2.98808
## Proportion of Variance 0.01181 0.01108 0.01012 0.00981 0.00956 0.00896
## Cumulative Proportion  0.61437 0.62545 0.63557 0.64538 0.65494 0.66390
##                           PC25    PC26   PC27    PC28    PC29    PC30
## Standard deviation     2.93068 2.84323 2.8245 2.77595 2.70204 2.63093
## Proportion of Variance 0.00861 0.00811 0.0080 0.00773 0.00732 0.00694
## Cumulative Proportion  0.67251 0.68062 0.6886 0.69635 0.70367 0.71062
##                           PC31    PC32    PC33   PC34    PC35    PC36
## Standard deviation     2.56323 2.50318 2.49690 2.4670 2.41403 2.35847
## Proportion of Variance 0.00659 0.00628 0.00625 0.0061 0.00585 0.00558
## Cumulative Proportion  0.71721 0.72349 0.72974 0.7359 0.74169 0.74727
##                           PC37    PC38    PC39    PC40    PC41    PC42
## Standard deviation     2.34659 2.28553 2.27168 2.21923 2.17348 2.12082
## Proportion of Variance 0.00552 0.00524 0.00518 0.00494 0.00474 0.00451
## Cumulative Proportion  0.75280 0.75804 0.76321 0.76815 0.77289 0.77740
##                           PC43    PC44    PC45    PC46    PC47    PC48
## Standard deviation     2.10431 2.05848 2.03009 2.01644 2.00243 1.95995
## Proportion of Variance 0.00444 0.00425 0.00413 0.00408 0.00402 0.00385
## Cumulative Proportion  0.78184 0.78609 0.79023 0.79430 0.79833 0.80218
##                           PC49    PC50    PC51    PC52    PC53    PC54
## Standard deviation     1.93334 1.89771 1.87407 1.85956 1.84949 1.84393
## Proportion of Variance 0.00375 0.00361 0.00352 0.00347 0.00343 0.00341
## Cumulative Proportion  0.80593 0.80954 0.81306 0.81653 0.81996 0.82337
##                           PC55    PC56    PC57    PC58    PC59    PC60
## Standard deviation     1.79992 1.78999 1.76172 1.74823 1.73631 1.71744
## Proportion of Variance 0.00325 0.00321 0.00311 0.00307 0.00302 0.00296
## Cumulative Proportion  0.82662 0.82984 0.83295 0.83601 0.83904 0.84200
##                           PC61    PC62    PC63    PC64    PC65    PC66
## Standard deviation     1.69538 1.69040 1.66717 1.66388 1.64684 1.63003
## Proportion of Variance 0.00288 0.00287 0.00279 0.00278 0.00272 0.00266
## Cumulative Proportion  0.84488 0.84775 0.85053 0.85331 0.85603 0.85869
##                           PC67    PC68    PC69    PC70    PC71    PC72
## Standard deviation     1.61982 1.61234 1.60146 1.59625 1.58969 1.55197
## Proportion of Variance 0.00263 0.00261 0.00257 0.00256 0.00253 0.00242
## Cumulative Proportion  0.86133 0.86393 0.86651 0.86906 0.87160 0.87401
##                           PC73    PC74    PC75    PC76   PC77    PC78
## Standard deviation     1.52909 1.51701 1.49565 1.48909 1.4813 1.47768
## Proportion of Variance 0.00235 0.00231 0.00224 0.00222 0.0022 0.00219
```

```
## Cumulative Proportion  0.87636 0.87867 0.88091 0.88313 0.8853 0.88752
##                           PC79    PC80    PC81   PC82    PC83    PC84
## Standard deviation     1.46229 1.43464 1.41523 1.4106 1.39964 1.39192
## Proportion of Variance 0.00214 0.00206 0.00201 0.0020 0.00196 0.00194
## Cumulative Proportion  0.88967 0.89173 0.89374 0.8957 0.89770 0.89965
##                           PC85    PC86    PC87    PC88   PC89    PC90
## Standard deviation     1.38667 1.37193 1.36849 1.35533 1.3400 1.32431
## Proportion of Variance 0.00193 0.00189 0.00188 0.00184 0.0018 0.00176
## Cumulative Proportion  0.90158 0.90346 0.90534 0.90718 0.9090 0.91074
##                           PC91    PC92    PC93    PC94    PC95    PC96
## Standard deviation     1.31085 1.30487 1.28815 1.28034 1.27943 1.27157
## Proportion of Variance 0.00172 0.00171 0.00166 0.00164 0.00164 0.00162
## Cumulative Proportion  0.91247 0.91418 0.91584 0.91748 0.91913 0.92075
##                           PC97    PC98    PC99   PC100  PC101   PC102
## Standard deviation     1.25797 1.24418 1.23699 1.22846 1.2236 1.21657
## Proportion of Variance 0.00159 0.00155 0.00153 0.00151 0.0015 0.00148
## Cumulative Proportion  0.92233 0.92389 0.92542 0.92694 0.9284 0.92992
##                          PC103   PC104   PC105   PC106   PC107   PC108
## Standard deviation     1.20154 1.18916 1.18413 1.17204 1.16240 1.15792
## Proportion of Variance 0.00145 0.00142 0.00141 0.00138 0.00136 0.00134
## Cumulative Proportion  0.93137 0.93279 0.93419 0.93557 0.93693 0.93827
##                          PC109   PC110   PC111   PC112   PC113  PC114
## Standard deviation     1.14994 1.14442 1.12339 1.11515 1.10080 1.0956
## Proportion of Variance 0.00133 0.00131 0.00127 0.00125 0.00122 0.0012
## Cumulative Proportion  0.93960 0.94091 0.94218 0.94343 0.94464 0.9458
##                          PC115   PC116   PC117   PC118   PC119   PC120
## Standard deviation     1.08729 1.08382 1.07936 1.06785 1.05956 1.04988
## Proportion of Variance 0.00119 0.00118 0.00117 0.00114 0.00113 0.00111
## Cumulative Proportion  0.94703 0.94821 0.94938 0.95052 0.95165 0.95275
##                          PC121   PC122   PC123   PC124   PC125   PC126
## Standard deviation     1.04221 1.03568 1.03452 1.01826 1.01342 1.00316
## Proportion of Variance 0.00109 0.00108 0.00107 0.00104 0.00103 0.00101
## Cumulative Proportion  0.95384 0.95492 0.95599 0.95703 0.95806 0.95907
##                          PC127   PC128   PC129   PC130   PC131   PC132
## Standard deviation     0.9981 0.98731 0.98433 0.97566 0.96670 0.96351
## Proportion of Variance 0.0010 0.00098 0.00097 0.00095 0.00094 0.00093
## Cumulative Proportion  0.9601 0.96105 0.96202 0.96297 0.96391 0.96484
##                          PC133   PC134   PC135   PC136   PC137   PC138
## Standard deviation     0.95455 0.94391 0.93408 0.92749 0.91314 0.91145
## Proportion of Variance 0.00091 0.00089 0.00088 0.00086 0.00084 0.00083
## Cumulative Proportion  0.96576 0.96665 0.96753 0.96839 0.96923 0.97006
##                          PC139   PC140  PC141   PC142   PC143   PC144
## Standard deviation     0.90726 0.89934 0.8931 0.88905 0.87957 0.87326
## Proportion of Variance 0.00083 0.00081 0.0008 0.00079 0.00078 0.00076
## Cumulative Proportion  0.97088 0.97170 0.9725 0.97329 0.97406 0.97483
##                          PC145   PC146   PC147   PC148  PC149   PC150
## Standard deviation     0.86696 0.86062 0.85557 0.85004 0.8369 0.82895
## Proportion of Variance 0.00075 0.00074 0.00073 0.00072 0.0007 0.00069
## Cumulative Proportion  0.97558 0.97633 0.97706 0.97778 0.9785 0.97918
##                          PC151   PC152   PC153   PC154   PC155   PC156
## Standard deviation     0.82558 0.81901 0.81675 0.81082 0.79796 0.78939
## Proportion of Variance 0.00068 0.00067 0.00067 0.00066 0.00064 0.00063
## Cumulative Proportion  0.97986 0.98053 0.98120 0.98186 0.98250 0.98312
##                          PC157   PC158  PC159   PC160   PC161   PC162
```

```
## Standard deviation     0.78315 0.77686 0.7709 0.76243 0.75296 0.74469
## Proportion of Variance 0.00062 0.00061 0.0006 0.00058 0.00057 0.00056
## Cumulative Proportion  0.98374 0.98435 0.9849 0.98552 0.98609 0.98665
##                           PC163   PC164   PC165   PC166   PC167   PC168
## Standard deviation     0.73969 0.73651 0.72873 0.71649 0.71351 0.71154
## Proportion of Variance 0.00055 0.00054 0.00053 0.00051 0.00051 0.00051
## Cumulative Proportion  0.98720 0.98774 0.98827 0.98879 0.98930 0.98981
##                           PC169   PC170   PC171   PC172   PC173   PC174
## Standard deviation     0.70164 0.69419 0.69055 0.68756 0.68055 0.66548
## Proportion of Variance 0.00049 0.00048 0.00048 0.00047 0.00046 0.00044
## Cumulative Proportion  0.99030 0.99079 0.99126 0.99174 0.99220 0.99265
##                           PC175   PC176   PC177   PC178   PC179   PC180
## Standard deviation     0.66236 0.65599 0.65058 0.64046 0.63954 0.62281
## Proportion of Variance 0.00044 0.00043 0.00042 0.00041 0.00041 0.00039
## Cumulative Proportion  0.99309 0.99352 0.99394 0.99435 0.99476 0.99515
##                           PC181   PC182   PC183   PC184   PC185   PC186
## Standard deviation     0.61904 0.61313 0.60746 0.60554 0.59554 0.58700
## Proportion of Variance 0.00038 0.00038 0.00037 0.00037 0.00036 0.00035
## Cumulative Proportion  0.99554 0.99591 0.99629 0.99665 0.99701 0.99735
##                           PC187   PC188   PC189   PC190  PC191   PC192
## Standard deviation     0.57247 0.56705 0.55747 0.55499 0.5469 0.53480
## Proportion of Variance 0.00033 0.00032 0.00031 0.00031 0.0003 0.00029
## Cumulative Proportion  0.99768 0.99801 0.99832 0.99863 0.9989 0.99921
##                           PC193   PC194   PC195     PC196
## Standard deviation     0.52393 0.51646 0.49349 2.441e-14
## Proportion of Variance 0.00028 0.00027 0.00024 0.000e+00
## Cumulative Proportion  0.99949 0.99976 1.00000 1.000e+00
```

```r
# Plot PC1 and PC2
plot(ABC_pca$x[,1],ABC_pca$x[,2], xlab="PC1 (9.6%)", ylab = "PC2 (8.7%)", main = "PC1 / PC2 - plot")
```