3/2/2010 36-402/608 ADA-II H. Seltman

Breakout #14 Results

Summary: To perform a principal components analysis in R, use prcomp() which produces a "prcomp" object with components "sdev" (sd of each principal component), "rotation" (loading coefficients), and "x" (scores / new variables). The most useful methods are print(), summary(), and plot(). Use plot(,type="l") for the traditional scree plot.

Don't use the old princomp() function.

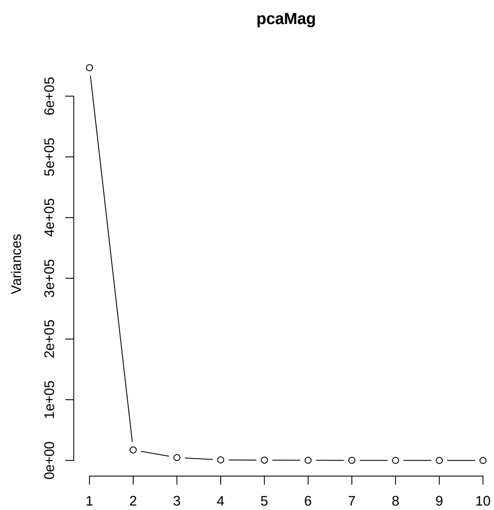
Sleuth example: The magnetic field on a printer component is tested at 11 locations along the component. A factorial experiment is performed at 3 currents, with two versions of the component, and with 4 materials. Before analysis of the effects of the three explanatory variables on the 11 outcomes, it is desirable to reduce the number of outcomes.

```
# Check correlation of outcomes
round(cor(mag[,1:11]),2)
#       L1   L2   L3   L4   L5   L6   L7   L8   L9  L10  L11
# L1  1.00 1.00 0.99 0.99 0.99 0.95 0.97 0.95 0.92 0.92 0.92
# L2  1.00 1.00 1.00 1.00 0.99 0.95 0.98 0.96 0.94 0.93 0.93
# L3  0.99 1.00 1.00 1.00 0.99 0.95 0.98 0.96 0.94 0.93 0.93
# L4  0.99 1.00 1.00 1.00 1.00 0.95 0.98 0.96 0.94 0.94 0.94
# L5  0.99 0.99 0.99 1.00 1.00 0.97 0.99 0.96 0.94 0.93 0.94
# L6  0.95 0.95 0.95 0.95 0.97 1.00 0.97 0.93 0.90 0.90 0.91
# L7  0.97 0.98 0.98 0.98 0.99 0.97 1.00 0.99 0.97 0.97 0.98
# L8  0.95 0.96 0.96 0.96 0.96 0.93 0.99 1.00 1.00 0.99 0.99
# L9  0.92 0.94 0.94 0.94 0.94 0.90 0.97 1.00 1.00 1.00 0.99
# L10 0.92 0.93 0.93 0.94 0.93 0.90 0.97 0.99 1.00 1.00 0.99
# L11 0.92 0.93 0.93 0.94 0.94 0.91 0.98 0.99 0.99 0.99 1.00

# Perform PCA
pcaMag = prcomp(mag[,1:11])
names(pcaMag)
# [1] "sdev"     "rotation" "center"   "scale"    "x"

# Examine relative variances
round( 100 * pcaMag$sdev^2 / sum(pcaMag$sdev^2), 2)
# [1] 96.46  2.57  0.70  0.14  0.07  0.04  0.01  0.01  0.00  0.00  0.00
```

```
summary(pcaMag)
# Importance of components:
#                            PC1      PC2      PC3      PC4      PC5      PC6
# Standard deviation      804.331 131.3274 68.43947 30.12473 21.21056 15.55544
# Proportion of Variance    0.965   0.0257  0.00698  0.00135  0.00067  0.00036
# Cumulative Proportion     0.965   0.9903  0.99730  0.99866  0.99933  0.99969
#                            PC7      PC8      PC9     PC10     PC11
# Standard deviation     9.64689  7.88855   5.20657  3.94252  3.48679
# Proportion of Variance 0.00014  0.00009   0.00004  0.00002  0.00002
# Cumulative Proportion  0.99983  0.99992   0.99996  0.99998  1.00000
#
# plot(pcaMag, type="l")
```



Question 1: What is the correlation pattern of the magnetic field at the different positions along the component? What does the code with `sdev^2` do? What do you learn from the scree plot?

```
# Examine the loadings
round(pcaMag$rotation,2)
#      PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10  PC11
# L1  0.22 -0.30  0.26 -0.09  0.73 -0.29 -0.22  0.16 -0.28  0.12 -0.04
# L2  0.23 -0.27  0.27 -0.05  0.20  0.33  0.43  0.06  0.62  0.07  0.27
# L3  0.24 -0.26  0.29 -0.01 -0.33  0.41  0.27  0.17 -0.62  0.13 -0.07
# L4  0.25 -0.26  0.29  0.02 -0.14 -0.08 -0.10 -0.43  0.16 -0.50 -0.54
# L5  0.26 -0.31  0.07  0.13 -0.37 -0.07 -0.60 -0.15  0.10  0.23  0.48
# L6  0.29 -0.39 -0.79 -0.21  0.10  0.09  0.13 -0.19 -0.11 -0.10  0.03
# L7  0.31 -0.08 -0.20  0.26 -0.22 -0.26  0.00  0.64  0.26  0.17 -0.41
# L8  0.34  0.18  0.08  0.09 -0.13 -0.57  0.41  0.01 -0.16 -0.36  0.43
# L9  0.38  0.37  0.05 -0.43 -0.08 -0.16  0.11 -0.34  0.05  0.57 -0.21
# L10 0.38  0.40  0.00 -0.39  0.04  0.34 -0.36  0.33  0.03 -0.41  0.09
# L11 0.36  0.34 -0.10  0.72  0.28  0.29 -0.03 -0.24 -0.08  0.07 -0.03
```

Question 2: How would you simplify the first 3 principal components into something more interpretable (using words and/or numbers?

```
# For educational purposes examine some properties
mag[1:2,1:11]
#    L1  L2  L3  L4  L5  L6  L7  L8   L9  L10  L11
# 1 136 142 139 131 122 118 134 138  148  149  171
# 2 639 723 782 756 804 804 909 962 1042 1058 1022
round(pcaMag$x[1:2,],1)
#         PC1  PC2   PC3  PC4   PC5  PC6  PC7  PC8  PC9 PC10 PC11
#[1,] -1428.0 65.2 -20.1  2.3   0.9  7.4 -1.7 -0.7 -2.2  1.3 -1.1
#[2,]  1011.2 23.0  13.5 18.3 -77.3 25.3  5.1 16.4 -5.5  0.2  7.2
round( scale(as.matrix(mag[,1:11]),scale=FALSE)[1:2,] %*% pcaMag$rotation[,], 1)
#          PC1  PC2   PC3  PC4   PC5  PC6  PC7  PC8  PC9 PC10 PC11
# [1,] -1428.0 65.2 -20.1  2.3   0.9  7.4 -1.7 -0.7 -2.2  1.3 -1.1
# [2,]  1011.2 23.0  13.5 18.3 -77.3 25.3  5.1 16.4 -5.5  0.2  7.2
#
sum(diag(cov(mag[,1:11]))) # [1] 670688.7
sum(diag(cov(pcaMag$x))) # [1] 670688.7
round(cor(pcaMag$x),2)[1:3,1:3]
    PC1 PC2 PC3
PC1   1   0   0
PC2   0   1   0
PC3   0   0   1
```

Question 3: Knowing that `scale(,scale=F)` centers the columns of a data.frame, what does the matrix multiplication do / tell us? What can you deduce about PCA from the rest of the code?

Here are some analyses using the PCA results and the original explanatory variables, current, configuration, and material:

```
anova(aov(as.matrix(mag[,1:11])~current+configur+material,mag))
#             Df  Pillai approx F num Df den Df    Pr(>F)
# (Intercept)  1 0.93546   35.578     11     27 3.432e-13 ***
# current      2 0.67314    1.291     22     56    0.2183
# configur     1 0.17424    0.518     11     27    0.8742
# material     3 1.02835    1.375     33     87    0.1221
# Residuals   37

anova(aov(pcaMag$x~current+configur+material,mag))
#             Df  Pillai approx F num Df den Df Pr(>F)
# (Intercept)  1 0.00000  0.00000     11     27 1.0000
# current      2 0.67314  1.29134     22     56 0.2183
# configur     1 0.17424  0.51793     11     27 0.8742
# material     3 1.02835  1.37504     33     87 0.1221
# Residuals   37

# summary(aov(pcaMag$x[,1]~current+configur+material,mag))
#             Df   Sum Sq Mean Sq F value Pr(>F)
# current      2   718795  359397  0.5107 0.6042
# configur     1   688980  688980  0.9791 0.3289
# material     3   373626  124542  0.1770 0.9113
# Residuals   37 26037380  703713
#
# summary(aov(pcaMag$x[,2]~current+configur+material,mag))
#             Df Sum Sq Mean Sq F value   Pr(>F)
# current      2 209183  104592  7.7885 0.001503 **
# configur     1   3212    3212  0.2392 0.627703
# material     3  32350   10783  0.8030 0.500219
# Residuals   37 496871   13429
#
```

```
# summary(aov(pcaMag$x[,3]~current+configur+material,mag))
#             Df Sum Sq Mean Sq F value Pr(>F)
# current      2    442   221.1  0.0444 0.9566
# configur     1   2347  2347.4  0.4712 0.4967
# material     3  14299  4766.5  0.9568 0.4233
# Residuals   37 184321  4981.7
# TukeyHSD(aov(pcaMag$x[,2]~current,mag))
#     95% family-wise confidence level
#                   diff        lwr      upr     p adj
# 250ma-0ma   144.449935   43.04051 245.8594 0.0035375
# 500ma-0ma   142.194723   40.78529 243.6042 0.0041169
# 500ma-250ma  -2.255213 -106.99042 102.4800 0.9984897
```

Question 4: What do you conclude from the ANOVAs?