Solutions for Exam 2013, Problem 1

ANYM, 20200918

1.1 - 3

We call the estimated eigenvalues of the correlation matrix $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_7$ .The arithmetic and the geometric average of the smallest 6 eigenvalues are respectively

$$\frac{1}{6}\sum_{i=2}^{7} \hat{\lambda}_i = 0.22461$$

$$\{\prod_{i=2}^{7} \hat{\lambda}_i\}^{\frac{1}{6}} = 0.10749$$

The usual test statistic for testing the hypothesis that the smallest 6 eigenvalues are equal against all alternatives is

We use theorem 6.8

$$Z_2 = -n \log \frac{\det \hat{\mathbf{R}}}{\hat{\lambda}_1 \cdot \ldots \cdot \hat{\lambda}_m \cdot \hat{\lambda}^{k-m}} = -n \log \frac{\hat{\lambda}_{m+1} \cdot \ldots \cdot \hat{\lambda}_k}{\hat{\lambda}_*^{k-m}},$$

$$\hat{\lambda}_* = (k - \hat{\lambda}_1 - \ldots - \hat{\lambda}_m)/(k-m) = (\hat{\lambda}_{m+1} + \ldots + \hat{\lambda}_k)/(k-m)$$

We have *n=109* observations

$$Z_2 = -109 \log \frac{\left((\prod_{i=2}^{7} \hat{\lambda}_i)^{\frac{1}{6}}\right)^6}{\left({}^1/_6 \sum_{i=2}^{7} \hat{\lambda}_i\right)^6} = -109 \times 6 \times \left\{\log\left(\prod_{i=2}^{7} \hat{\lambda}_i\right)^{\frac{1}{6}} - \log {}^1/_6 \sum_{i=2}^{7} \hat{\lambda}_i\right\}$$
$$= -109 \times 6 \times \{\log 0.10749 - \log 0.22461\}$$

1.2 - 2

We use teorem 6.8.

$$\{x_1, \ldots, x_n | z_2 > \chi^2(\frac{1}{2}(k-m+2)(k-m-1))_{1-\alpha}\}.$$

We have 7 eigenvalues, *k = 7*

We test the 6 smallest, *m = 1*

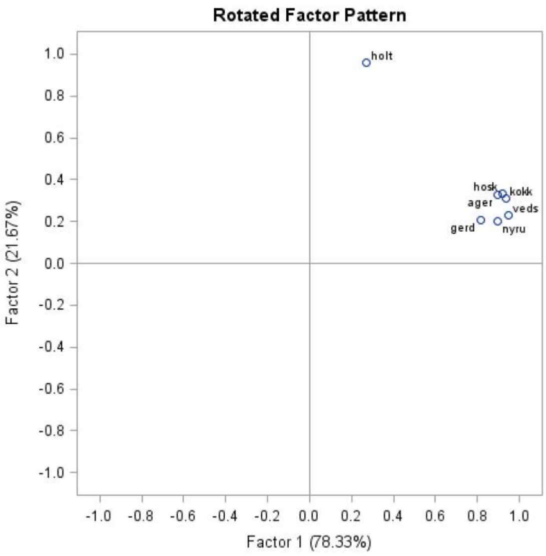
k-m=6

This yields:

$$\chi\left(\frac{1}{2}(k-m+2)(k-m-1)\right) = \chi\left(\frac{1}{2} \cdot 8 \cdot 5\right) = \chi(20)$$

1.3 - 5

We can use the component pattern plot



Or the coefficients directly

Rotated Factor Pattern		
	Factor1	Factor2
ager	0.89385	0.32474
holt	0.27020	0.96125
gerd	0.81424	0.20955
hosk	0.91898	0.33249
kokk	0.93737	0.30763
veds	0.95061	0.23021
nyru	0.89585	0.20365

We see from the figure, that *holt* is the station closest to the powerplant.

Factor 1 is then option D and factor 2 is option E

2.4 - 2

As we use the correlation matrix, all original variables have variance 1. The total variance is thus 7.

We have

Variance Explained by Each Factor	
Factor1	Factor2
4.9643460	1.3730299

Then (4.9643460+1.3730299)/7=0.905339

2.5 - 4

We find the loadings in the SAS output

Rotated Factor Pattern		
	Factor1	Factor2
ager	0.89385	0.32474
holt	0.27020	0.96125
gerd	0.81424	0.20955
hosk	0.91898	0.33249
kokk	0.93737	0.30763
veds	0.95061	0.23021
nyru	0.89585	0.20365

We have from page 405 that

Finally the (i, j)'th factor weight gives the correlation between the i'th variable and the j'th factor i.e.

$$\mathrm{Cov}(X_i, F_j) = \mathrm{Cov}(\sum_{\nu} a_{i\nu} F_\nu + G_i, F_j) = a_{ij}.$$

As we always use standardized data (i.e. correlation matrix) the covariance and correlation are equal.

Further we have from page 29

and the *squared coefficient of correlation represents the reduction in variance. i.e. the fraction of Y's variance, which can be explained by X,* since

$$\rho^2 = \frac{\mathrm{V}(Y) - \mathrm{V}(Y|X = x)}{\mathrm{V}(Y)}.$$

We thus need to square the loading of rotated factor 1 for *holt*

0.27020^2=0.073