Homework 5

Cody Frisby

4/10/2017

5.1

Assumptions of factor model:

- Uncorrelated common factors: $cov(f_i, f_j) = 0$ for $i \neq j$.
- Specific factors are uncorrelated with each other: $cov(\psi_i, \psi_j) = 0$ for $i \neq j$.
- Independence of common factors and specific variances: $cov(\psi_i, f_j) = 0$ for $i \neq j$.

The factor model can be written as $X = \Lambda f + u$. With the above assumptions we can write the population variance/covariance matrix, Σ, as

$$\Sigma = V(X) = V(\Lambda f + u) = \Lambda V(f) \Lambda^T + V(u) = \Lambda \Lambda^T + \Psi$$

since $cov(f_i, f_j) = 0$, $cov(u_i, u_j) = 0$, and $cov(f_i, u_j) = 0$.

If we let the common factors be correlated then $cov(f_i, f_j) \neq 0$ and the assumptions of the factor model are no longer valid. The (i, j) element of Σ (σ_{ij}) would be

$$\sigma_{ij} = cov(x_i, x_j) = cov(\lambda_{i1} f_1, \lambda_{j1} f_1) + cov(\lambda_{i2} f_2, \lambda_{j2} f_2) + ... + cov(\lambda_{ik} f_k, \lambda_{jk} f_k)$$

which under the model assumptions is $0 + 0 + ... + 0$ but if they are correlated this is not the case ■

5.2

Re-writing $X = \Lambda f + u$ as

$$X = \Lambda M M^T f + u = \Lambda^* f^* + u$$

and letting M be any orthonormal matrix, the original model formulation becomes

$$\Sigma = \Lambda^* (\Lambda^*)^T + \Psi = \Lambda M (\Lambda M)^T + \Psi = \Lambda M M^T \Lambda^T + \Psi = \Lambda \Lambda^T + \Psi$$

which is the result of the original model ■

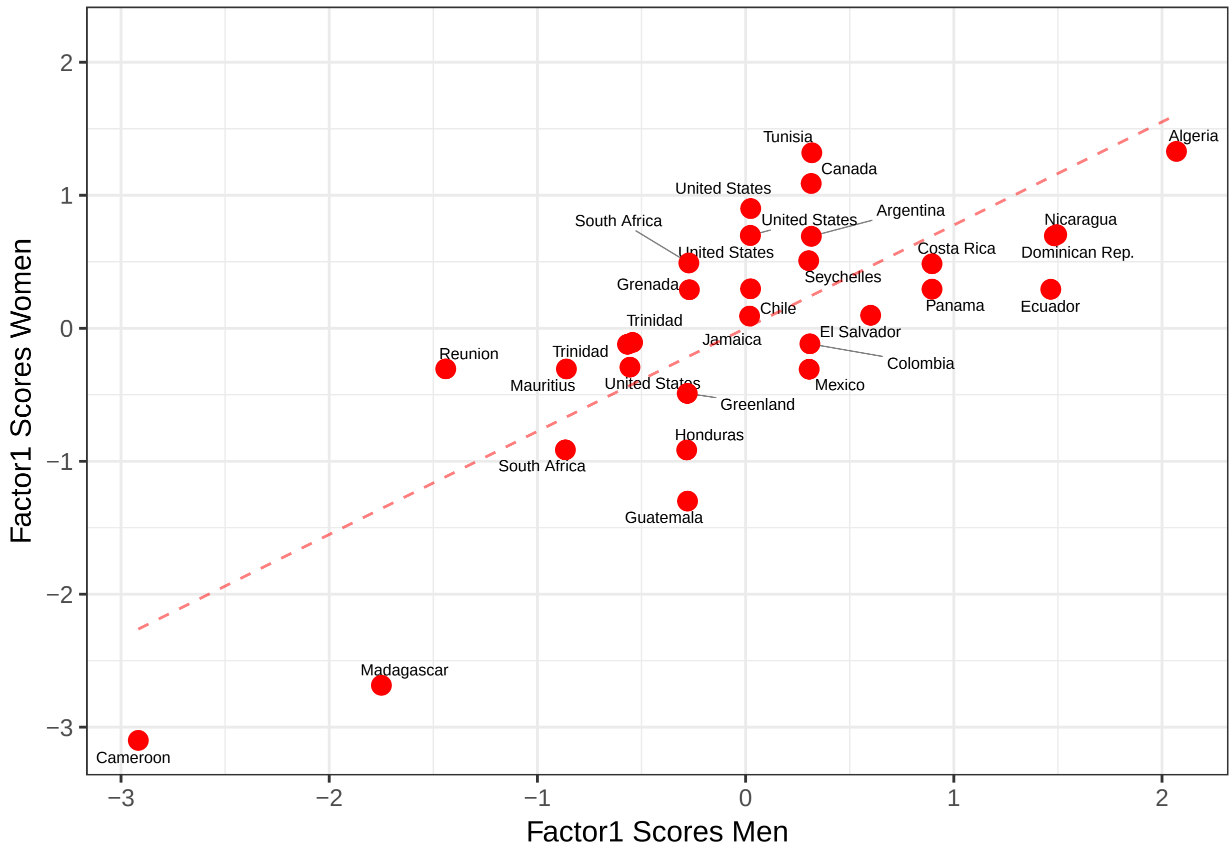
5.3

Let p_j equal the proportion of variance explained by the jth factor. Let e_j equal the jth eigen value from the correlation matrix of $\mathbf{X}$. Let n equal the number of eigen values.

$$p_j = \frac{e_j}{\sum_{i=1}^{n} e_i}, \ for \ j = 1, 2, ..., n$$

5.4

A one factor solution for both males and females is pictured here.



It looks like there is a linear relationship between the sexes' factor1 loading scores. The correlation between the two is 0.7755723. A fitted least squares line is displayed for reference. A bivariate box plot would identify **Cameroon**, **Madagascar**, and **Algeria** as potential outliers although there is no strong disagreement between the two factors as evidence by the linear relationship. Table 1 displays a comparison between the two factors for both men and women.

Table 1: Loadings

	Men	Women
0	0.6375392	0.8829280
25	0.6694452	0.9975803
50	0.9975012	0.9409332
75	0.7521583	0.6885076

We can see that *factor1* for women has large values for 0, 25, and 50; life expectancy in infancy, young adulthood, and even middle age. The largest value for men is for 50 and 75, which is life expectancy for middle aged and older.

5.5

A two-factor solution produces a p-value of 0.7026721. A large p-value indicated that a two-factor solution is sufficient. The *specific* or *unique* values are shown in table 2.

Table 2: Uniqueness

french	english	history	arithmetic	algebra	geometry
0.5083624	0.5945773	0.6439606	0.3772315	0.4402626	0.6283514

The derived loadings (without any rotation) are shown in table 3.

Table 3: Loadings

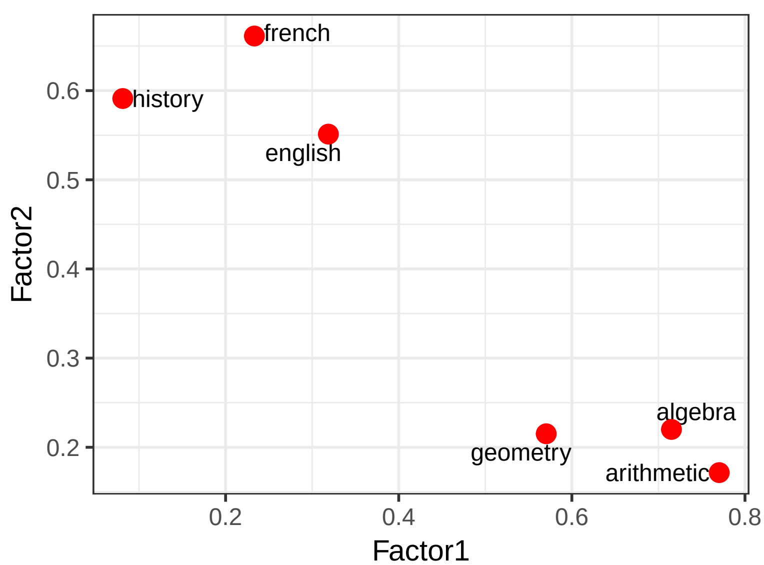
	Factor1	Factor2
french	0.5579783	0.4246165
english	0.5690101	0.2857435
history	0.3923861	0.4495243
arithmetic	0.7380265	-0.2794355
algebra	0.7184689	-0.2086671
geometry	0.5948773	-0.1333048

We could **rotate** this matrix by multiplying the loadings by an orthogonal matrix. The matrix in table 4 is used.

Table 4: Rotation matrix

1	2
0.8358396	0.5489738
-0.5489738	0.8358396

The derived rotation is pictured here.



There is some clear separation between the math related subjects and the others using an orthogonal rotation.

The rotated loadings that are plotted above are shown in table 5.

Table 5: Rotated Loadings

	Factor1	Factor2
french	0.2332771	0.6612268
english	0.3187354	0.5512073
history	0.0811948	0.5911398
arithmetic	0.7702745	0.1715939
algebra	0.7150775	0.2200083
geometry	0.5704028	0.2151506

Looking at the orthogonally rotated loadings it becomes clear that the math related subjects load heavily on *Factor1* while the others load heavily on *Factor2*.

5.6

Using `factanal` function in `R` and trying a few values of k ranging from 2 to 6 I found that the number of factors should be 3, $p = 0.0837743$.

The rotated loadings are shown in table 6. With so many factors it is going to take some heroism on my part to interpret them.

Table 6: Rotated Loadings

	Factor1	Factor2	Factor3
doc skill	0.6493130	-0.3720599	0.1903590
blame me	-0.1262411	0.1938827	0.6547915
what can docs do	0.7936806	-0.1441564	0.1164438
medical	0.7250306	-0.1057503	-0.0916411
exercise and diet	0.0152356	0.2919114	0.6447034
carelessness	-0.0833989	0.8245873	0.3766232
me	-0.2253364	0.5901192	0.3246421
doc control	0.8154517	0.0618640	-0.3040250
luck	0.4371164	-0.0756993	-0.2207399

Starting with *factor1* it appears this factor loads heavily on doctor involvement in the persons' pain, i.e. there are large values for "doc skill", "what can docs do", "medical" (seeking medical help), and "doc control". *Factor2* could be thought of as the "individual responsibility" factor with heavy loadings on "carelessness", their own personal responsibility, "me", and "exercise and diet" having the largest values. *Factor3* has a large score for for the "blame me" question. This factor could be thought of as the "my fault" factor.