PSY544 – Introduction to Factor Analysis

Homework assignment 2, Fall 2018

Due midnight, December 5, 2018

Data from a sample of N = 139 decathlon atletes were collected concerning their performance in each individual decathlon event (100m, Long Jump, Short Put, High Jump, 400m, 110m High Hurdles, Discus, Pole Vault, Javelin, 1500m). All performance variables were coded in such a way that the higher a person's score, the better his/her performance. A 10 x 10 correlation matrix was computed from the data:

	1CM	LJ	SP	HJ	4CM	HH	DS	PV	JV	15CM
1CM	1.00									
LJ	.59	1.00								
SP	.35	.42	1.00							
HJ	.34	.51	.38	1.00						
4CM	.63	.49	.19	.29	1.00					
HH	.40	.52	.36	.46	.34	1.00				
DS	.28	.31	.73	.27	.17	.32	1.00			
PV	.20	.36	.24	.39	.23	.33	.24	1.00		
JV	.11	.21	.44	.17	.13	.18	.34	.24	1.00	
15CM	-.07	.09	-.08	.18	.39	.00	-.02	.17	.00	1.00

Part 1

Obtain unrotated Maximum Likelihood (MWL) solutions for *m = 0, 2, 3* and *4* factors. Let CEFA save the iteration details and increase the maximum number of iterations to some higher than default number (like 300). You will encounter "parameters on the boundary" (likely Heywood cases) – ignore this problem for the rest of the homework and don't let it affect responses to the questions below.

a) For the 2-factor solution, explain (using your own words, but correctly and sensibly, so I understand) why the sample value of the discrepancy function $\hat{F}$ ("Disc Fun" in CEFA output) in the "iteration details" follows the pattern shown in the output, from one iteration to the next. Explain what is meant by "convergence" in this context.

b) For the *m = 2, 3* and *4* solutions, create a table showing the sample value of the discrepancy function $\hat{F}$, the likelihood-ratio test statistic (χ^2), the degrees of freedom (*df*), the effective number of parameters (*npar*), the estimated RMSEA and the associated 90% CI.

For the 3-factor solution…:

c) …comment on the meaning of each of the values you have put into the table for the previous question. Not what they imply for the model, but what do they represent.

d) …show the relationship between $\hat{F}$ and the likelihood-ratio test statistic.

e) …show the relationship between $\hat{F}$ and the population discrepancy function value, $\hat{F}_0$

f) …show the relationship between $\hat{F}_0$ and the point estimate of RMSEA

g) …compute the final communality for the second variable (Long Jump) using the factor loadings. What does this value represent?

h) Based on the results in your table and the residual correlations for each model, how many factors should be retained? Justify your response. Don't pay attention to interpretability of the solution for the purposes of this question.

Part 2

Obtain the eigenvalues of the sample correlation matrix ***R*** (CEFA gives you those, or compute them some other way).

a) What number of factors is suggested by the Kaiser criterion? Confront (and comment on) this with respect to your answer to **h)** in Part 1.

b) Draw a scree plot and use it to make a judgement on the number of factors to be retained. Again, confront (and comment on) this with respect to your answer to **h)** in Part 1.

Part 3

Conduct a rotation on the 4-factor OLS solution.

a) State what rotation method you used and whether it's an orthogonal or an oblique rotation. Explain the defining difference between the two.

b) In what sense is the rotated solution "better" than the unrotated one? Did the rotation help you in anything?

c) Briefly interpret the factors in the rotated solution.