PSY-GC 8867, Autumn 2022, Homework VII 1

In the file TV.txt is a correlation matrix of how much respondents “really like to watch” the 10 TV programs: Professional Boxing\*, This Week, Today, World of Sport\*, Grandstand\*, Line-Up, Match of the Day\*, Panorama, Rugby Special\*, 24 Hours, where those marked with \* are sport programs and the remaining are new programs. The sample size is quite large, but here we use  sample size 200 just for demonstrative purpose.

Use R for problems below. Show your code and outputs. Variables should be properly labeled with variable names. Don’t forget to answer questions and refer to parts of the outputs.

1. Create a screeplot and run parallel analysis (use 1, 000 replications, add error bars, and use 95th percentile for decision). What are the conclusions from screeplot and from parallel analysis?
2. Run PCA and retain 2 PCs, use an oblique target rotation. What is a proper target given the background of the 10 programs? What are the interpretation of the 2 components?
3. What proportion of total variance in the original variables is explained by the two components (as a whole)? At the individual level, which TV program has the largest proportion of variance  explained by the two PCs (and how much)? Which has the least (and how much)?
4. How much variance do the two PCs share with each other?
5. Express the PC1 in terms of the original standardize variables.
6. Write a regression equation that regresses Boxing on the two PCs.
7. What is the correlation between World Sport and PC1?