Stat ST465/665, Assignment 6

(73 points) This assignment is about factor analysis for the auction sales data for young bulls in bullsalesdata.xlsx *(R. Johnson and D. Wichern, Applied Multivariate Statistics, Pearson, 6th Edition, 2018).* The assignment involves 3 distinct factor analyses after an initial assessment.

For each factor analysis, display a table of the form.

Variables	load 1	load 2	load 3 (if specified)	commun.
YrHgt				
FtFrBody				
PrctFFB				
Frame				
BkFat				
SaleHt				
SaleWt				
Var. Acc. For				

where

- load j = j^{th} estimated loading = $\hat{\ell}_{ij}$

- commun = estimated communality = $\hat{h}_i^2$

- spec. var = specific variance = $\hat{\psi}_i$

- Var. Acc. For = Variance accounted for

1. **(18 points)** Preliminary investigation

   (a) Carry out a shortened initial investigation (steps 1, 2 and 3) based on the matrix scatter plot and box plot. Do not remove outliers or transform the data. Indicate if you had to process the data file in anyway. Explain any conclusions drawn from the evidence and backup your conclusions.

   (b) Explain why using the correlation matrix for the factor analysis is indicated.

   (c) Display the sample correlation matrix $\boldsymbol{R}$. Does the matrix $\boldsymbol{R}$ suggest the number of factors to use?

   (d) Perform a preliminary simplified principal component analysis using $\boldsymbol{R}$.

      i. List the eigenvalues and describe the percent contributions to the variance.

ii. Determine the number of principal components to retain and justify your answer by considering at least three methods. Note and comment if there is any disagreement between the methods.

(e) Include your code

2. **(32 points)** Factor analysis for 3 factors using the principal component method **without rotation** on the sample **correlation** matrix.

(a) Display the table of results.

(b) Show the error matrix $\boldsymbol{E} = \boldsymbol{R} - (\hat{\boldsymbol{L}} \, \hat{\boldsymbol{L}}^\top + \hat{\boldsymbol{\Psi}})$ and $\|\boldsymbol{E}\| = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} e_{ij}^2}$.

(c) Show plots of the loading vectors for the first two factors.

(d) Show the factor scores plot for the first two factors.

(e) Discuss the results.

(f) Include your code

3. **(16 points)** Factor analysis for 3 factors using the principal component method **with varimax rotation** on the sample **correlation** matrix.

(a) Display the table of results.

(b) Discuss the results.

(c) Include your code

4. **(7 points)** Factor analysis for 2 factors using the principal component method **without rotation** on the sample **correlation** matrix.

(a) Show the error matrix $\boldsymbol{E} = \boldsymbol{R} - (\hat{\boldsymbol{L}} \, \hat{\boldsymbol{L}}^\top + \hat{\boldsymbol{\Psi}})$ and $\|\boldsymbol{E}\| = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} e_{ij}^2}$.

(b) Discuss the results.

(c) Include your code