

## Math 5366 Homework 31

1. The file `Hw31data.txt` contains SAS code for generating two data sets. The first data set provides the correlation matrix of six measurements made on white leghorn fowls, including skull length (SL), skull breadth (SB), humerus length (HS), ulna length (UL), femur length (FL), and tibia length (TL).
  - (a) Perform a principal components analysis for this data set, and report the resulting eigenvalues and eigenvectors.
  - (b) How many principal components are required to explain at least 90% of the total variation in the data?
  - (c) Provide an intuitive interpretation for the principal components accounting for 90% of the total variation. (For example, the first principal component has large positive coefficients for all of the variables in the data set, so it roughly measures the overall size of a white leghorn fowl.)
2. Perform a factor analysis on the leghorn data.
  - (a) How many factors are retained using the MINEIGEN criterion?
  - (b) What is the overall RMS of off-diagonal residuals in this case?
3. Continuing with the leghorn data set, increase the number of factors until the overall residual RMS is less than 0.05.
  - (a) How many factors are required to achieve this?
  - (b) Report the estimated matrices  $\hat{L}$  and  $\hat{\Psi}$ .
  - (c) What is the communality for skull length?
  - (d) Find the unique variance of ulna length.
  - (e) What is the correlation between femur length and the 2nd factor?
4. The second data set in `Hw31data.txt` contains responses of 122 diabetes patients to 25 survey questions, on a Likert scale (a scale typically used on surveys, where 1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Neither Disagree Nor Agree, 4 = Somewhat Agree, and 5 = Strongly Agree).
  - (a) Perform a factor analysis with `nfact=17` on this data and store the factor scores in a data set.
  - (b) One of the assumptions of the factor model is that  $\text{cov}(f) = I_{k \times k}$ . Verify that the sample covariance matrix of the factor scores is equal to  $I$  (This occurs exactly, because we are using the principal component method for this factor analysis. There are other methods where this does not occur.)
  - (c) What does  $\text{cov}(f) = I_{k \times k}$  say about the correlation between two different factors? What would you expect the scatter plot of two different factors to look like?
  - (d) Create a scatterplot of  $f_1$  vs.  $f_2$ . Does this plot agree with your expectations?
  - (e) It would be interesting to interpret the factors in this problem, but in order to do that, we will need to consider rotations of the factors, so this will have to wait until a future homework assignment.