Assignment 05

Collinearity and PCA

This goal of this assignment is to give you experience diagnosing collinearity and using principal components analysis to create orthogonal composites that can be used in a regression model. Turn in a printed version of your responses to each of the questions on this assignment.

In addition, please adhere to the following guidelines for further formatting your assignment:

- All graphics should be set to an appropriate aspect ratio and sized so that they do not take up more room than necessary. They should also have an appropriate caption.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document.
- Syntax or computer output should not be included in your assignment unless it is specifically asked for.

This assignment is worth 22 points.

Data Set

In 2018, Iowa attorneys rated the 64 judges who were up for election on 12 different attributes in a Judicial Performance Review. They also indicated whether or not the judge should be retained. In this assignment, you are going to examine whether those ratings we can explain variation in the percentage of attorneys who endorsed retention using the data provided in the file *iowa-judges.csv* (see the data codebook).

Exploratory Analysis

1. Compute and report the eigenvalues of correlation matrix of the 12 predictors.

2. Based on the eigenvalues, comment on whether there may be any potential collinearity problems. Explain.

3. Compute and report the condition index for each of the 12 predictors.

4. Based on the condition indices, comment on whether there may be any potential collinearity problems. Explain.

5. Fit the standardized WLS model that regresses the standardized retention percentage on the 12 standardized predictors. Since the ratings are assigned based on different numbers of attorneys, use a weight equal to the number of respondents. Report the coefficient-level output, including the estimated coefficients (beta weights), standard errors, t-values, and p-values.

6. Compute and report the variance inflation factors.

7. Interpret the largest VIF value.

8. Based on the VIF values, comment on whether there may be any potential collinearity problems. Explain.

Principal Components Analysis

In this section you are going to carry out the principal components analysis by using singular value decomposition on the correlation matrix of the predictors.

9. Carry out the singular value decomposition on the variance–covariance matrix of the predictors. Report the **D** matrix.

10. Calculate and report the variance for the first principal component based on the values from the **D** matrix. Show your work.

11. Compute and interpret the proportion of variance accounted for by the first principal component. Show your work.

12. Compute the composite score based on the first principal component for the first observation (Judge John J. Bauercamper). Show your work.

Choosing the Number of Principal Components

Read the section on scree plots (Section 4) in this web article.

13. Create a scree plot showing the eigenvalues for the 12 principal components from the previous analysis.

14. Using the "elbow criterion", how many principal components are sufficient to describe the data? Explain by referring to your scree plot.

15. Using the "Kaiser criterion", how many principal components are sufficient to describe the data? Explain.

16. Using the "80% proportion of variance criterion", how many principal components are sufficient to describe the data? Explain.

Revisit the Regression Analysis

The evidence from the previous section suggests that the first two principal components are sufficient to explain the variation in the predictor space.

17. By examining the pattern of correlations (size and directions) in the first two principal components, try to identify which constructs the composites based on these components define. Explain.

18. Fit the regression analysis using the first two principal components as predictors of retention percentage. (Don't forget your weights.) Create and report the plot of the residuals vs. fitted values. Fit the same regression model, but this time also include a quadratic effect of the first principal component. Create and report the plot of the residuals vs. fitted values. Place these plots side-by-side and use the caption to comment on the assumption of linearity (the average residual is zero at each fitted value) for each model. (**2pts**)

19. Examine the coefficient-level output and re-fit the model by removing any non-significant predictors from the model. Re-examine the coefficient-level output and create a plot showing the fitted values from the model as a function of the first principal component. Use this plot to interpret the quadratic nature of the effect. (**2pts**)

20. Based on Cook's D, identify the name of any judges (and their Cook's D value) that are influential observations.