

# Assignment 05

## Collinearity and Dimension Reduction

This assignment is worth 16 points.

This assignment is worth 22 points.

### Exploratory Analysis

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

```
# A tibble: 12 x 13
  term      knowledge perception punctuality attention management demeanor
  <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1 knowledge    NA        0.967      0.722      0.915      0.867      0.708
2 perception  0.967      NA        0.755      0.953      0.910      0.811
3 punctuality 0.722      0.755      NA        0.810      0.827      0.661
4 attention   0.915      0.953      0.810      NA        0.935      0.855
5 management  0.867      0.910      0.827      0.935      NA        0.831
6 demeanor    0.708      0.811      0.661      0.855      0.831      NA
7 clarity     0.953      0.952      0.755      0.922      0.857      0.739
8 promptness  0.715      0.729      0.851      0.738      0.759      0.590
9 criticism   0.720      0.826      0.701      0.857      0.821      0.964
10 decision   0.839      0.915      0.752      0.923      0.868      0.876
11 courteous  0.642      0.754      0.603      0.800      0.777      0.985
12 equality    0.754      0.840      0.689      0.879      0.837      0.930
  clarity promptness criticism decision courteous equality
  <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1  0.953      0.715      0.720      0.839      0.642      0.754
2  0.952      0.729      0.826      0.915      0.754      0.840
3  0.755      0.851      0.701      0.752      0.603      0.689
4  0.922      0.738      0.857      0.923      0.800      0.879
5  0.857      0.759      0.821      0.868      0.777      0.837
6  0.739      0.590      0.964      0.876      0.985      0.930
7  NA        0.766      0.765      0.879      0.668      0.783
8  0.766      NA        0.592      0.677      0.520      0.587
9  0.765      0.592      NA        0.928      0.951      0.950
10 0.879      0.677      0.928      NA        0.835      0.940
11 0.668      0.520      0.951      0.835      NA        0.918
12 0.783      0.587      0.950      0.940      0.918      NA
```

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

The pairwise correlations are quite high (many above 0.70) indicating that there will likely be collinearity problems.

3. Compute and report the eigenvalues for the correlation matrix you created in Question 1.

```
[1] 9.93 1.01 0.49 0.17 0.15 0.08 0.05 0.04 0.04 0.02 0.01 0.01
```

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

Yes these values portend collinearity problems. They sum of the reciprocals of the eigenvalues (357.37) is not 0 and is greater than five times the number of predictors ( $5k = 60$ ).

## Initial Model

5. Report the coefficient-level output, including the estimated coefficients (beta weights), standard errors,  $t$ -values, and  $p$ -values.

Table 1: Coefficients, standard errors,  $t$ -values, and  $p$ -values for the predictors included in a standardized model to explain variation in judge retention voting.

Predictor	$\beta$	SE	$t$	$p$
Knowledge	0.23	0.210	1.11	0.273
Perception	0.40	0.254	1.58	0.120
Punctuality	-0.12	0.096	-1.20	0.235
Attention	0.12	0.171	0.72	0.474
Management	0.07	0.133	0.56	0.577
Demeanor	-0.11	0.328	-0.32	0.749
Clarity	-0.43	0.169	-2.52	0.015
Promptness	0.02	0.084	0.27	0.790
Criticism	0.96	0.233	4.10	0.000
Decision	0.05	0.204	0.27	0.790
Courteous	-0.40	0.291	-1.37	0.177
Equality	0.13	0.161	0.84	0.406

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

```
scale(knowledge) scale(perception) scale(punctuality) scale(attention)
33.517388      51.100146      8.494383      24.627348
scale(management) scale(demeanor) scale(clarity) scale(promptness)
14.894122      96.538137      22.412867      5.854343
scale(criticism) scale(decision) scale(courteous) scale(equality)
50.417915      35.905919      77.042138      24.323369
```

Yes these values portend collinearity problems. Most of the VIF values are greater than 10 indicating exacerbated estimates of the uncertainty associated with several of the coefficients.

7. Using the predictor with the largest VIF value, use the VIF value to indicate how the standard error for this predictor will be impacted by the collinearity.

The largest VIF value is 96.5 and is associated with the scaled attention variable. This suggests that the SE associated with the scaled attention coefficient is 9.8 times larger than it would be if all the variables in the model were independent.

## Principal Components Analysis

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

[1] 2.081978

$$\begin{aligned} PC_1 &= -0.286(-0.892) - 0.303(-1.072) - 0.264(0.177) - 0.308(-1.205) - 0.299(-0.070) - 0.290(-0.621) - 0.292(-1.232) \\ &\quad - 0.246(-0.437) - 0.294(-0.338) - 0.304(-0.466) - 0.276(-0.511) - 0.295(-0.428) \\ &\approx 2.08 \end{aligned}$$

*Note: Sign might be negative on the final answer.*

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

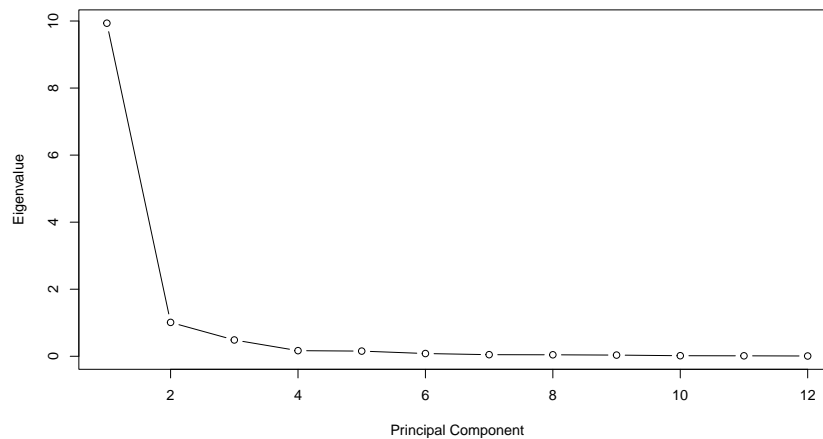


Figure 1: Scree plot show the eigenvalues for each of the principal components.

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

It looks like about two principal components are sufficient to describe the data based on where the elbow is located in the scree plot. The eigenvalues are pretty constant after the second principal component.

**11. Using the “Kaiser criterion”, how many principal components are sufficient to describe the data? Explain.**

```
[1] TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

Based on the Kaiser criterion, two principal components are sufficient to describe the data, as only two of the 12 components have an eigenvalue above 1.

**12. Using the “80% proportion of variance criterion”, how many principal components are sufficient to describe the data? Explain.**

```
[1] 0.8278805 0.9120898 0.9526008 0.9665619 0.9794068 0.9862351 0.9902118  
[8] 0.9938893 0.9968262 0.9982047 0.9993187 1.0000000
```

One principal components is sufficient to describe the data if we use the “80% proportion of variance criterion”, since the first PC explains more than 80% of the variation.

## Revisit the Regression Analysis

**13. By examining the pattern of correlations (size and directions) in the first two principal components, identify the construct defined by the composites of these two components. Explain.**

The first principal component seems to be a general measurement of judges’ demeanor and professional competence, as the signs of the weights are all in the same direction and the magnitudes of the weights are all of moderate size. The second PC seems to be measuring a contrast between the demeanor variables and the professional conduct variables, as the signs for these weights are in the opposite direction.

14. 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. What does this suggest about the validity of the linearity assumption?

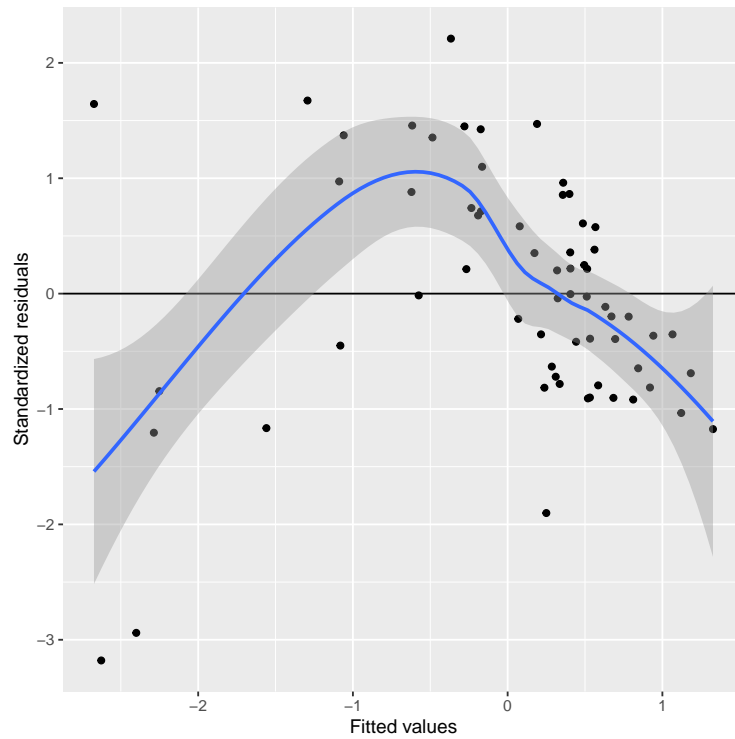


Figure 2: Plot of the standardized residuals versus the fitted values for the model that includes the first two principal components as predictors of retention.

The curvature of the loess smoother, and non-overlap with the  $Y = 0$  line, indicates that the average residual at each fitted value is not 0. This implies that the linearity assumption may be violated.

15. Again, fit the same regression model using the first two principal components as predictors of retention percentage, but this time also include a quadratic effect of the first principal component. Create and report the plot of the residuals vs. fitted values. What does this suggest about the validity of the linearity assumption?

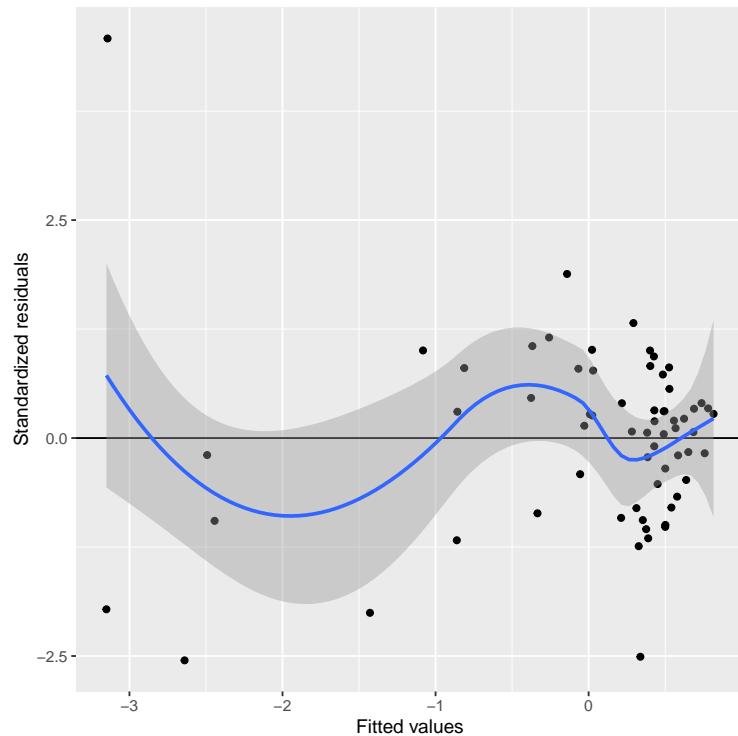


Figure 3: Plot of the standardized residuals versus the fitted values for the model that includes the first two principal components as predictors of retention.

The loess smoother aligns with the  $Y = 0$  line, indicating that the average residual at each fitted value is 0. This implies that the data are consistent with meeting the linearity assumption.

16. Use the coefficient-level output from the model fitted in Question 15, interpret the effect of the first principal component.

There is a positive quadratic effect of the first principal component on retention, controlling for differences in the second principal component. The same amount of changes in PC1 values at lower levels of the scale are associated with higher rate of change in retention voting (on average) and less change in retention voting at higher levels of the PC1.

## Influential Values

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

```
# A tibble: 5 x 2
  judge                .cooksD
  <chr>                <dbl>
1 Wittig, Monica      0.570
2 Brandt, Gregory D.  0.0742
3 Price, William A.   0.155
4 Seymour, Racheal    2.85
5 Farmer Minot, Deborah 0.420
```

*Note: If the decision is based on the index plot, Rachael Seymour, Monica Wittig, Deborah Farmer Minot, and William A. Price would likely be chosen.*

2em

18. Remove any influential observations identified in Question 17. Re-fit the same model. Based on comparing the model- and coefficient-level output for this model and the model which included all the observations, comment on how these observations were influencing the  $R^2$  value, the estimate of the quadratic effect of PC1, and the effect of PC2.

```
# A tibble: 4 x 5
  term      estimate std.error statistic  p.value
  <chr>      <dbl>    <dbl>    <dbl>    <dbl>
1 (Intercept) 0.158    0.0485     3.26 1.82e- 3
2 PC1         0.579    0.0485    11.9 1.70e-17
3 I(PC1^2)    -0.109    0.0211    -5.15 3.06e- 6
4 PC2         0.0763   0.0931     0.819 4.16e- 1
```

```
# A tibble: 4 x 5
  term      estimate std.error statistic  p.value
  <chr>      <dbl>    <dbl>    <dbl>    <dbl>
1 (Intercept) 0.190    0.0421     4.51 3.47e- 5
2 PC1         0.554    0.0418    13.3 9.13e-19
3 I(PC1^2)    -0.133    0.0250    -5.33 1.92e- 6
4 PC2         0.0355   0.0952     0.373 7.11e- 1
```

After the five influential observations were removed,

- The  $R^2$  value decreased from 0.920 to 0.870.
- The quadratic effect of PC1 value became stronger, decreasing from  $-0.109$  to  $-0.133$ .
- The effect of PC2 value became weaker, decreasing from 0.076 to 0.036.