Assignment 05

Variable Reduction

Answer Key

This assignment is worth 22 points.

Exploratory Analysis

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

```
# Create correlation matrix
X = cor(iowa[ , 6:17])

#Compute eigen values
round(eigen(X)$values, 2)
```

```
##  [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
```

2. 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 ($5k = 60$).

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

```
# Condition indices
round(sqrt(eigen(X)$values[1] / eigen(X)$values), 2)
```

```
##  [1]  1.00  3.14  4.52  7.70  8.03 11.01 14.43 15.00 16.79 24.51 27.26 34.86
```

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

Yes these values portend collinearity problems. Several of the condition indices are greater than 15, and one is larger than 30.

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.

```
# Fit model
lm.1 = lm(scale(retention) ~ -1 + scale(knowledge) + scale(perception) + scale(punctuality) +
            scale(attention) + scale(management) + scale(demeanor) + scale(clarity) +
            scale(promptness) + scale(criticism) + scale(decision) + scale(courteous) +
            scale(equality), data = iowa, weights = respondents)
```

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	β	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. Compute and report the variance inflation factors.

```
# Compute VIF
vif(lm.1)
```

```
##   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
```

7. Interpret the largest VIF value.

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

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

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.

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 correlation matrix of the predictors. Report the D matrix.

```
# Singular value decomposition
diag(round(svd(X)$d, 3))
```

```
##        [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10] [,11] [,12]
##  [1,] 9.935 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##  [2,] 0.000 1.011 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##  [3,] 0.000 0.000 0.486 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##  [4,] 0.000 0.000 0.000 0.168 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##  [5,] 0.000 0.000 0.000 0.000 0.154 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##  [6,] 0.000 0.000 0.000 0.000 0.000 0.082 0.000 0.000 0.000 0.000 0.000 0.000
##  [7,] 0.000 0.000 0.000 0.000 0.000 0.000 0.048 0.000 0.000 0.000 0.000 0.000
##  [8,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.044 0.000 0.000 0.000 0.000
##  [9,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.035 0.000 0.000 0.000
## [10,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.017 0.000 0.000
## [11,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.013 0.000
## [12,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.008
```

10. Report the standard deviation for the first principal component based on the values from the D matrix. Show your work.

$$\sqrt{\lambda} = \sqrt{9.935}$$

$$\approx 3.15$$

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

```
# Proportion of variance
svd(X)$d[1] / sum(svd(X)$d)
```

```
## [1] 0.8278805
```

The first principal component accounts for 82.8% of the variation in the predictors.

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

```
# Obtain vector of PC weights
pc1 = svd(X)$v[ , 1]

# Obtained scaled variable values for Obs. 1
judge_1 = scale(iowa[ , 6:17])[1, ]

# Compute composite value
sum(pc1 * judge_1)
```

```
## [1] 2.081978
```

$$\begin{aligned} \text{PC}_1 &= -0.286(-0.892) - 0.303(-1.072) - 0.264(0.177) - 0.308(-1.205) - 0.299(-0.070) - 0.290(-0.621) - 0.292(-1.23 \\ &\quad - 0.246(-0.437) - 0.294(-0.338) - 0.304(-0.466) - 0.276(-0.511) - 0.295(-0.428) \\ &\approx 2.08 \end{aligned}$$

Note: Sign might be negative on the final answer.

Choosing the Number of Principal Components

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

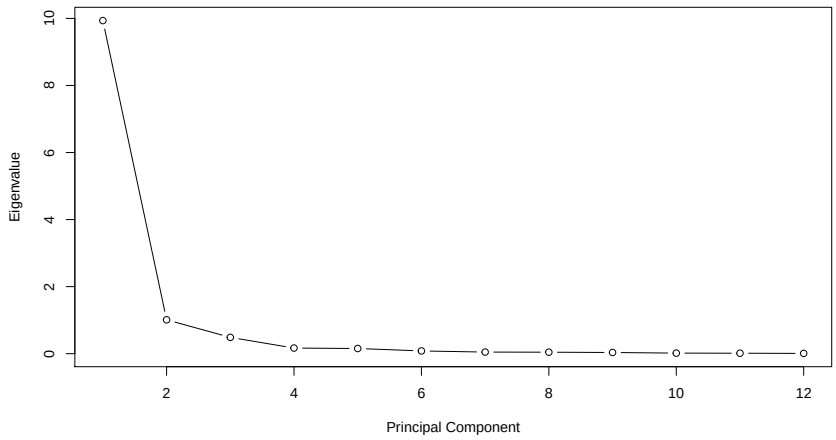
13. Create a scree plot showing the eigenvalues for the 12 principal components from the previous analysis.

```
plot(x = 1:12, y = svd(X)$d, type = "b", xlab = "Principal Component", ylab = "Eigenvalue")
```

14. 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.

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



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

```
svd(X)$d > 1
```

```
##  [1]  TRUE  TRUE FALSE 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.

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

```
# Compute cumulative proportion of variance
cumsum(svd(X)$d / sum(svd(X)$d))
```

```
##  [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

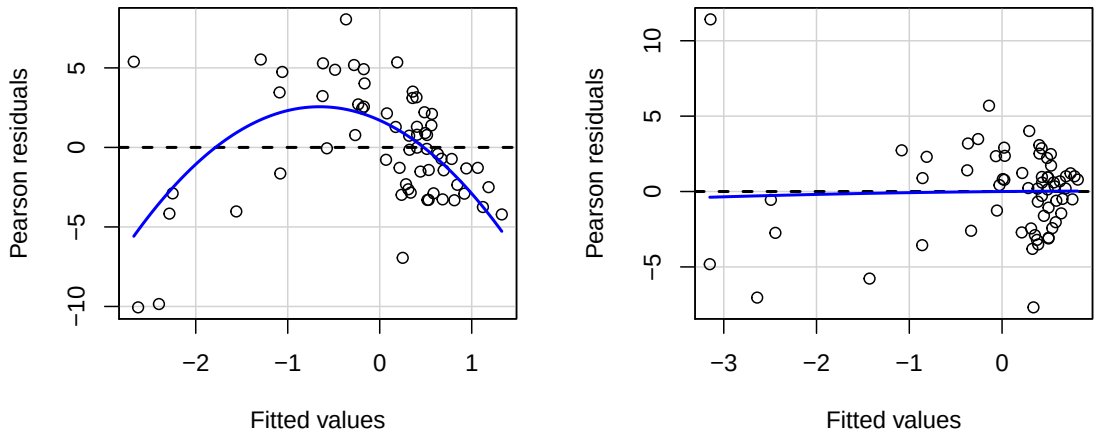
The evidence from the previous section suggests that the first two principal components are sufficient to explain the variation in the predictor space.

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.

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.

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)

Figure 2. *LEFT: Plot of the residuals versus the fitted values for Model 1 (linear effects). The curvature of the loess smoother indicates that the average residual at each fitted value is not 0. RIGHT: In contrast, the same plot for Model 2 (inc. quadratic effect of PC1) suggests that this assumption has been satisfied as the loess smoother follows the Y=0 line.*



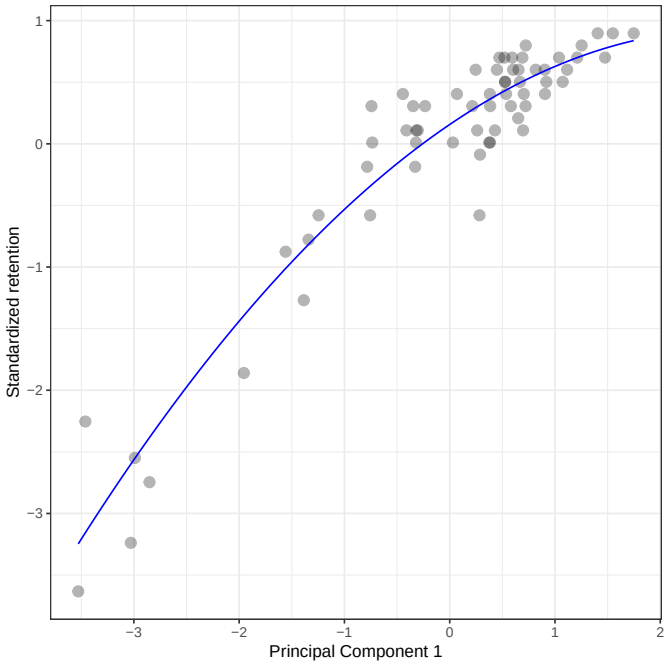
19. Examine the coefficient-level output and re-fit the model by removing any non-significant predictors from the model. Re-examine the coefficient-level output and create a plot showing the fitted values from the model as a function of the first principal component. Use this plot to interpret the quadratic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

```
# Remove PC2
lm.2.2 = lm(retention ~ PC1 + I(PC1 ^ 2), data = iowa_pc, weights = respondents)

# Coefficient-level output
tidy(lm.2.2)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)    0.157    0.0483      3.25 1.86e- 3
## 2 PC1            0.580    0.0483     12.0  1.07e-17
## 3 I(PC1^2)      -0.109    0.0210     -5.19 2.54e- 6
```

```
# Create plot
ggplot(data = iowa_pc, aes(x = PC1, y = retention)) +
  geom_point(size = 3, alpha = 0.3) +
  stat_function(
    fun = function(x) {0.157 + 0.580*x - 0.109 * x^2},
    color = "blue",
    linetype = "solid"
  ) +
  theme_bw() +
  xlab("Principal Component 1") +
  ylab("Standardized retention")
```



The effect of the first principal component on retention is non-linear. The same amount of changes in PC 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 PC.

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

```
# Cook's D cutoff
cutoff = 4/61

# Get judges with extreme Cook's D values
augment(lm.2.2) %>%
  mutate(judge = iowa$judge) %>%
  select(judge, .cooksd) %>%
  filter(.cooksd > cutoff)
```

```
## # A tibble: 4 x 2
##   judge                .cooksd
##   <chr>                  <dbl>
## 1 Wittig, Monica         0.203
## 2 Price, William A.      0.166
## 3 Seymour, Racheal       3.84
## 4 Farmer Minot, Deborah  0.486
```

Note: If the decision is based on the index plot, only Rachael Seymour (obs. 45) and Deborah Farmer Minot (Obs. 49) would likely be chosen.