Dimension reduction via PCA and FA

Student name

Principal Components

1. PMA6 14.1 (modified): For the depression data set, perform a PCA on the last seven variables `DRINK-CHRONILL`.

Use the covariance matrix, but *do not center or scale* the data. You should have the codebook for this data set open during this homework.

a) Determine the number of PC's to retain that contain 80% of the original variance.

b) Examine the loadings of the retained PC's using a heatmap. Interpret each PC as it relates to the individual questions.

2. Repeat question #1 using the correlation matrix instead of the covariance matrix.

Compare the results and comment. (Are the same number of PC's retained? Are the loadings different?)

3. Repeat question #1 after normalizing the data (centering and scaling). Use the `scale()` function here to help. Compare the results and comment. (Are the same number of PC's retained? Are the loadings different?)

4 (PMA6 14.2 modified): Perform a regression analysis of `CASES` on your retained PC's *from question 3* Interpret the results. Recall that PC's are on the scale of a standard deviation of the PC scale. So you can say "for every one standard deviation on the PC1 scale an individual is. . . .". But don't just call it "PC1". Use your interpretation from the loadings visualized in question 3.

5 (PMA6 14.11 modified): This question uses the Parental HIV data set.

a) If you were to conduct a PCA on the items of the Parental Bonding scale, *a priori* how many PC's would you expect to retain for this scale?

b) Perform this PCA. How many components should be retained based on the rules of thumb mentioned in Section 14.5 (% variance retained, Eigenvalues > 1, scree plot)?

Factor Analysis

1. (PMA6 15.1). The CESD scale items (`C1-C20`) from the depression data set were used to obtain the factor loadings listed in Table 15.7. The initial factor solution was obtained from the principal components method, and a varimax rotation was performed. Analyze this same data set by using an oblique rotation such as the direct quartimin procedure. Compare the results.

2. (PMA6 15.6) Separate the depression data set into two subgroups, men and women. Using four factors, repeat the factor analysis in Table 15.7. Compare the results of your two factor analyses to each other, and do the results in Table 15.7.

3. (PMA6 15.8) Perform a factor analysis on all of the items of the Parental Bonding scale for the Parental HIV data set. Retain two factors. Rotate the factors using an orthogonal rotation. Do the items with the highest loadings for each of the factors correspond to the items of the overprotection and care scale? Interpret the findings.

4. (PMA6 15.9) Repeat problem #3 (15.8) using an oblique rotation. DO the substantive conclusions change?