Assignment #7: Factor Analysis (0 points)

<u>Data:</u> The data for this assignment will be the correlation matrix input from the Stoetzel article in the course reserves.

'A Factor Analysis of Liquor Preference' by Stoetzel (Journal of Marketing Research)

<u>Assignment Instructions:</u>

Stoetzel's liquor preference application of factor analysis is a classic factor analysis example. It is also a useful example for those who are likely to see factor analysis in the context of marketing, and in particular marketing segmentation.

Begin the assignment by reading the article placed on course reserve. In particular note:

(a) Who comprises the sample population?
(b) How was the data collected?
(c) What is the data?

In this application the factor analysis is performed using the correlation matrix. What does this correlation matrix represent? The correlation matrix is in the paper. However, since the edge of the paper is difficult to read, we will provide the correlation matrix here.

1.000	0.210	0.370	-0.32	0.000	-0.31	-0.26	0.090	-0.38
0.210	1.000	0.090	-0.29	0.120	-0.30	-0.14	0.010	-0.39
0.370	0.090	1.000	-0.31	-0.04	-0.30	-0.11	0.120	-0.39
-0.32	-0.29	-0.31	1.00	-0.16	0.25	-0.13	-0.14	0.900
0.00	0.120	-0.04	-0.16	1.000	-0.20	-0.03	-0.08	-0.38
-0.31	-0.30	-0.30	0.25	-0.20	1.000	-0.24	-0.16	0.180
-0.26	-0.14	-0.11	-0.13	-0.03	-0.24	1.000	-0.20	0.040
0.090	0.010	0.120	-0.14	-0.08	-0.16	-0.20	1.000	-0.24
-0.38	-0.39	-0.39	0.900	-0.38	0.180	0.040	-0.24	1.000

(1) Let's begin by loading this correlation matrix into R. How do we do that? Start with a vector of values, and then read that vector of values into a matrix object.

```
cor.values <- c(1.000,0.210,0.370,-0.32,0.000,-0.31,-0.26,0.090,-0.38,
               0.210,1.000,0.090,-0.29,0.120,-0.30,-0.14,0.010,-0.39,
               0.370,0.090,1.000,-0.31,-0.04,-0.30,-0.11,0.120,-0.39,
               -0.32,-0.29,-0.31,1.00,-0.16,0.25,-0.13,-0.14,0.900,
               0.00,0.120,-0.04,-0.16,1.000,-0.20,-0.03,-0.08,-0.38,
               -0.31,-0.30,-0.30,0.25,-0.20,1.000,-0.24,-0.16,0.180,
               -0.26,-0.14,-0.11,-0.13,-0.03,-0.24,1.000,-0.20,0.040,
               0.090,0.010,0.120,-0.14,-0.08,-0.16,-0.20,1.000,-0.24,
               -0.38,-0.39,-0.39,0.900,-0.38,0.180,0.040,-0.24,1.000
               );
```

```
# How do we put these correlation values into a correlation matrix?;
help(matrix)
cor.matrix <- matrix(cor.values,nrow=9,ncol=9,byrow=TRUE);

# Check that object is a matrix object;
is.matrix(cor.matrix)
# Check that matrix is symmetric;
# This check helps check for typos;
isSymmetric(cor.matrix)
```

We can check most data types in R using an is.* function. We type cast in R using an as.* function.

(2) Stoetzel estimated a three factor model in his paper. We will estimate a three factor model with a VARIMAX rotation using maximum likelihood factor analysis. Did you get the same results as Stoetzel?

```
f.1 <- factanal(covmat=cor.matrix, n.obs=1442, factors=3, rotation='varimax');
names(f.1)
```

a. Are the results the same numerically, i.e. did you get the same factor loadings as Stoetzel? Do you expect to be able to reproduce a factor analysis and get the same factor loadings?

b. Are the results the same qualitatively, i.e. did you get the same factor interpretations as Stoetzel? Do you expect to be able to reproduce a factor analysis and get the same factor interpretation?

c. Does the statistical inference for the maximum likelihood factor analysis suggest that three factors are the correct number of factors to describe this correlation matrix? What is the null hypothesis for the chi-square test statistic? Do we reject or fail to reject this null hypothesis?

   Note that this hypothesis cannot be expressed in statistical notation like most hypotheses tests in Predict 410. (Hint: See Section 11.5 of Everitt.)

(3) Can we find the correct number of factors to describe this correlation matrix? Fit factor models for k=1 through 6. Do any of these models represent the correct number of factors based on the inference results?

(4) The VARIMAX factor rotation is an example of an orthogonal factor rotation. We also have oblique factor rotations. One example of an oblique factor rotation is the PROMAX rotation. Fit a three factor model with a PROMAX rotation using maximum likelihood factor analysis.

```
g.1 <- factanal(covmat=cor.matrix, n.obs=1442, factors=3, rotation='promax');
```

a. Does this model have better interpretability than the three factor model with the VARIMAX rotation?

b. Does the statistical inference for this maximum likelihood factor analysis suggest that three factors are the correct number of factors to describe this correlation matrix? Should the factor rotation affect the statistical inference for the number of factors?

(5) Use the factor loadings and the specific (or unique) variances to approximate the correlation matrix. Measure the fit of these approximations using the Mean Absolute Error of the residual matrix. Which factor model better approximates the correlation matrix?

```
gamma.f1 <- f.1$loadings;
approx.f1 <- gamma.f1%*%t(gamma.f1) + diag(f.1$uniqueness);
mae.f1 <- mean(abs(approx.f1-cor.matrix))

gamma.g1 <- g.1$loadings;
approx.g1 <- gamma.g1%*%t(gamma.g1) + diag(g.1$uniqueness);
mae.g1 <- mean(abs(approx.g1-cor.matrix))
```

Do we understand what we are computing? (Hint: It is in your pdf note packet on factor analysis.)

<u>Assignment Document:</u>

All assignment reports should conform to the standards and style of the report template provided to you. Results should be presented and discussed in an organized manner with the discussion in close proximity of the results. The report should not contain unnecessary results or information. The document should be submitted in pdf format. Name your file Assignment7_LastName.pdf.