Jiayuan Shi
MA684 HW10

1. As a first step in summarizing the information from the above questionnaire, perform a principal components analysis with a varimax (orthogonal) rotation.

```
> compout <-
princomp(~D1+D2+D3+D4+D5+D6+D7+D8+D9+D10+D11+D12+D13+D14+D15
+D16+D17+D18+D19+D20,
+ cor=TRUE)
> summary(compout)
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4
Standard deviation     2.7145877 1.8174272 1.5835824 0.94554542
Proportion of Variance 0.3684493 0.1651521 0.1253867 0.04470281
Cumulative Proportion  0.3684493 0.5336014 0.6589880 0.70369085
                           Comp.5     Comp.6     Comp.7     Comp.8
Standard deviation     0.90388394 0.82504997 0.79458115 0.74380537
Proportion of Variance 0.04085031 0.03403537 0.03156796 0.02766232
Cumulative Proportion  0.74454116 0.77857653 0.81014449 0.83780681
                           Comp.9    Comp.10    Comp.11    Comp.12
Standard deviation     0.70123071 0.65591710 0.59593519 0.57373484
Proportion of Variance 0.02458623 0.02151136 0.01775694 0.01645858
Cumulative Proportion  0.86239304 0.88390440 0.90166134 0.91811992
                          Comp.13    Comp.14    Comp.15    Comp.16
Standard deviation     0.56709196 0.55089241 0.50693687 0.44793824
Proportion of Variance 0.01607966 0.01517412 0.01284925 0.01003243
Cumulative Proportion  0.93419958 0.94937371 0.96222296 0.97225539
                           Comp.17     Comp.18     Comp.19     Comp.20
Standard deviation     0.424918743 0.389511104 0.360453014 0.304451949
Proportion of Variance 0.009027797 0.007585945 0.006496319 0.004634549
Cumulative Proportion  0.981283187 0.988869132 0.995365451 1.000000000

> vars <- data.frame(D1, D2, D3, D4, D5, D6, D7, D8, D9, D10,
+ D11, D12, D13, D14, D15, D16, D17, D18, D19, D20)
> principal(vars,nfactors=3,rotate="varimax")
Principal Components Analysis
Call: principal(r = vars, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
     PC1   PC2   PC3   h2   u2 com
D1   0.01  0.03  0.87 0.76 0.24 1.0
D2  -0.01  0.75  0.07 0.57 0.43 1.0
D3   0.73  0.19  0.04 0.57 0.43 1.1
D4   0.09  0.22  0.85 0.78 0.22 1.2
D5   0.62  0.39 -0.15 0.57 0.43 1.8
D6   0.31  0.21  0.67 0.59 0.41 1.6
D7   0.78 -0.10  0.17 0.64 0.36 1.1
D8   0.00  0.81  0.15 0.68 0.32 1.1
D9   0.68  0.25  0.05 0.53 0.47 1.3
```

D10	0.70	0.05	0.24	0.54	0.46 1.2
D11	0.86	0.14	0.03	0.76	0.24 1.1
D12	0.30	0.81	0.12	0.76	0.24 1.3
D13	0.84	-0.11	0.14	0.74	0.26 1.1
D14	0.83	0.16	-0.01	0.72	0.28 1.1
D15	0.61	0.33	0.18	0.52	0.48 1.8
D16	0.54	0.06	0.07	0.30	0.70 1.1
D17	0.07	0.06	0.92	0.85	0.15 1.0
D18	0.84	0.01	0.02	0.71	0.29 1.0
D19	0.19	0.83	0.12	0.74	0.26 1.1
D20	0.14	0.90	0.08	0.83	0.17 1.1

	PC1	PC2	PC3
SS loadings	6.25	3.91	3.02
Proportion Var	0.31	0.20	0.15
Cumulative Var	0.31	0.51	0.66
Proportion Explained	0.47	0.30	0.23
Cumulative Proportion	0.47	0.77	1.00

Mean item complexity = 1.2
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.05
with the empirical chi square 202.74 with prob < 9.3e-05

Fit based upon off diagonal values = 0.98

1A. How many components do you think are needed to summarize these 20 questions? Explain.
Solution: Based on the eigenvalue>1.0 criterion, we will retain 3 components.

1B*. How well do these components capture the information in these 20 questions? Explain.*
Solution: These 3 components capture (6.25+3.91+3.02)/20 = 65.9% of the information in the initial variables.

The 'h2' heading gives the communalities for each of the original variables, which describe proportion of variance of each initial variable captured by retained factors.
These 3 factors do a good job of explaining D17 (.85) and D20 (.83).
These 3 factors do a poorer job of explaining D16 (.30) and D15 (.52).
Therefore, these 3 factors capture 85% of the variability in the reason D17 (To forget about your problems), 83% of the variability in D20 (So you would not feel left out), only 30% of D16 (To celebrate on special occasions with friends) and only 52% of D15 (Because it would make you feel more confident and sure of yourself).

1C. Are there any questions that are not captured by the set of components? Explain.

Solution: No. All the questions are captured more or less by the set of three components. The 'h2' heading gives the communalities for each of the original variables, which describe proportion of variance of each initial variable captured by retained factors. No value in 'h2' column equals to 0, so all the questions are captured more or less by the set of three components.

2. From the principal components analysis, using the rotated solution, define the components - which items are related to which components? What do these components measure (suggest a name for the components)?

Solution:

The below component matrix (factor loadings) gives correlations between items and components. I'll focus on loadings > .5.

	PC1	PC2	PC3
D1	0.01	0.03	**0.87**
D2	-0.01	**0.75**	0.07
D3	**0.73**	0.19	0.04
D4	0.09	0.22	**0.85**
D5	**0.62**	0.39	-0.15
D6	0.31	0.21	**0.67**
D7	**0.78**	-0.10	0.17
D8	0.00	**0.81**	0.15
D9	**0.68**	0.25	0.05
D10	**0.70**	0.05	0.24
D11	**0.86**	0.14	0.03
D12	0.30	**0.81**	0.12
D13	**0.84**	-0.11	0.14
D14	**0.83**	0.16	-0.01
D15	**0.61**	0.33	0.18
D16	**0.54**	0.06	0.07
D17	0.07	0.06	**0.92**
D18	**0.84**	0.01	0.02
D19	0.19	**0.83**	0.12
D20	0.14	**0.90**	0.08

Component 1:

D3. Because it helped you enjoy a party
D5. To be sociable
D7. Because you liked the feeling
D9. Because it was exciting
D10. To get drunk
D11. Because it made social gatherings more fun
D13. Because it gave you a pleasant feeling
D14. Because it improved parties and celebrations
D15. Because it would make you feel more confident and sure of yourself
D16. To celebrate on special occasions with friends
D18. Because it was fun
=> Suggested Name: Have Fun.

Component 2:

D2. Because your friends pressured you to drink
D8. So that others wouldn't kid you about not drinking
D12. To fit in with a group you liked
D19. To be liked by others
D20. So you would not feel left out
=> Suggested Name: Be Social.

Component 3:
D1. To forget your worries
D4. Because it helped you when you felt depressed or nervous
D6. To cheer up when you were in a bad mood
D17. To forget about your problems
=> Suggested Name: Cheer Up.

3. We will use the principal components analysis as a guide in creating sub-scales by summing sets of items. Based on your findings above, calculate (have the computer calculate) sub-scales for the respondents by summing the responses to the appropriate items sfrom the questionnaire.

3A. Report the mean, standard deviation, and range for each of the sub-scales.

```
> havafun <- D3+D5+D7+D9+D10+D11+D13+D14+D15+D16+D18
> besocial <- D2+D8+D12+D19+D20
> cheerup <- D1+D4+D6+D17
```

Solution:

	Mean	Standard Deviation	Range
Have Fun	35.169	10.121	44
Be Social	9.900	4.889	20
Cheer Up	7.483	3.788	16

3B. Run a set of simple regressions, predicting each of your sub-scales from age of first drink. Based on these analyses, are there associations between drinking motives and age of first drink? Describe these associations.

```
> lm.havefun <- lm(havafun~AgeDrink)
> summary(lm.havefun)

Call:
lm(formula = havafun ~ AgeDrink)

Residuals:
    Min      1Q  Median      3Q     Max 
-24.7709  -6.7060   0.1967   7.2615  21.1643 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  52.2196     4.9860  10.473  < 2e-16 ***
AgeDrink     -0.9676     0.2802  -3.453 0.000676 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.855 on 199 degrees of freedom
Multiple R-squared:  0.05654,	Adjusted R-squared:  0.0518 
F-statistic: 11.93 on 1 and 199 DF,  p-value: 0.0006762

> lm.besocial <- lm(besocial~AgeDrink)
> summary(lm.besocial)

Call:
lm(formula = besocial ~ AgeDrink)

Residuals:
   Min     1Q Median     3Q    Max 
-5.640 -3.823 -1.415  2.973 15.585 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  13.4976     2.4661   5.473 1.32e-07 ***
AgeDrink     -0.2041     0.1386  -1.473    0.142    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.875 on 199 degrees of freedom
Multiple R-squared:  0.01079,	Adjusted R-squared:  0.005815 
F-statistic:  2.17 on 1 and 199 DF,  p-value: 0.1423

> lm.cheerup <- lm(cheerup~AgeDrink)
> summary(lm.cheerup)

Call:
lm(formula = cheerup ~ AgeDrink)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.0927 -3.0367 -0.6967  1.3593 12.5673 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   9.8087     1.9141   5.124 7.04e-07 ***
AgeDrink     -0.1320     0.1076  -1.227    0.221    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.783 on 199 degrees of freedom
Multiple R-squared:  0.007511,	Adjusted R-squared:  0.002524 
F-statistic: 1.506 on 1 and 199 DF,  p-value: 0.2212
```

Solution:

For the Have Fun group, the associations between drinking motive of having fun and age of first drink is significant, because the p-value for the age is less than 0.01. So we can reject the null hypothesis, which is there is no association between drinking motive of having fun and age of first drink, and conclude that the association is significant.
For each year increase in one's age, the drinking motive of having fun decreases by 0.9676 units.

For the Be Social group, the associations between drinking motive of being social and age of first drink is not significant, because the p-value for the age is 0.142, larger than 0.05. So we cannot reject the null hypothesis, which is there is no association between drinking motive of being social and age of first drink, and conclude that the association is not significant.
Although the association is not significant, we can describe it as for each year increase in one's age, the drinking motive of having fun decreases by 0.2041 units.

For the Cheer Up group, the associations between drinking motive of cheering up and age of first drink is not significant , because the p-value for the age is 0.221, larger than 0.05. So we cannot reject the null hypothesis, which is there is no association between drinking motive of cheering up and age of first drink, and conclude that the association is not significant.
Although the association is not significant, we can describe it as for each year increase in one's age, the drinking motive of having fun decreases by 0.1320 units.