

Name Jerry

Student Number \_\_\_\_\_

## STA 431 Quiz 9

1. Let

$$z_1 = \lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3 + e_1$$

$$z_2 = \lambda_4 F_1 + \lambda_5 F_2 + \lambda_6 F_3 + e_2$$

where all expected values are zero, variances of the factors equal one, and all the factors and error terms are independent. As the notation suggests,  $Var(z_1) = Var(z_2) = 1$ . Only  $z_1$  and  $z_2$  are observable. Please give Greek letter answers to the following.

(a) (2 points) What is  $Var(e_1)$ ? Show a bit of work. **Circle your answer.**

$$\begin{aligned} Var(z_1) &= 1 = \lambda_1^2 Var(F_1) + \lambda_2^2 Var(F_2) + \lambda_3^2 Var(F_3) + Var(e_1) \\ &= \lambda_1^2 + \lambda_2^2 + \lambda_3^2 + Var(e_1) \end{aligned}$$

$$\text{So } Var(e_1) = \boxed{1 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2}$$

(b) (1 point) What is  $Corr(z_1, F_3)$ ? Show a bit of work.

$$\begin{aligned} Corr(z_1, F_3) &= Cov(\lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3 + e_1, F_3) \\ &= 0 + Cov(\lambda_3 F_3, F_3) = \lambda_3 Var(F_3) \\ &= \lambda_3 \quad (\text{Don't have to show so much work}) \end{aligned}$$

(c) (1 point) What is the reliability of  $z_1$  as a measurement of  $F_1$ ?

$$\lambda_1^2$$

(d) (2 points) What is the reliability of  $s = z_1 + z_2$  as a measurement of  $F_1$ ? Show your work. **Circle your answer.**

$$\begin{aligned} \text{Reliability} &= (Corr(s, F_1))^2 = \left( \frac{Cov(z_1 + z_2, F_1)}{\sqrt{Var(s)} \cdot 1} \right)^2 \\ &= \left( \frac{1}{\sqrt{2}} Cov(\lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3 + e_1 + \lambda_4 F_1 + \lambda_5 F_2 + \lambda_6 F_3 + e_2, F_1) \right)^2 \\ &= \boxed{\frac{1}{2} (\lambda_1 + \lambda_4)^2} \end{aligned}$$

2. For the R part of the assignment (last question), you fit a 4-factor model to the `statclass` data.

- (a) (2 points) What is the estimated correlation between Factor 2 and mark on Computer Assignment 5? The answer is a number from your printout. Write the answer in the space below. Circle and label the number on your printout.

0.788

- (b) (2 points) What estimated proportion of the variance of the Midterm is explained by the common factors? The answer is a number from your printout. Write the answer in the space below. Circle and label the number on your printout.

0.21

Please attach your printout to the quiz paper. The printout should show your *complete* R input and output. Make sure your name and student number appear on the printout.

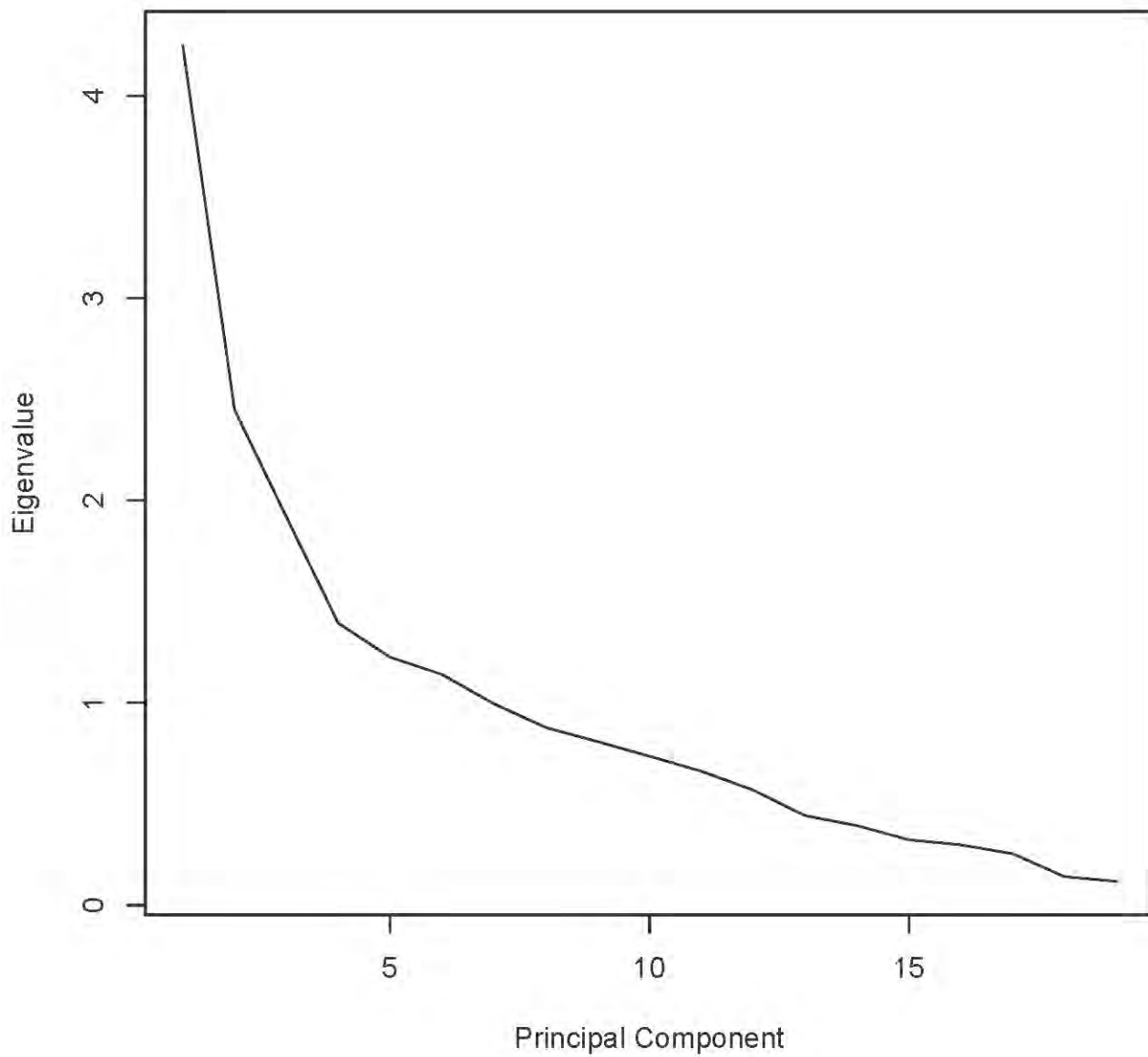
## Assignment 9, Question 6

```
> # EFA of the statclass data
>
> rm(list=ls())
> statclass =
read.table("https://www.utstat.toronto.edu/brunner/openSEM/data/statclass.data.txt",
header=TRUE)
> # dim(statclass); head(statclass)
> dat = statclass[,3:21]; dim(dat); head(dat)
[1] 58 19
  Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 C1 C2 C3 C4 C5 C6 C7 C8 C9 MT Final
1  9  1  7  8  4  3  5  2  6 10 10 10  5  0  0  0  0 55   43
2 10 10  5  9 10  8  6  8 10 10  8  9  9  9  9 10 10 66   79
3 10 10  5 10 10 10  9  8 10 10 10 10 10 10  9 10 10 94   67
4 10 10  8  9 10  7 10  9 10 10 10  9 10 10  9 10 10 81   65
5 10  6  7  9  8  8  5  7 10  9 10  9  5  6  4  8 10 57   52
6 10  9  5  8  9  8  5  6  8  7  5  6 10  6  5  9  9 77   64

> # (a) How many factors?
>
> pc = prcomp(dat, scale = T)
> Eigenvalue = pc$sdev^2 ; Eigenvalue
[1] 4.2527781 2.4483506 1.9187278 1.3940113 1.2256366 1.1402548 0.9955204 0.8785114
[9] 0.8083985 0.7363114 0.6619296 0.5688996 0.4426665 0.3932991 0.3232235 0.2984006
[17] 0.2544390 0.1409714 0.1176700
> # Could be 6
```

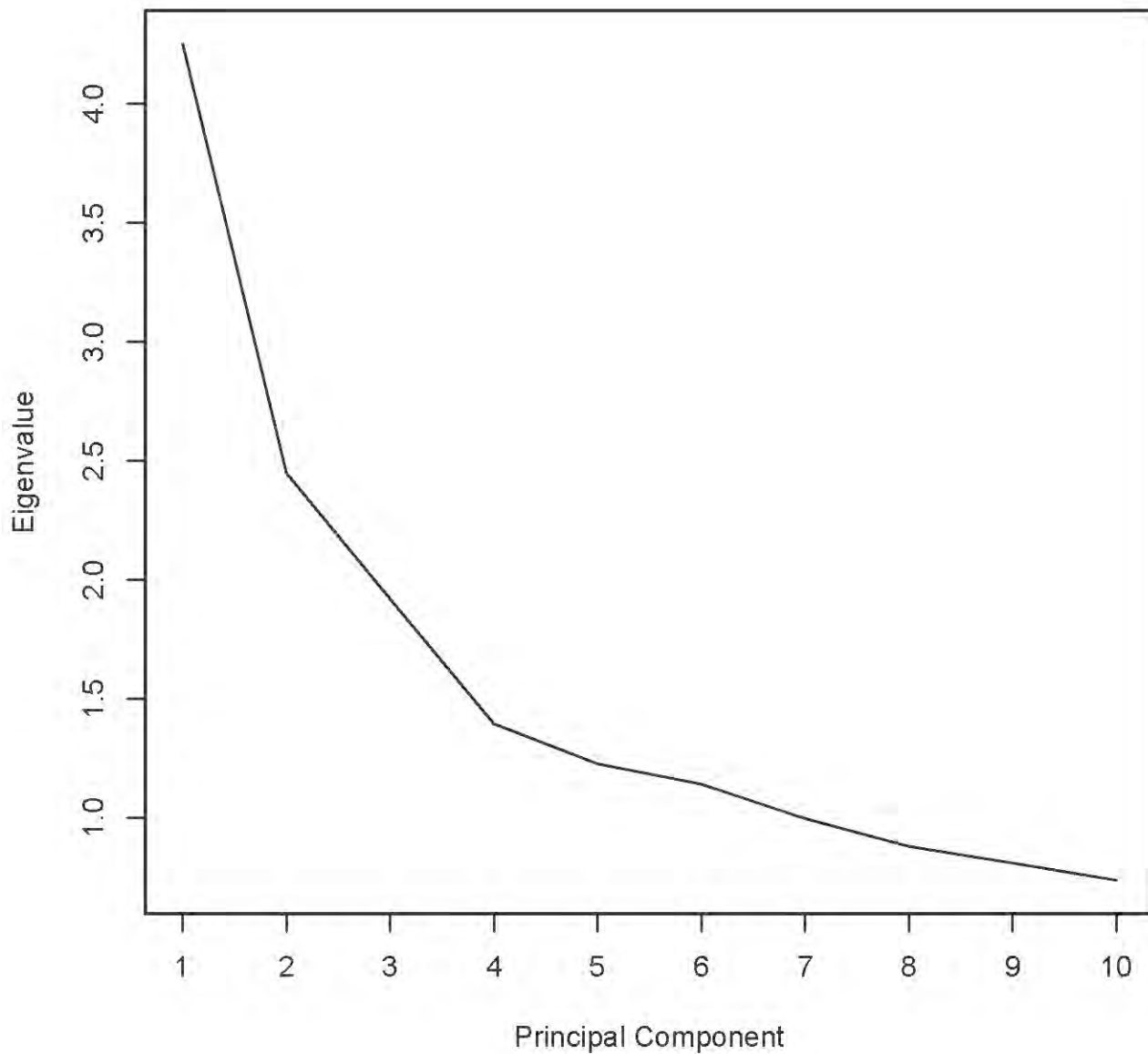
```
> # Scree plot  
> plot(1:19,Eigenvalue, xlab = "Principal Component", type = "l",  
+      main = "Scree Plot of the Stat Class Data")
```

**Scree Plot of the Stat Class Data**



```
> # Zoom in on the first ten
> plot(1:10,Eigenvalue[1:10], type = "l",
+      xaxp = c(1,10,9), # Tick marks on x axis
+      xlab = "Principal Component", ylab = "Eigenvalue",
+      main = "Scree Plot of the Stat Class Data")
```

**Scree Plot of the Stat Class Data**



```

> # help(factanal)
> fa4 = factanal(dat,factors=4) # rotation="varimax" is the default
> fa4

```

```

Call:
factanal(x = dat, factors = 4)

```

Uniquenesses:

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	C1	C2	C3	C4	C5	C6
0.918	0.527	0.855	0.121	0.795	0.534	0.525	0.805	0.666	0.523	0.005	0.823	0.324	0.284
C7	C8	C9	MT	Final									
0.542	0.374	0.945	0.790	0.432									

Loadings:

	Factor1	Factor2	Factor3	Factor4
Q1	0.189	-0.113		-0.177
Q2	0.654		0.208	
Q3	0.169		0.334	
Q4		0.203	0.909	
Q5	0.120	0.406	-0.156	
Q6	0.594	0.237	0.199	0.131
Q7	0.482	0.179	0.193	0.415
Q8	0.431			
C1	0.345	0.151	0.143	0.414
C2			0.666	0.178
C3	-0.170	-0.111	0.535	0.817
C4		-0.140	0.385	
C5	-0.119	0.788	0.196	
C6		0.724	0.159	0.398
C7	0.347	0.389		0.429
C8	0.496	0.571		0.227
C9	0.105	0.203		
MT	0.401	0.211		
Final	0.740		-0.114	

	Factor1	Factor2	Factor3	Factor4
SS loadings	2.536	2.083	2.069	1.523
Proportion Var	0.133	0.110	0.109	0.080
Cumulative Var	0.133	0.243	0.352	0.432

Test of the hypothesis that 4 factors are sufficient.  
The chi square statistic is 102.1 on 101 degrees of freedom.  
The p-value is 0.451

```
> L4 = fa4$loadings; print(L4,cutoff=0.3, sort="True")
```

Loadings:

	Factor1	Factor2	Factor3	Factor4
Q2	0.654			
Q6	0.594			
Final	0.740			
C5		0.788		
C6		0.724		0.398
C8	0.496	0.571		
Q4			0.909	
C2			0.666	
C3			0.535	0.817
Q1				
Q3			0.334	
Q5		0.406		
Q7	0.482			0.415
Q8	0.431			
C1	0.345			0.414
C4			0.385	
C7	0.347	0.389		0.429
C9				
MT	0.401			

	Factor1	Factor2	Factor3	Factor4
SS loadings	2.536	2.083	2.069	1.523
Proportion Var	0.133	0.110	0.109	0.080
Cumulative Var	0.133	0.243	0.352	0.432

```

>
> # Try 3 factors
> fa3 = factanal(dat,factors=3)
> print(fa3,cutoff=0.3, sort="True")

```

```

Call:
factanal(x = dat, factors = 3)

```

```

Uniquenesses:
  Q1    Q2    Q3    Q4    Q5    Q6    Q7    Q8    C1    C2    C3    C4    C5    C6
0.943 0.554 0.853 0.472 0.808 0.516 0.575 0.777 0.751 0.386 0.568 0.840 0.637 0.090
  C7    C8    C9    MT Final
0.577 0.381 0.942 0.796 0.565

```

```

Loadings:
      Factor1 Factor2 Factor3
Q2      0.654
Q6      0.662
Q7      0.519
C8      0.591    0.513
Final   0.642
C5              0.584
C6              0.916
C7      0.393    0.519
Q4              0.718
C2              0.777
C3              0.626
Q1
Q3
Q5              0.353
Q8      0.467
C1      0.423
C4              0.398
C9
MT      0.424

```

```

      Factor1 Factor2 Factor3
SS loadings      2.795    2.131    2.044
Proportion Var   0.147    0.112    0.108
Cumulative Var   0.147    0.259    0.367

```

Test of the hypothesis that 3 factors are sufficient.  
The chi square statistic is 137.51 on 117 degrees of freedom.  
The p-value is 0.0946



```
>
> # Try 2 factors
> fa2 = factanal(dat,factors=2)
> print(fa2,cutoff=0.3, sort="True")
```

```
Call:
factanal(x = dat, factors = 2)
```

```
Uniquenesses:
  Q1    Q2    Q3    Q4    Q5    Q6    Q7    Q8    C1    C2    C3    C4    C5    C6
1.000 0.761 0.924 0.479 0.842 0.594 0.619 0.826 0.734 0.417 0.564 0.842 0.830 0.571
  C7    C8    C9    MT Final
0.638 0.329 0.926 0.833 0.781
```

```
Loadings:
      Factor1 Factor2
Q6      0.606
Q7      0.509    0.350
C6      0.548    0.359
C7      0.589
C8      0.818
Q4              0.722
C2              0.764
C3              0.660
Q1
Q2      0.467
Q3
Q5      0.380
Q8      0.415
C1      0.467
C4              0.386
C5      0.331
C9
MT      0.406
Final   0.449
```

```

      Factor1 Factor2
SS loadings      3.257  2.234
Proportion Var   0.171  0.118
Cumulative Var   0.171  0.289
```

```
Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 187.03 on 134 degrees of freedom.
The p-value is 0.0017
```

```
> # (d) I guess I will settle for 4.
```

```
> # (e) Get communalities 2 different ways, and use cbind to display them
> # side by side.
> ls(fa4)
```

```
[1] "call"          "converged"     "correlation"   "criteria"      "dof"
[6] "factors"       "loadings"      "method"        "n.obs"         "PVAL"
[11] "rotmat"        "STATISTIC"     "uniquenesses"
```

```
> commun = diag(L4 %*% t(L4) )
> cbind(commun, 1-fa4$uniqueesses)
```

```

      commun
Q1    0.08240676 0.08240312
Q2    0.47306931 0.47307067
Q3    0.14456256 0.14456023
Q4    0.87859474 0.87859464
Q5    0.20510501 0.20511097
Q6    0.46588629 0.46588706
Q7    0.47453865 0.47453844
Q8    0.19479587 0.19479909
C1    0.33373977 0.33374219
C2    0.47704598 0.47704520
C3    0.99500540 0.99500000
C4    0.17678212 0.17678212
C5    0.67617032 0.67616986
C6    0.71553982 0.71553934
C7    0.45838861 0.45838767
C8    0.62609996 0.62610066
C9    0.05544739 0.05543382
MT    0.20960873 0.20961304
Final 0.56785720 0.56785778
```

Q2b

```
> # (f, g) See output above.
> # (h)
>
> print(L4,cutoff=0.3, sort="True")
```

```

Loadings:
      Factor1 Factor2 Factor3 Factor4
Q2      0.654
Q6      0.594
Final   0.740
C5      0.788
C6      0.724
C8      0.496 0.571
Q4      0.909
C2      0.666
C3      0.535 0.817
Q1      0.334
Q3      0.334
Q5      0.406
Q7      0.482
Q8      0.431
C1      0.345
C4      0.385 0.817
C7      0.347 0.389
C9      0.414
MT      0.429
```

Q2a

```

      Factor1 Factor2 Factor3 Factor4
SS loadings 2.536 2.083 2.069 1.523
Proportion Var 0.133 0.110 0.109 0.080
Cumulative Var 0.133 0.243 0.352 0.432
```

```
> # (i) Estimated reliability of the final exam as a measure of Factor 1
> 0.740^2
[1] 0.5476
```