STATISTICAL PROGRAMMING FOR BUSINESS ANALYTICS

Assignment No.8

TANAY BHALERAO

U47707491

APRIL 10, 2015

UNIVERSITY OF SOUTH FLORIDA
MANAGEMENT INFORMATION SYSTEMS

Homework Chapter 10 and 11

1. An instructor for an introductory SAS programming class has 25 students. Each week, he makes a list of 8 problems pertaining to the topics covered in class. He wants to assign 4 of the 8 questions to be turned in by each student, but he also wants to select those 4 problems at random for each student to help curtail cheating. Write a SAS program which uses PROC PLAN to randomly choose 4 of 8 homework problems to be done by each student. Make sure that your program does not assign the same 4 problems to every student.

```
PROC PLAN SEED =0;
FACTORS STUDENT=25 ORDERED QUES=4 OF 8 /NOPRINT;
OUTPUT OUT= RANDOMQ;
RUN;
PROC TRANSPOSE DATA= RANDOMQ OUT= RANDOMQ;
VAR QUES;
BY STUDENT;
TITLE "RANDOM QUESTIONS";
RUN;
PROC PRINT DATA = RANDOMQ;
      RUN;
```

RANDOM QUESTIONS

Obs	STUDENT	_NAME_	COL1	COL2	COL3	COL4
1	1	QUES	8	2	7	3
2	2	QUES	1	6	3	7
3	3	QUES	8	7	4	2
4	4	QUES	6	7	5	3
5	5	QUES	5	2	8	4
6	6	QUES	7	2	4	3
7	7	QUES	2	3	7	4
8	8	QUES	6	7	1	5
9	9	QUES	5	2	7	1
10	10	QUES	1	2	3	6
11	11	QUES	4	7	1	8
12	12	QUES	6	7	2	5
13	13	QUES	6	3	4	5
14	14	QUES	6	7	8	4
15	15	QUES	1	6	3	2
16	16	QUES	7	2	3	4
17	17	QUES	2	3	7	6
18	18	QUES	6	4	7	1
19	19	QUES	6	4	3	8
20	20	QUES	7	8	1	3
21	21	QUES	1	5	8	3
22	22	QUES	2	7	8	1
23	23	QUES	8	5	3	6
24	24	QUES	4	1	6	5
25	25	QUES	1	8	2	7

2. What does a *t* distribution with one degree of freedom look like? Create 1000 observations of the variables Z1 and Z2, where both Z1 and Z2 represent independent random observations from a normal distribution with mean 0 and variance 1. Next, calculate T=Z1/Z2, and use the PROC UNIVARIATE with the PLOT option to examine the distribution of T. (A *t* distribution with one degree of freedom is also known as a Cauchy distribution. It is symmetric about zero, and it is likely to have both positive and negative outlying values. It is very unlikely that you will reject a null hypothesis with a *t* test using only one degree of freedom.)

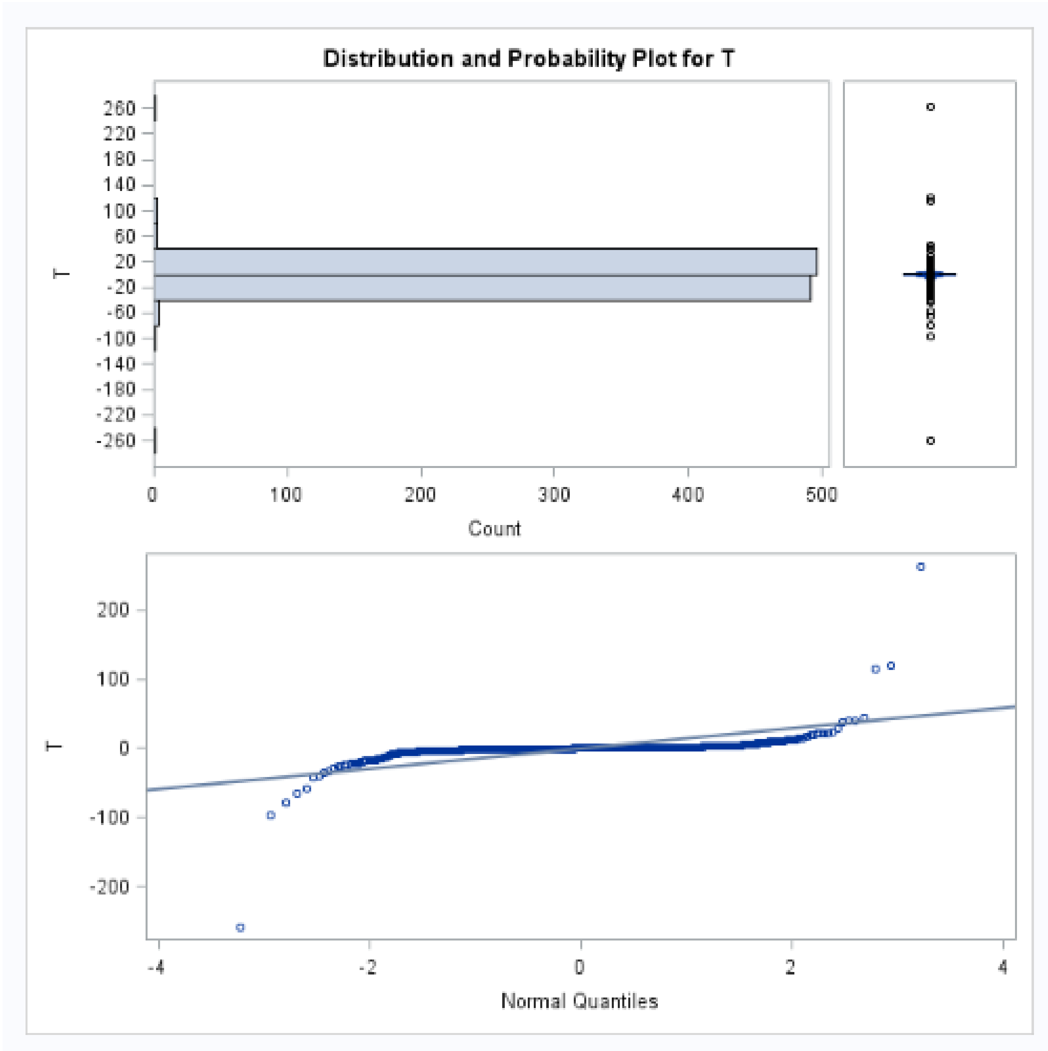
```
DATA OBS;
DO I= 1 TO 1000;
Z1= RANNOR(0);
Z2= RANNOR(100);
T= Z1/Z2;
OUTPUT;
END;
PROC UNIVARIATE DATA= OBS PLOT;
VAR T;
TITLE "CAUCHY DISTRIBUTION";
RUN;
```

CAUCHY DISTRIBUTION

The UNIVARIATE Procedure
Variable: T

Moments			
N	1000	Sum Weights	1000
Mean	-0.1050376	Sum Observations	-105.03755
Std Deviation	14.8951624	Variance	221.865863
Skewness	0.6774243	Kurtosis	200.461113
Uncorrected SS	221655.03	Corrected SS	221643.997
Coeff Variation	-14180.797	Std Error Mean	0.47102639

Basic Statistical Measures			
Location		Variability	
Mean	-0.10504	Std Deviation	14.89516
Median	0.00592	Variance	221.86586
Mode	.	Range	523.26098
		Interquartile Range	2.06403



As observed from the above plot the t-distribution is symmetric around zero. And also it has both positive and negative outlying points. Hence it truly is difficult to reject the null hypothesis

3. In the Florida Lotto, six balls are randomly chosen from a hopper of balls marked with the integers 1 through 49. Money is awarded to players who match 3, 4, 5, or 6 numbers. Suppose that you know someone who always bets on the numbers 3, 7, 8, 11, 19, and 29, and you would like to calculate the approximate odds of matching at least 3 of those numbers.Use PROC PLAN to draw 6 "balls" out of 49; do this 1000 times. Use NOPRINT after a slash in the FACTORS statement to prevent SAS from printing out many pages of output (for example, FACTORS A=1000 ORDERED B=6 OF $(/NOPRINT;). Then, in each draw, calculate the number of matches with 3, 7, 8, 11, 19, and 29. There could be no matches, 1 match, 2 matches, etc. up to 6 matches. (For this step, PROC TRANSPOSE and arrays will be helpful.) Finally, obtain a frequency table which shows the number of lotteries with no matches, 1 match, 2 matches, etc. Don't be surprised if you don't "win" anything; the probability of matching at least 3 numbers is only 1.9%, according to the Florida Lottery.

```
PROC PLAN SEED=99999;
FACTORS DRAW= 1000 ORDERED BALL=6 OF 49/NOPRINT;
OUTPUT OUT=LOTTO1;
RUN;
QUIT;
PROC TRANSPOSE DATA=LOTTO1 OUT=LOTTO2;
VAR BALL;
BY DRAW;
RUN;
DATA LOTTO2;
SET LOTTO2;
ARRAY BALLS(6) COL1-COL6;
MATCHES = 0;
DO i = 1 TO 6;
IF BALLS[i] = 3 THEN MATCHES = MATCHES + 1;
IF BALLS[i] = 7 THEN MATCHES = MATCHES + 1;
IF BALLS[i] = 8 THEN MATCHES = MATCHES + 1;
IF BALLS[i] = 11 THEN MATCHES = MATCHES + 1;
IF BALLS[i] = 19 THEN MATCHES = MATCHES + 1;
IF BALLS[i] = 29 THEN MATCHES = MATCHES + 1;
END;
RUN;
PROC FREQ DATA=LOTTO2;
TABLES MATCHES;
TITLE "LOTTO CHANCES";
RUN;
```

LOTTO CHANCES

The FREQ Procedure

MATCHES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	430	43.00	430	43.00
1	408	40.80	838	83.80
2	147	14.70	985	98.50
3	14	1.40	999	99.90
4	1	0.10	1000	100.00

4. Suppose that you want to use SAS to help you learn some strategies for playing bridge. There are 52 cards in a deck. In a game of bridge, four players each receive 13 cards. Write a SAS program which randomly deals 13 cards to each of four players. Remember that the cards are dealt without replacement, so make sure that your program cannot possibly deal the same card to two players at once. Refer to the four players as North, East, South, and West.

```
DATA CARD;
FORMAT CARD_VALUE $5. CARD_SUITE $10.;
DO CARD_SUITE = 'HEART','SPADE', 'DIAMOND','CLUB';
DO
CARD_VALUE='ACE','2','3','4','5','6','7','8','9','10','JACK','QUEEN','KING';
DO RANDOM=RANNOR(RANUNI(2356));
OUTPUT;
END;
```

```
END;
END;
RUN;
PROC SORT DATA =CARD;
BY RANDOM;
RUN;

DATA GAME_BRIDGE;
SET CARD;
IF _V_<=13 THEN PLAYER= 'NORTH';
ELSE IF 14<=_V_<=26 THEN PLAYER='EAST';
ELSE IF 27<=_V_<=39 THEN PLAYER= 'SOUTH';
ELSE IF _V_>=40 THEN PLAYER='WEST';
RUN;

PROC PRINT DATA = GAME_BRIDGE;
VAR PLAYER RANDOM CARD_VALUE CARD_SUITE;
TITLE "BRIDGE GAME- RANDOM DISTRIBUTION OF CARDS";
RUN;
```

BRIDGE GAME- RANDOM DISTRIBUTION OF CARDS

Obs	PLAYER	RANDOM	CARD_VALUE	CARD_SUITE
1	NORTH	-2.87104	2	SPADE
2	NORTH	-2.25912	QUEEN	SPADE
3	NORTH	-2.21085	JACK	CLUB
4	NORTH	-1.58944	6	HEART
5	NORTH	-1.36761	10	DIAMOND
6	NORTH	-1.22656	KING	SPADE
7	NORTH	-1.15505	2	HEART
8	NORTH	-0.92143	7	HEART
9	NORTH	-0.90088	JACK	DIAMOND
10	NORTH	-0.90027	5	DIAMOND
11	NORTH	-0.88705	JACK	SPADE
12	NORTH	-0.88663	7	CLUB
13	NORTH	-0.79546	5	CLUB
14	NORTH	-0.79299	6	SPADE
15	NORTH	-0.75081	6	CLUB
16	NORTH	-0.74874	3	SPADE
17	NORTH	-0.69519	5	SPADE
18	NORTH	-0.67659	2	CLUB
19	NORTH	-0.66419	10	CLUB
20	NORTH	-0.65531	ACE	SPADE
21	NORTH	-0.64997	7	SPADE
22	NORTH	-0.38208	10	SPADE
23	NORTH	-0.35685	3	HEART
24	NORTH	-0.26148	4	CLUB
25	NORTH	-0.25179	8	SPADE
26	NORTH	-0.08882	8	HEART
27	NORTH	-0.07458	QUEEN	HEART

5. Suppose that boys and girls are randomly paired together in a dance class for fifth graders. Assume that the girls' heights are normally distributed with mean 137.8 cm and standard deviation 6.8 cm, while the boys' heights are normally distributed with mean 137.1 cm and standard deviation 6.1 cm. In what proportion of couples will the girl be taller than the buy? Simulate 1000 pairs of boys and girls and find the proportion of those in which the girl is taller. The theoretical percentage of couples with taller girls is 53%.

```
PROC FORMAT ;
VALUE GIRL_TALL 1='YES' 2='NO';
RUN;
DATA COUPLES;
DO i =1 TO 1000;
GIRL_HEIGHT = 137.8 + 6.8 * RANNOR(2000);
BOY_HEIGHT = 137.1 + 6.1 *RANNOR(2000);
IF (GIRL_HEIGHT > BOY_HEIGHT) THEN HT = 1;
ELSE IF (BOY_HEIGHT >  GIRL_HEIGHT )  THEN HT =2;
OUTPUT;
END;
RUN;
PROC FREQ DATA=COUPLES;
TABLES HT;
FORMAT HT GIRL_TALL.;
TITLE "DANCE COUPLES-GIRL TALLER THAN BOY STATS";
RUN;
```

DANCE COUPLES-GIRL TALLER THAN BOY STATS

The FREQ Procedure

HT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
YES	556	55.60	556	55.60
NO	444	44.40	1000	100.00

6. Chapter 10: 10.2, 10.4 and Chapter 11: 11.2 and 11.4

10.2

```
DATA PRINCIPAL;
      DO SUBJ = 1 TO 200;
            X1 = ROUND(RANNOR(123)*50 + 500);
            X2 = ROUND(RANNOR(123)*50 + 100 + .8*X1);
            X3 = ROUND(RANNOR(123)*50 + 100 + X1 - .5*X2);
            X4 = ROUND(RANNOR(123)*50 + .3*X1 + .3*X2 + .3*X3);
      OUTPUT;
END;
RUN;
```

```
PROC FACTOR DATA = PRINCIPAL NFACTORS = 2 ROTATE = VARIMAX PLOT SCREE;
VAR X1-X4;
TITLE "FACTOR ANALYSIS";
RUN;
```

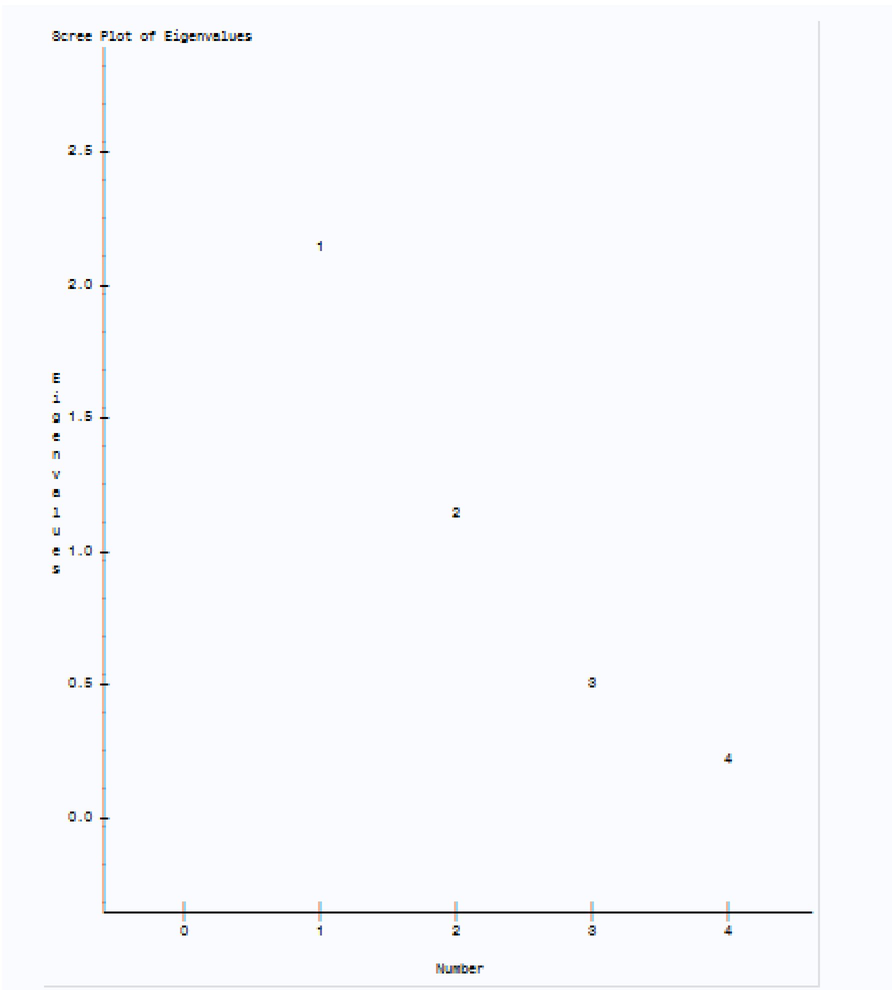
FACTOR ANALYSIS

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 4 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.12973163	0.97682870	0.5324	0.5324
2	1.15290294	0.62644662	0.2882	0.8207
3	0.52645632	0.33554720	0.1316	0.9523
4	0.19090911		0.0477	1.0000

4 factors will be retained by the NFACTOR criterion.

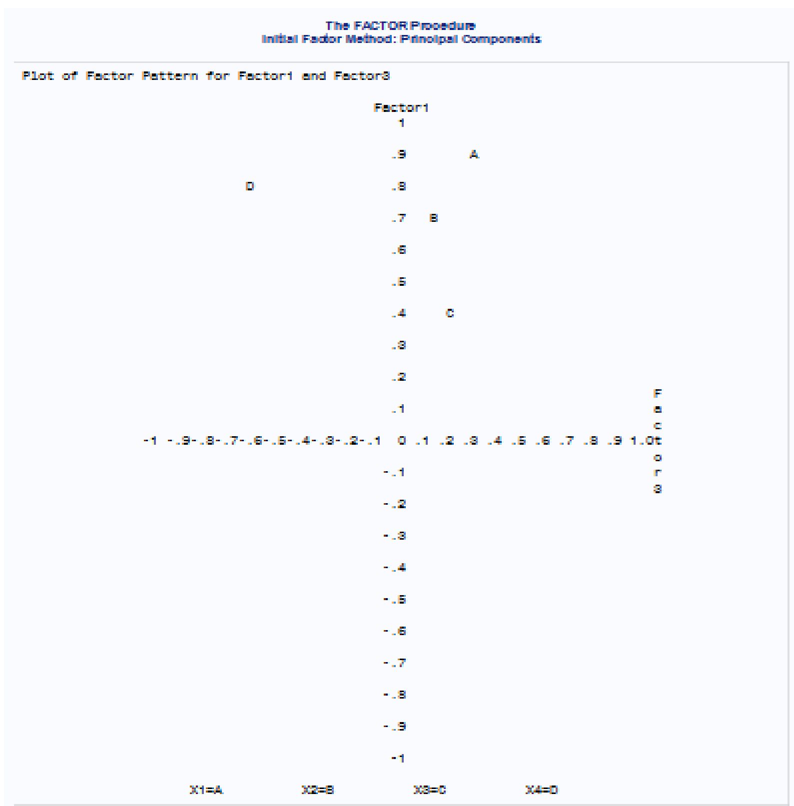


The scree plot graphs the eigenvalue against the factor number. Looking at the plot we will choose the factors 1 and 2. From the 3rd factor the line starts to flatten out. So the factors after this will account less to the total variance.

Factor Pattern				
	Factor1	Factor2	Factor3	Factor4
X1	0.90544	-0.01587	0.31324	-0.28600
X2	0.72071	-0.62142	0.14827	0.26913
X3	0.41579	0.86577	0.20298	0.19066
X4	0.78587	0.13011	-0.60428	-0.01818

Variance Explained by Each Factor			
Factor1	Factor2	Factor3	Factor4
2.1297316	1.1529029	0.5264563	0.1909091

Final Communality Estimates: Total = 4.000000			
X1	X2	X3	X4
1.0000000	1.0000000	1.0000000	1.0000000



Varimax rotatation is an orthogonal rotation which impose restriction on the factors after the initial extraction of factors that they cannot be correlated.

FACTOR ANALYSIS

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2	3	4
1	0.52029	0.30225	0.55377	0.57556
2	-0.56401	0.81738	0.11381	-0.02890
3	0.23460	0.29160	-0.82324	0.42687
4	0.59679	0.39434	-0.05161	-0.69690

Rotated Factor Pattern

	Factor1	Factor2	Factor3	Factor4
X1	0.38286	0.23926	0.25649	0.85463
X2	0.92086	-0.14073	0.19243	0.30850
X3	-0.11057	0.96771	0.15185	0.16807
X4	0.18289	0.16051	0.94840	0.20328

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4
1.0402418	1.0392811	1.0253418	0.8951352

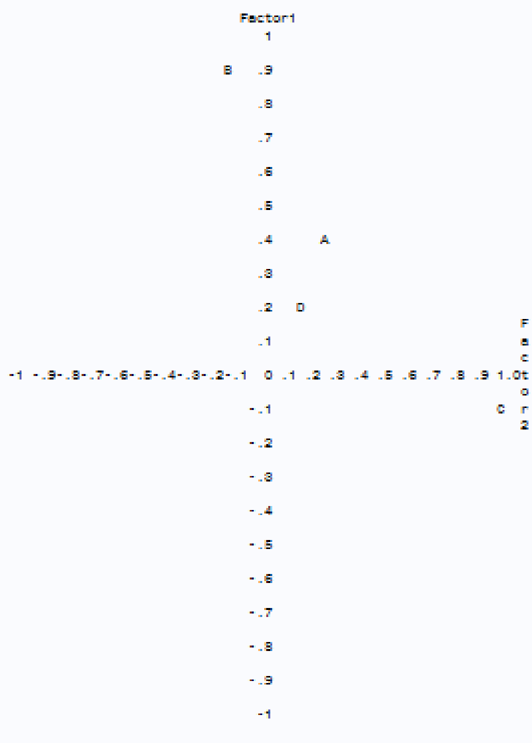
Final Communality Estimates: Total = 4.000000

X1	X2	X3	X4
1.0000000	1.0000000	1.0000000	1.0000000

FACTOR ANALYSIS

The FACTOR Procedure
Rotation Method: Varimax

Plot of Factor Pattern for Factor1 and Factor2



10.4

```
PROC FACTOR DATA = PRINCIPAL NFACTORS=2 ROTATE=PROMAX PLOT OUT=FACT_SCORE;
TITLE "NON-OBLIQUE ROTATION";
VAR x1--x4;
RUN;
PROC PRINT DATA = FACT_SCORE (OBS = 10);
RUN;
```

NON-OBLIQUE ROTATION

The FACTOR Procedure

Input Data Type	Raw Data
Number of Records Read	200
Number of Records Used	200
N for Significance Tests	200

NON-OBLIQUE ROTATION

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 4 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.12973163	0.97682870	0.5324	0.5324
2	1.15290294	0.62644662	0.2882	0.8207
3	0.52645632	0.33554720	0.1316	0.9523
4	0.19090911		0.0477	1.0000

Variance Explained by Each Factor	
Factor1	Factor2
2.1297316	1.1529029

Final Communality Estimates: Total = 3.282635			
X1	X2	X3	X4
0.82008254	0.90558427	0.92244850	0.63451926

NON-OBLIQUE ROTATION

The FACTOR Procedure
Rotation Method: Promax (power = 3)

Target Matrix for Procrustean Transformation		
	Factor1	Factor2
X1	0.99236	0.04089
X2	1.00000	-0.03856
X3	0.00060	1.00000
X4	0.76649	0.13187

Procrustean Transformation Matrix		
	1	2
1	1.14094936	-0.051234
2	-0.059132	0.82272224

Normalized Oblique Transformation Matrix		
	1	2
1	0.91911	0.30310
2	-0.41029	0.95981

Inter-Factor Correlations		
	Factor1	Factor2
Factor1	1.00000	0.11373
Factor2	0.11373	1.00000

Rotated Factor Pattern (Standardized Regression Coefficients)		
	Factor1	Factor2
X1	0.83872	0.25921
X2	0.91738	-0.37799
X3	0.02694	0.95700
X4	0.66892	0.36308

Reference Axis Correlations		
	Factor1	Factor2
Factor1	1.00000	-0.11373
Factor2	-0.11373	1.00000

Reference Structure (Semipartial Correlations)		
	Factor1	Factor2
X1	0.83328	0.25752
X2	0.91142	-0.37554
X3	0.02677	0.95080
X4	0.66458	0.36073

Variance Explained by Each Factor Eliminating Other Factors	
Factor1	Factor2
1.9674203	1.2414837

Factor Structure (Correlations)		
	Factor1	Factor2
X1	0.88820	0.35459
X2	0.87439	-0.27366
X3	0.13578	0.96007
X4	0.71021	0.43916

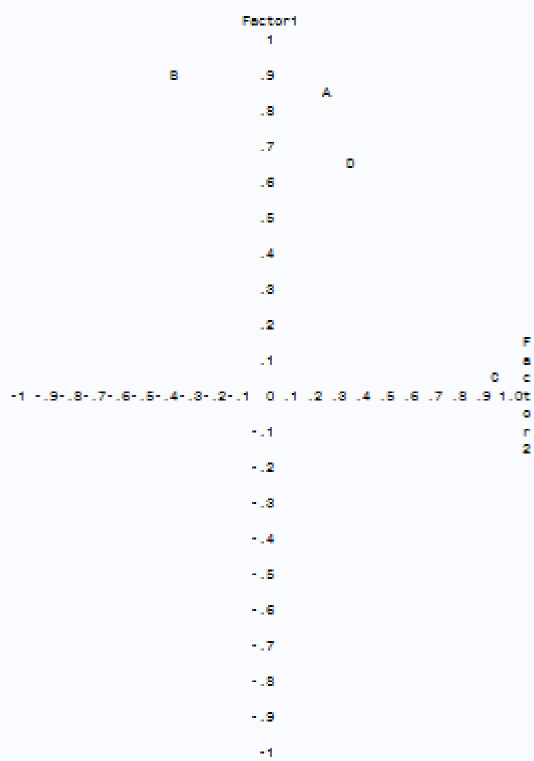
Variance Explained by Each Factor Ignoring Other Factors	
Factor1	Factor2
2.0411509	1.3152143

Final Communality Estimates: Total = 3.282635			
X1	X2	X3	X4
0.82008254	0.90558427	0.92244850	0.63451926

NON-OBLIQUE ROTATION

The FACTOR Procedure
Rotation Method: Promax (power = 3)

Plot of Factor Pattern for Factor1 and Factor2
Factor Axis Correlation = 0.1137 Angle = 83.4698



X1=A X2=B X3=C X4=D

NON-OBLIQUE ROTATION

Obs	SUBJ	X1	X2	X3	X4	Factor1	Factor2
1	1	484	564	315	381	0.23615	-0.87875
2	2	539	498	389	389	0.22281	0.66694
3	3	536	592	346	422	1.09974	-0.26346
4	4	552	449	396	399	0.02706	1.10985
5	5	489	438	454	391	-0.68407	1.67869
6	6	537	566	390	400	0.76492	0.35783
7	7	495	510	326	365	-0.16134	-0.47906
8	8	521	614	322	424	1.15865	-0.73787
9	9	551	517	352	347	0.25733	-0.06547
10	10	517	405	437	425	-0.48218	1.87921

Promax allows factors to correlate with one another.

```
DATA CORR_SCORE;
SET FACT_SCORE;
FACT_A= MEAN(OF X1,X2,X4);
FACT_B=X3;
RUN;
PROC CORR DATA = CORR_SCORE NOSIMPLE;
VAR FACTOR1 FACTOR2;
WITH FACT_A FACT_B;
RUN;
```

NON-OBLIQUE ROTATION

The CORR Procedure

2 With Variables:	FACT_A FACT_B
2 Variables:	Factor1 Factor2

Pearson Correlation Coefficients, N = 200
Prob > |r| under H0: Rho=0

	Factor1	Factor2
FACT_A	0.99565 <.0001	0.16932 0.0165
FACT_B	0.13578 0.0552	0.96007 <.0001

11.2

```
DATA TEST;
ARRAY KEY[10] $ 1 KEY1-KEY10; **key solutions;
ARRAY SOL[10] $ 1 SOL1-SOL10; **student's solution;
ARRAY S[10] 3 S1-S10;

RETAIN KEY1-KEY10;

IF _N_=1 THEN INPUT (KEY1-KEY10)($1.);
INPUT @1 ID 1-3
@5 (SOL1-SOL10) ($1.);

DO I = 1 TO 10;
S[I] = (SOL[I] EQ KEY[I]);
END;

RAW = SUM (OF S1-S10);
PERCENT = 100*RAW/10;
KEEP S1-S10 ID RAW PERCENT;
LABEL
ID = "ID"
RAW = "RAW_SCORE"
PERCENT = "PERCENTILE_SCORE";

DATALINES;
ABCDEABCDE
001 ABCDBECDBE
002 ABCDEABCDE
003 ABCDEABCDD
004 ABCEDABCCE
005 BBCDEBBCDE
006 CABEDACBED
007 DECAACEDAA
008 ABCDEBBBEE
009 DDDDDABCDE
010 ABECDABCDE
;

PROC PRINT DATA = TEST;
TITLE "SCORING TEST";
RUN;
PROC CORR DATA = TEST NOSIMPLE ALPHA;
VAR S1-S10;
RUN;
```

SCORING TEST

Obs	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	ID	RAW	PERCENT
1	1	1	1	1	0	0	0	0	0	1	1	5	50
2	1	1	1	1	1	1	1	1	1	1	2	10	100
3	1	1	1	1	1	1	1	1	1	0	3	9	90
4	1	1	1	0	0	1	1	1	0	1	4	7	70
5	0	1	1	1	1	0	1	1	1	1	5	8	80
6	0	0	0	0	0	1	0	0	0	0	6	1	10
7	0	0	1	0	0	0	0	0	0	0	7	1	10
8	1	1	1	1	1	0	1	0	0	1	8	7	70
9	0	0	0	1	0	1	1	1	1	1	9	6	60
10	1	1	0	0	0	1	1	1	1	1	10	7	70

The Point-Biserial Correlation Coefficient is a correlation measure of the strength of association between a continuous-level variable and a binary variable.

The CORR Procedure

10 Variables:	S1 S2 S3 S4 S5 S6 S7 S8 S9 S10

Cronbach Coefficient Alpha	
Variables	Alpha
Raw	0.804182
Standardized	0.805231

Cronbach Coefficient Alpha with Deleted Variable				
	Raw Variables		Standardized Variables	
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha
S1	0.467951	0.788321	0.477832	0.788435
S2	0.705117	0.761646	0.710814	0.761130
S3	0.173687	0.818717	0.177591	0.820728
S4	0.467951	0.788321	0.470471	0.789265
S5	0.571645	0.776097	0.574604	0.777339
S6	0.072855	0.831210	0.065535	0.831985
S7	0.803685	0.750000	0.799462	0.750202
S8	0.642990	0.767442	0.638393	0.769835
S9	0.543305	0.779367	0.536561	0.781742
S10	0.427669	0.792614	0.430979	0.793685

Pearson Correlation Coefficients, N = 10 Prob > |r| under H0: Rho=0										
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	1.00000	0.80178 0.0053	0.35635 0.3122	0.16667 0.6454	0.25000 0.4860	0.16667 0.6454	0.35635 0.3122	0.16667 0.6454	0.00000 1.0000	0.35635 0.3122
S2	0.80178 0.0053	1.00000	0.52381 0.1202	0.35635 0.3122	0.53452 0.1114	-0.08909 0.8067	0.52381 0.1202	0.35635 0.3122	0.21822 0.5447	0.52381 0.1202
S3	0.35635 0.3122	0.52381 0.1202	1.00000	0.35635 0.3122	0.53452 0.1114	-0.53452 0.1114	0.04762 0.8961	-0.08909 0.8067	-0.21822 0.5447	0.04762 0.8961
S4	0.16667 0.6454	0.35635 0.3122	0.35635 0.3122	1.00000	0.66667 0.0353	-0.25000 0.4860	0.35635 0.3122	0.16667 0.6454	0.40825 0.2415	0.35635 0.3122
S5	0.25000 0.4860	0.53452 0.1114	0.53452 0.1114	0.66667 0.0353	1.00000	-0.16667 0.6454	0.53452 0.1114	0.25000 0.4860	0.40825 0.2415	0.08909 0.8067
S6	0.16667 0.6454	-0.08909 0.8067	-0.53452 0.1114	-0.25000 0.4860	-0.16667 0.6454	1.00000	0.35635 0.3122	0.58333 0.0767	0.40825 0.2415	-0.08909 0.8067
S7	0.35635 0.3122	0.52381 0.1202	0.04762 0.8961	0.35635 0.3122	0.53452 0.1114	0.35635 0.3122	1.00000	0.80178 0.0053	0.65465 0.0400	0.52381 0.1202
S8	0.16667 0.6454	0.35635 0.3122	-0.08909 0.8067	0.16667 0.6454	0.25000 0.4860	0.58333 0.0767	0.80178 0.0053	1.00000	0.81650 0.0039	0.35635 0.3122
S9	0.00000 1.0000	0.21822 0.5447	-0.21822 0.5447	0.40825 0.2415	0.40825 0.2415	0.40825 0.2415	0.65465 0.0400	0.81650 0.0039	1.00000	0.21822 0.5447
S10	0.35635 0.3122	0.52381 0.1202	0.04762 0.8961	0.35635 0.3122	0.08909 0.8067	-0.08909 0.8067	0.52381 0.1202	0.35635 0.3122	0.21822 0.5447	1.00000

Like all Correlation Coefficients (e.g. Pearson's r), the Point-Biserial Correlation Coefficient measures the strength of association of two variables(raw and percentile score) in a single measure ranging from -1 to +1, where -1 indicates a perfect negative association, +1 indicates a perfect positive association and 0 indicates no association at all. All correlation coefficients are interdependency measures that do not express a causal relationship.

11.4

```
DATA SUICIDAL;
INPUT RATER_1 $ RATER_2 $ @@;
DATALINES;
S S  N N  S S  N N  S N  S S  N S  N N  N N  S N
S S  N N  N N  S S  N S  S S  N N  S S  S S
;

PROC PRINT DATA=SUICIDAL;
TITLE "KAPPA-PSYCHIATRIST STUDY";
RUN;

PROC FREQ DATA=SUICIDAL;
TABLE RATER_1*RATER_2/NOCUM NOPERCENT KAPPA;
RUN;
```

KAPPA-PSYCHIATRIST STUDY

Obs	RATER_1	RATER_2
1	S	S
2	N	N
3	S	S
4	N	N
5	S	N
6	S	S
7	N	S
8	N	N
9	N	N
10	S	N
11	S	S
12	N	N
13	N	N
14	S	S
15	N	S
16	S	S
17	N	N
18	S	S
19	S	S

KAPPA-PSYCHIATRIST STUDY

The FREQ Procedure

Frequency
Row Pct
Col Pct

Table of RATER_1 by RATER_2			
	RATER_2		
RATER_1	N	S	Total
N	7 77.78 77.78	2 22.22 20.00	9
S	2 20.00 22.22	8 80.00 80.00	10
Total	9	10	19

Statistics for Table of RATER_1 by RATER_2

McNemar's Test	
Statistic (S)	0.0000
DF	1
Pr > S	1.0000

Simple Kappa Coefficient	
Kappa	0.5778
ASE	0.1875
95% Lower Conf Limit	0.2102
95% Upper Conf Limit	0.9453