KUCEL
고려대학교인지인간공학연구실
Cognitive Ergonomics Laboratory

응용 통계 및 실습

Assignment #2

Prob 1.

KUCEL
고려대학교인지인간공학연구실
Cognitive Ergonomics Laboratory

1)

DAY

BATCH	1	2	3	4	5	TOTAL
1	A-8	B-7	D-1	C-7	E-3	26
2	C-11	E-2	A-7	D-3	B-8	31
3	B-4	A-9	C-10	E-1	D-5	29
4	D-6	C-8	E-6	B-6	A-10	36
5	E-4	D-2	B-3	A-8	C-8	25
TOTAL	33	28	27	25	34	147

TREATMENT TOTALS

A	B	C	D	E
42	28	44	17	16

A) Fill in how the data is to be entered into SAS and write the SAS data step.
Next write the SAS code to obtain the output on the reverse side.

B) Fill in the missing quanities on the reverse side. A-Z

Prob 1. SAS code

KUCEL
고려대학교인지인간공학연구실
Cognitive Ergonomics Laboratory

```
DATA p1;
    input day batch catalyst $ time @@;
    cards;
1 1 a 8 1 2 c 11 1 3 b 4 1 4 d 6 1 5 e 4
2 1 b 7 2 2 e 2 2 3 a 9 2 4 c 8 2 5 d 2
3 1 d 1 3 2 a 7 3 3 c 10 3 4 e 6 3 5 b 3
4 1 c 7 4 2 d 3 4 3 e 1 4 4 b 6 4 5 a 8
5 1 e 3 5 2 b 8 5 3 d 5 5 4 a 10 5 5 c 8
;
run;
proc anova data = p1;
       class batch day catalyst;
       model time = batch day catalyst;
       means batch day catalyst/snk;
run;
```

Class
- Model 에 사용할 변수 설정 (classification variables)
- Model 문 이전에 사용

Model
- Dependent Variables = independent effects

Means
- Mean 비교.
- Classification variables의 효과

Prob 1. SAS results

KUCEL 고려대학교인지인간공학연구실 Cognitive Ergonomics Laboratory

Dependent Variable: time

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	169.1200000	14.0933333	4.51	0.0072
Error	12	37.5200000	3.1266667		
Corrected Total	24	206.6400000			

R-Square	Coeff Var	Root MSE	time Mean
0.818428	30.07208	1.768238	5.880000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
batch	4	15.4400000	3.8600000	1.23	0.3476
day	4	12.2400000	3.0600000	0.98	0.4550
catalyst	4	141.4400000	35.3600000	11.31	0.0005

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	3.126667

Number of Means	2	3	4	5
Critical Range	2.4365401	2.9835232	3.3200994	3.5645713

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	catalyst
A	8.800	5	c
A			
A	8.400	5	a
B	5.600	5	b
B			
B	3.400	5	d
B			
B	3.200	5	e

Prob 2.

2)

	Percentage of Cotton				
	15	20	25	30	35
	7	12	14	19	7
	7	17	18	25	10
	15	12	18	22	11
	11	18	19	19	15
	9	18	19	23	11
TOTALS	49	77	88	108	54

A) Fill in how the data is to be entered into SAS and write the SAS data step. Next write the SAS code to obtain the output on the reverse side.

B) Fill in the missing quantities on the reverse side. A-P

C) Perform a Kruskal-Wallis test on the data with the pairwise comparisons. Do we obtain different results?

Prob 2. SAS code (a), (b)

KUCEL
고려대학교인지인간공학연구실
Cognitive Ergonomics Laboratory

```
Data p2;
        input percent strength @@;
        cards;
15 7 15 7 15 15 15 11 15 9
20 12 20 17 20 12 20 18 20 18
25 14 25 18 25 18 25 19 25 19
30 19 30 25 30 22 30 19 30 23
35 7 35 10 35 11 35 15 35 11
;
run;
proc glm data = p2;
        class percent;
        model strength = percent;
        means percent/lsd;
run;

proc npar1way data = p2 wilcoxon;
        class percent;
        var strength;
run;
proc rank data = p2 out = rp2;
        var strength;
run;
proc glm data = rp2;
        class percent;
        model strength = percent;
```

Prob 2. (a), (b) result

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	475.7600000	118.9400000	14.76	<.0001
Error	20	161.2000000	8.0600000		
Corrected Total	24	636.9600000			

R-Square	Coeff Var	Root MSE	strength Mean
0.746923	18.87642	2.839014	15.04000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
percent	4	475.7600000	118.9400000	14.76	<.0001

Means with the same letter are not significantly different.			
t Grouping	Mean	N	percent
A	21.600	5	30
B	17.600	5	25
B			
B	15.400	5	20
C	10.800	5	35
C			
C	9.800	5	15

Prob 2. (c) result

percent	strength LSMEAN	LSMEAN Number
15	5.5000000	1
20	13.2000000	2
25	17.0000000	3
30	22.6000000	4
35	6.7000000	5

Least Squares Means for effect percent
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: strength

i/j	1	2	3	4	5
1		0.0236	0.0006	<.0001	0.9841
2	0.0236		0.4833	0.0046	0.0698
3	0.0006	0.4833		0.1465	0.0019
4	<.0001	0.0046	0.1465		<.0001
5	0.9841	0.0698	0.0019	<.0001	

Percent	Strength	Grouping
30 %	22.6	
25 %	17.0	
20 %	13.2	
35 %	6.70	
15 %	5.50	

Prob 3.

Building	1	2	3	4	5	6	7	8	9
New Coke	3	1	23	11	8	31	28	3	4
Coke Classic	8	9	27	27	29	44	16	8	7
Pepsi	9	6	18	20	10	26	21	0	9

Analyze the data. Threat Building as a block. Perform a multiple comparison on brand. Which is the best selling? Which is the worst selling? Is there any difference?

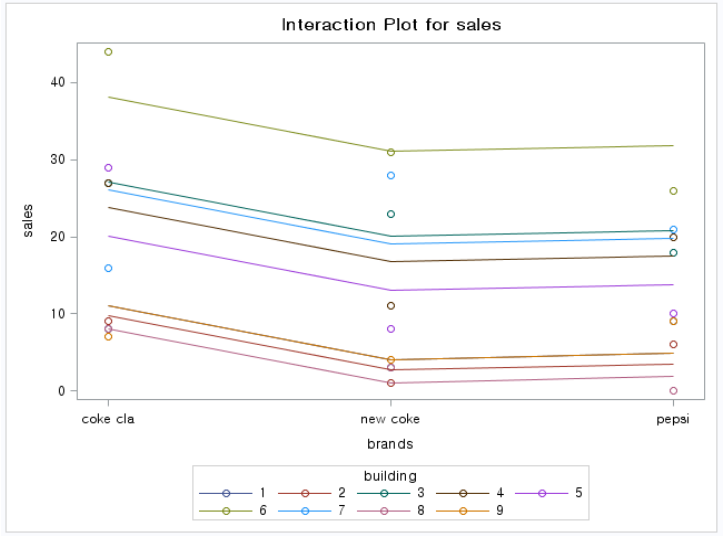
```
data p3;
Do brands ='new coke', 'coke classic', 'pepsi';
        Do building = 1 to 9;
                input sales @@;
                output;
        End; End;
Cards;
3 1 23 11 8 31 28 3 4 8 9 27 27 29 44 16 8 79 6 18 20 10 26 21 0 9
;
proc glm data = p3;
        class brands building;
        model sales = brands building;
        means brands/SNK LSD Tukey;
run;
```

Prob 3. result

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	2759.925926	275.992593	8.54	0.0001
Error	16	517.037037	32.314815		
Corrected Total	26	3276.962963			

R-Square	Coeff Var	Root MSE	sales Mean
0.842221	37.80407	5.684612	15.03704

Source	DF	Type I SS	Mean Square	F Value	Pr > F
brands	2	264.962963	132.481481	4.10	0.0365
building	8	2494.962963	311.870370	9.65	<.0001



Means with the same letter are not significantly different.				
Tukey Grouping		Mean	N	brands
	A	19.444	9	coke cla
	A			
B	A	13.222	9	pepsi
B				
B		12.444	9	new coke

Prob 4.

KUCEL
고려대학교인지인간공학연구실
Cognitive Ergonomics Laboratory

4) The Manager of fitness center wants to test whether three of her top athletes are of the same average performance level. The center has three identical exercise machines located at different places in the exercise hall. There are also three daily exercise times: morning, noontime, and evening. The manager assigns each of the athletes to a machine and to an exercise time according to the randomly chosen Latin square that follows. The manager measures the athletes' performance (number of pullups they can do in a specified time period.) The athletes are labeled A, B and C. Given the data in the Latin square, do all three athletes have the same average performance level?

	Machine		
Time	1	2	3
Morning	B=24	A=31	C=30
Noon	C=22	B=29	A=33
Evening	A=30	C=26	B=32

Prob 4. SAS code

KUCEL 고려대학교인지인간공학연구실 Cognitive Ergonomics Laboratory

```
data p4;
        input time $ machine athelete $ perform @@;
        cards;
morning 1 b 24 morning 2 a 31 morning 3 c 30
noon 1 c 22 noon 2 b 29 noon 3 a 33
evening 1 a 30 evening 2 c 26 evening 3 b 32
;
run;
proc glm data  = p4;
        class time machine athelete;
        model perform = time machine athelete;
        means time machine athelete/snk;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	106.0000000	17.6666667	5.68	0.1573
Error	2	6.2222222	3.1111111		
Corrected Total	8	112.2222222			

R-Square	Coeff Var	Root MSE	perform Mean
0.944554	6.176851	1.763834	28.55556

Source	DF	Type I SS	Mean Square	F Value	Pr > F
time	2	2.88888889	1.44444444	0.46	0.6829
machine	2	60.22222222	30.11111111	9.68	0.0936
athelete	2	42.88888889	21.44444444	6.89	0.1267

Prob 5. & SAS code

	Process		
Film	A	B	C
Kodak	32,34,31,30,37	26,29,27,30,31	28,28,27,30,32
Fuji	43,41,44,50,47	32,38,38,40,36	32,32,36,35,34
Agfa	23,24,25,21,26	27,30,25,25,27	25,27,26,22,25

Make sure you construct an interaction plot for the interaction between process and film brand. Perform a multiple comparison on both main effects.

```
data p5;
        Do obs = 1 to 5;
                Do film = 'kodak', 'fuji', 'agfa';
                        Do process = 'a', 'b', 'c';
                                input brightness @@;
                                output;
        end;end;end;
        cards;
32 26 28 43 32 32 23 27 25 34 29 28 41 38 32 24 25 27 31 27 27 44 38 36 25 25 26 30 30
30 50 40 35 21 25 22 37 31 32 47 36 34 26 27 25
;
run;
proc glm data = p5;
        class film process;
        model brightness = film process film*process;
        means film process film*process/tukey;
run;
```

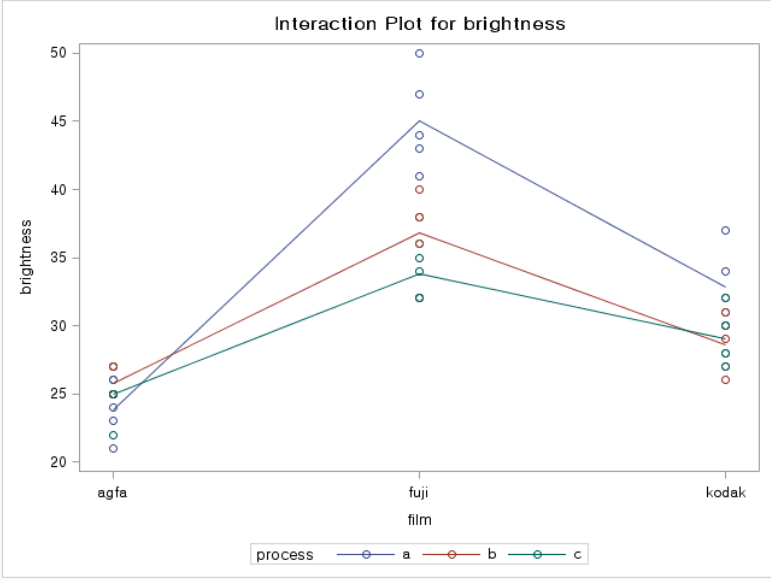
Prob 5. result

KUCEL 고려대학교인지인간공학연구실 Cognitive Ergonomics Laboratory

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	1825.377778	228.172222	41.65	<.0001
Error	36	197.200000	5.477778		
Corrected Total	44	2022.577778			

R-Square	Coeff Var	Root MSE	brightness Mean
0.902501	7.506838	2.340465	31.17778

Source	DF	Type I SS	Mean Square	F Value	Pr > F
film	2	1425.377778	712.688889	130.11	<.0001
process	2	172.311111	86.155556	15.73	<.0001
film*process	4	227.688889	56.922222	10.39	<.0001



Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	film
A	38.5333	15	fuji
B	30.1333	15	kodak
C	24.8667	15	agfa

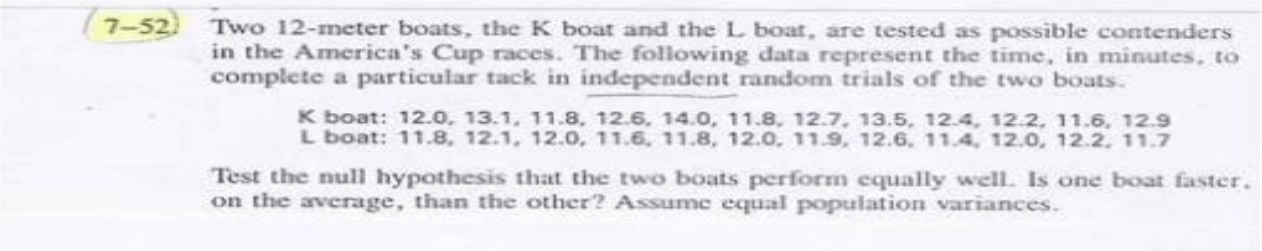
Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	process
A	33.8667	15	a
B	30.4000	15	b
B			
B	29.2667	15	c

Level of film	Level of process	N	brightness	
			Mean	Std Dev
agfa	a	5	23.8000000	1.92353841
agfa	b	5	25.8000000	1.09544512
agfa	c	5	25.0000000	1.87082869
fuji	a	5	45.0000000	3.53553391
fuji	b	5	36.8000000	3.03315018
fuji	c	5	33.8000000	1.78885438
kodak	a	5	32.8000000	2.77488739
kodak	b	5	28.6000000	2.07364414
kodak	c	5	29.0000000	2.00000000

Prob 6. & SAS code

6) Problem 7-52 of text. Test the assumption of normality of each population (PROC UNIVARIATE), and test the assumption of equal variance (PROC TTEST). Perform both the parametric and nonparametric test irrespective of the results of the assumption checks. Which is the best procedure of this problem and why?



7–52. Two 12-meter boats, the K boat and the L boat, are tested as possible contenders in the America's Cup races. The following data represent the time, in minutes, to complete a particular tack in independent random trials of the two boats.

K boat: 12.0, 13.1, 11.8, 12.6, 14.0, 11.8, 12.7, 13.5, 12.4, 12.2, 11.6, 12.9
L boat: 11.8, 12.1, 12.0, 11.6, 11.8, 12.0, 11.9, 12.6, 11.4, 12.0, 12.2, 11.7

Test the null hypothesis that the two boats perform equally well. Is one boat faster, on the average, than the other? Assume equal population variances.

```
data p6;
        input boat time @@;
        cards;
1 12.0 1 13.1 1 11.8 1 12.6 1 14.0 1 11.8
1 12.7 1 13.5 1 12.4 1 12.2 1 11.6 1 12.9
2 11.8 2 12.1 2 12.0 2 11.6 2 11.8 2 12.0
2 11.9 2 12.6 2 11.4 2 12.0 2 12.2 2 11.7
;
run;
proc univariate data =p6 normal;
        by boat;
        var time;
run;
```

<Boat 1>

정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.954713	Pr < W	0.7065
Kolmogorov-Smirnov	D	0.106431	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.027217	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.215211	Pr > A-Sq	>0.2500

<Boat 2>

정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.966104	Pr < W	0.8660
Kolmogorov-Smirnov	D	0.153761	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.038361	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.253389	Pr > A-Sq	>0.2500

Prob 6. SAS code

```
proc ttest data = p6;
        class boat;
        var time;
run;
proc rank data = p6 out = r;
        var time;
run;
proc ttest data = r;
        class boat;
        var time;
run;
proc npar1way data = p6 wilcoxon;
        class boat;
        var time;
run;
```

<Assumption of equal variance>

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	11	5.69	0.0076

<parametric test>

boat	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		12.5500	12.0835	13.0165	0.7342	0.5201	1.2466
2		11.9250	11.7294	12.1206	0.3079	0.2181	0.5227
Diff (1-2)	Pooled	0.6250	0.1484	1.1016	0.5630	0.4354	0.7968
Diff (1-2)	Satterthwaite	0.6250	0.1344	1.1156			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	22	2.72	0.0125
Satterthwaite	Unequal	14.752	2.72	0.0160

<non-parametric test>

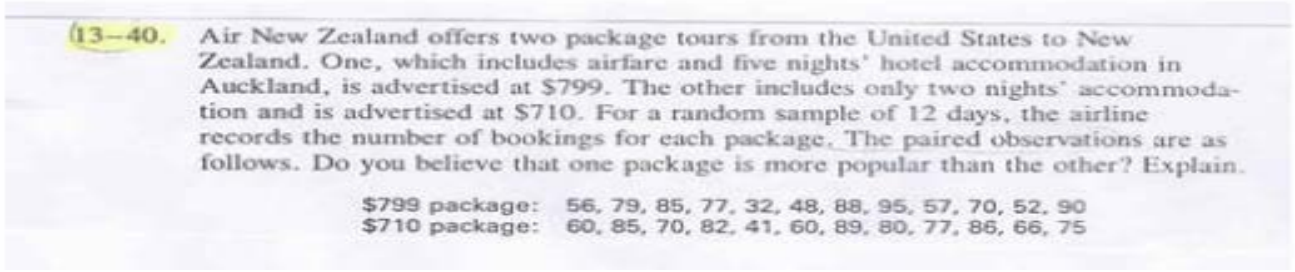
Wilcoxon Two-Sample Test	
Statistic	188.0000
Normal Approximation	
Z	2.1760
One-Sided Pr > Z	0.0148
Two-Sided Pr > |Z|	0.0296

t Approximation	
One-Sided Pr > Z	0.0200
Two-Sided Pr > |Z|	0.0401
Z includes a continuity correction of 0.5.	

Prob 7. & SAS code

7) Problem 13-40 of text. Do the same as specified in problem 8.

8) Problem 13-29 of text. Do both the parametric and nonparametric procedure. Which is more appropriate?



13–40. Air New Zealand offers two package tours from the United States to New Zealand. One, which includes airfare and five nights' hotel accommodation in Auckland, is advertised at $799. The other includes only two nights' accommodation and is advertised at $710. For a random sample of 12 days, the airline records the number of bookings for each package. The paired observations are as follows. Do you believe that one package is more popular than the other? Explain.

$799 package: 56, 79, 85, 77, 32, 48, 88, 95, 57, 70, 52, 90
$710 package: 60, 85, 70, 82, 41, 60, 89, 80, 77, 86, 66, 75

```
data p7;
        input dollar799 dollar710 @@;
        diff = dollar799 - dollar710;
        if diff < 0 then ind = 1;
        else ind = 0;
        absdiff = abs(diff);
        cards;
56 60 79 85 85 70 77 82 32 41 48 60 88 89
95 80 57 77 70 86 52 66 90 75
;
run;
proc univariate data = p7 normal;
        var diff;
run;
```

정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.881255	Pr < W	0.0910
Kolmogorov-Smirnov	D	0.182789	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.08451	Pr > W-Sq	0.1675
Anderson-Darling	A-Sq	0.586704	Pr > A-Sq	0.0994

Prob 8. SAS code & result

KUCEL 고려대학교인지인간공학연구실 Cognitive Ergonomics Laboratory

```
proc ttest data = p7;
        paired dollar799*dollar710;
run;
proc rank data = p7 out = rp7;
var absdiff;
run;
data rp7;
set rp7;
if ind = 1 then absdiff = -absdiff;
run;
proc univariate data = rp7;
        var absdiff;
run;
```

<Assumption of equal variance>

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	11	5.69	0.0076

<parametric test>

boat	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		12.5500	12.0835	13.0165	0.7342	0.5201	1.2466
2		11.9250	11.7294	12.1206	0.3079	0.2181	0.5227
Diff (1-2)	Pooled	0.6250	0.1484	1.1016	0.5630	0.4354	0.7968
Diff (1-2)	Satterthwaite	0.6250	0.1344	1.1156			

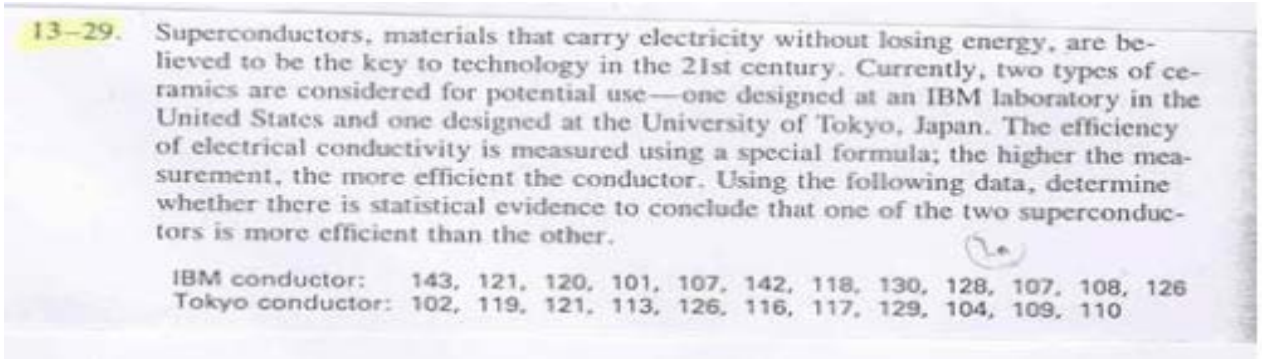
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	22	2.72	0.0125
Satterthwaite	Unequal	14.752	2.72	0.0160

<non-parametric test>

위치모수 검정: Mu0=0				
검정	통계량		p 값	
스튜던트의 t	t	-0.93808	Pr > |t|	0.3683
부호	M	-3	Pr >= |M|	0.1460
부호 순위	S	-12	Pr >= |S|	0.3799

Prob 8. & SAS code

8) Problem 13-29 of text. Do both the parametric and nonparametric procedure. Which is more appropriate?



13-29. Superconductors, materials that carry electricity without losing energy, are believed to be the key to technology in the 21st century. Currently, two types of ceramics are considered for potential use—one designed at an IBM laboratory in the United States and one designed at the University of Tokyo, Japan. The efficiency of electrical conductivity is measured using a special formula; the higher the measurement, the more efficient the conductor. Using the following data, determine whether there is statistical evidence to conclude that one of the two superconductors is more efficient than the other.

IBM conductor: 143, 121, 120, 101, 107, 142, 118, 130, 128, 107, 108, 126
Tokyo conductor: 102, 119, 121, 113, 126, 116, 117, 129, 104, 109, 110

```
data p8;
        input conductor efficiency @@;
        cards;
1 143 1 121 1 120 1 101 1 107 1 142 1 118
1 130 1 128 1 107 1 108 1 126
2 102 2 119 2 121 2 113 2 126 2 116 2 117
2 129 2 104 2 109 2 110
;
run;
proc univariate data = p8 normal;
        var efficiency;
        by  conductor;
run;
```

<conductor 1>

정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.977617	Pr < W	0.9515
Kolmogorov-Smirnov	D	0.088146	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.014829	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.124757	Pr > A-Sq	>0.2500

<conductor 2>

정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.977617	Pr < W	0.9515
Kolmogorov-Smirnov	D	0.088146	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.014829	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.124757	Pr > A-Sq	>0.2500

Prob 8. SAS code & result

KUCEL 고려대학교인지인간공학연구실 Cognitive Ergonomics Laboratory

```
proc ttest data = p8;
        class conductor;
        var efficiency;
run;
proc rank data = p8 out=rr;
        var efficiency;
run;
proc ttest data = rr;
        class conductor;
        var efficiency;
run;
proc npar1way data = rr;
        class conductor;
        var efficiency;
run;
```

<parametric test>

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	21	1.21	0.2386
Satterthwaite	Unequal	18.652	1.24	0.2313

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	10	2.56	0.1502

<non-parametric test>

Wilcoxon Two-Sample Test	
Statistic	116.0000
Normal Approximation	
Z	-0.9547
One-Sided Pr < Z	0.1699
Two-Sided Pr > |Z|	0.3397
t Approximation	
One-Sided Pr < Z	0.1751
Two-Sided Pr > |Z|	0.3501
Z includes a continuity correction of 0.5.	