Dong Yanli

One-way ANOVA Test and Kruskal-Wallist Test

Section I (Introduction)

Therapeutic agents are compounds with a beneficial and desirable effect when consumed or applied. These include a range of products, from topical aloe vera to soothe skin irritation to chemotherapy medications used to attack cancers. Humans have been using therapeutic agents for thousands of years, including some compounds which have continued in use through to the present day. Preparations made with poppies to address pain, for example, were the precursors of powerful synthetic opiates used for the same purpose by modern physicians. New compounds are in constant development, including biological and synthetic preparations to treat new diseases and improve quality of care for existing medical conditions. A therapeutic effect is a consequence of a medical treatment of any kind, the results of which are judged to be desirable and beneficial. This is true whether the result was expected, unexpected, or even an unintended consequence of the treatment. An adverse effect, on the other hand, is a harmful and undesired effect.

In our study, two investigators conducted a clinical trial to determine the effect of two doses of a new therapeutic agent on short-term memory function. A single oral dose of the test preparation was administered to subjects, who were then asked to recall items one hour after exposure to a list consisting of 15 items. A placebo group was included as a control in a parallel-group design.

There are 59 patients in this study, in which 23 patients are taken in placebo, 18 patients are taken 30mg therapeutic agent and 18 patients are taken 60mg therapeutic agent. In the data, variable is the item numbers that are recalled by patients who are given a single oral dose and then were asked to recall items one hour after exposure to a list consisting of 15 items. Observations are the patient number and the dose group name. We want test the hypothesis that μ, the median difference among the three dose types is 0 or not, so we set the null hypothesis is:

$H_0: \mu_{high} = \mu_{low} = \mu_{pacbo}$; Ha: not H0

where μhi means the mean of item numbers patients can recall after administered to μ60mg new therapeutic agent, μlow means the mean of item numbers that patients can recall after administered to 30mg of new therapeutic agent, and μpacbo means the mean of item numbers patients can recall after administered to μpacbo of new therapeutic agent. Our object is to find whether there is difference in average item numbers in the three group, if the null hypothesis not true, then we can conclude that there is two doses of a new therapeutic agent has effect on short-term memory function, otherwise, we do have no evidence that wo doses of a new therapeutic agent has effect on short-term memory function. The data is shown in the figure1.

dosegrp=high

Obs	patno	itemnum
1	1	11
2	2	7
3	3	7
4	4	11
5	5	9
6	6	9
7	7	12
8	8	13
9	9	9
10	10	13
11	11	10
12	12	12
13	13	9
14	14	15
15	15	12
16	16	14
17	17	15
18	18	12

dosegrp=low

Obs	patno	itemnum
19	1	8
20	2	12
21	3	7
22	4	8
23	5	5
24	6	6
25	7	6
26	8	5
27	9	3
28	10	6
29	11	9
30	12	6
31	13	11
32	14	8
33	15	6
34	16	9

dosegrp=low

Obs	patno	itemnum
35	17	11
36	18	5

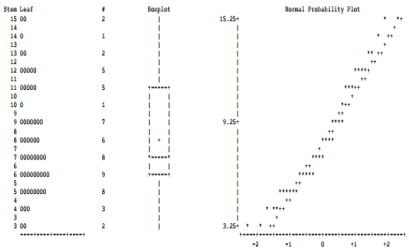
dosegrp=placebo

Obs	patno	itemnum
37	1	6
38	2	5
39	3	6
40	4	8
41	5	3
42	6	4
43	7	7
44	8	4
45	9	7
46	10	6
47	11	7
48	12	8
49	13	5
50	14	6
51	15	5
52	16	5
53	17	7
54	18	8
55	19	5
56	20	9
57	21	11
58	22	4
59	23	7

(figure1: data for 59 observations)

Section II (Summary Statistics)

The stem-leaf plot from display the quantitative variable for the 59 observation, it shows the relative density and shape of the data. As can be seen from figure2 majority of the measurements for variable itemnum lie in 5, 6, 7 ,8 9.



(figure2: stem-leaf plot for 59 observations)

We can see from the SAS statistics report from figure3 to figure5, the mean of the three groups high (60mg), low(30mg) and placebo is approximately 11.11,7.28, 6.22 respectively, it refters to the average item number that subjects in three dose groups . The median of the three groups high, low and placebo is approximately 11.5, 6.5, 6.00 respectively, which refers to the middle value in the three group datasets. Moreover, the standard deviation of the three groups is approximately 2.45,2.42,1.86 respectively, which refers to the measure of the dispersion of a set of data from its mean. The variance of the three groups are 5.99, 5.86, 3.45 respectiely, which refers to the expectation of the squared deviation of a random variable from its mean in the three groups data.

Basic Statistical Measures			
Location		Variability	
Mean	11.11111	Std Deviation	2.44682
Median	11.50000	Variance	5.98693
Mode	9.00000	Range	8.00000
		Interquartile Range	4.00000

(figure3: basic staistis for doesgrp-high)

Basic Statistical Measures			
Location		Variability	
Mean	7.277778	Std Deviation	2.42064
Median	6.500000	Variance	5.85948
Mode	6.000000	Range	9.00000
		Interquartile Range	3.00000

（figure4：staistis for doesgrp-low）

Basic Statistical Measures			
Location		Variability	
Mean	6.217391	Std Deviation	1.85758
Median	6.000000	Variance	3.45059
Mode	5.000000	Range	8.00000
		Interquartile Range	2.00000

(figure5: statistics for dosegrp-placebo)

In figure6, figure7, figure8, we can see that

In dose group high(60mg): Rang=15-7=8; Interquartile rang=Q3-Q1=13-9=4

In dose group low(30mg): Rang=12-3=9; Interquartile rang=Q3-Q1=9-6=3

In dose group placebo: Rang=11-3=8; Interquartile rang=Q3-Q1=7-5=2

Quantiles (Definition 5)	
Level	Quantile
100% Max	15.0
99%	15.0
95%	15.0
90%	15.0
75% Q3	13.0
50% Median	11.5
25% Q1	9.0
10%	7.0
5%	7.0
1%	7.0
0% Min	7.0

(figure6:dosegrp-high quantile)

Quantiles (Definition 5)	
Level	Quantile
100% Max	12.0
99%	12.0
95%	12.0
90%	11.0
75% Q3	9.0
50% Median	6.5
25% Q1	6.0
10%	5.0
5%	3.0
1%	3.0
0% Min	3.0

(figre7:dosegrp-low quantile)

Quantiles (Definition 5)	
Level	Quantile
100% Max	11
99%	11
95%	9
90%	8
75% Q3	7
50% Median	6
25% Q1	5
10%	4
5%	4
1%	3
0% Min	3

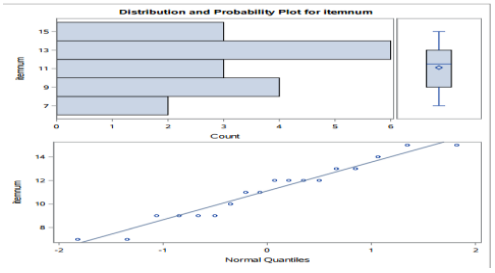
(figure8: dosegrp-placebo:quantile)

The characteristics of the graph

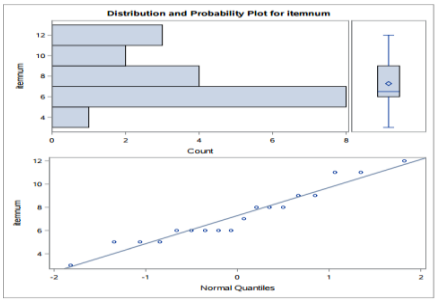
From figure9, we know that the distribution for dose group high(60mg) is approximately symmetric, the data in QQ plot closest to the regression line.

From figure10, we know that the distribution for dose group high(30mg) is not symmetric, it is right skew, the data in QQ plot closest to the regression line.

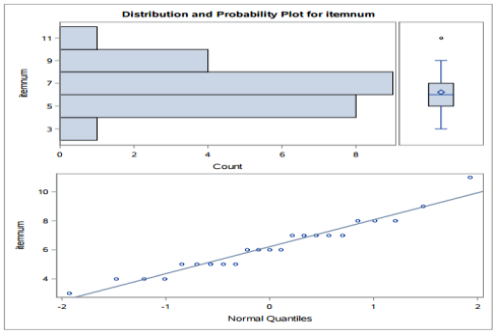
From figure11, we know that the distribution for dose group placebo is approximately symmetric,the data in QQ plot closest to the regression line.



(figure9: dosegrp=high)



(figure10: doesgrp=placebo)



(figure11: doesgrp=low)

Section III (Statistical Analysis)

In the normality test, we can see from figure12 that p-value for dosegrp=low, high and placebo are 0.4056, 0.3181 and 0.3982(>α=0.05), so we cannot reject the null hypothesis, so do not have enough evidence to support the claim that there is difference in the item number recalled in the three does groups patients.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.94874	Pr < W	0.4056
Kolmogorov-Smirnov	D	0.141803	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.053286	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.340855	Pr > A-Sq	>0.2500

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.94238	Pr < W	0.3181
Kolmogorov-Smirnov	D	0.201205	Pr > D	0.0520
Cramer-von Mises	W-Sq	0.08616	Pr > W-Sq	0.1644
Anderson-Darling	A-Sq	0.505471	Pr > A-Sq	0.1840

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.956595	Pr < W	0.3982
Kolmogorov-Smirnov	D	0.135188	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.066655	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.40868	Pr > A-Sq	>0.2500

(figure12: test for Normality dosegrp=low/high/placebo)

1. One-way ANOVA test

One-way ANOVA is to test data when there is only way to classify the population of interest. We call this categorical explanatory variable a factor. One-way analysis of variance is a statistical method for comparing several population means. One-way ANOVA test the null hypothesis that the population means are all equal, the alternative is that they are not all equal. The alternative could be true because all the mean are different or simply because one of them differs from the rest. If we reject the null hypothesis, we need to perform some further analysis to draw conclusion about which population means differ from which others and by how much. In our study, the model factor is Dose Group, one assumption is the normality assumption, one assumption is the homogeneity of variance (HOV) assumption.

we can see from the data in figure13 that the test summary, conducted at a significant level α=0.05 is show as:

Null Hypothesis: H0: μpacbo=μhi=μlow

Alternative Hypothesis: Ha: not H0

Test Statistics: F=128.315/4.952=25.92;

Decision rule: reject H0 if F>$F^2$59 (0.05) = 3.18

Conclusion: Because F = 25.92 > 3.18, we reject H0 and conclude that there is a significant difference in item numbers among the three dose group patients.

The GLM Procedure

Dependent Variable: itemnum

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	256.6302710	128.3151355	25.91	<.0001
Error	56	277.3019324	4.9518202		
Corrected Total	58	533.9322034			

(figure13)

Another method is to see the p-value, it is marginal (p-value<0.0001<α=0.05), so we reject H0 and conclude that there is a significant difference in item numbers among the three dose group patients.

2.Kruskal-Wallist test:

Kruskal-Wallist test is an alternative of the one-way ANOVA in the absence of normal data, it is used to compare population location parameter (mean, median, etc.) among two or more groups based on independent samples. It is based on the rank of the data of the data. According the data from evaluation of the three dose groups in figure14, with **Θ** representing the population location parameter, the test is summarized as shown:

Null Hypothesis: H0: **Θ**high=**Θ**low=**Θ**placebo

Alanative Hypothesis: Ha: not H0

Test statistics: h=26.6823

Decision rule: reject H0 if h> $\chi^2$2(0.05)=5.991

Conclusion: Because h=26.6823>5.991, theis is engough evidence to reject H0. We can conclude that data reveal a statistics difference in items number in the three doses group patients.

The Kruskal-Wallis Test
Evaluation in Three group

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable itemnum
Classified by Variable dosegrp

dosegrp	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
high	18	844.0	540.0	60.366081	46.888889
low	18	473.0	540.0	60.366081	26.277778
placebo	23	453.0	690.0	63.941080	19.695652

Average scores were used for ties.

Kruskal-Wallis Test	
Chi-Square	26.8623
DF	2
Pr > Chi-Square	<.0001

(figure 14)

Another method is to see the p-value, as can be seen from figure13, it is marginal (p-value<0.0001<α=0.05), so we reject H0 and conclude that there is a significant difference in item numbers among the three dose group patients.

3. Compare result obtained from One-way ANOVA test and Kruskal-Wallist test

In One-way ANOVA test we found that test statistics $F = 25.92 > 3.18$, and also p-value is marginal, so we have enough evidence to reject null hypotheis. In Kruskal-Wallist test, test statistics $h = 26.6823 > 5.991$, and p-value is marginal, we have enough evidence to reject null hypotheis. Therefore, both test provide enough evidene to support that there is difference in average items number in the three doses group subjects.

4. Test for Contrast

One-way ANOVA test and Kruskal-Wallist test gives a general answer to a general question about the significant difference among observed groups, however, it does not tell us specifically which means differ from each other. Usually, F test answers a very general question, it is less powerful than test contrast for designed to answer specific questions. A contrast statement is a combination of population mean of the form ψ=∑aiμi where the coefficient ai sum to 0. In our study we test for Contrast is following:

Null hypothesis: H0: 1/2low+1/2high =placebo

Alternative hypothesis: not H0.

From the figure15, we can see that the p-value of (Pr>F)=(Pr>25.12)<0.0001, so we reject the null hypothesis, and claim that we do not have enough evidence to conclude that there is significant difference in the three groups.

The GLM Procedure

Dependent Variable: itemnum

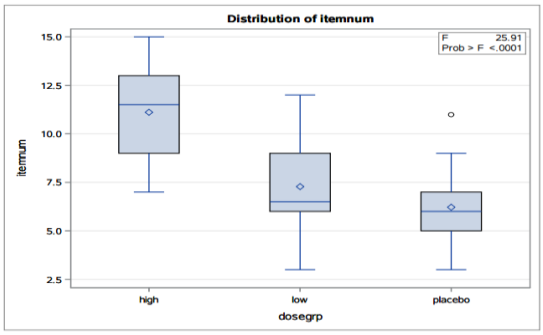
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
active vs placebo	1	124.3802710	124.3802710	25.12	<.0001

(figure15:test for contrast)

5. Multiple comparisons of means using pairwise t-test and Dunnett test.

The One-way ANOVA test and Kruskal-Wallist test tells you whether the means are significantly different from each other, but it does not tell which means differ from which other means. Multiple-comparison provides more detailed information about the differences among the means. The goal in multiple comparisons is to compare the average effects of three or more "treatments" to decide which treatments are better, which ones are worse. Boxplot can help check homogeneity of variance among groups,

as it depicts the range of values for each group shown by the vertical lines, the interquartile rang(25-75 percentile) shown by the shaded box, the group means shown by the diamond symbol, and the group medians shown by the horizontal lines within each box. From the figure 16, we can see that the mean for group high (60mg), low (30mg) and placebo are approximately 11.11, 7.28 and 6.217 respectively, in which the group high(60mg)has the biggest value. The median for the three groups high, low and placebo are 11.500, 6.50 and 6.00 respectively. The 25 interquartile for group high, low and placebo are approximately are 9.0, 6.0, 5 and 75 interquartile for group high, low and placebo are 13, 9, 7 respectively. Overall, we can see that the three group are differ from each other. Specifically, group high (60mg) with higher values in range than other two groups, and group low (30mg) has a little bigger value in mean and median than group placebo.



(figure16)

Section IV (Conclusion)

The test is summarized as follows:

Null Hypothesis: H0: μpacbo=μhi=μlow

Alternative Hypothesis: Ha: not H0

p-value(high)= 0.4056 (>$\boldsymbol{\alpha}$=0.05),

p-value(low)= 0.3181 (>$\boldsymbol{\alpha}$=0.05),

p-value(placebo)=0.3982(>$\boldsymbol{\alpha}$=0.05)

decision rule: reject H0 if p-value< $\boldsymbol{\alpha}$=0.05

So in normality test, we reject the null hypothesis H0 because all the p-value is great than $\boldsymbol{\alpha}$=0.05, so do not have sufficient evidence to say that we there is difference mean item number recalled by patients in the three dose group.

However, in One-way ANOVA test and Kruskal-Wallist test, we both cannot reject the null hypothesis H0, also the p-value in the two test is very marginal . Therefore, we do not have enough evidence to claim that there is difference in the

average number of items recalled by subjects who are given oral therapeutic agent by three does groups. As our data is not normal in our study, we choose the test for homogeneity of variance, we conclude that there is difference in the average number of items recalled by subjects who are given oral therapeutic agent by three does groups

To provide us with more specific information about the linear combination of the three group variable, we test Null hypothesis: H0: 1/2low+1/2high =placebo. because the p-value is marginal, we reject the null hypothesis, can say they do not such liner relationship, and the contrast SS is 124.

Multiple comparisons compare the average effects of three groups to decide which treatments are better. Through the analysis of boxplot, we found that there is much difference in the average items number recalled by subjects who are divided in three groups. Subjects who are given 60mg does new therapeutic agent can recall most items numbers, followed by the subjects who are given 30mg does new therapeutic agent. Therefore, we can make a conclusion that two doses of a new therapeutic agent on have better effect on short-term function.

Section V(Code)

```
data clinic;
input patno dosegrp $ itemnum @@;
datalines;
1 placebo 6   2 placebo 5 3 placebo 6 4 placebo 8 5 placebo 3 6 placebo 4 7 placebo 7
8 placebo 4 9 placebo 7   10 placebo 6   11 placebo 7 12 placebo 8   13 placebo 5 14
placebo 6   15 placebo 5 16 placebo 5     17 placebo 7   18 placebo 8 19 placebo 5 20
placebo 9 21 placebo 11   22 placebo 4 23 placebo 7 1   low 8 2 low 12   3 low 7 4
low 8 5 low 5 6 low 6 7 low 6   8 low 5 9   low 3   10 low 6   11   low 9 12   low 6 13
low 11 14 low 8 15 low 6 16 low 9   17 low 11 18 low 5 1 high 11 2 high 7 3 high 7 4
high 11 5 high 9 6   high 9 7 high 12 8 high 13 9 high 9 10 high 13 11 high 10 12 high
12 13 high 9 14 high 15   15 high 12   16 high 14 17 high 15 18 high 12
;
proc sort data=clinic;
by dosegrp;
proc print data=clinic;
by dosegrp;
run;

proc univariate data=clinic PLOT;
                    var itemnum ;
                    by dosegrp;
                    histogram itemnum ;
                    run;

                    proc boxplot data=clinic；
                    plot itemnum*dosegrp/boxstyle=schematicid;
                    run;
title "test for normality";
proc univariate normal data=clinic; by dosegrp;
var itemnum;
run;
title "summary statistics by treatment group";
proc means mean std n data=clinic;
by dosegrp;
var itemnum;
Title1 'one-way ANOVA';
run;
ods graphics on;
title 'comparsion significant at the 0.5 level and contrast';
proc glm data=clinic plots=boxplot;
class dosegrp;
model itemnum=dosegrp;
```

```
means dosegrp/hovtest Welch t dunnett('PB');
contrast 'active vs. placebo' dosegrp 0.5 0.5 -1;
run;
title1 'The Kruskal-Wallis Test';
title2 'Evaluation in Three group';
proc npar1way wilcoxon data=clinic   plots=wilcoxon;
class dosegrp;
var itemnum;

run;
        ods listing;
        ods graphics off;
        title "stem plot";
                proc univariate data=clinic PLOT;
                var itemnum ;
                run;
```