The George Washington University
6210,2018 spring

Report of Homework4

Jiaying Liu

March 13, 2018

1 Part A

1.1 Data set "DrugTest"

The data set contains 30 observation and there are three variable which are "Drug", "PreTreatment", "PostTreatment". To note that the variable of Drug is categorical variable and have three level.

Table 1: Original Data set.

Obs	Drug	PreTreatment	PostTreatment
1	A	11	6
2	A	8	0
3	A	5	2
4	A	14	8
5	A	19	11
6	A	6	4
7	A	10	13
8	A	6	1
9	A	11	8
10	A	3	0
11	D	6	0
12	D	6	2
13	D	7	3
14	D	8	1
15	D	18	18
16	D	8	4
17	D	19	14

Continued on next page

Table 1 – continued from previous page

Obs	Drug	PreTreatment	PreTreatment
18	D	8	9
19	D	5	1
20	D	15	9
21	F	16	13
22	F	13	10
23	F	11	18
24	F	9	5
25	F	21	23
26	F	16	12
27	F	12	5
28	F	12	16
29	F	7	1
30	F	12	20

In following step,we begin with some basic analysis of data. From the boxplot(figure 1 and figure 2), showing the distribution of different quantile of pretreatment and post-treatment under different drug. We can see that before treatment and after treatment, the boxplot do not change a lot under drug A, the boxplot compress under drug B, and the boxplot expand under drug C. We can first assume that the drug B and Drug C have significant effect on treatment.

Figure 1: Distribution of PerTreatment by Drug

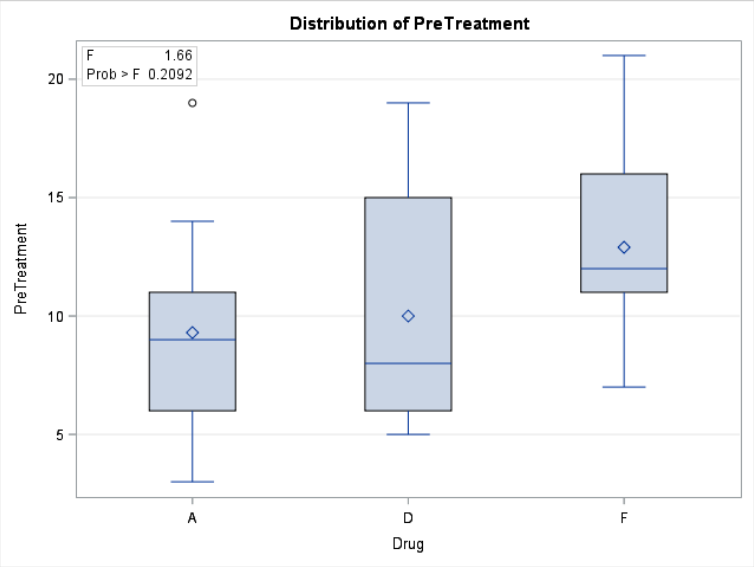
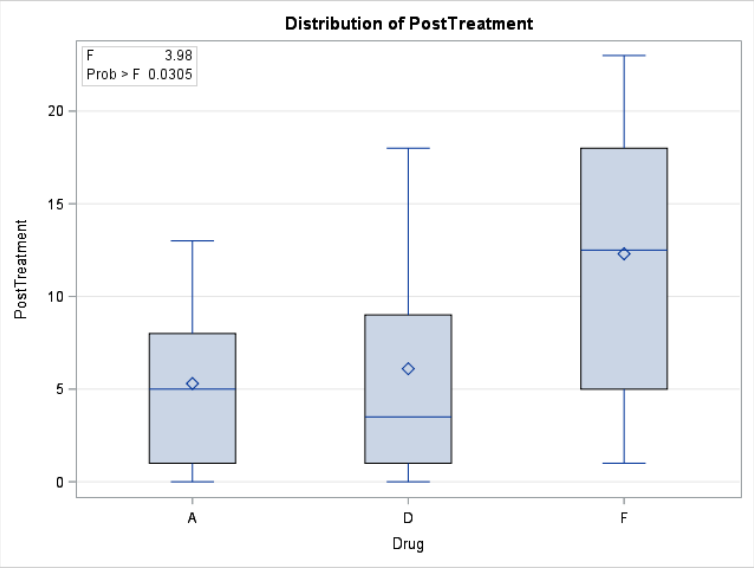


Figure 2: Distribution of PostTreatment by Drug



If we fit the generalized linear model, we can get the result (table 2, table 3, table 4) that the model p value is smaller than 0.05 which mean the model is significant.And when we check the type I sums of squares table. which represent the effect of adding a term to an existing model in a give order. The model become more significant when add the pretreatment variable. Regard to the type III sums of squares, it represent the contribution of each term to a model including all other possible terms. And we can find that when add the pretreatment to the model, the f value increase however when add the drug variable into model, the f value decrease. So we need to check the influence of drug.

Table 2: ANOVA TABLE with postTreatment.

Source	**DF**	**Sum of Squares**	**Mean Square**	**Fvalue**	**Pr>F**
Model	3	871.497403	290.499134	18.10	<.0001
Error	26	417.202597	16.046254		
Corrected Total	29	1288.700000			

Table 3: Tpye I sums of squares.

Source	**DF**	**Type I SS**	**Mean Square**	**Fvalue**	**Pr>F**
Drug	2	293.6000000	146.8000000	9.15	0.0010
PreTreatment	1	577.8974030	577.8974030	36.01	<.0001

Table 4: Tpye III sums of squares.

Source	DF	Type III SS	Mean Square	Fvalue	Pr>F
Drug	2	68.5537106	34.2768553	2.14	0.1384
PreTreatment	1	577.8974030	577.8974030	36.01	<.0001

Continued on next page

1.2 ANOVA TABLE between "Posttreatment" and "Drug"

1.2.1 one-two ANOVA (parametric)

Let's take the variable "PostTreatment" as response and variable "Drug" as factor. From the table 5, showing that the p value is 0.0305 which is small than the 0.05. So it reject the null hypothesis ($H0 = \alpha i$ is equal to zero). We can conclude that the different level of drug is significant influence the posttreatment.

Table 5: ANOVA Table with class variable drug.

Source	DF	Sum of Squares	Mean Square	Fvalue	Pr>F
Model	2	293.600000	146.800000	3.98	0.0305
Error	27	995.100000	36.855556		
Corrected Total	29	1288.700000			

1.2.2 One-way ANOVA with contrasts

Sometime we are not necessary to check the equality of all population means. We can check the subgroup by using the linear contrast. In this section we can check the influence of drug A and drug F or drug D and drug F or even the drug A and combine drug D and drug F.From the table 7, displaying the result of contrast variable. We find that all group of p value are smaller than 0.05. So we have reason to believe that they have different population mean. All level of Drug is significant.

Table 6: ANOVA Table with class variable drug.

Source	DF	Sum of Squares	Mean Square	Fvalue	Pr>F
Model	2	293.600000	146.800000	3.98	0.0305
Error	27	995.100000	36.855556		
Corrected Total	29	1288.700000			

Table 7: Contrast table.

Contrast	DF	Contrast SS	Mean Square	Fvalue	Pr>F
A vs. F	1	245.0000000	245.0000000	6.65	0.0157
D vs. F	1	192.2000000	192.2000000	5.21	0.0305
A and D vs. F	1	290.4000000	290.4000000	7.88	0.0092

1.2.3 One-way ANOVA Nonparametric

If the data do not follow the normal assumption or we do not sure the distribution. We can use the nonparametric method to do the ANOVA test.We use Kruskal-Wallis rank sum test. We can see the table 9 which is the result of Kruskal-wallis test. We find that the p value is 0.0483 which is smaller than 0.05. So we can reject the null hypothesis and believe that the different level of drug have different distribution and the variable is significant.

Table 8: Nonparametric one-way ANOVA.

Wilcoxon Scores for Variable PostTreatment Classified by Variable Drug					
Drug	**N**	**Sum of Scores**	**Expected under H0**	**Std Dev Under H0**	**Mean Score**
A	10	122.00	155.0	22.677144	12.200
D	10	132.50	155.0	22.677144	13.250
F	10	210.50	155.0	22.677144	21.050

Table 9: Kruskal-Wallis Test.

Source	Value
Chi-Square	6.0612
DF	2
Pr > Chi-Square	0.0483

1.2.4 conclusion

From above step, we preform the ANOVA TABLE of variable posttreatment as response and drug as factor. And under parametric, contrast and nonparametric situation, we all found that p value are all smaller than 0.05 which mean that the drug have significant influence on variable pretreatment. It prove the boxplot result(figure 1 and figure 2) that the quantile distribution is different between pretreatment and posttreatment after using drug. We have reason to believe that drug have significant influence on posttreatment.

1.3 ANOVA TABLE between "PreTreatment" and "Drug"

1.3.1 one-two ANOVA (parametric)

Let's take the variable "PreTreatment" as response and variable "Drug" as factor. From the table 5, showing that the p value is 0.2092 which is greater than the 0.05. So it cannot reject the null hypothesis ($H0 = \alpha i$ is equal to zero). Each levels of Drug have same α. We can conclude that the different levels of Drug have no influence on pretreatment.But it is make sense because we would not use drug in the process of pretreatment. So drug would not influence the variable pretreatment.

Table 10: ANOVA Table with class variable drug.

Source	DF	Sum of Squares	Mean Square	Fvalue	Pr>F
Model	2	72.8666667	36.4333333	1.66	0.2092
Error	27	593.0000000	21.9629630		
Corrected Total	29	665.8666667			

1.3.2 One-way ANOVA with contrasts

Sometime we are not necessary to check the equality of all population means. We can check the subgroup by using the linear contrast. In this section we can check the influence of drug A and drug F or drug D and drug F or even the drug A and combine drug D and drug F.From the table 12, displaying the result of contrast variable. We find that all group of p value are greater than 0.05. All level of Drug is insignificant in variable pretreatment.

Table 11: ANOVA Table with class variable drug.

Source	DF	Sum of Squares	Mean Square	Fvalue	Pr>F
Model	2	72.8666667	36.4333333	1.66	0.2092
Error	27	593.0000000	21.9629630		
Corrected Total	29	665.8666667			

Table 12: Contrast table.

Contrast	DF	Contrast SS	Mean Square	Fvalue	Pr>F
A vs. F	1	64.80000000	64.80000000	2.95	0.0973
D vs. F	1	42.05000000	42.05000000	1.91	0.1778
A and D vs. F	1	70.41666667	70.41666667	3.21	0.0846

1.3.3 One-way ANOVA Nonparametric

If the data do not follow the normal assumption or we do not sure the distribution. We can use the nonparametric method to do the ANOVA test.We use Kruskal-Wallis rank sum test. We can see the table 14 which is the result of Kruskal-wallis test. We find that the p value is 0.1077 which is greater than 0.05. So we can not reject the null hypothesis which each *alpha* have same value.

Table 13: Nonparametric one-way ANOVA.

Wilcoxon Scores for Variable PostTreatment Classified by Variable Drug					
Drug	**N**	**Sum of Scores**	**Expected under H0**	**Std Dev Under H0**	**Mean Score**
A	10	126.50	155.0	22.649250	12.650
D	10	136.00	155.0	22.649250	13.600
F	10	202.50	155.0	22.649250	20.250

Table 14: Kruskal-Wallis Test.

Source	Value
Chi-Square	4.4569
DF	2
Pr > Chi-Square	0.1077

1.3.4 conclusion

From above step, we preform the ANOVA TABLE of variable pretreatment as response and drug as factor. And under parametric, contrast and nonparametric situation, we all found that p value are all greater than 0.05 which mean that the drug have no significant influence on variable pretreatment. Again, we have the reason to believe that drug have no influence on drug cause we would not use drug in the process of pretreatment.

1.4 Overall correlation between the variables" pretreatment" and "posttreatment"

Table 15 displaying the result of pearson correlation and the table 16 displaying the nonparametric result of spearmen correlation. Under the parametric(normal assumption) situation, the correlation coefficients is 0.7834 and its p value <0.001 which mean the pretreatment and posttreatment are highly related. And we can check the situation under nonparametric, the spearman correlation coefficients is 0.82153 and its p-value is much smaller than 0.05. So we can conclude that pretreatment and posttreatment have strong linear relation. The scatter plot(figure 3) also verify this conclusion. So there exist a possible assumption that pretreatment is actually influence by pretreatment?

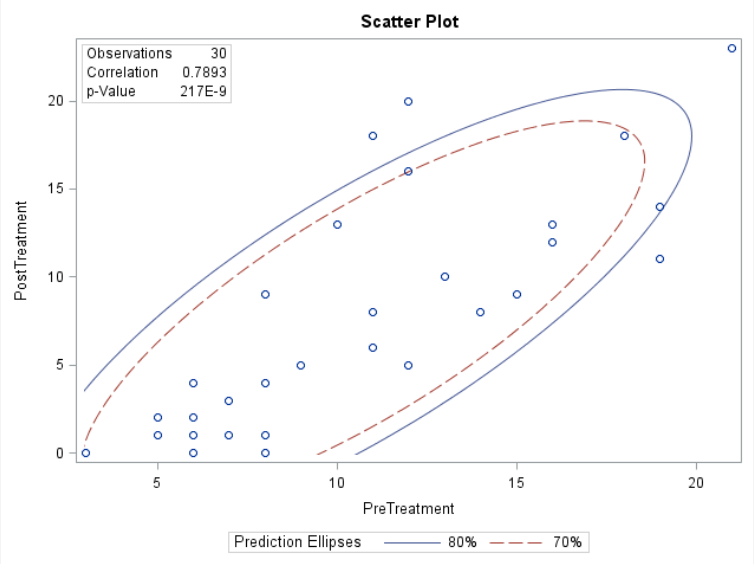
Table 15: Pearson Correlation Coefficients.

source	**PosrTreatment**	**Pretreatment**
PostTreatment	1.00000	0.78934 <.0001
PreTreatment	0.78934 < .0001	1.00000

Table 16: Spearman Correlation Coefficients.

source	**PosrTreatment**	**Pretreatment**
PostTreatment	1.00000	0.82153 <.0001
PreTreatment	0.82153 < .0001	1.00000

Figure 3: Scatter Plot of posttreatment and pretreatment



1.5 Within group(drug type) correlation between the variables" pretreatment" and "posttreatment"

Let's consider the correlation coefficient between pretreatment and posttreatment when control the partial variable drug. We can see that even control the variable drug, the pearson and spearman correlation coefficient are still highly related.

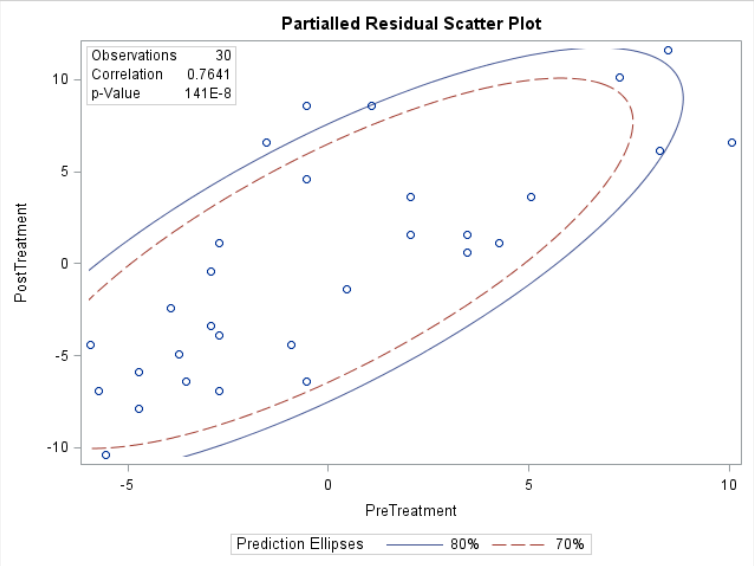
Table 17: Pearson Correlation Coefficients.

source	PosrTreatment	Pretreatment
PostTreatment	1.00000	0.76410 <.0001
PreTreatment	0.76410 < .0001	1.00000

Table 18: Spearman Correlation Coefficients.

source	PosrTreatment	Pretreatment
PostTreatment	1.00000	0.79180 <.0001
PreTreatment	0.79180 < .0001	1.00000

Figure 4: Scatter Plot of posttreatment and pretreatment



1.6 Conclusion

From the above analysis, we can conclude that if we just test the ANOVA TABLE between posttreatment and drug, variable drug have significant influence on response posttreatment. And when we explore the correlation between pretreatment and posttreatment, we find that posttreatment and pretreatment have strong linear relation. It can raise a guess that the influence of pretreatment is much significant than drug. And we can see that in the table 4 (type III SS table) pretreatment contribution is more significant when the model including all other possible terms.When we controlling the effect of pretreatment and then add the variable of drug, from the table 19 type I SS, it is obvious that the drug is no longer significant. The reason is that the posttreatment is highly relate to pretreatment, the effect is more strong than drug.

Table 19: Tpye I sums of squares.

Source	DF	Type I SS	Mean Square	Fvalue	Pr>F
PreTreatment	1	802.9436924	802.9436924	50.04	<.0001
Drug	2	68.5537106	34.2768553	2.14	0.1384

Table 20: Tpye III sums of squares.

Source	DF	Type I SS	Mean Square	Fvalue	Pr>F
PreTreatment	1	802.9436924	802.9436924	50.04	<.0001
Drug	2	68.5537106	34.2768553	2.14	0.1384

2 Part B

2.1 Original Data

The data are from an experiment examining the effects of codeine and acupuncture on postoperative dental pain in male subjects. The variable codeine contain two level.The variable acupuncture contain two levels and pain level contain 8 levels. And the response variable variable is relief which the higher the score, the more relied the patient has.

Table 21: Pain Relief Data .

Obs	Pain level	Codeine	Acupuncture	Relief
1	1	1	1	0.0
2	1	2	1	0.5
3	1	1	2	0.6
4	1	2	2	1.2
5	2	1	1	0.3
6	2	2	1	0.6
7	2	1	2	0.7
8	2	2	2	1.3
9	3	1	1	0.4
10	3	2	1	0.8
11	3	1	2	0.8
12	3	2	2	1.6
13	4	1	1	0.4
14	4	2	1	0.7
15	4	1	2	0.9
16	4	2	2	1.5
17	5	1	1	0.6
18	5	2	1	1.0
19	5	1	2	1.5
20	5	2	2	1.9
21	6	1	1	0.9
22	6	2	1	1.4
23	6	1	2	1.6
24	6	2	2	2.3
25	7	1	1	1.0
26	7	2	1	1.8
27	7	1	2	1.7
28	7	2	2	2.1
29	8	1	1	1.2
30	8	2	1	1.7
31	8	1	2	1.6
32	8	2	2	2.4

2.2 Factorial ANOVA table

Since the variable of codeine, acupuncture a d pain level have several levels. We are interested in the effect of different variable. We are going to create a factorial ANOVA. From the

table 22 and 23 which are the result of ANOVA TABLE. First we can find that the model: $Relief = PainLevel + Codeine + Acupuncture + Codeine * Acupuncture$ is significant because its p value smaller than 0.0001. And from the Table 23 we can see the effect of each first order variable and second order interaction. It show tat the the variable of painlevel is significant, the variable of codeine is significant and the variable of acupuncture is also significant because both of there p value are smaller than 0.0001. However the second order interaction (codeine*Acupuncture) p value is 0.0923 which is greater than 0.05, so we can conclude that the second order interaction is not significant to relief pain.

Table 22: ANOVA TABLE.

Source	**DF**	**Sum of Square**	**Mean Square**	**F value**	**P > f**
Model	10	11.33500000	1.13350000	78.37	<.0001
Error	21	0.30375000	0.01446429		
Corrected Total	31	11.63875000			

Table 23: ANOVA TABLE.

Source	**DF**	**Anova SS**	**Mean Square**	**F value**	**P > f**
PainLevel	7	5.59875000	0.79982143	55.30	<.0001
Codeine	1	2.31125000	2.31125000	159.79	<.0001
Acupuncture	1	3.38000000	3.38000000	233.68	<.0001
Codeine*Acupuncture	1	0.04500000	0.04500000	3.11	0.0923

2.3 Scheffe's Method

The Scheffe Test is a post-hoc test used in Analysis of Variance. After we have run ANOVA and got a significant F-statistic then we can run Sheffe's test to find out which pairs of means are significant. The Scheffe test corrects alpha for simple and complex mean comparisons. Complex mean comparisons involve comparing more than one pair of means simultaneously. The table 24 show the result of scheffe test.The result was order from high to low by its group mean and those equal group mean group would be putted in the same letter. From the table 24 we can see that pain levels 8,7,6 are put in the same A, group 6,5 are put in the same B, group 5,3 are put in same C and the group 1,2,3,4 are put in same D. Different group mean they have different group mean.

Table 24: Scheffe's test of pain level.

Means with the same letter are not significantly different				
Scheffe Grouping		**Mean**	**N**	**PainLevel**
	A	1.72500	4	8
	A	1.65000	4	7
B	A	1.55000	4	6
B	C	1.25000	4	5
D	C	0.90000	4	3
D		0.87500	4	4
D		0.72500	4	2

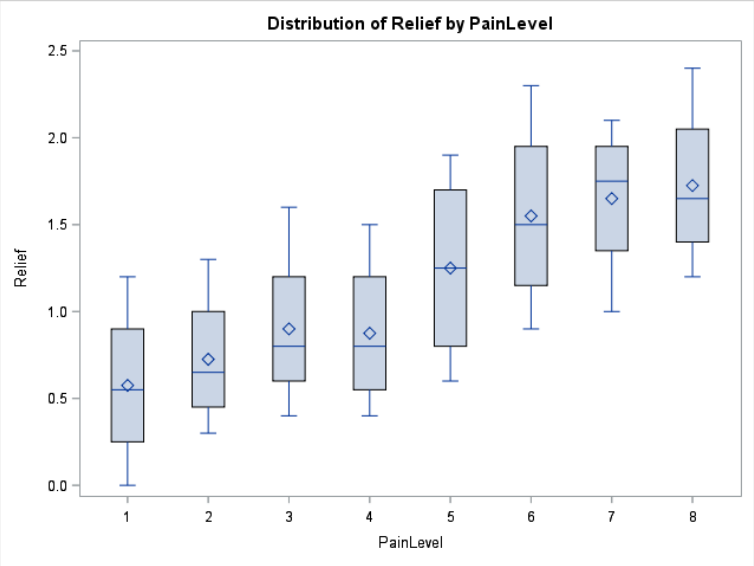
Table 24 – continued from previous page

Scheffe Grouping		Mean	N	PainLevel
D		0.57500	4	1

2.4 BoxPlot of Pain Relief with different Pain Levels

We create the boxplot of pain relief with different pain levels to directly explore the distribution of relief by pain levels. Combining the above result, we can see that the group 1,2,3,4 approximately have same group mean, the group 6,7,8 have same group mean and the group 5 is unique.

Figure 5: Distribution of Relief by PainLevel



2.5 Conclusion

Combining the result of Scheffe's test and the boxplot figure, we cab regroup group 1,2,3,4 as same group, group 5 as same group and the group 6,7,8 as same group in the low to high order.