HW 4 Report

Yongchao Qiao

Part A

For this part, we intend to understand further why the response "PostTreatment" is no longer significantly associated with the factor "Drug" after controlling the effect of "PreTreatment", so we need to reproduce the result first to verify this statement. Then we will take the four-step analysis to find the reason.

1. Reproduce the result of the statement

As it is shown in table 1, when controlling Pretreatment, the P-values of F-test using all three types' SS are the same as 0.1384, which indicates that variable Drug does not contribute significant variation for the PostTreatment when given PreTreatment. So the statement that the response "PostTreatment" is no longer significantly associated with the factor "Drug" after controlling the effect of "PreTreatment" is true.

Table 1: Three types of sum squares for Drug given PreTreatment with PostTreatment as response

Types	Source	DF	SS	Mean Square	F-Value	Pr>F
Type I	PreTreatment	1	802.9437	802.9437	50.04	<0.0001
	Drug	2	68.5537	34.2769	2.14	0.1384
Type II	PreTreatment	1	577.8974	577.8974	36.01	<0.0001
	Drug	2	68.5537	34.2769	2.14	0.1384
Type III	PreTreatment	1	577.8974	577.8974	36.01	<0.0001
	Drug	2	68.5537	34.2769	2.14	0.1384

2. Analyses of association between "PostTreatment" as response and "Drug" as factor

One-way ANOVA (parametric)

H_0: Different levels of the variable Drug have the same effect on the variable PostTreatment

H_1: Not all levels of the variable Drug have the same effect on the variable PostTreatment

Table 2: Analysis of variance

Source	DF	Sum of squares	Mean Square	F Value	Pr > F
Model	2	293.60	146.80	3.98	0.00305
Error	27	995.10	36.86		
Corrected Total	29	1288.70			

Table 3: ANOVA SS for factor Drug

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Drug	2	293.60	146.80	3.98	0.00305

The results are shown in table 2 and table 3. Since there is only one variable in the model, so the Anova analysis for the model is equal to that for the factor Drug. The P-values of the F test in this two tables are the same, as 0.00305, less than 0.05. So the null hypothesis that different levels of the variable Drug have the same effect on the response variable PostTreatment can be rejected. That is, there is an association between "PostTreatment" as response and "Drug" as factor.

O*ne-way ANOVA with contrasts: A vs. F, D vs. F, average of A & D vs. F*

A vs. F:

H_0: Group A and Group F have the same population mean

H_1: Group A and Group F do not have the same population mean

D vs. F:

H_0: Group D and Group F have the same population mean

H_1: Group D and Group F do not have the same population mean

Average of A & D vs. F:

H_0: Average of A & D and Group F have the same population mean

H_1: Average of A & D and Group F do not have the same population mean

Table 4: One-way ANOVA with contrasts

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
A vs. F	1	245.00	245.00	6.65	0.0157
D vs. F	1	192.20	192.20	5.21	0.0305
Average of A & D vs. F	1	290.40	290.40	7.88	0.0092

The results are shown in table 4. For contrast A vs. F, the P-value of F-test is 0.0157, less than 0.05, so the null hypothesis that Group A and Group F have the same population mean can be rejected. For contrast D vs. F, the P-value of F-test is 0.0305, less than 0.05, so the null hypothesis that Group D and Group F have the same population mean can be rejected. For contrast average of A & D vs. F, the P-value of F-test is 0.0092, less than 0.05, so the null hypothesis that average of A & D and Group F have the same population mean can be rejected. Based on these contrast tests, it can be concluded that there is an association between "PostTreatment" as response and "Drug" as factor.

Nonparametric one-way ANOVA (Kruskal-Wallis rank sum test)

H_0: Different levels of the variable Drug have the same effect on the variable PostTreatment

H_1: Not all levels of the variable Drug have the same effect on the variable PostTreatment

Table 5: Kruskal-Wallis rank sum test

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

The results are shown in table 5. The P-values of the Kruskal-Wallis rank sum test is 0.0483 less than 0.05. So the null hypothesis that different levels of the variable Drug have the same effect on the response variable PostTreatment can be rejected. That is, there is an association between "PostTreatment" as response and "Drug" as factor.

3. Analyses of association between "PreTreatment" as response and "Drug" as factor

One-way ANOVA (parametric)

H_0: Different levels of the variable Drug have the same effect on the variable PreTreatment

H_1: Not all levels of the variable Drug have the same effect on the variable PreTreatment

Table 6: Analysis of variance

Source	DF	Sum of squares	Mean Square	F Value	Pr > F
Model	2	72.87	36.43	1.66	0.2092
Error	27	593.00	21.96		
Corrected Total	29	665.87			

Table 7: ANOVA SS for factor Drug

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Drug	2	72.87	36.43	1.66	0.2092

The results are shown in table 6 and table 7. Since there is only one variable in the model, the ANOVA analysis for the model is equal to that for the factor Drug. The P-values of the F test in this two tables are the same, as 0.2092, greater than 0.05. So the null hypothesis that different levels of the variable Drug have the same effect on the response variable PreTreatment cannot be rejected. That is, there is no association between "PreTreatment" as response and "Drug" as factor.

O*ne-way ANOVA with contrasts: A vs. F, D vs. F, average of A & D vs. F*

A vs. F:

H_0: Group A and Group F have the same population mean

H_1: Group A and Group F do not have the same population mean

D vs. F:

H_0: Group D and Group F have the same population mean

H_1: Group D and Group F do not have the same population mean

Average of A & D vs. F:

H_0: Average of A & D and Group F have the same population mean

H_1: Average of A & D and Group F do not have the same population mean

Table 8: One-way ANOVA with contrasts

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
A vs. F	1	64.80	64.80	2.95	0.0973
D vs. F	1	42.05	42.05	1.91	0.1778
Average of A & D vs. F	1	70.42	70.42	3.21	0.0846

The results are shown in table 8. For contrast A vs. F, the P-value of F-test is 0.0973, greater than 0.05, so the null hypothesis that Group A and Group F have the same population mean cannot be rejected. For contrast D vs. F, the P-value of F-test is 0.1778, greater than 0.05, so the null hypothesis that Group D and Group F have the same population mean cannot be rejected. For contrast average of A & D vs. F, the P-value of F-test is 0.0846, greater than 0.05, so the null hypothesis that average of A & D and Group F have the same population mean cannot be rejected. Based on these contrast tests, it can be concluded that there is no association between "PreTreatment" as response and "Drug" as factor.

Nonparametric one-way ANOVA (Kruskal-Wallis rank sum test)

H_0: Different levels of the variable Drug have the same effect on the variable PreTreatmen

H_1: Not all levels of the variable Drug have the same effect on the variable PreTreatmen

Table 9: Kruskal-Wallis rank sum test

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

The results are shown in table 5. The P-values of the Kruskal-Wallis rank sum test is 0.1077 greater than 0.05. So the null hypothesis that different levels of the variable Drug have the same effect on the response variable PreTreatment cannot be rejected. That is, there is no association between "PreTreatment" as response and "Drug" as factor.

4. Evaluate the overall correlation between "PreTreatment" and "PostTreatment"

Parametric way

$$H_0 : \text{Rho} = 0, \quad H_1 : \text{Rho} \neq 0$$

Table 10: Pearson Correlation Test between PreTreatment and PostTreatment |

	PreTreatment	PostTreatment
PreTreatment	1.00000	0.78934(<0.0001)
PostTreatment	0.78934(<0.0001)	1.00000

Non-parametric way

Table 11: Spearman Correlation Test between PreTreatment and PostTreatment |

	PreTreatment	PostTreatment
PreTreatment	1.00000	0.82153(<0.0001)
PostTreatment	0.82153(<0.0001)	1.00000

The results are shown in table 10 and table 11. The Pearson correlation coefficient between PreTreatment and PostTreatment is 0.78934 with P value less than 0.05. Thus, it indicates that the null hypothesis that Rho = 0 can be rejected. That is, variable PreTreatment and variable PostTreatment have a significant linear correlation relationship, with Pearson correlation coefficient as 0.78934. The Spearman correlation coefficient between PreTreatment and PostTreatment is 0.82153 with P value less than 0.05. Thus, it indicates that the null hypothesis that Rho = 0 can be rejected. That is, variable PreTreatment and variable PostTreatment have a significant correlation relationship, with Spearman correlation coefficient as 0.82153. Therefore, there is an overall correlation relationship between variable PreTreatment and variable PostTreatment.

5. Evaluate the within-group correlation between "PreTreatment" and "PostTreatment"

Parametric way

$$H_0 : \text{Rho} = 0, \quad H_1 : \text{Rho} \neq 0$$

Table 12: Pearson Correlation Test between PreTreatment and PostTreatment |

Group		PreTreatment	PostTreatment
A	PreTreatment	1.00000	0.76418(0.0101)
	PostTreatment	0.76418(0.0101)	1.00000
D	PreTreatment	1.00000	0.91139(0.0002)
	PostTreatment	0.91139(0.0002)	1.00000
F	PreTreatment	1.00000	0.66100(0.0374)
	PostTreatment	0.66100(0.0374)	1.00000

Non-parametric way

Table 13: Spearman Correlation Test between PreTreatment and PostTreatment |

Group		PreTreatment	PostTreatment
A	PreTreatment	1.00000	0.75153(0.0122)
	PostTreatment	0.75153(0.0122)	1.00000
D	PreTreatment	1.00000	0.84833(0.0019)
	PostTreatment	0.84833(0.0019)	1.00000
F	PreTreatment	1.00000	0.51238(0.1300)
	PostTreatment	0.51238(0.1300)	1.00000

The results are shown in table 12 and table 13. The Pearson correlation coefficients between PreTreatment and PostTreatment in group A, D and F are 0.76418, 0.91139 and 0.66100 respectively, with all P values less than 0.05. Thus, it indicates that the null hypothesis that Rho = 0 within each group can be rejected. That is, variable PreTreatment and variable

PostTreatment have a significant linear correlation relationship in each group, with Pearson within-group correlation coefficients as 0.76418, 0.91139 and 0.66100 respectively for group A, D and F. The Spearman correlation coefficients between PreTreatment and PostTreatment in group A, D and F are 0.75153, 0.84833 and 0.51238 with first two P-values less than 0.05 and the last P-values greater than 0.05. Thus, it indicates that the null hypothesis that Rho = 0 can be rejected for first two groups but not for group F. That is, variable PreTreatment and variable PostTreatment have a significant correlation relationship in group A and group D, with Spearman with-group correlation coefficient as 0.75153 and 0.84833. Therefore, there is an within-group correlation relationship between variable PreTreatment and variable PostTreatment in group A, D and F respectively.

6. Conclusion

Based on above analysis results, it shows that there is an association between "PostTreatment" as response and "Drug" as factor but no association between "PreTreatment" as response and "Drug" as factor. And there is an overall as well as within-group correlation relationship between variable PreTreatment and variable PostTreatment. Then if we control the effect of variable PreTreatment, since there is a high overall and within-group correlation relationship between PostTreatment and PreTreatment, variable Drug will not contribute significant variation for the PostTreatment comparing to that of PreTreatment. Also, the reason can be illustrated as while there is a statistically significant difference for PostTreatment between different groups which are grouped by variable Drug, this difference is reduced to below the level of background noise when the variable PreTreatment is taken into consideration.

Part B

1. Factorial ANOVA

Table 14: Analysis of variance

Source	DF	Sum of squares	Mean Square	F Value	Pr > F
Model	10	11.34	1.33	78.37	<0.0001
Error	21	0.30	0.01		
Corrected Total	31	11.64			

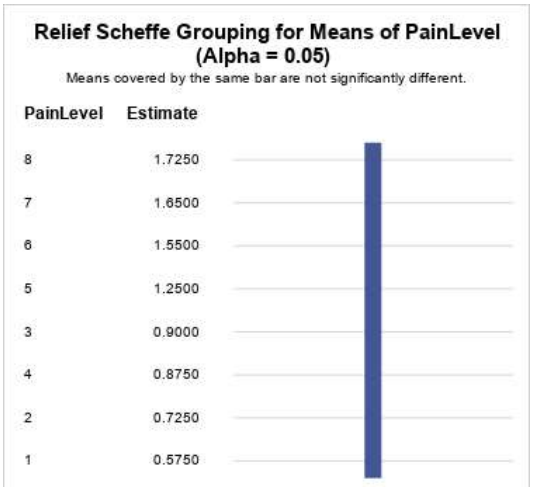
Table 15: ANOVA SS for factor Drug

Source	DF	Anova SS	Mean Square	F Value	Pr > F
PainLevel	7	5.60	0.80	55.30	<0.0001
Codeine	1	2.31	2.31	159.79	<0.0001
Acupuncture	1	3.38	3.38	233.68	<0.0001
Codeine*Acupuncture	1	0.05	0.05	3.11	0.0923

The results are shown in table 14 and table 15. The P-values of the F test for the whole model is less than 0.05. So the null hypothesis that different levels of specified variables have the same effect on the response variable Relief can be rejected. Then for variable PainLevel, the P-values of the F test is less than 0.05. So the null hypothesis that different levels of variable PainLevel have the same effect on the response variable Relief can be rejected; for variable Codeine, the P-values of the F test is less than 0.05. So the null hypothesis that different levels of variable Codeine have the same effect on the response variable Relief can be rejected; for variable Acupuncture, the P-values of the F test is less than 0.05. So the null hypothesis that different levels of variable Acupuncture have the same effect on the response variable Relief can be rejected; for interaction Codeine *Acupuncture, the P-values of the F test is greater than 0.05. So the null hypothesis that different levels of interaction Codeine *Acupuncture have the same effect on the response variable Relief cannot be rejected. In a word, variable PainLevel, Codeine and Acupuncture have the significant effects on the response variable Relief.

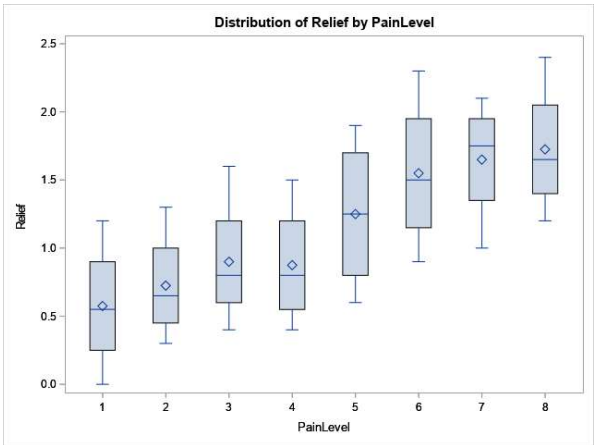
2. Multiple comparison by Scheffe's method

Figure 1: Multiple comparison by Scheffe's method



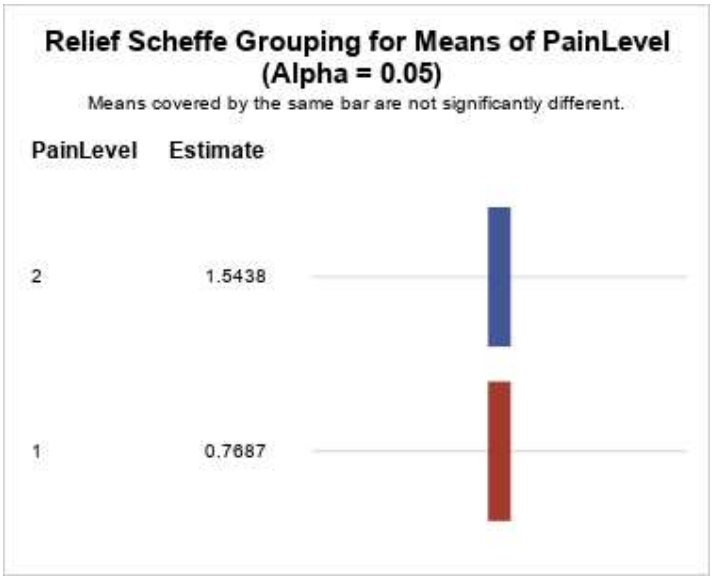
3. Boxplot for pain relief scores against different pain levels

Figure 2: Boxplot for pain relief scores against different pain levels



Based on above two plots, there is just one vertical band for all levels, which means all these 8 levels are not significantly different. The reason why we get this result is because of the sequential mechanism behind the multiple comparison in SAS. Then based on the boxplot, we can still regroup these different pain levels, one possible solution is that we can regroup level 1, 2, 3 and 4 into group 1 with putting the rest of levels into group 2, since first four levels visually have the same population mean while the last four groups visually have another same population mean based on the boxplot. Then a multiple comparison plot for the regrouped data is shown as below and there are two bands in the plots which verifies the possible solution.

Figure 3: Multiple comparison by Scheffe's method for regrouped data



Appendix

```
/* Part A */
  /* Read the data*/
data drugtest;
input Drug $ PreTreatment PostTreatment @@;
datalines;
A 11   6     A   8   0     A   5   2     A 14   8     A 19 11
A   6   4     A 10 13     A   6   1     A 11   8     A   3   0
D   6   0     D   6   2     D   7   3     D   8   1     D 18 18
D   8   4     D 19 14     D   8   9     D   5   1     D 15   9
F 16 13     F 13 10     F 11 18     F   9   5     F 21 23
F 16 12     F 12   5     F 12 16     F   7   1     F 12 20
;
/* Reproduce the result of the statement*/
proc glm data = drugtest;
class Drug;
model PostTreatment = PreTreatment Drug / ss1 ss2 ss3;
run;

/* ANOVA for “PostTreatment” as response and “Drug” as factor */
proc anova data = drugtest;
class Drug;
model PostTreatment = Drug;
run;
/* Sort the data by drug */
proc sort data=drugtest;
by Drug;
run;
/* Do the contrast test for different levels of Drug*/
proc glm data = drugtest;
class Drug;
model PostTreatment = Drug;
contrast 'A vs F' Drug 1 0 -1;
contrast 'D vs F' Drug 0 1 -1;
contrast 'A&D vs F' Drug   -1 -1 2;
run;
/* Nonparametric one-way ANOVA (Kruskal-Wallis rank sum test)t*/
proc npar1way data=drugtest wilcoxon;
class Drug;
var PostTreatment;
run;

/* ANOVA for “PreTreatment” as response and “Drug” as factor */
proc anova data = drugtest;
```

```
class Drug;
model PreTreatment = Drug;
run;
/* Do the contrast test for different levels of Drug*/
proc glm data = drugtest;
class Drug;
model PreTreatment = Drug;
contrast 'A vs F' Drug 1 0 -1;
contrast 'D vs F' Drug 0 1 -1;
contrast 'A&D vs F' Drug    -1 -1 2;
run;
/* Nonparametric one-way ANOVA (Kruskal-Wallis rank sum test)*/
proc npar1way data=drugtest wilcoxon;
class Drug;
var PreTreatment;
run;

/* Overall correlation between “PreTreatment” and “PostTreatment” */
proc corr data=drugtest;
var PreTreatment PostTreatment;    /* calculate the Pearson correlation coefficient and take
the Pearson correlation test */
run;
/* Nonparametrix Overall correlation between “PreTreatment” and “PostTreatment” */
proc corr data=drugtest spearman;
var PreTreatment PostTreatment; /* calculate the Spearman correlation coefficient and take
the Spearman correlation test */
run;

proc corr data=drugtest;
var PreTreatment PostTreatment;
by Drug;
run; /* calculate the partial Pearson correlation coefficient and take the partial Pearson
correlation test */
/* Nonparametrix*/
proc corr data=drugtest spearman;
var PreTreatment PostTreatment;
by Drug;
run;/* calculate the partial Spearman correlation coefficient and take the partial Spearman
correlation test */

/* Part B*/
/* Read the data*/
data PainRelief;
input PainLevel Codeine Acupuncture Relief @@;
```

```
datalines;
1 1 1 0.0   1 2 1 0.5   1 1 2 0.6   1 2 2 1.2
2 1 1 0.3   2 2 1 0.6   2 1 2 0.7   2 2 2 1.3
3 1 1 0.4   3 2 1 0.8   3 1 2 0.8   3 2 2 1.6
4 1 1 0.4   4 2 1 0.7   4 1 2 0.9   4 2 2 1.5
5 1 1 0.6   5 2 1 1.0   5 1 2 1.5   5 2 2 1.9
6 1 1 0.9   6 2 1 1.4   6 1 2 1.6   6 2 2 2.3
7 1 1 1.0   7 2 1 1.8   7 1 2 1.7   7 2 2 2.1
8 1 1 1.2   8 2 1 1.7   8 1 2 1.6   8 2 2 2.4
;
/* Implement the factorial ANOVA with specified variables*/
proc anova data=PainRelief;
class PainLevel Codeine Acupuncture;
model Relief = PainLevel Codeine|Acupuncture;
run;
/* Sort the data*/
proc sort data=PainRelief;
by PainLevel;
run;
/* Multiple comparison by scheffe method */
proc glm data = PainRelief;
class PainLevel;
model Relief = PainLevel;
means PainLevel / scheffe;
run;
/* Boxplot */
proc boxplot data=PainRelief;
plot Relief*PainLevel / boxstyle=schematic;
run;
/* Regrouped data */
data Regroup;
input PainLevel Codeine Acupuncture Relief @@;
datalines;
1 1 1 0.0   1 2 1 0.5   1 1 2 0.6   1 2 2 1.2
1 1 1 0.3   1 2 1 0.6   1 1 2 0.7   1 2 2 1.3
1 1 1 0.4   1 2 1 0.8   1 1 2 0.8   1 2 2 1.6
1 1 1 0.4   1 2 1 0.7   1 1 2 0.9   1 2 2 1.5
2 1 1 0.6   2 2 1 1.0   2 1 2 1.5   2 2 2 1.9
2 1 1 0.9   2 2 1 1.4   2 1 2 1.6   2 2 2 2.3
2 1 1 1.0   2 2 1 1.8   2 1 2 1.7   2 2 2 2.1
2 1 1 1.2   2 2 1 1.7   2 1 2 1.6   2 2 2 2.4
;
/* Multiple comparison by scheffe method for regrouped data*/
proc glm data = Regroup;
```

```
class PainLevel;
model Relief = PainLevel;
means PainLevel / scheffe;
run;
```