BIOSTAT 200B HW8

1. Lab 9 (two-way ANOVA) problems

Q1: What differences do you see in the relationships between the SS for A, B and A*B for balanced versus unbalanced data?

Number of Observations Read	18
Number of Observations Used	18

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	0.02354450	0.02354450	81.38	<.0001
B	2	0.00115811	0.00057906	2.00	0.1778
A*B	2	0.00084633	0.00042317	1.46	0.2701

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.02354450	0.02354450	81.38	<.0001
B	2	0.00115811	0.00057906	2.00	0.1778
A*B	2	0.00084633	0.00042317	1.46	0.2701

Number of Observations Read	14
Number of Observations Used	14

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	0.02101572	0.02101572	65.28	<.0001
B	2	0.00033302	0.00016651	0.52	0.6148
A*B	2	0.00018286	0.00009143	0.28	0.7600

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.01682504	0.01682504	52.27	<.0001
B	2	0.00045773	0.00022887	0.71	0.5198
A*B	2	0.00018286	0.00009143	0.28	0.7600

The SS table shown in the left is for the balanced data, and in the right is for the unbalanced data. We can see that the Type I SS and Type III SS for the balanced data are identical, since the data is balanced (cell sizes are equal) and the factors are uncorrelated predictors. For the unbalanced data, the Type I SS and Type III SS for A and B are different, that is because the factors are correlated predictors and thus the marginal and partial effects are different.

Q2: For the balanced and unbalanced data, provide a 2 x 3 table of cell means. By hand, compute the (unadjusted) factor means and least square (adjusted) means for each data set.

Cell means for balanced data:

B \ A	1	2
1	0.185	0.268
2	0.179	0.259
3	0.212	0.265

A	B	Y LSMEAN
1	1	0.18500000
1	2	0.17866667
1	3	0.21200000
2	1	0.26833333
2	2	0.25933333
2	3	0.26500000

Cell means for unbalanced data:

B \ A	1	2
1	0.187	0.268
2	0.179	0.259
3	0.202	0.265

A	B	Y LSMEAN
1	1	0.18700000
1	2	0.17866667
1	3	0.20200000
2	1	0.26800000
2	2	0.25933333
2	3	0.26500000

Balanced data:

B \ A	1	2	means
1	0.204, 0.17, 0.181	0.257, 0.279, 0.269	0.226
2	0.167, 0.182, 0.187	0.283, 0.235, 0.26	0.219
3	0.202, 0.198, 0.236	0.256, 0.281, 0.258	0.239
means	0.192	0.264	0.228

The unadjusted (factor) means and the adjusted (least squared) means are the same, because each cell size is equal in balanced data.

Unbalanced data:

B \ A	1	2	unadjusted means	adjusted means
1	0.204, 0.17	0.257, 0.279	0.228	0.228
2	0.167, 0.182, 0.187	0.283, 0.235, 0.26	0.219	0.219
3	0.202	0.256, 0.281, 0.258	0.234	0.249
unadjusted means	0.189	0.264		
adjusted means	0.185	0.264		0.230

The unadjusted (factor) means and the adjusted (least squared) means are not the same, because not every cell size is equal in unbalanced data.

Q3: Write code and estimate the contrast for comparing the difference between auditory and visual when elapse time is 10 sec to the difference between auditory and visual when elapse time is 15 sec.

```
proc glm data=rtime;
    class A B;
    model Y = A B A*B;
    contrast '(12-22) - (13-23)' A*B 0 1 -1 0 -1 1;
    estimate '(12-22) - (13-23)' A*B 0 1 -1 0 -1 1;
run;
```

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
(12-22) - (13-23)	1	0.00057408	0.00057408	1.98	0.1843

Parameter	Estimate	Standard Error	t Value	Pr > |t|
(12-22) - (13-23)	-0.02766667	0.01964123	-1.41	0.1843

The estimated difference between auditory and visual when elapse time is 10 sec to the difference between auditory and visual when elapse time is 15 sec is -0.02766667.

2.

a.

```
libname hw8 "/home/u48583412/Biostat 200B/LAB";
data survtime;
 set hw8.survtime;
 logsurvtime = log(survtime);
 run;
proc glm data=survtime;
	class A B;
	model logsurvtime = A B A*B;
run;
proc glm data=survtime;
	class A B;
	model logsurvtime = A B;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	2	5.23747262	2.61873631	48.43	<.0001
B	3	3.55717347	1.18572449	21.93	<.0001
A*B	6	0.39574668	0.06595778	1.22	0.3189

First, perform a natural log transformation of survival time, and conduct a two-way ANOVA model with an interaction. From the table above, the p-value of interaction A*B is 0.3189, which we can conclude there is no interaction between A and B. The table shown below is the ANOVA model without an interaction.

The GLM Procedure

Dependent Variable: logsurvtime

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	8.79464609	1.75892922	31.54	<.0001
Error	42	2.34226247	0.05576815		
Corrected Total	47	11.13690857			

R-Square	Coeff Var	Root MSE	logsurvtime Mean
0.789685	-27.59429	0.236153	-0.855803

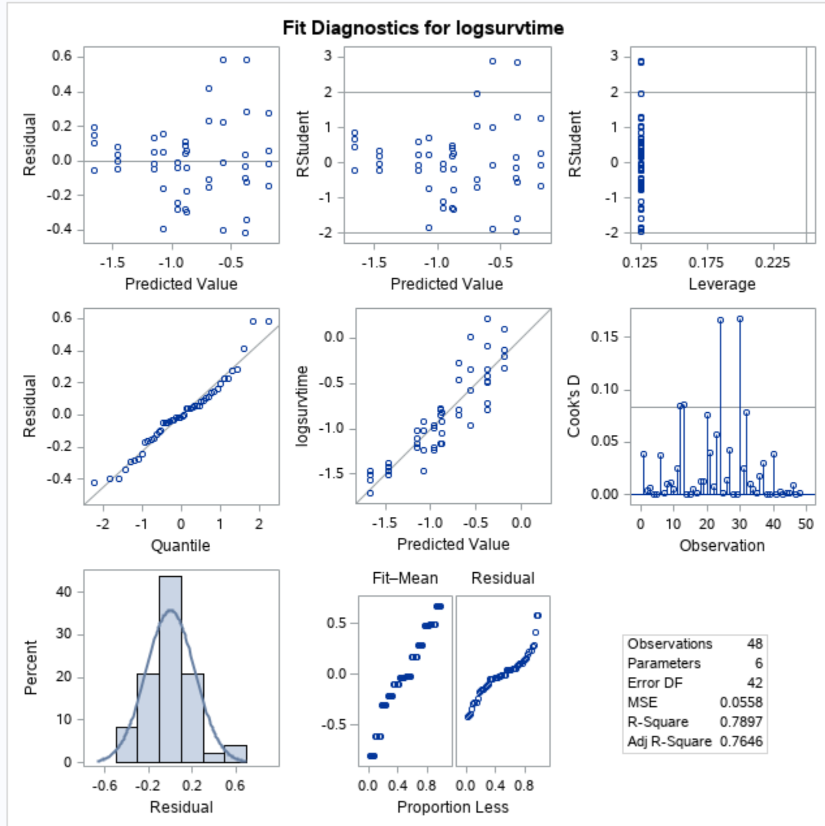
Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	2	5.23747262	2.61873631	46.96	<.0001
B	3	3.55717347	1.18572449	21.26	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	2	5.23747262	2.61873631	46.96	<.0001
B	3	3.55717347	1.18572449	21.26	<.0001

b.

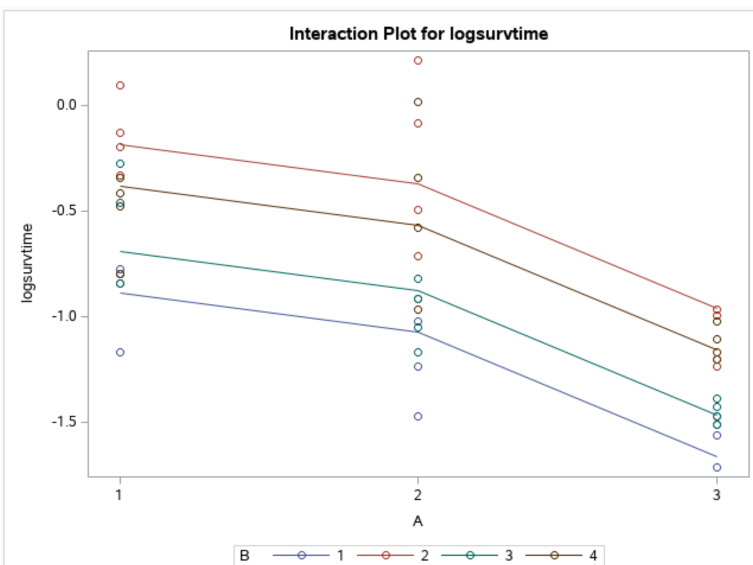
```
proc glm data=survtime plot=diagnostics;
	class A B;
	model logsurvtime = A B; run;
```

The diagnostics plot is shown beside. From the residual plot, we can observe the fact that the variances for error might not be constant because we do not see a random pattern in this plot. From the studentized, leverage and Cook's D plot, we see that there are 2 outliers in this data, and we might want to examine these 2 values and determine if they should be removed or not. For the normality, the Q-Q plot seems to suggest that the data is normally distributed.



c.

The interaction plot is shown below. As (b) suggested, there is no interaction between A and B, thus the mean lines are parallel.



d.

```
proc glm data=survtime plot=diagnostics;
    class A B;
    model logsurvtime = A B;
```

```
    estimate 'A1-A2' A 1 -1 0;
run;
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
A1-A2	0.18666302	0.08349263	2.24	0.0307

The point estimate for $\alpha1 - \alpha2$ is 0.1867, and the 95% confidence interval for it is [0.18666302-1.96*0.08349263, 0.18666302+1.96*0.08349263] = [0.023, 0.350]. Use exponential to back transform to the original scale, the point estimate for $\alpha1 - \alpha2$ will be 1.205 and the 95% confidence interval for it will be [1.023, 1.419]. That is to say, the difference of survival time in means for poison 1 and poison 2 is 1.205*10 = 12.05 hours, and its 95% confidence interval is [10.23, 14.19] hours.

e.

```
proc glm data=survtime plot=diagnostics;
    class A B;
    model logsurvtime = A B;
    estimate '(B1-B2)=(B3-B4)' B 1 -1 -1 1;
run;
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
(B1-B2)=(B3-B4)	-0.39429349	0.13634289	-2.89	0.0060

The point estimate for $\beta1 - \beta2 - \beta3 + \beta4$ is -0.394, and the 95% confidence interval for it is [-0.39429349-1.96*0.13634289, -0.39429349+1.96*0.13634289] = [-0.662, -0.127]. Use exponential to back transform to the original scale, the point estimate for $\beta1 - \beta2 - \beta3 + \beta4$ will be 0.674 and the 95% confidence interval for it will be [0.516, 0.881]. That is to say, the difference of survival time in means for poison 1 and poison 2 is 0.674*10 = 6.74 hours, and its 95% confidence interval is [5.16, 8.81] hours.

3.

$\mu = 10$, $\alpha1$ = -2.333, $\alpha2$ = 2.333, $\beta1$ = 5, $\beta2$ = -1, $\beta3$ = -4

$(\alpha\beta)11$ = 13-10+2.333-5 = 0.333

$(\alpha\beta)12$ = 6-10+2.333+1 = -0.667

$(\alpha\beta)13$ = 4-10+2.333+4 = 0.333

$(\alpha\beta)21$ = 17-10-2.333-5 = -0.333

$(\alpha\beta)22$ = 12-10-2.333+1 = 0.667

$(\alpha\beta)23$ = 8-10-2.333+4 = -0.333