hw8

AUTHOR
Hanbei Xiong

Question 1 (From lab10.1)

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

Answer:

Figure: SS in balanced data

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

Figure: SS in unbalanced data

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

In unbalanced data, the SS for A, B, and A*B in both I SS and III SS are smaller than the corresponding values in balanced data which should be due to the fact that we exclude 4 observations in the unbalanced data. The SSI and SSIII in balanced data are equivalent because factors are independent but the SSI and SSIII in unbalanced data are not equivalent because factors are not independent.

1.2: 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.

Answers:

Figure: Balanced Table

		B		
		1	2	3
A	1	0.204 0.17 0.181	0.167 0.182 0.187	0.202 0.198 0.236
	2	0.257 0.279 0.269	0.283 0.235 0.26	0.256 0.281 0.258

Unadjusted means:

$A_{1.}$: $\frac{0.204+0.17+0.181+0.167+0.182+0.187+0.202+0.198+0.236}{9} \approx 0.192$

$A_{2.}$: $\frac{0.257+0.279+0.269+0.283+0.235+0.26+0.256+0.281+0.258}{9} \approx 0.264$

$B_{.1}$: $\frac{0.204+0.17+0.181+0.257+0.279+0.269}{6} \approx 0.227$

$B_{.2}$: $\frac{0.167+0.283+0.182+0.235+0.187+0.26}{6} \approx 0.219$

$B_{.3}$: $\frac{0.202+0.256+0.198+0.281+0.236+0.258}{6} \approx 0.239$

adjusted mean:

$$\mu_{11} = \frac{0.204+0.17+0.181}{3} \approx 0.185$$

$$\mu_{12} = \frac{0.167+0.182+0.187}{3} \approx 0.179$$

$\mu_{13} = \frac{0.202+0.198+0.236}{3} \approx 0.212$

$\mu_{21} = \frac{0.257+0.279+0.269}{3} \approx 0.268$

$\mu_{22} = \frac{0.283+0.235+0.26}{3} \approx 0.259$

$\mu_{23} = \frac{0.256+0.281+0.258}{3} \approx 0.265$

$A_{1.} = \frac{0.185+0.179+0.212}{3} = 0.192$

$A_{2.} = \frac{0.268+0.259+0.265}{3} = 0.264$

$B_{.1} = \frac{0.185+0.268}{2} = 0.227$

$B_{.2} = \frac{0.179+0.259}{2} = 0.219$

$B_{.3} = \frac{0.212+0.265}{2} = 0.239$

Figure 2: Unbalanced Table

		B		
		1	2	3
A	1	0.204 0.17	0.167 0.182 0.187	0.202
	2	0.257 0.279	0.283 0.235 0.26	0.256 0.281 0.258

Unadjusted means:

$A_{1.}$: $\frac{0.204+0.17+0.167+0.182+0.187+0.202}{6} \approx 0.185$

$A_{2.}$: $\frac{0.257+0.279+0.283+0.235+0.26+0.256+0.281+0.258}{8} \approx 0.264$

$B_{.1}$: $\frac{0.204+0.17+0.257+0.279}{4} = 0.2275$

$B_{.2}$: $\frac{0.167+0.283+0.182+0.235+0.187+0.26}{6} = 0.219$

$B_{.3}$: $\frac{0.202+0.281+0.256+0.258}{4} = 0.24925$

adjusted mean:

$$\mu_{11} = \frac{0.204+0.17}{2} = 0.187$$

$$\mu_{12} = \frac{0.167+0.182+0.187}{3} \approx 0.179$$

$$\mu_{13} = \frac{0.202}{1} = 0.202$$

$$\mu_{21} = \frac{0.257+0.279}{2} = 0.268$$

$$\mu_{22} = \frac{0.283+0.235+0.26}{3} \approx 0.259$$

$$\mu_{23} = \frac{0.256+0.281+0.258}{3} \approx 0.265$$

$$A_{1.} = \frac{0.187+0.179+0.202}{3} = 0.189$$

$$A_{2.} = \frac{0.268+0.259+0.265}{3} = 0.264$$

$$B_{.1} = \frac{0.187+0.268}{2} = 0.2275$$

$$B_{.2} = \frac{0.179+0.259}{2} = 0.219$$

$$B_{.3} = \frac{0.202+0.265}{2} = 0.2335$$

1.3: 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.

Answers:

Figure: Code

```
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;
```

Figure: Output

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

Question 2

Two-way ANOVA: This problem uses the survtime dataset, which is from a 1964 paper by Box and Cox. The observations are from a balanced 3x4 ANOVA experiment on survival time of animals in which the factors are poison (3 types) and treatment (4 types). The outcome variable is survival time in units of 10 hours.

a. Using a natural log transformation of survival time, conduct a two-way ANOVA. First, fit a full model with an interaction. If the interaction is not significant, it can be dropped. (Hint: Drop it.)
b. Using the selected model from (a), examine model diagnostics to determine whether the model assumptions are reasonably met.
c. Produce a means plot (also called an interaction plot).
d. Obtain a point estimate and 95% confidence interval for $\alpha_1 - \alpha_2$ and backtransform to the original scale. How do you interpret this quantity?
e. Conduct a test of whether the difference in means (on the log-transformed scale) for treatments 1 and 2 is equal to the difference in means for treatments 3 and 4, i.e., $H_0: \beta_1 - \beta_2 = \beta_3 - \beta_4$. Obtain a point estimate and 95% confidence interval for $\beta_1 - \beta_2 - \beta_3 + \beta_4$.

(a) Answers:

Figure: Two way ANOVA with interaction

Source	DF	Type I 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

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

We drop the interaction term because it is not significant. Here is the model after dropping the interaction term.

Figure: Two way anova without 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) Answers:

Figure: Diagnostics Plots

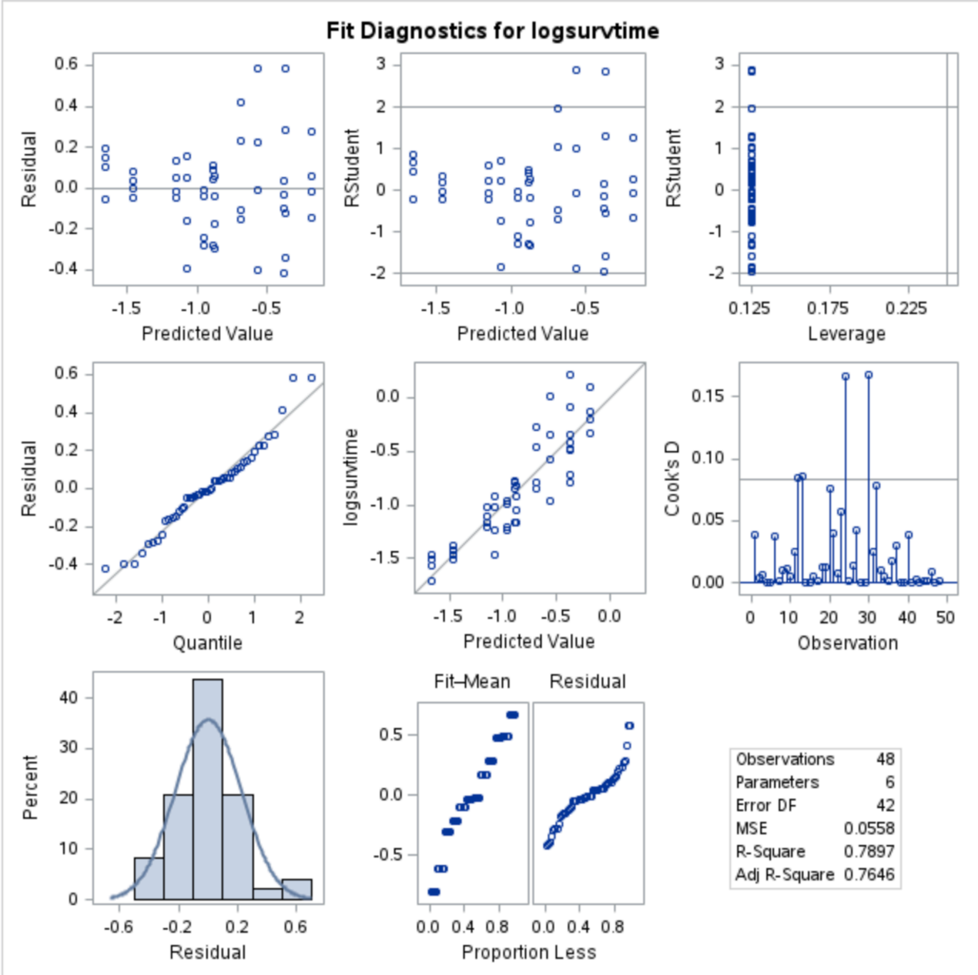


Figure: Distribution of B

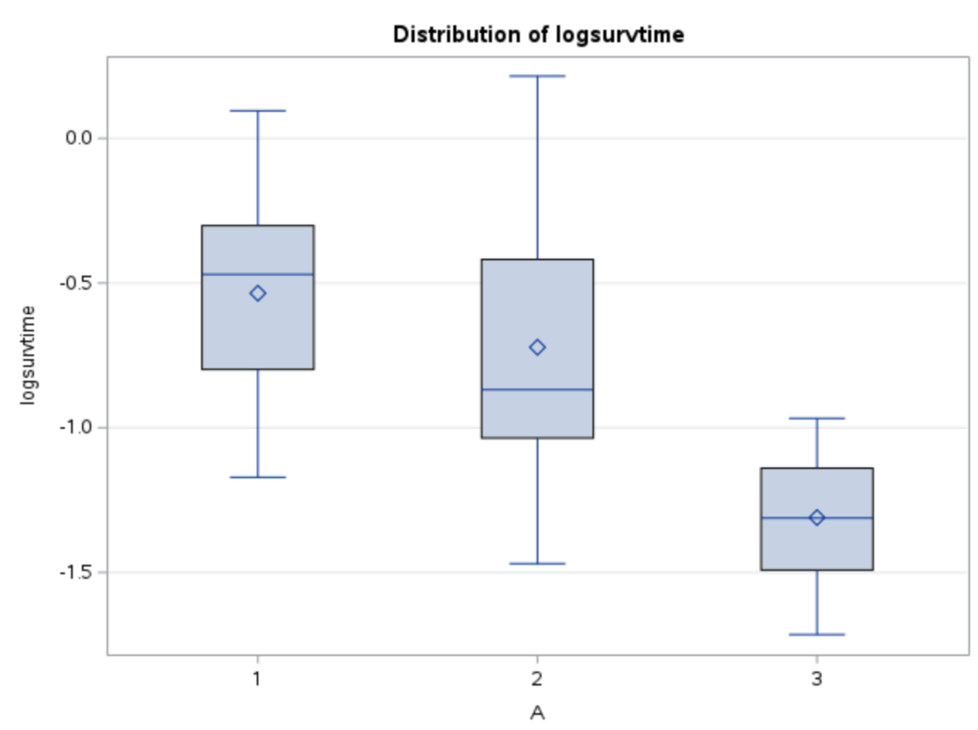
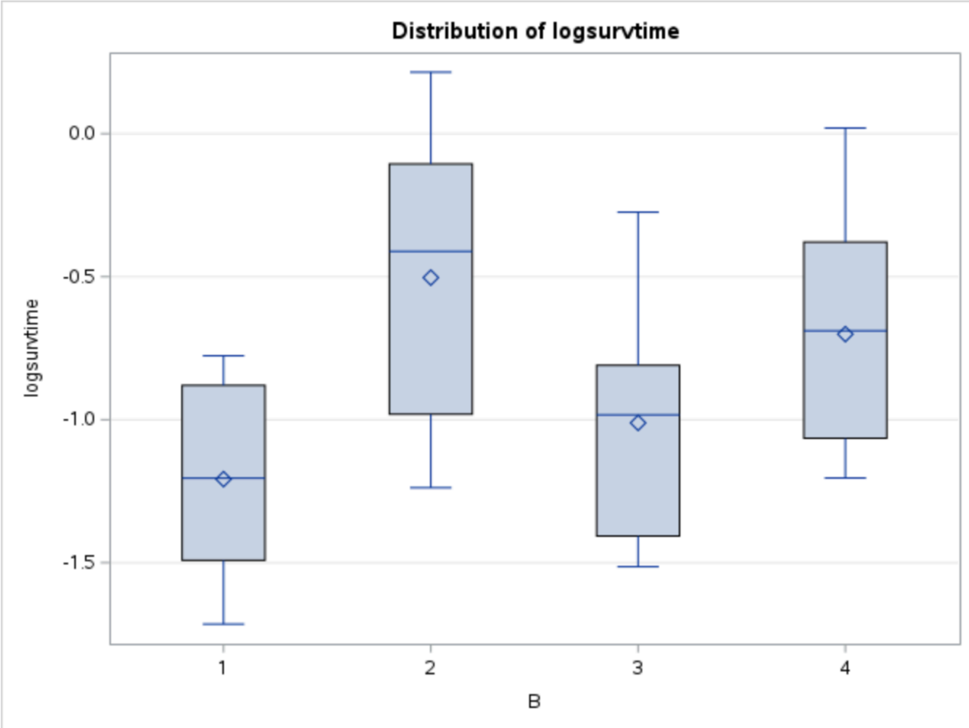
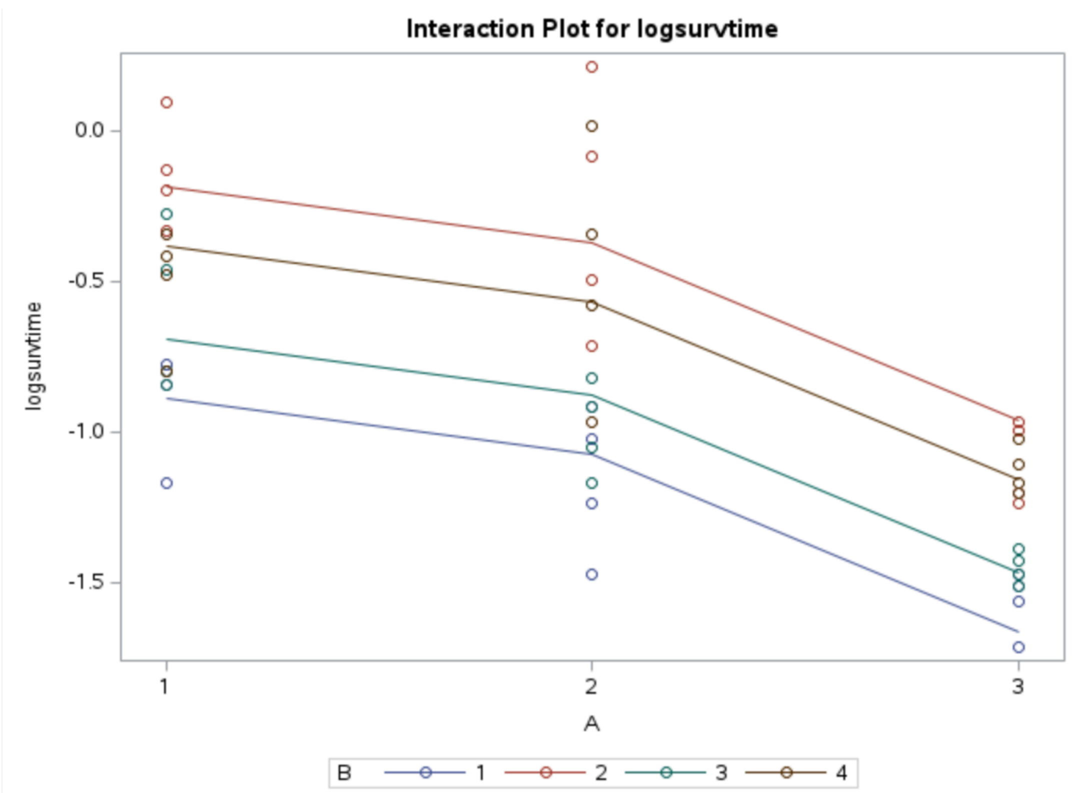


Figure: Distribution of A



By observing the above plots, we can see the normality is met but the constancy of variance across group is not met.

(c) Answers:



(d) Answers:

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
A1-A2	0.18666302	0.08349263	2.24	0.0307	0.01816807	0.35515797

$e^{0.187} \approx 1.206$

The median survival time for animal is about 20 percent higher for poison 1 compared to poison 2.

We are 95% confident that the true medians of animals between poison group 1 and poison group 2 lies in the interval (0.0182, 0.3552) on the original scale of survtime.

(e) Answers:

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
(1-2) - (3-4)	-0.39429349	0.13634289	-2.89	0.0060	-0.66944459	-0.11914239

The point estimate is roughly -0.394. The 95% confidence interval is (-0.669, -0.119)

Question 3

3. Consider the table of cell means from a balanced 2-way ANOVA experiment. Assume a two-way ANOVA model with interactions,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim indep\ N(0, \sigma^2)$$

with the constraints

$$\sum_i \alpha_i = 0, \quad \sum_j \beta_j = 0, \quad \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$$

Using this table of means, obtain estimates of the following quantities:
$\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, (\alpha\beta)_{11}, (\alpha\beta)_{12}, (\alpha\beta)_{13}, (\alpha\beta)_{21}, (\alpha\beta)_{22}, (\alpha\beta)_{23}$

		Factor B		
		1	2	3
Factor A	1	13	6	4
	2	17	12	8

Answer:

Here are quantities we need to use:

$\bar{Y}_{1..} = \frac{13+6+4}{3} \approx 7.67 \; \bar{Y}_{2..} = \frac{17+12+8}{3} \approx 12.33$

$\bar{Y}_{.1.} = \frac{13+17}{2} = 15 \; \bar{Y}_{.2.} = \frac{6+12}{2} = 9 \; \bar{Y}_{.3.} = \frac{4+8}{2} = 6$

Here are what we are looking for:

$\mu = \bar{Y}_{...} = \frac{13+6+4+17+12+8}{6} = 10$

$$\alpha_1 = \bar{Y}_{1..} - \bar{Y}_{...} = 7.67 - 10 = -2.33$$

$$\alpha_2 = \bar{Y}_{2..} - \bar{Y}_{...} = 12.33 - 10 = 2.33$$

$$\beta_1 = \bar{Y}_{.1.} - \bar{Y}_{...} = 15 - 10 = 5$$

$$\beta_2 = \bar{Y}_{.2.} - \bar{Y}_{...} = 9 - 10 = -1$$

$$\beta_3 = \bar{Y}_{.3.} - \bar{Y}_{...} = 6 - 10 = -4$$

$$(\alpha\beta)_{11} = \bar{Y}_{11} - (\mu + \alpha_1 + \beta_1) = 13 - (10 - 2.33 + 5) = 0.33$$
$$(\alpha\beta)_{12} = \bar{Y}_{12} - (\mu + \alpha_1 + \beta_2) = 6 - (10 - 2.33 - 1) = -0.67$$
$$(\alpha\beta)_{13} = \bar{Y}_{13} - (\mu + \alpha_1 + \beta_3) = 4 - (10 - 2.33 - 4) = 0.33$$
$$(\alpha\beta)_{21} = \bar{Y}_{21} - (\mu + \alpha_2 + \beta_1) = 17 - (10 + 2.33 + 5) = -0.33$$
$$(\alpha\beta)_{22} = \bar{Y}_{22} - (\mu + \alpha_2 + \beta_2) = 12 - (10 + 2.33 - 1) = 0.67$$
$$(\alpha\beta)_{23} = \bar{Y}_{23} - (\mu + \alpha_2 + \beta_3) = 8 - (10 + 2.33 - 4) = -0.33$$