

Stat 539 - Biostatistics I – Spring 2006

Homework 6 due March 9 **Solutions**

(assignment for these solutions can be found on the last page)

The code for this analysis comes directly from the labs website
http://www.stat.unm.edu/~erike/courses/stat539/stat539_lab7.do
 essentially just doing a find/replace of the three variables used.

I run each section of code, then discuss the results.

```
clear
* load the dataset
use stat539_hw6_stroke_lesion_data
sort hemisphere lesionsize score
list

* table of means, standard deviations, and frequencies
tabulate lesionsize hemisphere, summarize(score) means
```

Means of score

lesionsize	hemisphere		Total
	1	2	
2	10.111111	9.9285714	10
3	10.2	10.777778	10.571429
4	13	11.166667	11.705882
Total	10.894737	10.571429	10.685185

The table of means shows that almost all of the means are around 10, except for lesionsize 4 and hemisphere 1 with a 13.

```
* fit the two-way anova with interaction
anova score lesionsize hemisphere lesionsize*hemisphere
```

	Number of obs =	54	R-squared =	0.1774	
	Root MSE =	2.00956	Adj R-squared =	0.0917	
Source	Partial SS	df	MS	F	Prob > F
Model	41.8084656	5	8.36169312	2.07	0.0855
lesionsize	37.6513418	2	18.8256709	4.66	0.0141
hemisphere	2.66172479	1	2.66172479	0.66	0.4209
lesionsize*hemisphere	10.5400912	2	5.27004558	1.31	0.2806
Residual	193.839683	48	4.03832672		
Total	235.648148	53	4.44619147		

The ANOVA table indicates that there is no interaction between lesionsize and hemisphere. As we will see in the interaction plots below, there is indeed an interaction – which I pointed out in the table of means.

```
* create fitted values, yhat and residuals, residual
predict yhat, xb
predict residual,r
```

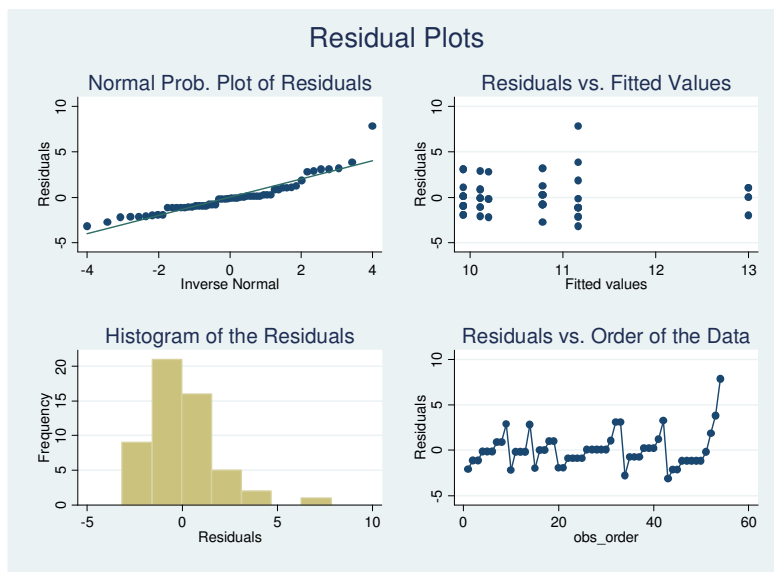
```

* check for equal variances by hemisphere
robvar(score), by(hemisphere)
* check for equal variances by lesionsize
robvar(score), by(lesionsize)

* Create a four-in-one plot
quietly qnorm residual, name(probplot,replace) nodraw title(Normal
Prob. Plot of Residuals)
quietly rvfplot, name(respredplot,replace) nodraw
title(Residuals vs. Fitted Values)
quietly hist residual, freq name(hist,replace) nodraw
title(Histogram of the Residuals)
generate obs_order = _n
quietly twoway connect residual obs_order, name(obs_order,replace) nodraw
title(Residuals vs. Order of the Data)
drop obs_order
graph combine probplot respredplot hist obs_order,
title(Residual Plots)

```

Skipping the tests for equal variance for now and turning our attention to the residual plots, we note one outlier in the right tail. This observation has an unawareness score of 19, much larger than any other score.



```

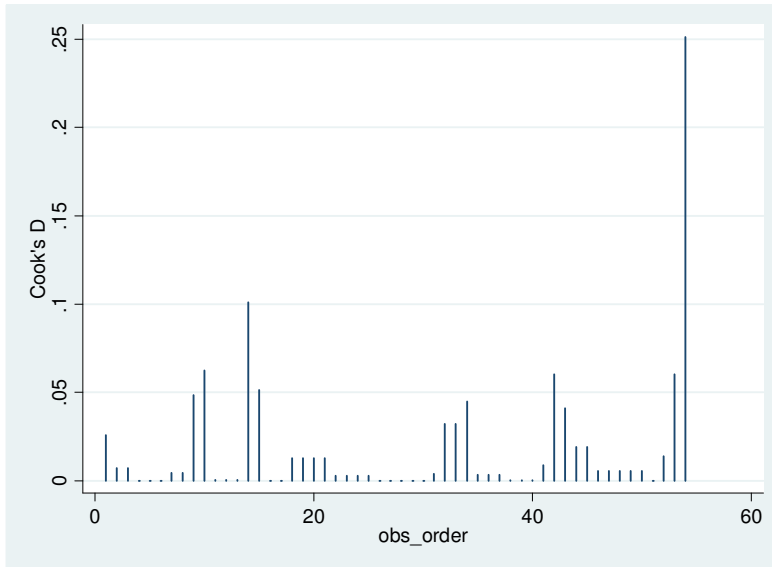
* print the Shapiro-Wilks normality test results
swilk residual

```

```

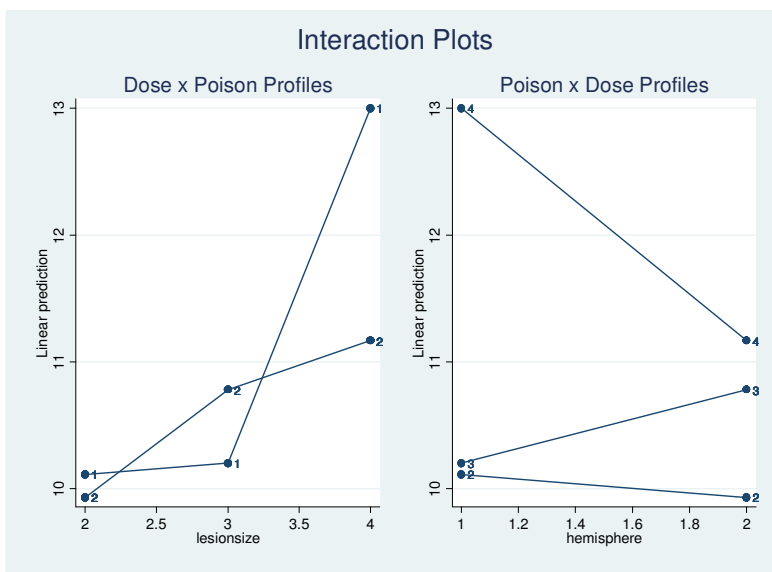
* Cook's distance to examine if any observations have great influence on the
regression
predict cooks,cooks
gene obs_order = _n
twoway spike cooks obs_order
drop obs_order cooks

```



Also, the influence of that one observation on the model is much greater than the other observations.

```
* interaction plot, hemisphere is lines, lesionsize is horizontal axis
sort hemisphere lesionsize
scatter yhat lesionsize, connect(ascending) mlabel(hemisphere) name(iplot1,replace)
nodraw title(Dose x Poison Profiles)
* interaction plot, hemisphere is lines, lesionsize is horizontal axis
sort lesionsize hemisphere
scatter yhat hemisphere, connect(ascending) mlabel(lesionsize) name(iplot2,replace)
nodraw title(Poison x Dose Profiles)
* two plots together
graph combine iplot1 iplot2, title(Interaction Plots)
```



Do these lines look parallel? No, absolutely not! So there is interaction. Look back at the ANOVA table where it indicates that the interaction is not significant. This is quite a

contradiction. This is the doing of the outlier. Let's remove the outlier (score=19) and repeat the analysis.

```
*****
* remove outlier (score=19) and repeat analysis above
drop in 54
list

*** Go back to the table of means above and repeat...

* table of means, standard deviations, and frequencies
tabulate lesionsize hemisphere, summarize(score) means
```

Means of score

lesionsize	hemisphere		Total
	1	2	
2	10.111111	9.9285714	10
3	10.2	10.777778	10.571429
4	13	10.454545	11.25
Total	10.894737	10.323529	10.528302

This table has not changed much, the lesionsize 4, hemisphere 2 has decreased from 11.16 to 10.45. The big advantage is the interaction that we'll see again in the interaction plot below now appears in the ANOVA table.

```
* fit the two-way anova with interaction
anova score lesionsize hemisphere lesionsize*hemisphere
```

	Number of obs =	53	R-squared =	0.2319
	Root MSE =	1.64317	Adj R-squared =	0.1502

Source	Partial SS	df	MS	F	Prob > F
Model	38.3072586	5	7.66145171	2.84	0.0255
lesionsize	24.9690109	2	12.4845054	4.62	0.0147
hemisphere	5.89302471	1	5.89302471	2.18	0.1462
lesionsize*hemisphere	18.3925075	2	9.19625375	3.41	0.0416
Residual	126.900289	47	2.70000614		
Total	165.207547	52	3.17706821		

Now the interaction is significant (at a 0.05 level). There also appears to be a main effect due to lesionsize.

```
* create fitted values, yhat and residuals, residual
drop yhat
drop residual
predict yhat, xb
predict residual,r

* check for equal variances by hemisphere
robvar(score), by(hemisphere)
* check for equal variances by lesionsize
robvar(score), by(lesionsize)
```

hemisphere	Summary of score		
	Mean	Std. Dev.	Freq.
1	10.894737	1.64317	10
2	10.323529	1.64317	10

1	10.894737	1.9117978	19
2	10.323529	1.7006343	34
Total	10.528302	1.7824332	53

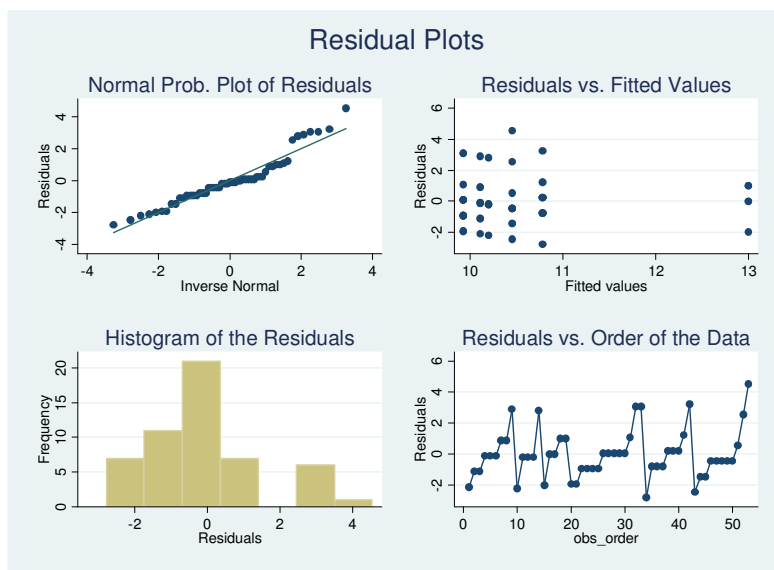
W0 = 1.0077365 df(1, 51) Pr > F = .32018504
W50 = .98252855 df(1, 51) Pr > F = .32625507
W10 = 1.4506962 df(1, 51) Pr > F = .23397622

lesionsize	Summary of score		
	Mean	Std. Dev.	Freq.
2	10	1.4770979	23
3	10.571429	1.6508406	14
4	11.25	2.113449	16
Total	10.528302	1.7824332	53

W0 = 2.7814534 df(2, 50) Pr > F = .07155272
W50 = 1.7733544 df(2, 50) Pr > F = .18027295
W10 = 2.6293606 df(2, 50) Pr > F = .08207849

All tests for unequal variance are nonsignificant, so the assumption of equal variance appears to be met.

```
* Create a four-in-one plot
quietly qnorm residual, name(probplot,replace) nodraw title(Normal
Prob. Plot of Residuals)
quietly rvfplot, name(respredplot,replace) nodraw
title(Residuals vs. Fitted Values)
quietly hist residual, freq name(hist,replace) nodraw
title(Histogram of the Residuals)
generate obs_order = _n
quietly twoway connect residual obs_order, name(obs_order,replace) nodraw
title(Residuals vs. Order of the Data)
drop obs_order
graph combine probplot respredplot hist obs_order,
title(Residual Plots)
```



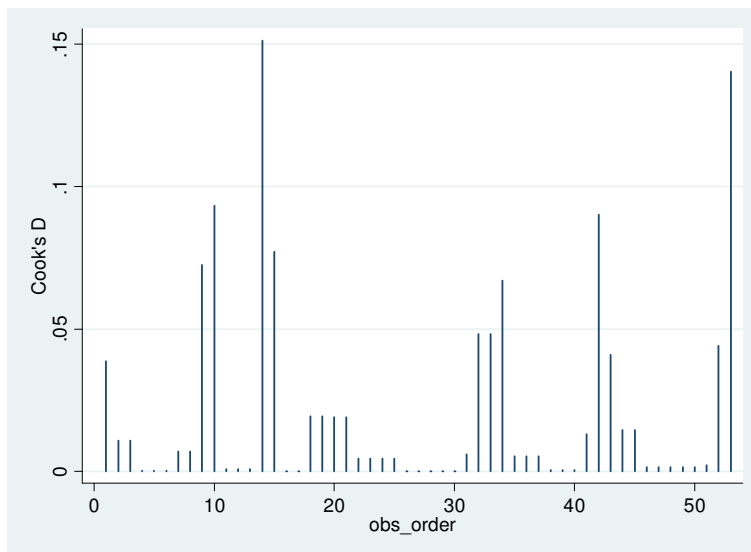
The residual plot appears much better now. There is a little jump in the histogram and the normal plot, but that may be in part that the responses are integers and do not take any real number. These residuals do not look bad, so we'll continue.

```
* print the Shapiro-Wilks normality test results
swilk residual
```

Variable	Shapiro-Wilk W test for normal data				
	Obs	W	V	z	Prob>z
residual	53	0.93339	3.280	2.542	0.00551

The residuals are not normal based on the Shapiro-Wilk test. We'll continue in light of this failure of the normality assumption. While the data are not normal, they are roughly symmetric and unimodal. Without looking too closely why it failed normality, it is likely because the distribution is too peaked and slightly right skewed.

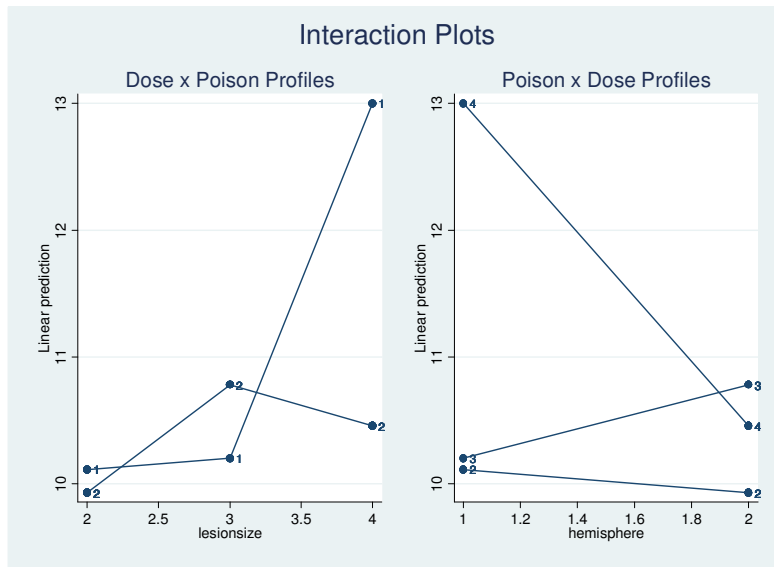
```
* Cook's distance to examine if any observations have great influence on the
regression
predict cooks_d,cooks_d
gene obs_order = _n
twoway spike cooks_d obs_order
drop obs_order cooks_d
```



There no longer appears to be any observations having a much greater influence on the model than others.

```
* interaction plot, hemisphere is lines, lesionsize is horizontal axis
sort hemisphere lesionsize
scatter yhat lesionsize, connect(ascending) mlabel(hemisphere) name(iplot1,replace)
nodraw title(Dose x Poison Profiles)
* interaction plot, hemisphere is lines, lesionsize is horizontal axis
sort lesionsize hemisphere
scatter yhat hemisphere, connect(ascending) mlabel(lesionsize) name(iplot2,replace)
nodraw title(Poison x Dose Profiles)
* two plots together
```

```
graph combine iplot1 iplot2, title(Interaction Plots)
```



The interaction is still present, which was significant in the ANOVA (now that the score=19 observation is removed). Now the ANOVA and interaction plot is showing the same information. It is important for the evidence we find in our analysis to point in the same general direction. It is important for one observation to not unduely influence the analysis.

Based on the interaction plots, it does not appear that there is any real main effect. The ANOVA indicates there is a main effect due to lesionsize. However, look at the left lesionsize interaction plot. For left(1) hemisphere, the response increases as lesionsize increases from 2 to 3 to 4. However, for right(2) hemisphere, the response increases as lesionsize goes from 2 to 3, but decreases from 3 to 4. Everything that is going on appears to be in the interaction.

This is one of those pure interaction examples where lesionsize doesn't make any real difference in the right(2) hemisphere but makes a huge difference (if lesionsize is big enough = 4) in the left(1) hemisphere.

Stat 539 Spring 2006 Homework 6 due March 9

Do a complete analysis of the data for effects of lesion size, hemisphere affected, and interactions between those effects.

The aim of a study by Hartman-Maier et al. (A-33) was to evaluate the awareness of deficit profiles among stroke patients undergoing rehabilitation. She studied 35 patients with a stroke lesion in the right hemisphere and 19 patients with a lesion on the left hemisphere. She also grouped lesion size as

2 = 1–3 cm, 3 = 3–5 cm, and 4 = 5 cm or greater.

One of the outcome variables was a measure of each patient's total unawareness of their own limitations. Scores ranged from 8 to 24, with higher scores indicating more unawareness.

Lesion Size Group	Unawareness Score		
	Left Hemisphere	Right Hemisphere	
2	11	10	8
	13	11	10
	10	13	9
	11	10	9
	9	13	9
	10	10	
	9	10	
	8	9	
	10	8	
3	13	11	10
	8	10	11
	10	10	12
	10	14	11
	10	8	
4	11	10	11
	13	13	9
	14	10	19
	13	10	10
	14	15	9
		8	10

SOURCE: Adina Hartman-Maier, Ph.D., O.T.R. Used with permission.