36-201 Lab #7 – Partial Solutions

Question #2: This is an experiment, since the researchers controlled who got which treatment.

Question #3: Imip: 38/109 = 35%; Li: 37/109 = 34%; Pl: 34/109 = 31%

♣ **Question #4:** The randomization does look as if it was effective in assigning approximately an equal number of patients per treatment group.

Question #6: Aside from AcuteT, we should also check that

- GAS score
- Age
- Gender

are distributed equally among the treatment groups by the randomization. It would also be interesting to know whether the hospitals differ greatly in GAS, age, gender or AcuteT; if so this might mean we are running the risk of a simpson's paradox phenomenon if we analyze the hospitals all together—especially if we also discover that GAS, age, gender or AcuteT are associated with success or time to recurrence of depression!

Question #7: OUTCOME is response, qualitative. TREAT is explanatory, qualitative.

Question #8: A contingency table is an effective method of looking at the relationship between two qualitative variables.

Question #9:

	Observed Table					Expected Table		
	Imip	Li	Pl	Total		Imip	Li	Pl
Success	27 (71%)	14 (38%)	11 (32%)	52	S	18.1	17.7	16.2
Failure	11 (29%)	23 (62%)	23 (68%)	57	F	19.9	19.3	17.8
Total	38 (100%)	37 (100%)	34 (100%)	109		38	37	34

♣ **Question #11:** It looks as if Imip (71% success rate) is more effective than either Li (38%) or Pl (32%). Also it looks as if there is very little difference in effectiveness between Li and Pl.

Question #12: The standardized residuals look like this:

```
    Imip     Li    Pl
S   2.08  -0.87 -1.30
F  -1.99   0.83  1.24
```

The largest standardized residual is 2.08, which suggests there are significantly more successful imiprimine patients than there would be if success/failure were indepenent of treatment. The smallest, -1.99, similarly suggests there were too few failures on imiprimine, relative to what would be expected under independence.

Question #13: TIME is response, quantitative. TREAT is explanatory, qualitative.

Question #14: Side by side boxplots are good for comparing quantitative responses with qualitative explanatory variables.

♣ **Question #15:** Median time in weeks: Imip: 71; Li: 22; Pl: 18. Based on the median time followed in the study more than 50% of the patients on Imip were followed longer than 71 weeks. Furthermore Q_1=21

for the Imip group which is about equal to the median time followed in the other two groups. Since longer times followed in the study are evidence that the treatment prevented the recurrence of depression we would conclude based on this analysis that Imip was more effective than Li or Pl. Interestingly, the distribution of time in the study looks about the same in the Li and the Pl groups.

Question #17:

(I) Imiprimine was more effective in preventing recurrence.

(II) Imiprimine was also more effective in delaying the recurrence of depression.

♣ **Question #18:** Both analyses point to imiprimine: it is better at preventing recurrence of depression overall, and it is better at delaying a recurrence among patients who do have recurrences.

Question #19: (i) Suppose just by bad luck the more severely ill patients ended up in the Li and Pl group. Then we may be observing poorer performance on these two treatments simply because the patients were sicker. (ii) Suppose older patients have more recurrences of depression. Then it might appear that Imip was the more effective treatment simply because more older patients were in the Li and Pl groups and the younger patients were all in the Imip group.

Both of these examples as well as others are examples of confounding variables that could explain the relationship between "Treatment" and "Outcome". However, we hope that by having randomized patients to treatments we have controlled for the effects of confounding or lurking variables by distributing these variables equally across the treatment groups and thereby eliminating systematic differences between the groups.