Exam 1 Sample
KEY

STA209-04: Applied Statistics

February 15, 2019

1) [10 pts] As the consulting statistician within a research hospital, you are approached by a physician interested in conducting a study to compare the efficacy of two kidney stone treatments. The first treatment, Treatment A, is an invasive surgical procedure whereas the second treatment (B) involves only a small incision. Both treatments are known to be efficacious, but interest is in determining which of the two is more effective. Recognizing their own lack of statistical training, the physician is looking to you for guidance and has prepared several questions she feels are important to get started:

a) Which study design - observational study or randomized experiment - do you recommend to determine which of these two treatments leads to an improved outcome? Why?

A randomized experiment should be recommended. In contrast to an observational study, a randomized experiment mitigates bias that might arise from one or more confounding factors. Since we are comparing two treatments, we want to be able to claim any difference in efficacy are attributed directly to the treatment and not some lurking variable.

b) There are a few patients of mine that have shared with me an interest in receiving Treatment B. They've also provided me with their consent to participate in the study. Should I include them in my study and administer them treatment B? More generally, how should I recruit study participants?

These patients should not be recruited and assigned to their preferred treatment. Doing so would bias results since patients who prefer Treatment B will likely view the results of the procedure favorably (i.e. confirmation bias). Simple random sampling should be used in recruiting patients.

c) Which population(s) can I generalize my results to?

Assuming that simple random sampling was used, the study results may be generalized to kidney stone patients at that hospital/surrounding area.

2) [20 pts] A few months after your initial consultation, the physician returns with the following data:

	Treatment A		**Treatment B**	
Kidney Stone Size	**# Improved**	**Total**	**# Improved**	**Total**
Small	81	87	234	270
Large	192	263	55	80

Hoping that you could impart more of your statistical wisdom and expertise, the physician asks you to tell her which treatment is supported by the data. Towards this end, you consider the following questions:

a) What are the cases for these data? What are the variables?

The cases are the kidney stone patients and the variables are stone size, treatment, and whether the patient improved.

b) What would be a good way to visualize these data?

A clustered bar chart capturing each level of treatment and stone size would be a good way of visualizing these data.

c) What are the marginal proportions for each treatment's success? What can we conclude based on these proportions?

The marginal success proportions are: $p_A = (81 + 192)/(87 + 263) = 0.78$; $p_B = (234 + 55)/(270 + 80) = 0.83$. From these proportions, we could conclude that Treatment B is more effective.

d) What are the success proportions for each treatment when we consider large and small kidney stones separately? What conclusion(s) can we draw from these proportions?

The kidney stone size conditional success proportions for each treatment are: $p_{SA} = (81)/(87) = 0.93$; $p_{SB} = (234)/(270) = 0.87$; $p_{SA} = (192)/(263) = 0.73$; $p_{SB} = (55)/(80) = 0.69$. For each type of kidney stone, Treatment A has a higher success proportions and thus is more effective.

e) Based on these analyses, what would you tell the physician? Is there evidence that one treatment is more effective than the other? If so, which treatment and why?

There is evidence to support the conclusion that Treatment A is more effective than Treatment B. Initial analyses indicated the reverse, but these results were confounded by the severity of kidney stone (i.e. kidney stone size). When we stratified by kidney size, we see that Treatment A yields higher success rates and is therefore more effective.