## A4. EXERCISES FROM CHAPTER 4

1. Suppose that an author reports that hospital in-patients who watch T.V. in their room rate their stay an average of two points higher than those who do not. There are two clear problems with this statistical conclusion. Identify them.

   First, the item for rating here is almost certainly a Likert item. This results in ordinal data; taking an average is inappropriate and the median should be used.

   Second, even if averages were appropriate, the author has incorporated the sample average (two) as if it will automatically be the same as the population average. It won't. So instead, the author should be providing a confidence interval instead, perhaps indicating that it is between 1.3 and 2.7 points higher with 95% confidence.

2. A study on the impact of exercise on soda intake finds, with 95% confidence, exercisers take in on average 40 to 80 more ounces of soda per day, as compared to non-exercisers. Identify at least three other things that one would need to know in order to clinically interpret this result.

   Among the information we would be interested in: Should 40 oz. be considered a lot of soda? Is 80 oz. a small enough amount of soda that we wouldn't be interested? How do these numbers compare to overall amounts? (The perception is a lot different depending on whether the overall average amount of soda in a day is 50 oz., 500 oz., or 5000 oz.) Also, it seems clear that too much soda is probably bad, but how much soda is "too much" in a day? Confidence intervals for each group individually would helpful to know whether either group average exceeds that amount.

3. A researcher reports that, at significance level 0.05, she has found evidence that the average salt intake for those who exercise is lower than the average salt intake for those who do not. Because of this fact, she states *that exercise is necessary in order to lower salt intake*. There are a number of reasons that her conclusion does not follow from her statistical analysis. Identify at least two of them.

   The biggest issue here is that the researcher is likely using observational data to draw a cause-effect conclusion. You simply can't do that. It may simply be that exercise and salt intake are both related to a third factor (one's general outlook on health comes to mind). Those who are "health conscious" probably both try to get more exercise and at the same time not consume too much salt. A secondary issue is that even if a cause effect relationship exists, perhaps it goes the other way. Maybe lower salt intake leads one to a greater desire for exercise. Nothing about this observational study will be able to clarify this for us. In order to consider cause/effect, we'd need an experimental design in which we randomly assign people to exercise groups (and ensure they do what we've assigned and only that) and then measure their salt intake.

4. Authors often provide a lengthy table of demographic information for their study groups. Along with this, they may include p-values for two-sample tests. Unless an author is primarily interested in comparing the groups across certain demographics, these p-values are unnecessary. Explain why.

   Such demographics are provided because it is often important that the groups are similar in various ways. When they are dissimilar in some extraneous way that relates to the response variable, confounding will result. For example, if the response variable is the time to elimination of symptoms, but the treatment group is substantially younger than the control group, it will be impossible to tell whether the treatment was effective (since it is entirely possible that the younger group would automatically get better more quickly without treatment).

   P-values are unnecessary here because it is **samples**, not **populations**, that are to be compared. In fact, in the example above there is really only one population, from which a sample has been divided into two parts to form comparison groups. Comparison of those two groups (you want to see similarities across other variables that could impact the response) can be done simply by looking at distributions (often via graphs) and sample statistics. In a perfect world, if we have a 40-year old in one group we'll also have a 40-year old in the other group to "match". Of course observational studies never turn out quite that nice, but randomization of participants to treatment groups should help avoid large differences.

5. You wish to make a comparison of two groups across an interval-ratio response variable. You have sample sizes of 17 and 22. If there is a difference between the groups it is very important that you find it. Would you utilize the T-test or the Mann Whitney test? Explain.

   Use the T-test here, since it is statistically more powerful. Recognize, however, that the normality assumption of the T-test might not quite be satisfied. Therefore your confidence will not be as high as expected; and your false positive rate would be higher. In all likelihood, you should consider this "pilot" data. If a difference is found (or suggested to be possible), you may want to do another study with larger samples to follow up.

6. Consider the research study discussed in Section 4.10.

   > Reference: Usher, K., Park, T., Foster, K., and Buettner, P. (2013). A randomized controlled trial undertaken to test a nurse-led weight management and exercise intervention designed for people with serious mental illness who take second generation antipsychotics. *Journal of Advanced Nursing*, 69(7),1539-1548.

   Address the following issues:

   a. It was noted in section 4.10.2 that the chi-square test is probably not appropriate for the fourth variable (medication compliance) in the table on page 1545 of the manuscript. What statistical procedure would be appropriate here? Explain.

Percentages for each of the two groups are being compared here. Most effective would be the two-sample Z-test for proportions, perhaps with a confidence interval. In Chapter 6 you will learn that the chi-square test will only establish whether improvement depends on the group. The two-sample proportions will allow us to develop a conclusion that would assess the strength of such dependence (and therefore carry greater weight in clinical interpretation).

b. Refer to part (a). Use an appropriate software to conduct the correct test and interpret your results.

A two-sided Z-test for proportions shows no evidence of a difference in the proportion who improve when comparing the treatment to control ($Z = 0.048$, p-value = 0.9616). The 95% confidence interval for the difference in proportions is (-0.15,+0.16). This basically tells us there could be up to a 15% difference in either direction (i.e. with either group having the greater success rate). If the difference were as high as 15%, that would probably be important. Hence a future study with larger samples is recommended.

c. Consider the BMI response variable. Compute and interpret a confidence interval (roughly 95%) that compares the changes for the intervention and control groups. Hint: See calculation in Section 4.10.2. This calculation should be quite similar.

The difference in sample means (intervention minus control) is 0.19. The standard error calculation is:

$$SE_{\text{mean difference}} = \sqrt{\frac{(1.34)^2}{51} + \frac{(1.17)^2}{50}} \approx 0.25$$

To get a roughly 95% confidence interval, add and subtract twice the standard error. This results in an interval from -0.39 to +0.61. **With BMI's typically in the 20's and 30's, such a small difference is may not be all that clinically relevant.**

d. In their fourth table (page 1544), consider the test comparing BMI. Give a statistical interpretation for the p-value for this test (0.601). Connect your interpretation to the fact that these groups were selected by random division of all available participants into two equal groups. Use this connection to explain why the use of statistical methods represents a contradiction in ideas.

There is no evidence of a difference between the control population and the intervention population when it comes to average BMI. Note already the problem with this statement! There is no such thing as a "control population" and an "intervention population". Rather, there is a single population, from which we have collected a single sample and randomly allocated our participants into two groups. These two groups may differ, but they would differ at the "sample" level. As such, caring only about the samples here, no statistical inference would be necessary.