Homework 6: Sections 5.1 - 6.5

STA209-04: Applied Statistics

Assigned: 03/08/2019
Due: 04/01/2019

From the Book:

Questions: 5.31, 5.46, 6.37, 6.60, 6.75, 6.107, 6.124, 6.145, 6.173, 6.195, 6.221, 6.255, 6.257

5.31 In a smoking cessation program, over 2000 smokers who were trying to quit were randomly assigned to either a group program or an individual program. After six months in the program, 148 of the 1080 in the group program were successfully abstaining from smoking, while 120 of the 990 in the individual program were successful. We wish to test to see if this data provide evidence of a difference in the proportion able to quit smoking in between smokers in a group program and smokers in an individual program.

a) State the null and alternative hypotheses, and give the notation and value of the sample statistic.

b) Use a randomization distribution and the observed sample statistic to find the p-value.

c) Give the mean and standard error of the normal distribution that most closely matches the randomization distribution, and then use this normal distribution with the observed sample statistic to find the p-value.

d) Use the standard error found from the randomization distribution in part (b) to find the standardized test statistic, and then use that test statistic to find the p-value using a standard normal distribution.

e) Compare the p-values from parts (b), (c), and (d). Use any of these p-values to give the conclusion of the test.

5.46 Hospital admissions for asthma in children younger than 15 years was studied in Scotland both before and after comprehensive smoke-free legislation was passed in March 2006. Monthly records were kept of the annualized percent change in asthma admissions. For the sample studied, before the legislation, admissions for asthma were increasing at a mean rate of 5.2% per year. The standard error for this estimate is 0.7% per year. After the legislation, admissions were decreasing at a mean rate of 18.2% per year, with a standard error for this mean of 1.79%. In both cases, the sample size is large enough to use a normal distribution.

a) Find and interpret a 95% confidence interval for the mean annual percent rate of change in childhood asthma hospital admissions in Scotland before the smoke-free legislation.

b) Find a 95% confidence interval for the same quantity after the legislation.

c) Is this an experiment or an observational study?

d) The evidence is quite compelling. Can we conclude cause and effect?

6.37 A survey of 1000 adults in the US conducted in March 2011 asked "Do you favor or oppose 'sin taxes' on soda and junk food?" The proportion in favor of taxing these foods was 32%.

a) Find a 95% confidence interval for the proportion of US adults favoring taxes on soda and junk food.

b) What is the margin of error?

c) If we want a margin of error of only 1% (with 95% confidence), what sample size is needed?

6.60 Can babies reason probabilistically? A study investigates this by showing ten- t twelve-month-old infants two jars of lollipop-shaped objects colored pink or black. Each infant first crawled or walked to whichever color they wanted, determining their "preferred" color. They were then given teh choice between two jars that had the same number of preferred objects, but that differed in their *probability* of getting the preferred color; each jar had 12 in the preferred color and either 4 or 36 in the other color. Babies choosing randomly or based on the absolute number of their preferred color would choose equally between the two jars, while babies understand probability would more often choose the jar with the higher proportion of their preferred color. Of the 24 infants studied, 18 chose the jar with the higher proportion of their preferred color. Are infants more likely to choose the jar with the higher proportion of their referred color?

a) State the null and alternative hypotheses.

b) Give the relevant sample statistic, using correct notation.

c) Which of the following should be used to calculate a p-value for this dataset? A randomization test, a test using the normal distribution, or either one? Why?

d) Find a p-value using a method appropriate for this data situation.

e) Make a conclusion in context, using $\alpha = 0.05$.

6.76 Compute the standard error for sample means from a population with mean $\mu = 100$ and standard deviation $\sigma = 25$ for sample sizes of $n = 30$, $n = 200$, and $n = 1000$. What effect does increasing the sample size have on the standard error? Using this information about the effect on the standard error, discuss the effect of increasing the sample size on the accuracy of using a sample mean to estimate a population mean.

6.107 Plastic microparticles are contaminating the world's shorelines, and much of this pollution appears to come from fibers from washing polyester clothes. The worst offender appears to be fleece, and a recent study found that the mean number of polyester fibers discharged into wastewater from washing fleece was 290 fibers per liter of wastewater, with a standard deviation of 87.6 and a sample size of 120.

a) Find and interpret a 99% confidence interval for the mean number of polyester microfibers per liter of wastewater when washing fleece.

b) What is the margin of error?

c) If we want a margin of error of only ± 5 with 99% confidence, what sample size is needed?

6.124 During the National Football League's 2014 AFC championship game, officials measured the air pressure on 11 of the game footballs being used by the New England Patriots. They found that the balls had an average air pressure of 11.1 psi, with a standard deviation of 0.40 psi.

a) Assuming this is a representative sample of all footballs used by the Patriots in the 2014 season, perform the appropriate test to determine if the average air pressure in footballs used by t he Patriots was significantly less than the allowable limit of 12.5 psi. There is no extreme skewness or outliers in the data, so it is appropriate to use the t-distribution.

b) Is it fair to assume that this sample is representative of all footballs used by the Patriots during the 2014 season?

6.145 Errors in medical prescription occur, and a study examined whether electronic prescribing may help reduce errors. Two groups of doctors used written prescriptions and had similar error rates before the study. One group switched to e-prescriptions while the other continued to use written prescriptions, and error rates were measured one year later. The results are given in Table 6.4. Find and interpret a 95% confidence interval for the difference in proportion of errors between the two groups. Is it plausible that there is no difference?

Table 6.4

	Error	**No Error**	**Total**
Electronic	254	3594	3848
Written	1478	2370	3848

6.173 Are malaria parasites able to control mosquito behavior to their advantage? A study investigated this question by taking mosquitoes and giving them the opportunity to have their first "blood meal" from a mouse. The mosquitoes were randomized to either eat from a mouse infected with malaria or an uninfected mouse. At several time points after this, mosquitoes were put into a cage with a human and it was recorded whether or not each mosquito approached the human (presumably to bite, although mosquitoes were caught before biting). Once infected, the malaria parasites in the mosquitoes go through two stages: the Oocyst stage in which the mosquito has been infected but it is not yet infectious to others and then the Sporozite stage in which the mosquito is infectious to others. Malaria parasites would benefit if mosquitoes sought risky blood meals (such as biting a human) *less* often in the Oocyst stage (because mosquitoes are often killed while attempting a blood meal) and *more* often in the Sporozite stage after becoming infectious (because this is one of the primary ways in which malaria is transmitted). Does exposing mosquitoes to malaria actually impact their behavior in this way?

a) In the Oocyst stage (after eating from mouse but before becoming infectious), 20 out of 113 mosquitoes in the group exposed to malaria approached the human and 36 out of 117 mosquitoes in the group not exposed to malaria approached the human. Calculate the z-statistic.

b) Calculate the p-value for testing whether this provides evidence that the proportion of mosquitoes in the Oocyst stage approaching the human is lower in the group exposed in malaria.

c) In the Sporozite stage (after becoming infectious), 37 out of 149 mosquitoes in the group exposed to malaria approached the human and 14 out of 144 mosquitoes in the group not exposed to malaria approached the human. Calculate the z-statistic.

d) Calculate the p-value for testing whether this provides evidence that the proportion of mosquitoes in the Sporozite stage approaching the human is higher in the group exposed to malaria.

e) Based on your p-values, male conclusions about what you have learned about mosquito behavior, stage of infection, and exposure to malaria or not.

f) Can we conclude that being exposed to malaria (as opposed to not being exposed to malaria) *causes* these behavior changes in mosquitoes? Why or why not?

6.195 A study examines chocolate's effects on blood vessel function in healthy people. In the randomized, double-blind, placebo-controlled study, 11 people received 46 grams (1.6 ounces) of dark chocolate (which is naturally flavonoid-rich) every day for two weeks, while a control group of 10 people received a placebo consisting of dark chocolate with low flavonoid content. Participants had their vascular health measured (by means of flow-mediated dilation) before and after the two-week study. The increase over the two-week period was measured, with larger numbers indicating greater vascular health. For the group getting the good dark chocolate, the mean increase was 1.3 with a standard deviation of 2.32, while the control group had a mean change of -0.96 with a standard deviation of 1.58.

a) Explain what "randomized, double-blind, placebo-controlled study" means.

b) Find and interpret a 95% confidence interval for the difference in means between the two groups. Be sure to clearly define the parameters you are estimating. You may assume that neither sample shows significant departures from normality.

c) Is it plausible that there is "no difference" between the two kinds of chocolate? Justify your answer using the confidence interval found in (b).

6.221 Drinking tea appears to offer a strong boost to the immune system. IN a study introduced in Exercise 3.91 on page 239, we see that production of interferon gamma, a molecule that fights bacteria, viruses, and tumors, appears to be enhanced in tea drinkers. In the study, eleven healthy non-tea-drinking individuals were asked to drink five or six cups of tea a day, while ten healthy non-tea- and non-coffee-drinkers were asked to drink the same amount of coffee, which has caffeine but not the L-theanine that is in tea. The groups were randomly assigned. After two weeks, blood samples were exposed to an antigen and production of interferon gamma was measured. The results are shown in Table 6.19 and are available in ImmuneTea. The question of interest is whether the data provide evidence that production is enhanced in tea drinkers.

Table 6.19

Tea	Coffee
5	0
11	0
13	3
18	11
20	15
47	16
48	21
52	21
55	38
56	42
58	

a) Is this an experiment or an observational study?

b) What are the null and alternative hypotheses?

c) Find a standardized test statistic and use the t-distribution to find the p-value and make a conclusion.

d) Always plot your data! Look at a graph of the data. Does it appear to satisfy a normality condition?

e) A randomization test might be a more appropriate test to use in this case. Construct a randomization distribution for this test and use it to find a p-value and make a conclusion.

f) What conclusion can we draw?

6.255 As part of the same study described in Exercise 6.254, the researchers were also interested in whether babies preferred singing or speech. Forty-eight of the original fifty infants were exposed to both singing and speed by the same woman. Interest was again measured by the amount of time the baby looked at the woman while she made noise. In this case the mean time while speaking was 66.97 with standard deviation of 43.42, and the mean for singing was 56.58 with a standard deviation of 31.57 seconds. The mean of the differences was 10.39 more seconds for the speaking treatment with a standard deviation of 55.37 seconds. Perform the appropriate test to determine if this is sufficient evidence to conclude that babies have a preference (either way) between speaking and singing.

6.257 Table 6.25 gives a sample of grades on the first two quizzes in an introductory statistics course. We are interested in testing whether the mean grade on the second quiz is significantly higher than the mean grade on the first quiz.

Table 6.25

First Quiz	Second Quiz
72	78
95	96
56	72
87	89
80	80
98	95
74	86
85	87
77	82
62	75

a) Complete the test if we assume that the grades from the first quiz come from a random sample of 10 students in the course and the grades on the second quiz come from a different separate random sample of 10 students in the course. Clearly state the conclusion.

b) Now conduct the test if we assume that the grades recorded for the first quiz and the second quiz are from the same 10 students in the same order. (So the first student got a 72 on the first quiz and a 78 on the second quiz.)

c) Why are the results so different? Which is a better way to collect the data to answer the question of whether grades are higher on the second quiz?