

Project 7: Hypothesis Testing

Statistics 401: Fall 2016

Due 1:40pm Tuesday, November 29

Turn in your answers and work in a type-written report. Give all R-output in an appendix.

1. On October 25, the LA Times ran a story with the title “there is no evidence that thousands of non-citizens are illegally voting and swinging elections” (<http://www.latimes.com/politics/la-na-pol-noncitizen-voters-20161025-snap-story.html>).
 - (a) Which of the following two sets of hypotheses, (i) or (ii), is the LA Times statement consistent with?
 - i. H_0 : a thousand illegal immigrants (or less) vote
 H_a : more than a thousand illegal immigrants vote
 - ii. H_0 : a thousand illegal immigrants (or more) vote
 H_a : less than a thousand illegal immigrants vote
 - (b) Does the LA Times statement indicate that H_0 was rejected or not rejected? Justify your answer.
 - (c) Give a sample design that describes how you would collect data to test the hypotheses that you identified in 1a.
2. In 11/17’s class, to illustrate the use of R’s **prop.test** function, we collected data from all $n = 12$ students in class to estimate the true percentage of all Montanan college students who support a ban on trapping on public lands in Montana. We found that $\hat{p} = 7/12$ students supported the ban.
 - (a) Unfortunately, the test and CI we calculated on 11/17 regarding the proportion of all college students is not reliable because the assumptions of the Central Limit Theorem were not satisfied. Specify which assumptions were violated.
 - (b) Even though we realize that the CLT should not be applied (see 2a), for practice, calculate a 95% CI for the true proportion of Montana college students who support the ban on hunting. Perform the calculations “by hand” and give all relevant work and answers.
 - (c) Give a conclusion about the CI in terms of the problem.
 - (d) On 11/8’s ballot, Montanan voters recently rejected a ballot initiative seeking to ban trapping on public lands in Montana, with only 37.2% of voters supporting the ban. Use the CI from 2b to test whether the proportion of college students who support the ban is larger than 0.372. State the statistical hypotheses that you are testing, explain how you are using the CI to conduct the test, and give the test’s conclusion in terms of the problem.
 - (e) You want to take a SRS of $n = 30$ students to test the hypotheses in 2d at a significance level of $\alpha = 0.05$. If the true proportion of college students who support banning trapping is 58.3% (i.e., $H_a : p = 0.583$), calculate the power of the test. See the example power calculation for a proportion on the STAT401 website.
3. In an observational study of middle aged Finnish men (see the abstract for *Coffee Drinking Is Dose-Dependently Related to the Risk of Acute Coronary Events in Middle-Aged Men* from the September 2004 Journal of Nutritional Epidemiology, available at the STAT401 web page), the researchers were interested in determining if there is a relationship between coffee consumption and white blood cell count (white blood cells or *leukocytes* defend the body against infection and disease by ingesting foreign materials and cellular debris, by destroying infectious agents and cancer cells, or by producing antibodies). The white blood cell count of $n_1 = 77$ non-coffee

drinkers was compared to $n_2 = 351$ heavy coffee drinkers (those who drink on average 950 ml of coffee each day). The study reported that $\bar{x}_1 = 5.2$, $s_1 = 1.4$, $\bar{x}_2 = 6.0$, and $s_2 = 1.7$. These white blood cell counts are in billions per liter.

- (a) How many 8oz cups of coffee does 950ml correspond to? Does this make you a heavy coffee drinker?
- (b) Explain why a pooled procedure (instead of an un-pooled) is appropriate to analyze these data?
- (c) Test whether drinking coffee is associated with higher white blood cell counts, on the average, in middle aged men. Perform all SIX steps of the test “by hand” and give a conclusion in terms of the problem?
- (d) Describe a Type I error in terms of this problem.
- (e) Describe a Type II error in terms of this problem.
- (f) The study was really interested in determining the effects of coffee consumption on coronary heart disease. In light of this being an observational study, comment on the assertion in the abstract that “heavy coffee consumption increases the short-term risk of acute myocardial infarction or coronary death.”

4. Cancer is a disease of uncontrolled cell growth and can occur in nearly any tissue in the body, including breast tissue. A breast cancer diagnosis is among the most frightening medical experiences a woman can have. In 2012 alone, more than a quarter of a million in the US were diagnosed with the disease. Approximately one in eight women will develop breast cancer in her life. Ductal carcinoma in situ (DCIS) is the most common type of non-invasive breast cancer. “Ductal” means that the cancer is inside the milk ducts. “Non-invasive” means that it hasn’t spread beyond the ducts. DCIS is vigorously treated by surgery and radiation by today’s healthcare professionals. One reason for this is based on four separate observational studies from 1980 to 2005 that reported that when DCIS went untreated in $n = 136$ women, it developed into invasive cancer 25% of the time (if you are interested, these studies and others are in the review article by Boughey et al. (2007) in *The Oncologist* available at <http://theoncologist.alphamedpress.org/content/12/11/1276.full>). In the 1990’s, a completely randomized experiment (called a **randomized clinical trial** in the epidemiological literature) showed that of $n = 408$ women who received a partial mastectomy, 17% eventually developed invasive cancer, and of $n = 410$ women with DCIS who received a partial mastectomy and radiation, 8% developed invasive cancer.

- (a) Perform a 2-sample test of proportions to compare the observational study of $n_1 = 136$ women to the treatment group in the CRD of $n_2 = 408$ women who received only a partial mastectomy. Do the data support the hypothesis that a partial mastectomy alone reduces the risk of invasive breast cancer?
- (b) Describe a Type I error in terms of this problem.
- (c) Describe a Type II error in terms of this problem.
- (d) Use R’s prop.test to test whether partial mastectomy and radiation reduces the risk of an invasive cancer compared to partial mastectomy alone. Can we conclude that the addition of radiation caused a decrease in the lower proportion of invasive breast cancer? Explain.