Homework 5

Fall 2016, Math 107, Prof. Adam Loy

Due Tuesday, October 11 by 4 p.m.

Problem 1

Write the null and alternative hypotheses in words and then symbols (notation) for each of the following situations.

a. The 2004 National Technology Readiness Survey sponsored by the Smith School of Business at the University of Maryland surveyed 418 randomly sampled Americans, asking them how many spam emails they receive per day. The survey was repeated on a new random sample of 499 Americans in 2009. Has the average number of spam emails per day has changed from 2004 to 2009?
b. In a 1994 study, 164 pregnant, HIV-positive women were randomly assigned to receive the drug AZT during pregnancy and 160 such women were randomly assigned to a control group that received a placebo. It turned out that 40 of the mothers in the control group gave birth to babies who were HIV-positive, compared to only 13 in the AZT group. Does this provide evidence that AZT reduces the proportion of HIV-positive babies?

Problem 2

Verizon is the primary local telephone company (incumbent local exchange carrier, ILEC) for a large area of the eastern United States. As such, it is responsible for providing repair service for the customers of other telephone companies known as competing local exchange carriers (CLECs) in this region. Verizon is subject to fines if the repair times (the time it takes to fix a problem) for CLEC customers are substantially worse than those for Verizon customers. The data set `Verizon.csv` contains a random sample of repair times for 1664 ILEC and 23 CLEC customers recorded in hours.

Please use R markdown to complete this part of the assignment. You should **knit to Word or PDF** and submit a hard copy of this file along with your solution to Problem 1.

a. Using correct mathematical notation, write down the hypotheses for a permutation (i.e. randomization) test to determine whether Verizon is discriminating against CELCs in this region—that is, whether repairs times for CELC customers are substantially longer, on average. b.Use R to calculate the mean repair time for the 1664 ILEC customers and the 23 CLEC customers, and the difference in mean repair times.
b. Use R to create a permutation distribution consisting of 9999 simulations. Create a histogram of the permutation distribution with a superimposed marker representing the observed difference in mean repair times between ILEC and CELC customers (the `permTest` function does this for you). Include this plot in your homework submission along with the R code you used to generate it.
c. Describe the shape of the permutation distribution you created in the previous part.
d. Report the p-value from your permutation distribution.
e. Based on your p-value, does it appear that there is evidence that Verizon is discriminating against CELCs in this region? Briefly justify your answer.