Homework №6

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) Among a simple random sample of 331 American adults who do not have a four-year college degree and are not currently enrolled in school, 48% said they decided not to go to college because they could not afford school. A newspaper article states that only a minority of the Americans who decide not to go to college do so because they cannot afford it and uses the point estimate from this survey as evidence. We wish to test this claim.

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.

(a) Using correct mathematical notation, write down the hypotheses for a permutation 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) Calculate the mean repair time for the 1664 ILEC customers and the 23 CLEC customers, and the difference in mean repair times.

(c) Use R to create a permutation distribution consisting of 1000 simulations. Create a histogram of the permutation distribution with a superimposed vertical line representing the observed difference in mean repair times between ILEC and CELC customers. Include this plot in your homework submission along with the R code you used to generate it.

(d) Describe the shape of the randomization distribution you created in the previous part.

(e) Use R to calculate a the p-value from your permutation distribution. Include the code you used to perform your calculation and report the p-value that you obtained.

(f) 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.

Problem 3

Some people claim that they can tell the difference between a diet soda and a regular soda in the first sip. A researcher wanting to test this claim randomly sampled 80 such people. They then filled 80 plain white cups with soda, half diet and half regular through random assignment, and asked each person to take one sip from their cup and identify the soda as diet or regular. 53 participants correctly identified the soda.

(a) Is this study an experiment or observational study? Briefly justify your answer.

(b) Using correct mathematical notation, write down the hypotheses for a permutation test determining whether these people are able to detect the difference between diet and regular soda on the first sip.

(c) Calculate the observed sample proportion of people that correctly identified the soda.

(d) Use R to create a permutation distribution consisting of 1000 simulations. Create a histogram of the permutation distribution with a superimposed vertical line representing the proportion of people that correctly identified the soda. Include this plot in your homework submission along with the R code you used to generate it.

(e) Describe the shape of the randomization distribution you created in the previous part.

(f) Use R to calculate a the p-value from your permutation distribution. Include the code you used to perform your calculation and report the p-value that you obtained.

(g) Based on your p-value, do these data provide strong evidence that these people are able to detect the difference between diet and regular soda? Briefly justify your answer.