QBUS3830
Advanced Analytics

Semester 2, 2018

Homework Task 2: Hypothesis Testing

1 Case study: Benford's Law

1. Implement a Python function that performs Pearson's χ^2 test for multinomial data and returns the test statistic and p-value.

2. The `GDP` dataset contains a list of countries ranked by their GDP in 2017 (in millions of dollars), according to the International Monetary Fund (IMF). Make a basic table to discuss how well the data conforms to Benford's law for the first digit. Perform the χ^2 test and discuss the results.

3. The `Fraud` dataset contains three series. One is a real financial variable for a random sample of companies listed in the New York Stock Exchange (NYSE). The other two are the same series, but with random modifications of digits. Repeat the exercise above for each series, and identify the two "fraudulent" series.

2 Case study: Verizon Repair Times

The `Verizon` dataset contains data from a court case that involved the American telecommunications company. 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 to provide repair services for the customers of other telephone companies know as competing local exchange carriers (CLECs). Verizon is subject to fines if the repair times for CLEC customers are worse than those for Verizon customers.

Assume a significance level of 1%.

1. Conduct a two-sample test based on large-sample theory and discuss the results.

2. Implement a Python function that conducts a permutation test based on the mean.

3. Conduct the permutation test, plot the permutation distribution, and discuss the results. Compare the permutation test to the large-sample test, and discuss which one is more appropriate for this problem.

3 Rules

Do not use package versions of the χ^2 and permutation tests. Do not look for similar code in the internet. The code must be your own work.

4 Rubric

You will get the full marks if you follow the instructions, obtain the correct p-values, and interpret the results correctly.