Chapter 4   Hypothesis Tests and Sampling

4.1   Chi-Square Goodness of Fit Test

The chi-square goodness of fit test is shown below.

$$\sum \frac{(O_i - E_i)^2}{E_i}$$

The purpose of this test is to show whether a group of observations follows an expected distribution.

Hypothesis

For a chi-square goodness of fit test, the hypotheses take the following form. - Ho: The data are consistent with a specified distribution. - Ha: The data are not consistent with a specified distribution.

These are then to be compared to an α (alpha) of your choosing (0.01, 0.05, 0.1)

Assumptions

The chi-square goodness of fit test is appropriate when the following conditions are met: - The data are from a simple random sample - The groups that are being looked at is a categorical, i.e. Qualitative - There are at least five expected observations per group.

Benford's Law

Let's take a look at Benford's law as an example of how to use the chi-square goodness of fit test. Benford's law states that the first digits of a random group of numbers always follows a certain pattern.
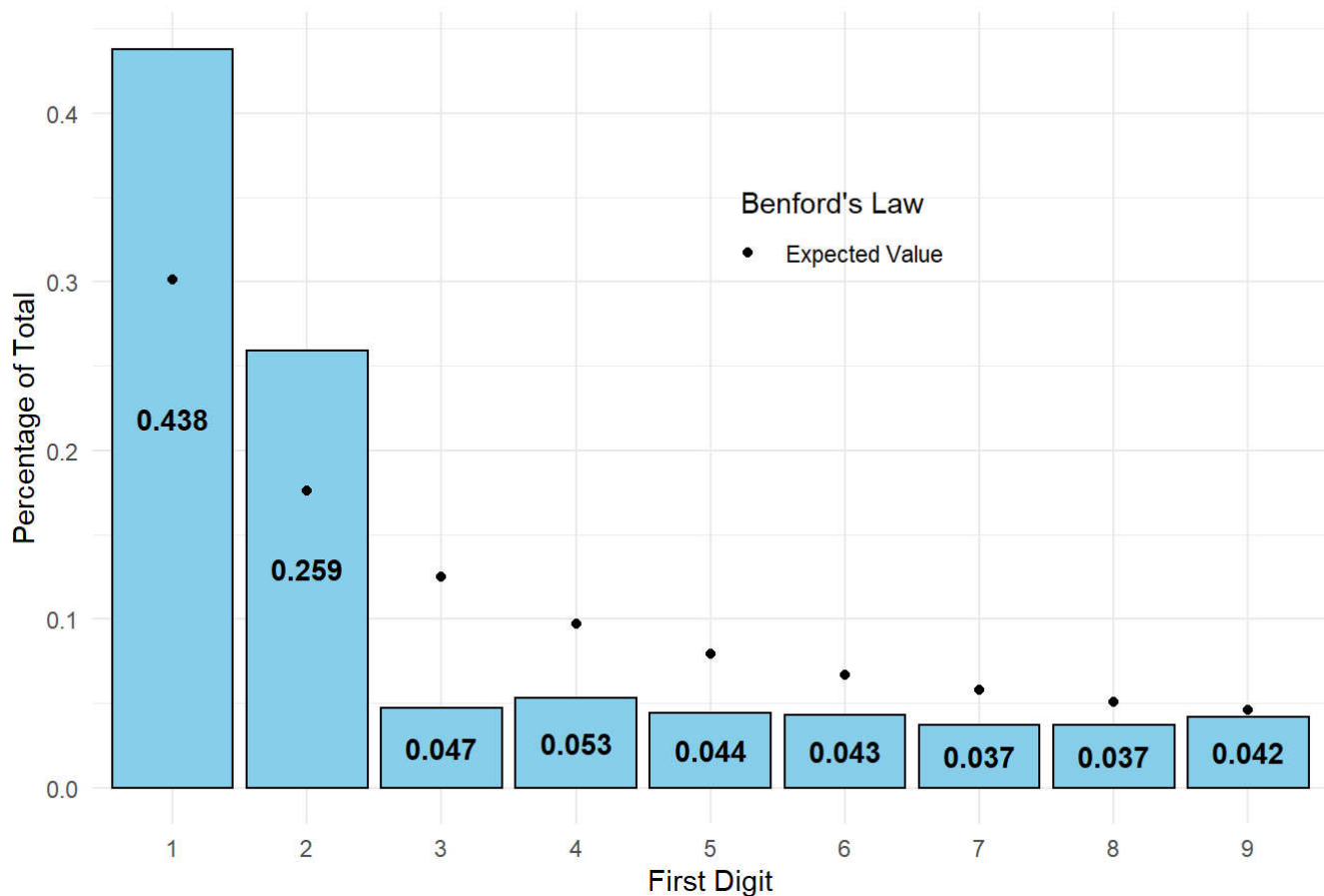
Benford's Law
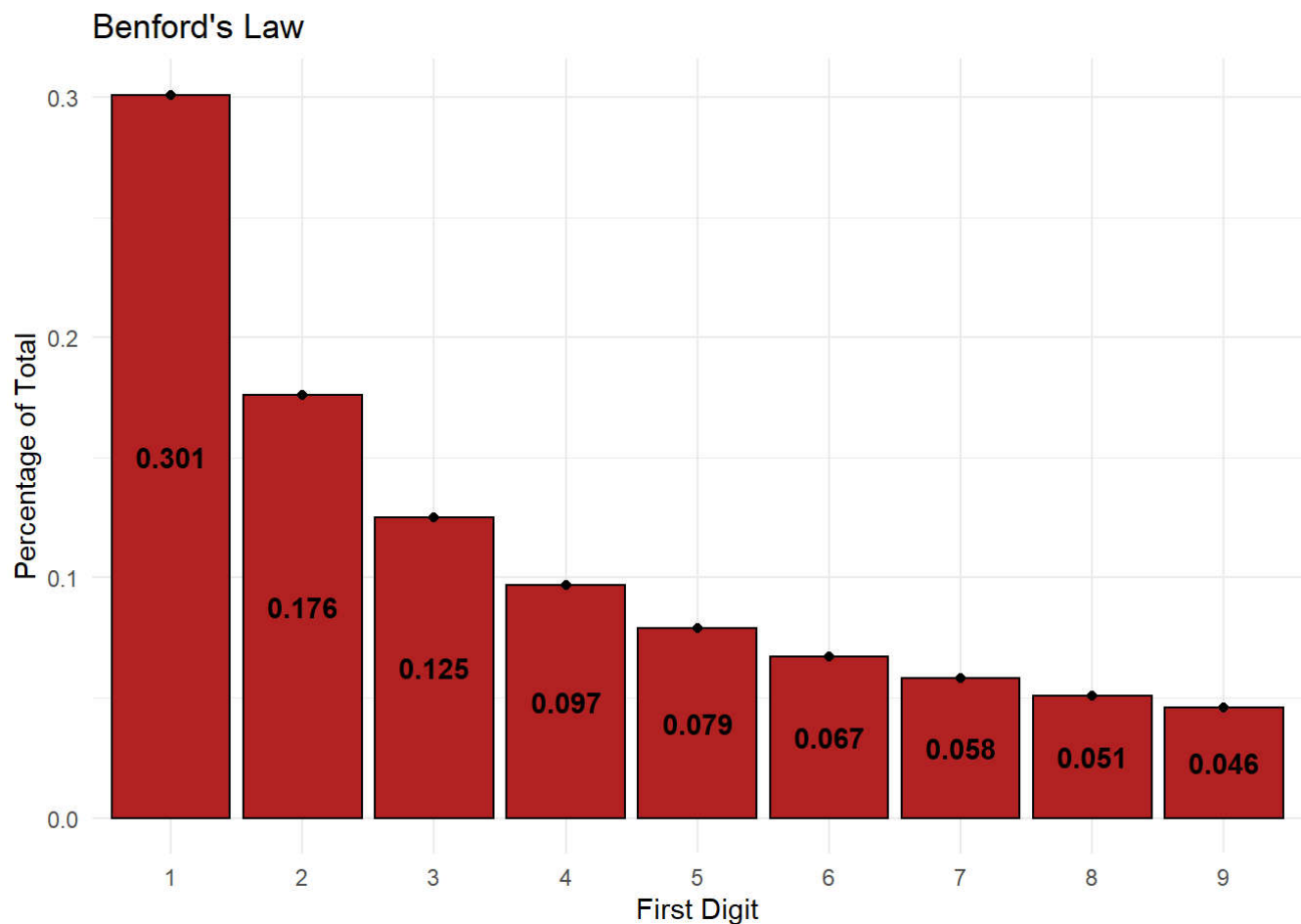
We can assume that if I were to take a set of numbers and pull their first digits from that set then we should see this same distribution or something close to it. This is the purpose of the chi-square goodness of fit test.

Example

Let's take a look at the distribution of the first digits of the Walmart stock. This is a good test set because there are a lot of records of what the stock has been over many years, and we also know that it's truly random. From there we will randomly subset all of walmart data and we will

perform a goodness of fit test.

Comparing Walmart Stock to Benford's Law

As we can se it looks pretty similar to Benford's Law, but lets take a look at the actual p-value calculated from the goodness of fit test. The first thing we need to do is find the expected counts of what Benford's Law says the distribution should be. To do this, we multiply Benford's distribution percentages by the total.

301, 176, 125, 97, 79, 67, 58, 51 and *46*

This should be the expected count, now let's calculate the chi-squared value based on the formula at the beginning of this page. In Laman's - The observed values minus the expected values all squared / divided by the expected values, all summed.

$$\sum \frac{(O_i - E_i)^2}{E_i}$$

The chi-square value is *206*

We can now calculate the the p-value by using a chi-square test. The p-value that is produced is: *3.47e-40*

As wee can see the p-value is less than our α of .05, therefore we fail to reject and we can say that the random sample of numbers from the Walmart stock follows Benford's Law.