Chapter 3: Inference for a Categorical Variable with More than Two Categories

The analyses we completed in Chapter 2 were for a single categorical variable with only two outcomes. For example, in the helper/hinderer study, the babies were choosing either one toy or the other. In the gender discrimination example, each employee selected for management was either male or female. Next, we'll consider problems involving a single categorical variable which has more than two categories.

Example 3.1: Crime Statistics

The Minneapolis Police Department posts regular updates on crime statistics on their website. A colleague of mine has collected this data for the past few years on all neighborhoods in Minneapolis. A portion of the data set and the precinct map are shown below.

Source: http://www.minneapolismn.gov/police/crime-statistics/

Suppose the police chief for Precinct #2 has received a complaint from a permanent resident who lives in a neighborhood near the University of Minnesota. This resident has asked for additional patrol to take place in his neighborhood as he believes that crime rates vary over the course of the year.

Crime rates are reported by month, so we will use the following definitions for the Seasons:

- Fall: September, October, and November
- Winter: December, January, and February
- Spring: March, April, and May
- Summer: June, July, and August

Crimes classified as Murder, Rape, Robbery, Aggravated Assault, Burglary, Larceny, Auto Theft, and Arson are all used in reporting the Total.

The Minneapolis Police Department reported that a total of 103 crimes occurred in the University of Minnesota neighborhood last year.

Research Question	Is there evidence to suggest that crime patterns in the University of Minnesota neighborhood differ over the four seasons of the year?
Parameters of Interest	The four parameters of interest are defined as follows: $_{\text{fall}}$ = the probability of a crime occurring in the Fall $_{\text{winter}}$ = the probability of a crime occurring in the Winter $_{\text{spring}}$ = the probability of a crime occurring in the Spring $_{\text{summer}}$ = the probability of a crime occurring in the Summer
Hypotheses	H_o: Crimes are equally dispersed over the four seasons H_a: Crimes are not occurring equally over the four seasons

The approach we take here to address the research question is very similar to what we have done previously. We will assume the crime patterns are occurring equally across the four seasons (i.e., that the null hypothesis is true) and then get a good idea of what outcomes we would expect to see if this were really the case. Then, we will check to see if the observed outcomes given in the data are consistent (or inconsistent) with what we expected to see under the null hypothesis. If the observed data are inconsistent with the outcomes expected under the null, then we have sufficient statistical evidence to say crime rates vary across the four seasons.

1. Find the expected number of total crimes for each season under the assumption that crimes are occurring equally over the four seasons. How did you obtain these values?