ST 516: Foundations of Data Analytics

Case Study Example—Recyclables

Example: Recyclables

Taken from Chapter 5 of *An Introduction to Statistical Methods and Data Analysis*

> *The sanitation department if a large city wants to investigate ways to reduce the amount of recylcable materials that are placed in the city's landfill. From an analysis of the recycling records from other cities, it is determined that if more than 50% of households and businesses have a weekly amount of recyclable material that is five pounds or more, a commercial recycling firm could make a profit collecting the material. To determine the feasibility of a new recylcing plan, a random sample of 25 households and businesses is selected, and the weekly weight of recyclable material (in pounds/week) is recorded. Is it safe to say that more than 50% of households and businesses produce five or more pounds of recyclable material per week?*

Think About an Approach

Some things to consider:

1. The question is specifically about a median (50% of households).
2. The households and businesses were randomly selected, so we will be able to generalize our findings to the entire city.
3. There is no notion of a treatment here, we are just going to observe a feature of these households and businesses—it is an observational study.

Perform an Analysis

We want to know if the population median is 5 pounds, so we'll perform a sign test.

```
pounds <- c(14.2, 5.3, 2.9, 4.2, 1.2, 4.3, 1.1, 2.6, 6.7, 7.8, 25.9,
  43.8, 2.7, 5.6, 7.8, 3.9, 6.5, 29.5, 2.1, 34.8, 3.6, 5.8, 4.5, 6.7)
signs <- ifelse(pounds >= 5,1,0)
binom.test(sum(signs),length(pounds))
```

```
##
##  Exact binomial test
##
## data:  sum(signs) and length(pounds)
## number of successes = 13, number of trials = 24, p-value = 0.8388
## alternative hypothesis: true probability of success is not equal to
## 95 percent confidence interval:
##  0.3282081 0.7444698
## sample estimates:
## probability of success
##              0.5416667
```

Perform an Analysis

We can't reject the possibility that the population median is 5 pounds. Let's construct a 95% confidence interval for the median:

we need to calculate $\frac{n}{2} - \frac{1.96\sqrt{n}}{2}$ and $\frac{n}{2} + \frac{1.96\sqrt{n}}{2} + 1$

```
n <- 25
round(c(n/2 - 1.96*sqrt(n)/2, n/2 + 1.96*sqrt(n)/2 + 1))
```

```
## [1]  8 18
```

So we need, the 8^{th} and 18^{th} observations:

```
spounds <- sort(pounds)
spounds[c(8, 18)]
```

```
## [1] 3.9 7.8
```

Write a Summary

There is no evidence to suggest that the median weekly amount of recyclables per household or business is different from 5 pounds per week ($p = 0.8$ from a sign-test, $n = 25$). An approximate 95% confidence interval for the population median runs from 3.9 pounds to 7.8 pounds. This means that it is possible that the population median is less than 5 pounds. To decide on the feasibility of a new managment plan, the city should collect data on more households and businesses before proceeding.