ST 516: Foundations of Data Analytics

Case Study Example—Sun Protection Factor

First Example: Sun Protection Factor

Taken from exercise 30 in Chapter 4 of The Statistical Sleuth.

> *A sunscreen sunlight protection factor (SPF) of 5 means that a person who can tolerate Y minutes of sunlight without the sunscreen can tolerate $5Y$ minutes of sunlight with the sunscreen. The data in the file ex0430 of the Sleuth3 library are times (in minutes) that 13 patients could tolerate the sun (a) before receiving treatment and (b) after receiving a particular sunscreen treatment. Analyze the data to estimate and provide a 95% confidence interval for the sunlight protection factor.*

Take a Look at the Data

```
library(Sleuth3)
head(ex0430)
```

```
##   PreTreatment Sunscreen
## 1           30        120
## 2           45        240
## 3          180        480
## 4           15        150
## 5          200        480
## 6           20        270
```

Think About an Approach

Some things to consider:

1. There are two samples of data—'PreTreatment' and 'Sunscreen'. But notice that they are *paired*.
2. The sample size is small, there are $n = 13$ pairs.
3. The question is about a multiplicative change (SPF), so we should consider the paired quotients, rather than the paired differences. But, notice that if we use quotients the key hypothesis of interest is about $\mu_0 = 1$.
4. It's not clear from the data description how the patients came to be in the study—we will assume there was some random selection mechanism.
5. Each patient receives the treatment, so there is no notion of a randomly asssigned treatment here.

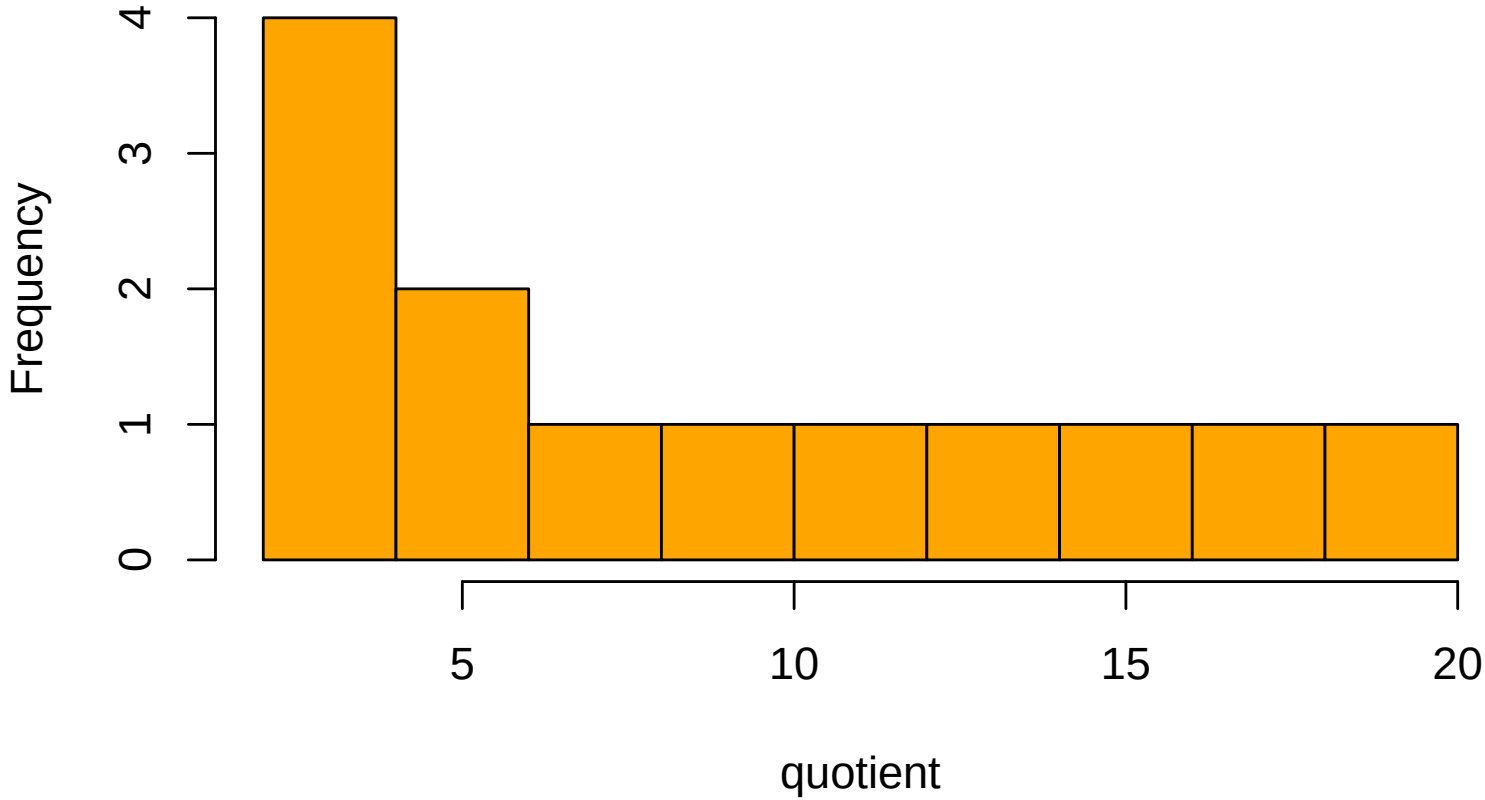
Perform an Analysis

```
quotient <- ex0430$Sunscreen/ex0430$PreTreatment
t.test(quotient, mu = 1)
```

```
##
##  One Sample t-test
##
## data:  quotient
## t = 4.9431, df = 12, p-value = 0.0003403
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##   5.598494 12.847660
## sample estimates:
## mean of x
##  9.223077
```

Consider Alternatives

Since the sample size is relatively small, we should just check to make sure that the paired quotients are not highly skewed:



This sample distribution does have some skewness, but remember that we're only looking at 13 observations here, so evaluating skewness in the population distribution is challenging.

Write a Summary

There is convincing evidence ($p = 0.0003$, paired t-test on 12 df) that the sunscreen is associated with an positive mean SPF. A 95% confidence interval for the mean SPF is (5.9, 12.8). There may be confounding factors that make it difficult to attribute this increase in mean SPF directly to the sunscreen. Most notably, the patients used the sunscreen after first being in the in sun for a while—this could have conditioned their skin to be better protected from the sun the next time they went out.