

## Lab 4

Remember: to get help on any command you can type `?command`.

### 1. Constructing confidence intervals for slope

I want to examine, as a class, whether confidence intervals (CIs) for the slope constructed using the method we covered in lectures have the correct “coverage properties” (e.g. do 95% CIs contain the true value 95% of the time?). With the person next to you discuss how we can do this as a class. Then we’ll discuss it as a class, before agreeing on a procedure to follow.

### 2. Testing for significant effects: correlation vs causality

a) The file

```
http://www.stat.washington.edu/stephens/S423/firstletter.txt
```

contains data from New York Times, 17th Jan 2002, relating to 88 US cities. The first column (“FirstLetter”) contains the first letter of the city name (coded 1-26), and the second column (“MaxTemp”) contains the max temperature for the previous day.

Read the data into R. I think you can use

```
weather <- read.table('http://www.stat.washington.edu/stephens/S423/firstletter.txt',  
                      header=TRUE)
```

Use linear regression to investigate whether there appears to be any evidence for a connection between the first letter of the city name and its maximum temperature. What do you conclude?

### 3. The distribution of $p$ values under $H_0$

- a) Do you know what the distribution of a  $p$  value is, under the null hypothesis? (Do you know what this question means? Think about it.)
- b) Try simulating data for  $Y$  and  $X$  under the null hypothesis that the slope is 0. Find a  $p$ -value for testing the hypothesis that the slope is 0, and write it on the board. Plot a histogram of the class results. What do you find?