

Lab 5: Confounding and Sampling

Oct 30, 2017

1. Consider the `state.x77` pre-loaded R dataset that presents some statistics about the fifty states, published by the Bureau of the Census in 1977. We will consider the relationships between life expectancy, income and illiteracy in this data.

1a. Convert the `state.x77` data set to a 'data frame' format using the command `st = as.data.frame(state.x77)`. What is the dimension of `st` (use the `dim` command)? Describe the first four column variables in the data (very briefly in words, but don't just list them).

1b. Plot life expectancy, `Life Exp`, against `Income` in one plot and `Life Exp` against `Illiteracy` in another plot. (`Life Exp` should be on the y axis). Add a trend line to the plots as described in the note at the end of this lab. Use the `par(mfrow = c(1,2))` command so that your plots show up next to each other. Title the plots appropriately using the `main` argument in `plot`. In a sentence, compare the relationships shown in the plots.

```
#a
?state.x77
st = as.data.frame(state.x77)
```

```
#b
```

1c. Create plots of `Life Exp` vs. `Illiteracy` for 1) only the states with below-average `Income` and 2) Only the states with above-average `Income`. Again, add a trend line and appropriate title to each plot and use the `par(mfrow = c(1,2))` to show the plots side-by-side. Repeat for `Life Exp` vs. `Income`, broken down by below and above-average `Illiteracy`. In 2-3 sentences: Describe how the relationships shown in 1b do or do not persist in these plots.

```
#c
```