

Intro to Social Science Data Analysis

Week 11 Seminar: Simple Linear Regression

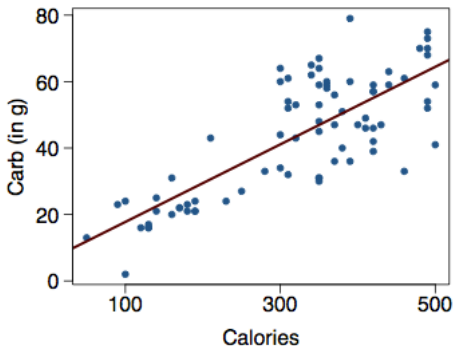
Christopher Gandrud

November 15, 2012

Interpretation

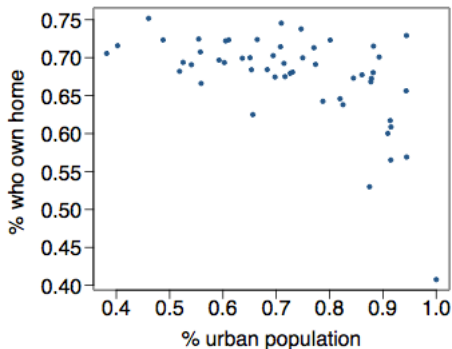
With a partner,

- ▶ describe the relationship between the number of calories a food item has and its carbohydrates,
- ▶ roughly estimate how many carbohydrates an item with 300 calories would have,
- ▶ what are the independent and dependent variables?



Outliers

This graph plots states in the United States.
The outlier on the bottom right is Washington, DC.
What should we do with the outlier?



Inference from Intercepts

What does the point estimate of α mean?

How interested are we in making inferences from $\hat{\alpha}$?

Interpreting Dummy Variables

The variable `Elect` is a dummy variable equalling 1 when a country has an elected legislature and 0 when it does not.

Predict its `GDPperCapita` in thousands of US dollars for countries with and without elected legislatures using the following simple linear regression equation:

$$\text{GDPperCapita} = 12 + 14(\text{Elect})$$

Note: this is not based on real data.

Load Data

Load data on 3143 US counties from the *openintro* package.

```
# Load library
library(openintro)

# Load Data
data(county)

# Show variable names
names(county)
```

##	[1]	"name"	"state"
##	[3]	"pop2000"	"pop2010"
##	[5]	"fed_spend"	"poverty"
##	[7]	"homeownership"	"multiunit"
##	[9]	"income"	"med_income"

Practice

Pick two continuous variables. Explore the relationship between the variables, including:

- ▶ their correlation,
- ▶ the direction of the relationship,
- ▶ the strength of the relationship,
- ▶ predict three values of one variable based on three values of the other,
- ▶ statistical inferences from a linear regression equation.

Could the two variables be causally related? Why/why not?

Practice

Pick two continuous variables. Explore the relationship between the variables, including:

- ▶ their correlation,
- ▶ the direction of the relationship,
- ▶ the strength of the relationship,
- ▶ predict three values of one variable based on three values of the other,
- ▶ statistical inferences from a linear regression equation.

Could the two variables be causally related? Why/why not?

Practice

Pick two continuous variables. Explore the relationship between the variables, including:

- ▶ their correlation,
- ▶ the direction of the relationship,
- ▶ the strength of the relationship,
- ▶ predict three values of one variable based on three values of the other,
- ▶ statistical inferences from a linear regression equation.

Could the two variables be causally related? Why/why not?

Practice

Pick two continuous variables. Explore the relationship between the variables, including:

- ▶ their correlation,
- ▶ the direction of the relationship,
- ▶ the strength of the relationship,
- ▶ predict three values of one variable based on three values of the other,
- ▶ statistical inferences from a linear regression equation.

Could the two variables be causally related? Why/why not?

Practice

Pick two continuous variables. Explore the relationship between the variables, including:

- ▶ their correlation,
- ▶ the direction of the relationship,
- ▶ the strength of the relationship,
- ▶ predict three values of one variable based on three values of the other,
- ▶ statistical inferences from a linear regression equation.

Could the two variables be causally related? Why/why not?