Intro to Social Science Data Analysis

Week 12 Seminar: Multivariate Linear Regression & Presenting Regression Results

Christopher Gandrud

November 22, 2012

Due: Friday 30 November

Research Design

With your partner plan your research by answering the following questions:

1. What **difference** do you want to explain?
2. What is your **best guess** explanation (i.e. thesis statement)?
3. Can you **test your hypothesis using data**? If so, what data do you need to collect and what tests could you use?
4. What **rival explanations** are their?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation**.

Questionnaire from: modified from Cheryl Schonhardt-Bailey

Due: Friday 30 November

Research Design

With your partner plan your research by answering the following questions:

1. What **difference** do you want to explain?
2. What is your **best guess** explanation (i.e. thesis statement)?
3. Can you **test your hypothesis using data**? If so, what data do you need to collect and what tests could you use?
4. What **rival explanations** are their?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation**.

Questionnaire from: modified from Cheryl Schonhardt-Bailey

Due: Friday 30 November

Research Design

With your partner plan your research by answering the following questions:

1. What **difference** do you want to explain?
2. What is your **best guess** explanation (i.e. thesis statement)?
3. Can you **test your hypothesis using data**? If so, what data do you need to collect and what tests could you use?
4. What **rival explanations** are their?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation**.

Questionnaire from: modified from Cheryl Schonhardt-Bailey

Due: Friday 30 November

Research Design

With your partner plan your research by answering the following questions:

1. What **difference** do you want to explain?
2. What is your **best guess** explanation (i.e. thesis statement)?
3. Can you **test your hypothesis using data**? If so, what data do you need to collect and what tests could you use?
4. What **rival explanations** are their?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation**.

Questionnaire from: modified from Cheryl Schonhardt-Bailey

Due: Friday 30 November

Research Design

With your partner plan your research by answering the following questions:

1. What **difference** do you want to explain?
2. What is your **best guess** explanation (i.e. thesis statement)?
3. Can you **test your hypothesis using data**? If so, what data do you need to collect and what tests could you use?
4. What **rival explanations** are their?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation**.

Questionnaire from: modified from Cheryl Schonhardt-Bailey

Load Data

```r
# Load openintro package
library(openintro)

# Load Data
data(census)

# Show variable names
names(census)
```

```
## [1] "censusYear"          "stateFIPScode"
## [3] "totalFamilyIncome"   "age"
## [5] "sex"                 "raceGeneral"
## [7] "maritalStatus"       "totalPersonalIncome"
```

Model

With a partner, hypothesize what the likely associations between the variables:

- age,
- `sex`,
- `raceGeneral`,
- `maritalStatus`,

with `totalPersonalIncome`.

Model

With a partner, hypothesize what the likely associations between the variables:

- age,
- sex,
- raceGeneral,
- maritalStatus,

with totalPersonalIncome.

Model

With a partner, hypothesize what the likely associations between the variables:

- `age`,
- `sex`,
- `raceGeneral`,
- `maritalStatus`,

with `totalPersonalIncome`.

Model

With a partner, hypothesize what the likely associations between the variables:

- `age`,
- `sex`,
- `raceGeneral`,
- `maritalStatus`,

with `totalPersonalIncome`.

Predict the Effect

Using all of the variables in the data set create a parsimonious, but comprehensive linear regression model to find a **point estimate** of the total presonal income of a white widowed women who is 32 years old.

Write the linear regression equation and make the prediction.

Simulations

Simulate expected total family incomes, with associated uncertainty, for a range of ages.