

Intro to Social Science Data Analysis

Week 12 Seminar: Multivariate Linear Regression & Presenting Regression Results

Christopher Gandrud

November 12, 2012

Assignment 4

Due: Friday 30 November

Research Design

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

1. What difference or anomaly do you want to explain?
2. What is your best guess explanation? Draw your best guess in a diagram.
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 there?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation** if possible.
6. What other factors may influence the relationship you observe?

Assignment 4

Due: Friday 30 November

Research Design

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

1. What difference or anomaly do you want to explain?
2. What is your best guess explanation? Draw your best guess in a diagram.
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 there?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation** if possible.
6. What other factors may influence the relationship you observe?

Assignment 4

Due: Friday 30 November

Research Design

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

1. What difference or anomaly do you want to explain?
2. What is your best guess explanation? Draw your best guess in a diagram.
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 there?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation** if possible.
6. What other factors may influence the relationship you observe?

Assignment 4

Due: Friday 30 November

Research Design

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

1. What difference or anomaly do you want to explain?
2. What is your best guess explanation? Draw your best guess in a diagram.
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** if possible.
6. What other factors may influence the relationship you observe?

Assignment 4

Due: Friday 30 November

Research Design

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

1. What difference or anomaly do you want to explain?
2. What is your best guess explanation? Draw your best guess in a diagram.
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 there?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation** if possible.
6. What other factors may influence the relationship you observe?

Assignment 4

Due: Friday 30 November

Research Design

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

1. What difference or anomaly do you want to explain?
2. What is your best guess explanation? Draw your best guess in a diagram.
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 there?
5. How could you use data to test whether your best guess or the rival explanations are better? Write this as an **equation** if possible.
6. What other factors may influence the relationship you observe?

Load Data

```
# 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,
- ▶ totalPersonalIncome

on totalFamilyIncome.

Model

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

- ▶ age,
- ▶ sex,
- ▶ raceGeneral,
- ▶ maritalStatus,
- ▶ totalPersonalIncome

on totalFamilyIncome.

Model

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

- ▶ age,
- ▶ sex,
- ▶ raceGeneral,
- ▶ maritalStatus,
- ▶ totalPersonalIncome

on totalFamilyIncome.

Model

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

- ▶ age,
- ▶ sex,
- ▶ raceGeneral,
- ▶ maritalStatus,
- ▶ totalPersonalIncome

on totalFamilyIncome.

Model

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

- ▶ age,
- ▶ sex,
- ▶ raceGeneral,
- ▶ maritalStatus,
- ▶ totalPersonalIncome

on totalFamilyIncome.

Predict the Effect

Create a parsimonious, but comprehensive linear regression model to predict the total family income of a white, 32 year old women, with an personal income of \$25,000.

Write the linear regression equation and make the prediction.

Simulations

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