Regression Analysis

Logistic Regression

Nicoleta Serban, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

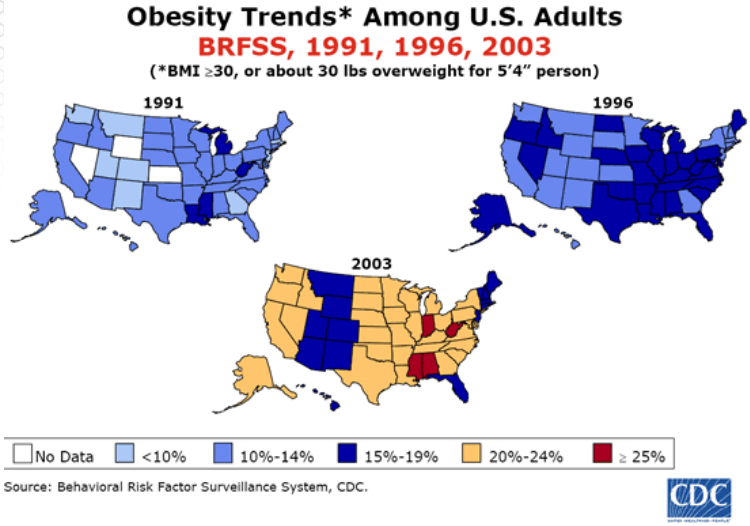
Case Study: The Demographics of Obesity

Georgia Tech

About This Lesson

Georgia Tech

Obesity in the United States



Georgia Tech

Case Study Overview

Objective:

- Use National Health and Nutrition Examination Survey (NHANES) to create a model used to predict likelihood of obesity.
- Identify predicting factors with predictive power

Variables:

- Response Variable: Whether an adult is classified as obese
- Predicting Variables: Age, Education Level, Gender

Data
- Training Data (4,314 Observations)
- Testing Data (1,000 Observations)

Georgia Tech

Case Study Overview

Objective:

- Use National Health and Nutrition Examination Survey (NHANES) to create a model used to predict likelihood of obesity.
- Identify predicting factors with predictive power

Variables:

- Response Variable: Whether an adult is classified as obese
- Predicting Variables: Age, Education Level, Gender

Age variable

- Present as a continuous variable in the data
- Recoded into classes (or ranges) like
  - Class 1: 18-24 years
  - Class 2: 25-34 years
  - etc.

Georgia Tech

Obesity Data

```
## Read data in R
obdata = read.table("obesitydata.txt", h=T)
attach(obdata)

## Data before aggregation
obesityind = factor(Obesity, labels=c("NotObese", "Obese"))
agegr = factor(AgeGroup,
    labels=c("18to24", "25to34", "35to44", "45to64", "65+"))
gender = factor(Gender, labels=c("Male", "Female"))
edu = factor(Education,
    labels=c("<9thGrade", "9to11Grade", "HighSchool", "SomeCollege", "College+"))
```

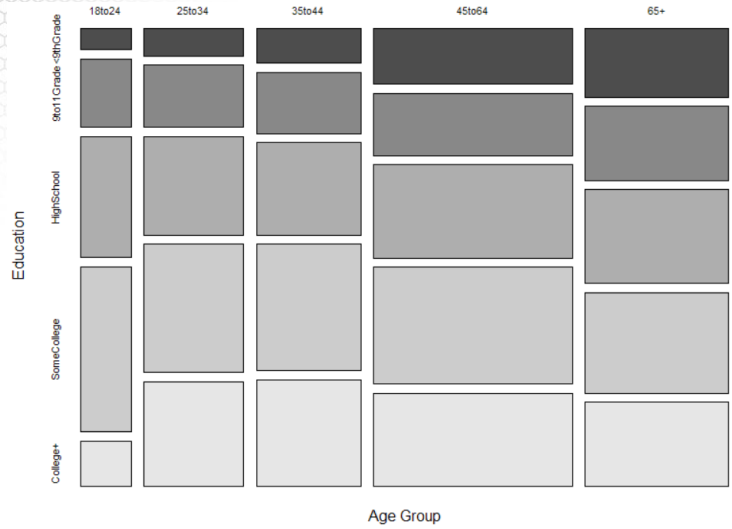
Georgia Tech

Exploratory Data Analysis

```
## Exploratory data analysis:  Categorical Predictors
tb_ageedu = xtabs(~agegr+edu)
library(vcd)
mosaicplot(tb_ageedu, xlab="Age Group", ylab="Education", color=TRUE, main="")
```

Georgia Tech

Exploratory Data Analysis



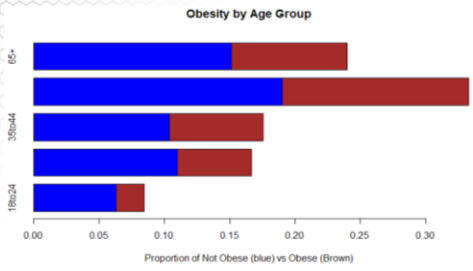
Georgia Tech

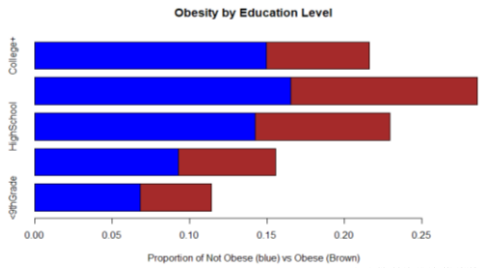
Exploratory Data Analysis

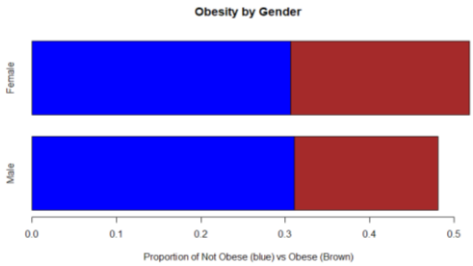
```
## Exploratory data analysis: Response vs Predictors
tb_obage = xtabs(~obesityind+agegr)
tb_obgender = xtabs(~obesityind+gender)
tb_obedu = xtabs(~obesityind+edu)

barplot(prop.table(tb_obage), axes=T, space=0.3, horiz=T,
    xlab="Proportion of Not Obese (blue) vs Obese (Brown)",
    col=c("blue","brown"), main="Obesity by Age Group")

barplot(prop.table(tb_obgender), axes=T, space=0.3, horiz=T,
    xlab="Proportion of Not Obese (blue) vs Obese (Brown)",
    col=c("blue","brown"), main="Obesity by Gender")

barplot(prop.table(tb_obedu), axes=T, space=0.3, horiz=T,
    xlab="Proportion of Not Obese (blue) vs Obese (Brown)",
    col=c("blue","brown"), main="Obesity by Education Level")
```

Georgia Tech

Exploratory Data Analysis







Georgia Tech

Summary