Regression Analysis

Logistic Regression

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

Model Estimation: Data Example

Georgia Tech

About This Lesson

Georgia Tech

Data Example: Smoking

- Between 1972 and 1974, a survey was taken in Whickham, a mixed urban and rural district near Newcastle upon Tyne, United Kingdom.
  - Among the information obtained originally was whether a person was a smoker or not.
- Twenty years later a follow-up study was conducted.
  - 76.12% of the 582 smokers were still alive, while only 68.58% of 732 nonsmokers were still alive.

Smokers had a higher survival rate than nonsmokers!
Call Philip Morris, smoking leads to a longer life span!

Acknowledgement: This example was provided by Dr. Jeffrey Simonoff from New York University.

Georgia Tech

Data Example in R

Data: $\boldsymbol{Y}_i$ binary responses ~ Binomial($\mathbf{p}_i$, $\boldsymbol{n}_i$)
- $\boldsymbol{Y}_i$ number of people at risk who survived (Survived)
- $\boldsymbol{n}_i$ number of people at risk (At.risk)

```
## Fit a logistic regression model
smoke1 = glm(Survived/At.risk ~ Smoker, weights=At.risk, family=binomial)
summary(smoke1)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.78052	0.07962	9.803	< 2e-16 ***
Smoker	0.37858	0.12566	3.013	0.00259 **

$\widehat{\boldsymbol{\beta}}_{smoker} = \mathbf{0.378}$: The log odds of survival increases by 0.378 for smokers versus non-smokers OR the odds of survival are 46% higher for smokers than for non-smokers (the odds ratio is 1.459).

Georgia Tech

Data Example in R

```
## Fit a logistic regression model
smoke2 = glm(Survived/At.risk ~ Smoker + Age, weights=At.risk, family=binomial)
summary(smoke2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	7.785001	0.454999	17.110	<2e-16 ***
Smoker	-0.240831	0.167885	-1.435	0.151
Age	-0.127419	0.007397	-17.227	<2e-16 ***

$\widehat{\boldsymbol{\beta}}_{smoker} = -\mathbf{0.24}$ The odds of survival is 27.2% higher for non-smokers than for smokers (odds ratio for non-smokers versus smokers is 1/exp(-0.24) =1.272).

Georgia Tech

Summary

Georgia Tech