Regression Analysis

Logistic Regression

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

Statistical Inference:
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

```
## 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 **

Null deviance: 641.5 on 13 degrees of freedom
Residual deviance: 632.3 on 12 degrees of freedom

```
1 - pchisq(smoke1$null.deviance-smoke1$deviance, 1)
[1] 0.002419817
```

Test for significance: β_{smoker} P-value = **0.0025**, thus statistically significant
Test for overall regression: Null deviance - Residual Deviance = 9.2
P-value = $\Pr(\chi^2_1 > 9.2)$ = **0.0024**

Georgia Tech

Data Example in R (cont'd)

```
## 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 ***

Null deviance: 641.496 on 13 degrees of freedom
Residual deviance: 43.459 on 11 degrees of freedom

Test for significance: β_{smoker} P-value = **0.151**, not statistically significant
Test for significance: β_{age} P-value ≈ **0**, statistically significant

Georgia Tech

Summary

Georgia Tech