Regression Analysis
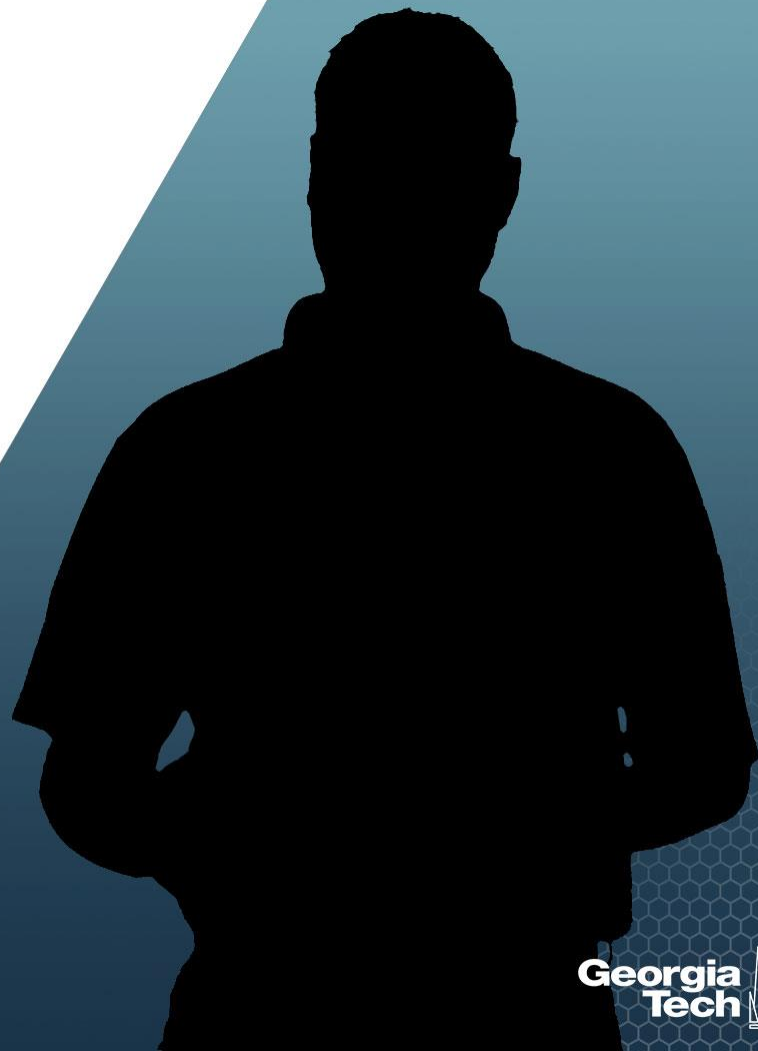## Logistic Regression

Nicoleta Serban, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Data Example

Georgia Tech

About This Lesson

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!

Georgia Tech

Data Example in R

Data Example in R

```
smoking <- read.table("CIGARETT.dat",sep="",row.names=NULL)
names(smoking) <- c("Age","Smoker","Survived","At.risk")
attach(smoking)
```

```
plot(Age,Survived/At.risk, xlab="Age", ylab="Survival Proportion", col=c("red","blue"),lwd=3)
legend(30,0.2, legend=c("Smokers","Non-smokers"),pch=1, col=c("red","blue"))
```
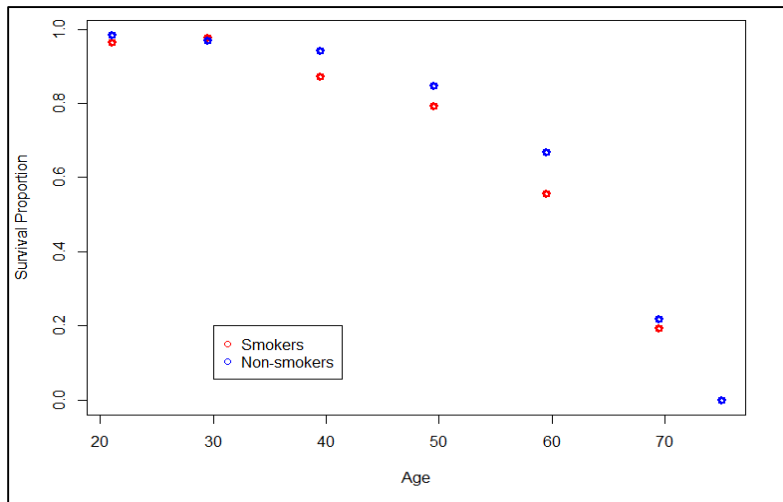
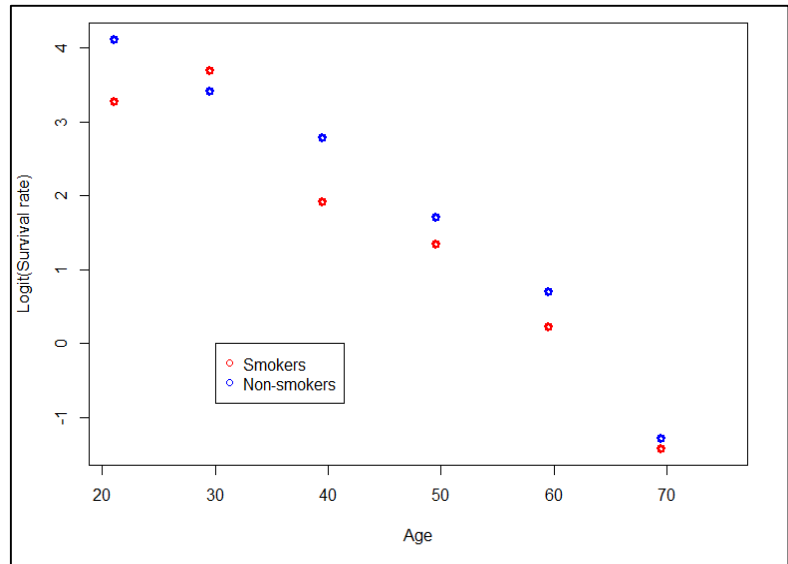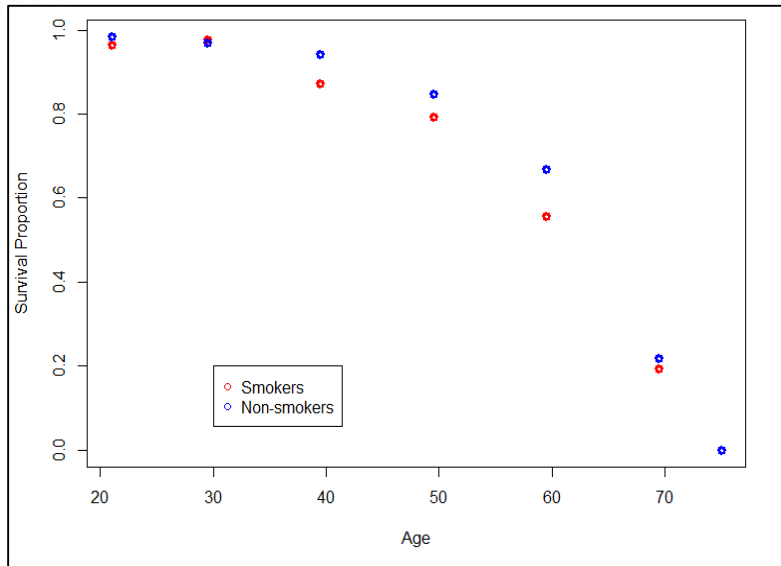# Data Example in R (cont'd)

Plot of logit transformation of the proportion survival
prop.survival <-Survived/At.risk
plot(Age,log(prop.survival/(1-prop.survival)), col=c("red","blue"), xlab="Age", ylab="Logit(Survival Proportion)", lwd=3)
legend(30,0, legend=c("Smokers","Non-smokers"), pch=1, col=c("red","blue"))

Summary