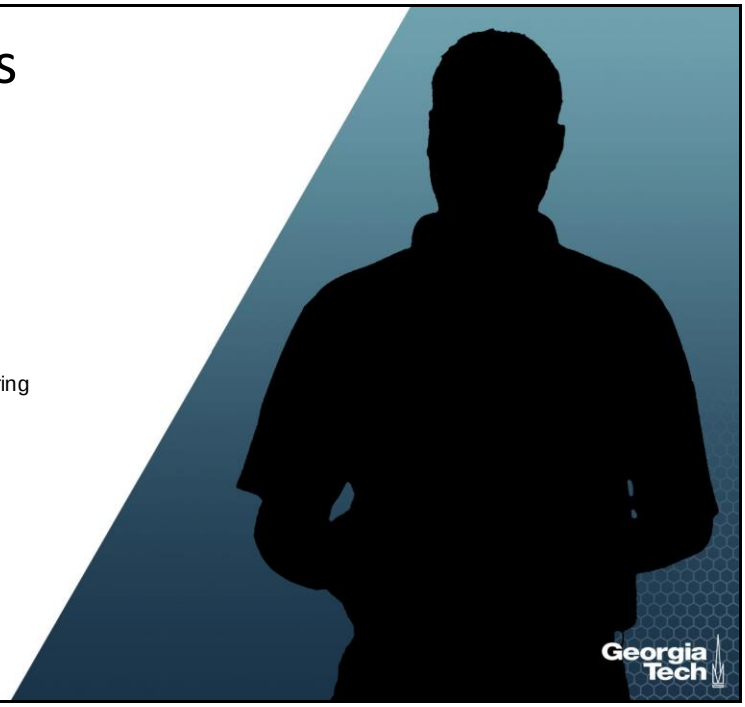
Regression Analysis

Poisson Regression

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



Goodness of Fit Assessment:
Data Examples

Georgia Tech

1

About This Lesson

Georgia Tech

2

Data Example 1: High School Awards

Objective: To model and predict the number of awards earned by students at one high school for multiple high schools.

Response Variable: The number of awards earned by students at a high school per *year*

Predicting Variables:

- The type of program in which the student was enrolled, with three levels: 1 = "General", 2 = "Academic" and 3 ="Vocational"; and
- The score on the final exam in math.

Acknowledgement: This data example was acquired from the Institute for Digital Research and Education at University of California, Los Angeles.

Georgia Tech

3

Goodness-Of-Fit

```
## Deviance Test for GOF
with(m1, cbind(res.deviance = deviance, df = df.residual,
               p = 1- pchisq(deviance, df.residual)))
     res.deviance  df     p
[1,]     189.4496 196 0.6182274
```

Test for goodness-of-fit:

- Using deviance residuals: p-value = 0.61
- Do not reject the null hypothesis of good fit.

Georgia Tech

4

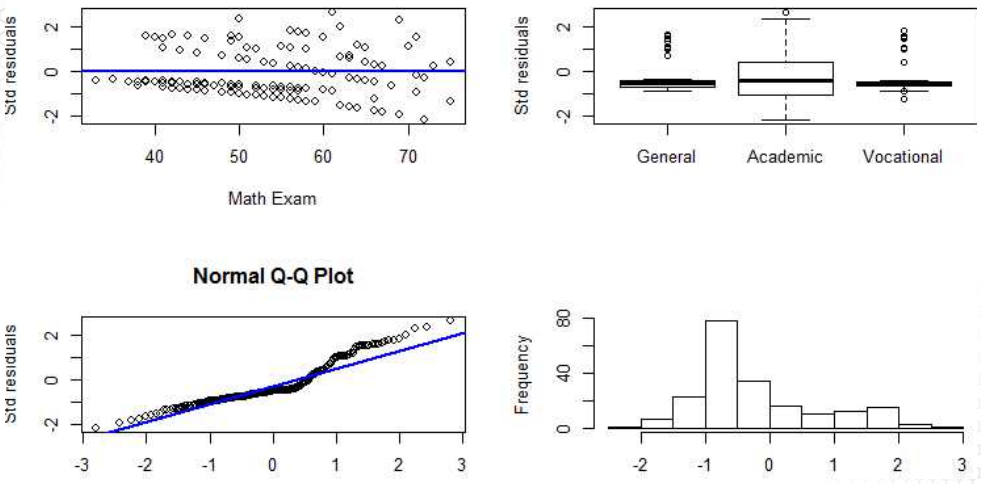
Residual Analysis

```
## Residual Plots
res = resid(m1,type="deviance")
par(mfrow=c(2,2))
plot(awardsdata$math,res,ylab="Std residuals",xlab="Math Exam")
abline(0,0,col="blue",lwd=2)
boxplot(res~prog,ylab = "Std residuals")
qqnorm(res, ylab="Std residuals")
qqline(res,col="blue",lwd=2)
hist(res,10,xlab="Std residuals", main="")
```

Georgia Tech

5

Residual Analysis



Georgia Tech

6

Modeling Nonlinear Relationships

```
## Fit a logistic regression model with math nonlinearly associated to awards count
library(mgcv)
m2 = gam(num_awards ~ prog + s(math), family="poisson", data=awardsdata)
```

- The residuals vs math: downward trend: Consider a **non-parametric** transformation of 'math' predicting variable
- *Nonparametric association*: not specifying the transformation but allowing the data to best identify/fit the transformation
- For this example, we do not see an improvement in the fit.

Georgia Tech

7

Summary

Georgia Tech

8