

Regression Analysis

Regression Analysis in Practice

Nicoleta Serban, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Emergency Department Healthcare
Costs: Exploratory Data Analysis

About This Lesson



Exploratory Data Analysis: Response Variable

```
## Read the data using read.csv() R command
```

```
dataAdult = read.csv("DataADULT.csv", header=TRUE)  
attach(dataAdult)
```

```
## Rescale outcome/response variable
```

```
EDCost.pppm = EDCost/PMPM
```

```
## Rescale utilization
```

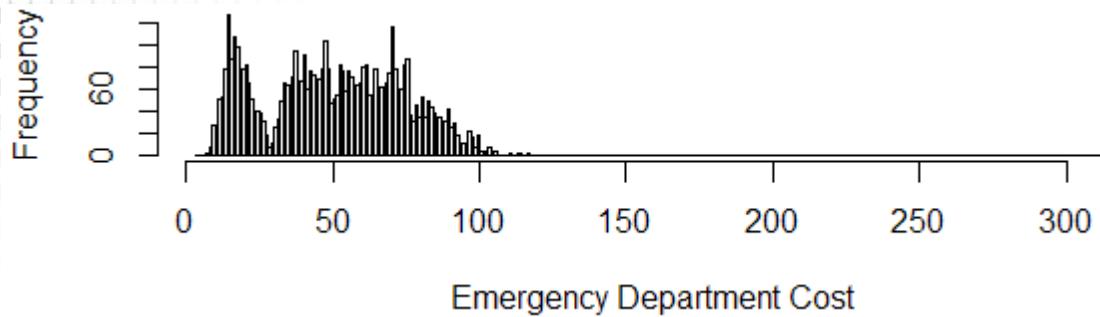
```
dataAdult$PO = PO/PMPM  
dataAdult$HO = HO/PMPM
```

```
## Histogram of the response variable
```

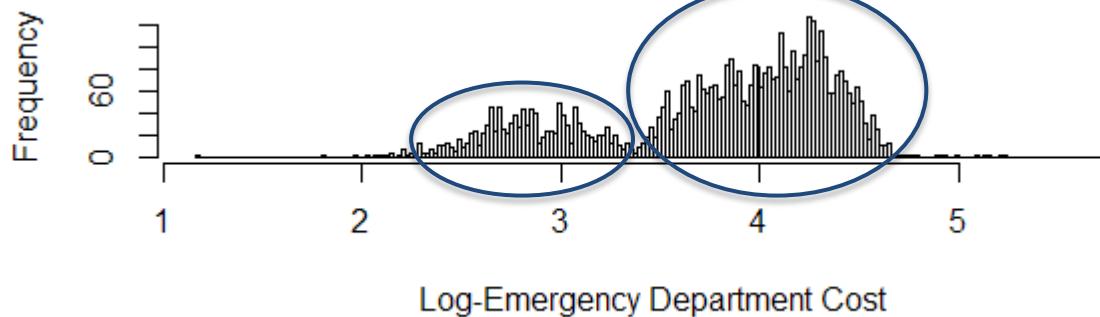
```
par(mfrow=c(2,1))
```

```
hist(EDCost.pppm, breaks=300, xlab="Emergency Department Cost", main="")  
hist(log(EDCost.pppm), breaks=300, xlab="Log-Emergency Department Cost", main="")
```

Exploratory Data Analysis: Response Variable



Emergency Department Cost



Log-Emergency Department Cost

Exploratory Data Analysis: Response vs Qualitative Predictors

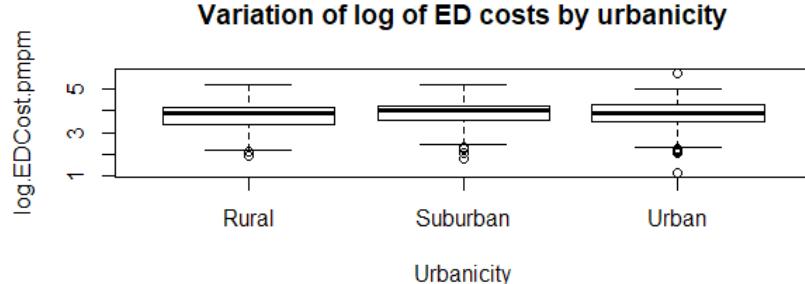
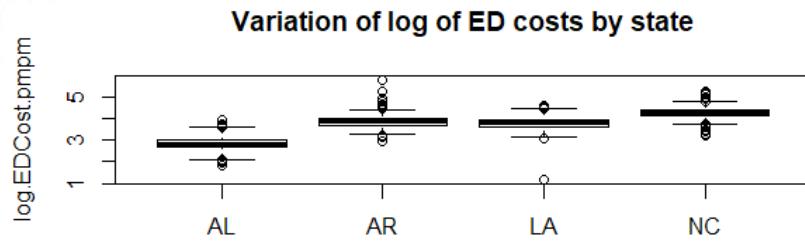
$\log.EDCost.pppm = \log(EDCost.pppm)$

Response variable vs categorical predicating variables

`par(mfrow=c(2, 1))`

`boxplot(log.EDCost.pppm ~ State, main = "Variation of log of ED costs by state")`

`boxplot(log.EDCost.pppm ~ Urbanicity, main = "Variation of log of ED costs by urbanicity")`



Exploratory Data Analysis: Response vs Qualitative Predictors

Scatterplot matrix plots

```
library(car)
```

Response vs Utilization

```
scatterplotMatrix(~ log(EDCost.pppm) + HO + PO, smooth=FALSE)
```

Response vs Population Characteristics

```
scatterplotMatrix(~ log(EDCost.pppm) + WhitePop + BlackPop + OtherPop + HealthyPop +  
ChronicPop + ComplexPop, smooth=FALSE)
```

Response vs Socioeconomic and Environmental Characteristics

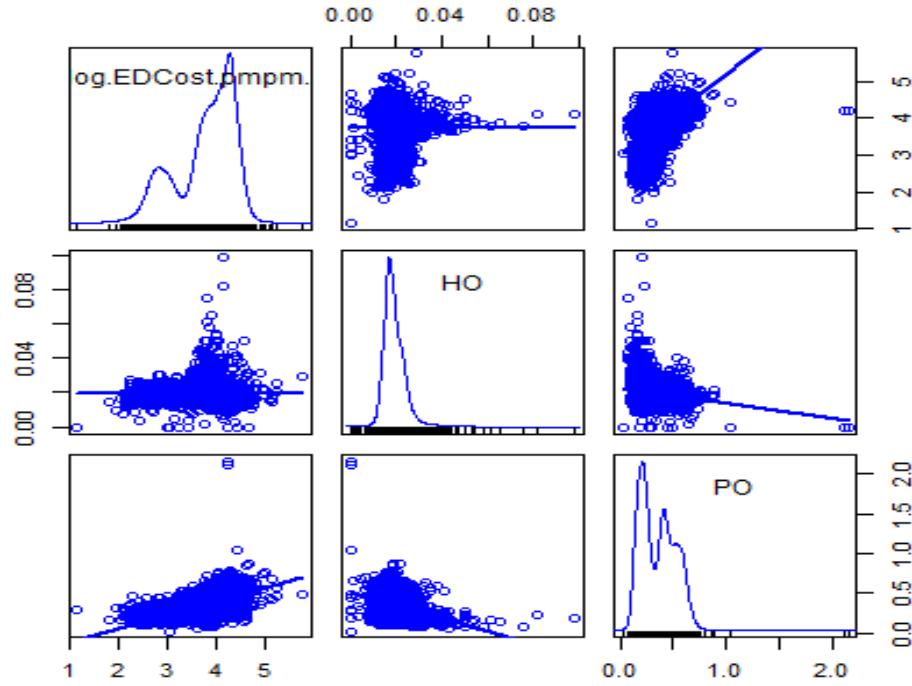
```
scatterplotMatrix(~ log(EDCost.pppm) + Unemployment + Income + Poverty + Education +  
Accessibility + Availability + ProvDensity, smooth=FALSE)
```

Response vs County Health Rankings

```
scatterplotMatrix(~ log(EDCost.pppm) + RankingsPCP + RankingsFood + RankingsHousing +  
RankingsExercise + RankingsSocial, smooth=FALSE)
```

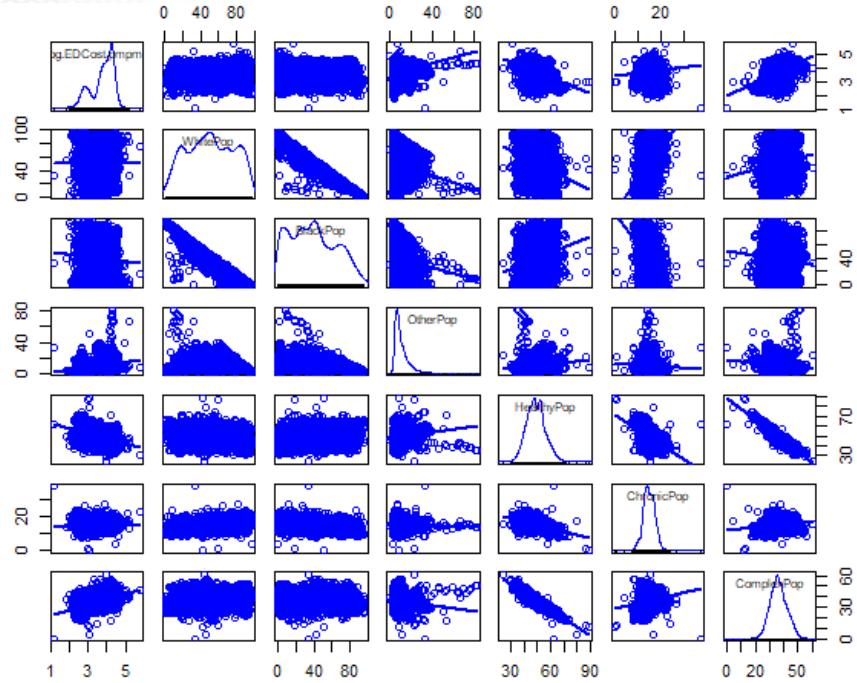
Response vs Quantitative Predictors

ED Cost vs. Utilization Measures: Number of Claims for HO and PO



Response vs Quantitative Predictors

ED Cost vs. Population Characteristics: *WhitePop, BlackPop, OtherPop, HealthyPop, ChronicPop, ComplexPop*



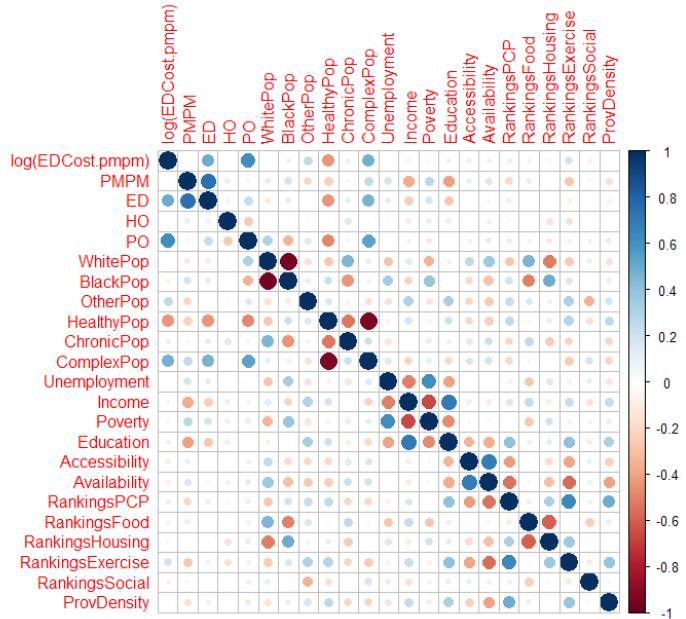
Response vs. Predicting Variables: Correlation Matrix Plot

Correlation matrix plot

```
library(corrplot)
```

```
corr = cor(cbind(log(EDCost.pppm), dataAdult[,-c(1, 2, 3, 18)]))
```

```
corrplot(corr)
```



Summary

