Regression Analysis

Multiple Linear Regression

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

Assumptions and Diagnostics: Data Example

About This Lesson

Linear Regression: Example 1

Quantitative Predicting Variables:

X_1 = The amount (in hundreds of dollars) spent on advertising in 1999

X_2 = The total amount of bonuses paid in 1999

X_3 = The market share in each territory

X_4 = The largest competitor's sales

Qualitative Predicting Variable:

X_5 = Indicates the region of the office (1 = south, 2 = west, 3 = midwest)

Response Variable:

Y = Sales (in thousands of dollars) in 1999

Georgia Tech

Residual Analysis: Example 1

a. Do the assumptions hold? Provide the graphical displays needed to support the diagnostics. Interpret.

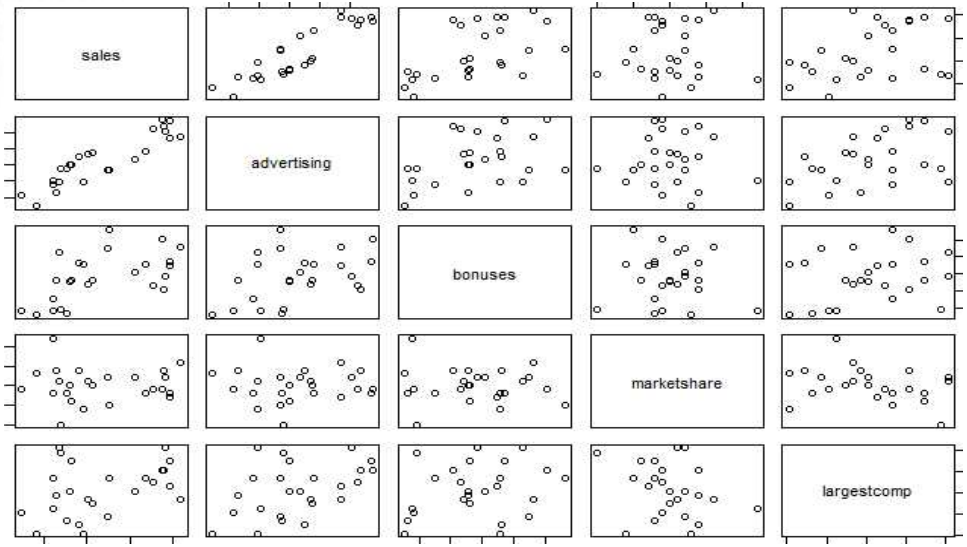
b. If one or more assumptions do not hold, what transformations do you suggest? Did the residual diagnoses improve with the suggested transformations?

c. Do you identify any outliers?

Linearity Assumption

Scatter plot matrix of sales and numeric predicting variables

plot(meddcor[,1:5])

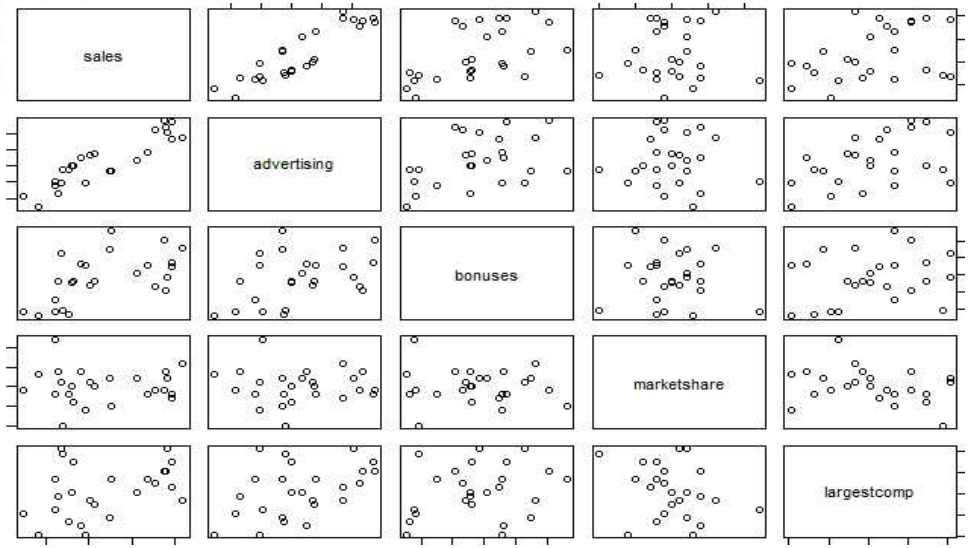


Georgia Tech

Linearity Assumption

Scatter plot matrix of sales and numeric predicting variables

plot(meddcor[,1:5])



- **Linearity assumption holds for all predicting variables.**
- **For advertisement expenditure, bonus amount, and competitor's sales, the relationship with sales is strongly linear.**

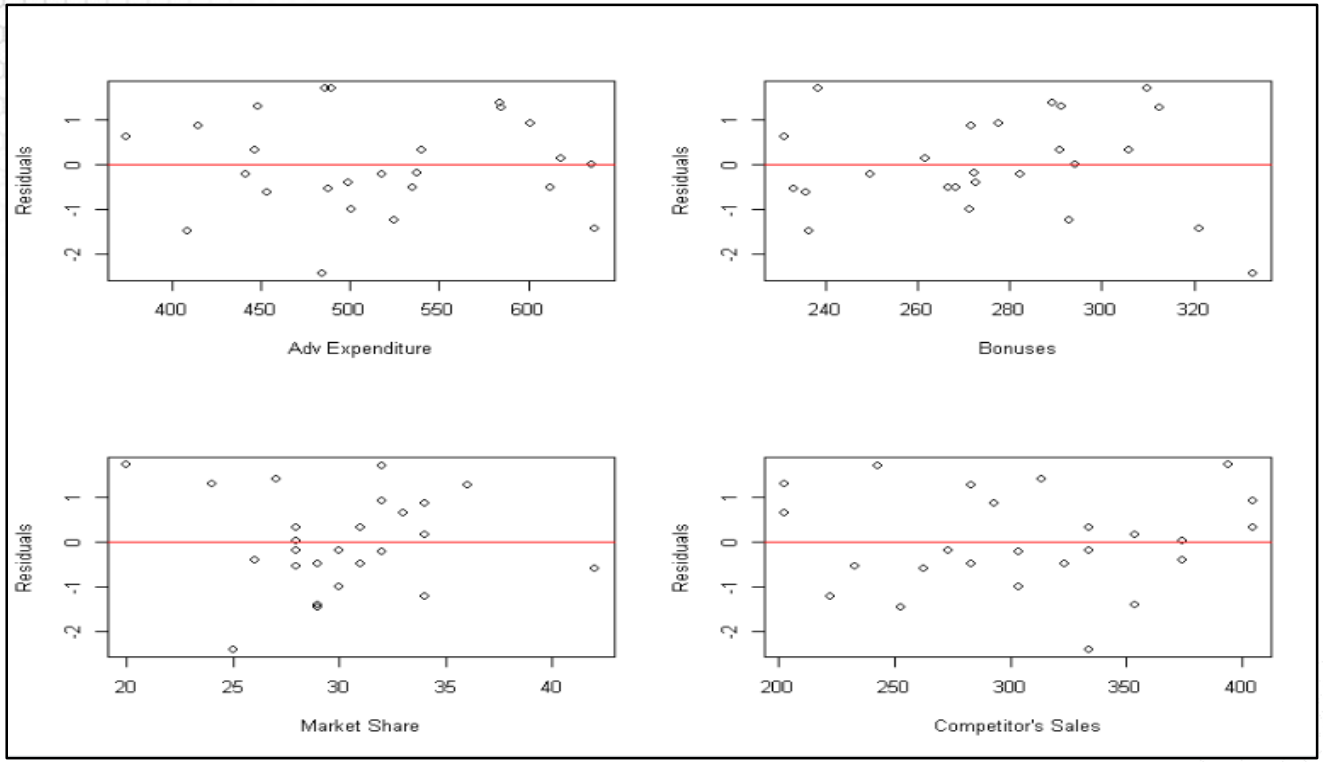
Georgia Tech

Linearity Assumption

```
## Standardized Residuals versus individual predicting variables
resids = stdres(model)
par(mfrow =c(2,2))
plot(meddcor[,2],resids,xlab="Adv Expenditure",ylab="Residuals")
abline(0,0,col="red")
plot(meddcor[,3],resids,xlab="Bonuses",ylab="Residuals")
abline(0,0,col="red")
plot(meddcor[,4],resids,xlab="Market Share",ylab="Residuals")
abline(0,0,col="red")
plot(meddcor[,5],resids,xlab="Competitor's Sales",ylab="Residuals")
abline(0,0,col="red")
```

Georgia Tech

Linearity Assumption



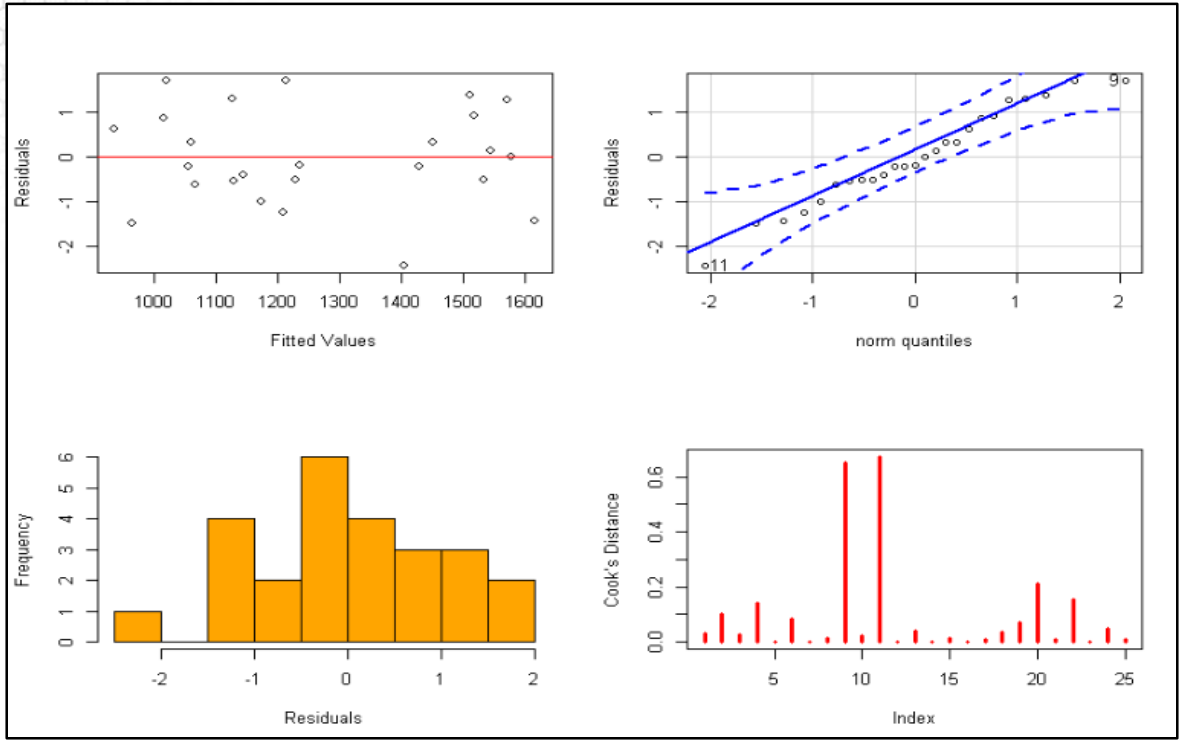
Georgia Tech

Residual Analysis: Other Assumptions

```
library(car)
fits = model$fitted
cook = cooks.distance(model)
par(mfrow =c(2,2))
plot(fits, resids, xlab="Fitted Values",ylab="Residuals")
abline(0,0,col="red")
qqPlot(resids, ylab="Residuals", main = "")
hist(resids, xlab="Residuals", main = "",nclass=10,col="orange")
plot(cook,type="h",lwd=3,col="red", ylab = "Cook's Distance")
```

Georgia Tech

Residual Analysis: Other Assumptions



Summary