

Regression Analysis

Multiple Linear Regression

Nicoleta Serban, Ph.D.

Professor

School of Industrial and Systems Engineering

Assumptions and Diagnostics:
Data Example



1

About This Lesson



2

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



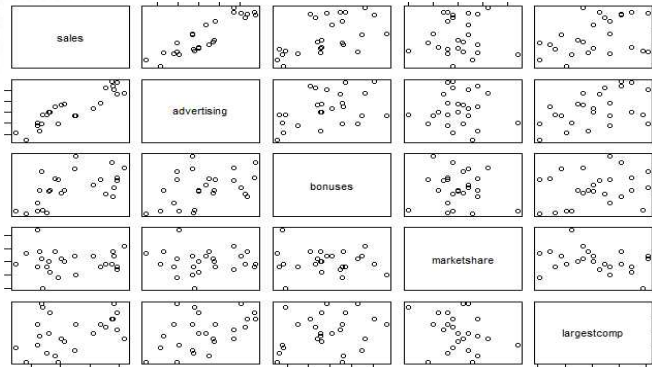
Residual Analysis: Example 1

- 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

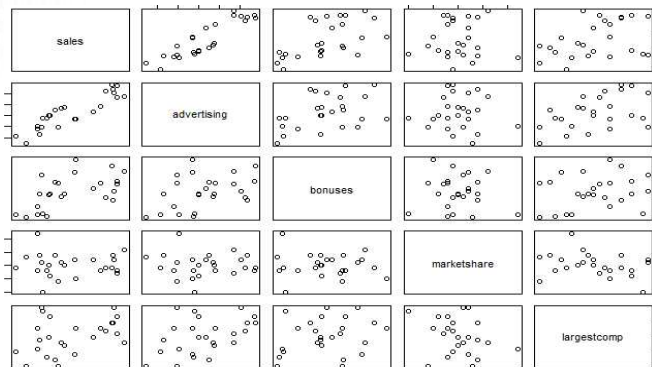
```
plot(medddcor[,1:5])
```



Linearity Assumption

Scatter plot matrix of sales and numeric predicting variables

```
plot(medddcor[,1:5])
```



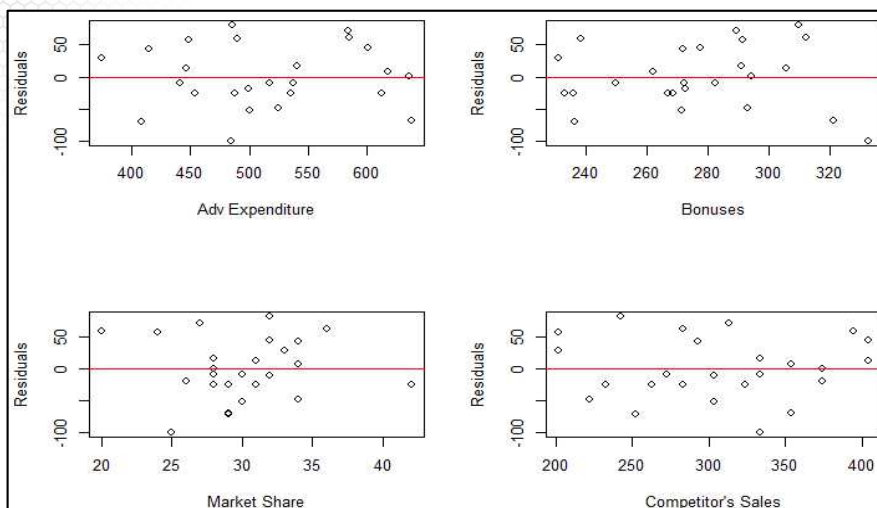
- Linearity assumption holds for all predicting variables.
- For advertisement expenditure, bonus amount, and competitor's sales, the relationship with sales is strongly linear.

Linearity Assumption

Standardized Residuals versus individual predicting variables

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

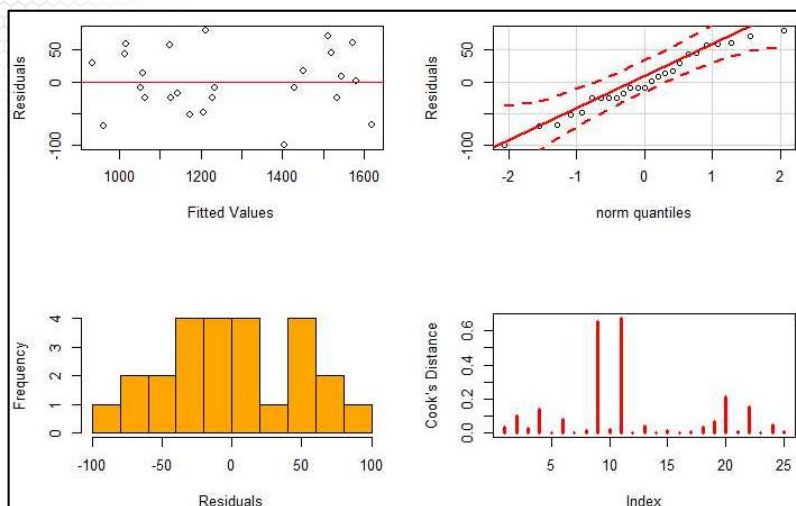
Linearity Assumption



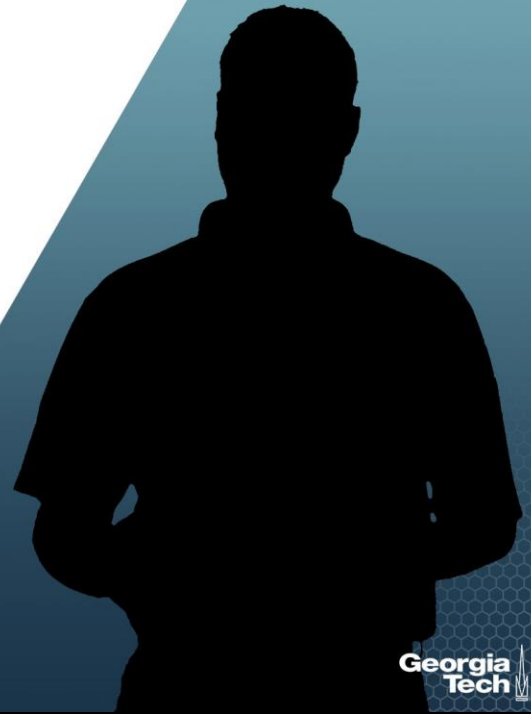
Residual Analysis: Other Assumptions

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

Residual Analysis: Other Assumptions



Summary



Georgia
Tech