Regression Analysis

Multiple Linear Regression

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

Multicollinearity: Data Example

1

About This Lesson

2

Model Evaluation: Example 1

Quantitative Predicting Variables:

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

X_2 = the total amount of bonuses paid

X_3 = the market share in each territory

X_4 = the largest competitor's sales

Qualitative Predicting Variable:

X_5 = a variable to indicate the region in which office is located (1 = south, 2 = west, 3 = midwest)

Response Variable:

Y = yearly sales (in thousands of dollars)

Georgia Tech

3

Model Evaluation: Example 1

a. What are the correlation coefficients between the quantitative predicting variables? Any potential multicollinearity?

b. Obtain the variance inflation factors for the quantitative predicting variables. Any potential multicollinearity?

c. What is the coefficient of determination? Interpret.

Georgia Tech

4

Model Evaluation: Example 1

```
cor(meddcor[,2:5])
```

	Advertising	bonuses	marketshare	largestcomp
advertising	1.00000000	0.41868215	-0.02029937	0.4524897
bonuses	0.41868215	1.00000000	-0.08484673	0.2286563
marketshare	-0.02029937	-0.08484673	1.00000000	-0.2872159
largestcomp	0.45248974	0.22865628	-0.28721592	1.0000000

a. The maximum correlation between predicting variables is **0.452**.

Georgia Tech

5

Model Evaluation: Example 1

```
cor(meddcor[,2:5])
```

	Advertising	bonuses	marketshare	largestcomp
advertising	1.00000000	0.41868215	-0.02029937	0.4524897
bonuses	0.41868215	1.00000000	-0.08484673	0.2286563
marketshare	-0.02029937	-0.08484673	1.00000000	-0.2872159
largestcomp	0.45248974	0.22865628	-0.28721592	1.0000000

```
vif(model)
```

	GVIF	Df	GVIF^(1/(2*Df))
advertising	3.081657	1	1.755465
bonuses	1.359601	1	1.166019
marketshare	1.311265	1	1.145105
largestcomp	1.569851	1	1.252937
region	3.784660	2	1.394783

b. The R function vif() outputs the generalized VIF (GVIF), which specializes to the usual VIF in the case of a single coefficient.

Georgia Tech

6

Model Evaluation: Example 1

```
cor(meddcor[,2:5])
             Advertising    bonuses     marketshare  largestcomp
advertising   1.00000000 0.41868215 -0.02029937   0.4524897
bonuses       0.41868215 1.00000000 -0.08484673   0.2286563
marketshare  -0.02029937 -0.08484673 1.00000000  -0.2872159
largestcomp   0.45248974 0.22865628 -0.28721592   1.0000000

vif(model)
             GVIF       Df GVIF^(1/(2*Df))
advertising  3.081657   1  1.755465
bonuses      1.359601   1  1.166019
marketshare  1.311265   1  1.145105
largestcomp  1.569851   1  1.252937
region       3.784660   2  1.394783

summary(model)$r.squared
0.9555032
```

c. The coefficient of determination is **0.955**. Thus the model explains 95.5% of the variability in sales.

Georgia Tech

7

Summary

Georgia Tech

8