# Regression Analysis

## Multiple Linear Regression

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

Estimating Regression Line & Predicting a New Response: Data Example


1

# About This Lesson


2

# Linear Regression: Example 1

**Quantitative Predicting Variables:**

- $X_1$ = amount (in hundreds of dollars) spent on advertising in 1999
- $X_2$ = total amount of bonuses paid in 1999
- $X_3$ = market share in each territory
- $X_4$ = largest competitor's sales (thousands)

**Qualitative Predicting Variable:**

- $X_5$ = indicates region where office is located (1 = south, 2 = west, 3 = midwest)

**Response Variable:**

- $Y$ = yearly sales (in thousands of dollars)


Georgia Tech

3

# Example 1: Mean Response & Prediction

a. For <u>all</u> offices with the characteristics such as those of the first office:
   - What is the <u>average estimated sales</u>?
   - What is the standard deviation?
   - What is the 95% confidence interval for this mean response?

b. If the first office's largest competitor's sales increase to $303,000 (assuming everything else fixed):
   - What sales would you predict for the first office?
   - What is its standard deviation?
   - What is the 95% prediction interval for this prediction?

Georgia Tech

4

# Example 1: Mean Response Estimation

```
s2 = summary(model)$sigma^2  # Variance estimate
X = model.matrix(model) # Design Matrix
xstar = X[1,] # First office data for formula
resp.var = s2*(xstar%*%solve(t(X)%*%X)%*%xstar) # Variance formula
sqrt(resp.var)

        [,1]
[1,] 33.19118

newdata = xstar[-1] # First office data for confidence interval
predict(model, newdata, interval="confidence") # Confidence Interval

          fit      lwr      upr
1 934.7767   865.0446 1004.509
```

Georgia Tech

5

# Example 1: Mean Response Estimation

```
s2 = summary(model)$sigma^2  # Variance estimate
X = model.matrix(model) # Design Matrix
xstar = X[1,] # First office data for formula
resp.var = s2*(xstar%*%solve(t(X)%*%X)%*%xstar) # Variance formula
sqrt(resp.var)

        [,1]
[1,] 33.19118

newdata = xstar[-1] # First office data for confidence interval
predict(model, newdata, interval="confidence") # Confidence Interval

          fit      lwr      upr
1 934.7767   865.0446 1004.509
```

**a. Average estimated sales (mean response for sales):**

$\hat{y} =$ **934.777**

**Estimated standard deviation:**

$\text{se}(\hat{y}) =$ **33.191**

**95% Confidence Interval:**

**(865.045, 1004.509)**

**Interpretation:** For offices with the same characteristics as the first, the average estimated sales are $934,777, with a lower bound of $865,045 and an upper bound of $1,004,509.

Georgia Tech

6

# Example 1: Mean Response Estimation

```
## Change the competitor's sales for prediction of future observation
xstar.new = xstar
xstar.new[5] = 303

# Variance formula
pred.var = s2*(1+xstar.new%*%solve(t(X)%*%X)%*%xstar.new)
sqrt(pred.var)

       [,1]
[1,] 64.31099

# Prediction Interval
predict(model, xstar.new[-1], interval="prediction")

        fit      lwr      upr
1 911.0569 775.9446 1046.169
```



# Example 1: Mean Response Estimation

```
## Change the competitor's sales for prediction of future observation
xstar.new = xstar
xstar.new[5] = 303

# Variance formula
pred.var = s2*(1+xstar.new%*%solve(t(X)%*%X)%*%xstar.new)
sqrt(pred.var)

       [,1]
[1,] 64.31099

# Prediction Interval
predict(model, xstar.new[-1], interval="prediction")

        fit      lwr      upr
1 911.0569 775.9446 1046.169
```

**b. Predicted sales of the first office given the higher competitor's sales:**

$\hat{y} =$ **911.057**

**Estimated standard deviation:**

$\text{se}(\hat{y}) =$ **64.311**

**95% Confidence Interval:**

**(775.945, 1046.169)**

**Interpretation:** If the competitor's sales increase to \$303,000 (from \$202,220), the predicted sales reduce by \$23,720 (from \$934,777 to \$911,057). Since this is prediction, the standard deviation increases.

# Summary


9