

Regression Analysis

Simple Linear Regression

Nicoleta Serban, Ph.D.

Professor

School of Industrial and Systems Engineering

Regression Concepts: Regression
Line and Prediction Examples



1

About This Lesson



2

Linear Regression: Example in R

A company, which sells medical supplies to hospitals, clinics, and doctor's offices, had considered the effectiveness of a new advertising program. Management wants to know if the advertisement is related to sales.

This company intends to increase the sales with an effective advertising program.

What inferences can be made on the prediction of the sales given a targeted advertisement expenditure?



3

Example in R: Estimating Regression Line & Prediction

- What sales would you predict for an advertisement expenditure of **\$30,000**?
- What is the variance estimate of the estimated predicted sales for an advertisement expenditure of **\$30,000**?
- What are the lower and upper limits of predicted sales for an advertisement expenditure of **\$30,000** at **99%** confidence level? How will the limits change if we lower the confidence level to **95%**?
- Compare the confidence intervals of the estimated regression line versus the predicted regression line. Interpret.



4

Example in R

```
summary(model)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -157.3301  145.1912  -1.084    0.29
adv           2.7721    0.2794    9.921 8.87e-10
---
Residual standard error: 101.4 on 23 degrees of freedom
xbar = mean(ADV)
n = 23+2
mse = 101.4^2
var.beta1 = 0.2794^2
sxx = mse/var.beta1
pred.var = mse*(1+1/n+(xbar-300)^2/sxx)
pred.var
[1] 14286.16
```



5

Example in R

```
summary(model)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -157.3301  145.1912  -1.084    0.29
adv           2.7721    0.2794    9.921 8.87e-10
---
Residual standard error: 101.4 on 23 degrees of freedom
xbar = mean(ADV)
n = 23+2
mse = 101.4^2
var.beta1 = 0.2794^2
sxx = mse/var.beta1
pred.var = mse*(1+1/n+(xbar-300)^2/sxx)
pred.var
[1] 14286.16
```

a. For advertising expenditure of \$30,000, the predicted sales is:
 $-157.33 + 300 \times 2.77$
 $= 673.67$ thousand

b. The variance of the predicted sales is

$$\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}} \right) = 14286.16$$



6

Example in R (cont'd)

```
new = data.frame(adv = 300)
predict.lm(model, new, interval = "predict", level = 0.99)
      fit      lwr      upr
1 674.3047 338.712 1009.897
predict.lm(model, new, interval = "predict", level = 0.95)
      fit      lwr      upr
1 674.3047 427.0146 921.5948
predict.lm(model, new, interval = "confidence", level = 0.99)
      fit      lwr      upr
1 674.3047 496.6497 851.9596
predict.lm(model, new, interval = "confidence", level = 0.95)
      fit      lwr      upr
1 674.3047 543.395 805.2143
```

7

Example in R (cont'd)

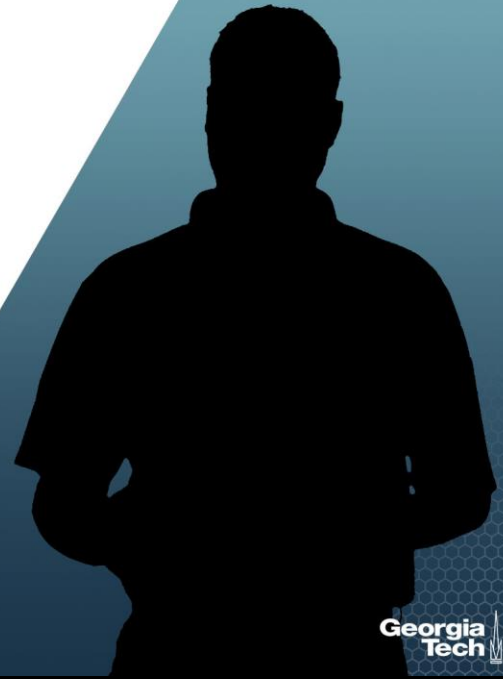
```
new = data.frame(adv = 300)
predict.lm(model, new, interval = "predict", level = 0.99)
      fit      lwr      upr
1 674.3047 338.712 1009.897
predict.lm(model, new, interval = "predict", level = 0.95)
      fit      lwr      upr
1 674.3047 427.0146 921.5948
predict.lm(model, new, interval = "confidence", level = 0.99)
      fit      lwr      upr
1 674.3047 496.6497 851.9596
predict.lm(model, new, interval = "confidence", level = 0.95)
      fit      lwr      upr
1 674.3047 543.395 805.2143
```

- c. A 99% **prediction** interval at an advertisement expenditure of \$30,000 is (338.712, 1009.897). A 95% interval is (427.014, 921.594).
- d. A 99% **confidence** interval at an advertisement expenditure of \$30,000 is (496.649, 851.959). A 95% interval is (543.395, 805.214).

The confidence intervals are narrower than the prediction intervals because the prediction intervals have additional variance from the variation of a new measurement.

8

Summary



Georgia
Tech