Regression Analysis

Multiple Linear Regression

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

Estimating the Regression Line
and Predicting a New Response

1

About This Lesson

2

Estimating the Regression Line

At some selected value of $\boldsymbol{x}$, say $\boldsymbol{x}^*$, estimate the "mean response" of y (the regression line) via

$$\boxed{\hat{Y}|x^* = \hat{\beta}_0 + \hat{\beta}_1 x^*{}_1 + \hat{\beta}_2 x^*{}_2 + \cdots + \hat{\beta}_p x^*{}_p = \boldsymbol{x}^{*\mathrm{T}}\hat{\boldsymbol{\beta}}}$$

- Because the estimators of $\boldsymbol{\beta}$ are normally distributed, so is $\hat{Y}$.
- If we know the expected value and variance, we can use the normal distribution of $\hat{Y}$ to draw inferences on the regression line.

3

Estimating the Regression Line

$\hat{y}$ has a normal distribution with

$$\mathrm{E}(\hat{Y}|\boldsymbol{x}^*) = \boldsymbol{x}^{*\boldsymbol{T}}\boldsymbol{\beta} = \beta_0 + \beta_1 x^*{}_1 + \beta_2 x^*{}_2 + \cdots + \beta_p x^*{}_p$$

$$\mathrm{Var}(\hat{Y}|\boldsymbol{x}^*) = \sigma^2 \boldsymbol{x}^{*\boldsymbol{T}} \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-\mathbf{1}} \boldsymbol{x}^*$$

If we replace the unknown variance with its estimator, $\hat{\sigma}^2$ = MSE, the sampling distribution becomes a *t*-distribution with *n*-*p*-1 degrees of freedom.

4

Confidence Interval for Regression Line

The (1 - α) **Confidence Interval** for the *mean response* (or regression line) for one instance of predicting variables $\mathring{\boldsymbol{x}}$ is:

$$\hat{y}|\mathring{\boldsymbol{x}} \pm t_{\alpha/2,\, n-p-1} \sqrt{\hat{\sigma}^2 \mathring{\boldsymbol{x}}^{\mathrm{T}} (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1} \mathring{\boldsymbol{x}}}$$

The (1 - α) **Confidence Surface** for all possible instances of the predicting variables is:

$$\hat{y}|\mathring{\boldsymbol{x}} \pm \sqrt{(p+1)\mathrm{F}_{\alpha,\, p+1,\, n-p-1}} \sqrt{\hat{\sigma}^2 \mathring{\boldsymbol{x}}^{\mathrm{T}} (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1} \mathring{\boldsymbol{x}}}$$

Georgia Tech

5

Predicting a New Response

- One of the primary motivations for regression is to use the regression equation to predict future responses.
- The predicted regression line is the same as the estimated regression line.
- But a prediction is not the same as the regression line estimation. The prediction contains *two* sources of uncertainty:
  - From the parameter estimates (of βs)
  - From the new observation(s)

Georgia Tech

6

Predicting a New Response *(cont'd)*

1. Variation of the estimated regression line: $\sigma^2 x^{*\mathrm{T}} \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1} x^*$
2. Variation of a new measurement: σ^2

The new observation is independent of the regression data, so the total variation in predicting $\boldsymbol{y}^*|\boldsymbol{x}^*$ is

$$\mathrm{Var}\left(\hat{Y}|\boldsymbol{x}^*\right) = \sigma^2\, \boldsymbol{x}^{*\mathrm{T}}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{x}^* + \sigma^2 = \hat{\sigma}^2\left(1 + \boldsymbol{x}^{*\mathrm{T}}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{x}^*\right)$$

Georgia Tech

7

Predicting a New Response *(cont'd)*

The (1 - α) **Prediction Interval** for <u>one new</u> (future) $\hat{y}^*$ (at $\boldsymbol{x}^*$) is

$$\boxed{\boldsymbol{x}^{*\mathrm{T}}\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,\, n-p-1}\sqrt{\hat{\sigma}^2\left(1 + \boldsymbol{x}^{*\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{x}^*\right)}}$$

$\hat{y} = \boldsymbol{x}^{*\mathrm{T}}\hat{\boldsymbol{\beta}}$ is the same as the line estimate, but the *Prediction Interval* is wider than the *Confidence Interval* for the mean response.

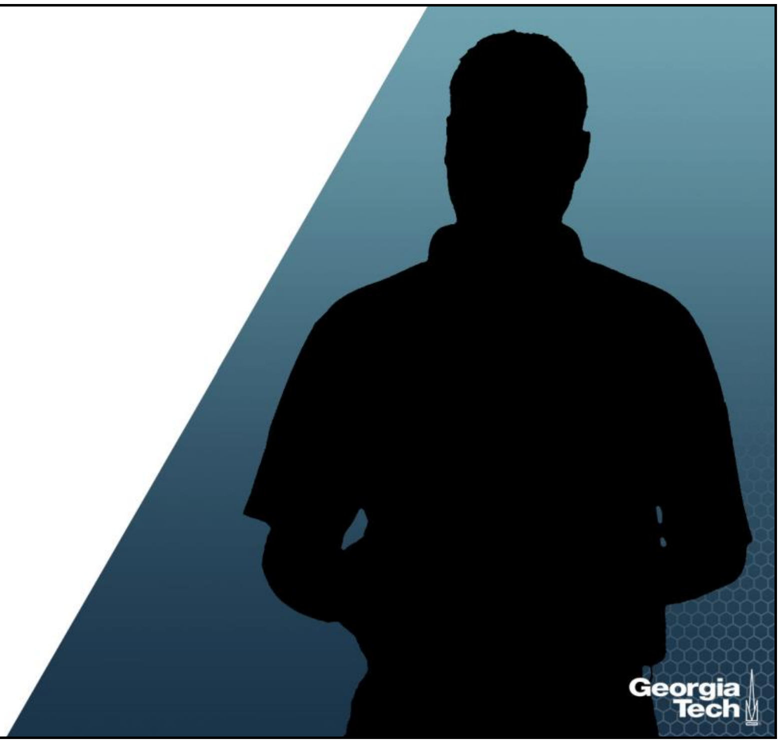
The (1 - α) **Prediction Interval** for <u>*m* new</u> (future) $\hat{y}^*$s (at $\boldsymbol{x}^*$) is

$$\boxed{\hat{y}|\boldsymbol{x}^* \pm \sqrt{m\mathrm{F}_{\alpha,\, m,\, n-p-1}}\sqrt{\hat{\sigma}^2\left(1 + \boldsymbol{x}^{*\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{x}^*\right)}}$$

Georgia Tech

8

Summary



9