

Regression Analysis

Multiple Linear Regression

Nicoleta Serban, Ph.D.

Professor

School of Industrial and Systems Engineering

Inference for Regression Parameters



1

About This Lesson



2

Sampling Distribution

$$E(\hat{\beta}) = \beta$$

$$V(\hat{\beta}) = \sigma^2 (X^T X)^{-1} = \Sigma$$

Furthermore, $\hat{\beta}$ is a linear combination of $\{y_1, \dots, y_n\}$. If we assume that $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$, then $\hat{\beta}$ is also distributed as $\hat{\beta} \sim \text{Normal}(0, \sigma^2)$.

$$\hat{\beta} \sim N(\beta, \Sigma)$$

Properties of Regression Estimators

$$\hat{\beta} \sim N(\beta, \Sigma)$$

σ^2 is unknown!

Replace σ^2 with $\hat{\sigma}^2 = \text{MSE}$

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n-p-1} \sim \chi_{n-p-1}^2$$

(chi-squared
distribution with
 $n-p-1$ degrees
of freedom)

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}} \sim t_{n-p-1}$$

(t-distribution with
 $n-p-1$ degrees of
freedom)

Confidence Interval Estimation

We can derive confidence intervals for β_j using this t sampling distribution:

$$\hat{\beta}_j \pm (t_{\alpha/2, n-p-1})(SE(\hat{\beta}_j))$$

Is β_j statistically significant?

- Check whether zero is in the confidence interval

Why is this a t -interval?

Confidence Interval Estimation

Why is this a t -interval?

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}} = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim T_{n-p-1} \longrightarrow t\text{-interval for } \beta_j$$

$$1-\alpha \text{ Confidence Interval for } \beta_j \longrightarrow \frac{\hat{\beta}_j}{\text{Estimate of } \beta_j} \pm \frac{(t_{\alpha/2, n-p-1})}{t\text{-critical point}} \frac{(SE(\hat{\beta}_j))}{\text{Standard Deviation/Error of } \hat{\beta}_j}$$

Testing Statistical Significance

To test for statistical significance of β_j given all other predicting variables in the model, use a t -test for H_0 and H_a :

$$H_0: \beta_j = 0$$

vs.

$$H_a: \beta_j \neq 0$$

$$t\text{-value} = \frac{\hat{\beta}_j - 0}{SE(\hat{\beta}_j)}$$

- Reject H_0 if $|t\text{-value}|$ gets too large
- Interpret rejecting the null hypothesis as β_j being statistically significant



7

Testing Statistical Significance

How will the procedure change if

we test

$$t\text{-value} = \frac{\hat{\beta}_j - 0}{SE(\hat{\beta}_j)}$$

$$H_0: \beta_j = 0$$

vs.

$$H_a: \beta_j \neq 0$$

for some known b ?

- Reject H_0 if $|t\text{-value}|$ is large
 - For significance level α , if $|t\text{-value}| > t_{\alpha/2, n-p-1} \longrightarrow$ reject H_0
- Alternatively, compute a p-value based on the probability that the t distribution is greater than the t -value:

$$\text{p-value} = 2\text{Prob}(T_{n-p-1} > |t\text{-value}|)$$

- If p-value is small (e.g., < 0.01) \longrightarrow reject H_0



8

Testing Statistical Significance

How will the procedure change if we test whether a coefficient is statistically positive or negative?

Test for Statistically Positive

$$H_0: \beta_j \leq 0$$

vs.

$$H_a: \beta_j > 0$$

$$\text{p-value} = \text{Prob}(T_{n-p-1} > t\text{-value})$$

Test for Statistically Negative

$$H_0: \beta_j \geq 0$$

vs.

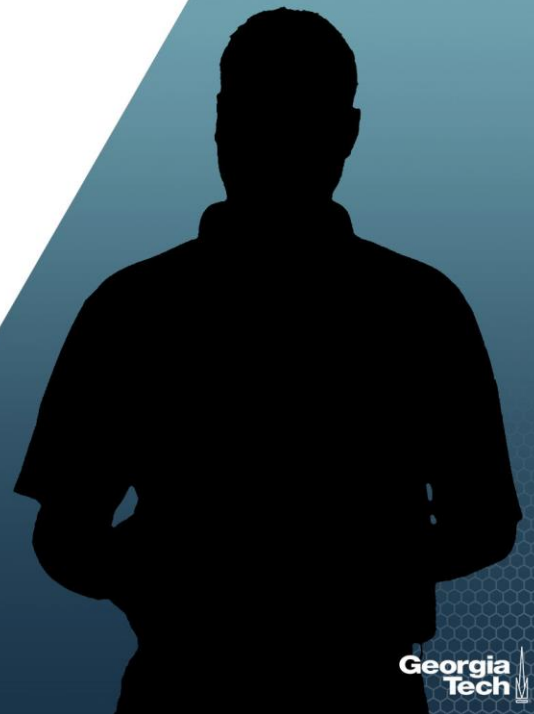
$$H_a: \beta_j < 0$$

$$\text{p-value} = \text{Prob}(T_{n-p-1} < t\text{-value})$$



9

Summary



10