Simple Linear Regression Model – Part III

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 2 – Lecture 2B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Inferential Statistics for Parameters in Simple Linear Regression

In Module 1, we learn how to construct a point estimator of a model parameter (e.g., regression coefficient, residual variance) and derive expectation and variance of the estimator. Also, estimate the variance of the point estimator.

The point estimator and its variance estimator are descriptive statistics.

Q: How to construct *inferential* statistics on these estimators?

Basics for Inferential Statistics

Inferential statistics are often relied on to make statistical inference on the targeted parameter.

Here is a general framework:

θ : parameter targeted for inference

$\hat{\theta}$: point estimator, often unbiased for θ, i.e., $E(\hat{\theta}) = \theta$

$Var(\hat{\theta})$: variance of $\hat{\theta}$

$\hat{V}(\hat{\theta})$: estimator of $Var(\hat{\theta})$, often unbiased for $Var(\hat{\theta})$

that is, $E(\ \hat{V}(\hat{\theta})\ ) = Var(\hat{\theta})$

Basics for Inferential Statistics

Inferential statistics are: **Confidence Interval** for interval estimation, **Test Statistic** for statistical testing

What is needed: Statistical distribution of $\hat{\theta}$. This distribution may come from the distribution of the response variable y or may be obtained approximately via Central Limit Theorem. For example, it is often stipulated that the 1st order statistic (e.g., sample mean) satisfies:

$$\frac{\hat{\theta}-\theta}{\sqrt{Var(\hat{\theta})}} \sim N(0,1) \qquad \frac{\hat{\theta}-\theta}{\sqrt{\hat{V}(\hat{\theta})}} \sim t_{df}$$

Basics for Inferential Statistics

Interval Estimation:

$$\frac{\hat{\theta}-\theta}{\sqrt{\hat{V}(\hat{\theta})}} \sim t_{df} \Rightarrow \Pr\left(-t_{\alpha/2;df} \le \frac{\hat{\theta}-\theta}{\sqrt{\hat{V}(\hat{\theta})}} \le t_{\alpha/2;df} \right) = 1-\alpha$$

where $t_{\alpha/2;df}$ is upper $(\alpha/2)$ percentile of t distribution with df degrees of freedom.

Thus, $100(1 - \alpha)\%$ confidence interval (CI) for θ :

$$\hat{\theta} \pm t_{\alpha/2;df}\sqrt{\hat{V}(\hat{\theta})}$$

Basics for Inferential Statistics

Statistical Testing:

$$H_0: \theta = \theta_0 \text{ (specified)} \quad \text{vs.} \quad H_1: \theta \neq \theta_0$$

Test Statistic: $$\frac{\hat{\theta} - \theta_0}{\sqrt{\hat{V}(\hat{\theta})}}$$

Critical Region (i.e., Rejection Region for H_0):

$$\left| \frac{\hat{\theta} - \theta_0}{\sqrt{\hat{V}(\hat{\theta})}} \right| > t_{\alpha/2;df}$$

Statistical Hypothesis for Slope β_1

$$y_i \mid x_i \sim \text{NIND}(\beta_0 + \beta_1 x_i, \sigma^2),$$
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, \ldots, n$$

Wish to test

$$H_0 : \beta_1 = \beta_{10} \quad , \quad H_1 : \beta_1 \neq \beta_{10}$$

β_{10} is a pre-specified value

Test Statistic and Critical Region for Slope β_1

$$Z_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\sigma^2 / S_{xx}}}$$

is standard normal under H_0. If σ^2 is unknown, use the test statistic

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{MS_{\text{Res}} / S_{xx}}}$$

which is t distribution with df = n - 2 under H_0 . If $|t_0| > t_{\alpha/2, n-2}$, reject H_0 . This is a critical region for β_1 .

Test Statistic and Critical Region for Intercept β_0

Likewise, to test $H_0\colon \beta_0 = \beta_{00}, \quad H_1\colon \beta_0 \neq \beta_{00}$

the test statistic is

$$t_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{MS_{\text{Res}}\left(1/n + \bar{x}^2/S_{xx}\right)}} = \frac{\hat{\beta}_0 - \beta_{00}}{\text{se}\left(\hat{\beta}_0\right)}$$

which is also t distribution with df = n - 2 under H_0 .

If $|t_0| > t_{\alpha/2,n-2}$, reject H_0 . Critical value for β_0 .

P-value can be calculated for the test statistic t_0 .

Rocket Propellant Data (Ex 2.3, page 25 of Textbook)

$$\hat{\beta}_1 = -37.15 \qquad MS_{\text{Res}} = \hat{\sigma}^2 = 9244.59$$

$$\text{se}\left(\hat{\beta}_1\right) = \sqrt{\frac{MS_{\text{Res}}}{S_{xx}}} = \sqrt{\frac{9244.59}{1106.56}} = 2.89$$

$$t_0 = \frac{\hat{\beta}_1}{\text{se}\left(\hat{\beta}_1\right)} = \frac{-37.15}{2.89} = -12.85$$

$$t_{0.025,18} = 2.101 \text{ (critical value)}$$

Reject H_0: $\beta_1 = 0$.

ANOVA (Analysis of Variance) for Testing Slope

Decomposition: $y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$SS_{\mathrm{T}} = SS_{\mathrm{R}} + SS_{\mathrm{Res}} \qquad SS_{\mathrm{R}} = \hat{\beta}_1 S_{xy}$$

$$df_{\mathrm{T}} = df_{\mathrm{R}} + df_{\mathrm{Res}}$$

$$n - 1 = 1 + (n - 2)$$

ANOVA F Test for $H_0: \beta_1 = 0$:

$$F_0 = \frac{SS_R/df_R}{SS_{Res}/df_{Res}} = \frac{SS_R/1}{SS_{Res}/(n-2)} = \frac{MS_R}{MS_{Res}}$$

which is F distribution with dfs being 1 and (n-2) under H_0: $\beta_1 = 0$.

TABLE 2.4 Analysis of Variance for Testing Significance of Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	MS_R/MS_{Res}
Residual	$SS_{Res} = SS_T - \hat{\beta}_1 S_{xy}$	$n-2$	MS_{Res}	
Total	SS_T	$n-1$		

Rocket Propellant Data (Ex 2.4, page 28 of Textbook)

TABLE 2.5 Analysis-of-Variance Table for the Rocket Propellant Regression Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P value
Regression	1,527,334.95	1	1,527,334.95	165.21	1.66×10^{-10}
Residual	166,402.65	18	9,244.59		
Total	1,693,737.60	19			

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING