Estimated Simple Regression Equation

If we choose the parameters a and β in the simple linear regression model so as to minimize the sum of squares of the error term ϵ, we will have the so called **estimated simple regression equation**. It allows us to compute **fitted values** of y based on values of x.

$$\hat{y} = a + bx$$

Problem

Apply the simple linear regression model for the data set faithful, and estimate the next eruption duration if the waiting time since the last eruption has been 80 minutes.

Solution

We apply the lm function to a formula that describes the variable eruptions by the variable waiting, and save the linear regression model in a new variable eruption.lm.

```
> eruption.lm = lm(eruptions ~ waiting, data=faithful)
```

Then we extract the parameters of the estimated regression equation with the coefficients function.

```
> coeffs = coefficients(eruption.lm); coeffs
(Intercept)     waiting
  -1.874016    0.075628
```

We now fit the eruption duration using the estimated regression equation.

```
> waiting = 80           # the waiting time
> duration = coeffs[1] + coeffs[2]*waiting
```

```
> duration
(Intercept)
     4.1762
```

Answer

Based on the simple linear regression model, if the waiting time since the last eruption has been 80 minutes, we expect the next one to last 4.1762 minutes.

Alternative Solution

We wrap the waiting parameter value inside a new data frame named newdata.

```
> newdata = data.frame(waiting=80) # wrap the parameter
```

Then we apply the predict function to eruption.lm along with newdata.

```
> predict(eruption.lm, newdata)    # apply predict
     1
4.1762
```

‹ Simple Linear Regression up Coefficient of Determination ›

Tags: Elementary Statistics with R error term estimated regression equation fitted value linear regression coefficients data.frame lm predict faithful

Hypothesis Testing
Type II Error
Inference About Two Populations
Goodness of Fit
Analysis of Variance
Non-parametric Methods
Simple Linear Regression
Estimated Simple Regression Equation
Coefficient of Determination
Significance Test for Linear Regression
Confidence Interval for Linear Regression
Prediction Interval for Linear Regression
Residual Plot
Standardized Residual
Normal Probability Plot of Residuals
Multiple Linear Regression
Logistic Regression
GPU Computing with R

Recent Articles

- Deep Learning in R
  August 14, 2016

- Installing CUDA Toolkit 7.5 on Fedora 21 Linux
  September 10, 2015
- Installing CUDA Toolkit 7.5 on Ubuntu 14.04 Linux
  September 10, 2015
- Hierarchical Linear Model
  July 22, 2013