Assessing the Utility of the Model

Te Rutherford

March 10, 2015

Agenda

- Last time
  - Sum of Squared Error
  - Fitting the linear regression model with `lm`
- Today
  - Assessing the utility of the model

Fitting the linear regression model

- Method of Least Square is used by `lm` function.

```
d = read.csv('http://www.typ-stats.com/datasets/pressure.cs
lm(compression ~ pressure, data=d)
```

```
##
## Call:
## lm(formula = compression ~ pressure, data = d)
##
## Coefficients:
## (Intercept)     pressure
##        -0.1          0.7
```

Should we need β_1?

$$y = \beta_0 + \beta_1 x + \epsilon$$

- β_0 is called intercept term.
- β_1 is called the slope.

$$y = \beta_0 + \epsilon$$

- Just the intercept term

The P-value for the slope term

```
d = read.csv('http://www.typ-stats.com/datasets/pressure.cs
insanelinearmodel = lm(compression ~ pressure, data=d)
summary(insanelinearmodel)
```

```
##
## Call:
## lm(formula = compression ~ pressure, data = d)
##
## Residuals:
##         1         2         3         4         5
##  4.00e-01 -3.00e-01 -3.89e-16 -7.00e-01  6.00e-01
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.100      0.635   -0.16    0.885
## pressure       0.700      0.192    3.66    0.035 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

The P-value

- If p-value for the slope term is less than 0.05, then the slope term significantly contributes to the prediction of the dependent variable y using the linear model.
- The p-value is at 0.0354. It even has a star next to it.

R-Squared or the coefficient of determination

- SSE is meaningful when comparing models. (relative term)
- R-Squared or the coefficient of determination indicates how well the model does in the absolute term.

What is R-Squared?

- The improvement on SSE from adding β_1.
- Compute SSE from Model 1

$$y = \beta_0 + \beta_1 x + \epsilon$$

- Compute SSE from Model 2

$$y = \beta_0 + \epsilon$$

- Compute the percent reduction of SSE.

Understanding R-Squared

- What happens when R-Squared is low?
- What happens when R-Squared is high?