



Rensselaer

why not change the world?®

Evaluating Regression Models

Ahmed Eleish

Data Analytics ITWS/MGMT-4600/6600, CSCI-4600, BCBP-4600

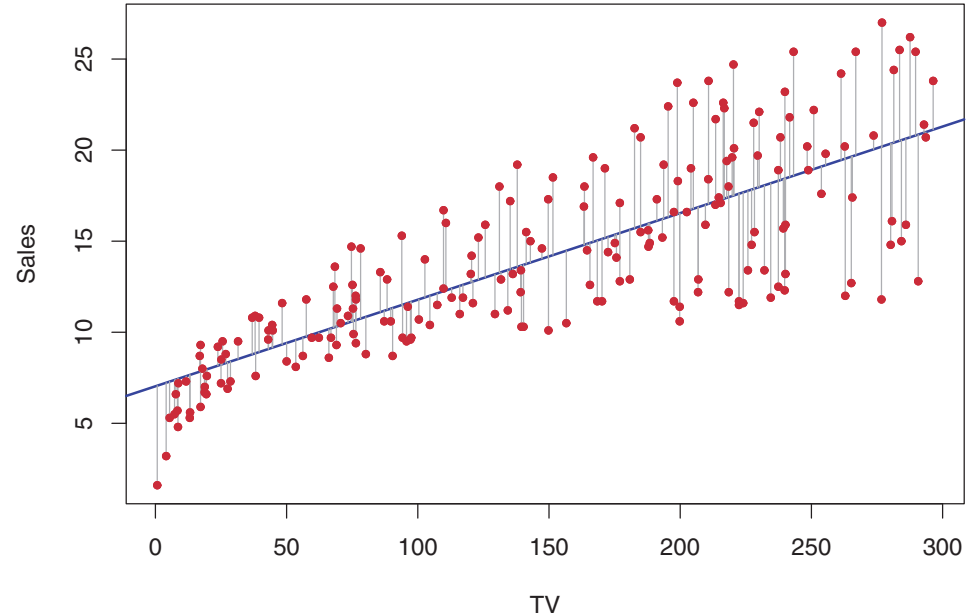
November 3rd 2025

Tetherless World Constellation
Rensselaer Polytechnic Institute



Evaluating Linear Models

- Sales vs. TV ad spending
- Sales in 1000s of units
- TV ad spending in 1000s of \$



Evaluating Linear Models

	Coefficient	Std. error	t-statistic	p-value
Intercept	7.0325	0.4578	15.36	< 0.0001
TV	0.0475	0.0027	17.67	< 0.0001

Values of coefficients >> their Std. errors

High t-statistic

Very low p-value

$$t = \frac{\hat{\beta}_1 - 0}{\text{SE}(\hat{\beta}_1)},$$

Residual Standard Error

- Mean sales $\approx 14,000$ units

RSE = 3.26 = 3,260 units
good/bad?

- R^2
- measures the proportion of the variability in Y that can be explained using X
 - has a value between 0,1

Quantity	Value
Residual standard error	3.26
R^2	0.612
F-statistic	312.1

$$RSE = \sqrt{\frac{1}{n-2}RSS} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$TSS = \sum (y_i - \bar{y})^2$$

Residual Sum of Squares (RSS)

For given data $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R} \times \mathbb{R}$,

- Residual Sum of Squares (RSS), the i th residual $e_i = y_i - \hat{y}_i$

$$\text{RSS} = e_1^2 + e_2^2 + \dots + e_n^2$$

Mean Absolute Error

- Mean(||Predicted value - Real value||)

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$

Mean Squared Error

- Mean((Predicted value - Real value)²)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Root Mean Squared Error

- SquareRoot(Mean((Predicted value - Real value)²))

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n}}$$

Thanks!

