Rensselaer

why not change the world?®

Evaluating Regression Models, Error Estimation

Ahmed Eleish

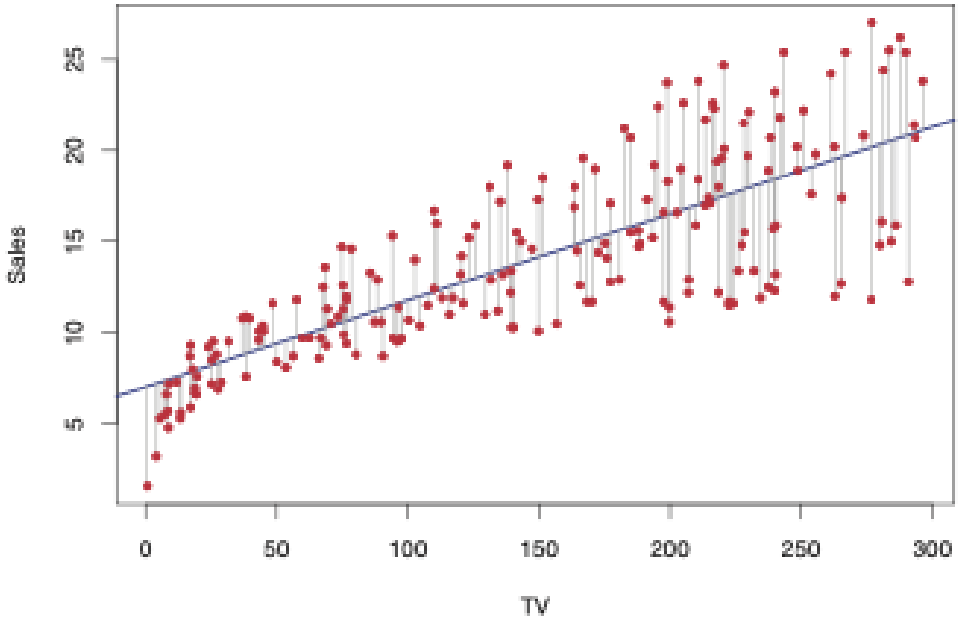
Data Analytics ITWS-4600/ITWS-6600/MATP-4450/CSCI-4960

November 19th 2024

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

Values of coefficients >> their Std. errors

High t-statistic

Very low p-value

	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

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

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

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

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

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

Mean Absolute Error

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

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}$$

Mean Squared Error

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

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Root Mean Squared Error

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

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

in-class exercise

https://rpi.box.com/s/91ur50ga346s7dyxp8vg09723pu9rv2c

Thanks!