Root-Mean-Square Error (RMSE)

- The root-mean-square error (RMSE) is a frequently used measure of the difference between values (sample and population values) predicted by a model or an estimator and the values actually observed.
- The RMSE represents the sample standard deviation of the differences between predicted values and observed values.
- These individual differences are called ***residuals*** when the calculations are performed over the data sample that was used for estimation, and are called ***prediction errors*** when computed out-of-sample.
- The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power.
- RMSE is a measure of accuracy, to compare forecasting errors of different models for a particular data and not between datasets, as it is scale-dependent.
- (Outliers) RMSE is the square root of the average of squared errors. The effect of each error on RMSE is proportional to the size of the squared error; thus larger errors have a disproportionately large effect on RMSE. Consequently, RMSE is sensitive to outliers.

Mean Absolute Error (MAE)

- The mean absolute error (MAE) is a measure of difference between two continuous variables.
- Assume X and Y are variables of paired observations that express the same phenomenon.
    - Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement.

Consider a scatter plot of n points, where point i has coordinates (x_i, y_i).... The Mean Absolute Error (MAE) is the average vertical distance between each point and the Y=X line, which is also known as the One-to-One line. MAE is also the average horizontal distance between each point and the Y=X line.

The Mean Absolute Error is given by:

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

The Mean Error is given by:

$$\mathrm{ME} = \frac{\sum_{i=1}^{n} y_i - x_i}{n}.$$

It is also possible to identify the types of difference by looking at an (x, y) plot. Allocation difference exists if and only if points reside on both sides of the Y=X line. Quantity difference exists when the average of the X values does not equal the average of the Y values.

```
> library(modelr)
>
> car.model <- lm(mpg ~ cyl + wt, data=mtcars)
>
> rmse(car.model,mtcars)
[1] 2.444202
>
> mae(car.model,mtcars)
[1] 1.921012
>
> rsquare(car.model,mtcars)
[1] 0.8302274
```

Summary

- Some researchers have recommended the use of Mean Absolute Error (MAE) instead of Root Mean Square Error.
- MAE possesses advantages in interpretability over RMSE. MAE is the average absolute difference between two variables designated X and Y. MAE is fundamentally easier to understand than the square root of the average of squared errors.
- Furthermore, each error influences MAE in direct proportion to the absolute value of the error, which is not the case for RMSE.