Exercise 2.1

Part 1. Concepts

$$\textbf{RMSE}\ (\text{Root Mean Square Error}) = \sqrt{\frac{\sum_{i=1}^{i=n} (\hat{y}_i - y_i)^2}{n}}$$

Q2.1. Explain why RMSE is a good metric of model predictive accuracy for continuous Y.

Q2.2. Can we use RMSE for categorical Y? Explain.

Q2.3. Netflix used RMSE in their US$1 million prize. What is the implication?

Confusion Matrix

		Actual	
		Not Fraud	Fraud
Model Prediction	Not Fraud	10	17
	Fraud	3	20

Q2.4: Is this a good result? Comment.

Q2.5: What is the (a) true positive, (b) false positive?

Q2.6: What is the (a) true negative, (b) false negative?

Q2.7: What is the overall error rate? Some algorithm and researcher reported only the overall error rate. Ok or not?

Q2.8 If Model Prediction Error = 0, it means the model is excellent for use. True/False? Explain.

Q2.9 We must always do Train-Test split in every analytics model. True/False? Explain.

Part 2. Run R Script

R script file: BA1w2 baby.R

Dataset: baby.csv

Objectives

- Learn how to use the comments operator #
  - Annotate and explain your code to a human
  - Record your results [optional]
- Learn how to create data within R.
- Learn how to create a dataset within R.
- Learn how to calculate simple statistics in R.
- Learn how to do simple plot in R.
- Learn how to export the final dataset in R to CSV format.

Run BA1w2 baby.R

- Run code one line at a time to check if there is any error in that line.