Modeling with Logistic Regression Hands-On

Testing assumptions

1. Meets minimum sample size

After predicting the presence or absence of gold, the following confusion matrix was generated.

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 34  8
         1  2 20

               Accuracy : 0.8438
                 95% CI : (0.7314, 0.9224)
    No Information Rate : 0.5625
    P-Value [Acc > NIR] : 1.615e-06

                  Kappa : 0.6748

 Mcnemar's Test P-Value : 0.1138

            Sensitivity : 0.9444
            Specificity : 0.7143
         Pos Pred Value : 0.8095
         Neg Pred Value : 0.9091
             Prevalence : 0.5625
         Detection Rate : 0.5312
   Detection Prevalence : 0.6562
      Balanced Accuracy : 0.8294

       'Positive' Class : 0
```

Since one of the cells in the confusion matrix has a value less than 5, this **dataset DOES NOT meet the minimum sample size for binary logistic regression.**

Also, an accuracy rate of 0.8438 shows that our predictions are right 84% of the time.

2. Logit linearity
   From the graph below, the predictor variable shows a strong linear relationship, hence passes this assumption.