LING 3802 Final Project

Eric Perez, Ben Rogers, Spencer Black

4/14/2021

We want to figure out how well our SPAM detector works. We will be testing whether or not our SPAM detector has an accuracy of greater than 80% (detecting whether an email is SPAM or not SPAM). In order to do so we will set up a model that p_0 is the true proportion of correct detection. Our model will thus be...

$$H_0 : p_0 = 0.80 \text{ and } H_a : p_0 > 0.80$$

```
#The effect of Yate's correction is to prevent overestimation of statistical significance
#for small data. For this reason I set correct = FALSE
#since we are working with large data.

prop.test(7480,9147,p = .8, alternative = "greater",
          correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  7480 out of 9147, null probability 0.8
## X-squared = 18.021, df = 1, p-value = 1.093e-05
## alternative hypothesis: true p is greater than 0.8
## 95 percent confidence interval:
##  0.8110214 1.0000000
## sample estimates:
##         p
## 0.8177545
```

The p-value is 1.093e-05 which is approximately zero. This means at $\alpha = 0.05$ we can say with strong evidence that we can reject the null hypothesis. This means that there is strong evidence that the true proportion of success or accuracy of our SPAM detector is greater than 80%.