BurnettC Final

Clay Burnett

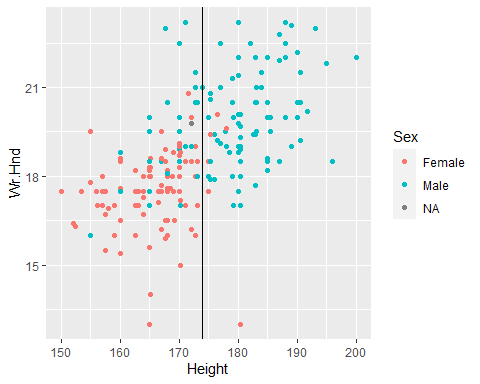
5/6/2022

## Introduction: Predicting male vs female

This report will use the “survey” data. This is a data set contains responses of 237 Statistics students at the University of AdelaideIn to a number of questions. In this report we will be examining height (the height of the student) wr.hnd (the span of the students writing hand in centimeters), and sex (the sex of the student). This report we will be using machine learning to determine if a given student is male or female. We will test our accuracy using three factors, Accuracy (the number correctly predicted divided by the total), Sensitivity (predicted female vs actual female), and specificity (predicted male vs actual male). Our first trial will use one predictor and our second trial will attempt to improve on the first using two predictors.

## Model 1 Vizualization

## Warning: Removed 29 rows containing missing values (geom\_point).



Using height and wr.hnd we can see a pretty clear split in the data. We elect to start with height and after some trial and error establish that 174 is a spot for our deciding line.

## Model 1 Results

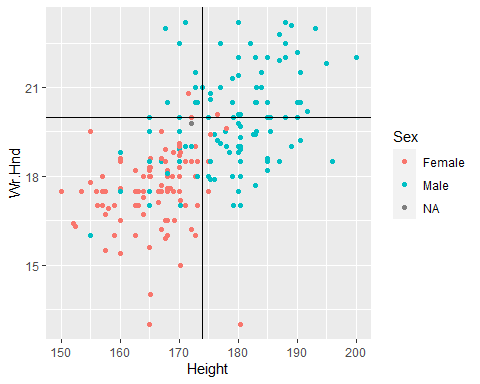
## Actual  
## Predict Female Male  
## 0 95 29  
## 1 7 77

## Metric Value  
## 1 Accuracy 0.827  
## 2 Sensitivity 0.726  
## 3 Specificity 0.931

For our first trial we get an excellent specificity. This is expected and can be seen in our plot as almost all of the points above 174 are male. This is good but we would like to improve on the other numbers.

## Model 2 visualization

## Warning: Removed 29 rows containing missing values (geom\_point).



We will now incorporate the writing hand data into our prediction and again after some trial and error have chosen 20 to be our deciding point.

## Model 2 results

## Actual  
## Predicted Female Male  
## Female 94 20  
## Male 9 87

## Metric Value  
## 1 Accuracy 0.862  
## 2 Sensitivity 0.813  
## 3 Specificity 0.913

These numbers look much better as there is not as much disparity between them. We maintained a high number for specificity and increased both other metrics. The biggest thing we accomplished here was raising our sensitivity from .726 to .813.

## Conclusion

We used height for our first predictor stating that if height was less than 174 we would classify the point as female. This gave us the metrics of Accuracy .827, Sensitivity .726, and Specificity .931. These are decent metrics but we would like to improve specifically on sensitivity. In our second test we incorporated the span of the writing hand stating that we would predict female if height was below 174 and the span of the writing hand was below 20, and for all other points predict male. This gave us the metrics of Accuracy .862, Sensitivity .813, and Specificity .913. While these metrics are not perfect they are much better than our first trial and by sacrificing only a little specificity we improved both accuracy and sensitivity. Ideally we would like to have all three predictors above .9 but for this particular data set the sample size causes a few outliers to have a large affect on our numbers.