Something About Prostate Cancer

Andrew Bates

November 15, 2018

Executive Summary

This is the executive summary. Unfortunately, it has to be typed in the .yaml header. The only other thing i can think of is to have a separate file for the abstract. I'm not sure I want to do this. But actually, this may not be so bad. Each section could have a different file. this might make things a bit easier to edit. because then i would only have to look at say, the conclusion file instead of having to scroll all the way down, passing it, srolling up, passing it, etc.

1 Introduction

In this paper, we investigate the relationship between the results of various prostae related exams and whether tumor penetration of the prostatic capsule has occured. The objective of this analysis is to determine if there are any factors that have a particularly influential relationship with prostatic capsule penetration. Additionally, we wish to develop a model to predict capsule penetration so it can be used as a diagnostic tool for future patients.

2 Methods

In this analysis we examine a subset of data collected by the Ohio State University Comprehensive Cancer Center as part of a study to determine the potential of standard exam results to predict whether a tumor will penetrate the prostatic capsule. Out of 380 patients, 153 have experienced capsule penetration and 227 have not. The data set contains six explanatory variables: the patients age and race, results of a digital rectal exam, whether capsular involvement was detected, Prostatic Specific Antigen (PSA) value, and total Gleason score. For a detailed description of each variable see Table 1. Observe that race is recorded as only Black or White. This may be the reason why three observations contain missing values for race. Perhaps three of the patients were neither Black nor White. However, without access to details of the study and the population considered, we can only speculate. Furthermore, the other variables in these observations do not point to any particular reason as to why race is not recorded[1]. Because of this, we decided to omit these three observations. The data set used for analysis then, consists of 377 observations. Of these, 151 patients have experienced tumor penetration of the prostatic capsule, and 226 have not.

To investigate the relationship between the covariates and tumor penetration status, we use a logistic regression model. The response is especially well balanced with 40% of the observations having capsular penetration and 60% not having penetration. As such, procedures designed to assist with class imbalances, such as downsampling, are not considered in this analysis. In addition to the explanatory variables included in the data set, all two-way interaction terms are examined. Using this as a starting point, the model is chosen via a backwards elimination procedure using Akaike

[1] As in, some have experienced capusle penetration, the ages and PSA scores vary significantly across the observations, etc.

Table 1: Description of variables in the data set.

Name	Description	Details
penetrate	Tumor penetration of prostatic capsule?	Yes, no
age	Patient age	Years
race	Patient race	Black, White
dre	Results of digital rectal exam	No nodule, unilobar left, unilobar right, bilobar
caps	Detection of capsular involvement?	Yes, no
psa	Prostatic Specific Antigen value	mg/ml
gleason	Total Gleason score	0-10

Information Criteria (AIC) as the model selection criterion. (variable name) is not significant at the 0.05 level so it is subsequently removed, along with its corresponding interaction terms.

Diagnostics take the form of examining explanatory variable patterns. also do blah test.

log odds ratios were calculated, and predictions were made. predictive power was examined through a confusion matrix. What's more important here, false negatives or false positives?

3 Analysis

3.1 Exploratory Analysis

We begin by inspecting contingency tables.

3.2 Modeling and Diagnostics

blah blah blah blah blah blah blah blah blah

4 Conclusion

Appendix

A Exploratory Analysis

A.1 Tables

A.2 Figures

B Model Building and Diagnostics

B.1 Intermediate Models

B.2 Diagnostics

C R Code