Predicting Cases of Cervical Cancer in the 'Hospital Universitario de Caracas'

Ben Herndon-Miller and Nicole Jaiyesimi

11/25/2018

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Introduction

Data Description

The data was collected at Hospital Universitario (which is located in Caracas Venezuela) in 2017. There are 36 attributes in the dataset, all of which are either integers or booleans (e.g. age, number of sexual partners, whether or not the individual smokes, whether or not the individual has AIDS). There are 858 instances of the 36 attributes, but there are some missing values; we will conduct a coverage analysis of the data in our final project. We examine the number of missing values for each variable.

Variable	Missing_Values	Percent_Missing
Age	0	0.0000000
Number.of.sexual.partners	26	0.0303030
First.sexual.intercourse	7	0.0081585
Num.of.pregnancies	56	0.0652681
Smokes	13	0.0151515
Smokes..years.	13	0.0151515
Smokes..packs.year.	13	0.0151515
Hormonal.Contraceptives	108	0.1258741
Hormonal.Contraceptives..years.	108	0.1258741
IUD	117	0.1363636
IUD..years.	117	0.1363636
STDs	105	0.1223776
STDs..number.	105	0.1223776
STDs.condylomatosis	105	0.1223776
STDs.cervical.condylomatosis	105	0.1223776
STDs.vaginal.condylomatosis	105	0.1223776
STDs.vulvo.perineal.condylomatosis	105	0.1223776
STDs.syphilis	105	0.1223776
STDs.pelvic.inflammatory.disease	105	0.1223776
STDs.genital.herpes	105	0.1223776
STDs.molluscum.contagiosum	105	0.1223776
STDs.AIDS	105	0.1223776

Variable	Missing_Values	Percent_Missing
STDs.HIV	105	0.1223776
STDs.Hepatitis.B	105	0.1223776
STDs.HPV	105	0.1223776
STDs..Number.of.diagnosis	0	0.0000000
STDs..Time.since.first.diagnosis	787	0.9172494
STDs..Time.since.last.diagnosis	787	0.9172494
Dx.Cancer	0	0.0000000
Dx.CIN	0	0.0000000
Dx.HPV	0	0.0000000
Dx	0	0.0000000
Hinselmann	0	0.0000000
Schiller	0	0.0000000
Citology	0	0.0000000
Biopsy	0	0.0000000