MSCR 509: Homework Assignment 1

Anish Shah, MD

January 27, 2020

Assignment Description

The treatment regimen for patients who have been diagnosed as having cancer of the prostate is crucially dependent upon whether or not the cancer has spread to the surrounding lymph nodes. Indeed, a laparotomy (a surgical incision into the abdominal cavity) may be performed to ascertain the extent of this nodal involvement. However, there are a number of variables that are indicative of nodal involvement that can be measured noninvasively, and the aim of a study reported by Brown (1980) was to determine whether a combination of variables could be used to forecast whether or not the cancer has spread to the lymph nodes. The response variable (Y) is the presence or absence of nodal involvement (Y: 1 = present, 0 = absent).

The prognostic variables considered are:

- AGE – age of patient at diagnosis (in years)
- ACID – level of serum acid phosphatase (in King-Armstrong units)
- XRAY – the result of an X-ray examination (0 = negative, 1 = positive)
- SIZE – the size of the tumor as determined by a rectal examination (0 = small, 1 = large)
- GRADE – a summary of the pathological grade of the tumor determined from a biopsy (0 = less serious, 1 = more serious)
- SES – income (high, middle, low)

Data intake

The data frame to be loaded can be seen as below, in its raw form.

```
# Show first few lines
head(prostate)
```

```
## # A tibble: 6 x 11
##      age  acid  xray  size grade     y    id income  qual enzyme ses
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl>  <dbl> <chr>
## 1    66  0.48     0     0     0     0     1      1 0.122   2.40 high
## 2    68  0.56     0     0     0     0     2      1 0.387   3.12 high
## 3    66  0.5      0     0     0     0     3      1 0.970   2.55 high
## 4    56  0.52     0     0     0     0     4      1 0.353   2.50 high
## 5    58  0.5      0     0     0     0     5      1 0.119   3.21 high
## 6    60  0.49     0     0     0     0     6      1 0.648   3.23 high
```

This will need to be modified to be *publication ready* by adding in the appropriate labels.

Table presentation

```
# Data set
df <- prostate
```

```r
## Relabel the variables for prettier data

# Outcome variable
df$y <- factor(df$y, levels = c(0,1),
               labels = c("No Spread", "Lymph Node Involvment"))
attr(df$y, "label") <- "Nodal Spread"

# Age
attr(df$age, "label") <- "Age (years)"

# Acid phosphatase
attr(df$acid, "label") <- "Acid Phosphatase (King-Armstrong units)"

# Xray
df$xray <- factor(df$xray, levels = c(0,1),
                  labels = c("Normal x-Ray", "Abnormal X-Ray"))
attr(df$xray, "label") <- "X-Ray Findings"

# Size
df$size <- factor(df$size, levels = c(0,1),
                  labels = c("Small", "Larger"))
attr(df$size, "label") <- "Tumor Size"

# Grade
df$grade <- factor(df$grade, levels = c(0,1),
                   labels = c("Less Serious Pathology", "Serious Pathology"))
attr(df$grade, "label") <- "Pathological Grade Finding"

# SES
df$ses <- ordered(df$ses, levels = c("low", "middle", "high"),
                  labels = c("Low SES", "Middle SES", "High SES"))
attr(df$ses, "label") <- "Socioeconomic Class (SES)"

# Correct shape of table
compareGroups(y ~ age + acid + xray + size + grade + ses, data = df) %>%
    createTable(., show.p.overall = TRUE) %>%
    export2md(., size = 8,
              caption = "Difference in Characteristics Between Prostate Cancer with and Without Nodal
```

Table 1: Difference in Characteristics Between Prostate Cancer with and Without Nodal Involvement

	No Spread N=56	Lymph Node Involvment N=36	p.overall
Age (years)	60.0 (5.54)	58.2 (6.80)	0.201
Acid Phosphatase (King-Armstrong units)	0.66 (0.29)	0.77 (0.22)	0.046
X-Ray Findings:			<0.001
Normal x-Ray	51 (91.1%)	17 (47.2%)	
Abnormal X-Ray	5 (8.93%)	19 (52.8%)	
Tumor Size:			0.005
Small	32 (57.1%)	9 (25.0%)	
Larger	24 (42.9%)	27 (75.0%)	
Pathological Grade Finding:			0.045
Less Serious Pathology	38 (67.9%)	16 (44.4%)	
Serious Pathology	18 (32.1%)	20 (55.6%)	
Socioeconomic Class (SES):			0.005
Low SES	23 (41.1%)	19 (52.8%)	
Middle SES	8 (14.3%)	12 (33.3%)	
High SES	25 (44.6%)	5 (13.9%)	