STA 440 Case 2

Annie Lott, Wuming Zhang, William Yang

October 1, 2017

Goals and Data Description

Getting treatment quickly after a stroke is crucial to a positive long-term prognosis. For emergency room patients who may have had a stroke, we are exploring the variables that influence the time it takes for these patients to receive a neurological assessment. The time from getting to the emergency room to receiving a neurological assessment, such as a CT scan, factors into the total time it takes for the patient to get treated for a stroke. This has a direct impact on the stroke patient's subsequent neurological health and survival.

Based on data from 335 emergency room patients with mild to moderate motor impairment, possibly indicative of a stroke, we are analyzing whether sex, race, ethnicity, and the number of symptoms displayed affects the time until the patients receive a neurological assessment, or whether or not they receive a neurological assessment at all. Sex is given as a binary variable of whether the patient is male (1) or female (0), race is a binary variable of whether the patient is black (1) or not (0), ethnicity is a binary variable of whether the patient is hispanic (1) or not (0), and the number of symtoms ranges from zero to four, with binary variables for if the patient exhibited one, two, three, or four symptoms (1 for each number of symptoms if they did exhibit this number, 0 for each number if they didn't). The four possible symptoms include a headache, loss of motor skills or weakness, trouble talking or understanding, and vision problems. Thus, given these variables, our goal is to build a model that predicts the amount of time until neurological assessment for potential stroke patients based on clinical presentation (how many symptoms the patient seems to have), gender and race and ethnicity, and to perform inference on the impact of these variables based on the model.

For ease of analysis, we cleaned the data by introducing a numerical variable for the number of symptoms for each patient- zero, one, two, three, or four- rather than using binary variables for whether a particular number of symptoms were displayed or not. There were no missing data to recode for our analysis.

```
data <- read.table("kellydat.txt", header = TRUE)



ggplot(data = data) + geom_histogram(aes(x = nctdel)) +
  facet_wrap(~male)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
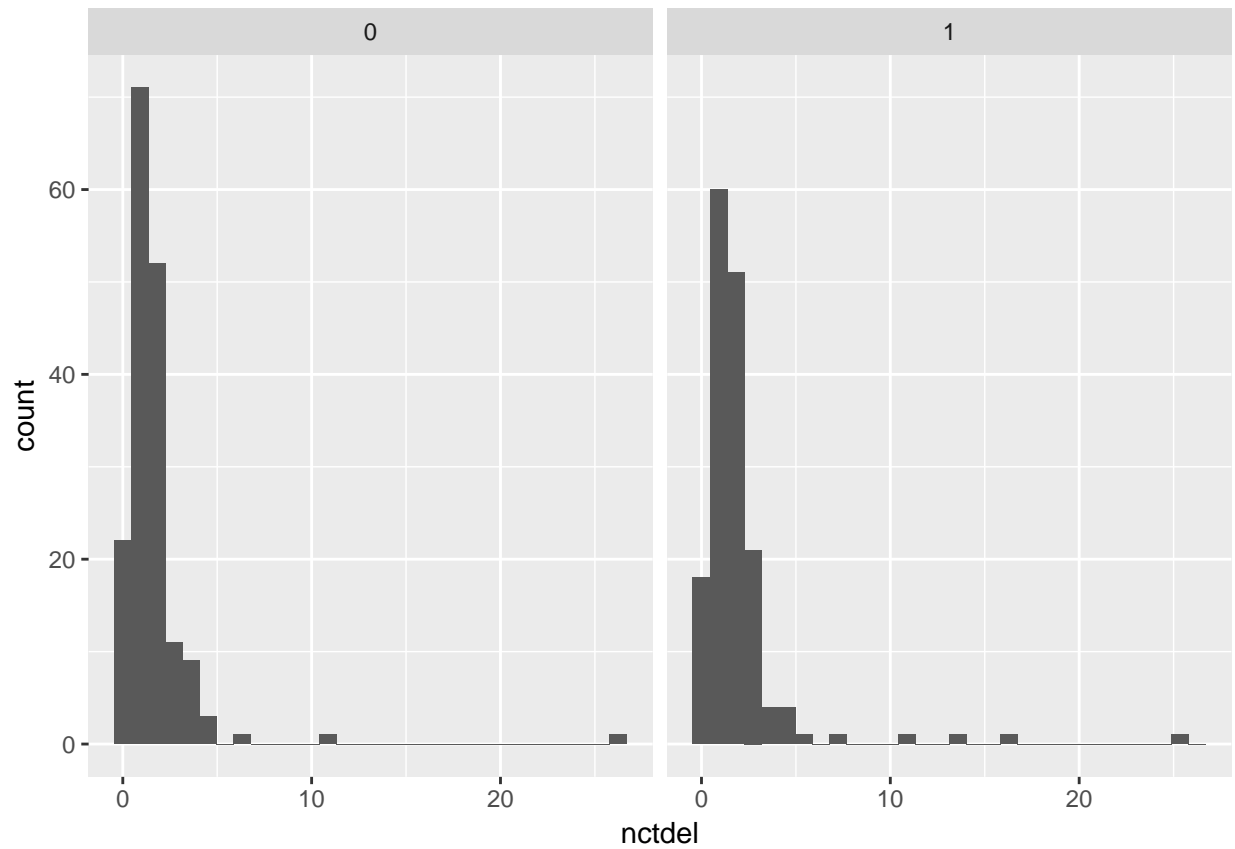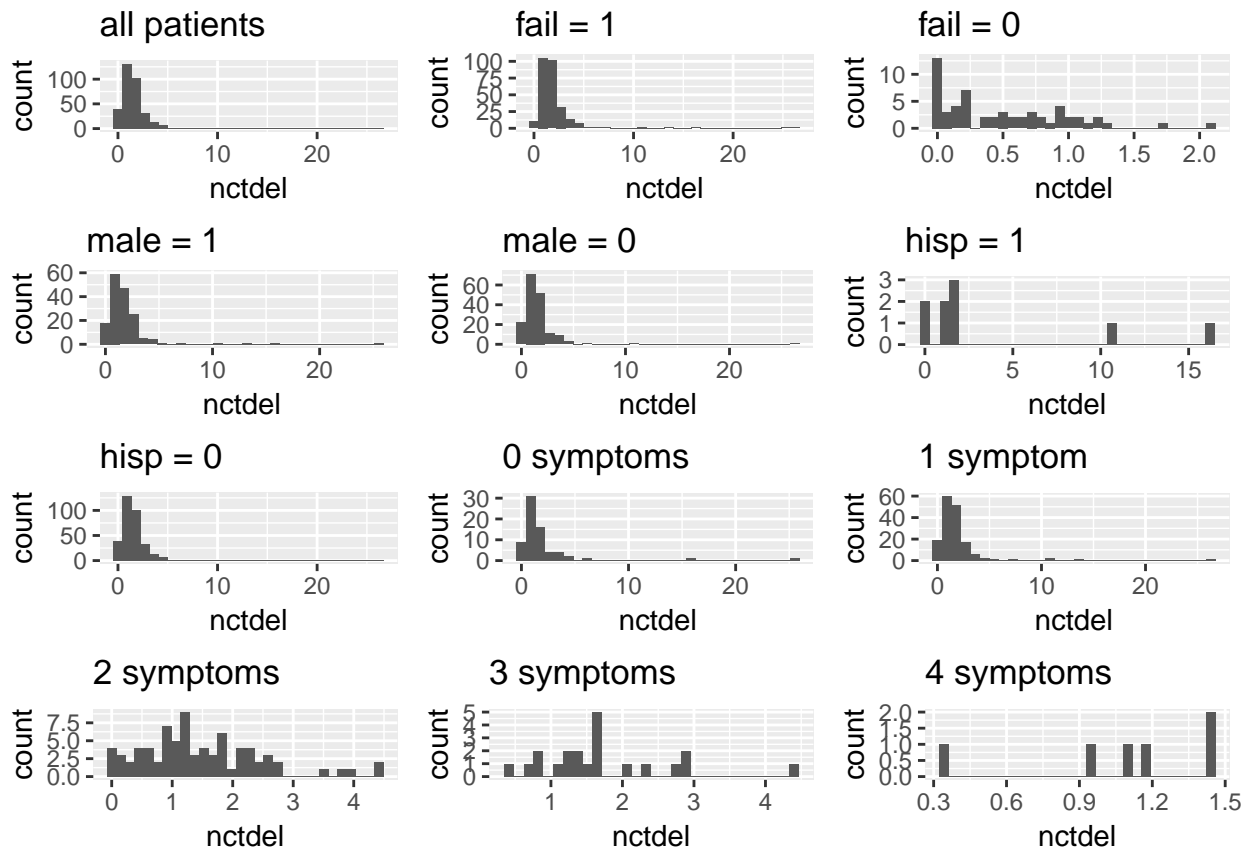
```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
kellydat = read.table("kellydat.txt", header = TRUE)
kellydat$sn0 = ifelse(kellydat$sn1 + kellydat$sn2 + kellydat$sn3 + kellydat$all4, 0, 1)
```

Exploratory Data Analysis

Histograms of each Indicator

To get a better sense of the distributions in the data, we first plot histograms of the nctdel times for each level of all the variables. We notice some general trends here - nctdel time seems to go down with more symptoms, the overall distribution of times for all patients is right-skewed with a mean around 1.5. However, we note that there are only 9 data points for hispanic patients, and only 6 for patients with all 4 symptoms. Inference around these variables will therefore have high variance.