

FinalStatsProject,Option 4: Stent data

Amber Vasquez, kav835

Background: This experiment studies the effectiveness of stents in treating patients at risk of stroke. The main question is “Does the use of stents reduce the risk of stroke?” 451 at-risk patients were in the experiment and assigned to a treatment group or control group randomly. It turned out the treatment group had more strokes than the control group, which is the opposite of what doctors expected. The second statistical question is whether the data shows a real difference between the groups or were the results just by chance? "The larger the difference we observe, the less believable it is that the difference is due to chance. So what we are really asking is the following: is the difference so large that we should reject the notion that it was due to chance?"

Read in data

```
data<-read.csv("stent30.csv")
```

Install dplyr package

```
#install.packages("dplyr")
library(dplyr)
```

```
##
## 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
```

Separate data

```
treatment <-filter(data,group == "treatment")
control <-filter(data,group == "control")
```

Count strokes in each group

```
treatment_stroke<-filter(treatment,outcome == "stroke")
control_stroke<-filter(control,outcome == "stroke")
number_treatment_stroke<-length(treatment_stroke$outcome)
number_control_stroke<-length(control_stroke$outcome)
```

Count no events in each group

```
treatment_noevent<-filter(treatment, outcome == "no event")
control_noevent<-filter(control, outcome == "no event")
number_treatment_noevent<-length(treatment_noevent$outcome)
number_control_noevent<-length(control_noevent$outcome)
number_treatment_noevent
```

```
## [1] 191
```

```
n_treatment <- (number_treatment_noevent+number_treatment_stroke)
n_control <- (number_control_noevent+number_control_stroke)
n_treatment
```

```
## [1] 224
```

```
n_control
```

```
## [1] 227
```

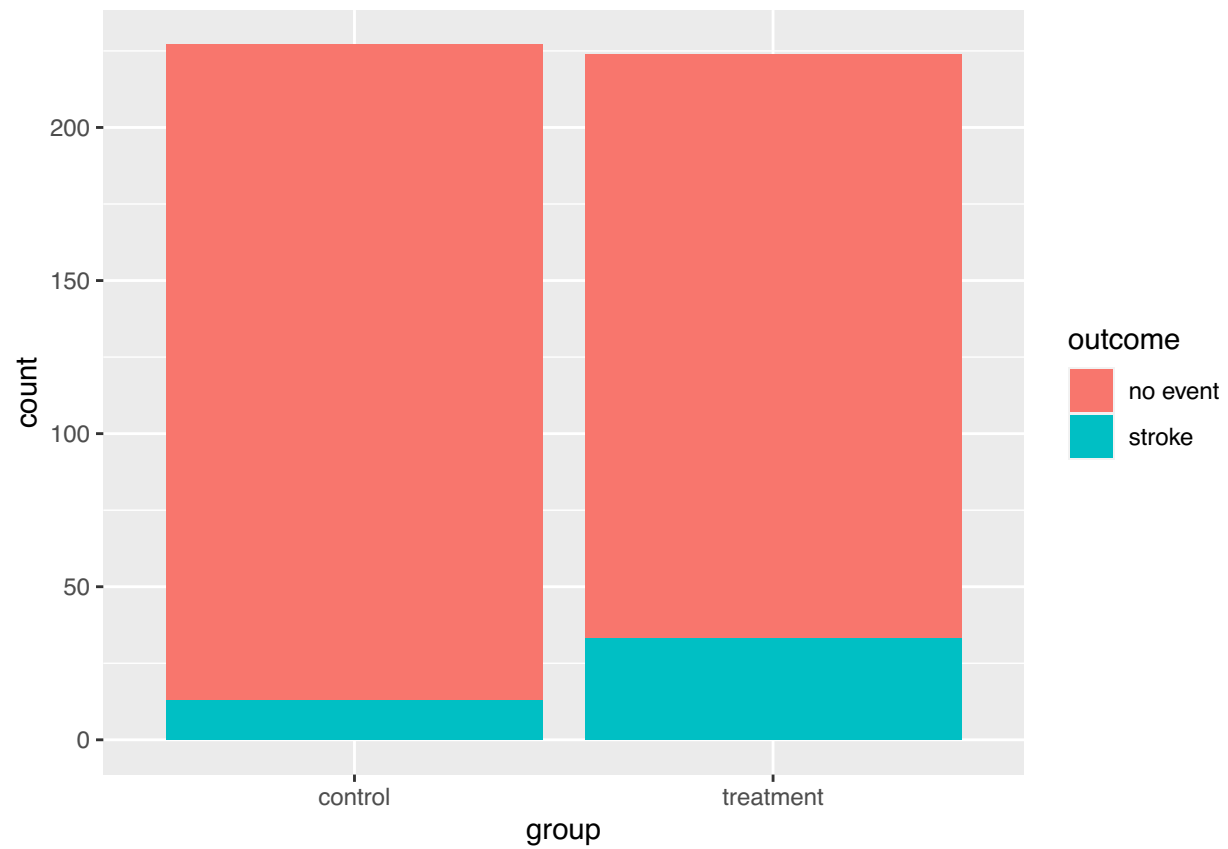
Call ggplot library to prepare environment to create a stacked bar chart

```
library(ggplot2)
```

Stacked Barplot to visualize data

```
ggplot(data=data) +
  geom_bar(mapping = aes(x = group, fill=outcome, position = "fill"))
```

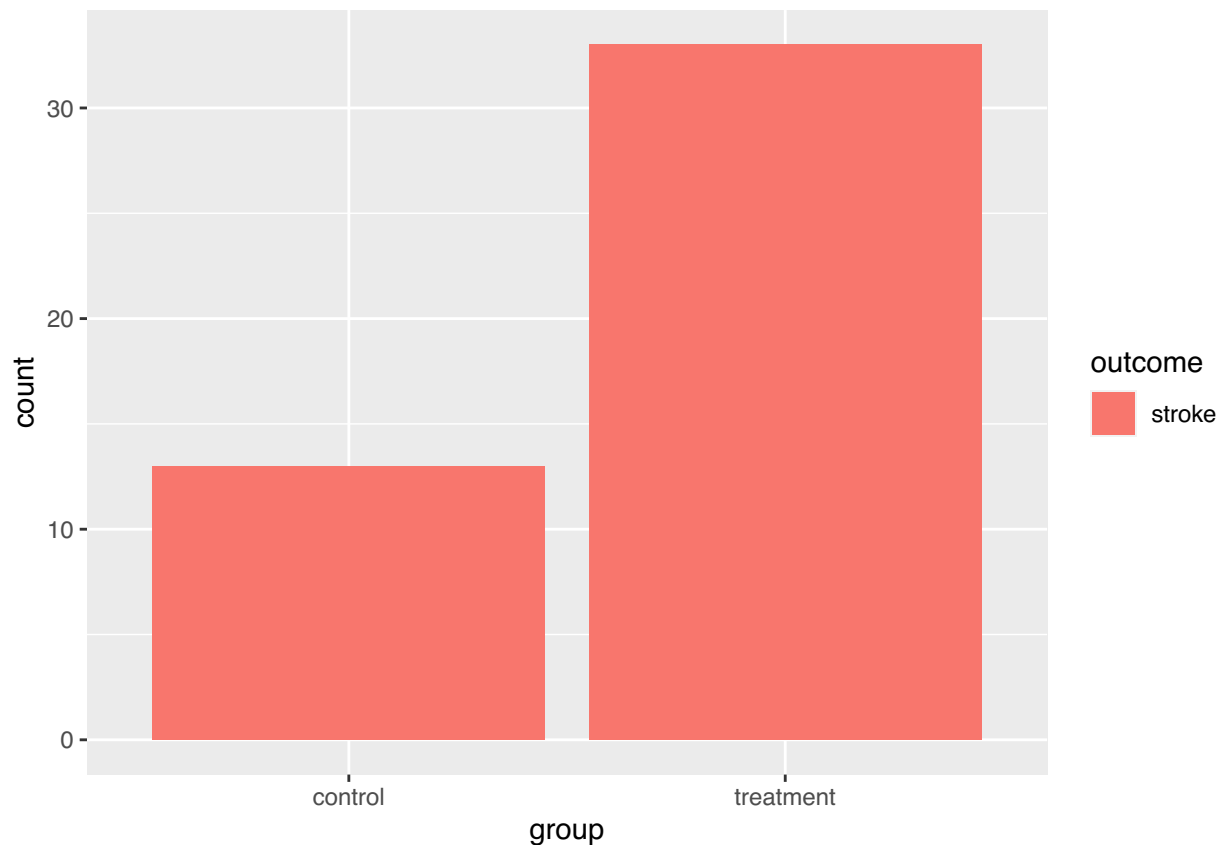
```
## Warning: Ignoring unknown aesthetics: position
```



```
data2 <- filter(data,outcome == "stroke")
```

Another view of how many strokes in each group

```
ggplot(data=data2) + geom_bar(mapping = aes(x=group,fill=outcome))
```



Conduct a 2-sample proportions test

```
p_treatment <- number_treatment_stroke/n_treatment
p_control <- number_control_stroke/n_control
prop.test(x=c(number_treatment_stroke, number_control_stroke), n = c(n_treatment, n_control),
          ,alternative = "two.sided")
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: c(number_treatment_stroke, number_control_stroke) out of c(n_treatment, n_control)
## X-squared = 9.0233, df = 1, p-value = 0.002666
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.03022922 0.14987619
## sample estimates:
##      prop 1      prop 2
## 0.14732143 0.05726872
```

Table of results

```
table<-matrix(c("", "C", "T", "S", number_control_stroke, number_treatment_stroke, "NoEvent", number_control_n, number_treatment_n), nrow=4, ncol=8, byrow=TRUE)
```

	V1	V2	V3
1		S	NoEvent
2	C	13	214
3	T	33	191

Our confidence Interval at 95% (0.03022,0.1498) Here, the p-value of 0.002666 is small we reject the null. It is significant that the two groups are different.

The treatment group has significantly more strokes than the control group. It is unlikely this is due to chance, stents are making patients worse and at higher risk of stroke. Somehow the treatment is working negatively, though it was originally expected it would help patients. According to the statistical analysis, the groups do have a significant difference, not due to natural variability.