

Clumping Hypothesis

Beau Meche & Robert McKenzie

2/12/2020

```
# ignored messages because geom_histogram was complaining about
# the nature of our discrete data

num_men <- rep(0,14)
num_women <- rep(1,16)
class_1 <- seq(1, 7)
class_2 <- seq(7,8)

x = c(num_men,num_women)

reps = 10000

compositions <- vector("numeric", reps)

for (q in 1:reps) {

compositions[q] <- data.frame(gender = sample(x, 30),
                             table = c(rep(class_1, 4), 8,8)) %>%
  group_by(table) %>%
  summarize(num_women = sum(gender)) %>%
  mutate(clumps = ifelse(num_women %in% c(0,4), yes = 1, no = 0)) %>%
  summarise(prop_clumps = mean(clumps))
}

compositions <- data.frame(t(as.data.frame(compositions))*8, seq_along(t(as.data.frame
(compositions))))

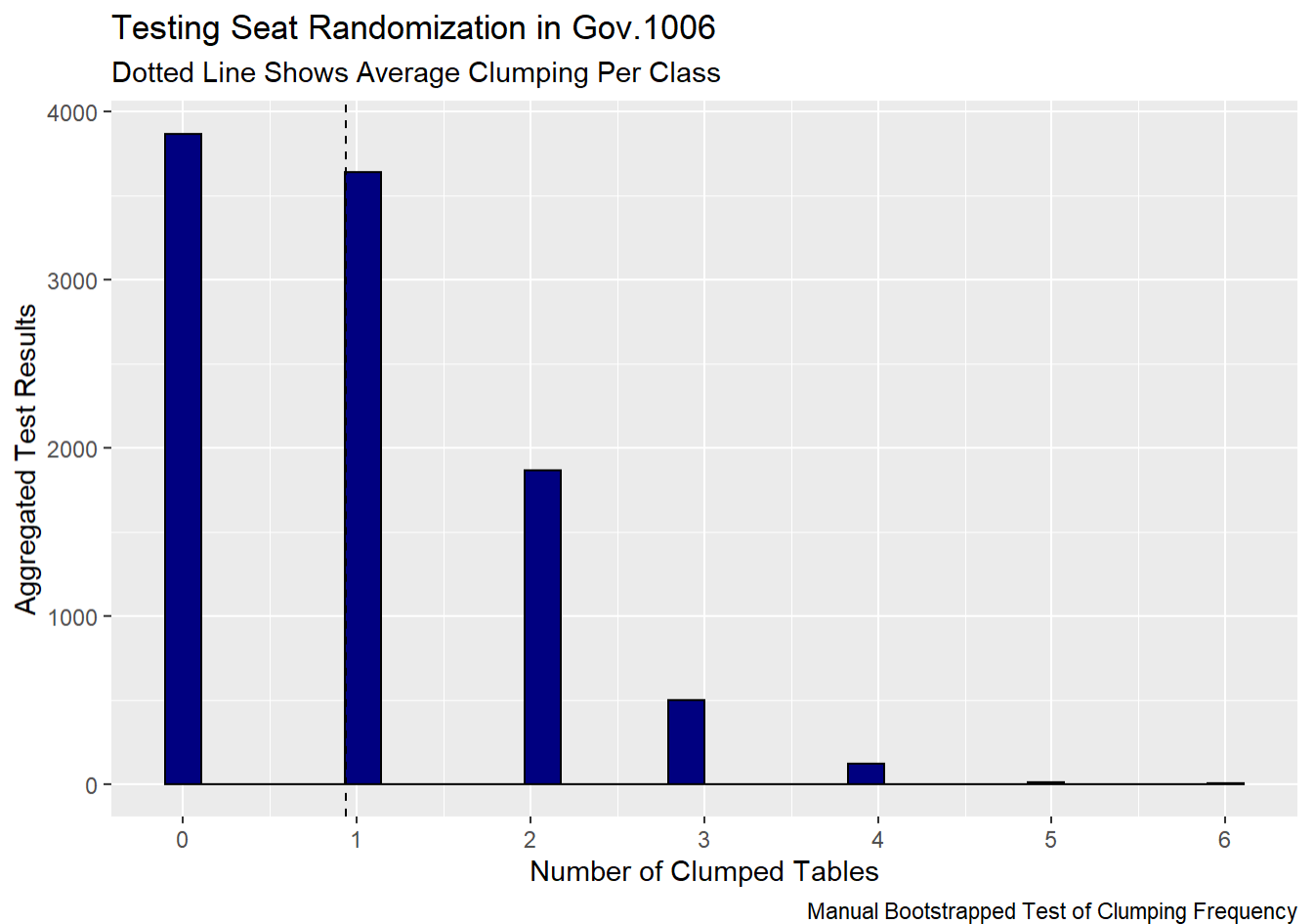
# plot here

mn <- mean(compositions$t.as.data.frame.compositions.....8)
```

```

compositions %>%
  ggplot(aes(x = t.as.data.frame(compositions)....8)) +
  geom_histogram(color = "black", fill = "navy") +
  geom_vline(xintercept = mn, linetype = 2) +
  labs(title = "Testing Seat Randomization in Gov.1006",
       subtitle = "Dotted Line Shows Average Clumping Per Class",
       x = "Number of Clumped Tables",
       y = "Aggregated Test Results",
       caption = "Manual Bootstrapped Test of Clumping Frequency") +
  scale_x_continuous(breaks = seq(0,6, by = 1))

```



We would expect a class of 14 men and 16 women to clump into about 0.9408 single gender tables per class meeting, given 4 seats per table with the exception of the last table, which has 2 seats.