

Lab 6B: Testing for Goodness of Fit

2009 Iran Election

On June 12 2009, the Republic of Iran [held an election](#) where President Mahmoud Ahmadinejad sought re-election against three challengers. When it was announced that Ahmadinejad had won with 62% of the vote, there were widespread allegations of election fraud.

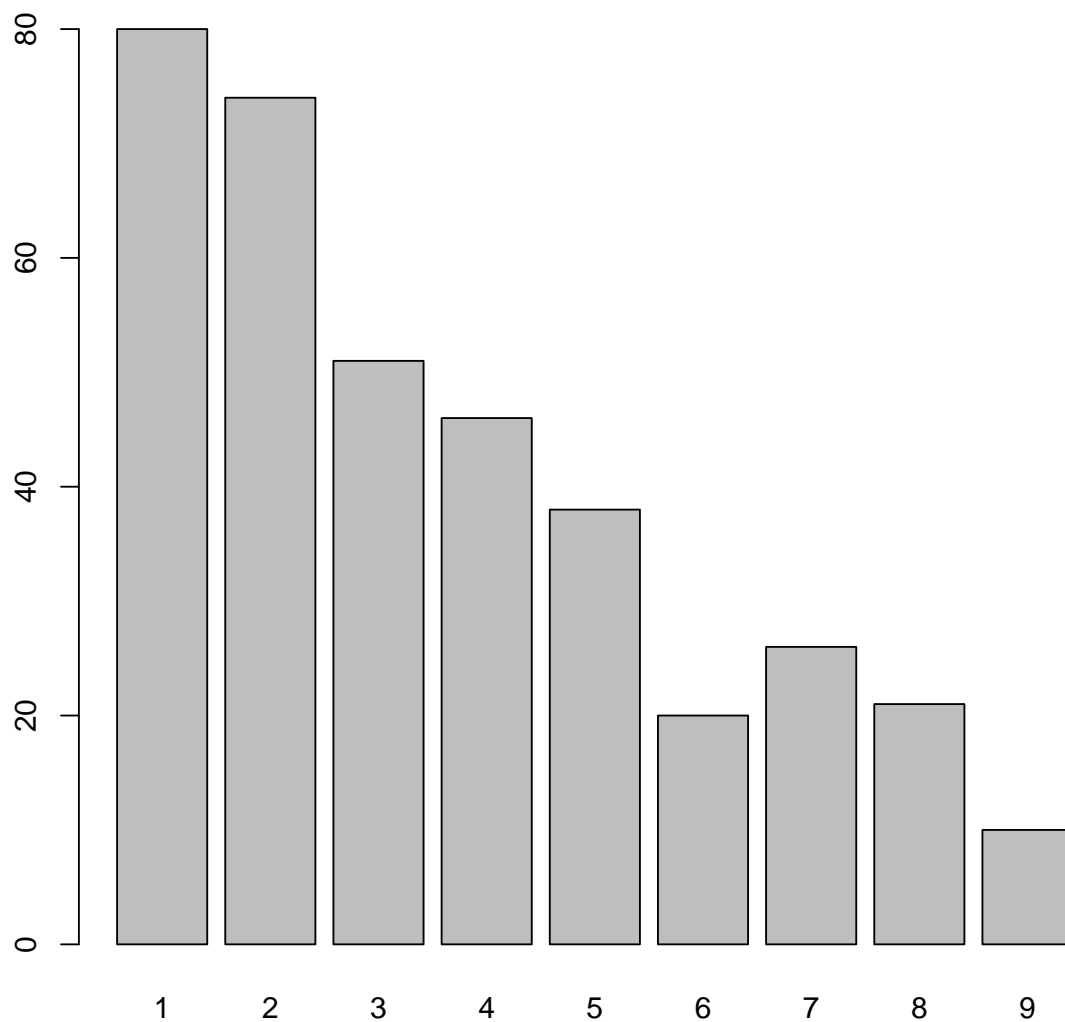
There are many methods, both quantitative and qualitative, to detect election fraud. In this lab we will explore just one proposed method.

The Data

The election commission released total vote counts for each candidate by region. Let's load up this data.

```
get_digit <- function(x, which_digit = "first") {  
  if (is.na(match(which_digit, c("first", "last")))) {  
    stop("which_digit must be first or last")  
  }  
  if (which_digit == "first") {  
    as.integer(head(strsplit(as.character(x), "")[[1]], n = 1))  
  } else {  
    as.integer(tail(strsplit(as.character(x), "")[[1]], n = 1))  
  }  
}  
  
# read in iran data  
iran <- read.csv("iran.csv", header = TRUE)  
digits_iran <- sapply(iran$total_votes_cast, FUN = get_digit)  
obs <- table(digits_iran)  
barplot(obs)
```

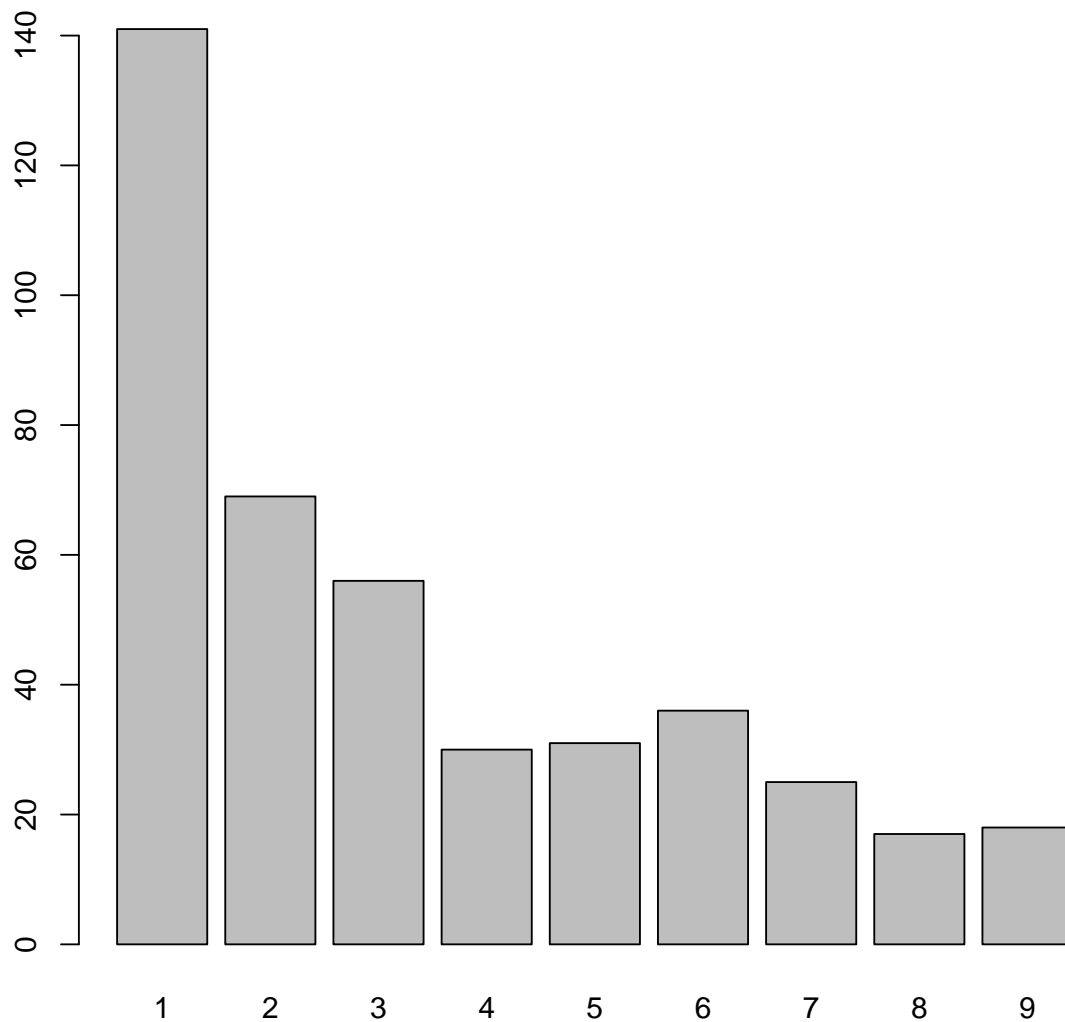
This is a product of OpenIntro that is released under a Creative Commons Attribution-ShareAlike 3.0 Unported (<http://creativecommons.org/licenses/by-sa/3.0>). This lab was written by Andrew Bray and Mine Çetinkaya-Rundel.



```
# compare to Benfords dist
n <- dim(iran)[1]
benfords_p <- log10(1 + 1/1:9)
expected <- benfords_p * n
chistat <- sum(((obs - expected)^2)/expected)
pchisq(chistat, df = 8, lower.tail = FALSE)

## [1] 0.006836

# read in norway data
norway <- read.csv("norway.csv", header = TRUE)
digits_norway <- sapply(norway$total_votes_cast, FUN = get_digit)
obs <- table(digits_norway)
barplot(obs)
```



```
# compare to Benfords dist
n <- dim(norway)[1]
expected <- benfords_p * n
chistat <- sum(((obs - expected)^2)/expected)
pchisq(chistat, df = 8, lower.tail = FALSE)
## [1] 0.3979
```

David: do these calculations look right? I haven't worked with this data for awhile, but for some reason I remember that the iran data didn't mesh with benfords, suggesting fraud but then the norway data didn't either, suggesting . . . that the assumptions of benfords don't really hold in vote total settings. Running, it here, though, the Norwegian data looks fine by Benfords. I guess I could try getting county level data from the US and trying there...