Andrew Goldstein
March 2012

Horse Racing Violations

```
x <- read.csv("http://www.stat.yale.edu/~jay/230/Homework/230horses.csv",as.is = TRUE)
x$max <- 0  # initialize new column for the outermost position (the number of horses running in each race)
x$outer <- FALSE  # initialize new column for whether a horse is in the two outermost positions


###################### FILL IN (x$max) ########################################

max <- x[1,2]                    #set 'max' variable to the first value
n <- 1                           #set 'counter' variable to 1

for (b in 2:2950){
  if (x[b,1] == x[b-1,1]){       #determine whether row of dataset is within the same race as previous row
    if (as.numeric(x[b,2]) > max) {max <- as.numeric(x[b,2])}   #determine highest draw position in a race
    n = n+1
  } else {
    x$max[(b-n):(b-1)] <- max    #fill in x$max for previous horse race with the value of 'max'
    max <- as.numeric(x[b,2])    #reset 'max' variable to first value of the next race
    n <- 1                       #reset 'counter' variable
  }
}
x$max[(b-n+1):(b)] <- max        #fill in x$max for final horse race (of data set) with the value of 'max'


#################### FILL IN (x$outer) ########################################

for (b in 1:2950){
  if (x[b,2]    == x[b,5]) {x[b,6] <- TRUE}   #fill in x$outer as 'TRUE' for outermost position
  if (x[b,2] +1 == x[b,5]) {x[b,6] <- TRUE}   #fill in x$outer as 'TRUE' for second outermost position
}


> table(x$outer)

FALSE  TRUE
 2394   556

> table(x$runner.f, x$outer)
```

```
            FALSE TRUE
  Non-Runner   227  109
  Runner      2167  447

############## Fisher's Exact Test for Count Data ###############

> fisher.test(table(x$runner.f, x$outer))

data:  table(x$runner.f, x$outer)
p-value = 1.811e-10
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.3326139 0.5574114
sample estimates:
odds ratio
 0.4297322

################## Pearson's Chi-squared test #################

> chisq.test(table(x$runner.f, x$outer), correct = FALSE)

data:  table(x$runner.f, x$outer)
X-squared = 45.8071, df = 1, p-value = 1.305e-11

################### PERMUTATION TEST ########################
> horsesnotrun <- rep(0, 10000)
> for (i in 1:10000) {
+ x$SimOutcome <- rep("FALSE", nrow(x))
+ notrunthese <- sample(1:2950, 336, replace = FALSE)
+ x$SimOutcome[notrunthese] <- "TRUE"
+ horsesnotrun[i] <- sum(x$outer == "TRUE" & x$SimOutcome == "TRUE")
+ }
> max(horsesnotrun)
[1] 90
```

The extremely low p-values for the Fisher Exact Test and the Chi-squared test provide evidence to support the conclusion that horses are more likely to drop out of a race if they draw one of the outer track positions. Additionally, the permutation test above is run 10,000 times and not a single natural permutation yields more than 90 horses dropping out from the outer tracks whereas the true number is 109 (bolded in the table above). These tests imply that while horse racing rules stipulate that owners are not supposed to drop their horses from a race based solely on track position draw – only due to injury, it appears that some owners are breaking these rules.