HOMEWORK 2

1. Reporting a binomial test

When sampling datives constructions from a large corpus of American English spontaneous phone conversations, Bresnan et al. (2007) found 501 prepositional datives and 1,859 double objects. This was significant at α = .05 (binom.test, two-tailed p-value < 2.2e-16, CI: .196, .229), which means we can reject the null hypothesis that the two constructions are equiprobable in American English.

```
x <- c(501,1859)
binom.test(x, p=0.5)

##
##  Exact binomial test
##
## data:  x
## number of successes = 501, number of trials = 2360, p-value < 2.2e-16
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.1959431 0.2293504
## sample estimates:
## probability of success
##              0.2122881
```

2. McNemar's test

```
data <-
read.table(url("http://wellformedness.com/courses/LING82100/Data/PTB.tsv"),
header = TRUE, comment.char="#")

Stanford.correct <- data$gold.tag == data$Stanford.tag
NLP4J.correct <- data$gold.tag == data$NLP4J.tag
```

Number of "wins" for the Stanford tagger:

```
x1 <- sum(Stanford.correct & !NLP4J.correct)
x1

## [1] 943
```

Number of "wins" for the NLP4J tagger:

```
x2 <- sum(NLP4J.correct & !Stanford.correct)
x2

## [1] 1016
```

```r
x <- min(x1,x2)
n <- x1 + x2
binom.test(x, n, 0.5)

##
##  Exact binomial test
##
## data:  x and n
## number of successes = 943, number of trials = 1959, p-value = 0.1038
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.459029 0.503763
## sample estimates:
## probability of success
##               0.481368
```

The test is non-significant at $\alpha = .05$ (two-tail p-value = .104, CI: .459, .503), so it is not possible to reject the null hypothesis of p=0.5 and say that one tagger is significantly better than the other.