Statistics for Linguistic Research
MP 2
Yohamy Polanco

Part One

This test looks at the actual amount of times a specific event occurs in a given amount of trials. The prepositional dative was recorded to have occurred 501 times out of 2360 trials. The test statistic is 50%.

```
n <- 501 + 1859
x <- 501
p <- 0.5
binom.test(x, n, p)

	#summary p-value = 2.2e-16, it is smaller than 0.05 so this rejects the null hypothesis
	#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
```

The p-value is smaller than 0.05 so this rejects the null hypothesis.

Part Two

```
df1 <- read.table(file='PTB.tsv', sep='\t', header = TRUE, comment.char = "@" )

stan <- df1$Stanford.tag
nlp <- df1$NLP4J.tag
gold <- df1$gold.tag
nlp.correct <- gold == nlp
stan.correct <- gold == stan

stan_wins <- sum(stan.correct & !nlp.correct)
nlp_wins <- sum(nlp.correct & !stan.correct)
```

```
x <- min(stan_wins, nlp_wins)
n <- sum(stan_wins, nlp_wins)
p <- 0.5
binom.test(x, n, p)

 #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
```

Since the p-value is greater than the significance level $\alpha=0.05$, we cannot reject the null hypothesis.