

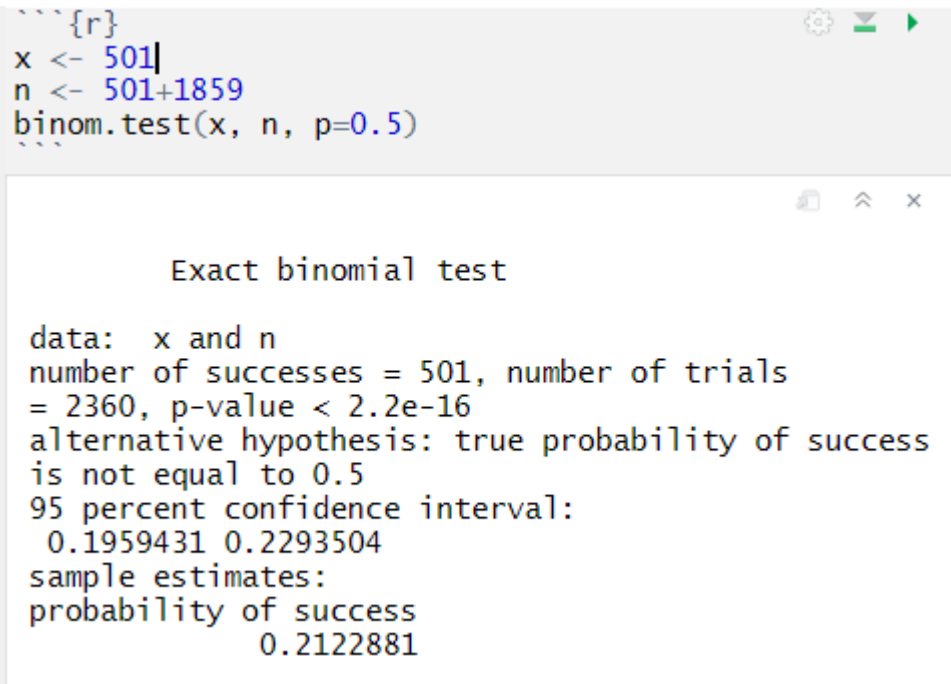
HW #2

Bora Seo

3/4/20

Q1.

The null hypothesis (h_0) is that the prepositional datives and the double objects can be found equiprobably. To test h_0 , we put the number of prepositional datives (501) as x , the sum of both constructions (501+1859) as n , and 0.5 from h_0 as p . When we run `binom.test` in R, we can see the result as below:



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

Exact binomial test

data: x and n
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 95% confidence interval here is from 0.196 to 0.223 which includes the true probability of success. The p-value is less than 2.2×10^{-16} which entails that it is undoubtedly less than α (0.05) at 95% CI, thus the test is statistically significant. We reject the h_0 , which means that the given constructions do not occur equiprobably.

Q2.

The R expressions are as below:

```
ptb_df <- read.table(file='/Users/velos/Documents/r_project/PTB.tsv', header = TRUE)

standford.correct <- ptb_df$gold.tag == ptb_df$Stanford.tag
nlp4j.correct <- ptb_df$gold.tag == ptb_df$NLP4J.tag

#The number of "wins" for the Stanford tagger over NLP4J tagger
stand_win <- sum(standford.correct & !nlp4j.correct)
stand_win

#The number of "wins" for the NLP4J tagger over Stanford tagger
nlp4j_win <- sum(nlp4j.correct & !standford.correct)
nlp4j_win

x <- min(stand_win, nlp4j_win)
n <- stand_win+nlp4j_win

#Binomial test
binom.test(x, n, p=0.5)
```

```
[1] 943
[1] 1016
```

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 was insignificant at α (0.05) at 95% CI since the p-value (0.104) is greater than α . Thus, we do not reject the null hypothesis. However, it is difficult to say one tagger is significantly better than other at this point. In order to determine this, we need to conduct more tests.