

**Homework 2***Binomial test*

```
```{r}
binom.test(501, 2360, .5)
```
```

Exact binomial test

```
data: 501 and 2360
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
```

To test the null hypothesis that the two potential outcomes of a trial (prepositional dative vs. double object) are equiprobable, this experiment included 2360 total trials with 501 successes. The p-value of the test is  $2.2 \times 10^{-16}$ . The 95% confidence interval was .1959431—.2293504. This is a statistically significant result. We therefore reject the null hypothesis.

*McNemar's Test*

```
```{r}
PTB <- read.table("/tmp/mozilla_meredith0/hw02-MeredithHilliard-master/PTB.tsv", header = TRUE)
PTB

Stanford.correct <- PTB$gold.tag == PTB$Stanford.tag
NLP4J.correct <- PTB$gold.tag == PTB$NLP4J.tag
x1 <- sum(NLP4J.correct & !Stanford.correct)
x1
x2 <- sum(Stanford.correct & !NLP4J.correct)
x2
```

[1] 1016
[1] 943
```

The number of times that NLP4J was correct when Stanford was wrong is 1016; the number of times the Stanford was correct when NLP4J was wrong is 943.

Exact binomial test

```
data: min(x1, x2) and x1 + x2
```

```
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 p-value is larger than .05, and therefore we cannot reject the null hypothesis. Neither of the taggers is significantly better than the other.