Task 1

There are two possible outcomes: prepositional dative (a) or double object (b).
With the null-hypothesis not rejected, both would be equally likely.
With 501 observances of outcome (a) and 1859 observances of outcome (b):
n total number of trials:
501 + 1859 = 2360
x (number of successes) is outcome (a)
p (probability of getting a success on trial) is 0.5
Probability density function (probability of getting exactly 501 desired outcomes out of a total number of trials):

$$\binom{2360}{501} \times 0.5^{2360-501} \times 0.5^{501}$$

= 5.434782092197675322174266855388319171820382999439890... × 10^-183 (very small)
Is this result significantly far enough from the mean?

- **Description**

This experiment tests the use of prepositional dative (a) over the double object (b) structure in modern English. The two constructions are assumed equiprobable. From a large corpus of American English spontaneous phone conversations 501 occurrences of (a) and 1859 occurrences of (b) were extracted.

- **Test statistics (p) (**Probability to reject the null hypothesis)

```
> 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
```

- **95% confidence intervals**

CI: .195, .229

- **Whether the test was significant at** α **= .05** (threshold for rejecting)

This was significant at α = .05, null-hypothesis rejected

Task 2

Mc Nemar's test

- Stanford Tagger over NLP4J

943

- NLP4J over Stanford

1016

- The McNemar test results; is one tagger significantly better than the other at α = .05

This was non-significant at α = .05 (Mc Nemar's test, two-tailed p = .103, 95% CI: .459, .503). Neither tagger is significantly better than the other.

R Commands

```
> data <- read.table("/Users/biatris/Desktop/Stats/HW2/PTB.tsv", header = TRUE, sep = "\t", comment.char = "")
> head(data)

  gold.tag TnT.tag Collins.tag Stanford.tag LAPOS.tag NLP4J.tag
1       JJ      JJ          JJ          JJ        JJ        JJ
2      NNS     NNS         NNS         NNS       NNS       NNS
3       IN      IN          IN          IN        IN        IN
4       DT      DT          DT          DT        DT        DT
5      NNP     NNP         NNP         NNP       NNP       NNP
6      NNP    NNPS         NNP         NNP      NNPS      NNPS

> Stanford.correct <- data$gold.tag == data$Stanford.tag
> NLP4J.correct <- data$gold.tag == data$NLP4J.tag
> x1 <- sum(Stanford.correct & !NLP4J.correct)
> x2 <- sum(NLP4J.correct & !Stanford.correct)
> x1
[1] 943
> x2
[1] 1016

> x <- min(x1, x2)
> n <- x1 + x2
> p <- .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