Yuying Ren
LING82100: Statistics in Linguistics
Homework 02
Mar/03

1. Reporting a binominal test

Report:

Description: The experiment is detecting a total number of 2,360 American English spontaneous phone conversations, to find whether people use the prepositional datives(A) or the double objects(B) in these conversations. The outcome is: 501 prepositional datives and 1,859 double objects.

Test statistic: according to the hypothesis that the two constructions are equiprobable, $p = 0.5$.

95% confidence intervals:

- The 95% confidence intervals of A: 0.1959431 0.2293504
- The 95% confidence intervals of B: 0.7706496 0.8040569

The test was significant at $\alpha = 0.5$, p-value < 2.2e-16.

R expressions:

```
> binom.test(501, 501+1859, 0.5)

        Exact binomial test

data:  501 and 501 + 1859
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

> binom.test(1859, 501+1859, 0.5)

        Exact binomial test

data:  1859 and 501 + 1859
number of successes = 1859, 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:
 0.7706496 0.8040569
sample estimates:
probability of success
        0.7877119
```

2. McNemar's test

	> d <- read.table("https://raw.githubusercontent.com/statistics-for-linguistic-research/hw02-yuyingren/master/PTB.tsv", comment.char = "#", header = TRUE)	#Load the TSV file to R
	> Staf.correct <- d$gold.tag == d$Stanford.tag > sum(Staf.correct) [1] 126203 > NLP4correct <- d$gold.tag == d$NLP4J.tag > sum(!NLP4.correct) [1] 2104	#Get the data for x1(the "wins" for Standford tagger over NLP4J tagger)
1.	> x1 <- sum(Staf.correct & !NLP4.correct) > x1 [1] 943	Answer for Q1: the number of "wins" for Standford tagger over NLP4J is **943.**
2.	> x2 <- sum(NLP4.correct & !Staf.correct) > x2 [1] 2275	Answer for Q2: the number of "wins" for NLP4J tagger over Standford tagger is **2275.**
3.	> binom.test(x1, x1+x2, 0.5) Exact binomial test data: x1 and x1 + x2 number of successes = 943, number of trials = 3218, p-value < 2.2e-16 alternative hypothesis: true probability of success is not equal to 0.5 95 percent confidence interval: 0.2773518 0.3091035 sample estimates: probability of success 0.2930392 > binom.test(x2, x1+x2, 0.5) Exact binomial test data: x2 and x1 + x2 number of successes = 2275, number of trials = 3218, p-value < 2.2e-16 alternative hypothesis: true probability of success is not equal to 0.5 95 percent confidence interval: 0.6908965 0.7226482 sample estimates: probability of success 0.7069608	Answer for Q3: According to the results, we don't see that one tagger is significantly better than the other at α = 0.05. Because the p-value for both taggers is significantly smaller than α, and the 95 percent confidence interval is about the same.