MAST20005/MAST90058: Week 5 Lab

Goals: (i) Compute and interpret confidence intervals for proportions; (ii) Study the accuracy of confidence intervals by simulation.

1 Confidence intervals for proportions

Let p be the proportion of runners who suffered a training-related overuse injury during last year. Out of 20 runners who responded to a small survey, 6 indicated that they had suffered such an injury during the past year. Give a point estimate and an approximate 90% confidence interval for p.

1.1 Wald approximation

One method consists of using the CLT approximation for $\hat{p}$ and also estimating the margin of error by substituting $\hat{p}$ in place of p, giving $\hat{p} \pm c\sqrt{\hat{p}(1 - \hat{p})/n}$. This is sometimes called the Wald approximation.

```
n <- 20
p.hat <- 6 / n
p.hat

## [1] 0.3

z0 <- qnorm(0.95)
p.hat + c(-1, 1) * z0 * sqrt(p.hat * (1 - p.hat) / n)

## [1] 0.1314527 0.4685473
```

1.2 Quadratic approximation

A second method also uses the CLT approximation but avoids approximating the margin of error. See question 8 from the tutorial problems for more details. We calculate this in R as follows.

```
num <- p.hat + z0^2 / (2 * n) +
       c(-1, 1) * z0 * sqrt(p.hat * (1 - p.hat) / n + z0^2 / (4 * n^2))
den <- 1 + z0^2 / n
num / den

## [1] 0.1638594 0.4838037
```

This is the default method implemented in the R function `prop.test`.

```
prop.test(x = 6, n = 20, conf.level = 0.90, correct = FALSE)

##
##  1-sample proportions test without continuity correction
##
```

```
## data:  6 out of 20, null probability 0.5
## X-squared = 3.2, df = 1, p-value = 0.07364
## alternative hypothesis: true p is not equal to 0.5
## 90 percent confidence interval:
##   0.1638594 0.4838037
## sample estimates:
##   p
## 0.3
```

1.3   Exact method

When the sample size is small we can directly compute the quantiles of the binomial distribution. However, the discrete nature of the binomial distribution requires some approximation for the binomial quantiles, which affects the actual confidence level of the test. The method proposed by Clopper and Pearson (1934) is as follows. Let x be the number of successes observed in the sample; compute $S_1 \cap S_2$ where

$$S_1 = \{p : \Pr(X \leqslant x) > \alpha/2\}$$

and

$$S_2 = \{p : \Pr(X \geqslant x) > \alpha/2\}$$

where the above probabilities refer to $X \sim \mathrm{Bi}(n, p)$. In R this is implemented by the function `binom.test`.

```
t <- binom.test(x = 6, n = 20, conf.level = 0.90)
t


##
##   Exact binomial test
##
## data:  6 and 20
## number of successes = 6, number of trials = 20, p-value = 0.1153
## alternative hypothesis: true probability of success is not equal to 0.5
## 90 percent confidence interval:
##   0.1395537 0.5078184
## sample estimates:
## probability of success
##                    0.3

names(t)    # shows names of various objects returned by binom.test()

## [1] "statistic"   "parameter"   "p.value"     "conf.int"    "estimate"
## [6] "null.value"  "alternative" "method"      "data.name"

t$conf.int  # extracts the confidence interval

## [1] 0.1395537 0.5078184
## attr(,"conf.level")
## [1] 0.9
```

Although this interval guarantees a confidence level of at least $1 - \alpha$ for the population proportion, p, it is usually too conservative (i.e. can be too large).

2

2 Comparing confidence intervals

From theory we know that the first two methods in Section 1 will work if n is large. But what is their performance when n is moderate or small? In this case the true confidence level might differ from the desired probability of $1 - \alpha$. Next we investigate the performance of the procedures introduced in Section 1 by simulationg samples from a $Bi(n, p)$ distribution. The goal is to estimate the coverage probability for each method.

```r
p <- 0.3            # set the true proportion
n <- 15             # set number of trials
z0 <- qnorm(0.95)   # quantile for 90% confidence level
B <- 2500           # number of simulated samples


result1 <- result2 <- result3 <- 1:B  # empty vectors, to store results
for (b in 1:B) {
    y <- rbinom(1, size = n, prob = p) # generate a binomial outcome
    p.hat <- y / n  # sample proportion

    # Wald interval.
    int1 <- p.hat + c(-1, 1) * z0 * sqrt(p.hat * (1 - p.hat) / n)

    # Interval based on quadratic equation.
    num <- p.hat + z0^2 / (2 * n) +
            c(-1, 1) * z0 * sqrt(p.hat * (1 - p.hat) / n + z0^2 / (4 * n^2))
    den <- 1 + z0^2 / n
    int2 <- num / den

    # Clopper-Pearson "exact" interval.
    t <- binom.test(y, n, conf.level = 0.90)
    int3 <- t$conf.int

    # Check if p is in the intervals and store results.
    result1[b] <- (int1[1] < p) & (p < int1[2])
    result2[b] <- (int2[1] < p) & (p < int2[2])
    result3[b] <- (int3[1] < p) & (p < int3[2])
}
```

The last few lines of the above code gives two logical vectors with elements equal to `TRUE` or `FALSE` depending on whether the bth interval is found to contain the true proportion p. The coverage probabilities are found as follows:

```r
mean(result1)  # coverage Wald CI

## [1] 0.8332

mean(result2)  # coverage quadratic CI

## [1] 0.9136

mean(result3)  # coverage Clopper-Pearson CI

## [1] 0.9408
```

3

Note that mathematical operators applied to logical vectors treat `TRUE` or `FALSE` as 0 and 1, respectively. Thus, the last three commands compute simulated estimates for the coverage probability. Note that the first method gives a much smaller coverage than the nominal coverage of $1 - \alpha = 0.9$. Therefore, the method based on the quadratic equation should be preferred in practice.

Try varying the settings of your simulations and explore what happens to the coverage probability for different values of n and p. Particularly, what happens when the sample size is small and p gets closer to 0?

Much has been written about this problem. A general advice is to never use the Wald normal approximation when n is moderate or small, as it has terrible coverage properties. After installing the package `binom` (e.g. type `install.packages("binom")`), run the following illustrative R code to see the coverage probability for p ranging between 0 and 1. For example, let us compare the Wald interval with the Agresti-Coull correction. The Agresti-Coull correction uses the biased estimator $\tilde{p} = (Y + 2)/(n + 4)$ instead of the sample proportion $\hat{p} = Y/n$.

```r
library(binom)
p <- seq(0, 1, 0.001)
coverage.wald    <- binom.coverage(p, 25, method = "asymptotic")$coverage
coverage.agresti <- binom.coverage(p, 25, method = "agresti-coull")$coverage
plot( p, coverage.wald, type = "l", ylab = "coverage")
lines(p, coverage.agresti, col = "blue", lwd = 2)
abline(h = 0.95, col = "red") # desired coverage
```
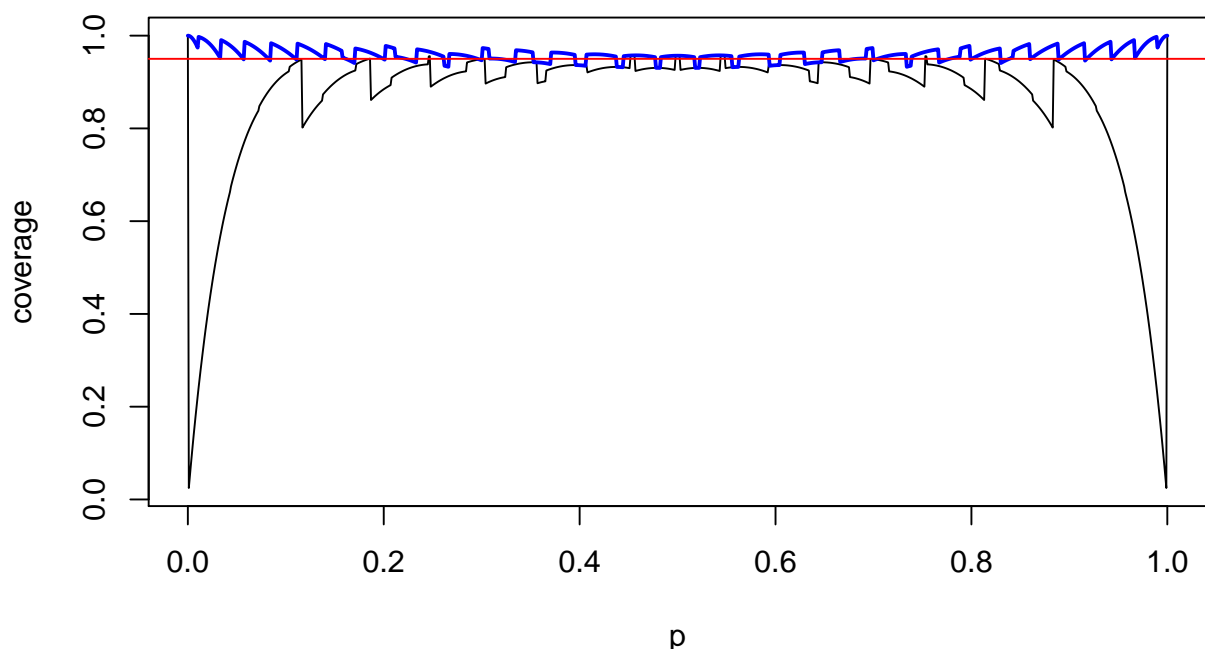

Comment on accuracy of Wald and Agresti's CIs based the above plot. For small success probabilities, with the Wald CI (black line) you might ask for a 95% confidence interval, but actually you get, say, a 10% confidence interval! Agresti's formula gives much better results with coverage probability very close or just a above 95% (thick blue line).

3 Confidence intervals for two proportions

An environmental survey contained a question asking what respondent thought was the major cause of air pollution in this country, giving the following choices: cars, factories, and incinerators. Two versions of the test, A and B, were used. Let p_A and p_B be the respective proportions of people using forms A and B who select factories. If 141 out of 460 people who used version A chose factories and 170 out of 440 people who used version B choose factories. We are interested in assessing whether the two versions of the test are consistent concerning this answer.

1. Find a 95% confidence interval for $d = p_A - p_B$.

```
prop.test(x = c(141, 170), n = c(460, 440), correct = FALSE)


##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(141, 170) out of c(460, 440)
## X-squared = 6.3393, df = 1, p-value = 0.01181
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.14185034 -0.01783346
## sample estimates:
##    prop 1    prop 2
## 0.3065217 0.3863636
```

2. Do the versions seem to be consistent concerning the answer? The interval does not include the value $p_A - p_B = 0$, suggesting that the two versions are inconsistent.

3. Suppose that form B is known to be more difficult than form A. In this case we are only interested in a one-sided interval, i.e. we are only interested in computing an upper bound for $p_A - p_B$.

```
prop.test(x = c(141, 170), n = c(460, 440),
          alternative = "less", correct = FALSE)


##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(141, 170) out of c(460, 440)
## X-squared = 6.3393, df = 1, p-value = 0.005905
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.00000000 -0.02780277
## sample estimates:
##    prop 1    prop 2
## 0.3065217 0.3863636
```

The above output suggests that $d < -0.028$ at the 95% confidence level.

Exercises

1. Do question 4 from the tutorial problems.

2. Do question 2 from the tutorial problems.
   (You might need to revise some information from last week's lab.)

3. Do question 7 from the tutorial problems.

4. Suppose we take a random sample of size 100 from $X \sim \mathrm{N}(\mu, \sigma^2)$. How do you calculate a 95% confidence interval for μ? Do a simulation to calculate the coverage of this interval when $\mu = 10$ and $\sigma = 2$.