

Name: _____

Computing in Statistics, STAT:5400
Midterm 2, Fall 2017

1 Simulation study

The `prop.test` function, with the Yates correction argument set to `FALSE`, computes large sample normal-theory confidence intervals. With appropriate adjustment to the numeric arguments, it can also compute plus-four method intervals.

Conduct a simulation study to compare the true coverage probabilities of large-sample normal theory intervals versus plus-four method intervals, when the nominal coverage is 90%, the sample size is 25 and the true success probability is 0.16.

Use a large enough number of replicate datasets so that the standard error of your estimated coverage probabilities will be no greater than 0.0025.

Time how long it takes to run your simulation study.

Submit:

- Your R code.
- Your R output containing the results of the simulation study.
- R output showing how long it took to run the simulation study.
- A brief paragraph interpreting the results.

```
> doit <- function( n, conf, se, truep, seed)
+ {
+   getInt <- function(v)
+   {
+     prop.test( v[1], v[2], conf.level = conf,
+               correct = FALSE)$conf.int
+   }
+
+   # how many replicate datasets, based on required stand error
+   S <- 1/(4 * se^2)
+
+   # generate data
+   set.seed( seed )
+
+   mydat <- rbinom( S , size = n, prob = truep )
+   normal <- apply( cbind(mydat, rep(n,S) ), 1, getInt )
```

```

+      plus4 <- apply( cbind(mydat+2, rep(n+4,S) ), 1, getInt )
+
+      normalcov <- 0.16 > normal[1,] & 0.16 < normal[2,]
+      plus4cov <- 0.16 > plus4[1,] & 0.16 < plus4[2,]
+
+
+      coverage <- c(mean(normalcov), mean(plus4cov) )
+      names(coverage) = c("normal", "plus4")
+      coverage
+ }
> system.time(coverage <- doit( 25, .90, .0025, 0.16, 17 ))

```

```

      user  system elapsed
7.720    0.002    7.741

```

```

> print(coverage)

```

```

      normal  plus4
0.950775 0.796700

```

```

>

```

This is one of the fairly rare conditions under which the plus-four method performs more poorly than the large-sample, normal-theory method. The plus-four actually is anti-conservative here – its actual coverage is lower than its nominal coverage.

However, `prop.test` must not be doing exactly what is advertised. If you code the computations yourself, you will see both methods coming in right around 90% coverage!

```

> mydat <- rbinom(40000, 25, 0.16)
> z <- qnorm(0.95)
> computeInt <- function( y, n)
+ {
+   phat <- y/n
+   se <- sqrt( phat * (1-phat) / n )
+   mult <- c(-1,1)
+   interval <- phat + mult * z * se
+ }
> normInts <- plus4Ints <- matrix(0, nrow=40000, ncol=2)
> for(i in 1:40000)
+ {
+ normInts[i,] <- computeInt( mydat[i], 25 )
+
+ plus4Ints[i,] <- computeInt( mydat[i] + 2, 29)

```

```

+ }
> normcov <- mean( .16 > normInts[,1] & .16 < normInts[,2])
> plus4cov <- mean( .16 > plus4Ints[,1] & .16 < plus4Ints[,2])
> print(c(normcov, plus4cov))

[1] 0.891025 0.894875

>

```

2 Root-finding

The function

$$f(x) = x^4 - 2x^3 + 1$$

has two real roots in the interval $[0.5, 2.0]$.

Use the `uniroot` function in such a way that you are able to find both roots.

Submit:

- Your R code.
- R output containing results for each root.
- Any other R output that helped you do this.

```

> myfunc <- function(x) {x^4 - 2 * x^3 + 1}
> x <- seq(0.5,2.0, 0.01)
> plot(x, myfunc(x), type="l")
> result1 <- uniroot( myfunc, c(0.5, 1.5) )
> result2 <- uniroot( myfunc, c(1.5, 2.0) )
> result1

```

```

$root
[1] 1.000017

```

```

$f.root
[1] -3.476541e-05

```

```

$iter
[1] 4

```

```

$init.it
[1] NA

```

```

$estim.prec
[1] 6.103516e-05

```

```
> result2

$root
[1] 1.839287

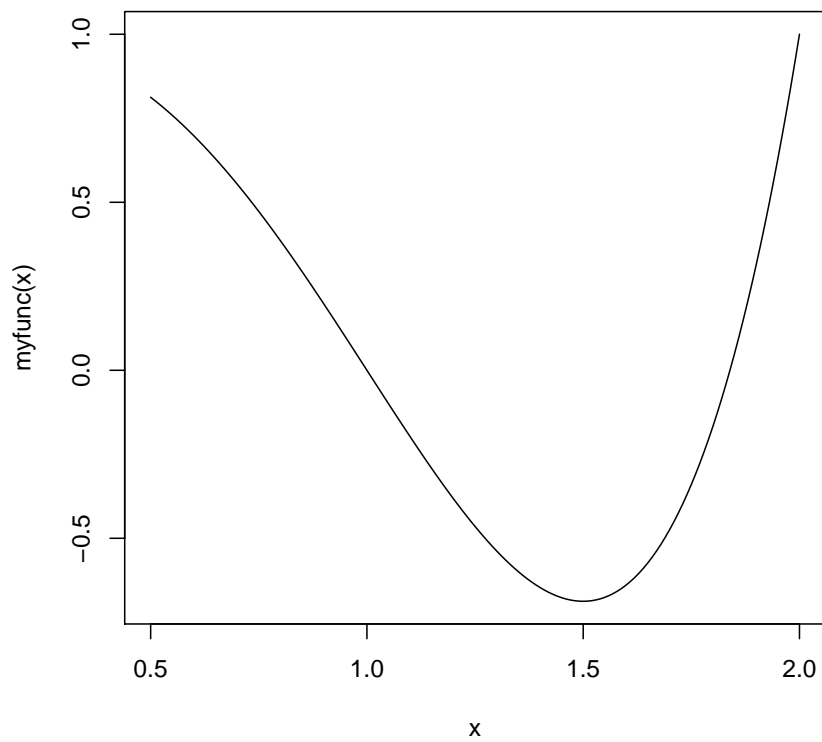
$f.root
[1] 1.702496e-07

$iter
[1] 5

$init.it
[1] NA

$estim.prec
[1] 0.0001039337

>
```



3 Relational database structures

A journal editor wishes to store information about the papers published in her journal and the authors of those papers. Below are the attributes that she wishes to store. Develop a relational database structure, in third normal form, for these data. For each table, indicate any primary and/or foreign keys. Note that the same person could be an author of more than one paper, and a paper may have more than one author.

This is a many-to-many relationship, so a linking table will be needed.

Paper table

Paper ID (primary key)

Paper title

Paper volume number

Paper issue number

Paper starting page number

Author table

Author ID (primary key)

Author name

Author affiliation

Author email

Linking table

(Primary key is combination of the two fields)

Paper ID (foreign key)

Author ID (foreign key)