

**Example:** In R it is very easy to simulate the effects of different sampling schemes and there are several ways to do this.

Consider a red/green marble experiment, where we first define the sample size (number of marbles to be drawn) and set up the sequences of red and green marbles along with the entire set:

```
n<-25
numred<-10
numgreen<-30
green<-rep("G",numgreen)
red<-rep("R",numred)
bowl<-sample(c(green,red))
```

Note the use of the sample command to shuffle the elements of the vector.

The next step is to draw a sample of  $n$  marbles, with replacement:

```
outcome<-sample(bowl,n,replace=T)
```

Note that the arguments to the sample command are first the elements to be sampled and then the sample size. The default is to sample without replacement so we need to change this behaviour.

Next take a look at the outcome:

```
outcome
[1] "G" "G" "R" "G" "R" "R" "G" "G" "G" "R" "G" "G" "G" "G" "G" "R" "G" "G" "G"
[20] "G" "R" "R" "R" "R" "G"
```

This basically says that the sequence obtained was GGRGRRGGGRGGGGGRGGGGR-RRRG.

Suppose that we are primarily interested in the number of red marbles obtained. This can be computed by first recoding "R" to 1 and "G" to 0 and then adding the outcome:

```
x<-sum(ifelse(outcome=="R",1,0))
x
[1] 9

phat<-x/n
phat
[1] 0.36
```

The next step is to do this a few times to see how  $\hat{p}$  varies.

Finally, the real interest is in looking at how this sampling method performs in general. Thus we add a loop to take 1000 such samples and investigate the histogram of  $\hat{p}$ . This can be done with the following code:

```

n<-25
numred<-10
numgreen<-30
green<-rep("G",numgreen)
red<-rep("R",numred)
bowl<-sample(c(green,red))
B<-1000
phat<-rep(NA,B)
for(b in 1:B){
  outcome<-sample(bowl,n,replace=T)
  x<-sum(ifelse(outcome=="R",1,0))
  phat[b]<-x/n
}
hist(phat)

```

In addition to plotting the output, we can compare how well the simulated results fit the theory:

```

truep<-numred/(numgreen+numred)
truep
[1] 0.25
truevar<-truep*(1-truep)/n
truevar
[1] 0.0075
var(phat)
[1] 0.007487801

```

**Example:** Consider an evaluation of whether the confidence interval for a proportion has the correct coverage probability.

```

n<-1200
x<-350
phat<-x/n

p<-0.4
alpha<-0.05
zstar<-qnorm(1-alpha/2)
B<-10000
success<-rep(NA,B)

for(b in 1:B){
  smpl<-ifelse(runif(n)<p,1,0)
  x<-sum(smpl)
  phat<-x/n
  se<-sqrt(phat*(1-phat)/n)
  lo<-phat-zstar*se
  up<-phat+zstar*se
}

```

```
    success[b]<-ifelse(lo<p&p<up,1,0)
  }
  cat("Coverage probability of the confidence interval\n")
  print(table(success)/B)
```

A typical session would look as follows:

```
source("simci.r")
Coverage probability of the confidence interval
success
      0      1
0.0533 0.9467
```