

# Vectorizing code

```
library(downloader)
url <- "https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/femaleControlsPopulation.csv"
filename <- "femaleControlsPopulation.csv"
if (!file.exists(filename)) download(url,destfile=filename)
set.seed(1)
population = unlist( read.csv("femaleControlsPopulation.csv") )
```

To give an example of how we can simulate  $V$   $S$  and we constructed a simulation with :

```
alpha <- 0.05
N <- 12
m <- 10000
p0 <- 0.90 ##10% of diets work, 90% don't
m0 <- m*p0
m1 <- m-m0
nullHypothesis <- c( rep(TRUE,m0), rep(FALSE,m1))
delta <- 3
```

We then ran a Monte Carlo simulation by repeating a procedure in which 10,000 tests were run one by one using sapply.

```
B <- 10 ##number of simulations
system.time(
VandS <- replicate(B,{
  calls <- sapply(1:m, function(i){
    control <- sample(population,N)
    treatment <- sample(population,N)
    if(!nullHypothesis[i]) treatment <- treatment + delta
    t.test(treatment,control)$p.val < alpha
  })
  c(sum(nullHypothesis & calls),sum(!nullHypothesis & calls))
})
)
```

```
##    user  system elapsed
## 12.568   0.004  12.582
```

In each iteration we checked if that iteration was associated with the null or alternative hypothesis. We did this with the line

```
if(!nullHypothesis[i]) treatment <- treatment + delta
```

In R, operations based on matrices are typically much faster than operations performed within ‘ This code runs several times faster than the code above, which is necessary here due to the fact that we will be generating several simulations. Understanding this chunk of code and how it is equivalent to the code above using sapply will take a you long way in helping you code efficiently in R.

```
library(genefilter) ##rowttests is here
set.seed(1)
##Define groups to be used with rowttests
g <- factor( c(rep(0,N),rep(1,N)) )
B <- 10 ##number of simulations
system.time(
VandS <- replicate(B,{
```

```

##matrix with control data (rows are tests, columns are mice)
controls <- matrix(sample(population, N*m, replace=TRUE),nrow=m)

##matrix with control data (rows are tests, columns are mice)
treatments <- matrix(sample(population, N*m, replace=TRUE),nrow=m)

##add effect to 10% of them
treatments[which(!nullHypothesis),]<-treatments[which(!nullHypothesis),]+delta

##combine to form one matrix
dat <- cbind(controls,treatments)

calls <- rowttests(dat,g)$p.value < alpha

  c(sum(nullHypothesis & calls),sum(!nullHypothesis & calls))
})
)

##      user  system elapsed
## 0.616   0.000   0.617

```