BIN784 Week1 Practices

2024-03-03

Şevval Aktürk

```r
#Loading libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.0      v tibble    3.2.1
## v lubridate 1.9.3      v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts --------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(hierfstat)
```

Exercise 1

```r
#Generating three pops
three_pops <- sim.genot(nbpop=3,nbloc=500,nbal=2,size=1000,N=2500)
three_dos<- biall2dos(three_pops[,-1])

#Calculating allele frequencies
popN<-3
N<-nrow(three_dos)/popN
lociN<- 500
size<- 1000

n<-seq(from=0,nrow(three_dos),size)

three_dos_q<-matrix(nrow = popN, ncol = lociN)
colnames(three_dos_q) <- colnames(three_pops[-1])
row.names(three_dos_q) <- paste0("pop", 1:3)

for (i in 1:popN){
  from<-n[i]+1
  to<-n[i+1]
  for (k in 1:lociN){
```

```
    three_dos_q[i,k] <- sum(three_dos[from:to,k])/(2*N)
  }
}

three_dos_p <- 1- three_dos_q

three_dos_q[,1:10]
```

```
##      loc.1  loc.2  loc.3 loc.4  loc.5  loc.6 loc.7  loc.8  loc.9 loc.10
## pop1 0.954 0.8530 0.3995 0.832 0.2080 0.3800 0.618 0.1645 0.5000 0.4255
## pop2 0.942 0.7595 0.4805 0.769 0.3920 0.4520 0.680 0.0805 0.6590 0.3180
## pop3 0.986 0.8345 0.6400 0.796 0.3685 0.5565 0.714 0.2340 0.3015 0.2635
```

```
three_dos_p[,1:10]
```

```
##      loc.1  loc.2  loc.3 loc.4  loc.5  loc.6 loc.7  loc.8  loc.9 loc.10
## pop1 0.046 0.1470 0.6005 0.168 0.7920 0.6200 0.382 0.8355 0.5000 0.5745
## pop2 0.058 0.2405 0.5195 0.231 0.6080 0.5480 0.320 0.9195 0.3410 0.6820
## pop3 0.014 0.1655 0.3600 0.204 0.6315 0.4435 0.286 0.7660 0.6985 0.7365
```

```
#Calculating genotype frequencies
popN<-3
N<-nrow(three_dos)/popN
lociN<- 500
n<-seq(from=0,nrow(three_dos),size)

three_dos_geno<-NULL
temp<-matrix(nrow = popN, ncol = lociN)

for (i in 1:popN){
  from<-n[i]+1
  to<-n[i+1]
  subset_dos<-three_dos[from:to,]
  for (k in 1:lociN){
    temp[1,k] <- sum(subset_dos[,k]==0)/N
    temp[2,k] <- sum(subset_dos[,k]==1)/N
    temp[3,k] <- sum(subset_dos[,k]==2)/N
    rownames(temp)<-paste0("pop",i,c("_homref","_het","_homalt"))
    colnames(temp) <- colnames(three_pops[-1])
  }
  three_dos_geno<-rbind(three_dos_geno,temp)
}

three_dos_geno[,1:10]
```

```
##             loc.1 loc.2 loc.3 loc.4 loc.5 loc.6 loc.7 loc.8 loc.9 loc.10
## pop1_homref 0.001 0.019 0.364 0.032 0.635 0.372 0.151 0.699 0.243  0.329
## pop1_het    0.090 0.256 0.473 0.272 0.314 0.496 0.462 0.273 0.514  0.491
## pop1_homalt 0.909 0.725 0.163 0.696 0.051 0.132 0.387 0.028 0.243  0.180
## pop2_homref 0.005 0.058 0.279 0.045 0.381 0.306 0.103 0.846 0.119  0.469
## pop2_het    0.106 0.365 0.481 0.372 0.454 0.484 0.434 0.147 0.444  0.426
## pop2_homalt 0.889 0.577 0.240 0.583 0.165 0.210 0.463 0.007 0.437  0.105
```

```
## pop3_homref 0.001 0.026 0.120 0.043 0.396 0.194 0.078 0.575 0.485  0.535
## pop3_het    0.026 0.279 0.480 0.322 0.471 0.499 0.416 0.382 0.427  0.403
## pop3_homalt 0.973 0.695 0.400 0.635 0.133 0.307 0.506 0.043 0.088  0.062
```

```
#Calculating heterozygosity
```

```
#Heterozygosity=fAa
```

```
three_dos_geno["pop1_het",1:10]
```

```
##  loc.1  loc.2  loc.3  loc.4  loc.5  loc.6  loc.7  loc.8  loc.9 loc.10
##  0.090  0.256  0.473  0.272  0.314  0.496  0.462  0.273  0.514  0.491
```

```
three_dos_geno["pop2_het",1:10]
```

```
##  loc.1  loc.2  loc.3  loc.4  loc.5  loc.6  loc.7  loc.8  loc.9 loc.10
##  0.106  0.365  0.481  0.372  0.454  0.484  0.434  0.147  0.444  0.426
```

```
three_dos_geno["pop3_het",1:10]
```

```
##  loc.1  loc.2  loc.3  loc.4  loc.5  loc.6  loc.7  loc.8  loc.9 loc.10
##  0.026  0.279  0.480  0.322  0.471  0.499  0.416  0.382  0.427  0.403
```

Exercise 2

```
CC<-25
CT<-5
TT<-0
```

```
#Calculating genotype frequencies
geno_freq <- function(homref,het,homalt) {
  ind<-c(homref,het,homalt)
  geno_data <- data.frame(ind)
  geno_data <- geno_data%>%
    mutate(GenotypeF=ind/sum(ind))
  return(geno_data$GenotypeF)
}
```

```
print(geno_freq(CC,CT,TT))
```

```
## [1] 0.8333333 0.1666667 0.0000000
```

```
#Calculating allele frequencies
al_freq <- function(homref,het,homalt) {
  fC <- (2*homref+het)/(2*(homref+het+homalt))
  fT <- 1-fC
  return(c(fC,fT))
}
```

```
print(al_freq(CC,CT,TT))
```

```
## [1] 0.91666667 0.08333333
```

Exercise 3

```
fA2 <- 0.0002
fA <- sqrt(fA2)
het<- 2*fA*(1-fA)
het
```

```
## [1] 0.02788427
```

Exercise 4

```
#Inbreeding coefficient=2fAfa-fAa/2fAfa
#Heterozygous frequency=2*p*q

F<-NULL
for (i in 1:popN){
  exp_het<-2*three_dos_p[paste0("pop",i),]*three_dos_q[paste0("pop",i),]
  obs_het<-three_dos_geno[paste0("pop",i,"_het"),]
  temp<-(exp_het-obs_het)/(exp_het)
  F<- rbind(F,temp)
}

rownames(F)<-paste0("pop",1:3)
F[,1:10]
```

```
##            loc.1         loc.2       loc.3        loc.4        loc.5       loc.6
## pop1 -0.02543068 -0.0208069160  0.01417154  0.027014652  0.04695998 -0.05263158
## pop2  0.02994363  0.0008745769  0.03653457 -0.047067367  0.04756310  0.02299593
## pop3  0.05824399 -0.0100662698 -0.04166667  0.008523007 -0.01199896 -0.01090829
##            loc.7        loc.8       loc.9        loc.10
## pop1  0.021501550  0.006837542 -0.02800000 -0.004296384
## pop2  0.002757353  0.007021775  0.01209956  0.017871964
## pop3 -0.018589254 -0.065586575 -0.01378101 -0.038297193
```

Exercise 5

```
#Defining inbreeding coefficient function
F_inb <- function(homref,het,homalt){
  fA <- homalt+(het/2)
  fa <- 1-fA
  F=(2*fA*fa-het)/(2*fA*fa)
  return(F)
}

#Calculating inbreeding coefficient
fAA<-0.58
fAa<-0.07
faa<-0.35
```

```r
print(F_inb(fAA,fAa,faa))
```

```
## [1] 0.8521803
```