trueMean1 <- 5

trueSD1 <- 5

population1 <- rnorm(1e6, trueMean1, trueSD1)

trueMean2 <- 4

trueSD2 <-5

population2 <- rnorm (1e6, trueMean2, trueSD2)

Size <- 50

Sample1 <- sample(population1, Size)

Sample2 <- sample(population2, Size)

#yes the two samples are different, and the populations were also different.

boxplot(Sample1, Sample2)

source("http://jonsmitchell.com/code/simFxn04.R")

MatGrandma <- makeFounder("grandma\_mom")

MatGrandpa <- makeFounder("grandpa\_mom")

PatGrandma <- makeFounder("grandma\_da")

PatGrandpa <- makeFounder("grandpa\_da")

Alan <- makeBaby(PatGrandma, PatGrandpa)

Brenda <- makeBaby(MatGrandma, MatGrandpa)

Focus <- makeBaby(Brenda, Alan)

#the number should be 50% or 0.5.

ToMom <- length(grep("mom", Focus))/length(Focus)

#these numbers are .3076 and .1924. This does not match my expectation

ToMomMom <- length(grep("grandma\_mom", Focus))/ length(Focus)

ToMomDad <- length(grep("grandpa\_mom", Focus))/ length(Focus)

#Focus is not equally realted to each maternal grandparent or each paternal grandparent. The averagre relatedness fo Focus to all four grandparents is .25.

Sibling\_01 <- makeBaby(Brenda, Alan)

#I would expect the sibling to share 50% of DNA with the focus. Actually it shares .47.

ToSib <- length(intersect(Focus, Sibling\_01))/ length(Focus)

ManySiblings <- replicate(1e3, length(intersect(Focus, makeBaby(Brenda, Alan)))/ length(Focus))

#Focus shares different amounts of genes with each of the 1,000 siblings.

quantile(ManySiblings)

mean(ManySiblings)

plot(density(ManySiblings), main="", xlab="proportion shared genes")

HWE <- function(p) {

aa <- p^2

ab <- 2 \* p \* (1-p)

bb <- (1-p)^2

return(c(aa=aa, ab=ab, bb=bb))

}

HWE(0.5)

plot(1, 1, type="n", xlim=c(0, 1), ylim=c(0, 1), xlab="freq.allele a", ylab="geno.freq")

p <- seq(from=0, to=1, by=0.01)

GenoFreq <- t(sapply(p, HWE))

lines(p, GenoFreq[,"aa"], lwd=2, col="red")

#the frequesncy of aa increases as the frequence of allele a increases in the population. As a decreases aa frequency decreases. Time is not shown in the plot or geographic space.

lines(p, GenoFreq[,"bb"], lwd=2, col="blue")

legend("top", legend=c("aa", "ab", "bb"), col=c("red", "purple", "blue"), lty=1, lwd=2, bty="n")

Pop <- simPop(500)

points(Pop[, "freqa"], Pop[, "Genotypes.aa"]/50, pch=22, bg="red")

#the frequency of aa genotypes does not match the expectation from Hardy-Weinberg.

Pop <- simPop(50)

points(Pop[, "freqa"], Pop[, "Genotypes.aa"]/50, pch=22, bg="red")

#there are a lot more points on the graph. This is because the population size changed to create a higher frequency since there is a smaller population.

install.packages("learnPopGen")

library(learnPopGen)

x <- genetic.drift(Ne=200, nrep=5, pause=0.01)

PopSizes<- 5:50

Samples <- rep(PopSizes, 5)

tExt <- sapply(Samples, function(x) nrow(simPop(x, 500)))

Line <- lm(tExt ~Samples)

summary(Line)

Line$coef

plot(Samples, tExt)

abline(Line)

Line <- lm(tExt ~ Samples + 0)

summary(Line)

Line$coef

plot(Samples, tExt)

abline(Line)

#as the population size increases the distance from the points to the line increases as well. This means that as the population size increases there is more extinction for alleles.

#Extra Credit