Stats4GeneticsHW5

Jesse Elder

11/5/2020

Using the **FAMuSS** data, estimate the resistin haplotype frequencies for Caucasians and African Americans separately.

Problem 1

```
#A Haplotype is a list of ALL COMBINATIONS OF SNPs and the frequency of those combinations

#Take only resistin data drom FAMuSS
llamo<-names(fms)[substr(names(fms),1,8)=="resistin"]
NewGen<-fms[,is.element(names(fms),llamo)]
#NewGen

#resistin haplotype frequencies
Geno <- cbind(substr(resistin_c30t,1,1), substr(resistin_c30t,2,2), substr(resistin_c398t,1,1), substr(

###Do Expectation-Maximization on haplotypes
#caucasians
Geno.C <- Geno[Race=="Caucasian" & !is.na(Race),]
#Geno.C

HaploEM <- haplo.em(Geno.C,locus.label=llamo,control=haplo.em.control(min.posterior=1e-4))
#HaploEM


#african americans
Geno.AA <- Geno[Race=="African Am" & !is.na(Race),]
HaploEM2 <- haplo.em(Geno.AA,locus.label=llamo,control=haplo.em.control(min.posterior=1e-4))
#HaploEM2[5]
#HaploEM2
```

Example 5.3 Haplotype Functions

```
HapDesign <- function(HaploEM){
    Nobs <- length(unique(HaploEM$indx.subj)) # number of observations
    Nhap <- length(HaploEM$hap.prob)    # number of haplotypes
    XmatHap <- matrix(data=0,nrow=Nobs,ncol=Nhap)
    for (i in 1:Nobs){
        IDSeq <- seq(1:sum(HaploEM$nreps))[HaploEM$indx.subj==i]
        for (j in 1:length(IDSeq)){
            XmatHap[i,HaploEM$hap1code[IDSeq][j]] <-
                XmatHap[i,HaploEM$hap1code[IDSeq][j]] +
                HaploEM$post[IDSeq][j]
            XmatHap[i,HaploEM$hap2code[IDSeq][j]] <-
                XmatHap[i,HaploEM$hap2code[IDSeq][j]] +
                HaploEM$post[IDSeq][j]
            }
        }
    return(XmatHap)
}

HapFreqSE <- function(HaploEM){
    HapMat <- HapDesign(HaploEM)
    Nobs <- length(unique(HaploEM$indx.subj)) # number of observations
    Nhap <- length(HaploEM$hap.prob)    # number of haplotypes
    S.Full<-matrix(data=0, nrow=Nobs, ncol=Nhap-1)
    for(i in 1:Nobs){
        for(k in 1:(Nhap-1)){
        S.Full[i,k]<-HapMat[i,k]/HaploEM$hap.prob[k]-
            HapMat[i,Nhap]/HaploEM$hap.prob[Nhap]
        }
    }
    Score<-t(S.Full)%*%S.Full
    invScore<-solve(Score)
    HapSE<-c(sqrt(diag(invScore)),
        sqrt(t(rep(1,Nhap-1))%*%invScore%*%rep(1,Nhap-1)))
    return(HapSE)
    }
```

Finishing up Problem 1

```
#Find the most common haplotype and then find the matching haplotypes between groups
check<-HaploEM$haplotype[which.max(HaploEM$hap.prob),]

HaploEM$haplotype[8,]==check  #Yup it worked
```

```
##   resistin_c30t resistin_c398t resistin_g540a resistin_c980g resistin_c180g
## 8          TRUE           TRUE           TRUE           TRUE           TRUE
##   resistin_a537c
## 8           TRUE
```

```
HaploEM2$haplotype[5,]==check #EM1 row 8 is EM2 row 5
```

```
##   resistin_c30t resistin_c398t resistin_g540a resistin_c980g resistin_c180g
## 5          TRUE           TRUE           TRUE           TRUE           TRUE
##   resistin_a537c
```

```
## 5           TRUE
```

```
FreqDiff <- HaploEM2$hap.prob[5] - HaploEM$hap.prob[8]
s1 <- HapFreqSE(HaploEM)[8]
s2 <- HapFreqSE(HaploEM2)[5]
SE <- sqrt(s1^2 + s2^2)
CI <- c(FreqDiff - 1.96*SE, FreqDiff + 1.96*SE)
CI
```

```
## [1] -0.3315211 -0.1535679
```

Since the confidence interval doesn't contain zero, we can conclude that the haplotype for african americans that is also most common in caucasians is significantly less common in african americans.

Based on the **HGDP** data, estimate the **AKT1** haplotype frequencies within groups defined by the variable *Population*.

Problem 2

```
hags<-hgdp_akt1[,7:10]

#Names of SNPs
llamo<-c("AKT1.C0756A", "AKT1.C6024T", "AKT1.G2347T", "AKT1.G2375A")


###Change this into 2X24 akt1 SNPs
Geno <- cbind(substr(hgdp_akt1$AKT1.C0756A,1,1), substr(hgdp_akt1$AKT1.C0756A,2,2),
              substr(hgdp_akt1$AKT1.C6024T,1,1), substr(hgdp_akt1$AKT1.C6024T,2,2),
              substr(hgdp_akt1$AKT1.G2347T,1,1), substr(hgdp_akt1$AKT1.G2347T,2,2),
              substr(hgdp_akt1$AKT1.G2375A,1,1), substr(hgdp_akt1$AKT1.G2375A,2,2))

###Change this into 2X24 akt1 SNPs


###Do EM looping through levels of "Population" variable
levels(as.factor(Population))
```

```
##  [1] "Adygei"         "Balochi"        "Bantu"          "Bedouin"
##  [5] "Biaka Pygmies"  "Biaka Pygmies " "Brahui"         "Burusho"
##  [9] "Cambodian"      "Colombian"      "Dai"            "Daur"
## [13] "Druze"          "French"         "French Basque"  "Han"
## [17] "Hazara"         "Hezhen"         "Japanese"       "Kalash"
## [21] "Karitiana"      "Lahu"           "Makrani"        "Mandenka"
## [25] "Maya"           "Mbuti Pygmies"  "Miaozu"         "Mongola"
## [29] "Mozabite"       "NAN Melanesian" "Naxi"           "North Italian"
## [33] "Orcadian"       "Oroqen"         "Palestinian"    "Papuan"
## [37] "Pathan"         "Pima"           "Russian"        "San "
## [41] "Sardinian"      "She"            "Sindhi"         "Surui"
## [45] "Tu"             "Tujia"          "Tuscan"         "Uygur"
## [49] "Xibo"           "Yakut"          "Yizu"           "Yoruba"
```

```
listy<-list()
for(i in 1:length(levels(as.factor(Population)))){
  Geno.C <- Geno[Population==levels(as.factor(Population))[i] & !is.na(Population),]

  HaploEM <- haplo.em(Geno.C,locus.label=llamo,control=haplo.em.control(min.posterior=1e-4))
  listy[[i]]<-cbind(HaploEM$haplotype,hap.prob=HaploEM$hap.prob)
}
listy[[52]]  #It works! This is for population 52
```

```
##   AKT1.C0756A AKT1.C6024T AKT1.G2347T AKT1.G2375A hap.prob
## 1           A           C           G           G     0.02
## 2           A           T           T           A     0.28
## 3           A           T           T           G     0.02
## 4           C           C           G           G     0.42
## 5           C           C           T           G     0.22
## 6           C           T           T           G     0.04
```

```
#Finding most common haplotype by population
listy2<-list()
for(j in 1:52){
  listy2[[j]]<-listy[[j]][which.max(listy[[j]]$hap.prob),] #Look at the most common haplotype for each
}
###Do EM looping through levels of "Population" variable

listy2[[1]]
```

```
##   AKT1.C0756A AKT1.C6024T AKT1.G2347T AKT1.G2375A  hap.prob
## 2           C           C           G           G 0.8235294
```

Is there an association between *Gender* and *Geographic.origin* in the **HGDP** data?

Problem 3

```
lads<-ifelse(hgdp_akt1$Gender=="M",1,0)
hmod<-lm(lads~hgdp_akt1$Geographic.origin);summary(hmod)
```

```
##
## Call:
## lm(formula = lads ~ hgdp_akt1$Geographic.origin)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9167 -0.3636  0.1200  0.3044  0.7083
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       6.667e-01  8.080e-02
## hgdp_akt1$Geographic.originBougainville          -3.030e-01  1.242e-01
## hgdp_akt1$Geographic.originBrazil                -2.000e-01  1.043e-01
## hgdp_akt1$Geographic.originCambodia              -1.212e-01  1.560e-01
```

```
## hgdp_akt1$Geographic.originCentral African Republic      2.500e-01  1.094e-01
## hgdp_akt1$Geographic.originChina                         2.899e-02  8.714e-02
## hgdp_akt1$Geographic.originColombia                     -2.821e-01  1.470e-01
## hgdp_akt1$Geographic.originDemocratic Republic of Congo  2.000e-01  1.400e-01
## hgdp_akt1$Geographic.originFrance                       -1.384e-01  1.011e-01
## hgdp_akt1$Geographic.originIsrael (Carmel)              -3.750e-01  1.030e-01
## hgdp_akt1$Geographic.originIsrael (Central)             -3.333e-01  1.018e-01
## hgdp_akt1$Geographic.originIsrael (Negev)               -9.524e-02  1.026e-01
## hgdp_akt1$Geographic.originItaly                        -5.556e-02  1.094e-01
## hgdp_akt1$Geographic.originItaly (Bergamo)              -2.381e-02  1.432e-01
## hgdp_akt1$Geographic.originJapan                         7.527e-02  1.133e-01
## hgdp_akt1$Geographic.originKenya                         2.500e-01  1.512e-01
## hgdp_akt1$Geographic.originMexico                       -3.267e-01  1.022e-01
## hgdp_akt1$Geographic.originNamidia                       3.333e-01  1.858e-01
## hgdp_akt1$Geographic.originNew Guinea                    9.804e-02  1.344e-01
## hgdp_akt1$Geographic.originNigeria                      -1.467e-01  1.198e-01
## hgdp_akt1$Geographic.originOrkney Islands               -2.292e-01  1.370e-01
## hgdp_akt1$Geographic.originPakistan                      2.133e-01  8.665e-02
## hgdp_akt1$Geographic.originRussia                       -2.667e-02  1.198e-01
## hgdp_akt1$Geographic.originRussia Caucasus              -2.549e-01  1.344e-01
## hgdp_akt1$Geographic.originSenegal                      -1.275e-15  1.212e-01
## hgdp_akt1$Geographic.originSiberia                       5.333e-02  1.198e-01
## hgdp_akt1$Geographic.originSouth Africa                  3.333e-01  1.761e-01
##                                                        t value Pr(>|t|)
## (Intercept)                                              8.251 4.74e-16 ***
## hgdp_akt1$Geographic.originBougainville                 -2.439 0.014879 *
## hgdp_akt1$Geographic.originBrazil                       -1.917 0.055473 .
## hgdp_akt1$Geographic.originCambodia                     -0.777 0.437320
## hgdp_akt1$Geographic.originCentral African Republic      2.285 0.022508 *
## hgdp_akt1$Geographic.originChina                         0.333 0.739476
## hgdp_akt1$Geographic.originColombia                     -1.919 0.055216 .
## hgdp_akt1$Geographic.originDemocratic Republic of Congo  1.429 0.153284
## hgdp_akt1$Geographic.originFrance                       -1.368 0.171486
## hgdp_akt1$Geographic.originIsrael (Carmel)              -3.641 0.000285 ***
## hgdp_akt1$Geographic.originIsrael (Central)             -3.273 0.001097 **
## hgdp_akt1$Geographic.originIsrael (Negev)               -0.928 0.353476
## hgdp_akt1$Geographic.originItaly                        -0.508 0.611702
## hgdp_akt1$Geographic.originItaly (Bergamo)              -0.166 0.868019
## hgdp_akt1$Geographic.originJapan                         0.664 0.506790
## hgdp_akt1$Geographic.originKenya                         1.654 0.098464 .
## hgdp_akt1$Geographic.originMexico                       -3.196 0.001435 **
## hgdp_akt1$Geographic.originNamidia                       1.794 0.073045 .
## hgdp_akt1$Geographic.originNew Guinea                    0.730 0.465721
## hgdp_akt1$Geographic.originNigeria                      -1.224 0.221310
## hgdp_akt1$Geographic.originOrkney Islands               -1.673 0.094688 .
## hgdp_akt1$Geographic.originPakistan                      2.462 0.013977 *
## hgdp_akt1$Geographic.originRussia                       -0.223 0.823963
## hgdp_akt1$Geographic.originRussia Caucasus              -1.897 0.058067 .
## hgdp_akt1$Geographic.originSenegal                       0.000 1.000000
## hgdp_akt1$Geographic.originSiberia                       0.445 0.656403
## hgdp_akt1$Geographic.originSouth Africa                  1.893 0.058655 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4426 on 1037 degrees of freedom
## Multiple R-squared:  0.1686, Adjusted R-squared:  0.1477
## F-statistic: 8.086 on 26 and 1037 DF,  p-value: < 2.2e-16
```

```
tab<-table(hgdp_akt1$Gender,hgdp_akt1$Geographic.origin) #Looks like this is right

chisq.test(tab)
```

```
## Warning in chisq.test(tab): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab
## X-squared = 179.35, df = 26, p-value < 2.2e-16
```

There is a significant relationship between gender and geographic origin. This means that it's important to stratify by gender when finding associations between haplotypes and a given trait.

Apply **haploytpe trend regression** (**HTR**) to determine if there is an association between the *resistin haplotypes* and change in *nondominant arm muscle strength* within *African Americans* using the FAMuSS data.

Problem 4

```
llamo<-names(fms)[substr(names(fms),1,8)=="resistin"] #only resistin
Geno <- cbind(substr(resistin_c30t,1,1), substr(resistin_c30t,2,2), substr(resistin_c398t,1,1), substr(
```

```
Geno.AA <- Geno[Race=="African Am" & !is.na(Race),] #only data for african americans
Geno.AA <- setupGeno(Geno.AA)

#Use expectation-maximization
HaploEM <- haplo.em(Geno.AA,locus.label=llamo,control=haplo.em.control(min.posterior=1e-4))

HapMat <- HapDesign(HaploEM)
Trait <- NDRM.CH[Race=="African Am" & !is.na(Race)]
mod1 <- (lm(Trait~HapMat))
mod2 <- (lm(Trait~1))        #the null model

summary(mod1)
```

```
##
## Call:
## lm(formula = Trait ~ HapMat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -60.097 -17.741  -1.463  23.159  54.030
##
```

```
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -16.944     71.160  -0.238    0.813
## HapMat1       46.870     38.267   1.225    0.229
## HapMat2       30.130     37.799   0.797    0.431
## HapMat3       34.808     37.386   0.931    0.358
## HapMat4       32.784     39.576   0.828    0.413
## HapMat5       42.714     37.629   1.135    0.264
## HapMat6       25.310     40.638   0.623    0.538
## HapMat7       20.940     51.527   0.406    0.687
## HapMat8        2.393     44.736   0.053    0.958
## HapMat9           NA         NA      NA       NA
##
## Residual standard error: 34.71 on 34 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1267, Adjusted R-squared:  -0.07883
## F-statistic: 0.6164 on 8 and 34 DF,  p-value: 0.7579
```

```
summary(mod2)
```

```
##
## Call:
## lm(formula = Trait ~ 1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -44.71 -29.41 -13.01  22.94  63.69
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   53.012      5.097    10.4 3.42e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.42 on 42 degrees of freedom
##   (1 observation deleted due to missingness)
```

```
anova(mod1,mod2)
```

```
## Analysis of Variance Table
##
## Model 1: Trait ~ HapMat
## Model 2: Trait ~ 1
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     34 40973
## 2     42 46915 -8   -5942.2 0.6164 0.7579
```

The HapMat model is not significant for any variable. Additionally, the results of the anova suggest that the addition of the HapMat variables to the model does not significantly improve it compared to the null model.

Using the expectation-maximization approach of the haplo.glm() function, determine if there is an association between the *resistin haplotypes* and change in *nondominant arm muscle strength*, as measured by NDRM.CH, within African Americans, based on the FAMuSS data. Consider both dominant and additive

genetic models.

Problem 5

```
llamo<-names(fms)[substr(names(fms),1,8)=="resistin"] #only resistin

Geno <- cbind(substr(resistin_c30t,1,1), substr(resistin_c30t,2,2), substr(resistin_c398t,1,1), substr(

#african americans
Geno.AA <- Geno[Race=="African Am" & !is.na(Race),]
Geno.AA <- setupGeno(Geno.AA)
Trait <- NDRM.CH[Race=="African Am" & !is.na(Race)]


Dat <- data.frame(Geno.AA=Geno.AA, Trait=Trait)

#haplo.glm's!!!
hap1<-haplo.glm(Trait~Geno.AA,data=Dat,allele.lev=attributes(Geno.AA)$unique.alleles,control=haplo.glm.
hap2<-haplo.glm(Trait~Geno.AA,data=Dat,allele.lev=attributes(Geno.AA)$unique.alleles,control=haplo.glm.

summary(hap1)
```

```
##
## Call:
## haplo.glm(formula = Trait ~ Geno.AA, data = Dat, control = haplo.glm.control(haplo.effect = "dominan
##     allele.lev = attributes(Geno.AA)$unique.alleles)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -59.823  -14.773   -1.656   31.335   52.699
##
## Coefficients:
##                 coef       se   t.stat  pval
## (Intercept)  76.5226  18.1015   4.2274 0.001
## Geno.AA.2   -14.3549  16.8762  -0.8506 0.406
## Geno.AA.3   -14.8665  23.0647  -0.6446 0.527
## Geno.AA.4   -12.9256  20.9743  -0.6163 0.545
## Geno.AA.6     4.4851  21.5891   0.2077 0.838
## Geno.AA.8   -24.2813  23.8472  -1.0182 0.322
## Geno.AA.10  -24.6677  43.0137  -0.5735 0.573
## Geno.AA.11  -36.3150  31.0929  -1.1680 0.258
## Geno.AA.14  -43.2226  43.3262  -0.9976 0.332
##
## (Dispersion parameter for gaussian family taken to be 1549.496)
##
##     Null deviance: 33957  on 26  degrees of freedom
## Residual deviance: 27891  on 18  degrees of freedom
## AIC: 284.01
##
## Number of Fisher Scoring iterations: 17
##
##
```

```
## Haplotypes:
##            loc.1 loc.2 loc.3 loc.4 loc.5 loc.6 hap.freq
## Geno.AA.2      C     C     A     C     G     A  0.22488
## Geno.AA.3      C     C     A     C     G     C  0.11111
## Geno.AA.4      C     C     G     C     C     A  0.10846
## Geno.AA.6      C     C     G     G     C     A  0.12963
## Geno.AA.8      C     T     A     C     G     A  0.07142
## Geno.AA.10     C     T     A     G     G     C  0.01852
## Geno.AA.11     C     T     G     C     C     A  0.03969
## Geno.AA.14     T     C     G     C     C     A  0.01852
## haplo.base     C     C     A     C     C     A  0.27778
```

```
summary(hap2)
```

```
##
## Call:
## haplo.glm(formula = Trait ~ Geno.AA, data = Dat, control = haplo.glm.control(haplo.effect = "additive
##     allele.lev = attributes(Geno.AA)$unique.alleles)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -63.155  -17.158   -1.301   32.476   51.165
##
## Coefficients:
##                 coef       se   t.stat  pval
## (Intercept)  79.8554  18.3521   4.3513 0.000
## Geno.AA.2   -14.0920  17.1301  -0.8226 0.421
## Geno.AA.3   -11.8015  18.1505  -0.6502 0.524
## Geno.AA.4   -16.9284  20.9168  -0.8093 0.429
## Geno.AA.6    -4.4623  17.2247  -0.2591 0.799
## Geno.AA.8   -24.5515  23.9795  -1.0239 0.319
## Geno.AA.10  -28.2634  42.9271  -0.6584 0.519
## Geno.AA.11  -39.7466  31.1476  -1.2761 0.218
## Geno.AA.13  -46.5554  43.6458  -1.0667 0.300
##
## (Dispersion parameter for gaussian family taken to be 1568.156)
##
##     Null deviance: 33957  on 26  degrees of freedom
## Residual deviance: 28227  on 18  degrees of freedom
## AIC: 284.33
##
## Number of Fisher Scoring iterations: 18
##
##
## Haplotypes:
##            loc.1 loc.2 loc.3 loc.4 loc.5 loc.6 hap.freq
## Geno.AA.2      C     C     A     C     G     A  0.22456
## Geno.AA.3      C     C     A     C     G     C  0.11111
## Geno.AA.4      C     C     G     C     C     A  0.10878
## Geno.AA.6      C     C     G     G     C     A  0.12963
## Geno.AA.8      C     T     A     C     G     A  0.07174
## Geno.AA.10     C     T     A     G     G     C  0.01852
## Geno.AA.11     C     T     G     C     C     A  0.03937
## Geno.AA.13     T     C     G     C     C     A  0.01852
## haplo.base     C     C     A     C     C     A  0.27778
```

```
#Sadly neither are at all significant
```

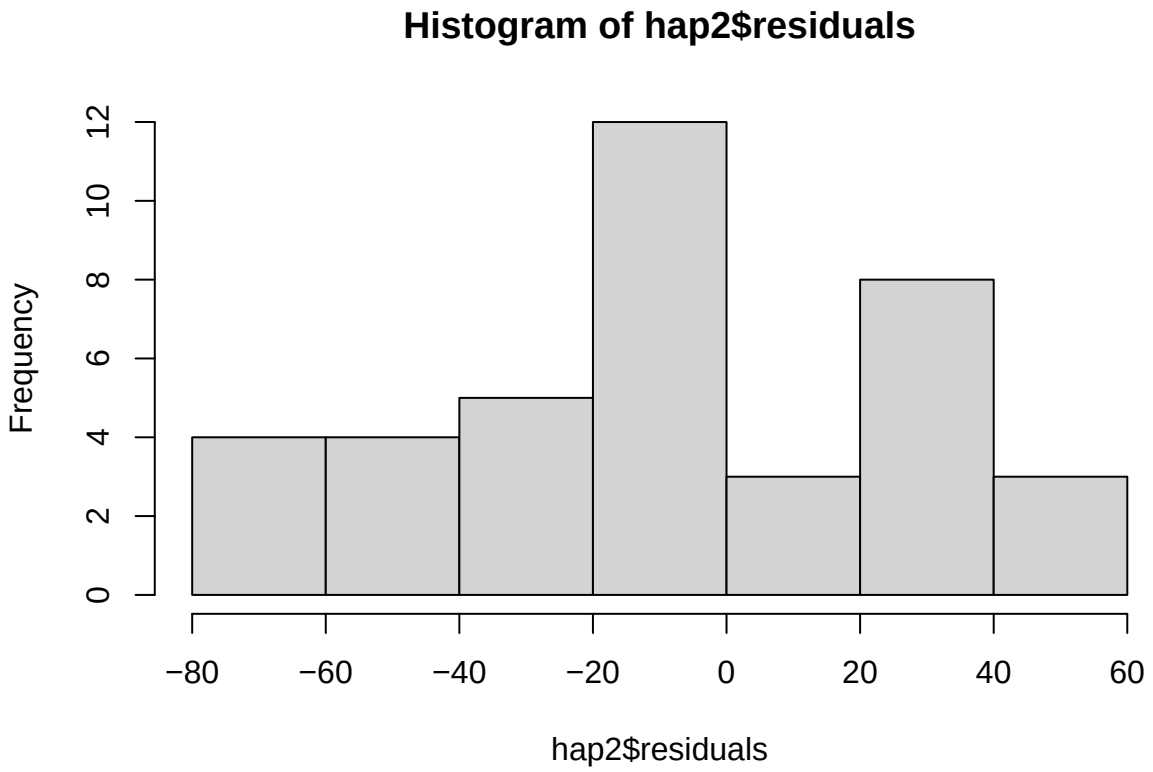
For both dominant and additive models there is no significant association between haplotype and non-dominant arm strength within African Americans

Examine the fit of the model you fit in the *previous problem* – do the residuals look approximately normal? Compare to a model that uses a logarithmic transformation. Does gender have an impact on the change in the non-dominant arm muscle strength in this subpopulation? Is there an impact of gender when you consider the effect of the resistin haplotypes? Explain your findings.

Problem 6

```
hist(hap1$residuals)
```


```
hist(hap2$residuals)
```




```
#haplo.glm's!!!
hap1log<-haplo.glm(log(Trait)~Geno.AA,data=Dat,allele.lev=attributes(Geno.AA)$unique.alleles,control=hap
hap2log<-haplo.glm(log(Trait)~Geno.AA,data=Dat,allele.lev=attributes(Geno.AA)$unique.alleles,control=hap

summary(hap1log)
```

```
##
## Call:
## haplo.glm(formula = log(Trait) ~ Geno.AA, data = Dat, control = haplo.glm.control(haplo.effect = "dom
##     allele.lev = attributes(Geno.AA)$unique.alleles)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7484  -0.4273   0.1313   0.5213   1.0083
##
## Coefficients:
##                coef      se  t.stat  pval
## (Intercept)  4.1290  0.4000 10.3217 0.000
## Geno.AA.2   -0.2676  0.3750 -0.7137 0.485
## Geno.AA.3   -0.2647  0.5098 -0.5192 0.610
## Geno.AA.4   -0.2632  0.4649 -0.5661 0.578
## Geno.AA.6    0.1013  0.4771  0.2123 0.834
## Geno.AA.8   -0.3981  0.5278 -0.7541 0.461
## Geno.AA.10  -0.2371  0.9510 -0.2493 0.806
## Geno.AA.11  -0.5817  0.6975 -0.8341 0.415
## Geno.AA.14  -0.6235  0.9575 -0.6512 0.523
##
## (Dispersion parameter for gaussian family taken to be 0.7567325)
##
##     Null deviance: 15.322  on 26  degrees of freedom
```

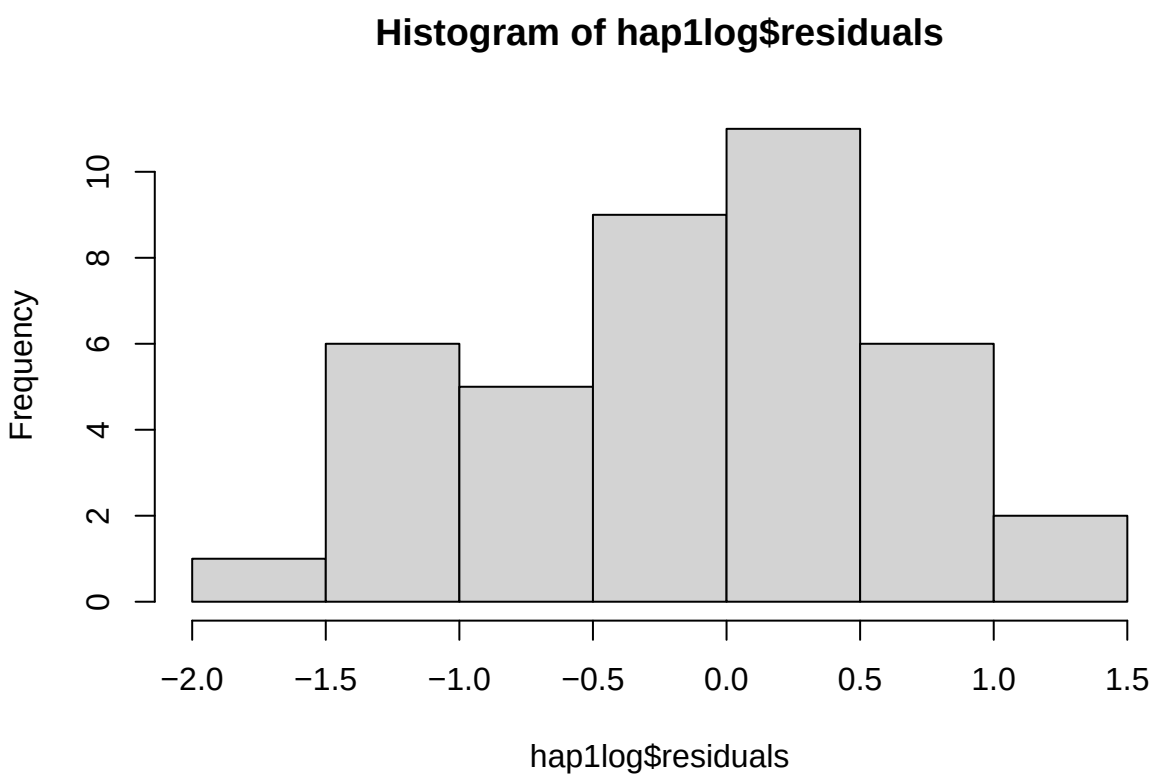
```
## Residual deviance: 13.621  on 18  degrees of freedom
## AIC: 78.149
##
## Number of Fisher Scoring iterations: 18
##
##
## Haplotypes:
##            loc.1 loc.2 loc.3 loc.4 loc.5 loc.6 hap.freq
## Geno.AA.2      C     C     A     C     G     A  0.22446
## Geno.AA.3      C     C     A     C     G     C  0.11111
## Geno.AA.4      C     C     G     C     C     A  0.10888
## Geno.AA.6      C     C     G     G     C     A  0.12963
## Geno.AA.8      C     T     A     C     G     A  0.07184
## Geno.AA.10     C     T     A     G     G     C  0.01852
## Geno.AA.11     C     T     G     C     C     A  0.03927
## Geno.AA.14     T     C     G     C     C     A  0.01852
## haplo.base     C     C     A     C     C     A  0.27778
```

```
summary(hap2log)
```

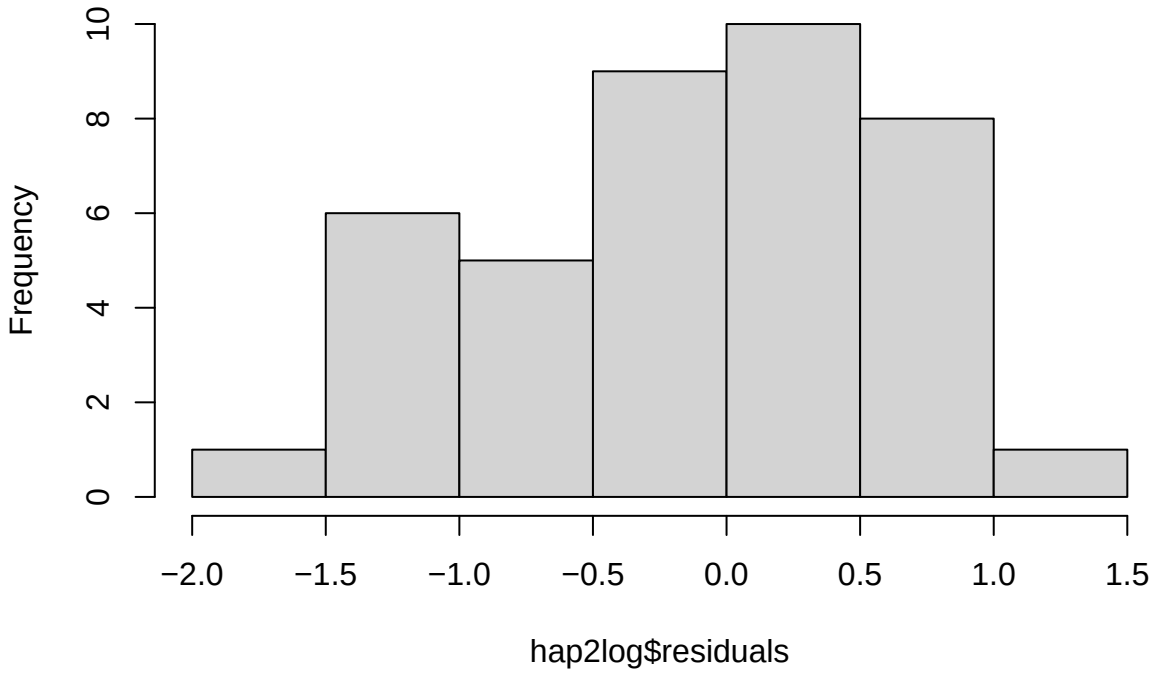
```
##
## Call:
## haplo.glm(formula = log(Trait) ~ Geno.AA, data = Dat, control = haplo.glm.control(haplo.effect = "ad
##     allele.lev = attributes(Geno.AA)$unique.alleles)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8745  -0.4584   0.1035   0.5684   1.0188
##
## Coefficients:
##                 coef       se   t.stat  pval
## (Intercept)  4.14440  0.40654 10.19425 0.000
## Geno.AA.2   -0.24436  0.38164 -0.64028 0.530
## Geno.AA.3   -0.15387  0.40212 -0.38263 0.706
## Geno.AA.4   -0.31380  0.46454 -0.67551 0.508
## Geno.AA.6   -0.02369  0.38154 -0.06208 0.951
## Geno.AA.8   -0.37809  0.53194 -0.71078 0.486
## Geno.AA.10  -0.27571  0.95134 -0.28981 0.775
## Geno.AA.11  -0.60661  0.70217 -0.86392 0.399
## Geno.AA.14  -0.63885  0.96682 -0.66077 0.517
##
## (Dispersion parameter for gaussian family taken to be 0.7694718)
##
##     Null deviance: 15.322  on 26  degrees of freedom
## Residual deviance: 13.850  on 18  degrees of freedom
## AIC: 78.6
##
## Number of Fisher Scoring iterations: 18
##
##
## Haplotypes:
##            loc.1 loc.2 loc.3 loc.4 loc.5 loc.6 hap.freq
## Geno.AA.2      C     C     A     C     G     A  0.22408
## Geno.AA.3      C     C     A     C     G     C  0.11111
## Geno.AA.4      C     C     G     C     C     A  0.10925
```

```
## Geno.AA.6      C     C     G     G     C     A  0.12963
## Geno.AA.8      C     T     A     C     G     A  0.07221
## Geno.AA.10     C     T     A     G     G     C  0.01852
## Geno.AA.11     C     T     G     C     C     A  0.03890
## Geno.AA.14     T     C     G     C     C     A  0.01852
## haplo.base     C     C     A     C     C     A  0.27778
```

```
hist(hap1log$residuals)
```




```
hist(hap2log$residuals)
```




```
Gendre<-fms$Gender[Race=="African Am" & !is.na(Race)]
linmod<-lm(Trait~Gendre); summary(linmod) #There is a significant relationship
```

```
##
## Call:
## lm(formula = Trait ~ Gendre)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.764 -22.364  -2.464  18.361  58.636
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   58.064      5.504  10.550 4.04e-13 ***
## GendreMale   -27.786     11.890  -2.337   0.0245 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31.62 on 40 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1201, Adjusted R-squared:  0.09814
## F-statistic: 5.461 on 1 and 40 DF,  p-value: 0.02453
```

```
hap1Gen<-haplo.glm(Trait~Geno.AA+Gendre,data=Dat,allele.lev=attributes(Geno.AA)$unique.alleles,control=
hap2Gen<-haplo.glm(Trait~Geno.AA+Gendre,data=Dat,allele.lev=attributes(Geno.AA)$unique.alleles,control=

summary(hap1Gen)
```

```
##
## Call:
```

```
## haplo.glm(formula = Trait ~ Geno.AA + Gendre, data = Dat, control = haplo.glm.control(haplo.effect =
##     allele.lev = attributes(Geno.AA)$unique.alleles)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -62.43  -10.65    0.00   30.34   46.13
##
## Coefficients:
##                 coef       se   t.stat  pval
## (Intercept)  79.1308  17.8520   4.4326 0.000
## Geno.AA.2    -5.3709  17.7865  -0.3020 0.766
## Geno.AA.3   -19.8889  22.9111  -0.8681 0.397
## Geno.AA.4    -9.8387  20.2179  -0.4866 0.633
## Geno.AA.6    -5.0091  22.2903  -0.2247 0.825
## Geno.AA.8    -9.9178  25.2330  -0.3930 0.699
## Geno.AA.10  -36.2599  42.9230  -0.8448 0.410
## Geno.AA.11   -6.1457  39.0468  -0.1574 0.877
## Geno.AA.14  -45.8308  42.4880  -1.0787 0.296
## GendreMale  -36.6321  27.3252  -1.3406 0.198
##
## (Dispersion parameter for gaussian family taken to be 1486.538)
##
##     Null deviance: 33957  on 26  degrees of freedom
## Residual deviance: 25271  on 17  degrees of freedom
## AIC: 283.35
##
## Number of Fisher Scoring iterations: 17
##
##
## Haplotypes:
##            loc.1 loc.2 loc.3 loc.4 loc.5 loc.6 hap.freq
## Geno.AA.2      C     C     A     C     G     A  0.22131
## Geno.AA.3      C     C     A     C     G     C  0.11111
## Geno.AA.4      C     C     G     C     C     A  0.11202
## Geno.AA.6      C     C     G     G     C     A  0.12963
## Geno.AA.8      C     T     A     C     G     A  0.07498
## Geno.AA.10     C     T     A     G     G     C  0.01852
## Geno.AA.11     C     T     G     C     C     A  0.03613
## Geno.AA.14     T     C     G     C     C     A  0.01852
## haplo.base     C     C     A     C     C     A  0.27778
```

```
summary(hap2Gen)
```

```
##
## Call:
## haplo.glm(formula = Trait ~ Geno.AA + Gendre, data = Dat, control = haplo.glm.control(haplo.effect =
##     allele.lev = attributes(Geno.AA)$unique.alleles)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -64.275  -12.318   -0.646   29.795   40.052
##
## Coefficients:
##                 coef       se   t.stat  pval
## (Intercept)  80.9754  17.9269   4.5170 0.000
```

```
## Geno.AA.2    -5.3293  17.7882  -0.2996 0.768
## Geno.AA.3   -14.2978  17.7884  -0.8038 0.433
## Geno.AA.4   -12.4405  20.2330  -0.6149 0.547
## Geno.AA.6    -9.9274  17.2255  -0.5763 0.572
## Geno.AA.7    -9.4935  25.2657  -0.3757 0.712
## Geno.AA.9   -38.1461  42.3494  -0.9007 0.380
## Geno.AA.10   -7.2244  39.2761  -0.1839 0.856
## Geno.AA.12  -47.6754  42.5553  -1.1203 0.278
## GendreMale  -37.4077  26.5944  -1.4066 0.178
##
## (Dispersion parameter for gaussian family taken to be 1489.579)
##
##     Null deviance: 33957  on 26  degrees of freedom
## Residual deviance: 25323  on 17  degrees of freedom
## AIC: 283.4
##
## Number of Fisher Scoring iterations: 15
##
##
## Haplotypes:
##            loc.1 loc.2 loc.3 loc.4 loc.5 loc.6 hap.freq
## Geno.AA.2      C     C     A     C     G     A  0.22129
## Geno.AA.3      C     C     A     C     G     C  0.11111
## Geno.AA.4      C     C     G     C     C     A  0.11204
## Geno.AA.6      C     C     G     G     C     A  0.12963
## Geno.AA.7      C     T     A     C     G     A  0.07501
## Geno.AA.9      C     T     A     G     G     C  0.01852
## Geno.AA.10     C     T     G     C     C     A  0.03610
## Geno.AA.12     T     C     G     C     C     A  0.01852
## haplo.base     C     C     A     C     C     A  0.27778
```

They both look a *little* bit normal. The dominant model is less so. The dominant model almost looks trimodal. However, the additive model looks roughly normal and can be considered so. When log transformed, both of the models' residuals more normally distributed. Gender has a significant impact on change in non-domninant arm strength. Specifically males have a significantly lower change in non-dom arm strength. When you consider gender along with the the haplotypes, neither the difference in gender or haplotype have a significant impact on change in non-dom arm strength.