GENE613 - Homework 5

Daniel Osorio - dcosorioh@tamu.edu
Department of Veterinary Integrative Biosciences
Texas A&M University

1. Express

   a) The regression coefficient of y on x as a function of the correlation of variables y and x.

   $$\beta = r(\frac{s_y}{s_x}) \tag{1}$$

   Where β is the regression coefficient of y on x, r is the correlation between the variables x and y, and s_x and s_y are the standard deviation of the x and y variables respectively.

   b) The correlation coefficient of y and x as a function of the regression of y on x

   $$r = \frac{\beta s_x}{s_y} \tag{2}$$

   Where β is the regression coefficient of y on x, r is the correlation between the variables x and y, and s_x and s_y are the standard deviation of the x and y variables respectively.

2. Commercial bovine chip arrays of $\sim$ 50,000 marker loci or $\sim$ 770,000 marker loci were designed with average r^2 (as the measurement of linkage disequilibrium) between adjacent markers of 0,2 and 0,6 respectively. Why? Because the number of marker loci that can be identified are positively correlated with the r^2 value, lower number of r^2 requires fewer markers to cover the entire genome due to 80 % of the time the identified markers will be segregated together.

   Use the Dorper sheep data that follows:

```
> dorperSheep <- read.csv("../data/dorperSheep.csv")
> attach(dorperSheep)
> dorperSheep
```

```
   ID AGE YEAR BW FAMACHA   FEC   PCV
1   1   M    1 NA       3    NA    NA
2   1   M    2 47       3    NA    NA
3   1   M    3 NA       3    NA    NA
4   1   M    4 NA       3 11900 17.75
5   8   M    2 36       3     1    NA
6   8   M    3 NA       3    NA    NA
7   8   M    4 NA       3  7800 14.25
8   9   M    1 NA       2  4100 19.00
9   9   M    2 40       2    NA 24.00
```

```
10  9   M    3 NA       3    NA    NA
11 11   M    2 48       2    NA    NA
12 11   M    3 NA       2    NA    NA
13 11   M    4 NA       3  3400 14.75
14 20   Y    1 NA      NA 11100    NA
15 20   Y    2 38       4  8400 18.00
16 20   Y    3 NA       3  1200 24.00
17 20   Y    4 NA       3    NA    NA
18 26   Y    1 NA       1    NA    NA
19 26   Y    2 37       2    NA    NA
20 26   Y    3 NA       2    NA    NA
21 26   Y    4 NA       3  6200 14.00
22 27   Y    1 NA       4  9900 18.00
23 27   Y    2 44       3  2400 20.00
24 27   Y    3 NA       4    NA    NA
25 27   Y    4 NA       4  3400    NA
26 31   Y    1 NA       2    NA    NA
27 31   Y    2 39       2    NA    NA
28 31   Y    3 NA       3  7000 15.00
29 32   Y    1 NA       1  1700 25.75
30 32   Y    2 40       1    NA    NA
31 32   Y    3 NA       2   600 22.25
32 32   Y    4 NA       2 10400    NA
```

3. Adjust BW using age group averages. What does this do and why? This procedure center the values to 0 by subtracting the mean of each group to the original values. Adjust the values allow us to identify the deviation with respect to the mean of each group.

```
> adjust <- function (values, groups) {
+   newValues <- values
+   for (group in unique(groups)) {
+     groupElements <- groups %in% group
+     groupCenter <- values[groupElements]
+     groupCenter <- groupCenter[!is.na(groupCenter)]
+     groupCenter <- mean(groupCenter, na.rm = TRUE)
+     newValues[groupElements] <- (values[groupElements] - groupCenter)
+   }
+   return(newValues)
+ }
> adjust(values = BW, groups = AGE)

 [1]    NA  4.25    NA    NA -6.75    NA    NA    NA -2.75    NA  5.25    NA
[13]    NA    NA -1.60    NA    NA    NA -2.60    NA    NA    NA  4.40    NA
[25]    NA    NA -0.60    NA    NA  0.40    NA    NA
```

4. Predict breeding values and accuracies for BW given $h^2 = 0{,}35$ and $R = 0{,}59$

```
> BV <- adjust(values = BW, groups = AGE) * 0.35
> ACC <- ifelse(test = !is.na(BV),yes = sqrt(0.35),no = NA)
> cbind(BV,ACC)
```

```
           BV      ACC
 [1,]      NA       NA
 [2,]  1.4875 0.591608
 [3,]      NA       NA
 [4,]      NA       NA
 [5,] -2.3625 0.591608
 [6,]      NA       NA
 [7,]      NA       NA
 [8,]      NA       NA
 [9,] -0.9625 0.591608
[10,]      NA       NA
[11,]  1.8375 0.591608
[12,]      NA       NA
[13,]      NA       NA
[14,]      NA       NA
[15,] -0.5600 0.591608
[16,]      NA       NA
[17,]      NA       NA
[18,]      NA       NA
[19,] -0.9100 0.591608
[20,]      NA       NA
[21,]      NA       NA
[22,]      NA       NA
[23,]  1.5400 0.591608
[24,]      NA       NA
[25,]      NA       NA
[26,]      NA       NA
[27,] -0.2100 0.591608
[28,]      NA       NA
[29,]      NA       NA
[30,]  0.1400 0.591608
[31,]      NA       NA
[32,]      NA       NA
```

5. Adjust $FAMACHA$ values using year averages (don't use ewe age).

```
> adjust(values = FAMACHA, groups = YEAR)

 [1]  0.8333333  0.5555556  0.2222222  0.0000000  0.5555556  0.2222222
 [7]  0.0000000 -0.1666667 -0.4444444  0.2222222 -0.4444444 -0.7777778
[13]  0.0000000         NA  1.5555556  0.2222222  0.0000000 -1.1666667
[19] -0.4444444 -0.7777778  0.0000000  1.8333333  0.5555556  1.2222222
[25]  1.0000000 -0.1666667 -0.4444444  0.2222222 -1.1666667 -1.4444444
[31] -0.7777778 -1.0000000
```

6. Predict producing abilities and corresponding accuracies for each ewe for eacth trait.

```
> computeERPA <- function(measurement, R) {
+   variable <- adjust(values = measurement, groups = YEAR)
+   ERPA <- t(sapply(unique(ID), function(id) {
+     value <- !is.na(variable)
```

```
+     validRecord <- ID %in% id & value
+     n <- max(order(YEAR[validRecord]))
+     vRn <- (n * R / (1 + (n - 1) * R))
+     year <- YEAR[validRecord]
+     nValues <-
+       round(mean(variable[validRecord][year %in% seq_len(max(year))]), 2)
+     ERPA <- round(nValues * vRn, 2)
+     ACC <- round(sqrt(vRn), 2)
+     cbind(id,
+           nValues,
+           n,
+           ERPA,
+           ACC)
+   }))
+   colnames(ERPA) <- c("ID", "AVERAGE", "n", "ERPA", "ACC")
+   return(ERPA)
+ }
> computeERPA(measurement = BW, R = 0.59)

      ID AVERAGE n  ERPA  ACC
 [1,]  1       6 1  3.54 0.77
 [2,]  8      -5 1 -2.95 0.77
 [3,]  9      -1 1 -0.59 0.77
 [4,] 11       7 1  4.13 0.77
 [5,] 20      -3 1 -1.77 0.77
 [6,] 26      -4 1 -2.36 0.77
 [7,] 27       3 1  1.77 0.77
 [8,] 31      -2 1 -1.18 0.77
 [9,] 32      -1 1 -0.59 0.77

> computeERPA(measurement = FAMACHA, R = 0.5)

      ID AVERAGE n  ERPA  ACC
 [1,]  1    0.40 4  0.32 0.89
 [2,]  8    0.26 3  0.20 0.87
 [3,]  9   -0.13 3 -0.10 0.87
 [4,] 11   -0.41 3 -0.31 0.87
 [5,] 20    0.59 3  0.44 0.87
 [6,] 26   -0.60 4 -0.48 0.89
 [7,] 27    1.15 4  0.92 0.89
 [8,] 31   -0.13 3 -0.10 0.87
 [9,] 32   -1.10 4 -0.88 0.89

> computeERPA(measurement = FEC, R = 0.1)

      ID  AVERAGE n    ERPA  ACC
 [1,]  1  4716.67 1  471.67 0.32
 [2,]  8 -1491.33 2 -271.15 0.43
 [3,]  9 -2600.00 1 -260.00 0.32
 [4,] 11 -3783.33 1 -378.33 0.32
 [5,] 20  2488.78 3  622.20 0.50
```

```
 [6,] 26  -983.33 1  -98.33 0.32
 [7,] 27  -594.56 3 -148.64 0.50
 [8,] 31  4066.67 1  406.67 0.32
 [9,] 32 -1372.22 3 -343.06 0.50
```

```
> computeERPA(measurement = PCV, R = 0.15)
```

```
      ID AVERAGE n  ERPA  ACC
 [1,]  1    2.56 1  0.38 0.39
 [2,]  8   -0.94 1 -0.14 0.39
 [3,]  9    0.71 2  0.19 0.51
 [4,] 11   -0.44 1 -0.07 0.39
 [5,] 20    0.46 2  0.12 0.51
 [6,] 26   -1.19 1 -0.18 0.39
 [7,] 27   -1.79 2 -0.47 0.51
 [8,] 31   -5.42 1 -0.81 0.39
 [9,] 32    3.33 2  0.87 0.51
```