Livestock Breeding and Genomics - Solution 3

Peter von Rohr

2019-10-11

Problem 1: Breeding Value

We are considering a quantitative trait that depends on a given bi-allelic locus G. The frequency of the favorable allele corresponds to 0.08. Suppose that genotype frequencies follow the Hardy-Weinberg equilibrium. The difference between the homozygous genotypes corresponds to 15. The heterozygous genotype has a value of -1.5.

a) Compute the breeding values and the dominance deviations for the three genotypes.
b) Because of selecting the positive allele the frequency has increased to 0.096. How does this increased allele frequency change the breeding values?

Hint: Have a look at the summary table of all values in the course notes.

Solution

According to the summary table of all values in the course notes, the breeding values depend on the term α. Therefore, we start by computing α first.

$$\alpha = a + (q - p)d$$

Based on the problem description, we know that $a = 7.5$, $p = 0.08$, $q = 1 - p = 0.92$ and $d = -1.5$. Therefore

$$\alpha = 7.5 + (0.92 - 0.08) * (-1.5) = 6.24$$

a) The summary table of all values then looks as follows.

Genotype	Genotypic Value	Breeding Value	Dominance Deviation
G_1G_1	$a = 7.5$	$2q\alpha = 11.4816$	$-2q^2d = 2.5392$
G_1G_2	$d = -1.5$	$(q-p)\alpha = 5.2416$	$2pqd = -0.2208$
G_2G_2	$-a = -7.5$	$-2p\alpha = -0.9984$	$-2p^2d = 0.0192$

b) Based on the change in the allele frequency to $p = 0.096$ and $q = 0.904$. The value of α changes to $\alpha = 6.288$. This has consequences for the whole summary table.

Genotype	Genotypic Value	Breeding Value	Dominance Deviation
G_1G_1	$a = 7.5$	$2q\alpha = 11.368704$	$-2q^2d = 2.451648$
G_1G_2	$d = -1.5$	$(q-p)\alpha = 5.080704$	$2pqd = -0.260352$
G_2G_2	$-a = -7.5$	$-2p\alpha = -1.207296$	$-2p^2d = 0.027648$

Due to the increment in the allele frequency p from 0.08 to 0.096 the value of α got bigger. But the breeding values decreased, because the negative influence of incrementing p on the breeding values was bigger than the positive change of α.

Problem 2: Allele Substitution

What is the meaning of the term **allele substituion** an how big is it in 1a) and 1b)?

Solution

The effect of **allele substitution** occurs in the difference of the breeding values between two genotypes where on of these genotypes has one favorable allele more than the other. For a single bi-allelic locus there are two possible differences that fullfill the requirement from the previous sentence, namely $BV_{12} - BV22$ and $BV_{11} - BV_{12}$. The result of both differences is the same and corresponds to $\alpha = a + (p - q) * d$.

The allele substitution (α) in 1a) corresponds to 6.24 in 1b) the value is 6.288.

Problem 3: Reading Data into R

You can download a file in csv-format from the course website. The URL is https://charlotte-ngs.github.io/LBGFS2019/ex/w04/iris.csv. Read the data from that csv-file into R using the function `read.csv2()`. Test the consequences of specifying the option `stringsAsFactors=FALSE`.

Hints:

1. You can first download the csv-file to your local computer and then read the data, or you can directly indicate the URL when reading the data. You get more information with the command `?read.csv2` at the R-console.
2. Assign the result of `read.csv2()` to a variable
3. Use the function `str()` on the result of `read.csv2()` to see the difference between the two results of reading the data.
4. Use the description at https://bookdown.org/rdpeng/rprogdatascience/getting-data-in-and-out-of-r.html as a reference to read data into R. There is also a video on the same subject under https://youtu.be/Z_dc_FADyi4.

Solution

- Read the data directly using the URL:

```
dfIris1 <- read.csv2(file = "https://charlotte-ngs.github.io/LBGFS2019/ex/w04/iris.csv")
str(dfIris1)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
dfIris2 <- read.csv2(file = "https://charlotte-ngs.github.io/LBGFS2019/ex/w04/iris.csv",
                     stringsAsFactors = FALSE)
str(dfIris2)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : chr  "setosa" "setosa" "setosa" "setosa" ...
```

- Read from the local computer

```
dfIris1 <- read.csv2(file = "iris.csv")
str(dfIris1)
dfIris2 <- read.csv2(file = "iris.csv", stringsAsFactors = FALSE)
str(dfIris2)
```

Additional Problem: Create a plot in R

Plot the values in the columns `Sepal.Length` and `Petal.Length` of the Iris data set. The plot should look like the following figure.

Solution

```
plot(dfIris2$Sepal.Length, dfIris2$Petal.Length)
```