

## Chapter 10

# Supplement 2: Solutions to Exercises

**Exercise 10.1.** On February 1, 1898, Hermon Bumpus gathered sparrows that had been harmed by a severe winter storm outside of his laboratory at Brown University. Of the 136 birds brought inside, 72 revived and the rest died. In order to examine whether survival was related to characteristics of the birds, Bumpus measured several aspects of the birds' size. Some of the traits Bumpus measured included length from the tip of the beak to the tip of the tail, weight, and the length of the humerus bone in the wing. All of Bumpus' data is presented in the text file *bumpus.txt* in the companion website for this volume. (Data were obtained from the Field Museum website: <http://www.fieldmuseum.org/explore/hermon-bumpus-and-house-sparrows>.)

The variables in the dataset are:

line:	line number from Bumpus' tables
sex:	$m$ = male; $f$ = female
age:	for males only, $a$ = adult, $y$ = young; for females, the value is set to NA because this is how missing values are indicated in R
survival:	TRUE if survived, FALSE if perished
totallength:	"from tip of the beak to the tip of the tail" (mm)
alarextent:	"from tip to tip of the extended wings" (mm)
weight:	weight of the bird (g)
beakhead:	length of beak and head (mm), "from tip of the beak to the occiput"
humerus:	length of humerus (in.)
femur:	length of femur (in.)
tibiotarsus:	length of tibiotarsus (in.)
skullwidth:	width of skull (in.), "from the postorbital bone of one side to the postorbital bone of the other"
sternum:	length of keel of sternum (in.)

The following R commands will help get you started.

□

### QUESTIONS WITH SOLUTIONS EXERCISE 10.1

1. Save the data file in a convenient location and read the data into your R workspace using  

```
> bumpus=read.table(file=file.choose(), header=TRUE)
```
2. A column of the data array is referenced using both the table name and the column name. For example, the weight variable is `bumpus$weight`. You may prefer to break out the columns into separate vectors with shorter names; for example,  

```
> weight=bumpus$weight
```
3. Make histograms for weight and length of the humerus bone for all of the birds, and then for just the surviving birds. To limit data to only the surviving birds, use  

```
> hist(bumpus$weight[bumpus$survival==TRUE])
```

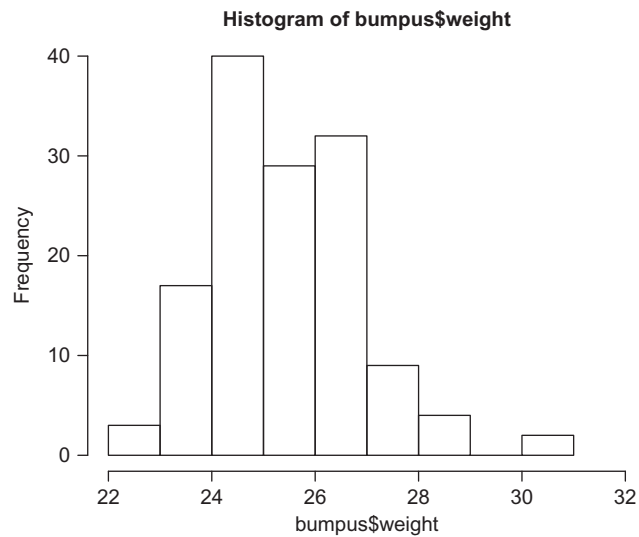
To help with comparisons, you can specify the  $x$ -axis scale. For example, set the range from 22 to 32 this way:

```
> hist(bumpus$weight[bumpus$survival==TRUE], xlim=c(22,32))
```

What differences do you note in shape, center, and spread? Use mean and standard deviation in your comparisons.

```
> hist(bumpus$weight)
```

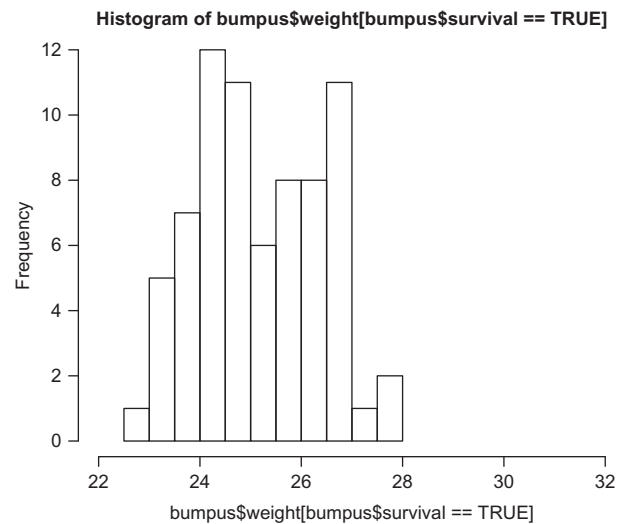
See Figure S10.7 for the resulting histogram.



**FIGURE S10.7** Histogram of weight for all birds; answer to Question 3 in Exercise 10.1.

```
> hist(bumpus$weight[bumpus$survival==TRUE], xlim=c(22,32))
```

See Figure S10.8 for the resulting histogram.



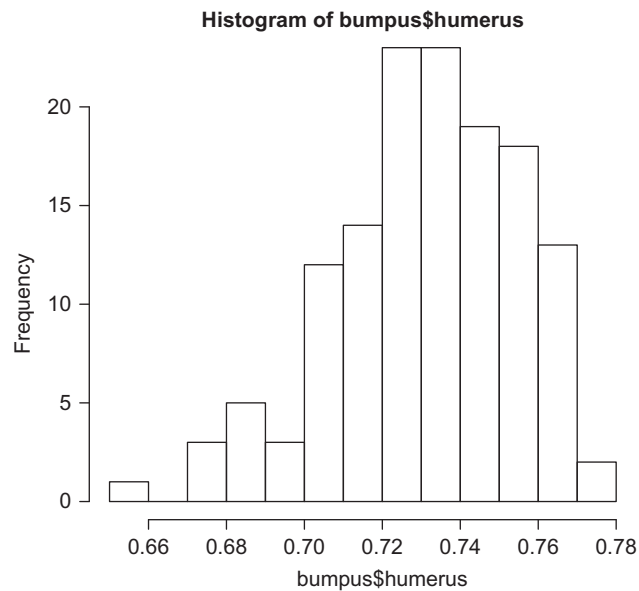
**FIGURE S10.8** Histogram of weight for surviving birds; answer to Question 3 in Exercise 10.1.

```
> hist(bumpus$humeral)
```

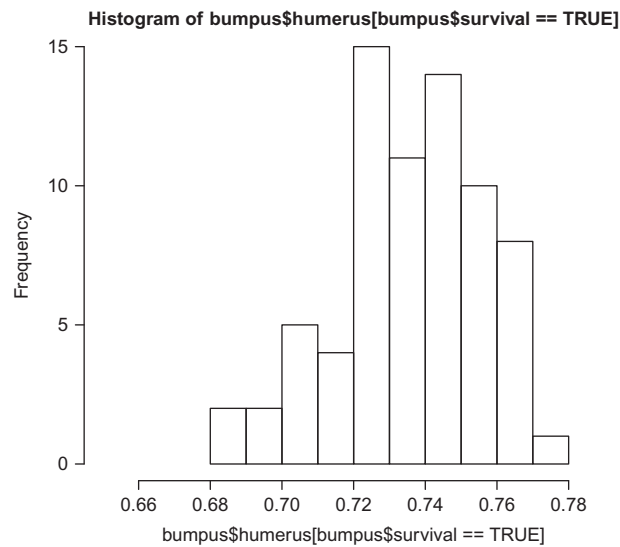
See Figure S10.9 for the resulting histogram.

```
> hist(bumpus$humeral[bumpus$survival==TRUE], xlim=c(0.65, 0.78))
```

See Figure S10.10 for the resulting histogram.



**FIGURE S10.9** Histogram of humerus length for all birds; answer to Question 3 in Exercise 10.1.



**FIGURE S10.10** Histogram of humerus length for surviving birds; answer to Question 3 in Exercise 10.1.

**SHAPE:** Distributions of weight are fairly symmetric. Both distributions of length of humerus are slightly skewed left. In both comparisons, the surviving sparrows' measurements are less variable than the sparrows taken all together.

**CENTERS:** Means of surviving birds are close to the means of all birds.

```
> mean(bumpus$weight)
[1] 25.525
```

Mean weight for all birds is 25.5 g.

```
> mean(bumpus$weight[bumpus$survival==TRUE])
[1] 25.22639
```

Mean weight for the surviving birds is 25.2 g.

```
> mean(bumpus$humerus)
[1] 0.7319412
```

Mean humerus length for all birds is 0.732 in.

```
> mean(bumpus$humerus[bumpus$survival==TRUE])
[1] 0.7359444
```

Mean humerus length for the surviving birds is 0.736 in.

**SPREAD:** Standard deviations for the surviving birds are smaller than for all the birds taken together.

```
> sd(bumpus$weight)
[1] 1.475215
```

Standard deviation of weight for all birds is 1.48 g.

```
> sd(bumpus$weight[bumpus$survival==TRUE])
[1] 1.257428
```

Standard deviation of weight for the surviving birds is 1.26 g. This is 85% of the standard deviation for weight in all birds.

```
> sd(bumpus$humerus)
[1] 0.02307821
```

Standard deviation of humerus length for all birds is 0.0231 in.

```
> sd(bumpus$humerus[bumpus$survival==TRUE])
[1] 0.02033445
```

Standard deviation of humerus length for the surviving birds is 0.0203 in. This is 88% of the standard deviation for humerus length in all birds.

4. Calculate the means of weight and length of the humerus bone for males and females separately. How do males and females differ?

*R note:* For the mean weight of females, use

```
> mean(bumpus$weight[bumpus$sex=="f"])
[1] 25.02857
```

Mean weight for all female birds is 25.0 g.

```
> mean(bumpus$weight[bumpus$sex=="m"])
[1] 25.8046
```

Mean weight for all male birds is 25.8 g.

Average weight of males is about 0.8 g larger than the average weight of females.

```
> mean(bumpus$humerus[bumpus$sex=="f"])
[1] 0.7270204
```

Mean humerus length for all female birds is 0.727 in.

```
> mean(bumpus$humerus[bumpus$sex=="m"])
[1] 0.7347126
```

Mean humerus length for all male birds is 0.735 in.

The mean lengths of the humerus bone for males and females are approximately equal.

5. Comparing means for weight and length of the humerus bone, are there differences between birds that survived and all birds? Make comparisons separately for males and females.

*R note:* To specify females that survived, use the logical “and” operator, denoted by &. The mean weight of that subgroup, for example, is obtained with

```
> mean(bumpus$weight[bumpus$sex=="f" & bumpus$survival==TRUE])
> mean(bumpus$weight[bumpus$sex=="f"])
[1] 25.02857
```

Mean weight for all females is 25.0 g.

```
> mean(bumpus$weight[bumpus$sex=="f"&bumpus$survival==TRUE])
[1] 24.61905
```

Mean weight for surviving females is 24.6 g.

```
> mean(bumpus$weight[bumpus$sex=="m"])
[1] 25.8046 g
```

Mean weight for all males is 25.8 g.

```
> mean(bumpus$weight[bumpus$sex=="m"&bumpus$survival==TRUE])
[1] 25.47647 g
```

Mean weight for surviving males is 25.5 g.

```
> mean(bumpus$humerus[bumpus$sex=="f"])
[1] 0.7270204
```

Mean humerus length for all females is 0.727 in.

```
> mean(bumpus$humerus[bumpus$sex=="f"&bumpus$survival==TRUE])
[1] 0.7283333
```

Mean humerus length for surviving females is 0.728 in.

```
> mean(bumpus$humerus[bumpus$sex=="m"])
[1] 0.7347126
```

Mean humerus length for all males is 0.735 in.

```
> mean(bumpus$humerus[bumpus$sex=="m"&bumpus$survival==TRUE])
[1] 0.7390784
```

Mean humerus length for surviving males is 0.739 in.

6. Calculate the variance and standard deviation for weight and length of the humerus bone for males and females separately.

```
> var(bumpus$weight[bumpus$sex=="f"])
[1] 2.115833
```

Variance of weight in all females is 2.1 g<sup>2</sup>.

```
> var(bumpus$weight[bumpus$sex=="m"])
[1] 2.015793
```

Variance of weight in all males is 2.0 g<sup>2</sup>.

```
> sd(bumpus$weight[bumpus$sex=="f"])
[1] 1.45459
```

Standard deviation of weight in all females is 1.5 g.

```
> sd(bumpus$weight[bumpus$sex=="m"])
[1] 1.419786
```

Standard deviation of weight in all males is 1.42 g.

```
> var(bumpus$humerus[bumpus$sex=="f"])
[1] 0.0004910204
```

Variance of humerus length in all females is 0.000491 in.<sup>2</sup>.

```
> var(bumpus$humerus[bumpus$sex=="m"])
[1] 0.0005404397
```

Variance of humerus length in all males is 0.000540 in.<sup>2</sup>.

```
> sd(bumpus$humerus[bumpus$sex=="f"])
[1] 0.02215898
```

Standard deviation of humerus length in all females is 0.0222 in.

```
> sd(bumpus$humerus[bumpus$sex=="m"])
[1] 0.02324736
```

Standard deviation of humerus length in all males is 0.0232 in.

7. Are the variances in weight and humerus length the same for surviving and all sparrows? (Again, make separate comparisons for males and females.) If not, speculate on how natural selection might be at work.

Weight in females:

```
> var(bumpus$weight[bumpus$sex=="f" & bumpus$survival==TRUE])
[1] 1.135619
```

Variance of weight in all females was 2.1 g<sup>2</sup>, and variance of weight in surviving females is 1.1 g<sup>2</sup>.

Weight in males:

```
> var(bumpus$weight[bumpus$sex=="m" & bumpus$survival==TRUE])
[1] 1.572235
```

Variance of weight in all males was 2.0 g<sup>2</sup>, and variance of weight in surviving males is 1.6 g<sup>2</sup>.

Variance of weight for survivors is smaller than for all birds, and the difference is particularly large for females.

Humerus length in females:

```
> var(bumpus$humeral[bumpus$sex=="f" & bumpus$survival==TRUE])
[1] 0.0002679333
```

Variance of humerus length in all females was 0.000491 in.<sup>2</sup>, and variance of humerus length in surviving females is 0.000268 in.<sup>2</sup>. Variance of humerus length for surviving females is about 55% of the variance of all females.

Humerus length in males:

```
> var(bumpus$humeral[bumpus$sex=="m" & bumpus$survival==TRUE])
[1] 0.0004456337
```

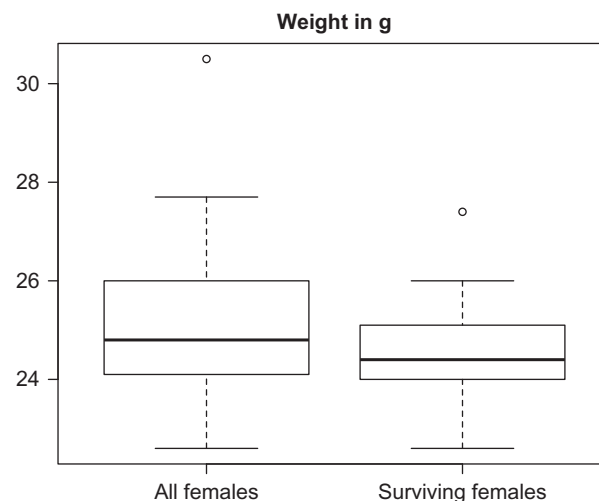
Variance of humerus length in all males was 0.000540 in.<sup>2</sup>, and variance of humerus length in surviving males is 0.000446 in.<sup>2</sup>. Variance of humerus length for surviving males is about 80% of the variance of all males.

It appears that stabilizing selection is at work, with individuals at the extremes of the distributions having less chance of survival.

8. Use the following line of R code to create a boxplot. Does it support your response to the previous question? What additional information does it provide? By modifying the code, create comparable boxplots for weight and humerus length for both males and females.

```
> boxplot(bumpus$weight[bumpus$sex=="f"], bumpus$weight[bumpus$sex=="f" &
bumpus$survival == TRUE], names=c("all females","surviving females"), main =
"weight in g")
```

It appears that the females at the upper range of weight are less likely to survive (see Figure S10.11).



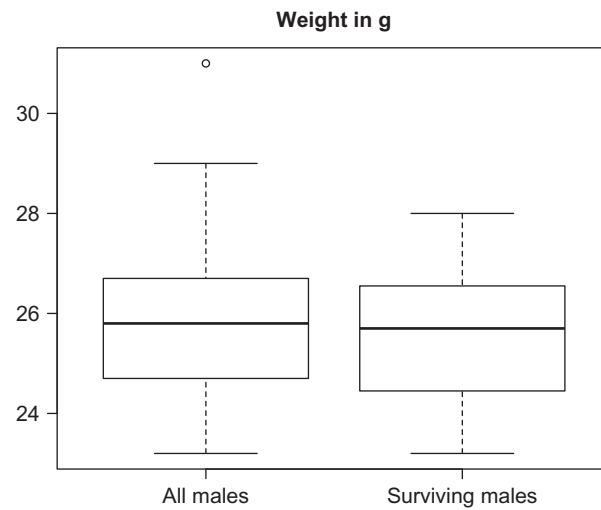
**FIGURE S10.11** Boxplot of weight for all female birds and surviving female birds; answer to Question 8 in Exercise 10.1.

```
> boxplot(bumpus$weight[bumpus$sex=="m"], bumpus$weight[bumpus$sex=="m" &
bumpus$survival == TRUE], names=c("all males","surviving males"),main =
"weight in g")
```

Similarly, the heavier males are less likely to survive (see Figure S10.12).

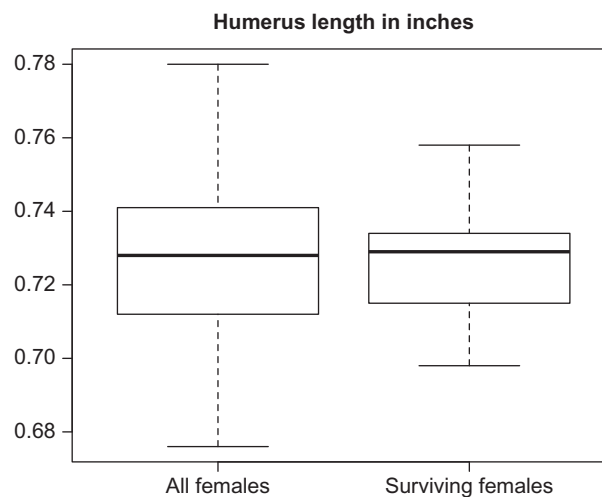
Humerus length:

```
> boxplot(bumpus$humeral[bumpus$sex=="f"], bumpus$humeral[bumpus$sex=="f" &
bumpus$survival == TRUE], names=c("all females","surviving females"),main =
"humerus length in inches")
```



**FIGURE S10.12** Boxplot of weight for all male birds and surviving male birds; answer to Question 8 in Exercise 10.1.

Females from both extremes of the distribution failed to survive (see Figure S10.13).



**FIGURE S10.13** Boxplot of humerus length for all female birds and surviving female birds; answer to Question 8 in Exercise 10.1.

```
> boxplot(bumpus$humerus[bumpus$sex=="m"], bumpus$humerus[bumpus$sex=="m" &
bumpus$survival == TRUE], names=c("all males","surviving males"),main =
"humerus length in inches")
```

The larger males survived at higher rates than the smaller males (see Figure S10.14). □

**Exercise 10.2.** Animals vary in the size of the meals they eat. Some animals, like snakes, eat only every few months but consume a prey item that is very large in comparison to their body size. Other animals, like many bird species, eat several small meals a day. Eating a large meal could mean more energy and resources for producing offspring, but it could also mean decreased mobility and therefore greater susceptibility to predation. Therefore, the amount of food consumed at one time could potentially be a trait correlated with fitness because it might affect both reproduction and survival. In 2010, Pruitt and Krauel investigated these ideas in a study on meal size in wolf spiders. The researchers observed wolf spiders in the lab, and determined how much weight they gained after eating as many crickets as they wanted in a single feeding bout. Then, they individually marked



**FIGURE S10.14** Boxplot of humerus length for all male birds and surviving male birds; answer to Question 8 in Exercise 10.1.

the same spiders, and released them into enclosures in the forest understory. In some enclosures, they removed or excluded all possible predatory insects and birds. In other enclosures, the spiders were exposed to predation. The researchers returned to the field site every other day for 40 days to look for egg cases, which the female carries on her abdomen. If a female had an egg case, the researchers recorded the day it was observed and its weight. The weight of the egg case is highly correlated with the number of eggs and the quality of offspring, and therefore is a good measure of fitness.

The data collected in this study are in the file *wolfspiders.txt*, available on the companion website for this volume. The columns are

treatment:	whether the spiders were exposed to predation or not
population:	the enclosed area where the spiders were kept in the field
startmass:	how much the spiders weighed before eating (g)
change:	how much weight the spiders gained after eating as many crickets as they chose (gluttonous feeding) (g)
threshold:	the satiation threshold, measured as the percent weight gain following gluttonous feeding
numattacked:	the number of crickets a spider attacked during gluttonous feeding
eggcase:	whether or not a spider produced an egg case
dayeggcase:	the first day that an egg case was observed
masseggcase:	the mass of the egg case made by the spider (g)

NA indicates missing data

In R, to simplify subscripting in the exercises, you may want to create a logical vector to distinguish the safe spiders from spiders under predation, as follows:

```
> safe=(wolfspiders$treatment=="safe")
```

This vector has the value TRUE for each safe spider and FALSE for each spider under predation.

The logical operator NOT is represented in R by `!`. The values of `!safe` are opposites of the corresponding values of `safe`. Now `wolfspiders$threshold[safe]` is a vector of threshold values for the safe spiders and `wolfspiders$threshold[!safe]` is a vector of threshold values for the spiders under predation.

The R tutorial provides more information about and practice with logical vectors. □



## QUESTIONS WITH SOLUTIONS EXERCISE 10.2

Using the dataset, do the following data analysis steps:

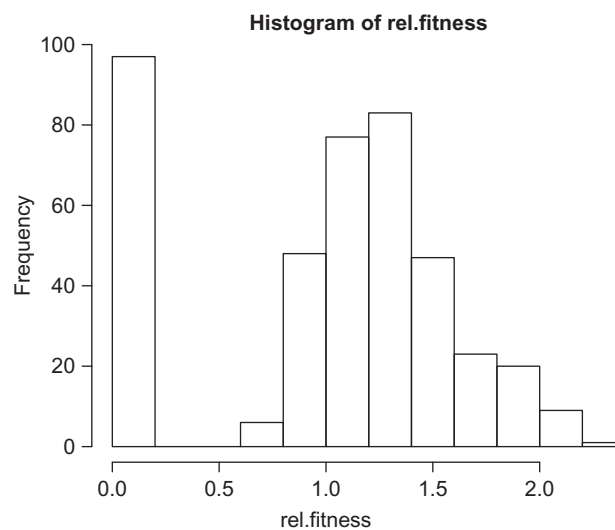
1. Calculate relative fitness for each individual in the dataset, using egg case mass as a measure of fitness.

```
> wolfspiders=read.table(file=file.choose(),header=TRUE)
> masseggcase=wolfspiders$masseggcase
> rel.fitness=masseggcase/mean(masseggcase)
```

2. Make a histogram of relative fitness. What is the shape of the distribution?

```
> hist(rel.fitness)
```

The distribution of nonzero values has a mound-shaped, slightly skewed distribution (Figure S10.15). The histogram also has a tall bar representing the 0 values.



**FIGURE S10.15** Histogram of relative fitness for wolf spiders; answer to Question 2 in Exercise 10.2.

3. Make separate histograms of relative fitness for the spiders that were exposed to predation and those that were safe. Does it look like the risk of predation affected fitness?

```
> hist(rel.fitness[wolfspiders$treatment=="predators"])
> hist(rel.fitness[wolfspiders$treatment=="safe"])
```

There were many more spiders with fitness equal to 0 (and without egg cases) among the spiders exposed to predators (Figure S10.16) than among the safe spiders (Figure S10.17). As a result, mean relative fitness was lower for spiders exposed to predators.

```
> mean(rel.fitness[wolfspiders$treatment=="predators"])
[1] 0.7777175
> mean(rel.fitness[wolfspiders$treatment=="safe"])
[1] 1.221203
```

4. Make a histogram of the satiation threshold. What is the shape of the distribution?

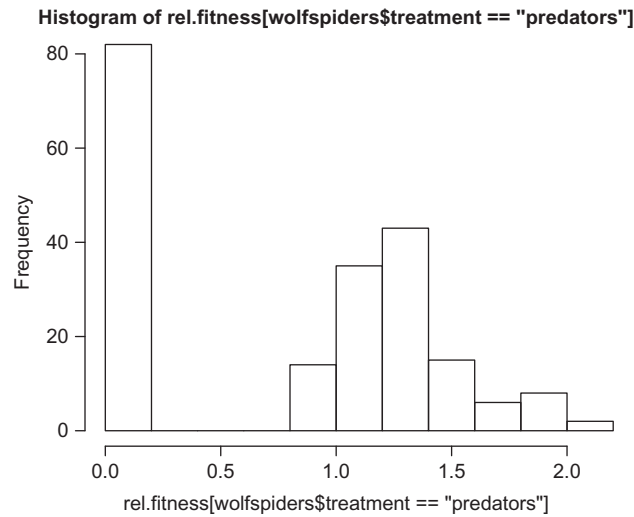
```
> hist(wolfspiders$threshold)
```

The shape is approximately normal (Figure S10.18).

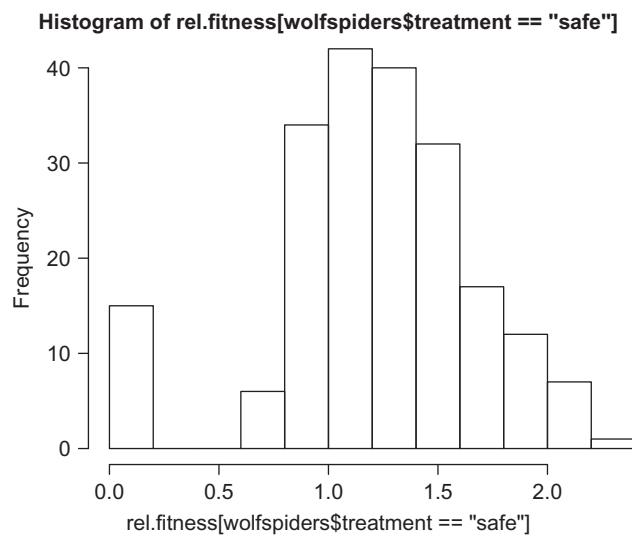
5. What is the covariance between relative fitness and satiation threshold? Do you think this trait is under selection?

```
> cov(wolfspiders$threshold, rel.fitness)
[1] -0.004129252
```

The covariance is very small,  $-0.0041$ , and trait does not appear to be under selection.



**FIGURE S10.16** Histogram of relative fitness for wolf spiders exposed to predation; answer to Question 3 in Exercise 10.2.

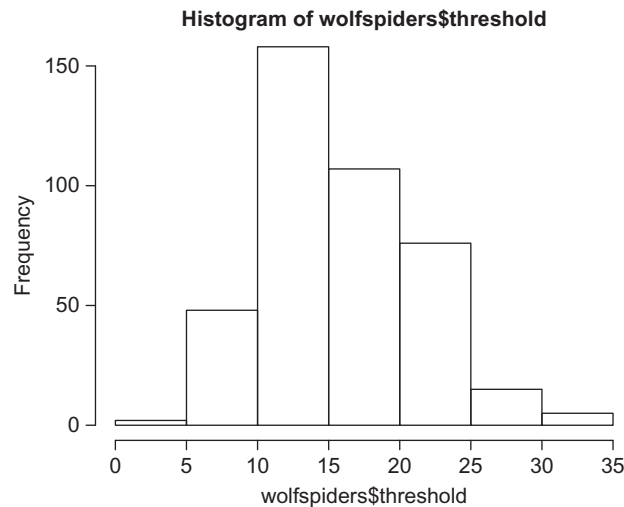


**FIGURE S10.17** Histogram of relative fitness for wolf spiders safe from predation; answer to Question 3 in Exercise 10.2.

6. Now, calculate the covariance between relative fitness and satiation threshold separately for safe spiders and spiders at risk of predation. How does the relationship differ between the two groups? What does this tell you about selection?

```
> safe=(wolfspiders$treatment=="safe")
> cov(wolfspiders$threshold[safe], rel.fitness[safe])
[1] 0.8312778
> cov(wolfspiders$threshold[!safe], rel.fitness[!safe])
[1] -0.3922264
```

Selection is operating in both groups, in opposite directions. Among the safe spiders, relative fitness and satiation threshold have a positive covariance of 0.83, suggesting that spiders that are more gluttonous have larger egg cases and therefore more offspring. Among spiders subject to predation, the covariance is  $-0.39$ . It is negative, suggesting that when predators are present, spiders that are more gluttonous are less likely to make an egg case and/or make smaller egg cases.



**FIGURE S10.18** Histogram of satiation threshold in wolf spiders; answer to Question 4 in Exercise 10.2.

**Exercise 10.2.** Exercise 10.2 demonstrated that there is variability in a wild population of wolf spiders for satiation threshold. It also established a relationship between satiation threshold and gluttony for wolf spiders that aren't threatened by predators. Both of these are required for natural selection to change satiation threshold over time. Natural selection also has a third requirement: a genetic basis for the trait. Pruitt and Krauel also collected data in the lab on satiation thresholds in mothers and their offspring. The mother wolf spiders carry their egg case until it hatches, and then the spiderlings climb up her legs and ride on her back and abdomen until they are big enough to live independently. Therefore, it was easy for the researchers to identify mother-offspring pairs. They measured satiation threshold in two offspring from each mother. The data are in the file *spiderheritability.txt*. The first column is labeled “dam” and is the satiation threshold value for the mother. The second column is labeled “offspring” and is the mean satiation threshold for two adult female offspring. □

### QUESTIONS WITH SOLUTIONS EXERCISE 10.3

1. Make a scatterplot with maternal satiation threshold on the x-axis and average offspring satiation threshold on the y-axis. Do the mothers and offspring appear to have similar variability? What is the relationship between mothers and offspring?

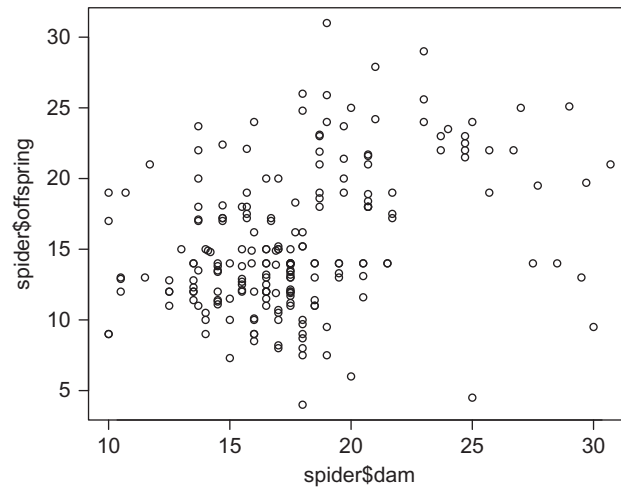
```
> spider=read.table(file=file.choose(), header = TRUE)
> plot(spider$dam,spider$offspring)
> sd(spider$dam)
[1] 4.042362
> sd(spider$offspring)
[1] 5.01502
```

There is a positive association between the satiation threshold of mothers and offspring (Figure S10.19). The offspring are a little more variable: the standard deviation of satiation threshold in the mothers was 4.04%, and it was 5.02% in the offspring.

2. Use regression analysis to determine whether the relationship between the mother's satiation threshold and that of her offspring is statistically significant. Report the estimated equation for the line, the  $p$ -value for the regression, and the correlation.

```
> spider.regress=lm(formula=spider$offspring ~ spider$dam)
> summary(spider.regress)
```

```
Call:
lm(formula = spider$offspring ~ spider$dam)
```



**FIGURE S10.19** Scatterplot of satiation threshold in mothers and their offspring; answer to Question 1 in Exercise 10.2.

Residuals:

Min	1Q	Median	3Q	Max
-14.1902	-2.9102	-0.3823	3.0734	14.9828

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.55265	1.47115	5.134	6.60e-07 ***
spider\$dam	0.44550	0.08107	5.495	1.15e-07 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.692 on 204 degrees of freedom

Multiple R-squared: 0.1289, Adjusted R-squared: 0.1247

F-statistic: 30.2 on 1 and 204 DF, p-value: 1.155e-07

The least squares regression line is  $\hat{y} = 7.55265 + 0.44550x$ . The  $p$ -value for the coefficient of spider\$dam is approximately 0; the linear relationship is highly significant.

```
> cor(spider$dam, spider$offspring)
[1] 0.359096
```

Correlation between satiation threshold of mother and offspring is 0.359.

3. Use the regression coefficient to determine heritability for satiation threshold.

Because the regression includes only the trait value from the mother, we multiply the regression coefficient by two to obtain heritability.

$$h^2 = 0.89$$

4. Estimate the heritability by dividing the covariance by the maternal variance. Do you get the same value?

```
> cov(spider$dam, spider$offspring)/var(spider$dam)
[1] 0.4455004
```

The covariance divided by the maternal variance is 0.446. Multiply by two to compensate for the fact that only one parent was used in the analysis.

$$h^2 = 0.89$$

This agrees with the first estimate of heritability.

**Exercise 10.3.** Theodore Garland and colleagues selected on voluntary wheel-running behavior for several generations in a laboratory strain of house mice. The mice were permitted to run as much as they liked on a wheel in their cage, and a sensor in the wheel counted the number of wheel revolutions per day. Each generation, they selected the 10 families that ran the most, and bred between families to obtain the next generation. Four different selection lines were established. Data for the first generation can be found in the `wheelrunning.txt` file on the companion website for this volume. The file contains the following variables:

<code>animal:</code>	a unique ID number for each mouse
<code>sex:</code>	whether the mouse was male (1) or female (0)
<code>gen:</code>	the generation. Generation 0 was the parental generation before selection, and generation 1 is the product of selection imposed in generation 0.
<code>line:</code>	the name for each line
<code>revol:</code>	the average number of revolutions per day that the mouse ran in the wheel
<code>pups:</code>	the number of pups each individual had
<code>bred:</code>	whether the animal was bred (1) or not bred (0)

*R note:* The function `tapply()` provides an efficient way to calculate means for all 4 lines of mice with one line of code. The generic form of `tapply` that we will use is `tapply(datavector, groupvector, statistic)`. To simplify notation, assign to the name `line` the value of `wheelrunning$line`, let `revol = wheelrunning$revol`, and let `gen = wheelrunning$gen`. If we wanted the mean number of revolutions for each line, regardless of generation, we would use

```
>tapply(revol, line, mean).
```

In order to specify a particular generation, say generation 0, we would use

```
>tapply(revol[gen == 0], line[gen == 0], mean).
```

□

## QUESTIONS WITH SOLUTIONS EXERCISES 10.4

1. Calculate the average number of revolutions for generation 0 for each line.

```
> wheelrunning=read.table(file=file.choose(), header = TRUE)
> line=wheelrunning$line
> gen=wheelrunning$gen
> revol=wheelrunning$revol
```

```
> gen0means = tapply(revol[gen==0],line[gen==0],mean)
> gen0means
      3      6      7      8
4549.295 4394.272 3850.034 4949.565
```

2. Calculate the average number of revolutions for generation 1 for each line.

```
> gen1means=tapply(revol[gen==1],line[gen==1],mean)
> gen1means
      3      6      7      8
5351.453 5521.902 5165.861 5577.906
```

3. For each line, calculate  $R$ , the response to selection. Did wheel running increase from one generation to the next?

```
> response.to.selection=gen1means - gen0means
> response.to.selection
      3      6      7      8
802.1576 1127.6297 1315.8271 628.3408
```

Mean wheel running increased in every line.

4. Calculate the relative fitness for each mouse in line 6 and generation 0, then calculate the covariance between relative fitness and number of revolutions for line 6, generation 0. This value is the selection differential,  $S$ , for line 6.

```
> tapply(wheelrunning$pups[gen==0], wheelrunning$line[gen==0], mean)
      3      6      7      8
2.700000 2.622222 2.490196 1.962617
> relfitness.line6=wheelrunning$pups[gen==0 & line==6]/2.62222
> cov(relfitness.line6,wheelrunning$revol[gen==0 & line ==6])
[1] 1831.597
```

The selection differential,  $S$ , for line 6 is 1832 revolutions per day.

The correlation between fitness and number of revolutions can be calculated using the following command:

```
> cor(relfitness.line6,wheelrunning$revol[gen==0 & line ==6])
[1] 0.5348557
```

5. Using the breeder's equation, calculate the realized heritability for wheel running behavior in selection line 6. For line 6,  $h^2 = R/S = 1127.6297/1831.597 = 0.616$

**Exercise 10.4.** White settlers have introduced many new plant species to North America over the last 400 years. Some of those plants grow very well in their new habitat, and outcompete native plants and disrupt ecosystems by growing aggressively. Reed canary grass was introduced from Europe around 1850, and has become a problematic invasive species in many parts of the United States. Once the plant establishes from seed, it sends out tillers that branch and root, increasing the size and number of stems on the plant. Calsbeek and colleagues measured several traits in plants from four populations of reed canary grass in an attempt to identify how the relationships between traits might change during an invasion. They measured traits that help a plant grow more aggressively, including how long it takes for the plant to emerge from an underground stem, how tall a plant can grow, and the number of tillers it has. They collected seeds from four source populations: one in the Czech Republic and one at a similar latitude in Vermont, one in France, and another comparison population in North Carolina. By using these four populations, they could compare differences between the native and invasive ranges of the species, as well as differences between populations at different latitudes. They grew and measured all plants in the same greenhouse. Data are in the file *reedcanary.txt* on the companion website for this volume.

The columns in the data file are

ID:	the identification number for each plant
range:	whether the population was in the native or invasive range for the species
location:	the source population for the seeds (Czech, France, NCarol, or Vermont)
emertime:	the number of days it took for the plant to emerge above the soil from an underground stem
maxht:	the maximum height of a stem at 78 days (cm)
tillernum:	the number of tillers on the plant at 78 days
abovewt:	the aboveground dry-weight biomass of the plant (g)
rootwt:	the belowground dry-weight biomass of the plant (g)
totalwt:	the total dry-weight biomass of the plant (g)

*R note:* As in some previous exercises, references to variables will be greatly simplified by the use of logical vectors. For example, create a logical string that is TRUE for plants located in France and FALSE for other locations:

```
> France=reedcanary$location=="France"
```

Now, create a matrix of data for plants from France with columns for emergence time, maximum height, and tiller number with the assignment

```
> France.data=cbind(reedcanary$emertime[France],reedcanary$maxht[France],
reedcanary$tillernum[France])
```

Refer to the R tutorial for help with creating matrices with `cbind()`, and help with the `cov()` and `cor()` functions. □

## QUESTIONS WITH SOLUTIONS EXERCISE 10.5

1. Calculate the phenotypic variance-covariance matrix and the correlation matrix for emergence time, maximum height, and tiller number in plants from the population in France.

```
> cov(France.data)
      [,1]      [,2]      [,3]
[1,]  5.762195 -18.670732 -5.517073
[2,] -18.670732 325.040244 -3.710976
[3,] -5.517073 -3.710976 30.593902
> cor(France.data)
      [,1]      [,2]      [,3]
[1,]  1.0000000 -0.43141853 -0.41552516
[2,] -0.4314185  1.00000000 -0.03721364
[3,] -0.4155252 -0.03721364  1.00000000
```

2. Calculate the phenotypic variance-covariance matrix and the correlation matrix for the same traits in plants from the population in North Carolina.

```
> NCarol=reedcanary$location=="NCarol"
> NCarol.data=cbind(reedcanary$emertime[NCarol], reedcanary$maxht[NCarol],
reedcanary$tillernum[NCarol])
> dim(NCarol.data)
[1] 95 3
> cov(NCarol.data)
      [,1]      [,2]      [,3]
[1,]  4.996641  6.444569 -5.271557
[2,]  6.444569 350.555599 -53.695577
[3,] -5.271557 -53.695577 62.244569
> cor(NCarol.data)
      [,1]      [,2]      [,3]
[1,]  1.0000000  0.1539843 -0.2989159
[2,]  0.1539843  1.0000000 -0.3635041
[3,] -0.2989159 -0.3635041  1.0000000
```

3. Within each matrix, which traits covary with each other? Can you give a biological explanation for the correlations?

In the plants located in France, there is a negative correlation between emergence time and height and between emergence time and number of tillers. This is reasonable because the later the plant emerges, the less time it has for growth. There is essentially no correlation between height and number of tillers. This could occur if there is no particular constraint to simultaneously being taller and having more tillers, and if taller plants aren't more likely to have more tillers.

In the North Carolina plants, there are negative correlations between number of tillers and time of emergence, and between number of tillers and maximum height, and very slight positive correlation between emergence time and maximum height. A biological explanation for the negative correlation between number of tillers and height might be that the plant has limited environmental resources so that it can grow in one direction or the other but not both.

4. Compare the variance-covariance matrices for the two populations. In what ways are they similar and different?

The variances of emergence time and maximum height are quite similar for the two locations but variance of number of tillers is about twice as large for North Carolina plants compared to French plants. Covariance between emergence time and maximum height is negative for the plants in France but positive for the plants in North Carolina. Covariance between emergence time and number of tillers is close to equal for the two locations. Covariance between maximum height and number of tillers is much larger for the North Carolina plants than the plants in France.

**Exercise 10.5.** Traits that are associated with achieving successful matings are a special case of natural selection called sexual selection. Often, traits of this nature are under relatively strong selection. In a study on introduced populations of guppies in Australia, Lindholm and colleagues measured relative male fitness and a variety of traits that female guppies find attractive. Females are generally attracted to male guppies that have larger areas of orange coloration and a greater number of black spots. The researchers collected guppies from the wild, and then measured the traits of interest and male fitness in controlled laboratory conditions. Data from one of the study populations, Mena Creek, are in the file *guppydata.txt*.

Columns in the file are

ID:	male ID number
offspring:	number of offspring sired by a given male
relfit:	relative fitness = number of offspring sired by a given male/mean fitness (average number of offspring per male within a trial)
body:	body area (mm <sup>2</sup> )
tail:	tail area (mm <sup>2</sup> )
black:	area of black color (mm <sup>2</sup> )
fuzzbl:	area of fuzzy black coloration (mm <sup>2</sup> )
orange:	area of carotenoid coloration (mm <sup>2</sup> )
totalirid:	area of combined iridescent coloration (mm <sup>2</sup> )

*R note:* Refer to the R tutorial for help with inverting matrices and solving simple matrix equations. □

## QUESTIONS WITH SOLUTIONS EXERCISE 10.6

1. Calculate the phenotypic variance-covariance matrix for the area of black color, the area of orange color, and the area of combined iridescent coloration.

```
> guppy=read.table(file=file.choose(), header = TRUE)

> cov(cbind(guppy$black, guppy$orange, guppy$totalirid))
      [,1]      [,2]      [,3]
[1,] 0.695541138 -0.002479101  1.2161087
[2,] -0.002479101  5.498873545  0.6103672
[3,] 1.216108730  0.610367196 14.9086365
```

2. Calculate the phenotypic correlation matrix for the same traits; which traits are correlated?

```
> cor(cbind(guppy$black, guppy$orange, guppy$totalirid))
      [,1]      [,2]      [,3]
[1,] 1.00000000 -0.00126764  0.37765221
[2,] -0.00126764  1.00000000  0.06741171
[3,] 0.37765221  0.06741171  1.00000000
```

None of the traits are strongly correlated. There is a positive correlation between iridescent coloration area and area of black.



3. Using the covariance between relative fitness and each trait, calculate  $\mathbf{S}$  for the area of black color, the area of orange color, and the area of combined iridescent coloration.

```
> cov(guppy$relfit, guppy$black)
[1] -0.2026249
> cov(guppy$relfit, guppy$orange)
[1] -0.07742458
> cov(guppy$relfit, guppy$totalirid)
[1] -1.134233
```

So

$$\mathbf{S} = \begin{bmatrix} -0.203 \\ -0.077 \\ -1.134 \end{bmatrix}$$

4. Using Eq. (10.16), calculate  $\beta$  for the area of black color, the area of orange color, and the area of combined iridescent coloration.

```
> guppy.matrix = cov(cbind(guppy$black, guppy$orange, guppy$totalirid))
> guppy.matrix
      [,1]      [,2]      [,3]
[1,] 0.695541138 -0.002479101 1.2161087
[2,] -0.002479101 5.498873545 0.6103672
[3,] 1.216108730 0.610367196 14.9086365
> guppyS=c(-0.2026249, -0.0774258, -1.134233)
> guppybeta=solve(guppy.matrix, guppyS)
> guppybeta
[1] -0.18528460 -0.00743056 -0.06066091
```

So

$$\beta = \begin{bmatrix} -0.185 \\ -0.007 \\ -0.061 \end{bmatrix}$$

5. How do the values in the  $\mathbf{S}$  and  $\beta$  vectors compare? Was accounting for covariance between traits in our selection gradient important?

For black color, the values in  $\mathbf{S}$  and  $\beta$  are quite similar.

For orange color, both values are small but the  $\beta$  value is about one-tenth the value in  $\mathbf{S}$ .

For iridescent coloration, the  $\beta$  value is much smaller than the  $\mathbf{S}$  value.

In  $\mathbf{S}$ , selection appears to be strongest on iridescent coloration. After accounting for covariance, selection is greatest on black color, less on iridescent color and least on orange color.

**Exercise 10.6.** In the study examined in Exercise 10.4, the researchers also estimated  $\mathbf{G}$  for the study populations from France and North Carolina (see Table 10.1).  $\square$

## QUESTIONS WITH SOLUTIONS EXERCISE 10.7

1. Using these estimates and Eq. (10.17), calculate the outcome of selection on emergence time, maximum height, and tiller number in these two populations for the following selection scenarios. Because comparisons are relative, the magnitude of values in  $\beta$  is not important; values between 1 and 10 will give answers that are intuitive to interpret.

**TABLE 10.1** G Matrix for Emergence Time, Height, and Tiller Number in Reed Canary Grass Plants from France and North Carolina

Location	Trait	Emergence Time	Height	Tiller Number
France	Emergence time	0	0	0
	Height	0	149.487902	-7.690694
	Tiller number	0	-7.690694	10.269224
North Carolina	Emergence time	0	0	0
	Height	0	171.80604	-35.41625
	Tiller number	0	-35.41625	24.71871

From Calsbeek et al. [24]

Equal selection on each trait (values of  $\beta$  are equal)

Selection on height only

Selection on tiller number only

Setting up the matrices:

```
> France=reedcanary$location=="France"
> NCarol=reedcanary$location=="NCarol"
> France.data=cbind(reedcanary$emertime[France],reedcanary$maxht[France],
reedcanary$tillernum[France])
> NCarol.data=cbind(reedcanary$emertime[NCarol],reedcanary$maxht[NCarol],
reedcanary$tillernum[NCarol])
> G.France=cbind(c(0,0,0), c(0, 149.487902, -7.690694), c(0, -7.690694,
10.269224))
> G.France
      [,1]      [,2]      [,3]
[1,]    0    0.000000  0.000000
[2,]    0 149.487902 -7.690694
[3,]    0 -7.690694 10.269224
> G.NCarol=cbind(c(0,0,0),c(0, 171.80604, -35.41625),c(0, -35.41625,
24.71871))
> G.NCarol
      [,1]      [,2]      [,3]
[1,]    0    0.00000  0.00000
[2,]    0 171.80604 -35.41625
[3,]    0 -35.41625  24.71871
```

Equal selection on each trait:

```
> beta=c(3,3,3)
> G.France %*% beta
      [,1]
[1,]  0.000000
[2,] 425.39162
[3,]  7.73559
```

So for France

$$\Delta \bar{z} = \begin{bmatrix} 0 \\ 425.4 \\ 7.736 \end{bmatrix}$$

```
> G.NCarol %*% beta
      [,1]
[1,]  0.000000
[2,] 409.16937
[3,] -32.09262
```

So for North Carolina

$$\Delta \bar{z} = \begin{bmatrix} 0 \\ 409.2 \\ -32.09 \end{bmatrix}$$

The outcome of selection is 0 on emergence time for both populations because, according to the **G** matrix, emergence time has no heritability and is not correlated with height or tiller number. The response to selection on maximum height is similar in both populations, and is a relatively large increase. The response to selection on number of tillers is positive in the France population but negative in the North Carolina population. Selection on height only:

```
> beta=c(0, 3, 0)
> G.France %*% beta
      [,1]
[1,]  0.000000
[2,] 448.46371
[3,] -23.07208
```

So for France,

$$\Delta \bar{z} = \begin{bmatrix} 0 \\ 448.5 \\ -23.07 \end{bmatrix}$$

```
> G.NCarol %*% beta
      [,1]
[1,]  0.0000
[2,] 515.4181
[3,] -106.2487
```

So for North Carolina,

$$\Delta \bar{z} = \begin{bmatrix} 0 \\ 515.4 \\ -106.2 \end{bmatrix}$$

In both populations, the response is similar to the scenario with equal selection, but the magnitude of response for height and tiller number is larger in the North Carolina population than when selection is equal on the three traits. Note that the number of tillers is expected to change, despite the fact that there was no direct selection on tiller number.

Selection on tiller number only:

```
> beta=c(0,0,3)
> G.France %*% beta
      [,1]
[1,]
```

```
[1,] 0.000000
[2,] -23.07208
[3,] 30.80767
```

So for France,

$$\Delta \bar{z} = \begin{bmatrix} 0 \\ -23.07 \\ 30.81 \end{bmatrix}$$

```
> G.NCarol %*% beta
      [,1]
[1,] 0.000000
[2,] -106.24875
[3,] 74.15613
```

So for North Carolina,

$$\Delta \bar{z} = \begin{bmatrix} 0 \\ -106.2 \\ 74.16 \end{bmatrix}$$

In this scenario, in the population from France, maximum height decreases and tiller number increases. The responses are the same direction in the North Carolina population, with greater magnitude.

2. How do the outcomes of selection differ for each scenario within each population?  
Quantitatively, results will depend on size of the  $\beta$  coefficients, but the direction and relative size of the response will be consistent with those in Table S10.3.

**TABLE S10.3** Response to Different Selection Scenarios in Two Populations of Reed Canary Grass; Answer for Question 2 in Exercise 10.6

	Equal selection	Selection on height	Selection on tiller number
<b>French population</b>			
Response in emergence time	0	0	0
Response in maximum height	425 (large increase)	448 (large increase)	−23 (small decrease)
Response in tiller number	7.7 (small increase)	−23 (small decrease)	31 (small increase)
<b>NC population</b>			
Response in emergence time	0	0	0
Response in maximum height	409 (large increase)	515 (large increase)	−106 (moderate decrease)
Response in tiller number	−32 (small decrease)	−106 (moderate decrease)	74 (moderate increase)

3. Do the two populations differ in the way they respond to the same selection scenarios?  
Yes, but only in magnitude, not in direction of the change.