ゲノム・オミックス統計解析　―回答―

P.10 Exercises

1. Read in the file femaleMiceWeights.csv and report the body weight of the mouse in the

exact name of the column containing the weights.

2. The [ and ] symbols can be used to extract specific rows and specific columns of the table.What is the entry in the 12th row and second column?

* 26.25

3. You should have learned how to use the $ character to extract a column from a table and return it as a vector. Use $ to extract the weight column and report the weight of the mouse in the 11th row.

* 26.91

4. The length function returns the number of elements in a vector. How many mice are included in our dataset?

* 24

5. To create a vector with the numbers 3 to 7, we can use seq(3,7) or, because they are consecutive, 3:7. View the data and determine what rows are associated with the high fat or hf diet. Then use the mean function to compute the average weight of these mice.

* 26.83417

6. One of the functions we will be using often is sample. Read the help file for sample using ?sample. Now take a random sample of size 1 from the numbers 13 to 24 and report back the weight of the mouse represented by that row. Make sure to type set.seed(1) to ensure that everybody gets the same answer.

* 25.34

P.30 Exercises

1. What is the average of these weights?

* 23.89338

2. After setting the seed at 1, set.seed(1) take a random sample of size 5. What is the absolute value (use abs) of the difference between the average of the sample and the average of all the values?

* Average of the sample : 24.164
* Absolute value of the difference between two : 0.2706222

3. After setting the seed at 5, set.seed(5) take a random sample of size 5. What is the absolute value of the difference between the average of the sample and the average of all the values?

* Average of the sample : 22.46
* Absolute value of the difference between two : 1.433378

4. Why are the answers from 2 and 3 different?

• A) Because we made a coding mistake.

• B) Because the average of the x is random.

• C) Because the average of the samples is a random variable.

• D) All of the above.

* C)

5. Set the seed at 1, then using a for-loop take a random sample of 5 mice 1,000 times. Save these averages. What percent of these 1,000 averages are more than 1 ounce away from the average of x ?

* 3.06%

6. We are now going to increase the number of times we redo the sample from 1,000 to 10,000. Set the seed at 1, then using a for-loop take a random sample of 5 mice 10,000 times. Save these averages. What percent of these 10,000 averages are more than 1 ounce away from the average of x ?

* 31.95%

7. Note that the answers to 4 and 5 barely changed. This is expected. The way we think about the random value distributions is as the distribution of the list of values obtained if we repeated the experiment an infinite number of times. On a computer, we can’t perform an infinite number of iterations so instead, for our examples, we consider 1,000 to be large enough, thus 10,000 is as well. Now if instead we change the sample size, then we change the random variable and thus its distribution.

Set the seed at 1, then using a for-loop take a random sample of 50 mice 1,000 times. Save these averages. What percent of these 1,000 averages are more than 1 ounce away from the average of x ?

* 0%hist

8. Use a histogram to “look” at the distribution of averages we get with a sample size of 5 and a sample size of 50. How would you say they differ?

• A) They are actually the same.

• B) They both look roughly normal, but with a sample size of 50 the spread is smaller.

• C) They both look roughly normal, but with a sample size of 50 the spread is larger.

• D) The second distribution does not look normal at all.

* B)

9. For the last set of averages, the ones obtained from a sample size of 50, what percent are between 23 and 25?

* 19.8%

10. Now ask the same question of a normal distribution with average 23.9 and standard

deviation 0.43.

* 16.92%