In-class Rmarkdown practice

Andriana Manousidaki

2023-09-11

Using Rmarkdown for in class assignments and homework

From now and on we will be writing the solution to inclass assignments and homework to Rmarkdown.

For a summary of the syntax rules go to this cheatsheet 1.

Aim of assignment:

In the next, tasks we are going to:

- create a simulated data set of variables drawn either from discrete or continuous distributions,
- then we will use this data set to practice on ggplot2 visualizations.

The simulated data set

The simulated data set will have variables that will express the *vitamin D concentration* in nmol/L of a simulated sample of *60 patients* that have normal and very low levels of vitamin D.

According to NIH:

"Because you get vitamin D from food, sunshine, and dietary supplements, one way to know if you're getting enough is a blood test that measures the amount of vitamin D in your blood. In the blood, a form of vitamin D known as 25-hydroxyvitamin D is measured in either nanomoles per liter (nmol/L) or nanograms per milliliter (ng/mL). One nmol/L is equal to 0.4 ng/mL. So, for example, 50 nmol/L is the same as 20 ng/mL.

- Levels of 50 nmol/L (20 ng/mL) or above are adequate for most people for bone and overall health.
- Levels below 30 nmol/L (12 ng/mL) are too low and might weaken your bones and affect your health.

Question 1

Create the variable, *D_group*, that will express in which category a patient belongs to: normal levels or low levels of vitamin D. Hint draw from the binomial distribution and take into consideration that the prevalence of vitamin D deficiency in the US is about 42%.

Question 2

Create the variable, *D_level*, that will express the level of vitamin D based on the group a patient belongs to. Hint: for Normal levels draw from the normal distribution with mean 60 and sd = 3,for low levels draw from the normal distribution with mean= 20 and sd = 3.

Question 3

Create the variable, *Sex*, that will express if a patient is Male or Female. Hint draw from the binomial distribution and assume that the sample has equal number of participants from each of the categories.

Question 4

Create the variable, *Age*, of the subject. Assume that we have sampled uniformly females of ages between 20 and 50 and male of ages between 20 and 50.

Question 5

Create the data frame, vitamin_D, with columns the variables created in question 1 through 4.

Question 6

Use ggplot2 to create side-by-side boxplots of the *D_level* of the different *D_groups*

Use ggplot2 to create side-by-side boxplots of the *D_level* of the different *Sex*

What do you observe regarding the vitamin D levels among the Sex and D_groups we created? Are they as you expected based on the way they are constructed?

Question 7

Use ggplot2 to create the histogram of *D_level*. Find the 50th quantile of the distribution used for the normal vitamin D levels and create a blue line that indicates it on the histogram. Find the 50th quantile of the distribution used for the low vitamin D levels and create a red line that indicates it on the histogram.