Week 2 DS Practice Problems

SIBDS @ Columbia

08 June, 2023

1. Create a new R project and name it something related to `Week 2 DS Recitation`.
2. Create a new R script within that project called `week2_DS_distribution`.

    a. Using the code at the end of Unit 3, repeat questions 5(b) to 5(f) of Recitation Unit 3, but using R this time.

    b. Using the code at the end of Unit 3, repeat questions 6 of Recitation Unit 3, but using R this time.

3. Create a new R script within that project called `week2_DS_screening`.

    a. Use the data in Recitation Unit 3, Problem 3. Write code to compute the PPV and NPV of the new screening test when the prevalence of disease is 0.10. Write your code in such a way that you can just update the values of sensitivity, specificity, and prevalence to get the revised PPV and NPV.

    b. Check that your code is flexible, by recomputing the PPV and NPV when the prevalence is 0.20 instead.

4. Create a new R script within that project called `week2_DS_plotting`.

    a. Use the code below to download the a package containing the `penguins` dataset. (You only need to run this command once to install the package, and you can do so directly in the console)

```
install.packages("palmerpenguins")
```

    b. Next, use the following code to load the penguins dataset:

```
data("penguins", package = "palmerpenguins")
```

    c. Make a scatterplot of `flipper_length_mm` (y) vs `bill_length_mm` (x); color points using the `species` variable (adding `color = ...` inside of `aes` in your `ggplot` code should help).

```
# The code you need for this problem will be something like:
ggplot(penguins, aes(x = ..., y = ..., color = species)) +
  geom_point()
```

    d. Export your first scatterplot to your project directory using `ggsave`.