Homework 11

due Apr 20, 2021

For all questions below, consider the `crabs` data set that we used in class:

```
crabs <- read.table(file = "crabs.tsv", header = T, sep = "\t")
dim(crabs); names(crabs)
```

```
## [1] 173   5
```

```
## [1] "color"  "spine"  "width"  "satell" "weight"
```

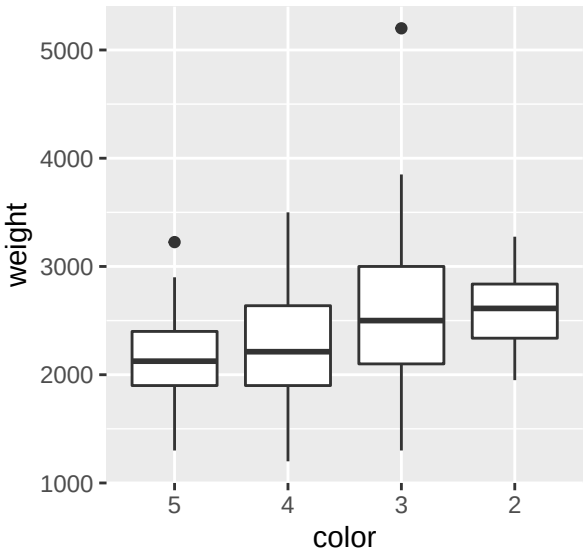
Consider the following three models for `satell`:

$$\textbf{Model 1: } \texttt{satell} = \beta_0 + \beta_1 \texttt{width},$$
$$\textbf{Model 2: } \texttt{satell} = \beta_0 + \beta_1 \texttt{weight},$$
$$\textbf{Model 3: } \texttt{satell} = \beta_0 + \beta_1 \texttt{width} + \beta_2 \texttt{weight}.$$

1. Compute R^2 for each model.

2. Which model is the best? Justify your answer.

3. Create the following box plots. Note the order of the labels in the x-axis. *Hint*: change `color` to a factor variable with levels 5, 4, 3, 2, in this order.



4. Carry out an appropriate regression between `weight` (response) and `color` (explanatory). How does the change in color from `5` to `3` affect `weight`?

5. How does the change in color from `2` to `4` affect `weight`? *Hint*: you might want to change `color` to a factor variable with levels 2, 3, 4, 5, in this order.