

Numerical Data Applications

Professor Bradley Warner

13 May, 2020

Exercises

Create an Rmd file for the work including headers, file creation data, and explanation of your work. Make sure your plots have a title and the axes are labeled.

1. Mammals exploratory

Data were collected on 39 species of mammals distributed over 13 orders. The data is in the `openintro` package as `mammals`

- Using `help`, report the units for the variable `BrainWt`.
- Using `inspect` how many variables are numeric?
- What type of variable is `Danger`?
- Create a density plot of `TotalSleep` and describe the distribution.
- Create a boxplot of `LifeSpan` and describe the distribution.
- Report the mean and median life span of a mammal.
- Calculate the summary statistics for `LifeSpan` broken down by `Danger`. What is the standard deviation of life span in danger outcome 5?

2. Mammals life spans

Continue using the `mammals` data set.

- Create side-by-side boxplots for `LifeSpan` broken down by `Exposure`. Note: you will have to change `Exposure` to a `factor()`. Report on any findings.
- What happened to the median and third quartile in exposure group 4?
- Create overlapping density plots. What are the shortcomings of this plot?
- Create a new variable `Exposed` that is a factor with level `Low` if exposure is 1 or 2 and `High` otherwise.
- Repeat part c with the new variable. Explain what you see in the plot.

3. Mammals life spans continued

- Create a scatterplot of life span versus length of gestation.

- b. What type of an association is apparent between life span and length of gestation?
- c. What type of an association would you expect to see if the axes of the plot were reversed, i.e. if we plotted length of gestation versus life span?
- d. Create the new scatterplot suggested in c.
- e. Are life span and length of gestation independent? Explain your reasoning.