Correlation Tutorial

FMU Biology Department

January 3, 2021

Pearson's Correlation in R

This documents includes code for creating a data object in R, displaying summary statistics, creating an informative plot, and running Pearson's correlation analysis.

Background

You have recently joined the lab of Pokemon Professor to study the relationship between nutrition and Pokemon fighting ability. This is an NSF funded project with the goal of improving Pokemon fighting abilities to create more dynamic battles at the Pokemon Gym. Your model organism is Pikachu, an electric Pokemon. Based on laboratory studies, Pokemon Professor has demonstrated that the electroplaques (cells that generate electricity) of Pikachu use sodium and potassium ions (charged atoms) to generate the electricity. The process is very similar to the way electric eels *Electrophorus electricus* generate electricity. This mechanism suggests that increased blood concentrations of sodium or potassium would result in an increase in electrical output from Pikachu's tail. Thus your research question is "Does increased sodium in the diet of Pikachu increase their electrical output?". To test this relationship, you have conducted a study that compares the output of electricity to the amount of sodium in Pokepuffs (food of Pokemon) fed to Pikachu.

Hypothesis: If the concentration of sodium influences Pikachu electrical output, then increasing sodium in Pikachu's diet will increase the amount of electricity Pikachu produces.

Null Hypothesis: The amount of sodium in Pikachu's diet has no effect on electrical output.

Data

The data frame being created below contain 20 observations (rows) and 2 variables (columns). Each row represents one individual Pikachu. Column one contains the concentration of sodium ("sodium"; milliequivalents per liter) in the treatment and column two contains the electrical output of Pikachu's tail ("elec"; watts).

Important points to note. The function is "data.frame" which creates a data frame object that is being assigned the name "dat3". Everything after "data.frame" are the arguments being passed to the function. The arguments include the column names (sodium and elec) and the data that go in each column.

Enter the following code in R:

```
dat3 = data.frame(sodium = c(95,104,88,97,102,89,104,105,111,88,
                             98,102,96,105,107,105,96,90,110,98),
                  elec = c(190,200,192,202,196,195,200,202,206,180,
                           205,199,193,205,208,210,204,197,206,200))
```

The next line of code will print the data contained in the dat3 object

```
dat3
```

```
##    sodium elec
## 1      95  190
```

```
## 2     104  200
## 3      88  192
## 4      97  202
## 5     102  196
## 6      89  195
## 7     104  200
## 8     105  202
## 9     111  206
## 10     88  180
## 11     98  205
## 12    102  199
## 13     96  193
## 14    105  205
## 15    107  208
## 16    105  210
## 17     96  204
## 18     90  197
## 19    110  206
## 20     98  200
```

Display summary data

Important things to note: "dat3$sodium" and "dat3$watts" references your data frame name (dat3) and the column name (e.g., sodium)

```
#Average and standard deviation of sodium
mean(dat3$sodium)
```

```
## [1] 99.5
```

```
sd(dat3$sodium)
```

```
## [1] 7.119284
```

```
#Average and standard deviation of elec
mean(dat3$elec)
```

```
## [1] 199.5
```

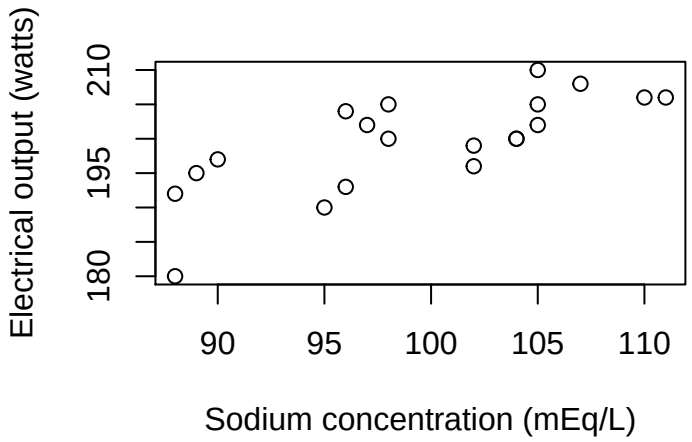
```
sd(dat3$elec)
```

```
## [1] 7.141428
```

Plot data

The next code chunk will display your data as a scatterplot with custom x- and y-axis labels.

```
plot(dat3$elec~dat3$sodium,
       ylab = "Electrical output (watts)", xlab = "Sodium concentration (mEq/L)")
```



Run correlation analysis

This code chuck performs a Pearson's correlation analysis and displays summary data.

```
#Perform a correlation analysis and print Pearson's r, p-value, and associated statistics
cor.test(dat3$sodium, dat3$elec, method = c("pearson"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  dat3$sodium and dat3$elec
## t = 4.7285, df = 18, p-value = 0.0001677
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4500103 0.8927751
## sample estimates:
##     cor
## 0.74431
```

Full code block

```
#Create data frame
dat3 = data.frame(sodium = c(95,104,88,97,102,89,104,105,111,88,
                             98,102,96,105,107,105,96,90,110,98),
                  elec = c(190,200,192,202,196,195,200,202,206,180,
                           205,199,193,205,208,210,204,197,206,200))
#Print data frame
dat3

#Summary statistics
#Average and standard deviation of sodium
mean(dat3$sodium)
sd(dat3$sodium)

#Average and standard deviation of watts
mean(dat3$elec)
sd(dat3$elec)

#Create scatterplot
plot(dat3$elec~dat3$sodium,
        ylab = "Electrical output (watts)", xlab = "Sodium concentration (mEq/L)")

#Perform a correlation analysis and print Pearson's r, p-value, and associated statistics
cor.test(dat3$sodium, dat3$elec, method = c("pearson"))
```