Laboratory 2 Skills – KEY

2025-01-17

Exercise 1

Sperm whales (Physeter macrocephalus) are among the deepest ocean divers among mammals. A team of marine mammal researchers attached time-depth recorders to sperm whales in the Pacific and Atlantic oceans to record individual diving depth for dives lasting longer than 30 minutes. The results of the study are presented below, showing dive depth in meters.

Pacific sperm whales: 420, 729, 442, 529, 484, 720, 453, 1002, 561, 980, 806, 263, 897, 652, 575, 346, 794, 553, 410, 417, 830, 1050, 1196, 687, 629, 1125, 496, 674, 998, 62, 735, 628, 1082, 1014, 732, 474, 111, 716, 567, 646, 286, 531, 550, 1346, 1401, 949, 644, 787, 929, 756, 763, 1035, 631, 1235, 395, 302, 804, 388, 574, 568, 491, 1389, 1125

Atlantic sperm whales: 614, 283, 415, 756, 288, 435, 473, 487, 512, 552, 795, 544, 472, 731, 290, 772, 674, 387, 670, 271, 648, 323, 344, 399, 812, 616, 396, 598, 1082, 594, 363, 332, 457, 456, 125, 566, 884, 801, 567, 442, 722, 374, 485, 370, 343, 658, 425, 429, 901, 489, 480, 431, 608, 576, 348, 331, 432, 538

Copy and paste each dataset into R. Don't forget to define the datasets with a name using the c() function, and remember: capitalization, spacing, and parentheses are important when using datasets and functions in R!

1. Use R to estimate the maximum, minimum, mean, median, variance, and standard deviation for each dataset using the max(), min(), mean(), median(), var(), and sd() functions.

```
# Your code here
```

Your final answer here

2. Calculate the range for each dataset.

```
# Your code here
```

Your final answer here

3. Calculate the coefficient of variation for each dataset, and then interpret the results.

```
# Your code here
```

Your final answer here

4. Assume that the data from both datasets are normally distributed. Compute the extent of diving depths (i.e., from x to y meters) that correspond to approximately 95% of the diving depths for each dataset. Remember, this is the empirical rule.

 # Your code here

Your final answer here

5. Calculate the lower (Q1), and upper (Q3) quartiles of the distribution for each dataset. You can do this by using the quantile() command.

 # Your code here

Your final answer here

6. Calculate the interquartile range (IQR) for each dataset.

 # Your code here

Your final answer here

7. Now compute the dive depth at the 95th and 5th percentile for each dataset. Note: The empirical rule tells us something subtly different from percentiles. Use quantile(data, probs = c(.05,.95)) to calculate percentiles. Note that we can also use this function to compute deciles.

 # Your code here

Your final answer here

8. Calculate the z-scores for a Pacific sperm whale that dove to a depth of 949 meters and an Atlantic sperm whale that dove to 538 meters, and then compare the results.

 # Your code here

Your final answer here

9. Determine if there are any outliers. In this case we want to look for z-scores less than -3 or greater than 3. Report the z-score as well as the depth associate with that z-score for any outliers.

 # Your code here

Your final answer here

10. Use the boxplot() command to create a box-whisker plot for both the Atlantic and Pacific sperm whale data side-by-side and copy and paste it below. Hint: The names argument in the plot() function lets you label the x-axis ticks.

 # Your code here

Your final answer here

Exercise 2

The gender and age of students from a past QSCI381 class were surveyed. The results of this survey can be found in the "lab2data.csv" file. REMEMBER: capitalization, spacing, and parentheses are important when using datasets and functions in R!

Include all code for each question, plus output from R or graphs, and answer all sub-questions.

1. Read in the "lab2data.csv" file using the read.csv() command and name it data. Look at the first six rows of data.

```
# Your code here
```

Your final answer here

2. Change the Age column values of data from numeric to character ("under21" , "over21") values using the ifelse() function. data$AgeFactor <- ifelse(data$AgeFactor < 21, "under21", "over21")

```
# Your code here
```

Your final answer here

3. Create a contingency table for Gender and Age from data and name it conting. Add the row and column totals to complete the contingency table.

```
# Your code here
```

Your final answer here

4. What is the probability of being female? Of being female and over 21? Of being female OR over 21?

a) P(Female)

```
# Your code here
```

Your final answer here

b) P(Female AND over21)

```
# Your code here
```

Your final answer here

c) P(Female OR over21)

```
# Your code here
```

Your final answer here

5. What is the probability of being male given one is over 21? What is the probability of being male given one is not over 21? Why do these two probabilities not sum to one?

a) P(Male|over21)

```
# Your code here
```

Your final answer here

b) P(Male|under21)

```
# Your code here
```

Your final answer here

c) Why do these two probabilities not sum to one?

Your final answer here