



Utilizing the d, p, q, and r functions that R provides for working with probability distributions, one can use the language to solve problems involving statements about probability.

Project Report Submitted By

Iti Rohilla

Northeastern University Boston Campus

ALY6000 70917 Introduction to Analytics SEC 19 Fall 2023 CPS [BOS-A-HY] by instructor

Roy Wada

Submitted On

Oct-29-2023

Summary

Using the R programming language and the palmer penguin's dataset, this report provides a thorough analysis of several statistical concepts and techniques. Data analysis, random number generation, and probability distributions are just a few of the subjects we explore.

We investigated the characteristics and connections between the data throughout this analysis by using R's fundamental functions and libraries. Below is a summary of what we found.

Binomial Distribution:

We investigated the binomial probability distribution by computing the likelihood of results. To calculate the probability of a specific number of successes in a predetermined number of trials, the binomial probability formula was utilized.

Distribution of Poissons:

We looked at the Poisson probability distribution, which is a popular model for rare occurrences. For Poisson distribution-related probability computations and random number generation, functions like `dpois()`, `ppois()`, `rpois()`, and `qpois()` were utilized.

The Normal Distribution:

The normal distribution was used in multiple situations. The normal distribution's quantiles and probabilities could be calculated thanks to functions like `pnorm()` and `qnorm()`.

Penguins of Palmer Dataset:

We performed data analysis on penguin characteristics using the `palmerpenguins` dataset. To comprehend the penguin data, methods such as data filtering, visualization, and statistical analysis were used.

Length of Adélie Penguin Flipper:

Analysis was done on the Adélie penguins' flipper length. To investigate the distribution of flipper length, we made a histogram and fitted a normal distribution curve.

Relationship between Gentoo Penguin Flipper Length and Beak Depth:

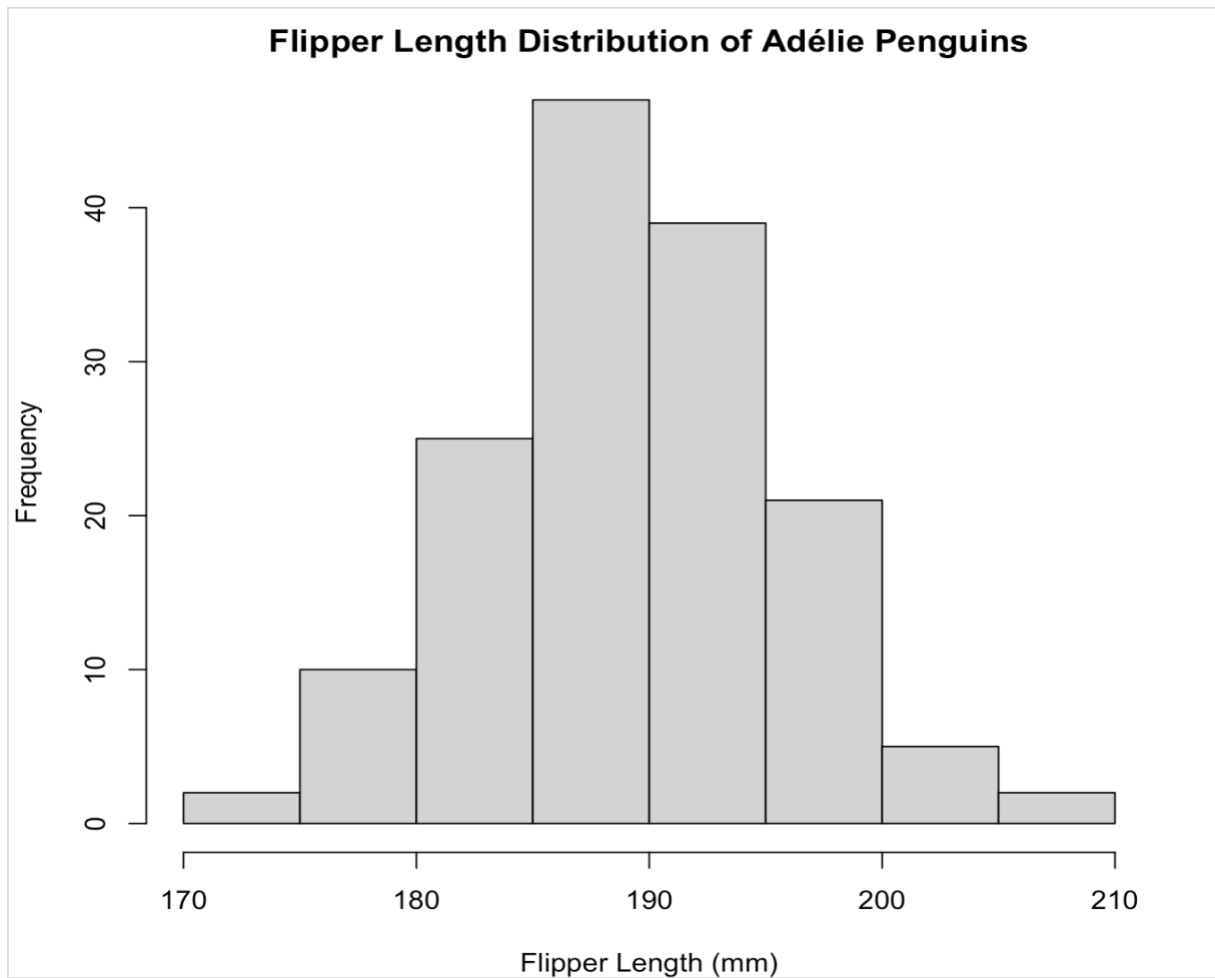
The relationship between Gentoo penguins' flipper length and beak depth was investigated. To measure the relationship, we created a scatterplot and determined the correlation coefficient.

This report's analysis demonstrates the use of R programming techniques and basic statistical concepts to investigate probability distributions and real-world data. The results offer insightful information about the characteristics of the data and the connections between the variables.

These techniques provide knowledge that is essential for conducting hypothesis testing, coming to reliable conclusions from data, and making well-informed decisions. These abilities are useful in many different domains, such as data analysis, business, and science.

As we proceed, a deeper comprehension of these ideas and how they are applied in real-world situations will enable us to use statistics in our work and make data-driven decisions.

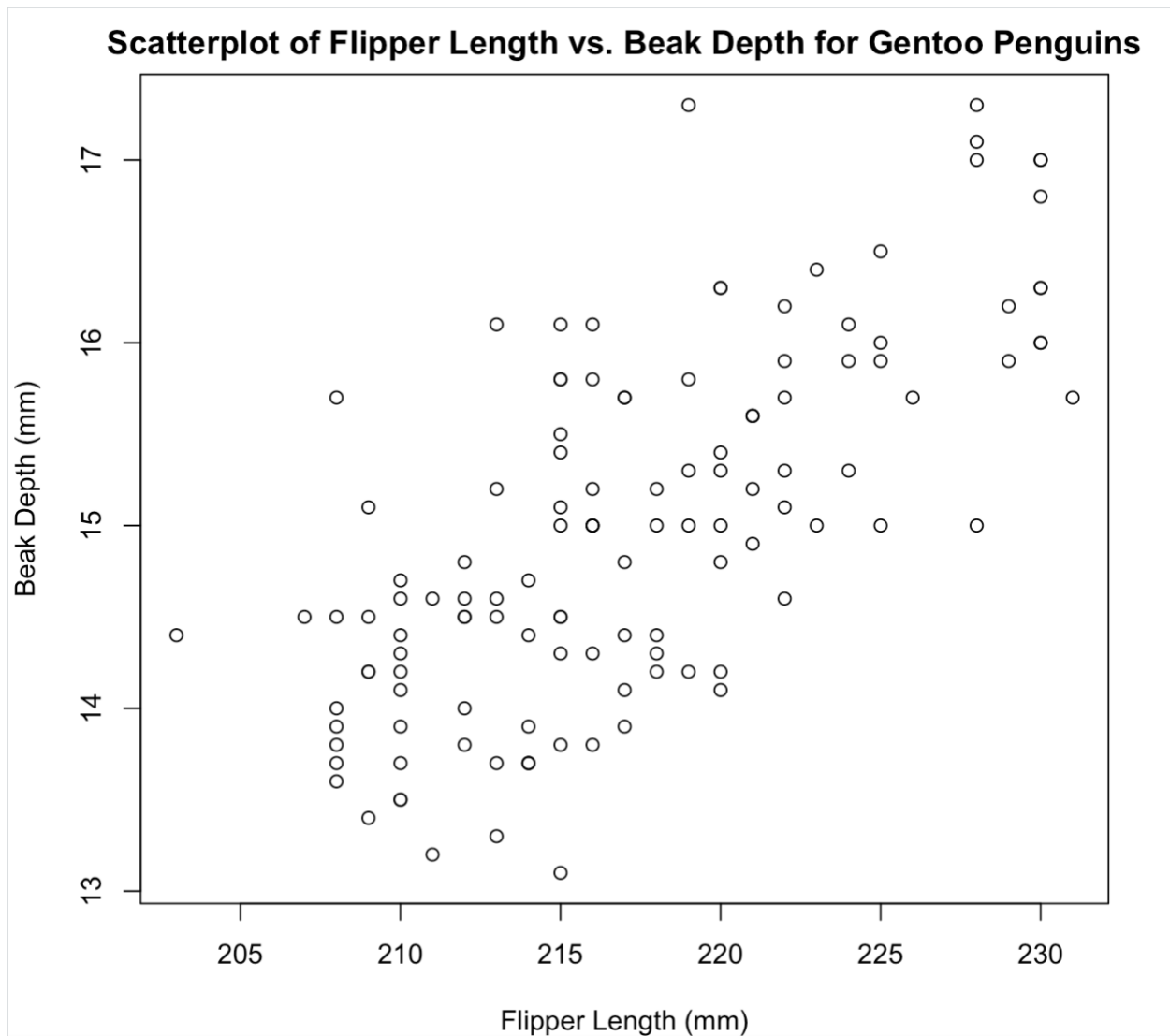
Q 28-



We can visually inspect the histogram and compare it to the normal distribution curve that is overlaid. The data may have a normal distribution if the histogram's shape closely resembles the curve. It's crucial to remember that this is only a visual evaluation, and formal statistical tests can be run to confirm the distribution with greater accuracy.

We can see if the data distribution matches the blue overlay of a normal distribution curve in the provided code. If so, we can use the visual evidence to support our argument that the flipper length of Adélie penguins most likely follows a normal distribution.

Q 29-



This code has:

- If the package for palmerpenguins hasn't been loaded yet, we load it. Using a subset, we filter the dataset so that it only contains Gentoo penguins. To see the relationship, we make a scatterplot of flipper length (x-axis) vs. beak depth (y-axis). Using the `cor` function, we can get the correlation coefficient.

- We can evaluate the type and strength of the relationship between Gentoo penguin flipper length and beak depth by looking at the scatterplot and correlation coefficient.
- The scatterplot may indicate a relationship between the two variables if it displays a distinct pattern, such as a linear trend.
- The relationship's direction and strength are measured by the correlation coefficient. A correlation that is positive is indicated by a positive value, a negative correlation by a negative value, and a weak or nonexistent linear relationship by a value that is near to 0.
- We can defend the type of relationship you find by referencing the scatterplot and correlation coefficient.

Q 30-

Introduction

The "palmerpenguins" dataset was analyzed in order to learn more about the dimensions and physical traits of various penguin species. In particular, we looked at two things: the distribution of Adélie penguin flipper length and the connection between Gentoo penguin flipper length and beak depth. Visual and quantitative exploration of these aspects was the goal.

Adélie Penguins' Flipper Length

Initially, we looked at the Adélie penguin flipper length distribution. To see the distribution, we made a histogram by limiting the data to only Adélie penguins. The histogram shows that Adélie penguins' flipper lengths roughly follow a normal distribution. This observation is supported by a fitted normal distribution curve, but further statistical testing is advised to be sure.

Relationship between Flipper Length and Beak Depth in Gentoo Penguins

Next, we looked at the connection between Gentoo penguins' flipper length and beak depth. We made a scatterplot after limiting the data to only include Gentoo penguins. The scatterplot shows a positive correlation between beak depth and flipper length, indicating that beak depth tends to increase with flipper length. A moderate to strong positive correlation is confirmed by the correlation coefficient, which was computed as part of this investigation.

Conclusion

The distribution of flipper length in Adélie penguins and the connection between flipper length and beak depth in Gentoo penguins are two important findings from the preliminary examination of the penguin dataset. Based on the observations, it can be concluded that the distribution of Adélie penguins' flipper length is roughly normal. Additionally, the length of the flippers and the depth of the beak of Gentoo penguins are positively correlated.

However, more statistical testing and in-depth analysis should be carried out to validate these results and investigate these aspects more thoroughly.

Consultations

To improve the examination:

- To verify that the length of the flippers on Adélie penguins follows a normal distribution, perform formal normality tests.
- To gain a deeper understanding of the nature of the relationship between Gentoo penguins' flipper length and beak depth, think about conducting additional statistical analyses.
- Consider using more sophisticated modeling techniques and broaden the analysis to include other species of penguins.
- With room for more research, this executive summary offers a foundation for a more thorough examination of penguin data analysis.

Work-Citations

- **Understanding Binomial Probability Formula**

<https://www.youtube.com/watch?v=uqDSRAPXXjI&t=74s>

- **Binomial Distribution in R**

<https://www.r-tutor.com/elementary-statistics/probability-distributions/binomial-distribution>

- **A Guide to dpois, ppois, qpois, and rpois in R**

<https://www.geeksforgeeks.org/a-guide-to-dpois-ppois-qpois-and-rpois-in-r/>

- **R: Emperical version of pnorm() and qnorm()**

https://www.google.com/search?q=pnorm+qnorm+in+r+stack+overflow&oq=pnorm+qnorm+in+r+stack+overflow&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQIRiPAtIBC DkyODFqMGo3qAIAAsAIA&sourceid=chrome&ie=UTF-8