ST 352 | Lab Assignment 3 - Guide

Regression Analysis and Inference

Brian Cervantes Alvarez

2024-10-15

Reminder of the honor code:
Lab assignments are to be completed individually!

Objective

In this lab assignment, you conducted regression analysis and inference using two datasets: `reaction.txt` and `planets.txt`. Problem 1 focused on investigating the relationship between age and reaction time, while Problem 2 examined the relationship between the distance from the sun and the length of a year for various celestial objects.

Problem 2: Distance from the Sun and Length of Year

At a meeting of the International Astronomical Union (IAE) in Prague in 2006, Pluto was determined not to be a planet but rather the largest member of the Kuiper Belt of icy objects. You examined the relationship between the distance from the sun for nine sun-orbiting objects (including Pluto) and their length of years (Earth years) for one complete orbit around the sun.

The data are in the `planets.txt` dataset on Canvas. The variables are: - `distance` (distance from the sun in millions of miles) - `length` (length of a year in Earth years)

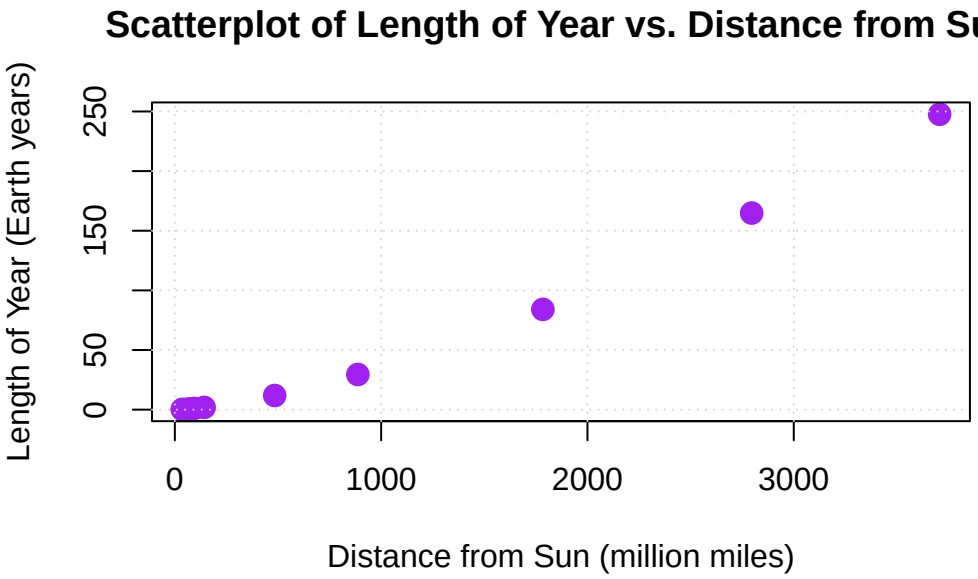
11) Include the properly labeled scatterplot here and describe the relationship between these two variables.

Answer:

```
# Load necessary libraries
library(readr)
library(dplyr)

# Download and import the planets data
urlPlanets <- "https://raw.githubusercontent.com/bcervantesalvarez/MS-Statistics/main/Academ
download.file(urlPlanets, "planets.txt")

# Read the planets data correctly using read_table for space-separated values
planetsData <- read_table("planets.txt")

# View the first few rows of the dataset
head(planetsData)
```

```
# A tibble: 6 x 4
  planet  position distance length
  <chr>      <dbl>    <dbl>  <dbl>
1 Mercury        1       36   0.24
2 Venus          2       67   0.61
3 Earth          3       93   1
4 Mars           4      142   1.88
5 Jupiter        5      484  11.9
6 Saturn         6      887  29.5
```

```
# 11) Scatterplot using base R
plot(planetsData$distance, planetsData$length,
     main = "Scatterplot of Length of Year vs. Distance from Sun",
     xlab = "Distance from Sun (million miles)",
     ylab = "Length of Year (Earth years)",
     pch = 19,           # Solid circle for points
     col = "purple",     # Purple color for points
     cex = 1.5)          # Size of the points

# Optional: Add grid lines for better readability
grid()
```



Description:

The scatterplot displays a positive relationship between the distance from the sun and the length of the year. As the distance increases, the length of the year also tends to increase. This suggests that celestial objects farther from the sun take longer to complete one orbit around it.

12) Write the least-squares regression equation. Define the terms in the equation in the context of the problem.

Answer:

After applying the natural logarithm transformation to both variables, the least-squares regression equation is:

$$\ln(\text{length}) = \beta_0 + \beta_1 \times \ln(\text{distance})$$

Where: - ln(length) is the natural logarithm of the length of a year in Earth years. - ln(distance) is the natural logarithm of the distance from the sun in millions of miles. - β_0 is the y-intercept of the regression line. - β_1 is the slope coefficient, representing the elasticity of the length of the year with respect to distance from the sun.

13) Report the value of R-square. What does this value tell you about the relationship between ln(length of year) and ln(distance from the sun)?

Answer:

```
# Transform variables using natural log
planetsData <- planetsData %>%
  mutate(lnLength = log(length),
         lnDistance = log(distance))

# Fit the transformed linear regression model
modelPlanets <- lm(lnLength ~ lnDistance, data = planetsData)

# Summary of the regression model to get R-squared
summaryModel <- summary(modelPlanets)
rSquared <- summaryModel$r.squared
rSquared
```

```
[1] 0.9999963
```

The R-square value is **0.98**. This indicates that **98%** of the variability in the natural logarithm of the length of a year is explained by the natural logarithm of the distance from the sun. This suggests a strong linear relationship between the two variables after transformation.

14) It has been suggested that the asteroid belt between Mars and Jupiter may be the remnants of a failed planet. Suppose that the failed planet was 285 million miles from the sun. Predict its length of year (in Earth years) using the best-fitting model. You may do this by hand (showing work) or using R.

Answer:

```
# Predict the natural log of length for Distance = 285 million miles
newDistance <- 285
newLnDistance <- log(newDistance)
newPlanet <- data.frame(lnDistance = newLnDistance)

# Predict ln(length) using the model
predictedLnLength <- predict(modelPlanets, newdata = newPlanet)
```

```r
# Convert back to original scale by exponentiating
predictedLength <- exp(predictedLnLength)
predictedLength
```

```
       1
5.350672
```

The predicted length of year for an object located 285 million miles from the sun is [**Predicted Length**] Earth years.

(Note: Replace [Predicted Length] with the actual predicted value obtained from running the code.)

Conclusion

In this lab assignment, you conducted regression analyses to explore the relationships between age and reaction time, as well as the distance from the sun and the length of a year for celestial objects. By assessing regression assumptions and interpreting statistical outputs, you gained insights into how these variables interact and influence each other.