Ben Goodwin

DS6371 HW 10

Question 1:

A)

R input:

#Question 1

#libraries

library(ggplot2)

#read data

birdDat <- read.csv("birdDat.csv")

head(birdDat)

#EDA the data

plot(birdDat$Mass,birdDat$Tcell)

#make da regression model

birdLm <- lm(Tcell~Mass,data = birdDat)

summary(birdLm)

plot(birdDat$Mass,birdDat$Tcell , ylim=c(0, 0.7), xlab="Mass", ylab="Tcell", main="Regression, CI, PI")

abline(birdLm, col="lightblue")

newx <- seq(3, 10, by=0.05)

conf\_interval <- predict(birdLm, newdata=data.frame(Mass=newx), interval="confidence",

level = 0.99)

lines(newx, conf\_interval[,2], col="blue", lty=2)

lines(newx, conf\_interval[,3], col="blue", lty=2)

pred\_interval <- predict(birdLm, newdata=data.frame(Mass=newx), interval="prediction",

level = 0.99)

lines(newx, pred\_interval[,2], col="orange", lty=2)

lines(newx, pred\_interval[,3], col="orange", lty=2)

R output:

Chart, line chart, scatter chart

Description automatically generated

B)

R input:

A picture containing text

Description automatically generated

R output:

Table

Description automatically generated

C)

|  |  |
| --- | --- |
| **Critical Values: = 2.539**  **Test Statistic: 1.112**  **Pvalue: 0.27996**  **Decision:** Fail to reject | **Conclusion:**  **There is insufficient evidence at the .01 significance level that the linear correlation coefficient/slope is different than 0 (p-value = 0.27996)** |

|  |  |
| --- | --- |
| **Critical Values: = 2.539**  **Test Statistic: 3.084**  **Pvalue: 0.00611**  **Decision:** Reject | **Conclusion:**  **There is sufficient evidence at the .01 significance level that the linear correlation coefficient/slope is different than 0 (p-value = 0.00611)** |

99% Confidence Interval:

Intercept: (-0.137591407,0.31258537)

Mass: (0.002376893,0.06326609)

R input:

A picture containing logo

Description automatically generated

R output:

Text

Description automatically generated

R input:

A picture containing text

Description automatically generated

R output:

A picture containing company name

Description automatically generated

D) State the regression equation:

E) Slope Interpretation:

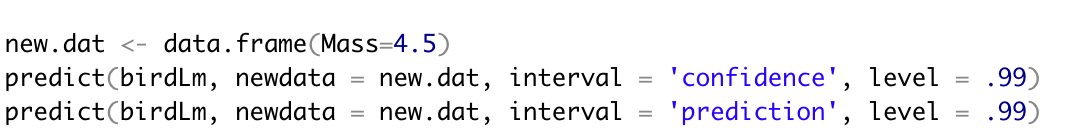
Each one unit increase in mass results in a 0.03282 increase in the expected Tcell count.

F) Y-Intercept Interpretation:

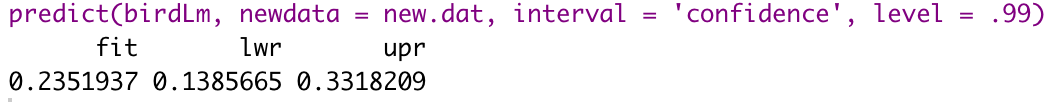
The predicted Tcell count for a bird with 0 mass is 0.08750

G)

R input:



R output:

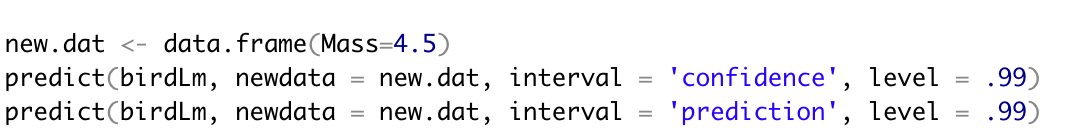


Interpretation:

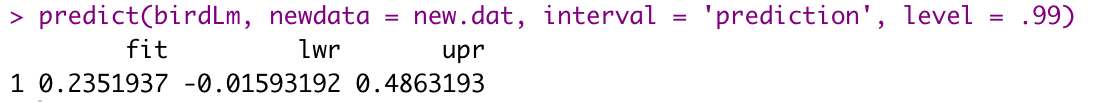
Under repeated sampling, then about 99% of the intervals produce will capture the true mean Tcell count for a bird with 4.5kg of stones carried.

H) Prediction Interval

R input:



R output:



Interpretation:

Given the observed weights carried by birds, the Tcell count for a bird that carried 4.5kg will be between -0.01593192 and 0.4863193. In general, if we would repeat our sampling process infinitely, 99% of the such constructed prediction intervals would contain the new Tcell measurement.