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",

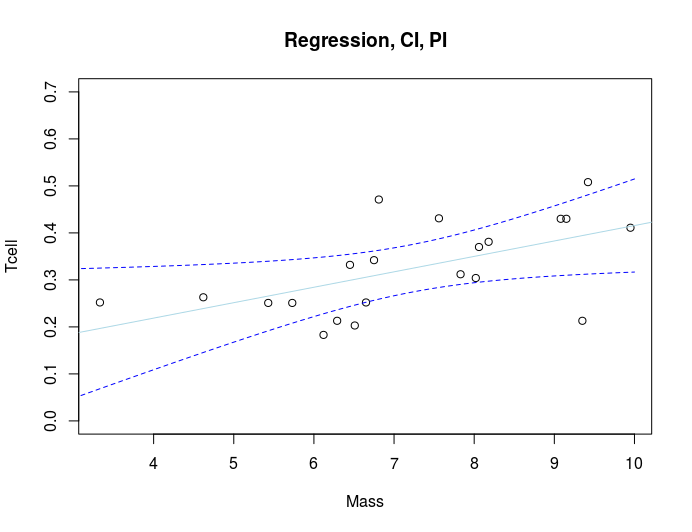
level = 0.99)

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

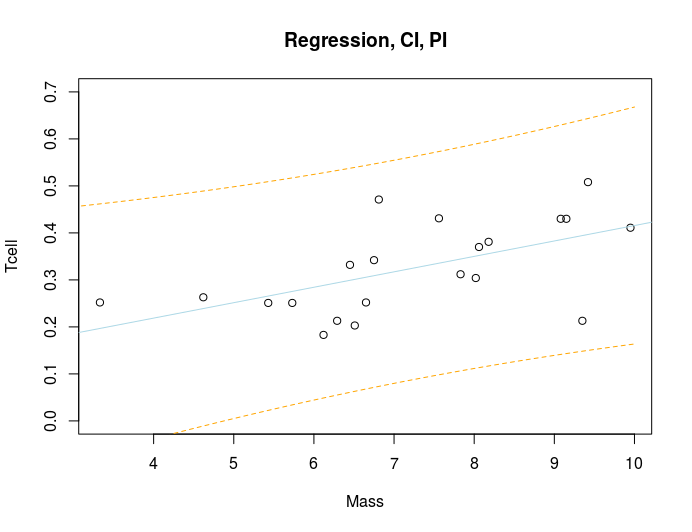
lines(newx, pred\_interval[,3], col="orange", lty=2)

R output:

99% CI plot:



99% PI plot:



Combined Plot

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)**  **Scope: The data does not specify if the sample was a random sample, so we cannot claim causation. However, other assumptions such as independence, and homogeneity of variance seem to be met.**  **99% CI: (0.002376893,0.06326609)** |

|  |  |
| --- | --- |
| **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)**  **Scope: : The data does not specify if the sample was a random sample, so we cannot claim causation. However, other assumptions such as independence, and homogeneity of variance seem to be met.**  **99% CI:** (-0.137591407,0.31258537) |

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:

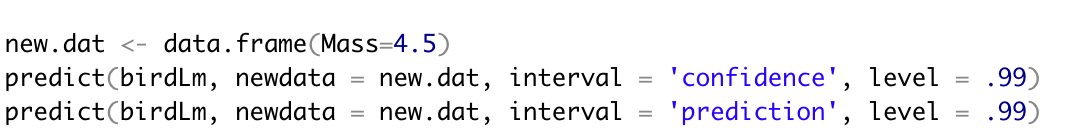
Each one unit increase in mass results in a 0.03282 increase in the expected Tcell count.

F) Y-Intercept Interpretation:

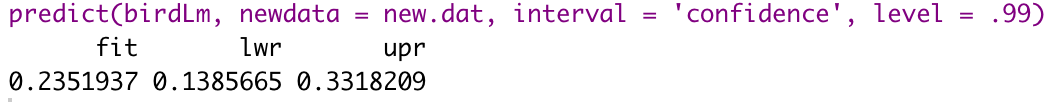
The predicted Tcell count for a bird with 0 mass is 0.08750

G)

R input:



R output:

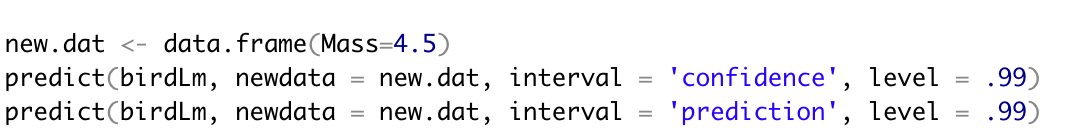


Interpretation:

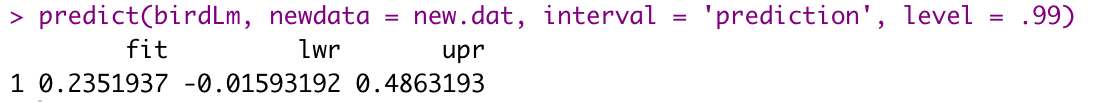
Under repeated sampling, then about 99% of the intervals produce will capture the true mean Tcell count for a bird with 4.5kg of stones carried.

H) Prediction Interval

R input:



R output:

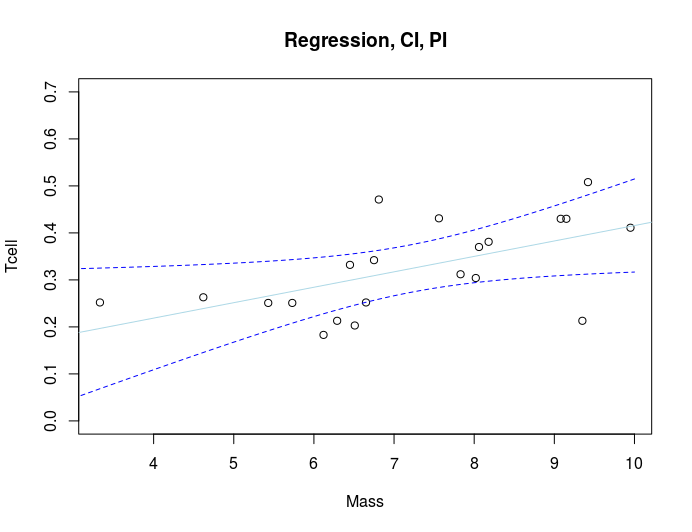


Interpretation:

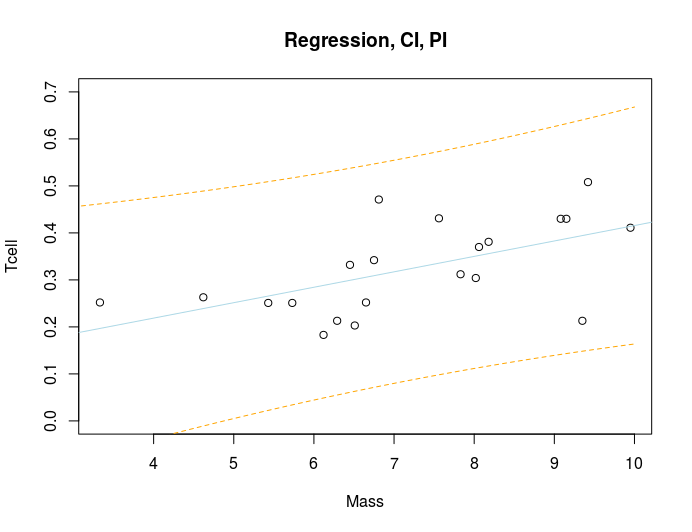
Given the observed weights carried by birds, the Tcell count for a bird that carried 4.5kg will be between -0.01593192 and 0.4863193. In general, if we would repeat our sampling process infinitely, 99% of the such constructed prediction intervals would contain the new Tcell measurement.

I)

A



B



Part II

A

R Input



R output:



B

R Input



R Output



Part III

1. Interpret:

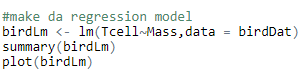
We are 99% confident that the estimated mass needed to obtain a mean Tcell count of 0.3 is between -4.389857 and 8.342649

1. Interpret”

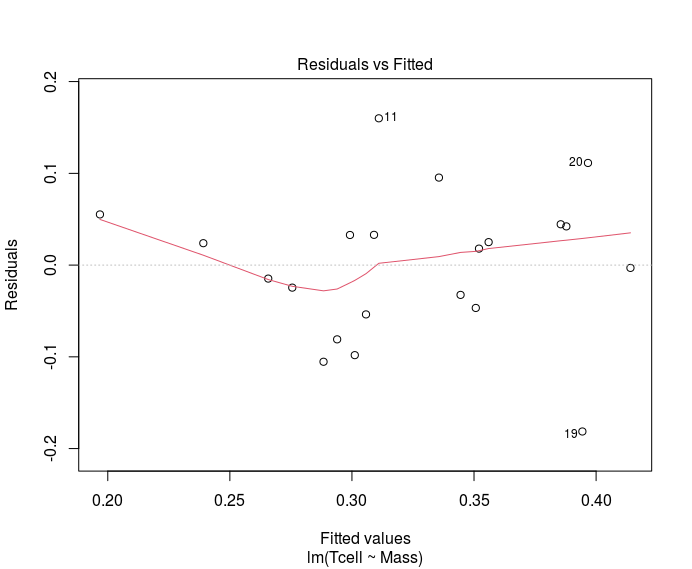
We are 99% confident that the mass required for a Tcell count of 0.3 is between -17.968869 and 21.921661

J

R Input:

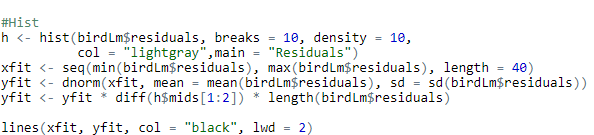


R output:

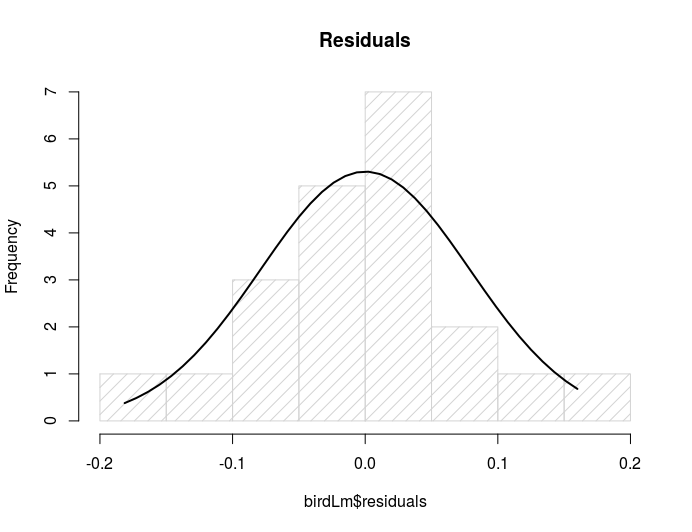


K

R input:

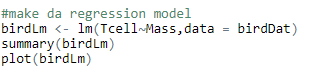


R output:

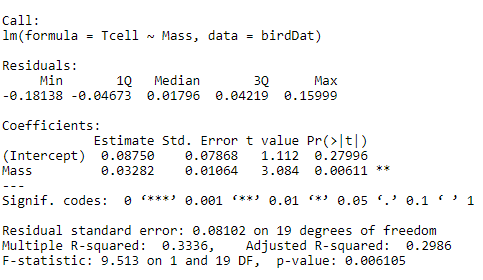


L

R input:



R output:



The regression model accounts for 33.36% of the variance.

Question 2

1. B\_0:0.08750

B\_1:0.03282

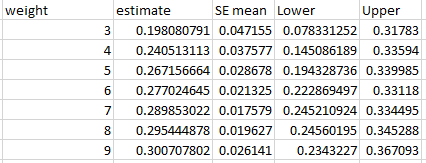
B)

|  |  |
| --- | --- |
| **Critical Values: = 2.539**  **Test Statistic: 1.163197**  **Pvalue: 0.259151**  **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 = 0.259151)**  **The data does not specify if the sample was a random sample, so we cannot claim causation. However, other assumptions such as independence, and homogeneity of variance seem to be met.** |

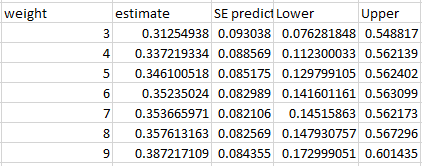
|  |  |
| --- | --- |
| **Critical Values: = 2.539**  **Test Statistic: 3.177562**  **Pvalue: 0.004957**  **Decision:** Reject | **Conclusion:**  **There is sufficient evidence at the .01 significance level that the linear correlation coefficient/slope is different than 0 (p-value = 0.004957)**  **Scope:The data does not specify if the sample was a random sample, so we cannot claim causation. However, other assumptions such as independence, and homogeneity of variance seem to be met.** |

1. 99% confidence intervals for the mean of Y when X = {3,4,5,6,7,8,9} grams.

Excel output:



1. 99% Prediction interval for the mean of Y when X = {3,4,5,6,7,8,9} grams.



E)

F)

A)



B)



G) A)

B)