UNIT 10 HW

1. Black-eared wheatears are small birds of Spain and Morocco. Males of the species demonstrate an exaggerated sexual display by carrying many heavy stones to nesting cavities. This 35-gram bird transports, on average, 3.1 kg of stones per nesting season! Different males carry somewhat different sized stones, prompting a study of whether larger stones may be a signal of higher health status. M. Soler et al. calculated the average stone mass (g) carried by each of 21 male black-eared wheatears, along with T-cell response measurements (in mm) reflecting their immune system strengths. Analyze the data and write a statistical report (by answering the questions below); treat the T-cell as the response and the stone mass as the explanatory variable. You may assume all criteria for regression and related t-tests are met. You can find the data for this problem on 2DS. (Male Display Data Set)

Analyze the data, providing the following:

* + 1. Provide a scatterplot with 99% confidence intervals of the regression line and 99% prediction intervals of the regression line. Please do this in R.

Code:

library(xlsx)

setwd("C:/Users/Marin Family/Desktop/Statistical Foundations for Data Science/Unit 10")

bird <- read.xlsx("MaleDisplayDataSet\_2\_2\_2.xlsx","Male Display Data Set")

plot(bird)

bird.lm=lm(Tcell~Mass, data = bird)

newx=bird$Mass

newx=sort(newx)

prd\_c=predict(bird.lm, newdata= data.frame(Mass = newx), interval=c("confidence"), type = c("response"), level=0.99)

prd\_c

prd\_p=predict(bird.lm, newdata= data.frame(Mass = newx), interval=c("prediction"), type = c("response"), level=0.99)

prd\_p

#Plot with confidence and prediction intervals

plot(bird[,1],bird[,2],xlim = c(0,10), ylim = c(0,.8),xlab = "Mass",ylab = "Tcell", main = "Mass versus Tcell")

abline(bird.lm, col = "red")

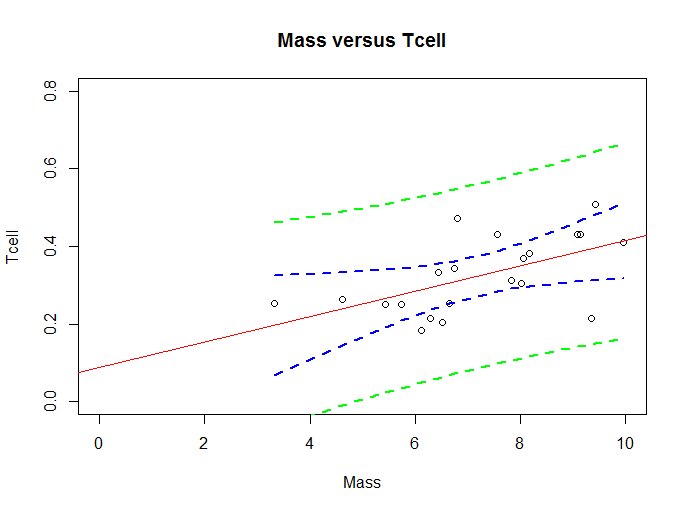
lines(newx,prd\_c[,2],col = "blue",lty = 2, lwd = 2)

lines(newx,prd\_c[,3],col = "blue", lty = 2, lwd = 2)

lines(newx,prd\_p[,2],col = "green", lty = 2, lwd = 2)

lines(newx,prd\_p[,3],col = "green", lty = 2, lwd = 2)

Output:



* + 1. Provide a table showing the t-statistics and p-values for the significance of the regression parameters (as different from 0). Please do this in R.

Code:

summary(bird.lm)

Output:

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.08750 0.07868 1.112 0.27996

Mass 0.03282 0.01064 3.084 0.00611 \*\*

* + 1. Using the output in (b), show all 6 steps of ***each*** hypothesis test. (That’s one test for and one test for .) Find critical values in R. Your conclusion should include a confidence interval. Use alpha = 0.01.

|  |  |
| --- | --- |
| Step 1:  Ho: B1 = 0  Ha: B1 <> 0  Step 2: Crit val = 2.86  Step 3: T-Value = 1.11  Step 4: P-val = .28  Step 5: Fail to Reject the Null  Step 6: There is not sufficient evidence at the alpha = .01 level of significance (p-value = .28) to suggest that the data are linearly correlated or the slope is nonzero with a confidence interval of [0, .31]. | Ho: B0 = 0  Ha: B0 <> 0  Step 2: Crit val = 2.86  Step 3: T-Value: 3.084  Step 4: P-val = <.006  Step 5: Reject the Null  Step 6: There is sufficient evidence at alpha = .01 level of significance (p-val= < .006) to suggest that the data are linearly correlated or the intercept is nonzero with a confidence interval of [.002,.06]. |

* + 1. State the regression equation. Be careful to use the mean Tcell or predicted Tcell, rather than just Tcell.

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.08750 0.07868 1.112 0.27996

Mass 0.03282 0.01064 3.084 0.00611 \*\*

Tcell = .033 \* Mass + .088

* + 1. Interpret the slope in the model (regression equation).

For an increase in Mass, the Tcell increases by .033 mm.

* + 1. Interpret the y-intercept in the model (regression equation).

For a stone weighing zero grams, the estimated TCell would be .088 mm.

* + 1. Find and interpret the 99% confidence interval for the mean t-cell response conditional on a stone mass of 4.5 grams. Please do this directly in R.

Code:

newpoint <- data.frame(Mass=4.5, TCell=NA)

predict(bird.lm, newpoint, interval = "confidence", level = .99)

Output:

fit lwr upr

1 0.2351937 0.1385665 0.3318209

We are 99% confident that when the stone mass is 4.5 grams, the mean TCell count is between [.14,.33].

* + 1. Find and interpret the 99% prediction interval for the predicted t-cell response given a stone mass of 4.5 grams. Please do this directly in R.

Code:

newpoint <- data.frame(Mass=4.5, TCell=NA)

predict(bird.lm, newpoint, interval = "prediction", level = .99)

Output:

fit lwr upr

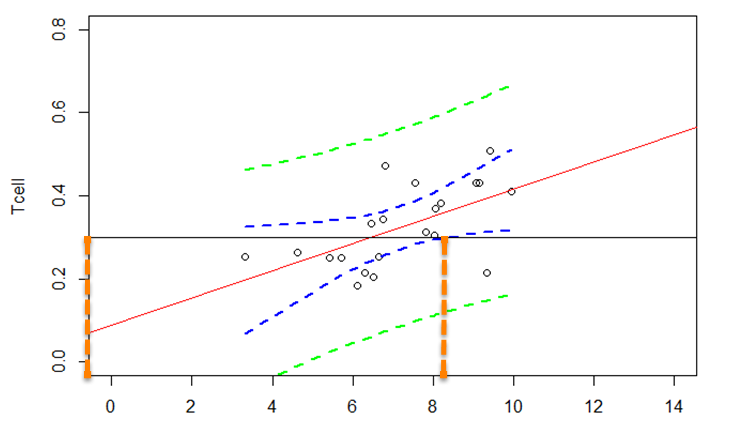
1 0.2351937 -0.01593192 0.4863193

We are 99% confident that that the Tcell (of a single bird) when the mass is 4.5 grams, will be between [0,.49]

* + 1. Calibration intervals:

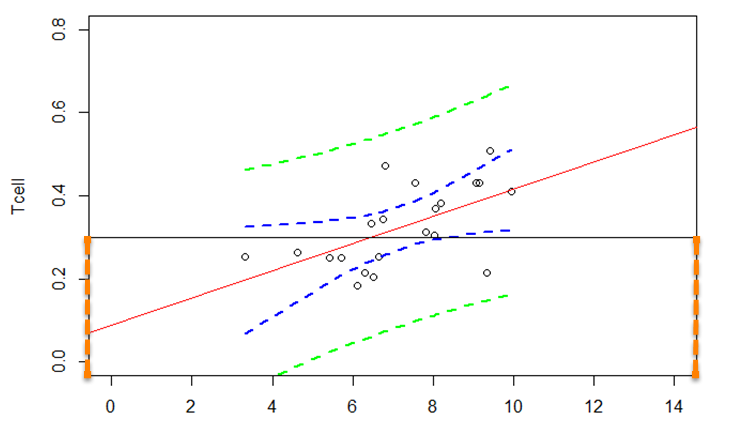
Using the **graphical method** (using your best judgment using the graphs from part (a)), find the following using R, as part (a) was done in R.

99% calibration interval for the **mean** t-cell response of 0.3.



Looks like between [0, 8.something]

99% calibration interval for a **single** t-cell response of 0.3.



Looks beyond what I can see from the data.

Using **software directly**, find the following using R, as SAS does not provide calibration intervals directly. (R: package investr)

99% calibration interval for the **mean** t-cell response of 0.3.

Code:

library(investr)

#Mean Response

calibrate(bird.lm, y0 = .3, interval = "inversion", level = .99, mean.response = TRUE)

Ouput:

estimate lower upper

6.474508 -4.389857 8.342649

99% calibration interval for a **single** t-cell response of 0.3.

Code:

#Single

calibrate(bird.lm, y0 = .3, interval = "inversion", level = .99, mean.response = FALSE)

Output:

estimate lower upper

6.474508 -17.968869 21.921661

**Interpret** the following using the results from (1) and (2) above.

99% calibration interval for the **mean** t-cell response of 0.3.

We are 99% confident that the required Mass to indicate a mean Tcell count of .3mm is between approximately [0,8.34].

99% calibration interval for a **single** t-cell response of 0.3.

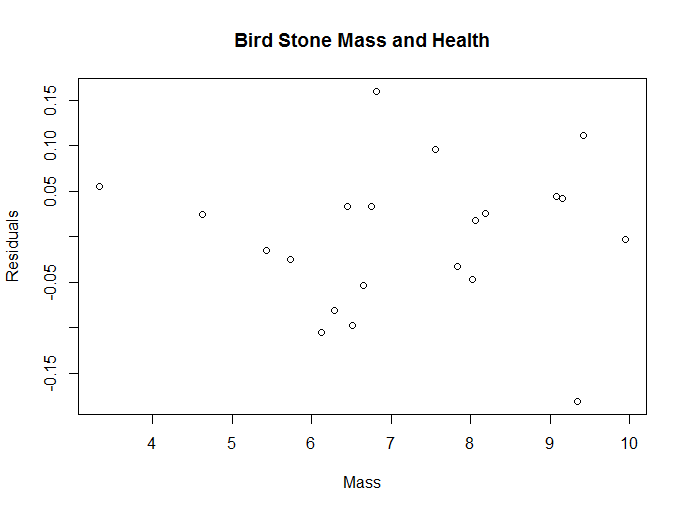
We are 99% confident that the required Mass to predict a Tcell count of .3mm is between approximately [0,21.92].

* + 1. Provide a scatterplot of residuals. Please do this in R.

Code:

plot(bird$Mass, bird.res, ylab = "Residuals", xlab = "Mass", main = "Bird Stone Mass and Health")

Output:

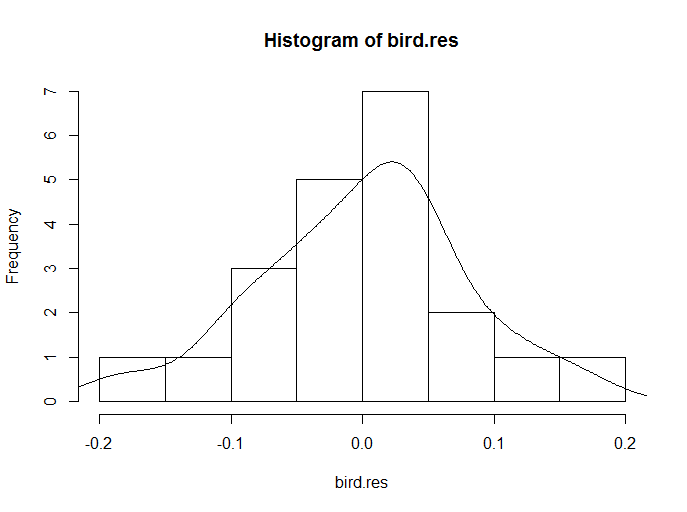


* + 1. Provide a histogram of residuals with a normal distribution superimposed. It might be helpful to use studentized residuals, rather than regular residuals, with a normal curve overlay. Use R. (You may need to research this, such as googling “histogram with normal curve in R.”)

Code: hist(bird.res)

hist(bird.res)

lines(density(bird.res))



* + 1. Provide a measure of the **proportion** of variation in the response that is accounted for by the explanatory variable. **Interpret** this measure. Use R.

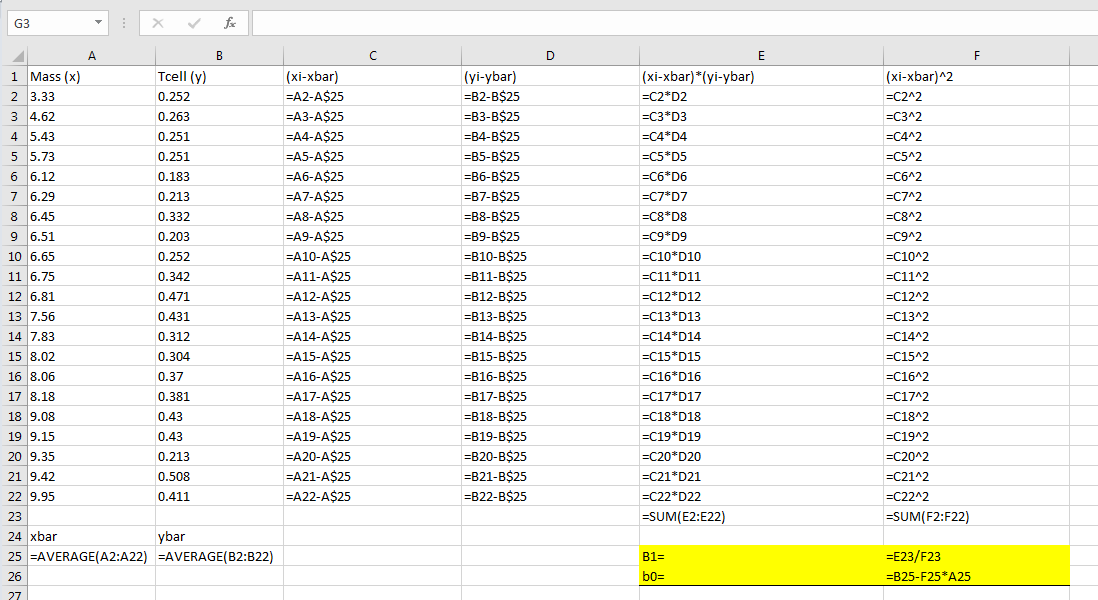
Code: summary(bird.lm)

R^2 = 0.3336

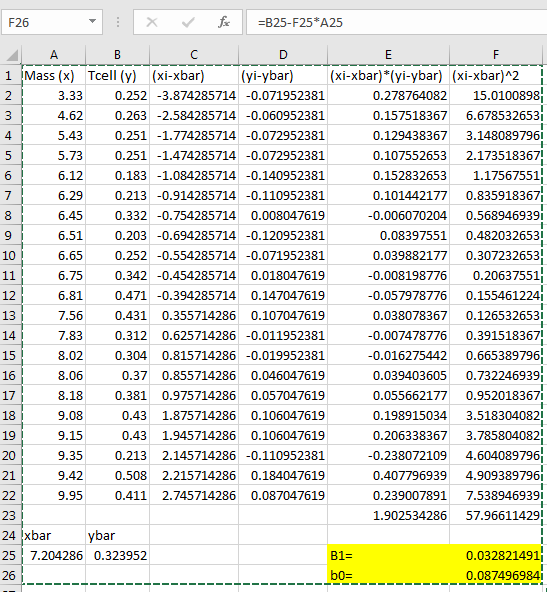
TCell explains 33.3% of the variation in Mass carried by these Birds.

1. Using the data for Black-eared Wheatears, calculate by “hand” (using Excel) the following elements. (An example of much of this was in the PowerPoints and in the videos below.)

Formulas:



Output:

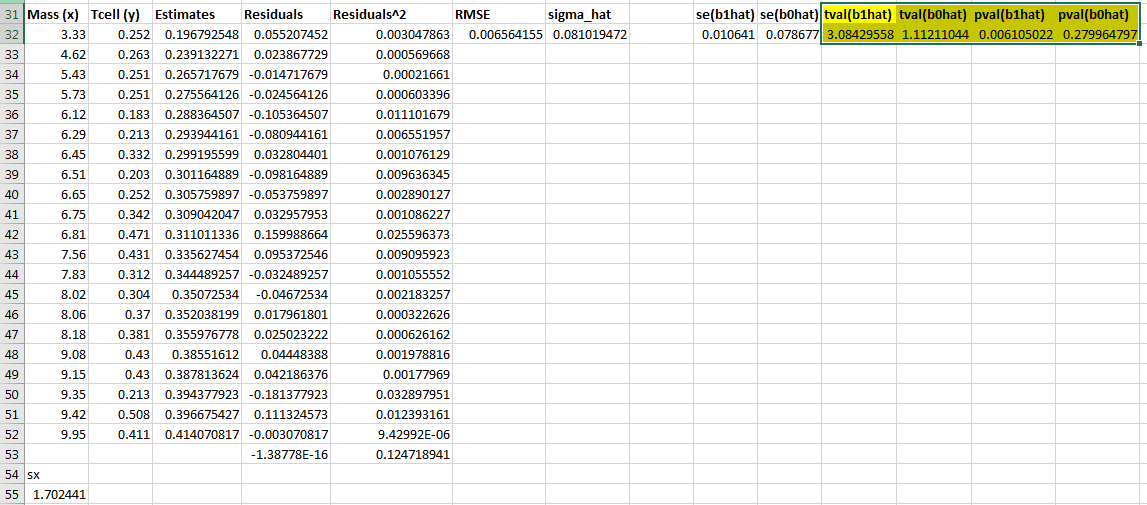


* 1. The t-statistics and p-values for the hypothesis tests ( and for ).

Formulas:

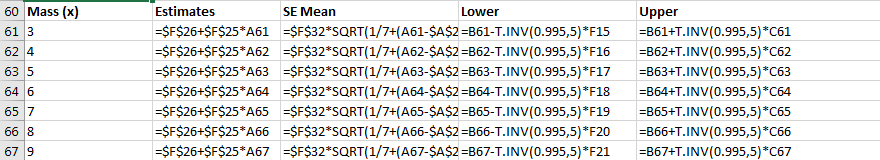


Output:

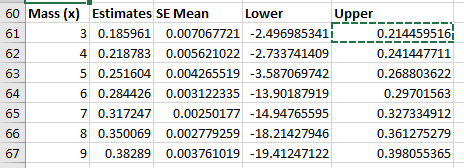


* 1. 99% confidence intervals for the mean of Y when X = {3,4,5,6,7,8,9} grams.

Formulas:

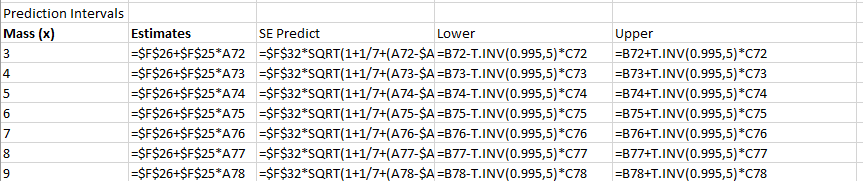


Output:

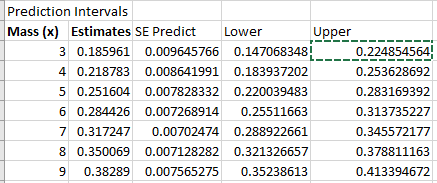


* 1. 99% prediction intervals for the predicted Y when X = {3,4,5,6,7,8,9} grams.

Formulas:



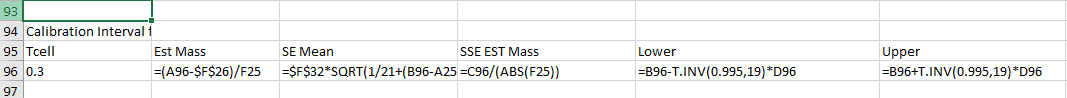
Output:



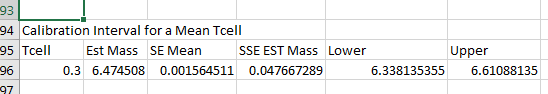
* 1. Provide a plot for the confidence intervals and prediction intervals using Excel. Fully label your graph. (Use the regression equation and parts (c) and (d) above to create the plot.)
  2. Calibration intervals:

1. Using the SE equations given in class and in the book (Version 3 page 194), find the following **analytically**. (Use Excel for calculations.)
   1. 99% calibration interval for the mean t-cell response of 0.3.

Formula:

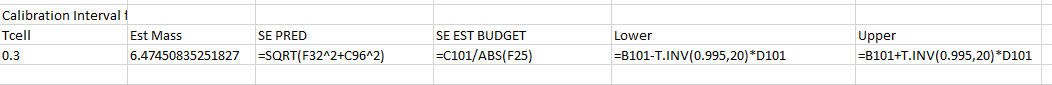


Output:

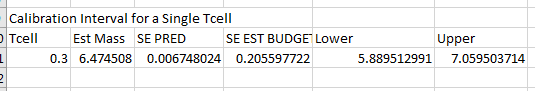


* 1. 99% calibration interval for a single t-cell response of 0.3.

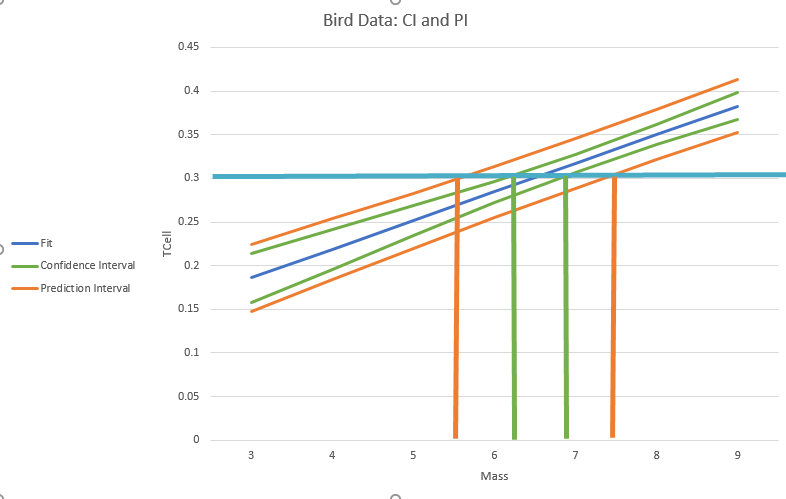
Formula:



Output:

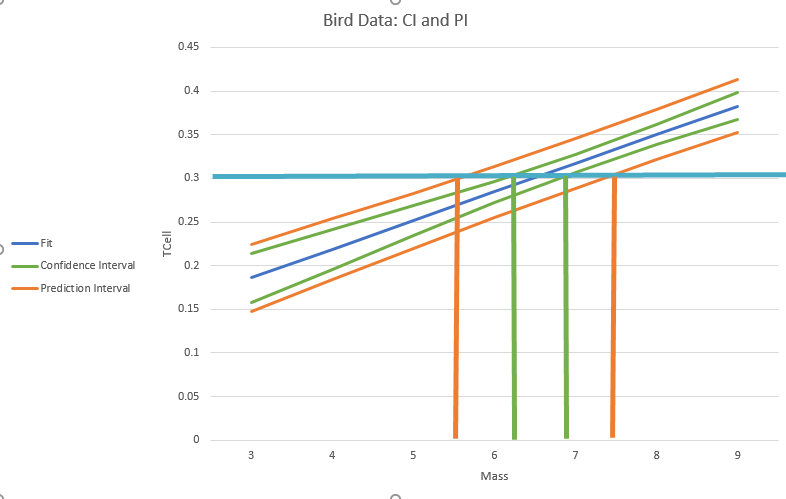


1. Using the **Excel graphs**, find the following. You may want to add data points to parts (c) and (d) so that the confidence and prediction limits extend well beyond the data range (although their interpretation is questionable outside the range).
   1. 99% calibration interval for the mean t-cell response of 0.3.



Green bars going down

* 1. 99% calibration interval for a single t-cell response of 0.3.



Orange bars going down.

1. Bonus!
   1. Repeat 1 (a) using SAS.
   2. Repeat 1 (b) using SAS.
   3. Repeat 1 (c) using SAS.
   4. Repeat 1 (g) using SAS.
   5. Repeat 1 (h) using SAS.
   6. Repeat 1 (j) using SAS.
   7. Repeat 1 (k) using SAS.
   8. Repeat 1 (l) using SAS.

Code used for bonus questions:

PROC IMPORT OUT= WORK.bird

DATAFILE= "/home/marinfamily1010/sasuser.v94/Data/MaleDisplayDataSet\_2\_2\_2.xlsx"

DBMS=xlsx REPLACE;

GETNAMES=YES;

DATAROW=2;

RUN;

data Work.Bird2;

input Mass Tcell;

datalines;

3 .

4 .

5 .

6 .

7 .

8 .

9 .

;

run;

proc sql;

create table work.bird3 as

select \* from work.bird

union

select \* from work.bird2

order by Tcell, Mass;

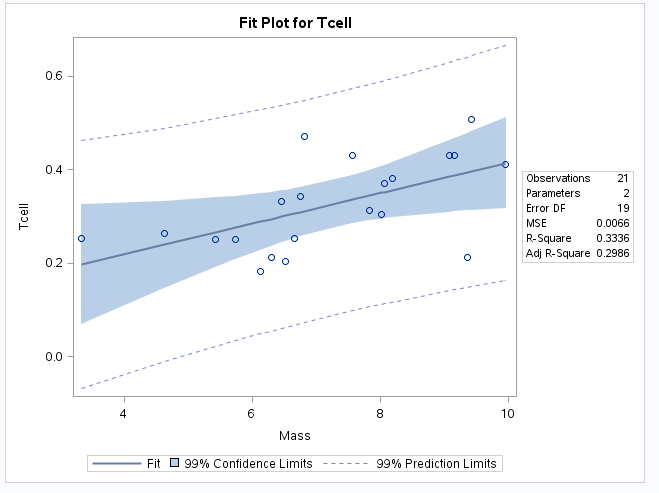
run;

proc reg data = work.bird3;

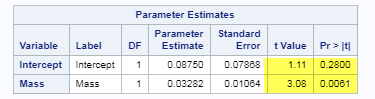
model Tcell = Mass /cli;

run;

* + 1. Provide a scatterplot with 99% confidence intervals of the regression line and 99% prediction intervals of the regression line.



* + 1. Provide a table showing the t-statistics and p-values for the significance of the regression parameters (as different from 0).



* + 1. Using the output in (b), show all 6 steps of ***each*** hypothesis test. (That’s one test for and one test for .) Find critical values in R. Your conclusion should include a confidence interval. Use alpha = 0.01.

|  |  |
| --- | --- |
| Step 1:  Ho: B1 = 0  Ha: B1 <> 0  Step 2: Crit val = 2.86  Step 3: T-Value = 1.11  Step 4: P-val = .28  Step 5: Fail to Reject the Null  Step 6: There is not sufficient evidence at the alpha = .01 level of significance (p-value = .28) to suggest that the data are linearly correlated or the slope is nonzero with a confidence interval of [0, .31]. | Ho: B0 = 0  Ha: B0 <> 0  Step 2: Crit val = 2.86  Step 3: T-Value: 3.08  Step 4: P-val = <.0061  Step 5: Reject the Null  Step 6: There is sufficient evidence at alpha = .01 level of significance (p-val= < .006) to suggest that the data are linearly correlated or the intercept is nonzero with a confidence interval of [.002,.06]. |

* + 1. State the regression equation. Be careful to use the mean Tcell or predicted Tcell, rather than just Tcell.

Tcell = .03282 \* Mass + .08750

* + 1. Interpret the slope in the model (regression equation).

For an increase in Mass, the Tcell increases by .033 mm.

* + 1. Interpret the y-intercept in the model (regression equation).

For a stone weighing zero grams, the estimated TCell would be .088 mm.

* + 1. Find and interpret the 99% confidence interval for the mean t-cell response conditional on a stone mass of 4.5 grams.

Code:

PROC IMPORT OUT= WORK.bird

DATAFILE= "/home/marinfamily1010/sasuser.v94/Data/MaleDisplayDataSet\_2\_2\_2.xlsx"

DBMS=xlsx REPLACE;

GETNAMES=YES;

DATAROW=2;

RUN;

data Work.Bird2;

input Mass Tcell;

datalines;

4.5 .

;

run;

proc sql;

create table work.bird3 as

select \* from work.bird

union

select \* from work.bird2

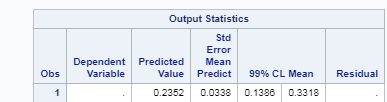
order by Tcell, Mass;

run;

proc reg data = work.bird3 alpha=.01;

model Tcell = Mass / clm;

run;



* + 1. Find and interpret the 99% prediction interval for the predicted t-cell response given a stone mass of 4.5 grams.

Updated Code from above to:

proc reg data = work.bird3 alpha=.01;

model Tcell = Mass / cli; /\*clm for mean cli for individual\*/

run;

Output:

