DDavieau\_UNIT\_10HW\_6371

Daniel Davieau

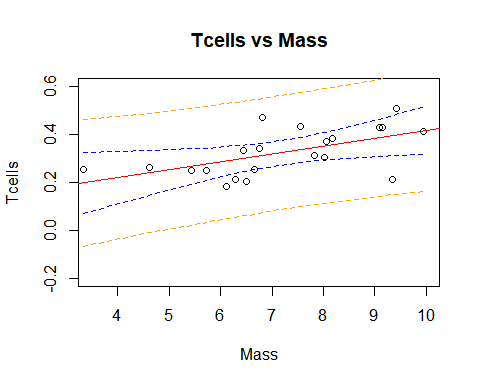
March 17, 2018

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.

MaleTest.lm<-lm(Tcell~Mass,data = Male\_Display\_Data\_Set)  
#qt(0.995, 198)  
#abs(qt(0.01/2, 198))  
#confint(MaleTest.lm, level = .99)  
NewMass=sort(Male\_Display\_Data\_Set$Mass)  
  
prd\_c=predict(MaleTest.lm, newdata= data.frame(Mass = NewMass), interval=c("confidence"), type = c("response"), level=0.99)  
  
prd\_p=predict(MaleTest.lm, newdata= data.frame(Mass = NewMass), interval=c("prediction"), type = c("response"), level=0.99)   
  
plot(x=Male\_Display\_Data\_Set$Mass,y=Male\_Display\_Data\_Set$Tcell,xlim = c(3.5,10), ylim = c(-.2,.6),xlab = "Mass",ylab = "Tcells", main = "Tcells vs Mass")  
abline(MaleTest.lm, col = "red")  
lines(NewMass,prd\_c[,2],col = "blue",lty = 2, lwd = 1)  
lines(NewMass,prd\_c[,3],col = "blue",lty = 2, lwd = 1)  
lines(NewMass,prd\_p[,2],col = "orange",lty = 2, lwd = 1)  
lines(NewMass,prd\_p[,3],col = "orange",lty = 2, lwd = 1)

 b. Provide a table showing the t-statistics and p-values for the significance of the regression parameters β\_(0 ) and β\_1 (as different from 0). Please do this in R.

summary(MaleTest.lm)

##   
## Call:  
## lm(formula = Tcell ~ Mass, data = Male\_Display\_Data\_Set)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.18138 -0.04673 0.01796 0.04219 0.15999   
##   
## 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 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08102 on 19 degrees of freedom  
## Multiple R-squared: 0.3336, Adjusted R-squared: 0.2986   
## F-statistic: 9.513 on 1 and 19 DF, p-value: 0.006105

1. Using the output in (b), show all 6 steps of each hypothesis test. (That’s one test for β\_(0 )=0 and one test for β\_1=0.) Find critical values in R. Your conclusion should include a confidence interval. Use alpha = 0.01.

qt(0.995, 19)

## [1] 2.860935

confint(MaleTest.lm, level=.99)

## 0.5 % 99.5 %  
## (Intercept) -0.137591407 0.31258537  
## Mass 0.002376893 0.06326609

1. State the regression equation. Be careful to use the mean Tcell or predicted Tcell, rather than just Tcell.
2. Interpret the slope in the model (regression equation).
3. Interpret the y-intercept in the model (regression equation).
4. 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.

newpoint <- data.frame(Tcell=NA, Mass=4.5)  
predict(MaleTest.lm, newpoint, interval="confidence", level = 0.99)

## fit lwr upr  
## 1 0.2351937 0.1385665 0.3318209

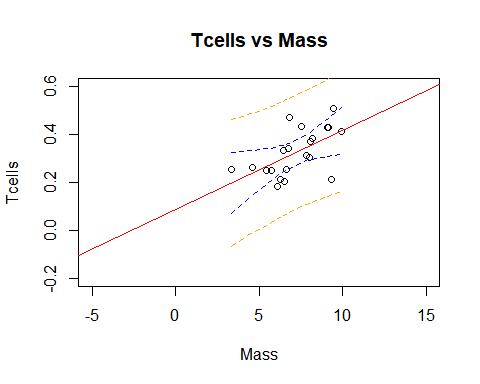
1. Find and interpret the 99% prediction interval for the predicted t-cell response given a stone mass of 4.5 grams. Please do this directly in R.

newpoint <- data.frame(Tcell=NA, Mass=4.5)  
predict(MaleTest.lm, newpoint, interval="prediction", level = 0.99)

## fit lwr upr  
## 1 0.2351937 -0.01593192 0.4863193

1. Calibration intervals:
   1. 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.
      1. 99% calibration interval for the mean t-cell response of 0.3.

plot(x=Male\_Display\_Data\_Set$Mass,y=Male\_Display\_Data\_Set$Tcell,xlim = c(-5,15), ylim = c(-.2,.6),xlab = "Mass",ylab = "Tcells", main = "Tcells vs Mass")  
abline(MaleTest.lm, col = "red")  
lines(NewMass,prd\_c[,2],col = "blue",lty = 2, lwd = 1)  
lines(NewMass,prd\_c[,3],col = "blue",lty = 2, lwd = 1)  
lines(NewMass,prd\_p[,2],col = "orange",lty = 2, lwd = 1)  
lines(NewMass,prd\_p[,3],col = "orange",lty = 2, lwd = 1)

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

2. Using software directly, find the following using R, as SAS does not provide calibration intervals directly. (R: package investr)   
 a. 99% calibration interval for the mean t-cell response of 0.3.

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

## estimate lower upper   
## 6.474508 -4.389857 8.342649

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

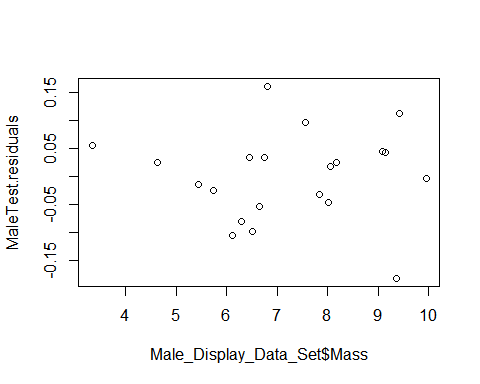
calibrate(MaleTest.lm, y0 = .3, interval = "inversion", level = .99)

## estimate lower upper   
## 6.474508 -17.968869 21.921661

3. Interpret the following using the results from (1) and (2) above.   
 a. 99% calibration interval for the mean t-cell response of 0.3.  
 b. 99% calibration interval for a single t-cell response of 0.3.

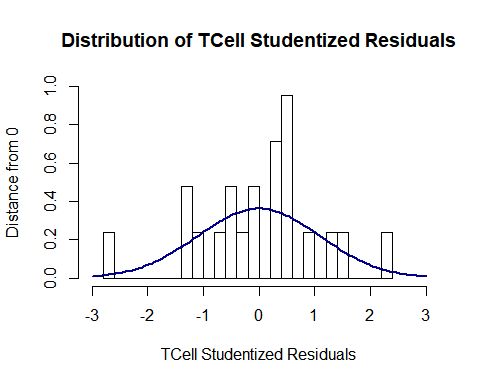
1. Provide a scatterplot of residuals. Please do this in R.

MaleTest.residuals<- resid(MaleTest.lm)  
plot(Male\_Display\_Data\_Set$Mass, MaleTest.residuals)



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.”)

MaleTest.StudentizedResiduals<-studres(MaleTest.lm)  
m<-mean(MaleTest.StudentizedResiduals)  
std<-sqrt(var(MaleTest.StudentizedResiduals))  
hist(MaleTest.StudentizedResiduals, breaks=20, prob=TRUE,   
xlab="TCell Studentized Residuals",ylab="Distance from 0",xlim = c(-3,3), ylim = c(-0,1), main="Distribution of TCell Studentized Residuals")  
curve(dnorm(x, mean=m, sd=std),  
col="darkblue", lwd=2, add=TRUE, yaxt="n")



summary(MaleTest.StudentizedResiduals)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.79179 -0.58416 0.22290 -0.00479 0.54248 2.22760

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.

summary(MaleTest.lm)$r.squared

## [1] 0.3336345

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.)
   1. β ̂) an$ β ̂
   2. The t-statistics and p-values for the hypothesis tests (β\_(0 )=0 and for β\_1=0).  
      c.99% confidence intervals for the mean of Y when X = {3,4,5,6,7,8,9} grams.
   3. 99% prediction intervals for the predicted Y when X = {3,4,5,6,7,8,9} grams.
   4. 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.)
   5. Calibration intervals:
   6. 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.
      2. 99% calibration interval for a single t-cell response of 0.3.
   7. 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.
      2. 99% calibration interval for a single t-cell response of 0.3.
   8. Bonus! Repeat 1 (a) using SAS. Repeat 1 (b) using SAS. Repeat 1 (c) using SAS. Repeat 1 (g) using SAS. Repeat 1 (h) using SAS. Repeat 1 (j) using SAS. Repeat 1 (k) using SAS. Repeat 1 (l) using SAS.

Videos for using Excel: β ̂\_(0 ) and β ̂\_1 <http://screencast.com/t/ztSxTImiOk6s> SE of β ̂\_(0 ) and β ̂\_1 and RMSE: <http://screencast.com/t/V9gnhSwb> Confidence Intervals: <https://www.screencast.com/t/ELiUGTe7Kc> Prediction Intervals: <https://www.screencast.com/t/ap8WETxsGUqN> CI and PI Plotting: <https://www.screencast.com/t/efrpHrqgYZnG> Calibration Mean Gross: <https://www.screencast.com/t/Yu7eqiiH0X> Calibration Single Movie: <https://www.screencast.com/t/2vS1lGqtJ>

