Name: Joe Hediger

HOMEWORK 1: Linear Models and Inference

WSCI 6390 – 002: Population Parameter Estimation

Due 11:59 PM Tuesday, January 30

Let’s practice fitting (i.e., estimating parameters and associated variances) and making inference from linear models.

INSTRUCTIONS:

We have a dataset on ornate box turtles, including home range size (ha; *obt.hr*), plastron length (mm; *plastron*), turtle sex (1 = male, 0 = female; *sex*), and the local density of a favored cover plant, yucca(plants/m2; *yucca*).

You can find the data in the file “turtle.csv” - import these data into your **R** environment using read.csv() and call it **turtle**.

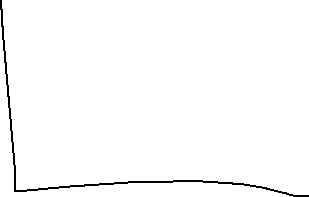
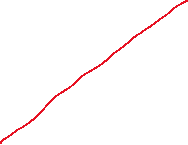
Be sure to install the packages “ggplot2” and “faraway” before completing this assignment.

ASSIGNMENT:

1. Use the dim() argument to tell me how many rows and columns are in this **turtle** dataset. You can use ?dim to learn about this function (or any other functions in R).
   1. Rows = 50
   2. Columns = 4
2. Use mean(turtle$COLUMN\_NAME) to tell me what the mean values are for (1) home range size, (2) plastron length, and (3) local density of yucca.
   1. Home range: 12.58 ha
   2. Plastron length: 105.86 mm
   3. Yucca: 1.5 plants/m2
3. Consider the model . Can you draw what that model will look like on the x and y axes (please label these axes appropriately) if we have an intercept () of 5 and a slope () of 1?



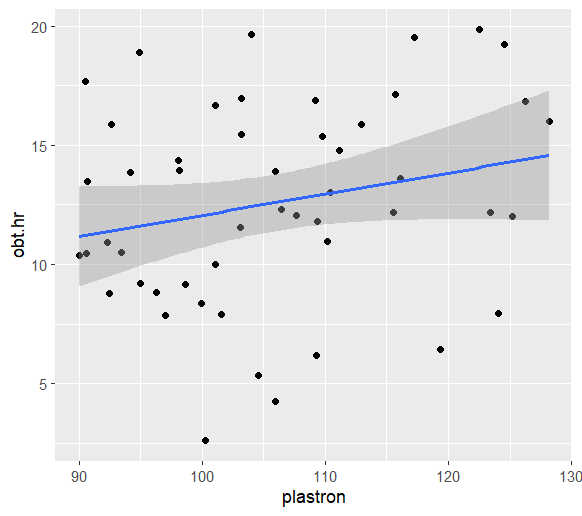
Home range (Ha)



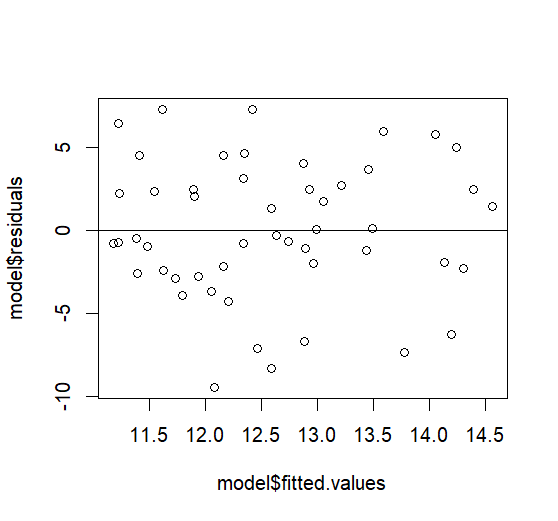
Plastron (mm)



1. Use the lm() argument in R to build this model. What is the value of the intercept? What is the value of the intercept standard error? Hint: use summary(MODEL\_NAME) to see these values.
   1. Intercept: 3.18
   2. Standard error of intercept: 5.8
2. What is the value of the beta coefficient for plastron length? What is the standard error?
   1. Beta: 0.09
   2. Standard error: 0.05
3. Is the slope (), the beta coefficient for plastron length, “significant” given a p-value of 0.05?
   1. Not significant
4. What is the R2 value for this model (let’s use Multiple R2 from the summary() argument) and what does that value mean? Do you consider this a low or high value?
   1. R2: 0.05
   2. This value describes how much the equation explains turtle home ranges. It is measure of how well the model fits the data.
   3. Low value
5. Please make a plot of observed home ranges against plastron length using ggplot2. There is an example of how to do this in the lecture slides. Paste it here.



1. Produce a plot of residuals versus fitted values (see lecture slides for how to do this). Be sure to add a line at so you can get a better view of what’s happening (abline(h = 1)). Do we meet the assumption of constant variance?
   1. Yes, constant variance



1. Let’s test the assumption of constant variance one other way. Fit the regression model Code is:

model.test <- lm(sqrt(abs(MODEL\_NAME $residuals)) ~ MODEL\_NAME $fitted)

summary(model.test)

Do we reach a different conclusion here?

* 1. No, P >0.05, not constant variance based upon equation.

1. Produce a Q-Q plot using

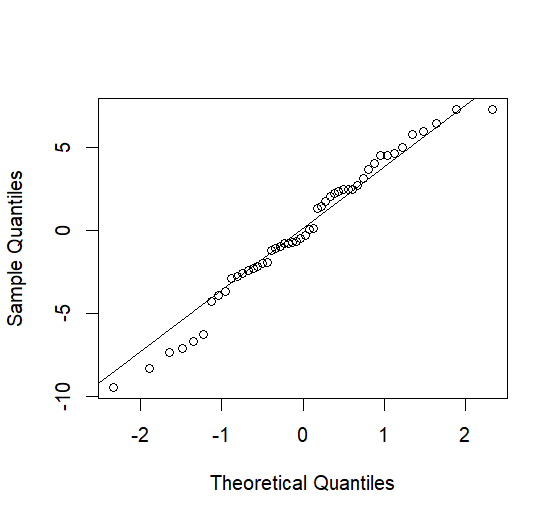
qqnorm(residuals(MODEL\_NAME), main= "")

qqline(residuals(MODEL\_NAME))

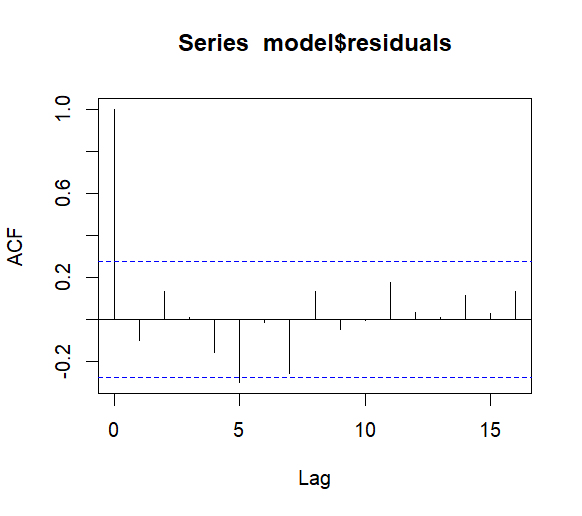
and paste it here.

Do we meet the assumption of normally distributed errors?

* 1. Yes we meet the assumption



1. Do we meet the assumption of independently distributed errors? Use acf(MODEL\_NAME$residuals) to see if there is any autocorrelation in the residuals after lag of 0 and paste that plot here.
   1. Yes errors are independently distributed



1. Do we have any influential observations that need a closer look? Use Cook’s D and plot a half-normal plot with these values to explore this. See final lecture slides for how to do this. Be sure to paste the plot with your answer.
   1. Yes, observations 14 and 15

