**Lab #4: Inference in Regression Analysis** STAT 415/615 Miller Wednesday February 2, 2022

**Instructions:** You will need R (or R Studio) for this lab assignment.

Also, keep in mind that in working with R (or R studio) that:

* There is often more than one way to achieve the same task (e.g., I will provide some instructions for creating plots, but it is also possible to get the same plots with a special package, such as ggplot2). If you have previous experience with R, feel free to use different code to achieve the required result.
* It is always helpful to start an R session with **library(tidyverse)**.

Guidelines for your lab submission:

* Your lab responses must be submitted as a single .**PDF (Portable Document Format)** file.
* Include your name at the top.
* Please copy and paste any R (or similar) output or graphics that you create into your lab document. All responses must be easy to read and labeled with the appropriate problem number.
* Please save your work regularly (both your work in R and the solutions to the lab questions).
* Submit the .PDF file with your lab responses via **Canvas** before you leave class (or by the 11:59 PM ET if you need more time once class is over).

**Instructions** In this lab, we will work with the data set from the notes and Assignment 3**-*“Mass\_Calorie\_Data.csv”***. Load the data into R.

* Set your working directory.
* Enter

Women<-read.csv("Mass\_Calorie\_Data.csv",header=T,sep=",")

* And then enter

attach(Women)

Here I will provide some notes on working with the data set from the text, **Copier Maintenance (ex 1.20).** It is posted on Canvas for reference**.** Use them as a guide to work with the ***“Mass\_Calorie\_Data.csv”*** data set and answer the questions below.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Notes:**

We will start by redefining the variables and **fitting a regression model** (you may not always need to do this if the variables in the data set have clear labels):

|  |
| --- |
| X<-V2  Y<-V1 |
| **reg** <- lm(Y ~ X) |

Remember: It helps to define this as ***reg*** (or another chosen name) for easy use with later functions.

# To see a summary of regression inference, including t-tests, analysis of residuals, F-test, and R2 enter

# summary(reg)

The results will look as follows:

Residuals:

Min 1Q Median 3Q Max

-7.6309 -3.2500 -0.2383 4.0235 6.6309

Coefficients:

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

**(Intercept)** -2.3221 2.5644 -0.906 0.379

**X** 14.7383 0.5193 28.383 4.1e-15 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.482 on 16 degrees of freedom

Multiple R-squared: 0.9805, Adjusted R-squared: 0.9793

F-statistic: 805.6 on 1 and 16 DF, p-value: 4.097e-15

# To find confidence intervals for regression coefficients enter:

**confint(reg)**

2.5 % 97.5 %

(Intercept) -7.758337 3.114041

X 13.637480 15.839030

By default, we get **95%** confidence intervals for β0 and β1 (as seen above). The percentages listed are the ***cumulative*** probabilities for the values listed below them.

To use a different confidence level, C=0.9 for example, enter

confint(reg,**level=0.90**)

5 % 95 %

(Intercept) -6.799213 2.154917

X 13.831693 15.644817

To look at Inference about regression coefficients, we may want to define those coefficients to use in other functions. First, let us look at the names of the different components in the output. Type:

**names(reg)**

[1] "coefficients" "residuals" "effects" "rank" "fitted.values" "assign" "qr" "df.residual" [9] "xlevels" "call" "terms" "model"

To save the regression coefficients, and :

**reg$coefficients**

(Intercept) X

-2.322148 14.738255

To save just the **slope**:

slope<-**reg$coefficients**[2]

**slope**

X

14.73826

Recall, to **predict a response** value based on a specific level of , for example , type:

**predict**(reg, data.frame(X=3))

1

41.89262

To look at the **confidence interval for mean response** based on a specific level of , for example , enter:

**predict**(reg, data.frame(X=3), **interval="confidence"**)

fit lwr upr

1 41.89262 39.11027 44.67496

This provides the predicted value and lower and upper limits of the 95% confidence interval.

For the problems below, we want to fit a regression model that predicts Calorie Rate (*RATE*) in calories per day based on Lean Body Mass (*LBM*) in kg.

1. Start by fitting the regression model. Find the values of the least-squares estimators (regression coefficients) and write the equation of the LSR line.
2. Plot the regression line (in a **color other than black**) with the data. Include the plot here.
3. Find a 98% confidence interval for the slope, . Paste your results below. Verify that your lower and upper limits match those from the notes.

con

1. What does the 98% confidence interval for the intercept, , show? State the lower and upper limits and discuss what they mean in context.
2. Is there a positive association between LBM and Calorie Rate? Use . State the hypotheses. Provide the value of , the statistic and p-value. Do they match those found in the notes? Explain.
3. Note the value of “Residual standard error” shown in the summary of the regression results. What does this correspond to? (Hint: Look in the notes.) Based on this value, what is the value of the *MSE*?