موسسه آموزش عالی آزاد توسعه

برگزار کننده دوره‌های تخصصی علم داده

Tihe.ac.ir

Homework 3: Winter 2019

Due date: ۱۲ دی

Please email your HWs to y.zerehsaz@gmail.com

Please append all your codes to your response

*****Please hand in your HWs as a word file with your email as the document's name. For instance, I would name my word file as y.zerehsaz@gmail.com.docx.**

*****Make sure to copy and paste the codes that you used for each question. I need to see your plots, results and conclusions but not the long output of your codes.**

*****When asked, please explain your results.**

Silver-Zinc Batteries

Common low capacity primary button cell versions are typically called **Silver Oxide** batteries. Higher capacity versions available as secondary cells are more often referred to as **Silver Zinc** batteries. They have an open circuit voltage of 1.6 Volts. Two types of Silver Oxide batteries are available, one type with a sodium hydroxide (NaOH) electrolyte and the other with a potassium hydroxide (KOH) electrolyte. Because of the high cost of sliver, they are available in either very small sizes as button cells where the amount of silver used is small and not a significant contributor to the overall product costs or they are available in very large sizes for critical applications where the superior performance characteristics of the silver oxide chemistry outweigh any cost considerations.

The dataset battery.csv contains some data regarding a satellite application of Silver-Zing battery. The variables use in this dataset are

Cycles : cycles to failure
Charge: charge rate (amps)
Discharge: discharge rate (amps)

Depth: % of rated ampere hours
Temperature: temperature (C)
Endvolt: end of charge voltage (volts)

The purpose is to determine

What factors affect the life of this battery?
Can we build a model to predict the life of this battery?
Does the low temperature adversely affect the life of this battery?

A) Please read the battery.csv file and place it in a data frame called df.

B) Get the summary and structure of df.

C) Are there any missing values in the data? (temperature can be zero)

D) Regress the response variable "cycles" on the rest of variables using "lm" function. Interpret the coefficients.

E) Perform the following hypothesis tests and comment on the results.

i. $H_0: \beta_{charge} = \beta_{discharge} = \beta_{temp} = \beta_{endvolt} = \beta_{depth} = 0$

ii. Single hypothesis tests for all coefficients

You do not need to use anova, just use the output of your model's summary

F) Check the normality and constant variance assumptions of your model. For normality use the Anderson-Darling test. For constant variance assumption, you can plot residuals (or the absolute value of the residuals) versus the fitted values **(perform the hypothesis test if you have doubts about the plot)**

G) Check for leverage points. You need to check this in two different ways:

i. Use library "faraway" and check the half normal plot of h_i values (Slide 17, Lecture 4).

ii. Compute the studentized residuals and plot them against the fitted values (Slide 21, Lecture 4).

H) Check for outliers. You need to perform the following procedure (Slides 25 to 28, Lecture 4)

i. Compute the externally studentized residuals (t_i values)
ii. Choose a specific value for α and compute the Bonferroni corrected α depending on your sample size.
iii. Compute the p-values for all the t_i values obtained in (i).
iv. Find the observations rejecting your null hypothesis (p-value < Bonferroni corrected α). These are outliers.

I) Check the influential points. Compute the cook's distance and check the half-normal plot.
Do not remove the outliers. Just do Part J and see what happens!

J) Check whether or not there is multicollinearity on the data.

K) A standard technique to address heterogenous variance problem is to transform the response variable. Apply the "log" function on the response variable and compute a new response. Call the new response "lcycles". Respond to parts D to I using the new response variable.