**STAT 43000/STAT 53001 Applied Statistics Spring 2023**

**Test 1-Part II**

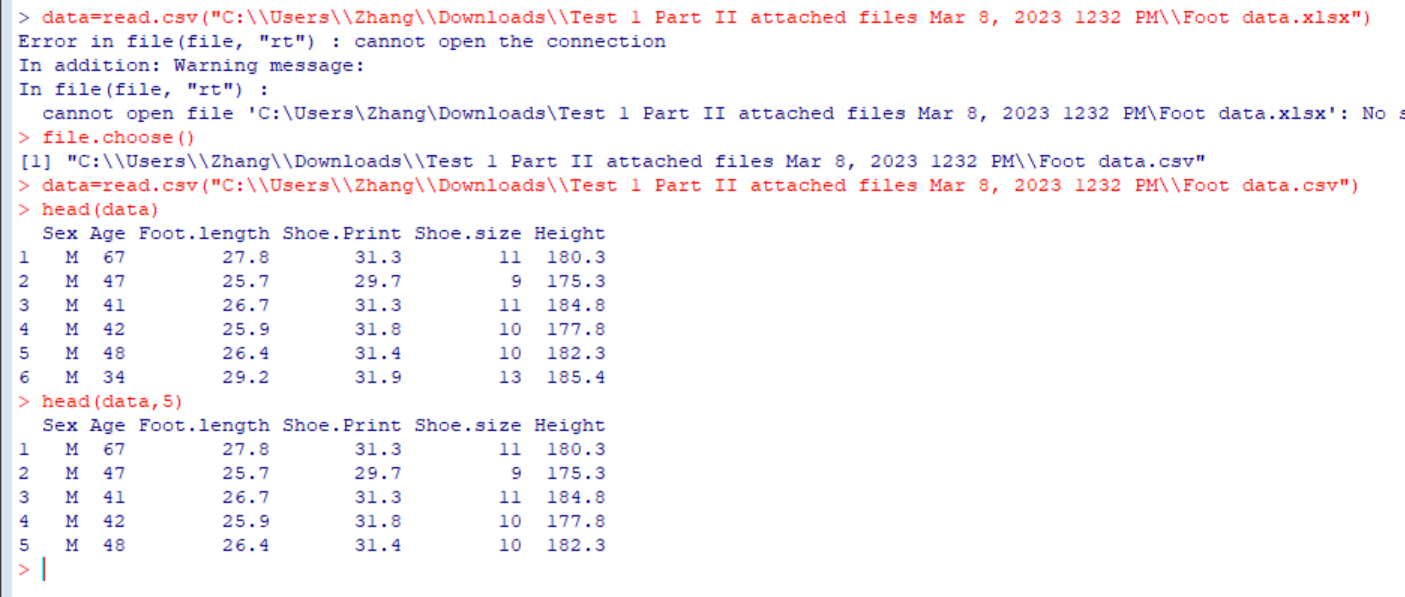
Due Date: March 8, 2023

Time: 2:00 PM (CST) Name:

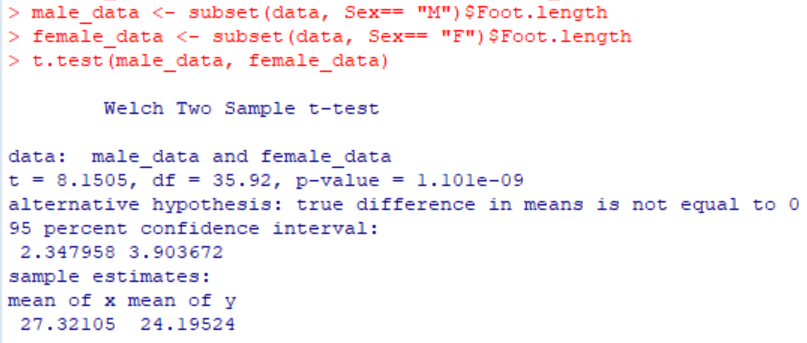
PUID:

**Q.N. 1 )** A data set containing data on 40 foot and height measurements of human is provided in the Brightspace (**Foot measurement**). This data is from ”Estimation of Stature from Foot Length: Applications in Forensic Science.

1. Import the data in R and print the first 5 observations.

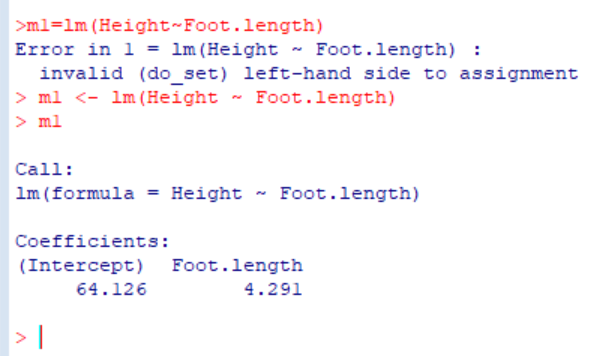


1. Is there a significant difference in the foot length of male and female?



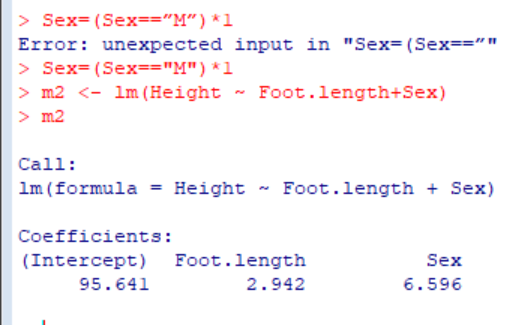
p value 1.101e-09 << 0.05, so there is a significant difference in the foot length of male and female

1. Fit a simple linear regression model using Foot length as a predictor variable and height as aresponse variable.



Height=4.291\*Foot.length+64.126

1. Update the fitted model in (c) by incorporating a binary variable Sex

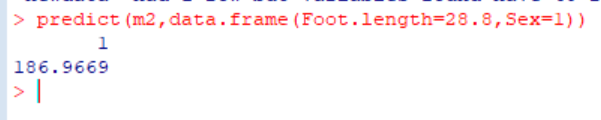


Set M=1 and F=0

So for male, Height=2.942 \*Foot.length+95.641+6.596

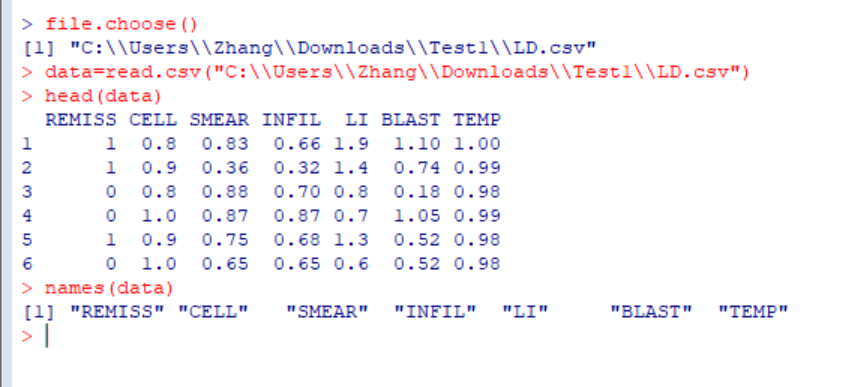
And for female, Height=2. 942 \*Foot.length+95.641

1. Predict the height of a male whose foot is 28.8 cm.

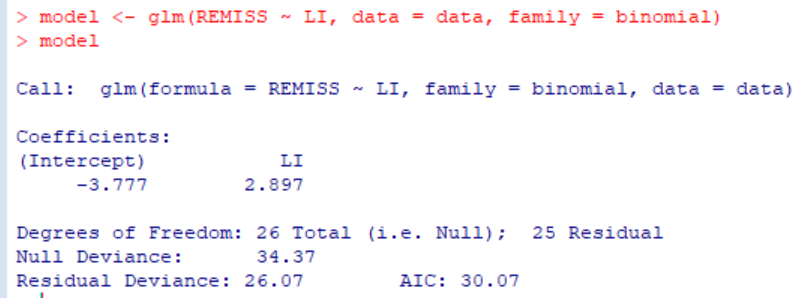


**Q.N. 2)** The **leukemia data** provided in the Brightspace provides the information of 27 patients.

The response variable of whether leukemia remission occurred (REMISS), which is given by a 1. a) Import the data in R and print the variables.

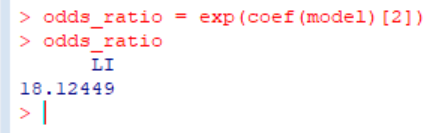


1. Fit a simple logistic regression model using percentage labeling index of the bone marrow leukemiacells (LI) as a predictor variable.

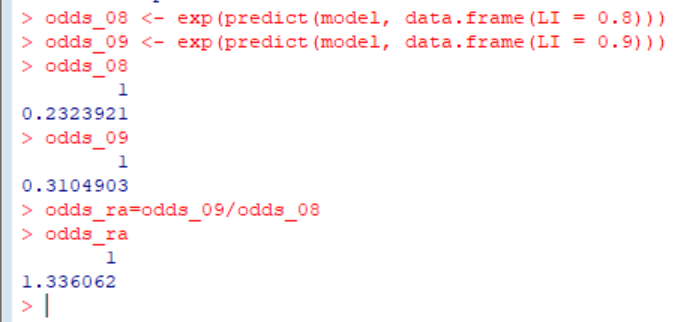


Π= =[1+exp(3.777-2.897LI)]

1. Calculate the odds ratio for LI.



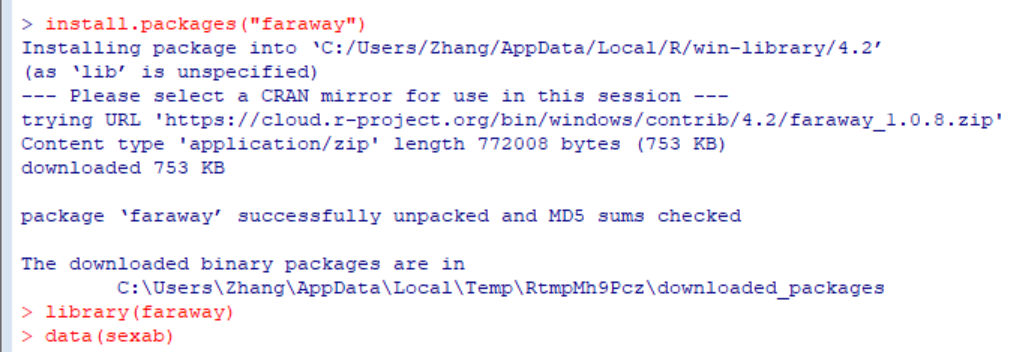
1. Calculate the estimated odds of leukemia remission at LI=0.8 and LI=0.9. Now, calculate theodds ratio using the odds at LI= 0.9 and LI=0.8. How do you interpret this value?

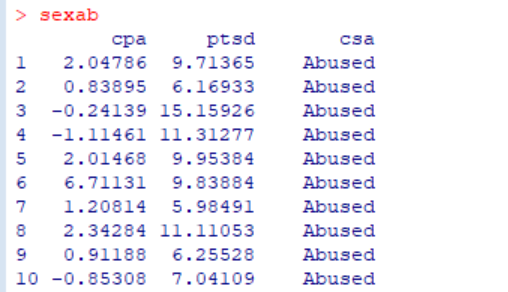


A value greater than 1 suggests that higher values of LI are associated with higher odds of leukemia

**Q.N. 3)** A data set sexab available in faraway package is related to a study of the effects of childhood sexual abuse on adult females reported by Rodriguez et al. (1997).

1. Install the library faraway and access the data sexab

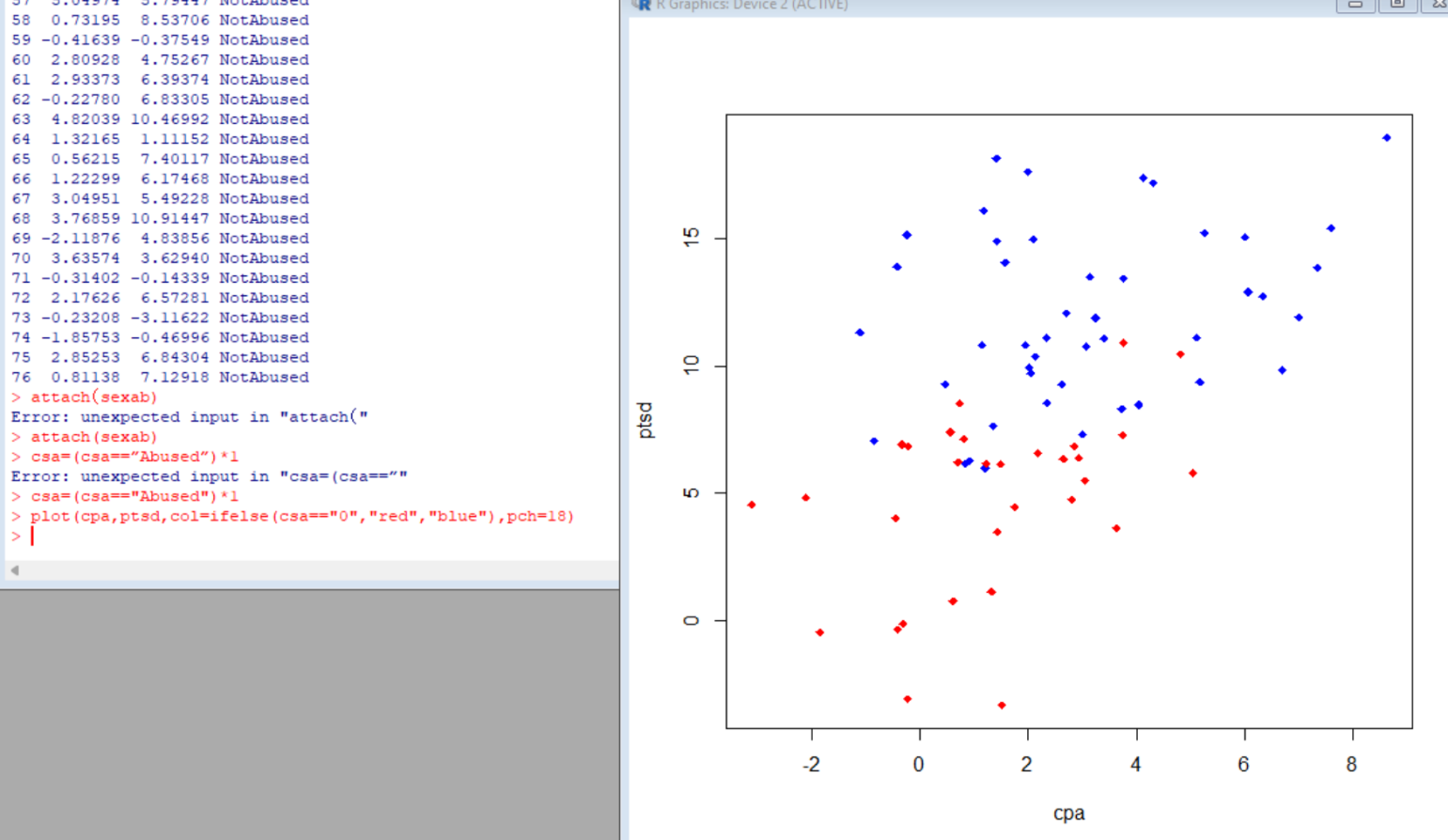




1. Note that the data include the variables:

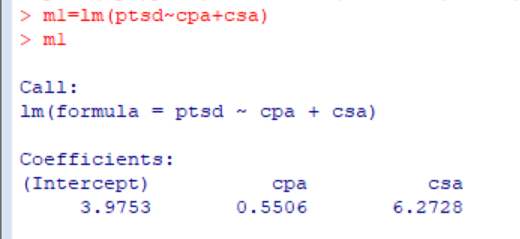
*cpa*-Childhood physical abuse on standard scale *csa*-Childhood sexual abuse - abused or not abused *ptsd*- Post-traumatic stress disorder on standard scale

Display the data graphically using different colors to csa variable: abused and not abused



1. Fit a linear regression model by choosing ptsd a response variable and using other variables as

Predictors

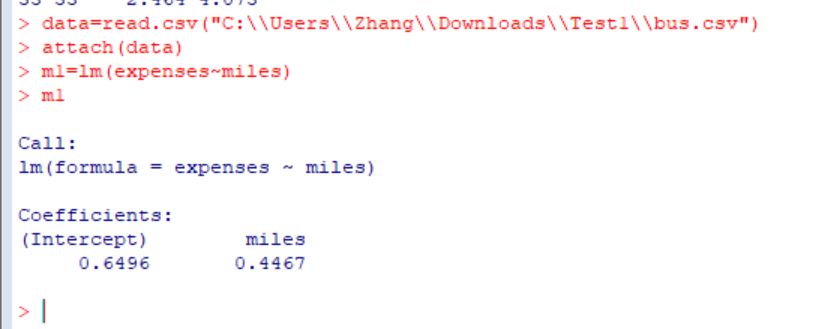


Ptsd=0.5506\*cpa+3.9753 (not abused)

Ptsd=0.5506\*cpa+3.9753 +6.2728(abused)

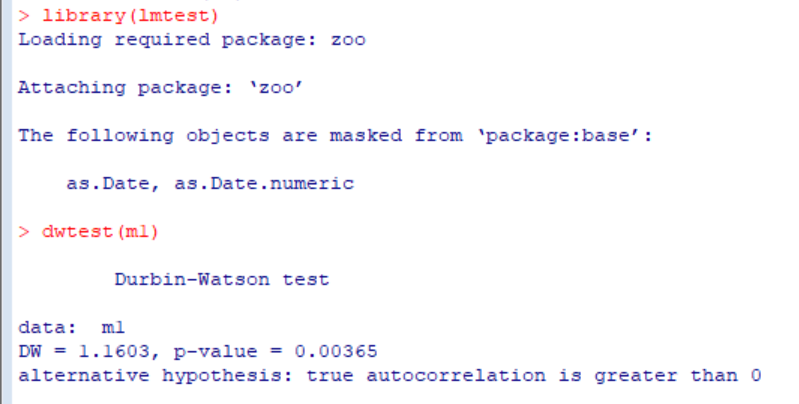
**Q.N. 4)** An economic study followed a British bus company for n = 33 time periods, recording y= Total Expenses (adjusted for inflation in 100,000s of pounds) and x=car miles(in millions). The data are available in the Brightspace (**Bus**)

1. Fit a simple linear regression model relating Total Expenses (y) to car miles (x).



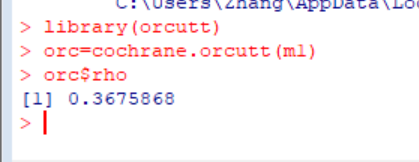
Expenses=0.4467\*miles+0.6496

1. Calculate the value of the Durbin-Watson test statistic. Do we have an evidence of auto correla-tion at *α* = 0*.*05.



Pvalue is much small than 0.05

1. Obtain estimates of the ˆ*ρ* based on the Cochrane-Orcutt procedure.



1. Obtain estimates of ˆ*ρ* based on the Hildreth-Lu procedure.

**Q.N. 5)** The transient points of an electronic inverter data are provided in the Brightspace as **inverter**. The variables under study are

y: Transient point (volts) of PMOS-NMOS inverters

X1: Width of the NMOS device

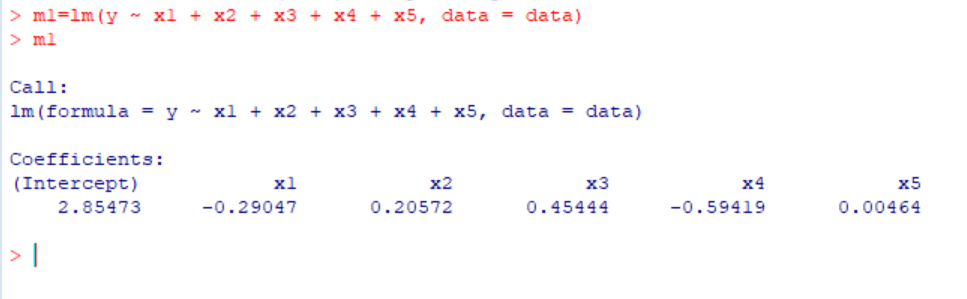
X2: Length of the NMOS device

X3: Width of the PMOS device

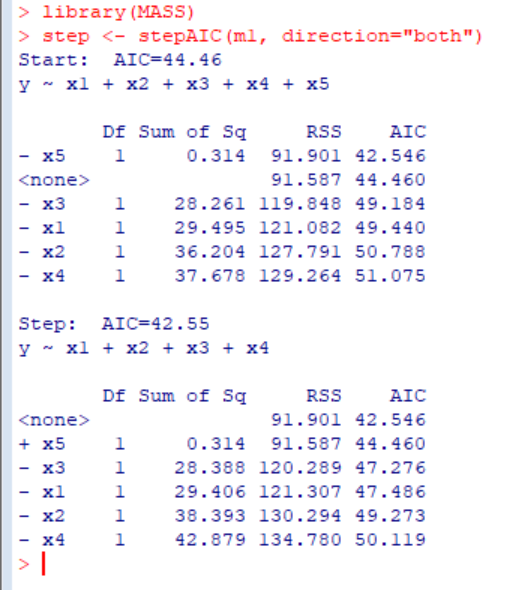
X4: Length of the PMOS device

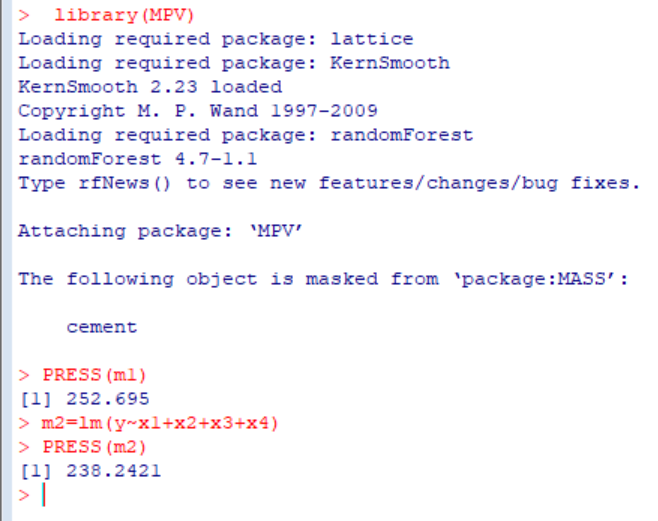
X5: Temperature (*oC*)

1. Fit a multiple linear regression model for this data.



1. Use stepwise regression criteria to find an appropriate regression model for these data .



1. Calculate the PRESS statistics for both models in (a) and (b). Which model would PRESSindicate is likely to be the best for predicting new response observations?  PRESS value of m2 is smaller so m2 is better

2