Bivariate analysis

Variance\_inflation\_factor

from statsmodels.stats.outliers\_influence import variance\_inflation\_factor -> this will import variance\_inflation factor function fromstatesmodels

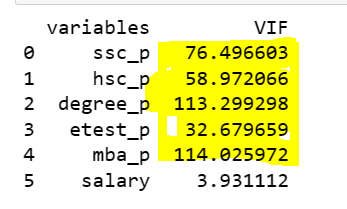
def calcvif(x): ->userdefind function to calculate multi collinearity dataframe with quantitative clumn are passed here

vif=pd.DataFrame() -> new empty dataframe is created

vif['variables']=x.columns ->in vif dataframe variables columns created it value will be all column name of input dataframe ex: ['ssc\_p', 'hsc\_p', 'degree\_p', 'etest\_p', 'mba\_p', 'salary']

vif['VIF']=[variance\_inflation\_factor(x.values,i)for i in range(x.shape[1])] -> variance\_inflation\_factor function is having dfvalues and which columns to consider , so based of the for loop each column values will be passed to function and it will be stored in vif column x.shape ill give number of columns

return vif -> finl dataframe will be returned



All are having multilinearity one can be predicted from other column except salary

Need to drop column which are having high vif and check others , high value means they can predicted by others.

Multicollinearity occurs when two or more independent variables in a data frame have a high correlation with one another in a regression model.

Multicollinearity can lead to unstable and unreliable coefficient estimates, making it harder to interpret the results and draw meaningful conclusions from the model. It is essential to detect and address multicollinearity to ensure the validity and robustness of regression models.

To predict multi co linearity we can use below

1.VIF

2.  regularization or feature selection

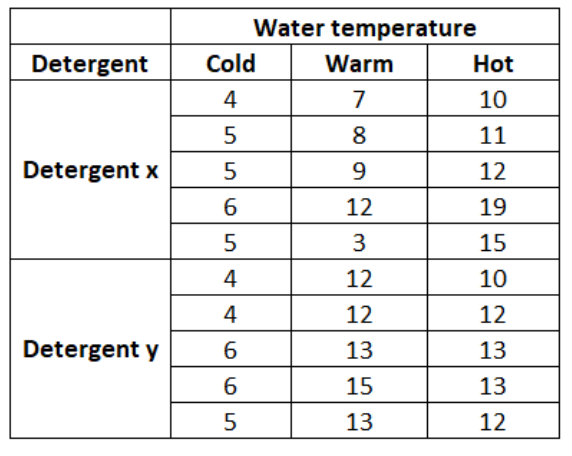
In homoscedasticity -> variance is equal among all error value

In heteroscedasticity -> variance is not equal among all error value we error are less amt few are more so it wont b aligned to mean value

Homoscedasticity is good because predicting good model will be easily because error is equal aligned to mean

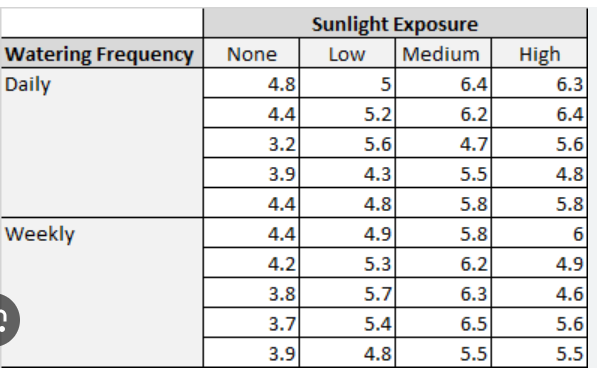
Hedro is difficult to get correct result

ANOVA



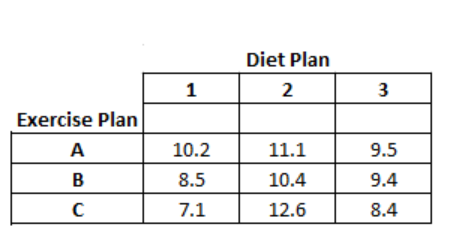
One way classification: find the difference of detergent x and y

Two way classification: find the difference of detergent and temperature



One way classification: find the watering frequency

Two way classification: find the significance of difference of watering frequency and sunlight exposure



One way classification: find the significance of difference of exercise plan accept at 1% level

Two way classification: find the significance of difference of exercise plan and diet plan accept at 1% level