Frequently Asked Questions in Advanced Statistics Project.

*** Please Go through the AS FAQs in detailed before raising any support request as we already have given all the necessary hints in project question and clear the doubts in this FAQ ***

- **How many files do I need to submit?**

2 files -

a. Business Report containing both the ANOVA & PCA answers (see instructions for more info)

b. Jupyter Files (ANOVA & PCA)

- **State the null and the alternate hypothesis for conducting one-way ANOVA for both Education and Occupation individually.**

In Q1.1 you have to correctly formulate null and the alternate hypothesis for conducting one-way ANOVA for both **Education** and **Occupation** individually

- **Do I need to check ANOVA assumptions?**

A. Please assume that the data qualifies all the assumptions for ANOVA. No need to check for assumptions.

- **Do I need to treat outliers in the ANOVA case study?**

A. No, it is not required.

- **What is the interaction between the two treatments? Analyse the effects of one variable on the other (Education and Occupation) with the help of an interaction plot.**

Use **Intersection_Plot** function from **statsmodel.graphics.factorplots** module and give inference from plot.

- **Perform Exploratory Data Analysis [both univariate and multivariate analysis to be performed]. What insight do you draw from the EDA?**

For Univariate, You need to analyse all the 17 variables as teach in the session and infer the result from it.

For Bi/Multivariate Analysis, You need to plot only **Correlation Heatmap** & **Pair plot** and infer the result from it.

- **Should I show the boxplot as one plot for variables?**

No. Always make separate boxplots when scales of variables are very different from each other.

- **Explain the business implications of performing ANOVA for this particular case study.**

You need to summarise the analysis of ANOVA on Problem 1 and write down the inference from it

- **Comment on the comparison between the covariance and the correlation matrices from this data. [on scaled data]"?**

You need to compare & show both the covariance and the correlation matrices of scaled data as either using **heatmap** or data frames and it is a must to write the interpretation from this comparison.

- **What do you mean by "Business Implications" of PCA?**

A. Please write about the interpretations of the PCs with respect to original features as taught in content videos.

- **Perform PCA and export the data of the Principal Component (eigenvectors) into a data frame with the original features**

There must be the Data Frame of all features and all PCs (up to the number of features).

- **Consider the cumulative values of the eigenvalues. How does it help you to decide on the optimum number of principal components? What do the eigenvectors indicate?**

You need to find the explained variance ratio and cumulative explained variance ratio and infer the results

- **Do I need to perform Outliers Treatment for Part 2: PCA?**

A. No, it is not required to perform outlier's treatment.

- **What does this mean "Perform PCA and export the data of the Principal Component (eigenvectors) into a data frame with the original features"?**

- Perform PCA
- Get the eigenvectors
- Put the eigenvectors into a data frame with All the PCs and Original Features.

- **Write down the explicit form of the first PC?**

You need to create a linear equation of PC in terms of eigenvectors and corresponding features with values rounded up to 2 decimal places
e.g. - 1.3* Variable 1 + 2.3* Variable 2 + 1.6* Variable 3 + 3.3* Variable 4 +.....

- **Extract the eigenvalues and eigenvectors?**

eigenvalues are **pca.explained_variance_** & eigenvectors are **pca.components_**

- **Can I put code in Business Report?**

No. You should not put code in the report. But you should put all the outputs & plots in the business report along with their respective inferences as per the rubric given.

- **What is a Business Report?**

A business report is a word document where all the questions are answered sequentially with all the relevant plots and outputs and their respective inferences. The business report word document can be converted to a pdf later and submitted. Both the .docx and .pdf are accepted.
Try to make your Business Report as professional as possible. It should have a proper Title Page, Table of Contents, Table of Figures, Proper Problem Headers (as in rubric), and respective answers under each. Even formatting should be taken care of. Just like you are submitting your Graduation Final Project Report.

- **Installation of the factor analyzer.**

Run the below code in Jupytur Notebook
pip install factor_analyzer
if this not work the try using
!pip install factor-analyzer --user
then restart the Jupytur Notebook (anaconda 3).