**ADVANCED STATISTICS PROJECT BUSINESS REPORT**

**Problem 1:**

**A research laboratory was developing a new compound for the relief of severe cases of hay fever. In an experiment with 36 volunteers, the amounts of the two active ingredients (A & B) in the compound were varied at three levels each. Randomization was used in assigning four volunteers to each of the nine treatments. The data on hours of relief can be found in the following .csv file:**[**Fever.csv**](https://olympus.greatlearning.in/courses/13136/files/909247/download?verifier=UV9URDqStNA73hrvKl23WB36i1pitQAb6larx7LC&wrap=1)

* 1. State the Null and Alternate Hypothesis for conducting one-way ANOVA for both the variables ‘A’ and ‘B’ individually. [both statement and statistical form like Ho=mu, Ha>mu]

**Solution:**

Hypothesis for A -

Ho : The relief hours do not differ by the levels of ingredient A

Ha : The relief hours differs by the levels of ingredient A

Ho : mu1 = mu2 = mu3

Ha : mu1 != mu2 != mu3

Hypothesis for B -

Ho : The relief hours do not differ by the levels of ingredient B

Hb : The relief hours differs by the levels of ingredient B

Ho : mu1 = mu2 = mu3

Hb : mu1 != mu2 != mu3

* 1. Perform one-way ANOVA for variable ‘A’ with respect to the variable ‘Relief’. State whether the Null Hypothesis is accepted or rejected based on the ANOVA results.

**Solution:**

Since p-value(4.578242e-07) < alpha(0.05), null hypothesis is rejected

Therefore, the relief hours significantly differ across the levels of the ingredient A

* 1. Perform one-way ANOVA for variable ‘B’ with respect to the variable ‘Relief’. State whether the Null Hypothesis is accepted or rejected based on the ANOVA results.

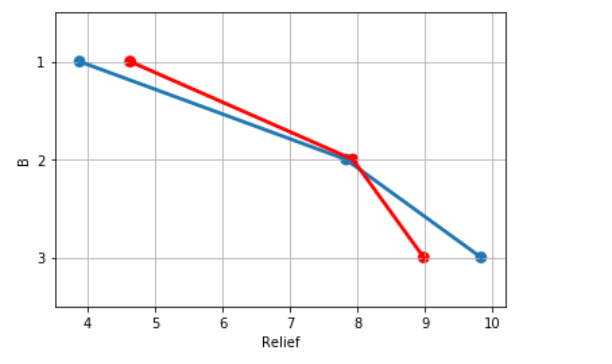
**Solution:**

Since p-value(0.00135) < alpha(0.05), null hypothesis is rejected

Therefore, the relief hours significantly differ across the levels of the ingredient B

* 1. Analyse the effects of one variable on another with the help of an interaction plot.  
     What is the interaction between the two treatments?  
     [hint: use the ‘pointplot’ function from the ‘seaborn’ function]

**Solution:**



The plot suggests that there is interaction between the levels of ingredient A and ingredient B.

* 1. Perform a two-way ANOVA based on the different ingredients (variable ‘A’ & ‘B’ along with their interaction 'A\*B') with the variable 'Relief' and state your results.

**Solution:**

Ho : There is no interaction between the levels of ingredient A and ingredient B

Ha : There is interaction between the levels of ingredient A and ingredient B

df sum\_sq mean\_sq F PR(>F)

C(A) 2.0 220.020 110.010000 1827.858462 1.514043e-29

C(B) 2.0 123.660 61.830000 1027.329231 3.348751e-26

C(A):C(B) 4.0 29.425 7.356250 122.226923 6.972083e-17

Residual 27.0 1.625 0.060185 NaN NaN

Since p-value(6.972083e-17) < alpha(0.05), null hypothesis is rejected

Therefore, we can conclude that there is significant interaction between the levels of ingredient A and ingredient B.

* 1. Mention the business implications of performing ANOVA for this particular case study.

**Solution:**

The business implications are -

1) It gives effect of different levels ingredient A on the relief hours

2) Similarly, it gives effect of different levels ingredient B on the relief hours

3) It gives effect on relief hours also when different levels of ingredients are used together in the compound.

Using this information manufacturer can add or remove proportion of ingredients from the compound for better/long relief.

**Problem 2:**

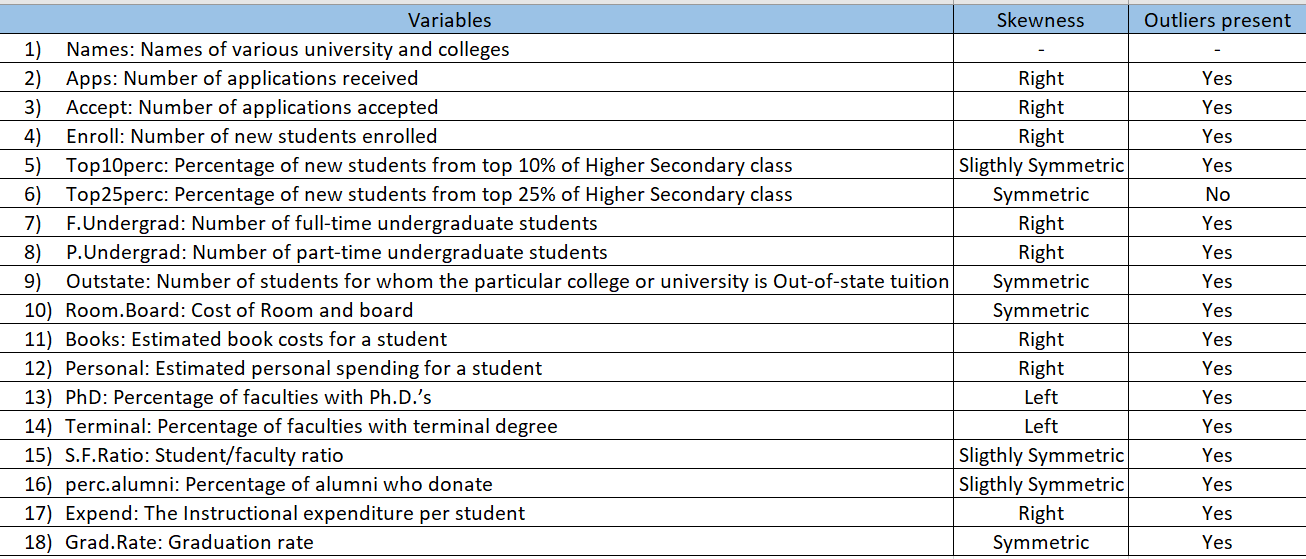
**The dataset**[**Education - Post 12th Standard.csv**](https://olympus.greatlearning.in/courses/13136/files/909246/download?verifier=Y3WOpNpJY1AbH2kCyKIXlRg6H0qpJ36fAFHDUTzE&wrap=1)**is a dataset that contains the names of various colleges. This particular case study is based on various parameters of various institutions. You are expected to do Principal Component Analysis for this case study according to the instructions given in the following rubric. The data dictionary of the 'Education - Post 12th Standard.csv' can be found in the following file:**[**Data Dictionary.xlsx**](https://olympus.greatlearning.in/courses/13136/files/909245/download?verifier=Y0n0S1jSl5A0YPz6YfEn45RIpUaDpnRkjfTFDKsi&wrap=1)**.**

2.1) Perform Exploratory Data Analysis [both univariate and multivariate analysis to be performed]. The inferences drawn from this should be properly documented.

**Solution:**

* **The data set has 777 observations and 18 variables in the data set.**
* **1 variable is of object datatype**
* **1 variable is of float datatype**
* **Rest 16 variables are of integer datatype**
* **There are no null values in the entire dataset**
* **There are no duplicate values**

**Univariate Analysis except class Names**



# Multi Variate Analysis

* The diagonal variables are highly correlated.
* The variables adjacent are also highly correlated
* Variables that are away from the diagonal are less correlated

2.2) Scale the variables and write the inference for using the type of scaling function for this case study.

**Solution:**

* Standard Scaler is used for scaling because outliers are present in most of the variables.
* Standard scaler will transform the data close to normal distribution.
* Which will further help to normalize most of the data in small intervals

2.3) Comment on the comparison between covariance and the correlation matrix.

**Solution:**

* Covariance matrix is equal to correlation matrix before scaling and also after scaling.
* With and Without standardisation, correlation matrix yields same result.

2.4) Check the dataset for outliers before and after scaling. Draw your inferences from this exercise.

**Solution:**

* Before scaling all variables were differently skewed and had outliers.
* After scaling the variables were transformed close to normal distribution and outliers were still present.
* After treating outliers using IQR method , outliers were brought to lower or upper whisker and now the all variables are almost symmetric

2.5) Build the covariance matrix, eigenvalues, and eigenvector.

**Solution:**

Covariance Matrix

%s [[ 3.92202587e-01 3.72045278e-01 3.44861102e-01 1.77882547e-01

2.28413447e-01 3.00372860e-01 1.53261035e-01 4.09309446e-02

1.16838149e-01 1.03276949e-01 1.29720101e-01 2.84046975e-01

2.67661215e-01 7.57322866e-02 -6.30526007e-02 9.90236070e-02

9.43096150e-02]

[ 3.72045278e-01 3.86719034e-01 3.57101106e-01 1.22742172e-01

1.70302604e-01 3.10747826e-01 1.67663695e-01 -3.11176328e-03

7.40055439e-02 9.06387432e-02 1.43597630e-01 2.59812502e-01

2.46777922e-01 1.12140752e-01 -1.02443762e-01 6.54924734e-02

4.90474515e-02]

[ 3.44861102e-01 3.57101106e-01 3.76969419e-01 9.32127566e-02

1.41572482e-01 3.30838988e-01 1.85453496e-01 -9.55996485e-02

-1.45695116e-02 8.66382517e-02 1.87681520e-01 2.29023882e-01

2.14034520e-01 1.61090704e-01 -1.36102632e-01 2.16676723e-02

-1.42555842e-02]

[ 1.77882547e-01 1.22742172e-01 9.32127566e-02 7.81305061e-01

8.08308058e-01 5.47613345e-02 -7.49074558e-02 4.97059995e-01

3.14346107e-01 9.47255125e-02 -9.29426027e-02 4.70148933e-01

4.40620887e-01 -3.28019751e-01 4.00987210e-01 3.78002594e-01

4.35749475e-01]

[ 2.28413447e-01 1.70302604e-01 1.41572482e-01 8.08308058e-01

1.00128866e+00 1.01001820e-01 -4.67766043e-02 4.90041139e-01

3.29590971e-01 1.18631669e-01 -7.82479424e-02 5.39489477e-01

5.19387111e-01 -2.84523563e-01 4.15135356e-01 3.73125871e-01

4.78619938e-01]

[ 3.00372860e-01 3.10747826e-01 3.30838988e-01 5.47613345e-02

1.01001820e-01 3.10315170e-01 1.82563117e-01 -1.26027953e-01

-3.01988039e-02 8.08715656e-02 1.80536480e-01 1.96913464e-01

1.83602609e-01 1.72926770e-01 -1.58267436e-01 1.34384529e-04

-4.57475459e-02]

[ 1.53261035e-01 1.67663695e-01 1.85453496e-01 -7.49074558e-02

-4.67766043e-02 1.82563117e-01 2.21636838e-01 -1.66812465e-01

-3.16881569e-02 4.02847905e-02 1.45904561e-01 5.87588044e-02

5.65696027e-02 1.66907512e-01 -1.96485156e-01 -6.18747157e-02

-1.24656448e-01]

[ 4.09309446e-02 -3.11176328e-03 -9.55996485e-02 4.97059995e-01

4.90041139e-01 -1.26027953e-01 -1.66812465e-01 1.00063964e+00

6.52513004e-01 3.56977546e-03 -2.93399069e-01 3.82701065e-01

4.05983213e-01 -5.48973759e-01 5.63250944e-01 5.04797524e-01

5.71836675e-01]

[ 1.16838149e-01 7.40055439e-02 -1.45695116e-02 3.14346107e-01

3.29590971e-01 -3.01988039e-02 -3.16881569e-02 6.52513004e-01

9.90307358e-01 7.56993714e-02 -1.96811114e-01 3.32218551e-01

3.71275537e-01 -3.58352687e-01 2.69792322e-01 3.76072607e-01

4.23126447e-01]

[ 1.03276949e-01 9.06387432e-02 8.66382517e-02 9.47255125e-02

1.18631669e-01 8.08715656e-02 4.02847905e-02 3.56977546e-03

7.56993714e-02 4.87714769e-01 1.50893084e-01 9.31226224e-02

1.09448483e-01 -5.70259896e-03 -2.97713834e-02 6.81740268e-02

-5.61443574e-03]

[ 1.29720101e-01 1.43597630e-01 1.87681520e-01 -9.29426027e-02

-7.82479424e-02 1.80536480e-01 1.45904561e-01 -2.93399069e-01

-1.96811114e-01 1.50893084e-01 8.11420703e-01 -1.02893106e-02

-2.83293174e-02 1.49863186e-01 -2.74121214e-01 -9.57249786e-02

-2.61665861e-01]

[ 2.84046975e-01 2.59812502e-01 2.29023882e-01 4.70148933e-01

5.39489477e-01 1.96913464e-01 5.87588044e-02 3.82701065e-01

3.32218551e-01 9.31226224e-02 -1.02893106e-02 9.55820539e-01

8.29899302e-01 -1.21012174e-01 2.42170983e-01 3.24864101e-01

3.02668013e-01]

[ 2.67661215e-01 2.46777922e-01 2.14034520e-01 4.40620887e-01

5.19387111e-01 1.83602609e-01 5.65696027e-02 4.05983213e-01

3.71275537e-01 1.09448483e-01 -2.83293174e-02 8.29899302e-01

9.67665090e-01 -1.42089297e-01 2.60463224e-01 3.35539361e-01

2.87625501e-01]

[ 7.57322866e-02 1.12140752e-01 1.61090704e-01 -3.28019751e-01

-2.84523563e-01 1.72926770e-01 1.66907512e-01 -5.48973759e-01

-3.58352687e-01 -5.70259896e-03 1.49863186e-01 -1.21012174e-01

-1.42089297e-01 9.15128308e-01 -3.92368123e-01 -4.07438139e-01

-2.94727250e-01]

[-6.30526007e-02 -1.02443762e-01 -1.36102632e-01 4.00987210e-01

4.15135356e-01 -1.58267436e-01 -1.96485156e-01 5.63250944e-01

2.69792322e-01 -2.97713834e-02 -2.74121214e-01 2.42170983e-01

2.60463224e-01 -3.92368123e-01 9.90599513e-01 2.99881835e-01

4.88405261e-01]

[ 9.90236070e-02 6.54924734e-02 2.16676723e-02 3.78002594e-01

3.73125871e-01 1.34384529e-04 -6.18747157e-02 5.04797524e-01

3.76072607e-01 6.81740268e-02 -9.57249786e-02 3.24864101e-01

3.35539361e-01 -4.07438139e-01 2.99881835e-01 4.23629634e-01

2.69919979e-01]

[ 9.43096150e-02 4.90474515e-02 -1.42555842e-02 4.35749475e-01

4.78619938e-01 -4.57475459e-02 -1.24656448e-01 5.71836675e-01

4.23126447e-01 -5.61443574e-03 -2.61665861e-01 3.02668013e-01

2.87625501e-01 -2.94727250e-01 4.88405261e-01 2.69919979e-01

9.97192469e-01]]

Eigen Vectors

%s [[ 0.0929684 0.32104652 0.06660652 -0.0129432 0.24674827 -0.00650339

0.2400326 -0.13180129 -0.01773119 -0.03400089 0.14346037 -0.59269472

0.5569348 0.03784437 0.22445438 0.11116423 -0.00914552]

[ 0.06592707 0.3319699 0.07883241 -0.03420729 0.22877472 -0.02487384

0.27288698 -0.12917757 -0.0195307 -0.06133927 -0.32336365 0.70710813

0.26755069 0.00907891 0.17576693 0.15058739 0.00281244]

[ 0.03166929 0.35033549 0.01381154 -0.01122623 0.19114851 -0.03094669

0.26523924 -0.12023338 -0.00760842 -0.00639234 0.69930928 0.13493584

-0.49315933 -0.01188448 0.04255122 0.01589244 0.02985794]

[ 0.33452816 0.06754279 -0.32328505 0.21141739 0.07741651 0.32096376

-0.09974811 0.02170441 0.13379303 -0.00700079 -0.0310452 0.02159112

0.00901027 0.08516089 0.14064341 -0.29511711 0.69375696]

[ 0.36427546 0.13142781 -0.41399578 0.19443136 0.11797053 0.37686172

-0.18713202 0.01531995 0.18486533 0.12749803 0.00890279 0.01682207

-0.00267915 -0.13131824 -0.18030241 0.29567976 -0.511526 ]

[ 0.01149875 0.32452837 0.02193807 -0.01744947 0.15029942 -0.01698553

0.21021832 -0.10021068 -0.01175077 0.02574847 -0.61814444 -0.35140689

-0.55093616 -0.01666731 -0.06116207 -0.05081468 -0.01136756]

[-0.04622402 0.20972858 0.1038968 -0.02676078 0.06989958 -0.00474311

0.08937877 -0.0537958 0.04181848 0.13366948 0.04589764 0.03885668

0.23891433 -0.08576745 -0.8706629 -0.24738219 0.14019587]

[ 0.37830181 -0.20665209 0.2446006 0.02000679 0.04640648 -0.05172404

0.05471221 -0.02283038 0.17673816 -0.75697305 -0.00367566 -0.04554559

-0.07289772 -0.0571668 -0.21583357 0.27224079 0.09997777]

[ 0.29777508 -0.07383062 0.65435548 -0.07736692 0.20589468 0.0558972

-0.31550426 -0.11989134 0.33073591 0.42976697 0.00556533 0.01161555

-0.06810512 0.05833278 0.09487917 0.01923638 0.02476745]

[ 0.0401252 0.13395669 0.06932308 0.29066279 0.05440207 -0.08091173

-0.50411092 -0.48393685 -0.6116081 -0.10870293 0.00416292 0.00902118

-0.01670698 0.0378997 -0.05038066 0.04789243 0.02826868]

[-0.10944135 0.29386253 0.02906196 0.60616613 -0.01326297 -0.52880595

-0.19661216 0.33142515 0.31893212 -0.03384734 -0.00630645 0.0047103

0.02161283 -0.01497856 0.03722249 0.00703647 -0.0047994 ]

[ 0.31241468 0.30728219 0.01051885 -0.21336602 -0.44256735 -0.07682269

-0.04689621 0.18837226 -0.0983305 0.02010444 0.00804 0.00548016

-0.00454371 0.70146927 -0.09315117 0.10037346 -0.0434578 ]

[ 0.31563577 0.28923834 0.07534077 -0.22034156 -0.48498252 -0.10434108

-0.06899073 0.09780965 -0.12229609 0.06046494 0.0048878 -0.00898686

0.02707478 -0.67932403 0.11394885 -0.01716347 0.08863951]

[-0.238936 0.27738993 -0.19726187 -0.50833033 0.128203 -0.06989583

-0.48926256 -0.16012666 0.36247736 -0.31211523 0.00662521 0.01073745

0.02278083 0.01040467 0.06491485 -0.20799672 -0.07241455]

[ 0.28566756 -0.26160327 -0.35032544 -0.05443422 -0.08731305 -0.56468516

0.1645982 -0.52854784 0.21127982 0.21023249 -0.01383698 0.00137953

0.01480452 0.03469458 -0.02561499 -0.01145122 0.00883861]

[ 0.24699418 -0.02920582 0.14253472 0.17754101 -0.0657708 0.05988252

0.1378192 -0.05801938 -0.02433122 -0.20131984 0.01193343 0.04641723

0.04412393 0.04429777 0.12534642 -0.76406889 -0.45737401]

[ 0.31117828 -0.12680266 -0.14343185 -0.26409767 0.54379453 -0.34192987

-0.11257333 0.47485834 -0.35725708 0.05772333 -0.00123347 0.00788293

-0.01406045 -0.03267103 -0.03283629 -0.10963549 -0.02419997]]

Eigen Values

%s [4.75579369 2.3800885 0.88497491 0.81453646 0.72423975 0.52688069

0.47958062 0.41127635 0.36620193 0.23942458 0.00793972 0.01435481

0.03582106 0.12943793 0.09751277 0.08189987 0.06059116]

2.6) Write the explicit form of the first PC (in terms of Eigen Vectors).

**Solution**:

PC0 = 0.09\*Apps + 0.07\*Accept + 0.03\*Enroll + 0.33\*Top10perc + 0.36\*Top25perc + 0.01\*F.Undergrad + -0.05\*P.Undergrad + 0.38\*Outstate + 0.30\*Room.Board + 0.04\*Books + -0.11\*Personal + 0.31\*PhD + 0.32\*Terminal + -0.24\*S.F.Ratio + 0.29\*perc.alumni + 0.25\*Expend + 0.31\*Grad.Rate

2.7) Discuss 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?  
Perform PCA and export the data of the Principal Component scores into a data frame.

**Solution**:

* First eight components explain 91.39% of variance in data
* The optimum number of principal components would be 8.
* The eigenvectors indicate eigen values which are insignificant can be removed from analysis.

2.8) Mention the business implication of using the Principal Component Analysis for this case study. [**Hint:**Write Interpretations of the Principal Components Obtained]

**Solution:**

It we employ 8 features we can capture 91% of the variance in the dataset

