**Mean, variance and standard diviation of the dataset :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **STG** | **SCG** | **STR** | **LPR** | **PEG** |
| **Mean** | **0.37** | **0.35** | **0.46** | **0.43** | **0.45** |
| **Standard div** | **0.21** | **0.21** | **0.24** | **0.24** | **0.25** |
| **Variance** | **0.044** | **0.044** | **0.057** | **0.057** | **0.062** |

**For each one of 50 sample :**

Mean :

[0.39 0.45 0.27 0.36 0.55 0.33 0.26 0.32 0.39 0.43 0.47 0.37 0.42 0.45

0.4 0.36 0.28 0.39 0.34 0.44 0.4 0.34 0.25 0.34 0.43 0.28 0.41 0.43

0.33 0.39 0.45 0.24 0.41 0.3 0.43 0.33 0.29 0.26 0.36 0.34 0.43 0.41

0.34 0.25 0.37 0.35 0.39 0.4 0.35 0.37]

Variance :

[0.03 0.01 0.02 0.02 0.03 0.09 0.03 0.05 0.09 0.05 0.07 0.07 0.04 0.02

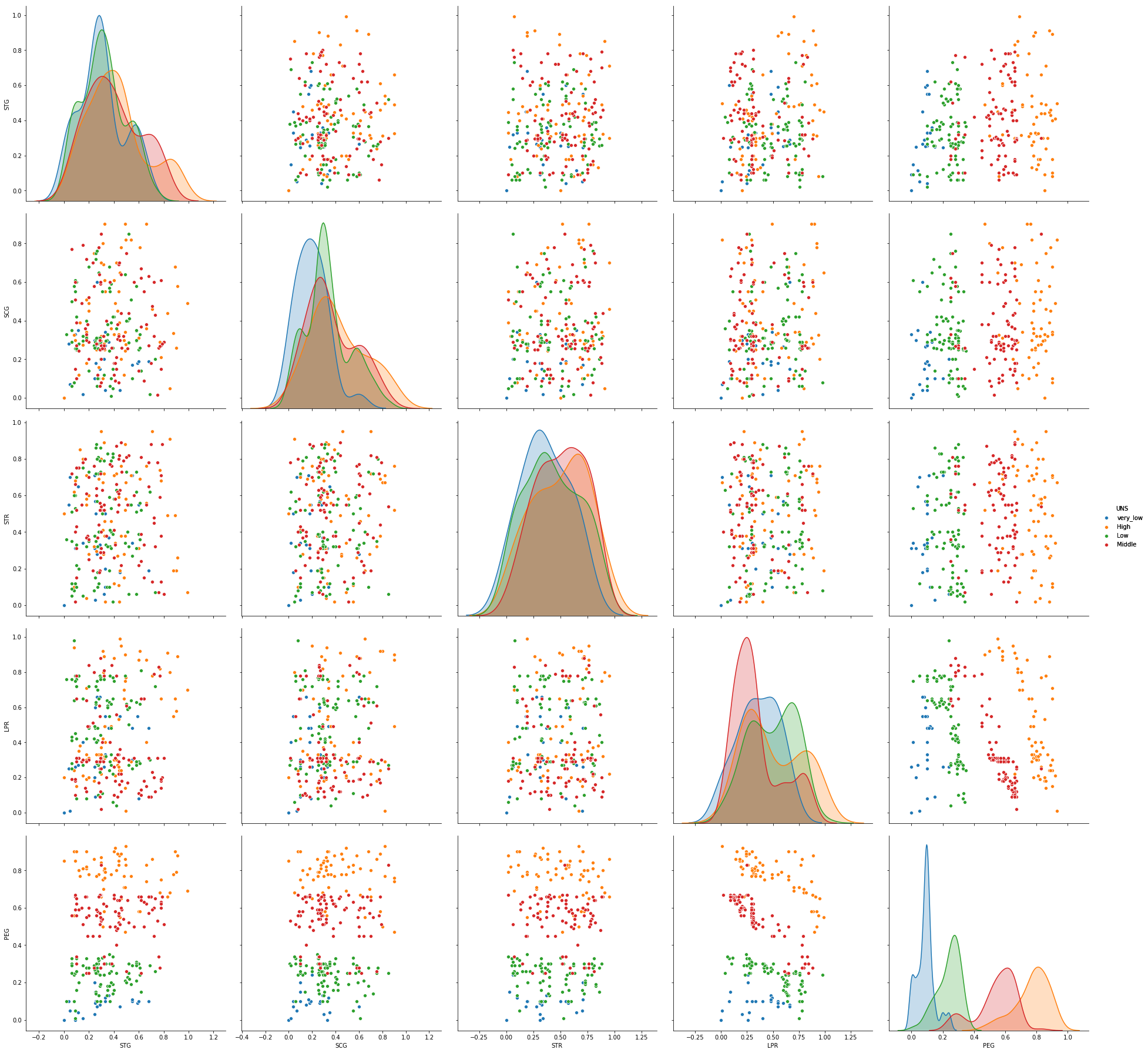
0.04 0.04 0.02 0.04 0.02 0.06 0.06 0.01 0.03 0.06 0.05 0.03 0.06 0.04

0.02 0.02 0.02 0.04 0.04 0.05 0.07 0.02 0.05 0.05 0.05 0.02 0.08 0.04

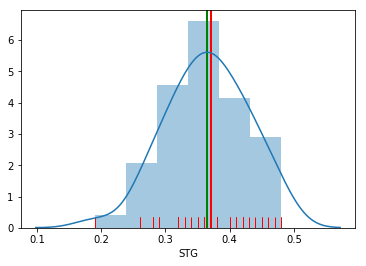
0.06 0.01 0.09 0.06 0.03 0.09 0.01 0.02]

**Covariance matrix :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **STG** | **SCG** | **STR** | **LPR** | **PEG** |
| **STG** | **0.044214** | **0.003612** | **0.002112** | **0.005193** | **0.011074** |
| **SCG** | **0.003612** | **0.044928** | **0.004365** | **0.005144** | **0.009888** |
| **STR** | **0.002112** | **0.004365** | **0.060486** | **0.002226** | **0.007581** |
| **LPR** | **0.005193** | **0.005144** | **0.002226** | **0.061558** | **-0.017069** |
| **PEG** | **0.011074** | **0.009888** | **0.007581** | **-0.017069** | **0.065133** |



**Mean and samples mean for STG**



**The red line is the dataset mean.**

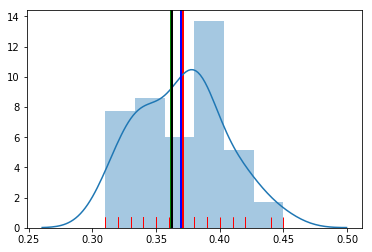
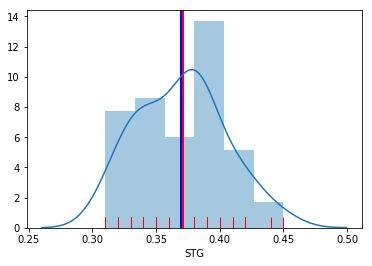
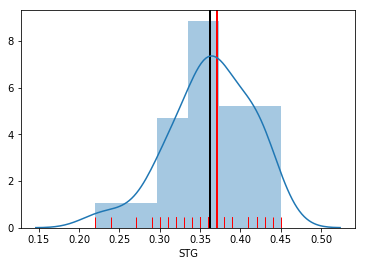
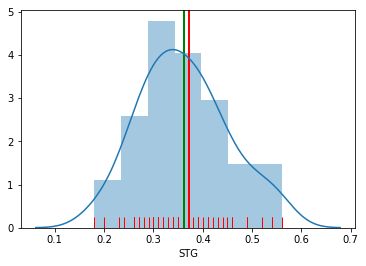
**The green line is the mean of samples mean.**

**This plot shows the difference between the mean of the samples mean and the dataset mean for STG.**

**The maen of the sample means of STG is : 0.3501**

**The mean of the dataset mean of STG is : 0.3711**

**Mean of samples means at different sample sizes (5 is green , 15 is black , 25 is blue) compared to real dataset mean of STG (red)**

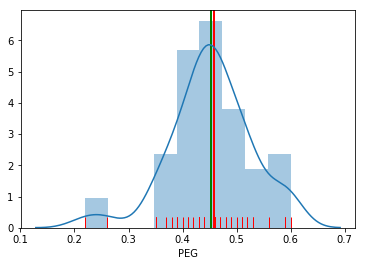


**Sample Means at sample size of 5 - variance : 0.0078**

**Sample Means at sample size of 15 - variance : 0.0025**

**Sample Means at sample size of 25 - variance : 0.0011**

**Mean and samples mean for PEG**



**The red line is the dataset mean.**

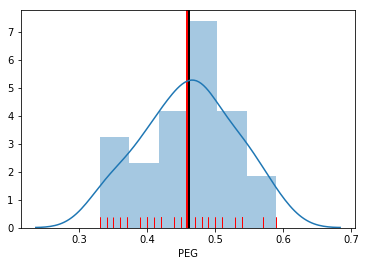
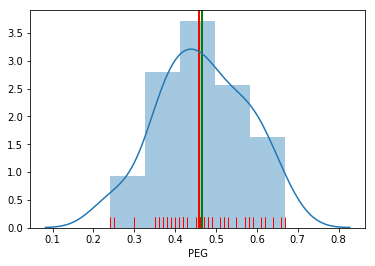
**The green line is the mean of samples mean.**

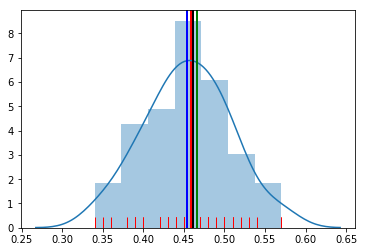
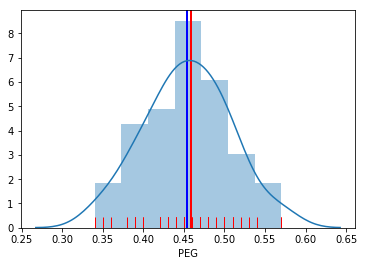
**This plot shows the difference between the mean of the samples mean and the dataset mean for PEG.**

**The maen of the sample means of PEG is : 0.4553**

**The mean of the dataset mean of PEG is : 0.4520**

**Mean of samples means at different sample sizes (5 is green , 15 is black , 25 is blue) compared to real dataset mean of PEG (red)**





**Sample Means at sample size of 5 - variance : 0.0114**

**Sample Means at sample size of 15 - variance : 0.0045**

**Sample Means at sample size of 25 - variance :** **0.0027**

**So we notice that the variance of samples means decreases as the size of the sample increases because the more observations the sample contains the more it gets closer to the full dataset and the spread of means becomes tight so the variance gets lower.**

**The PEG attribute tends to have the highest Covariance over all the other attributes.**

**The PEG has high covariance with all other attributes compared to the covariance of all the attributes compared to each other.**

**We could say that PEG has the greatest impact for the data to determine the UNS level.**

**We could obviously say that exam performance has the highest impact to the knowledge of the user and other attributes**

|  |
| --- |
| Attribute Information: |
| STG (The degree of study time for goal object materails), |
| SCG (The degree of repetition number of user for goal object materails) |
| STR (The degree of study time of user for related objects with goal object) |
| LPR (The exam performance of user for related objects with goal object) |
| PEG (The exam performance of user for goal objects) |
| UNS (The knowledge level of user) |

|  |
| --- |
| Data Set Characteristics: Multivariate |
| Number of Instances: 403 |
| Area: Education |
| Attribute Characteristics: Real |
| Number of Attributes: 5 |
| Associated Tasks: Classification |
|  |
| Class Distribution |
| Very Low: 50 |
| Low:129 |
| Middle: 122 |
| high 130 |
| Total:403 |