**Group 1 - Case Study #1**

|  |  |
| --- | --- |
| Kyle McGonigle - [kmcgonigle@horizon.csueastbay.edu](mailto:kmcgonigle@horizon.csueastbay.edu) | Joel Saetern - [jsaetern@horizon.csueastbay.edu](mailto:jsaetern@horizon.csueastbay.edu) |
| Dinesh Saraswat - [dsaraswat2@horizon.csueastbasy.edu](mailto:dsaraswat2@horizon.csueastbasy.edu) | Abhinav Baad - [abaad@horizon.csueastbay.edu](mailto:abaad@horizon.csueastbay.edu) |

**Question 1 →**

*Include a summary (mean, median, standard deviation, variance, range, and inter-quartile*

*range) for each quantitative variable in the dataset. Use a table to display the summary.*

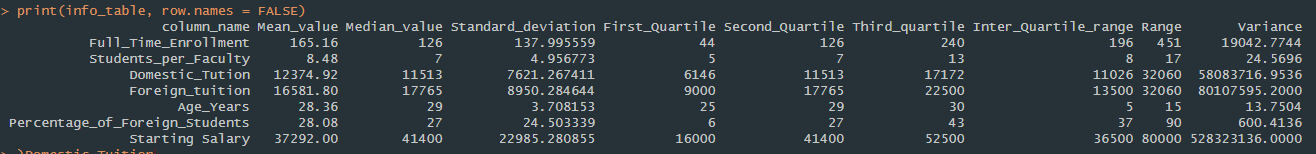
*measures of all the quantitative variables in one place. What insights do these descriptive*

*statistics provide regarding the Asia-Pacific business schools? (10 + 5 = 15 points)*

* **1A *- Summary*** *- we included a screenshot of the R-Studio along with the prettier version that was written to an excel file using a write to csv function in R.*

***NOTE*** *-* We treated the data sample we received as if this was not a population. Because this was a sample of 25 Asian universities, we used the formula for a sample standard deviation. Therefore, we had one less degree of freedom and did not use the built in sd() function in R and did these calculations ourselves. Kindly refer the R code file.

* ***R - Studio***

**

* ***Excel***

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***column\_name*** | ***Mean\_value*** | ***Median\_value*** | ***Standard\_deviation*** | ***First\_Quartile*** | ***Second\_Quartile*** | ***Third\_quartile*** | ***Inter\_Quartile\_range*** | ***Range*** | ***Variance*** |
| ***Full\_Time\_Enrollment*** | *165.16* | *126* | *137.995559348843* | *44* | *126* | *240* | *196* | *451* | *19042.7744* |
| ***Students\_per\_Faculty*** | *8.48* | *7* | *4.95677314389109* | *5* | *7* | *13* | *8* | *17* | *24.5696* |
| ***Domestic\_Tution*** | *12374.92* | *11513* | *7621.26741123811* | *6146* | *11513* | *17172* | *11026* | *32060* | *58083716.9536* |
| ***Foreign\_tuition*** | *16581.8* | *17765* | *8950.28464351833* | *9000* | *17765* | *22500* | *13500* | *32060* | *80107595.2* |
| ***Age\_Years*** | *28.36* | *29* | *3.70815317914457* | *25* | *29* | *30* | *5* | *15* | *13.7504* |
| ***Percentage\_of\_Foreign\_Students*** | *28.08* | *27* | *24.5033385480428* | *6* | *27* | *43* | *37* | *90* | *600.4136* |
| ***Starting Salary*** | *37292* | *41400* | *22985.2808553648* | *16000* | *41400* | *52500* | *36500* | *80000* | *528323136* |

* **1B *- Insights*** *-* What insights do these descriptive statistics provide regarding the Asia-Pacific business schools?
* **Foreign Tuition -** is almost 1.5 times the domestic tuition.
* **Age -** of the average student is around 29 which indicates that many of the students start graduate school after they’ve worked for a while.
* **Foreign students -** around ~28% of the students are foreign students.
* **Number of students to faculty-** is about 8.
* **Starting salary -** has a high range and high standard deviation for starting salary - which is because some colleges have low starting salaries and 1-2 may be a very high starting salary which is obvious by the fact that the mean is around 38000 salary.

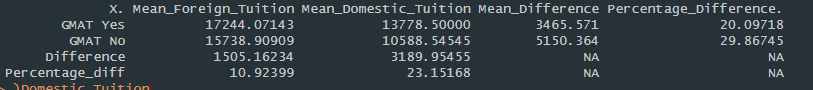
**Question 2 →**

Summarize the data to compare the following:

* **2A** - Mean differences between domestic and foreign tuitions for schools requiring

versus not requiring GMAT. (5 points)

* Here you can see that the cost is higher for universities that require GMAT for both foreign and domestic tuition. This lines up with the disparity in tuition here at CSUEB for international students.
  + Domestic GMAT is 23.15 % higher than Domestic NO GMAT
  + Foreign GMAT is 10.92 % higher than Foreign NO GMAT



* **2B** - Any difference between mean starting salaries for schools requiring versus not requiring work experience.
* Here the data shows that salaries of students that were required to have work experience prior to attending the university were about 68% higher. We interpret this to mean that people who have work experience are generally further along in their career.

A close up of white text

Description automatically generated

* **2C** - Any difference between mean starting salaries for schools requiring versus not requiring English test.
  + Here the data shows that salaries are inflated for those who have attended a university where an English test is required were about 34% higher. We interpret this to mean that people who have taken English tests are more proficient in English and have higher marketability to the job market worldwide.



**Question 3 →**

Does starting salary appear to be related to domestic tuition? Does starting salary appear to be associated with foreign tuition? Justify using both visual and numerical measures. (5 + 5 = 10 points)

* **3A** -Domestic Tuition
* **Numerical** - the correlation coefficient is at 78%, the closer the correlation coefficient is, the stronger the relationship is. Therefore, numerically this shows a rather strong positive relationship between domestic tuition and starting salary. Implying - schooling helps you get paid more.
* **Visual** - looking at the scatter plot, it shows that generally, the more the x - axis increases (tuition), the more the y axis increases (starting salary). The data points follow this trend with many points clustered to the trend line.

A graph with numbers and a line

Description automatically generated

* **3B** -Foreign Tuition
* **Numerical** - the correlation coefficient is at 66%, the closer the correlation coefficient is, the stronger the relationship is. Therefore, numerically this shows a rather strong positive relationship between foreign tuition and starting salary. Implying - schooling helps you get paid more but not as strongly correlated as domestic tuition.
* **Visual** - looking at the scatter plot, it shows that generally, the more the x - axis increases (tuition), the more the y axis increases (starting salary). The data points follow this trend with many points clustered to the trend line.

A graph with numbers and a line

Description automatically generated

**Question 4 →**

Draw a boxplot to graphically summarize the starting salary data. Identify the first, second and third quartiles from the boxplot. Are there any outliers? (4 + 4 + 2 = 10 points)

* **4A** - Draw a boxplot to graphically summarize the starting salary data.
  + Below is a boxplot drawn to summarize the data

A grey rectangular object with black lines

Description automatically generated

* **4B**- Identify the first, second and third quartiles from the boxplot.
  + First Quartile - Q1 - 16000
  + Second Quartile - Q2 - 41400
  + Third Quartile - Q3 - 52500
* **4C** - Are there any outliers.
  + No there are no outliers - using the formulas below we find out lower and upper limit and we deduce that no salary values lie outside the upper or lower limit.
    - Lower Limit = Q1 - 1.5(IQR) = -38750
    - Upper Limit = Q3+ 1.5(IQR) = 107250