

Project Report

FAST National University of Computer and Emerging Sciences

Statistical Analysis of Admission 2019 of NUCES

Submitted by:

Muhammad Waleed Usman

i192140

Submitted to:

Dr. Hammad Majeed

Q1

Description:

We were told that we have to find the relation/correlation between the students who perform good in matric exams and get offer admission by university. There are two columns/programs who get offer from the university. First thing to do is that we have to join these two columns, by putting data of one column in another let's say that column as "Offer". Now let's again focus on the "Offer" column, there are null cells means students were not selected, for the filled cells students were selected, I replace nulls with "Not Selected" and remaining with "Selected". Now we can see that from the matriculation marks that, total number and obtain marks are different of every student, for our easy let's take percentages. Once we are done with percentages. Another question rises here we cannot find probability/correlation between numerical and categorical data. Now let's change percentages into grades, so that we can easily access our statistics. Now both the data are in categorical form. We can easily interpret results.

Conclusion:

The average percentages of the students in matriculation marks is 76.4 so from the calculated probability the students who got greater than or equal to 80% have a very little chance with a probability of 0.34936460, whereas for percentages greater or equal to 70 also have way too little chance of being selected in university having probability 0.16207277. From this we can conclude that the student who got good grades in matriculation marks does not have a better chance of being selected in the university.

Results:

	Not Selected	Selected
A Grade	0.6506354	0.3493646
B Grade	0.8378233	0.1621767
C Grade	0.9426325	0.0573675

Barplot for Grades in Matric/O-level and Selection in Nukes

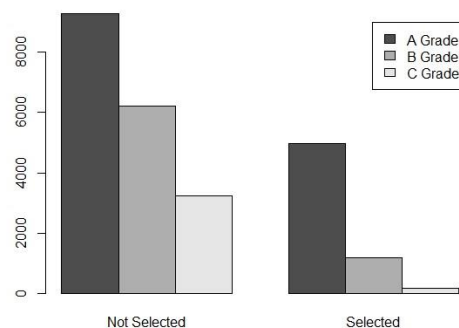


Figure 1 Question 1

Q2

Description:

We were told that we have to find the relation/correlation between the students who perform good in Intermediate exams and get offer admission by university. There are two columns/programs who get offer from the university. First thing to do is that we have to join these two columns, by putting data of one column in another let's say that column as "Offer". Now let's again focus on the "Offer" column, there are null cells means students were not selected, for the filled cells students were selected, I replace nulls with "Not Selected" and remaining with "Selected". Now we can see that from the Intermediate marks that, total number and obtain marks are different of every student, for our easy let's take percentages. Once we are done with percentages. Another question rises here we cannot find probability/correlation between numerical and categorical data. Now let's change percentages into grades, so that we can easily access our statistics. Now both the data are in categorical form. We can easily interpret results.

Conclusion:

The average percentages of the students in Intermediate marks is 105.7056, means there is some data which is not correct, lets ignore them and make our grades, so the students who got greater than or equal to 80% have a very little chance with a probability of 0.2817347, whereas for percentages greater or equal to 70 also have way too little chance of being selected in university having probability 0.1819277.

From this we can conclude that the student who got good grades in Intermediate marks does not have a better chance of being selected in the university.

Results:

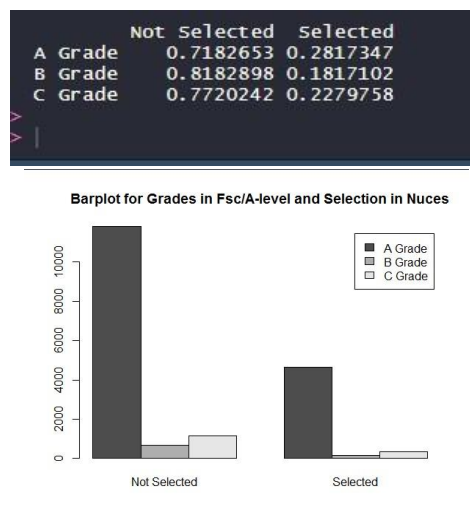


Figure 2 - Question 2

Q3

Description:

We were told to check probability that If the students got admission in their first preference joined university. If we think, we have to compare the values of the first preference in the discipline with that "Offer" column which I have discussed in Q1. Let's make a variable and put "Got Preference" if the values in first preference matches the values in the Offer column, otherwise put "Not g et preference". Now here nulls are the things who don't got admission. Then let's omit them, so that we can find the probability.

Conclusion:

From the results we can conclude that, those who got admission in their desired 1st preference does have a better chance to join university with probability 0.7779

Results:

Failed to get Choice	Got Choice
0.2220991	0.7779009

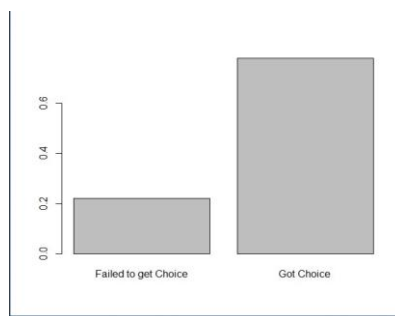


Figure 3- Question 3

Q4

Description:

In this question what I did is that, I grade the test marks of both NU and NTS test. I found the count of NU-Test with corresponding to the "Offer" as I have already discussed in Q1, and then found the probability.

Conclusion:

From the results I have concluded that those who performed well in NU Test have better chance of joining university with a probability 0.9366.

Results:

Probility_Nts	Probility_NU
0.06335267	0.93664733

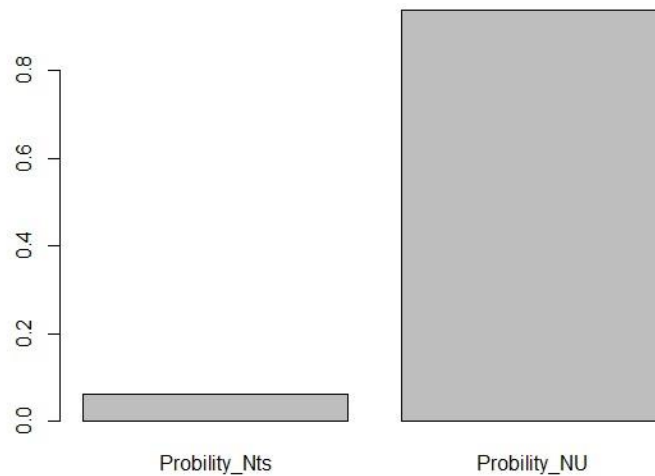


Figure 4- Question 4

Q5

Description:

What I did in question is that I found the count of each discipline with respect to campuses.

Conclusion:

From the results or table

Lahore campus's BS(CS) is much demanding than other discipline (count = 5731)

Islamabad campus's BS(CS) is much demanding than other discipline (count=4399)

Karachi campus's BS(CS) is much demanding than other discipline. (count=2686)

CFD campus's BS(CS) is much demanding than other discipline. (count=611)

PWR campus's BS(EE) is much demanding than other discipline. (count=335)

Results:

Camp	Deg	x
LHR	BS(CS)	5731
ISB	BS(CS)	4399
KHI	BS(CS)	2686
ISB	BS(EI)	1738
LHR	BS(EI)	1349
ISB	BBA	1176
LHR	BS(CV)	1144
LHR	BBA	819
KHI	BS(SE)	722
CFD	BS(CS)	611
KHI	BS(EI)	377
ISB	BS(AF)	352
PWR	BS(EI)	335
ISB	BS(CSDF)	305
ISB	BS(AI)	302
CFD	BS(SE)	293
PWR	BS(SE)	262
PWR	BS(CS)	260
CFD	BS(EI)	212
KHI	BBA	208
LHR	BS(AF)	205

Figure 5- Question 5

Q6

Description:

In this question, I calculated the percentages of both matriculation and Intermediate marks, After I have done with calculating the percentages, I found their correlation with NU test marks.

Conclusion:

The correlation between Matriculation Marks and NU Test in the BBA discipline is 0.36, which means there is no relation between Matriculation marks and NU Test.

The correlation between Matriculation Marks and NU Test in the BS discipline is 0.4857425 , means there is weak relation between Matriculation marks and NU Test.

The correlation between Intermediate Marks and NU Test in the BBA discipline is 0.1353845 , which means there is no relation between Matriculation marks and NU Test.

The correlation between Intermediate Marks and NU Test in the BS discipline is 0.2332956, which is no correlation between Matriculation marks and NU Test.

Results:

<p>Pearson's product-moment correlation</p> <p>data: Matriculation_Percentages and NUTest_BBA t = 21.315, df = 2877, p-value < 2.2e-16 alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.3373101 0.4004196 sample estimates: cor 0.3692905</p>	<p>Pearson's product-moment correlation</p> <p>data: Matriculation_Percentages and NUtest_BS t = 66.844, df = 14469, p-value < 2.2e-16 alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.4731943 0.4980936 sample estimates: cor 0.4857425</p>	<p>Pearson's product-moment correlation</p> <p>data: fsc_Percentages and NUTest_BBA t = 6.2124, df = 2067, p-value = 6.293e-10 alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.09283243 0.17744300 sample estimates: cor 0.1353845</p>	<p>Pearson's product-moment correlation</p> <p>data: fsc_Percentages and NUtest_BS t = 26.052, df = 11791, p-value < 2.2e-16 alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.2161571 0.2502903 sample estimates: cor 0.2332956</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 6 – Question 6

Q7

Description:

What I did in this question is that, I selected the boards from which the students are selected and aggregated the students according to their boards. And then the ascending order of the frame.

Conclusion:

From the conclusion,

The most successful board which were getting admission in NUCES was basically Lahore board followed by the federal board, then Karachi board and so on.

Results:

	D	x
15	Lahore	4443
10	Federal	3915
13	Karachi	1804
9	Faisalabad	1266
11	Gujranwala	996
27	Rawalpindi	940
22	Multan	663
24	Peshawar	592
6	Cambridge	524
28	Sahiwal	421
29	Sargodha	402
4	Bahawalpur	386
12	Hyderabad	367
1	Abbotabad	257
7	D.G.Khan	257
2	Agha Khan	234
21	Mirpur Khas	222
33	Sukkur	208

Figure 7 - Question 7

Q8

Description:

What I did in this question is that, I replace all the names in the intermediate board with province name, e.g. Lahore belongs to Punjab province so I replace Lahore with Punjab, and so on. Then I grade the NU's BS and BBA marks separately. After it I found the probability of province with these graded NU's BS and BBA marks.

Conclusion:

From the conclusion, there are way week/no correlation for the percentages greater then 20, but for less than 20 there is a strong correlation.

Results:

	A Grade	B Grade	C Grade	D Grade	F Grade
kpk	0.007649513	0.013908206	0.020862309	0.004867872	0.952712100
sindh	0.007007007	0.012679346	0.024691358	0.012012012	0.943610277
others	0.009345794	0.018691589	0.065420561	0.046728972	0.859813084
punjab	0.028815803	0.036147032	0.037979839	0.007636697	0.889420629
Federal	0.053840955	0.047533228	0.047082676	0.008785763	0.842757378
balochistan	0.000000000	0.000000000	0.027027027	0.000000000	0.972972973

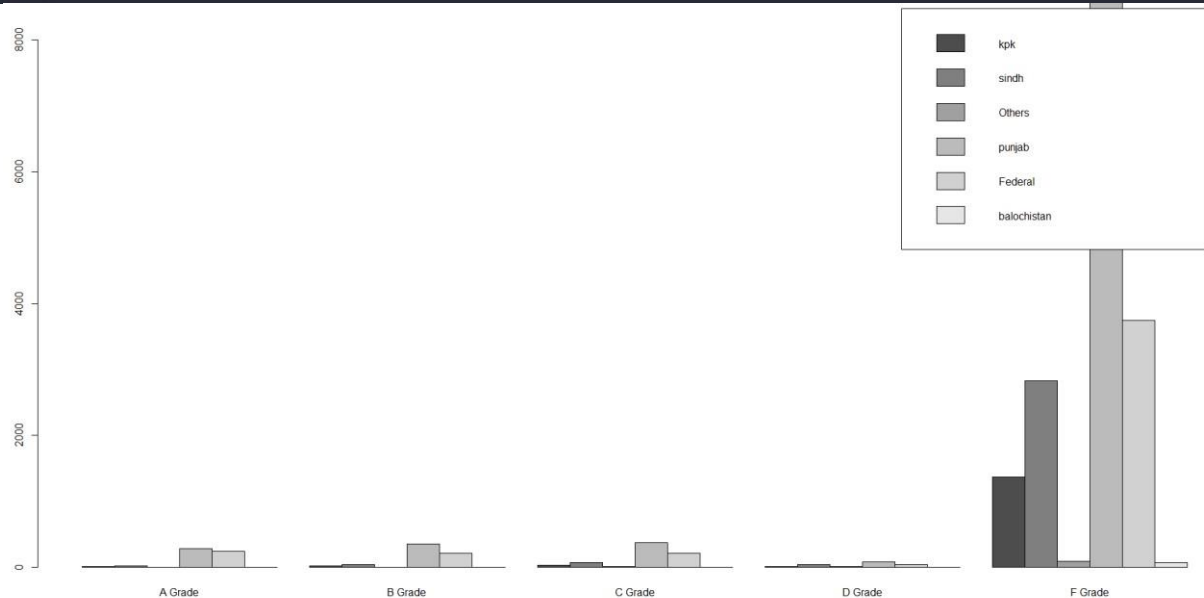


Figure 8 - Question 8

Q9

Description:

In this question I have selected the sample of 500 students from the population with respect to the intermediate population. But the data is not accurate because some of the total values are less than the obtained marks in intermediate column so I tested my algorithm on matriculation marks using t test.

Conclusion:

The results I got are:

Sample mean = 76.98779

Sample standard deviation = 17.36982

Calculated_Value = 8.995589

Critical_Value = -1.647913

Result:

As the calculated value is Greater than the critical value so We failed to reject Null Hypothesis.

```
one sample t-test

data: sample_Matrix
t = 8.0687, df = 499, p-value = 1
alternative hypothesis: true mean is less than 70
5 percent confidence interval:
 -Inf 75.71783
sample estimates:
mean of x
 77.18534

> # Calculating the t value
> t = (Sample_mean-70)/(Sample_standard_dev/sqrt(500))
> # Calculating the critical value for the test
> criticalvalue = qt(0.05, 499, lower.tail = TRUE)
> # Output matrix of the t test
> Output_test = c(Calculated_value=t, Critical_value=criticalvalue)
> print(Output_test)
Calculated_value Critical_value
      7.328779      -1.647913
> |
```

Figure 9 - Question 9

"As the calculated value is Greater than the critical value so We failed to reject Null Hypothesis"

Q10

Description:

In this question I have selected the sample of 500 students from the population with respect to the intermediate population. But the data is not accurate because some of the total values are less than the obtained marks in intermediate column so I tested my algorithm on matriculation marks using t test.

Conclusion:

The results I got are:

Sample mean = 47.86531

Sample standard deviation = 11.53497

Calculated_Value = -4.138119

Critical_Value = -1.647913

Result:

```

> # Calculating the critical value for the test for  $\mu$ 
> criticalvalue2 = qt(0.05, 499, lower.tail = TRUE)
> print(criticalvalue2)
[1] -1.647913
> output_test = c(Calculated_Value=t2,Critical_Value=criticalvalue2)
> print(output_test)
Calculated_Value    Critical_Value
          -4.232679           -1.647913
>

```

Figure 10 - Question 10

"As the calculated value is less than so We reject Null Hypothesis"