REPORT-:



OVERVIEW

* Data Source
* Motivation
* Conclusions
* Mathematical Arguments(T-test)
* Other resources

1. Data Source

The data of all the MPs are taken from the MP Track –PRS website.

2. Motivation

From the one of the course of 3rd semester CSP301 a course on Design Practices.

3. Test on The available data

Analysis of hypothesis using student’s T-test----:

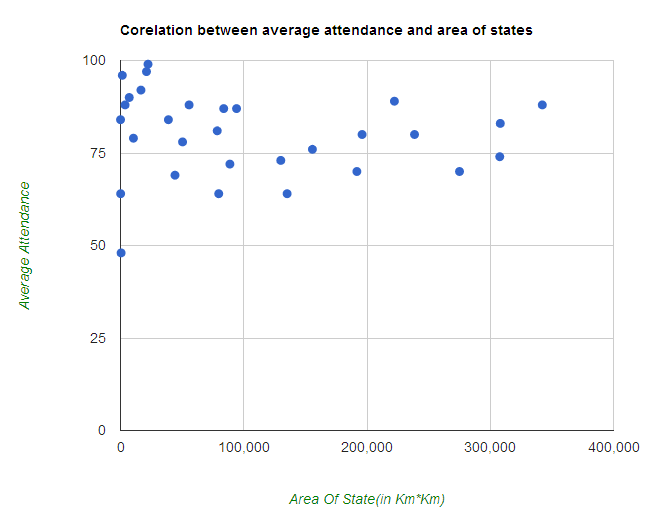
After analyzing the data taken from MP track site we found that in our parliament there are MPs whose age are in the range 30-91 and attendance between 1-100%. Average attendance and average age of all MPs in parliament is 77% and 56.43 years respectively.

1. Hypothesis-: MPs above certain age have low attendance

The plot of Age vs. Attendance for each MPs are as shown below where each bubble is corresponding for each MP and its position on graph is in accordance with his age and his attendance in parliament.

From the above plot we can draw following conclusions-:

* Large number of MPs of age above 62 years maintains attendance greater than 75%, which is nearly the average attendance of all MPs.
* Most of the MPs above 68 years maintain good attendance.
* Now for the Hypothesis let’s define our Threshold-age=68(as much as low as possible so that large number of people have low attendance) and Threshold-attendance=75%(as much as high possible, so that our hypothesis has more chance to be true).



Now let H0=`MPs of age above 68 has attendance lesser than 75%`. Sample size(Number of MPs of age above 68% )=N=67

Degree of freedom =f=N-1=66

So average attendance of MPs above 68 years=80%

Standard deviation of set of MPs above 68 years=s=0.169499545

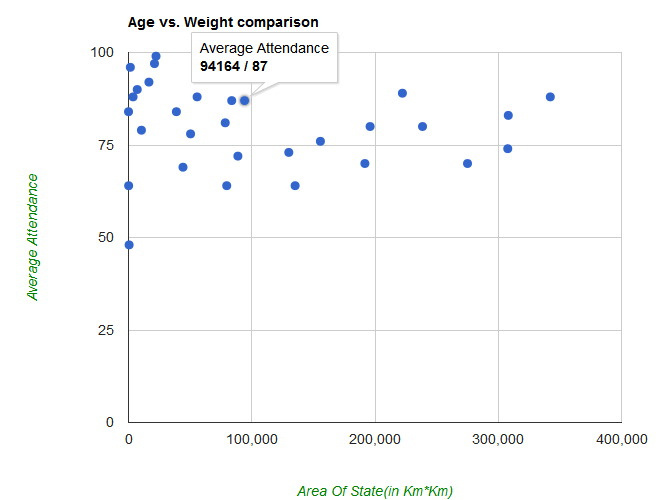
Difference in mean=80-75=5

95% confidence interval of this difference is From 4.95865582278 to 5.04134417722 and the corresponding t-value is 241.55 which is much large and not fall in the given range. Thus our hypothesis will rejected.

Also it can be seen from graph that the data four sample set {low attendance, low age}, {low attendance, high age}, {high attendance, low age}, {high attendance, high age} are disproportionate in the set {low attendance, low age} , and {high attendance, low age}.thus our hypothesis is not true that MPs above certain age has low attendance.

(2) Hypothesis: MPs from small states maintain better attendance in parliament-:

At 1st there is a problem how to decide which states are small as this data is not provided in excel sheet so I use data from Google to arrange the states in order of size according to the area of that particular state. Data is taken from the link given below, You can visit this link for further query <http://in.answers.yahoo.com/question/index?qid=20080218232455AA5Dgor> .

Now coming to the hypothesis part it can be analyzed from the graph that most of the states having area lesser then 100,000 KM2  maintain attendance greater than 75%, in fact some of them maintain even more than 90%. 

Thus visually we can conclude that small states have better attandence. To show it mathematically lets consider two sets-:

A=set of small states having area less than 100,000 km2

B=set of remaining states

And let our hypothesis is -:

H= MPs from set of small states that is from A maintain better attendance in parliament.

Mean(A)=81.42

d.o.f.(A)=19-1=18

Stdev(A)=13.48

Mean(B)=77

d.o.f.(B)=11-1=10

Stdev(B)=7.82

Now calculating T-value for two samples of diffent size and different mean.

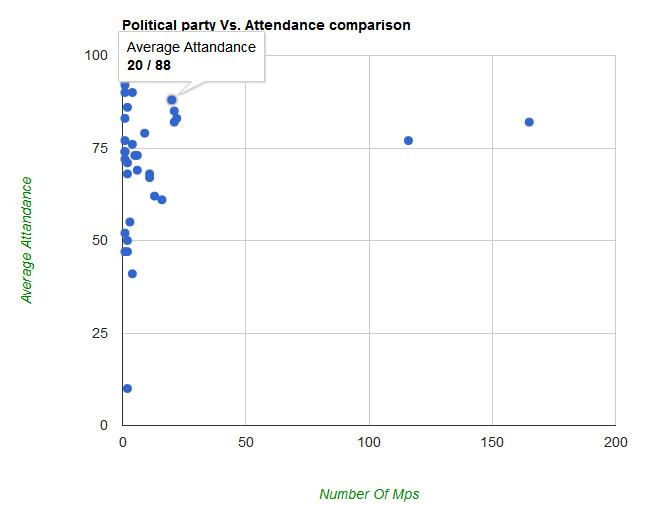
t-value= 0.9546

D.o.f.=28

95% confidence interval of this difference: From -4.7184 to 13.5584 and our t-value is in the range of this interval so our hypothesis is true.

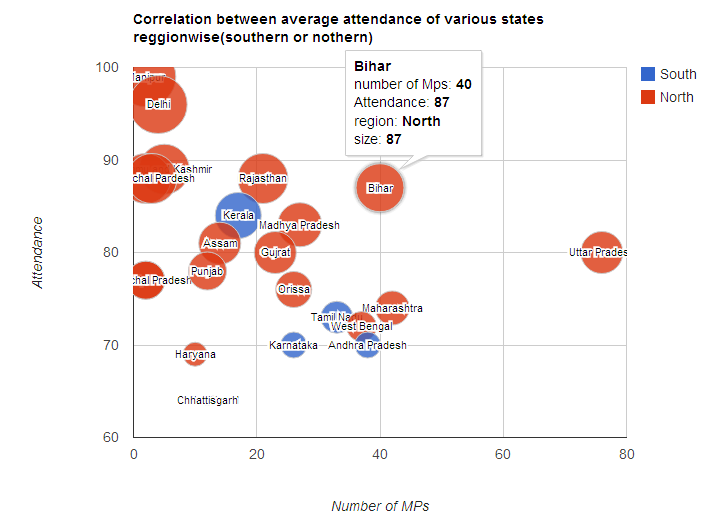
(3) Hypothesis: MPs from UPA (in power) or NDA (opposition) maintain better attendance.

Here we can see that Parties who are in power that is who have more number of MPs lie on the axis far from the other parties. We also see that two main parties who are in power and who are in opposition have attendance greater than 75% because they lie above the line Y=75 thus MPs of MPs of the two main party maintain better attendance.



4. MPs of North India maintain better attendance than MPs of south India-:

Among all the states of India few of them are southern and rest of them are northern. All southern states except Kerala(which has less number of MPs than other southern states) has attendance close to 70% but all northern states other than few has attendance greater than 75% so we can conclude from the graph that MPs of north India maintain better attendance than MPs of north India.



Mathematical test-:

Average attendance of southern MPs=72.96

Average attendance of southern MPs =78.81

Total number of southern MPs =114

Total number of southern MPs =371

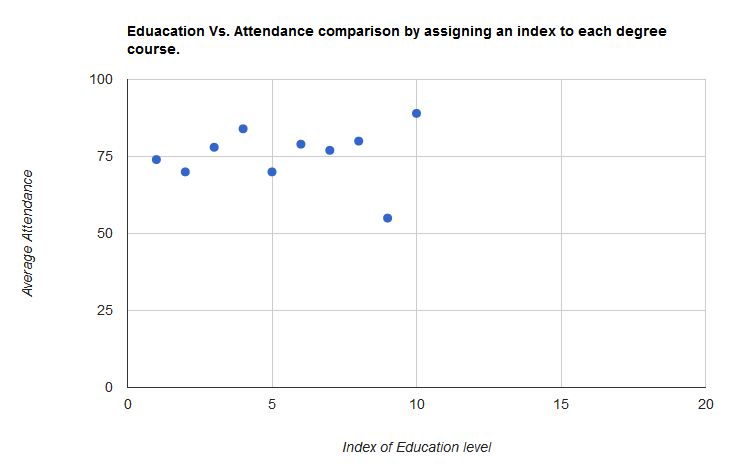
Standard deviation in attendance of southern MPs =23.09

Standard deviation in attendance of southern MPs =5.43

The difference in the mean of group1 and group 2(respectively) is equal to 5.8500 and 95% confidence corresponding to this interval lie in the range of 3.2916 to 8.4039.

T-value found to be equal to 4.5009 which lie in the range of critical range thus mathematically it also proved that MPs of Northern states maintain better attendance than Southern states.

5. More educated between MPs maintain better attendance-:



From here we can see that those who are highly educated(that whose index is highest) i.e. doctorates, undergraduates etc. maintain best attendance, but those who are post graduates does not attends much session held on the parliament so these MPs have lowest attendance

4. Difficulty faced-:

In the whole project I faced following difficulties which we efficiently reduce by taking few assumptions and analysis on the data.

1. To Read Data From Excel File-:For reading data form excel file I write a script in java-script and convert all the columns in an array and modify the array according to my need.
2. What to do in the case when data is not available-: In the case of N/A, at first I thought to assign the average of the data of particular property for which data is not available then suddenly someone raise the question of putting them zero which is awkward, as it will effect our data in terms of statistical analysis in average as well as standard deviation. So in that case what I finally decide to leave them and done the analysis for the rest of them because it will not effect our result much if we leave 1 MPs among large set or group of MPs and in case of small set or group case we are not drawing our conclusions form these sets.
3. Which method should I use for the test-: As I mentioned in discussion on piazza that which method should use among all t-Test, P-test, Z-test etc. sir mentioned us that use welch p-test for two different sets having different size. So I did this which draw the same conclusions as almost other test does.
4. Visualization is not clear in Scatter chart-: In the problem statement that I have to visualize all things using scatter chart but the Google API of scatter chart has only two field in the multidimensional array x-axis and y-axis. So I give user option to show the graphs on various option on bubble chart also which I think gives better visualization to the user due to the addition of color and size effect.
5. And others if any-:

5. Conclusions-:

Seeing all the data and after doing all the that was mentioned in the statement I start seeing interesting things in the data and find some interesting fact about this data which are the following-:

1. MPs of Kerala are more educated then MPs of Bihar-: whiling doing the visualization of MPs for different states it was found that MPs of Kerala are more educated then almost all other states or Bihar.

//here to show the png image of Kerala and Bihar state vs education of MPs.

1. High qualified MPs attends larger debates than those who are less qualified-: We can also show the visualization debates vs. education and can show that the average number of debates done by the people of different educational qualification.
2. MPs passes more bills attends larger number of debates.
3. There exist a correlation between the question and number of debates.
4. Others…………………

Other resources-:

* + Data sources-: PRS website Mp-track 15 loksabha.

<http://www.prsindia.org/index.php?name=mptracklok>

* + JavaScript Resources-:

<http://www.w3schools.com/js/default.asp>

* + T-test reference-: [http://en.wikipedia.org/wiki/Student%27s\_t-test](http://en.wikipedia.org/wiki/Student's_t-test)

<http://vassarstats.net/textbook/ch11pt1.html>