Module - 5

ALY 6010 Probability Theory and Statistics

RAHUL AVINASH JADHAV

Northeastern University

College of Professional Studies, Northeastern University, Boston, MA 02115

Contact: jadhav.ra@northeastern.edu

Submitted to Professor: Prof. Roy Wada

Date of submission: 12/12/2021

Introduction

In this assignment we are going to develop Correlation and regression table and plotting using the dataset “Guns_1999”. The dataset consists of 51 observations with 13 variables such as Year, Violent, Murder, Robbery, Prisoners etc.

We will also observe the difference we get after performing Correlation and regression and get understandings for the same.

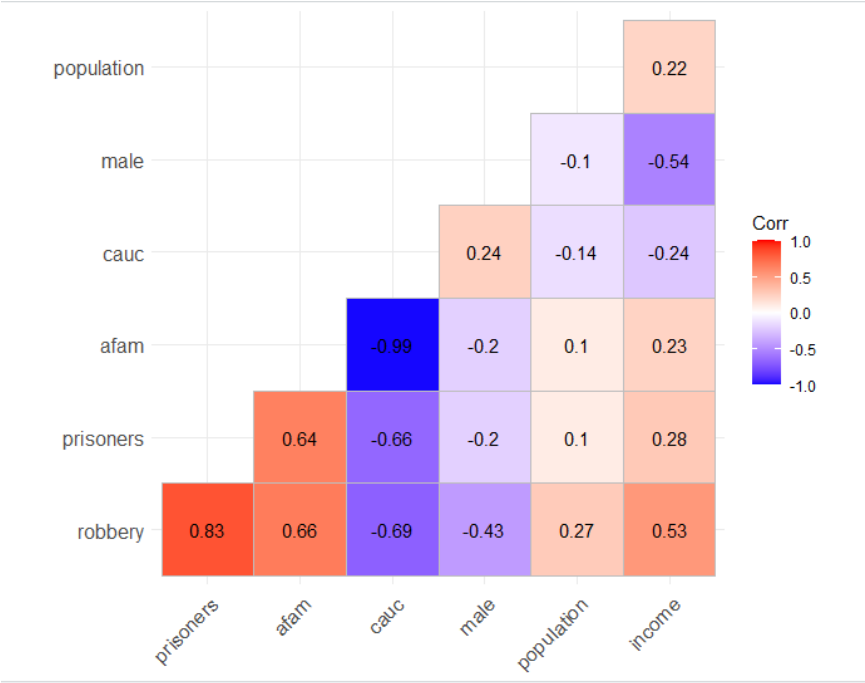
Analysis

Table 1: Correlation Metrics

	robbery	prisoners	afam	cauc	male	population	income
robbery	1	0.832647	0.655115	-0.68871	-0.43385	0.267119	0.531787
prisoners	0.832647	1	0.64236	-0.65595	-0.20358	0.101421	0.28009
afam	0.655115	0.64236	1	-0.98919	-0.20401	0.100594	0.234149
cauc	-0.68871	-0.65595	-0.98919	1	0.243404	-0.13898	-0.23757
male	-0.43385	-0.20358	-0.20401	0.243404	1	-0.10329	-0.53647
population	0.267119	0.101421	0.100594	-0.13898	-0.10329	1	0.221119
income	0.531787	0.28009	0.234149	-0.23757	-0.53647	0.221119	1

The above table provides correlation among various variables of the dataset. If properly observed the correlation of prisoners and robbery is highest.

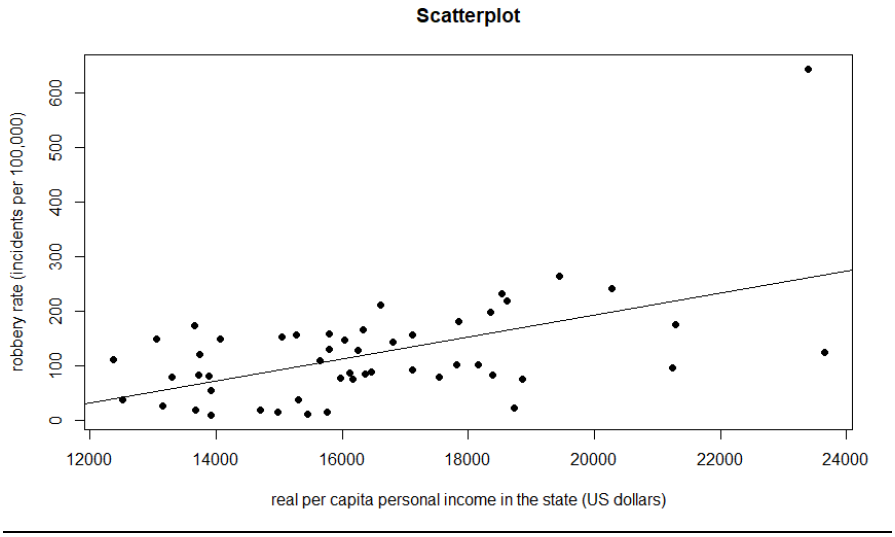
Plot 1: Correlation plot



From the above plot we can see the correlation between the robbery and prisoners is the highest. If we compare incarceration rate in the state in the previous year(prisoners) the population of African-American is higher as compared to Caucasian. Also, the population of African-American is more than Caucasian.

The real per capita income of African-American is more than Caucasian.

Plot 2: Scatter Plot based on per capita income and robbery rate



From the above scatter plot, we can see that, there is more robbery rate where the real per capita personal income in the state is more.

Plot 3: Scatter Plot based state wise pollution of male and robbery rate

Percentage of pollution of male of age between 10 to 29 is more, lesser the robbery rate can be concluded from the above plot. If we can see it from other side, the state where percentage of pollution of male of age group above 29 can have more rate of robbery.

Table 2: Regression table of Guns where dependent variable is robbery and independent variables are income and male.

Regression table of Guns

	Dependent variable:	
	robbery	
	(1)	(2)
income	0.020***	
	(0.005)	
male		-42.650***
		(12.653)
Constant	-210.719***	726.681***
	(76.439)	(180.074)
Observations	51	51
R^2	0.283	0.188
Adjusted R^2	0.268	0.172
Residual Std. Error (df = 49)	84.691	90.102
F Statistic (df = 1; 49)	19.321***	11.361***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Regression table of Guns

income	male

Summary

- Correlation is used for a quick and simple summary of direction and to find relationship between two or more numeric variables
- Regression is used when we are looking to predict between the variables (how x influences y)
- Interchanging x and y variables in correlation won't have any effect but if x and y variables are changed in regression there are difference in analysis

Bibliography

- Kabacoff, R. (2011). *R in action: Data analysis and graphics with R*. Manning.
- S, S. (2021, February 26). *Difference Between Correlation and Regression (with Comparison Chart)*. Key Differences. https://keydifferences.com/difference-between-correlation-and-regression.html