Assignment 11, MECO 6315

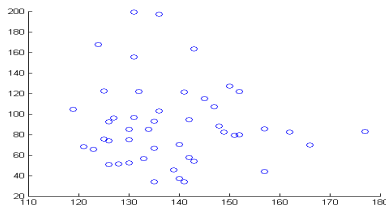Meisam Hejazinia

04/14/2013

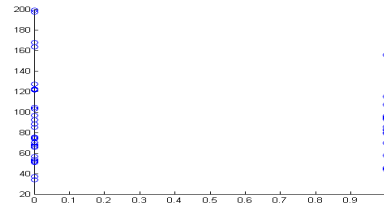Figure 1: Age preplot and outliers

Figure 2: Southern state preplot and outliers

1 Preplots

Preplots are shown in figure 1 to 12. If we want to be meticulous we can remove point with age 179, points with crime rate of 200, police expenditure greater than 140 in each of the years, unemployment greater than 55, population size greater than 140, transferable asset goods greater than 650, and non white proportion greater than 400 as outliers. However, these do not work as influence points due to their distance from the mean of population, so there would be no harm not removing them. Figure 13 shows the normal plot of crime rate, which confirms the assumption that data is normal. To be on the safe side we may need to remove outliers with crime rate greater than 140, but I did not remove it for my analysis.

2 Multiple Regression

Unfortunately MATLAB 2011 does not have forward and backward regression. I switched to R, and wrote the program, yet there also the criteria for selection is AIC and not p-value. Finally after checking different options, I had no choice other
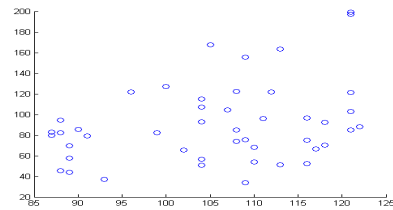

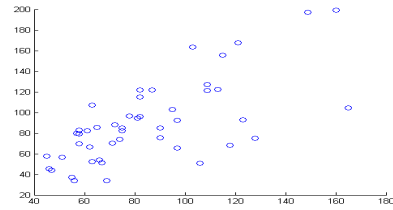Figure 3: Education preplot and outlier

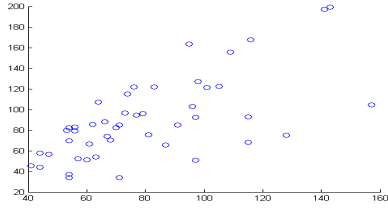Figure 4: Police expenditure 1960 preplot and outliers

1

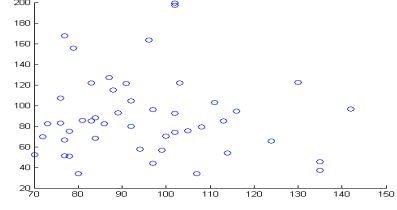Figure 5: Police expenditure 1959 preplot and outliers

Figure 6: Labor Force participation preplot and outliers

Figure 7: Population size preplot and outlier

Figure 8: Nonwhite portion of population preplot and outlier

Figure 9: Unemployment rate(14-24) preplot and outlier

Figure 10: Unemployment rate (25-35) preplot and outlier

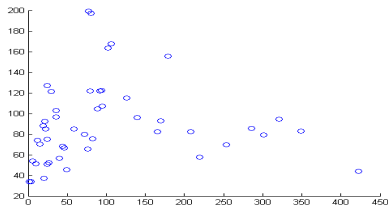Figure 11: Transferable goods and assets or family income preplot and outlier
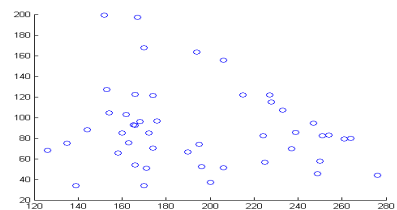

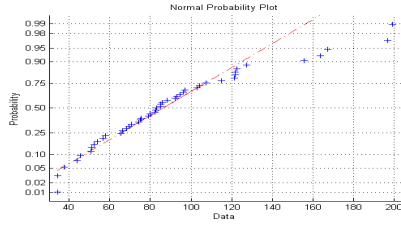Figure 12: Families earning below median income preplot and outlier

2

Figure 13: Crime rate normal probability plot

than implementing these two by myself from scratch, mean reinventing the wheel.

Result of backward regression is shown in table 1. The only remaining variables after step 10 would be age, education, police expenditure of 1960, and below median income families.

The result of forward multiple regression is shown in table 2. In the first step police expenditure of 1960 is added, on the second one below median income. On the Third step we have education, and on the fourth we have age.

Table 3 shows stepwise regression, built in in Matlab. Variables included are first police expenditure 1960, second number of median income families, third education, and fourth age. None of the variables are removed during steps. Table 4 shows which variables are included, and what are the coefficients and standard errors for them, resulting from stepwise regression.

Since three methods result in the same model, the coefficients of final model would be the same. The residuals of age, education, police expenditure 1960, and median income families is shown in figure 14-18. No fan like pattern could be seen in four plots, making us believe that the residuals are random, so our assumption was correct. Figure 19 shows normal probability plot of residuals confirming the normality assumption.

Next we will discuss about signs. Police expenditure in 1960 has positive effect on crime rate. This relation could be interpreted in reverse. In other word since the crime rate has increased, the police expenditure increased, yet the increase was not enough to dampen the crime rate. The number of families per
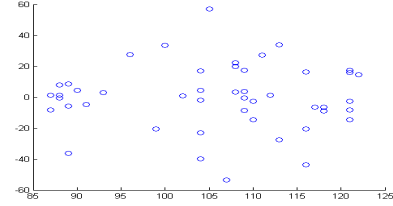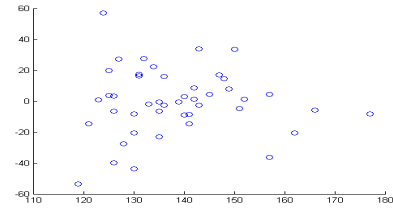

Figure 14: Education residual plot
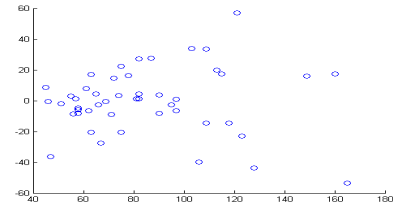

Figure 15: Age residual plot

Figure 16: Residual police expenditure 1960
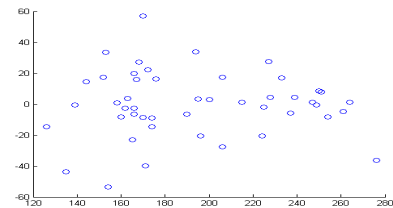
Figure 17: Residual lower than median income families plot

3

	Step1	Step 2	Step 3	Step 4	Step 6	Step 8	Step 10
Intercept	0.0001	0	0	0	0	0	0
Age	0.0251	0.0215	0.0135	0.0121	0.0111	0.0034	0.0383
South State	0.6675	0.6509	0.6563	0.6831			
Education	0.0152	0.0093	0.0084	0.0076	0.0059	0.0006	0.0008
Police Ex 60	0.1384	0.1283	0.1235	0.1245	0	0	0
Police EX 59	0.5809	0.5763	0.578	0.5833			
Labor force	0.9321						
Male female	0.414	0.358	0.3551	0.2259	0.1648		
Population	0.7911	0.7725	0.7745				
Nonwhite	0.8869	0.9012					
Unemp. 25	0.2083	0.1688	0.1637	0.1487	0.1504		
Unemp. 35	0.0586	0.055	0.0514	0.0487	0.0455	0.0814	
Trans. GDs	0.1717	0.1658	0.1592	0.1619	0.1495	0.073	
Bel. Med Inc.	0.0016	0.0011	0.0008	0.0005	0.0002	0	0.0001

Table 1: Backward regression steps. In each cell we have p-values. When there is no more element in the table it means that the variable is removed.

Var	Step1	Step 2	Step 3	Step 4	coeff
intercept	0.33	0.0045	0.0001	0	-430.1943
Police exp 60	0	0	0	0	1.3335
Bel. Med Inc.		0.0009	0	0.0001	0.669
Education			0.0018	0.0008	1.66
Age				0.0383	0.7344

Table 2: Forward Regression steps. At each cell of steps p-value could be seen.

Initial	columns	included:	none	
Step	1,	added column	4,	$p = 7.96673e - 008$
Step	2,	added column	13,	$p = 0.000881192$
Step	3,	added column	3,	$p = 0.00175702$
Step	4,	added column	1,	$p = 0.038288$
Final Col. Inc.	1	3	4	13

Table 3: Stepwise regression steps

Variables	Coeff	STD err	P-value	Status
Intercept	-430.1943			
Age	0.7344	0.3433	0.0383	In
South State	-2.4483	11.2563	0.8289	Out
Education	1.66	0.4568	7.55E-04	In
Police Ex 60	1.3335	0.1438	1.01E-11	In
Police EX 59	-1.099	1.0585	0.3052	Out
Labor force	0.0532	0.1029	0.6076	Out
Male female	0.1765	0.1294	0.1801	Out
Population	-0.0457	0.1124	0.6864	Out
Nonwhite	-0.0281	0.0535	0.6024	Out
Unemp. 25	0.1793	0.1896	0.3499	Out
Unemp. 35	0.8544	0.436	0.0568	Out
Trans. GDs	0.1923	0.0957	0.051	Out
Bel. Med Inc.	0.669	0.1523	7.46E-05	In

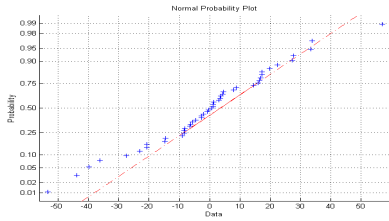Table 4: Stepwise regression coefficients and standard errors

Figure 18: Residuals normal probability plot

happened in the form that everybody find it fair. Therefore we see more crime. Finally age has positive effect on crime rate. This could be explained the same way we explained age in the class. As we have more seniors in the population, everybody is vigilant about who are wandering in the surrounding, yet as we have more young people, we do not have such phenomenon, and crime usually happens when people are not vigilant enough.

100 earning before half of the median income has positive effect on crime. This is intuitive since it is measure of difference of wealth between people. People who are under poverty may have motives of fairness perception make them to grab part of money of wealthy people, or really may have unsatisfied basic need making them engage in burglary. Education has also positive effect on crime. This again could be indogeneous and reverse effect. In other word, the relation could be interpreted in the form that usually educated people are more wealthy. Educated people move to places where there is more job, and more developed. In these places usually people do not know each other, but there are lots of resources. Crime usually happens in the places where there is lot of resource, yet resource allocation may have not