Huifen Zhou's Project

Multivariate Statistical Analysis of Crime Data in United States

(0) Definition of each crime

Murder is the unlawful killing of another human without justification or valid excuse, especially the unlawful killing of another human being with malice aforethought. This state of mind may, depending upon the jurisdiction, distinguish murder from other forms of unlawful homicide, such as manslaughter. Manslaughter is a killing committed in the absence of malice, brought about by reasonable provocation, or diminished capacity. Involuntary manslaughter, where it is recognized, is a killing that lacks all but the most attenuated guilty intent, recklessness.

Rape is a type of sexual assault usually involving sexual intercourse or other forms of sexual penetration carried out against a person without that person's consent. The act may be carried out by physical force, coercion, abuse of authority, or against a person who is incapable of giving valid consent, such as one who is unconscious, incapacitated, has an intellectual disability or is below the legal age of consent. The term rape is sometimes used interchangeably with the term sexual assault.

Robbery is the crime of taking or attempting to take anything of value by force, threat of force or by putting the victim in fear. According to common law, robbery is defined as taking the property of another, with the intent to permanently deprive the person of that property, by means of force or fear; that is to say, it is a larceny or theft accomplished by an assault. Precise definitions of the offence may vary between jurisdictions. Robbery is differentiated from other forms of theft (such as burglary, shoplifting or car theft) by its inherently violent nature (a violent crime); whereas many lesser forms of theft are punished as misdemeanors, robbery is always a felony in jurisdictions that distinguish between the two.

Assault is an attempt to initiate harmful or offensive contact with a person, or a threat to do so.[1]It is distinct from battery, which refers to the actual achievement of such contact. An assault is carried out by a threat of bodily harm coupled with an apparent, present ability to cause the harm. It is both a crime and a tort and, therefore, may result in either criminal and/or civil liability. Generally, the common law definition is the same in criminal and tort law. There is, however, an additional criminal law category of assault consisting of an attempted but unsuccessful battery. The term is often confused with battery, which involves physical contact. The specific meaning of assault varies between countries, but can refer to an act that causes another to apprehend immediate and personal violence, or in the more limited sense of a threat of violence caused by an immediate show of force. Assault in many US jurisdictions and Scotland is defined more broadly still as any intentional physical contact with another person without their consent; but in England and Wales and in most other common law jurisdictions in the world, this is defined instead as battery. Some jurisdictions have incorporated the definition of civil assault into the definition of the crime

making it a criminal assault intentionally to cause another person to apprehend a harmful or offensive contact.

Burglary (also called breaking and entering and sometimes housebreaking) is an unlawful entry into a building or other location for the purposes of committing an offence. Usually that offence is theft, but most jurisdictions include others within the ambit of burglary.

Larceny is a crime involving the unlawful taking of the personal property of another person or business. It was an offence under the common law of England and became an offence in jurisdictions which incorporated the common law of England into their own law.

Car theft or grand theft auto is the criminal act of stealing or attempting to steal any motor vehicle, usually an automobile.

Note: These definitions are quoted from Wikipedia.

(a) Principal analysis

Based on the summary of the data, this dataset contains 50 obs. and 7 variables. And the mean and standard deviation of this dataset are shown as below.

From the table Eigenvalues of the covariance Matrix and scree plot, it can be learned that the first two components can be used to explain the total variance, and they contribute more than 95 % of the total variance. Then the first two principal components can be used to explain the original data. Then no more than 4% of the total variance are contributed by the subsequent components.

The first two eigenvectors of the covariance matrix are

e_1^T =[0.000864, 0.008773, 0.056993, 0.059196, 0.465346, 0.872863, 0.121384]

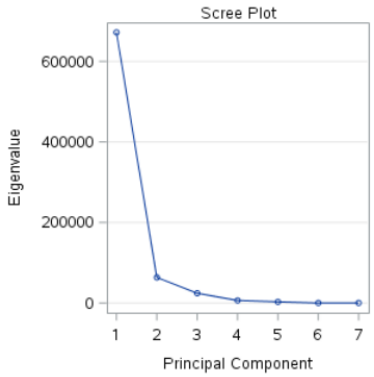
e_2^T =[[0.007077, 0.011477, 0.165921, 0.174243, 0.774439, -.481781, 0.331752]

The first component contains the variables of larceny (0.872863), burglary (0.465346) and maybe the auto (0.121384) because it shows very large eigen values for these three variables. Within the first component, all the other crimes also have small positive loading. The second component has a high positive loading on burglary (0.774439) and auto (0.331752), and high negative loading on larceny (-.481781). But the other variables' loadings are small and positive. The second component can be interpreted as "type of crimes" component.

Simple Statistics

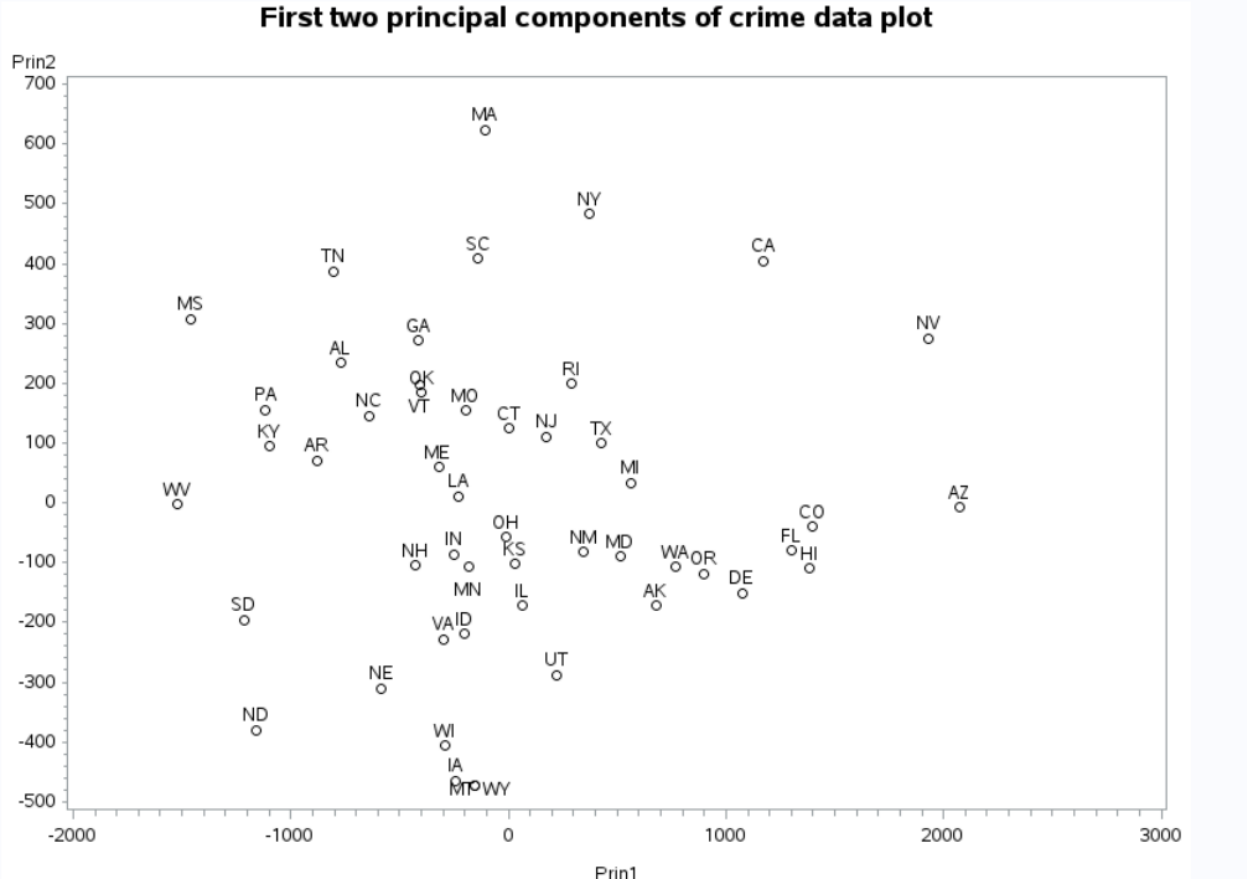
	murder	rape	robbery	assault	burglary	larceny	auto
Mean	7.444000000	25.73400000	124.0920000	211.3000000	1291.904000	2671.288000	377.5260000
StD	3.866768941	10.75962995	88.3485672	100.2530492	432.455711	725.908707	193.3944175

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	672099.938	608440.269	0.8736	0.8736
2	63659.669	39443.589	0.0827	0.9563
3	24216.080	17902.616	0.0315	0.9878
4	6313.464	3295.814	0.0082	0.9960
5	3017.650	2980.468	0.0039	0.9999
6	37.183	31.510	0.0000	1.0000
7	5.673		0.0000	1.0000



Eigenvectors							
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
murder	0.000864	0.007077	-.007375	0.022236	0.005032	0.184911	0.982437
rape	0.008773	0.011477	-.010400	0.051813	-.005986	0.981012	-.185953
robbery	0.056993	0.165921	0.110301	0.457211	0.864522	-.022240	-.011008
assault	0.059196	0.174243	-.150513	0.849046	-.468663	-.053603	-.009165
burglary	0.465346	0.774439	-.345505	-.253581	-.001246	-.003897	-.002102
larceny	0.872863	-.481781	0.059703	0.049105	0.001277	-.003608	0.002712
auto	0.121384	0.331752	0.917649	-.013601	-.181365	0.005253	0.004640

The plot of the first two principal components shows the first two PCs for the 50 states. It can be learned that the states with large number of crime due to burglary, larceny and maybe auto on the right hands side (NV, AZ) for the first PC. For the second PC, MA may have the highest chance of burglary and auto theft, and lowest chance of larceny.



(b) Factor Analysis

Based on the principal component analysis and the likelihood ratio test result, the number of factors should be m=2. From the table Factor Pattern, it can be learned that factor 1 shows the all crimes. That's because all kinds crimes load a large positive value. The factor 2shows that property crimes have the positive loading but violent crimes have the negative loading.

The communalities $h_j^2, j = 1 \ldots 7$, measure the part of variance of each variable that can be assigned to the common factors.

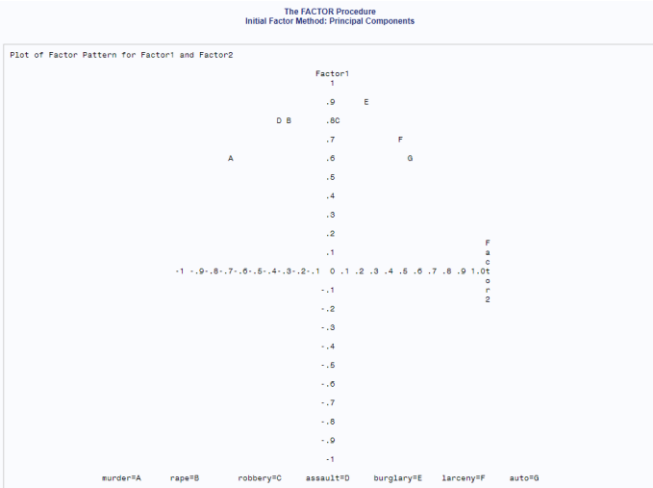
The estimated communalities are obtained

The $\widehat{h_j^2}^T = [0.86139693,0.80265644,0.65035880,0.79360069,0.84844347,0.72600469,0.67122032]$.

Factor Pattern		
	Factor1	Factor2
murder	0.60913	-0.70026
rape	0.87584	-0.18858
robbery	0.80508	0.04702
assault	0.80462	-0.38234
burglary	0.89287	0.22631
larceny	0.72492	0.44777
auto	0.59878	0.55918

Final Communality Estimates: Total = 5.353681						
murder	rape	robbery	assault	burglary	larceny	auto
0.86139693	0.80265644	0.65035880	0.79360069	0.84844347	0.72600469	0.67122032

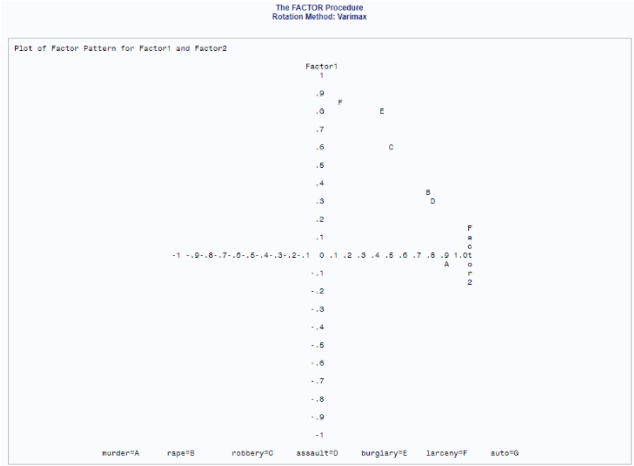
The Factor Procedure shows that some variables such as murder, auto are not clear in factor 1 or factor 2. Then the method of rotation is used.



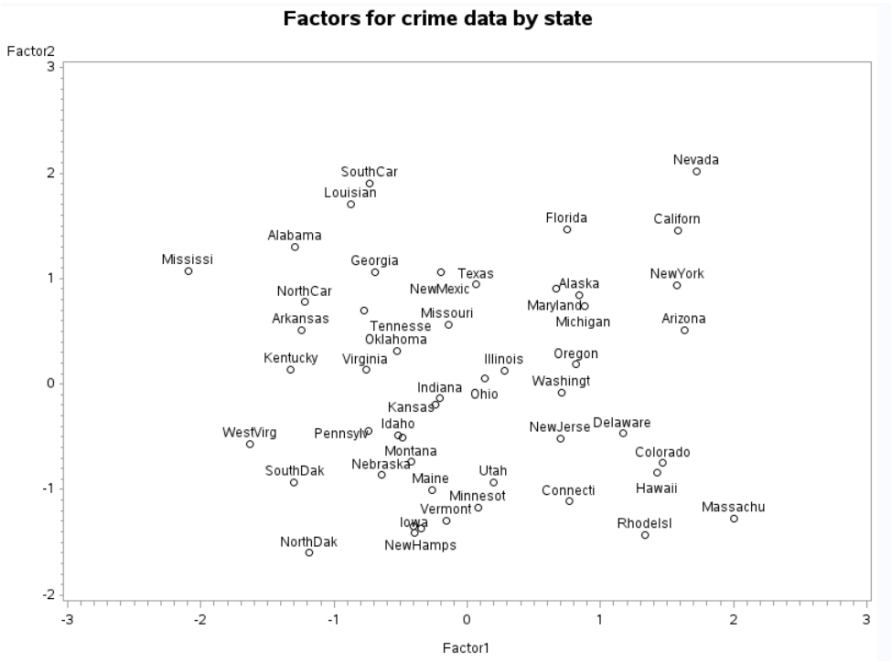
Rotation of factor loading

The Rotated Factor Pattern shows that larceny, auto theft and burglary have the large value of loading in factor 1 which means factor 1 represents property crimes. For factor 2, murder, assault and rape have the large value of loading. That means factor 2 represents violent crimes. The Factor Procedure plot also shows the same results to that. However, robbery isn't clear in both factor 1 and factor 2.

Rotated Factor Pattern		
	Factor1	Factor2
larceny	0.83310	0.17873
auto	0.81921	0.01100
burglary	0.80099	0.45481
robbery	0.61351	0.52341
murder	-0.04522	0.92701
assault	0.31595	0.83293
rape	0.50148	0.74241

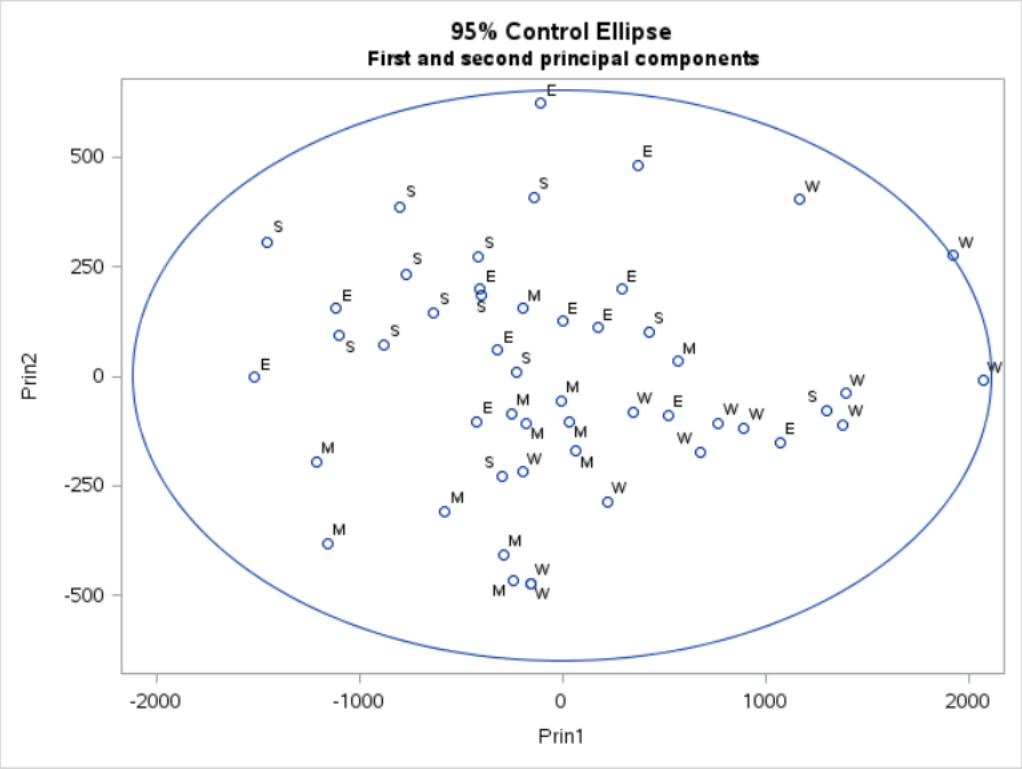


The plot of Factors for crime data shows that Nevada has high overall crimes, North Dakota has low crimes, Massachusetts has high property crimes but low violent crimes and Mississippi has high violent crimes but low property crimes.



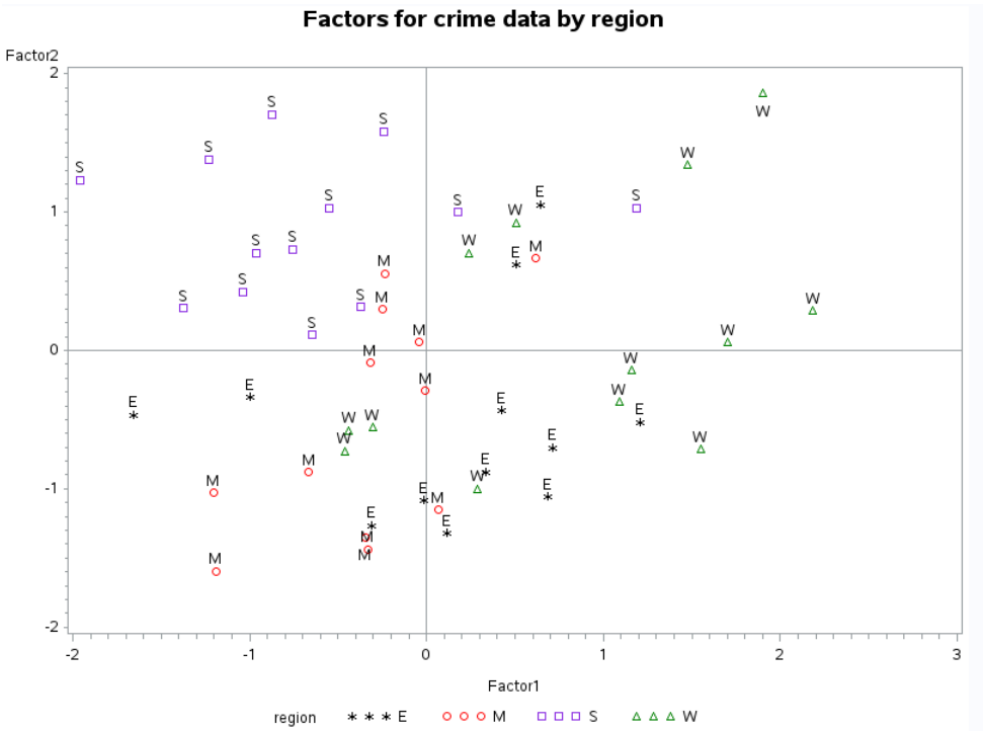
(c) Plot data by Region

1. PCA plot



The 95% control ellipse shows that all the data are controlled. Except 2 points of east region which have large value in PC2 and 2 points of west region which have large value in PC1. The region' plot of principal components shows that the states which locate in south region have positive value in principal 2 but negative value in principal 1. Only one state has negative value in PC2 and 2 states have positive value in PC1. And two points which belongs to west region show that they have larger chance in larceny, burglary and auto theft. West region's states have a positive and large value in PC 1 but some of them have positive value in PC2 and some of them have negative value in PC2.Most states of east region are around middle area. But there are 2 east states have extremely large value in PC2. Most states in middle region have negative value in PC1 and PC2. But they are more separate than south region.

2. Factors plot



Most of these points, stars, squares and triangles are in the middle of the plot and closes to its own region. The south region states locate the left upper area which means the positive loading of factor 2(violent crime) and negative loading of factor 1(property crime) and only one point locates in upper middle area. Most east region states' points in the bottom of the plot but there are 2 points locate in the upper side. Almost all west region's states have positive loading in factor 1 except 2 states have negative loading in factor 1. But some of them are have positive loading in factor2, some have negative loading in factor 2. That's means west region has a large chance in property crimes. For the middle region, most of the states' loading in the negative factor1 and facotr2, which means they have low chance of crimes.

(d) Characteristics of the regions with respect to the variables/ factors

Based on the explanation of(c), it indicates south region has more violent crimes than property crimes. Most east region states have a high loading in factor 1(property crime) but one or two points show that they have lower crimes both in property crimes and violent crimes. Most west region states are in the right side of vertical line which means west region has a large chance in property crimes. 2 points of west region has a large loading in both factor 1 and 2, which means they have all kinds of crimes. The middle region's states are distributed around 0 which means they have "average" chance of crimes. But most of these middle region's states' loading are negative in factor 1(property crimes) and factor 2 (violent crimes). That means middle region has low property crimes and violent crimes.

(e) Discriminant analysis

Based on the result of test of homogeneity, the P-value of Chi-square test is less than 0.001 which is smaller than significant level at 0.05. That result indicates that these variables have unequal variance.

From the table as below, the apparent error rate (APER) can be obtained : $APER = \frac{n_{1m}+n_{2m}+n_{3m}+n_{4m}}{n_1+n_2+n_3+n_4} = \frac{0}{50} = 0$, This result shows that there is no misclassification in this case.

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
165.655758	84	<.0001

Error Count Estimates for region					
	E	M	S	W	Total
Rate	0.0000	0.0000	0.0000	0.0000	0.0000
Priors	0.2500	0.2500	0.2500	0.2500	

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.CRIME
Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into region					
From region	E	M	S	W	Total
E	12 100.00	0 0.00	0 0.00	0 0.00	12 100.00
M	0 0.00	12 100.00	0 0.00	0 0.00	12 100.00
S	0 0.00	0 0.00	13 100.00	0 0.00	13 100.00
W	0 0.00	0 0.00	0 0.00	13 100.00	13 100.00
Total	12 24.00	12 24.00	13 26.00	13 26.00	50 100.00
Priors	0.25	0.25	0.25	0.25	

Then using two factors to do the discriminant analysis. The results are shown as below. The P-value of Chi-square is 0.2874 which is bigger than the significant level 0.05. Then the pooled covariance matrices can be used in discriminant analysis. The $APER = \frac{n_{1m}+n_{2m}+n_{3m}+n_{4m}}{n_1+n_2+n_3+n_4} = \frac{3+7+2+6}{50} = 0.36$.

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
10.831424	9	0.2874

Error Count Estimates for region					
	E	M	S	W	Total
Rate	0.5000	0.4167	0.0769	0.4615	0.3638
Priors	0.2500	0.2500	0.2500	0.2500	

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.FF
Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into region					
From region	E	M	S	W	Total
E	6 50.00	4 33.33	0 0.00	2 16.67	12 100.00
M	1 8.33	7 58.33	1 8.33	3 25.00	12 100.00
S	0 0.00	0 0.00	12 92.31	1 7.69	13 100.00
W	2 15.38	3 23.08	1 7.69	7 53.85	13 100.00
Total	9 18.00	14 28.00	14 28.00	13 26.00	50 100.00
Priors	0.25	0.25	0.25	0.25	

Then using two principal components to do the discriminant analysis. The results are shown as below. The P-value of Chi-square is 0.0922which is bigger than the significant level 0.05. Then the pooled covariance matrices can be used in discriminant analysis. The $APER = \frac{n_{1m}+n_{2m}+n_{3m}+n_{4m}}{n_1+n_2+n_3+n_4} = \frac{3+5+6+5}{50} = 0.38$.

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
14.954463	9	0.0922

Error Count Estimates for region					
	E	M	S	W	Total
Rate	0.6667	0.2500	0.3846	0.2308	0.3830
Priors	0.2500	0.2500	0.2500	0.2500	

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.AA
Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into region					
From region	E	M	S	W	Total
E	4 33.33	1 8.33	5 41.67	2 16.67	12 100.00
M	0 0.00	9 75.00	1 8.33	2 16.67	12 100.00
S	2 15.38	2 15.38	8 61.54	1 7.69	13 100.00
W	1 7.69	2 15.38	0 0.00	10 76.92	13 100.00
Total	7 14.00	14 28.00	14 28.00	15 30.00	50 100.00
Priors	0.25	0.25	0.25	0.25	

(f) Multivariate analysis of variance on region to variables and factors

Based on the results of variables which are shown as below, all test's P-value are less than 0.001 which are less than significant level at 0.05. The null hypothesis can be rejected. Then it can be concluded that these variables are not same in two different regions.

The GLM Procedure
Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for region
E = Error SSCP Matrix

		Characteristic Vector V'EV=1						
Characteristic Root	Percent	murder	rape	robbery	assault	burglary	larceny	auto
2.65110276	58.77	0.05871936	0.00794771	-0.00220498	0.00025858	0.00007934	-0.00002066	-0.00025870
1.49262362	33.09	-0.01111281	0.01696452	-0.00073937	-0.00070447	-0.00029115	0.00031129	-0.00001050
0.36736422	8.14	-0.00836249	-0.00321187	-0.00115988	0.00051791	0.00053353	-0.00014669	0.00043647
0.00000000	0.00	0.02493941	0.00145434	-0.00106718	-0.00001854	-0.00040262	0.00007993	0.00095045
0.00000000	0.00	-0.01748393	-0.00803047	-0.00016067	0.00229827	-0.00030042	0.00011700	0.00000000
0.00000000	0.00	0.01467732	-0.00389938	0.00160226	0.00013934	-0.00018752	0.00010093	0.00000000
0.00000000	0.00	0.03609437	-0.02074866	-0.00043601	-0.00009189	0.00001085	0.00023868	0.00000000

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall region Effect
H = Type III SSCP Matrix for region
E = Error SSCP Matrix

S=3 M=1.5 N=19

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.08035909	7.73	21	115.41	<.0001
Pillai's Trace	1.59359239	6.80	21	126	<.0001
Hotelling-Lawley Trace	4.51109061	8.38	21	77.224	<.0001
Roy's Greatest Root	2.65110276	15.91	7	42	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Based on the results of factors which are shown as below, all test's P-value are less than 0.001 which are less than significant level at 0.05. The null hypothesis can be rejected. Then it can be concluded that the 2 factors (property crimes and violent crimes) are not same in two different regions.

Appendix:

Code:

```
data crime;
 input state $1-15 murder rape robbery assault burglary larceny auto state $ region $;
 cards;
 Alabama         14.2 25.2 96.8 278.3 1135.5 1881.9 280.7 AL S
 Alaska          10.8 51.6 96.8 284.0 1331.7 3369.8 753.3 AK W
 Arizona          9.5 34.2 138.2 312.3 2346.1 4467.4 439.5 AZ W
 Arkansas         8.8 27.6 83.2 203.4 972.6 1862.1 183.4 AR S
 California      11.5 49.4 287.0 358.0 2139.4 3499.8 663.5 CA W
 Colorado         6.3 42.0 170.7 292.9 1935.2 3903.2 477.1 CO W
 Connecticut      4.2 16.8 129.5 131.8 1346.0 2620.7 593.2 CT E
 Delaware         6.0 24.9 157.0 194.2 1682.6 3678.4 467.0 DE E
 Florida         10.2 39.6 187.9 449.1 1859.9 3840.5 351.4 FL S
 Georgia          11.7 31.1 140.5 256.5 1351.1 2170.2 297.9 GA S
 Hawaii           7.2 25.5 128.0 64.1 1911.5 3920.4 489.4 HI W
 Idaho           5.5 19.4 39.6 172.5 1050.8 2599.6 237.6 ID W
 Illinois        9.9 21.8 211.3 209.0 1085.0 2828.5 528.6 IL M
 Indiana          7.4 26.5 123.2 153.5 1086.2 2498.7 377.4 IN M
 Iowa            2.3 10.6 41.2 89.8 812.5 2685.1 219.9 IA M
 Kansas           6.6 22.0 100.7 180.5 1270.4 2739.3 244.3 KS M
 Kentucky         10.1 19.1 81.1 123.3 872.2 1662.1 245.4 KY S
 Louisiana        15.5 30.9 142.9 335.5 1165.5 2469.9 337.7 LA S
 Maine           2.4 13.5 38.7 170.0 1253.1 2350.7 246.9 ME E
 Maryland         8.0 34.8 292.1 358.9 1400.0 3177.7 428.5 MD E
 Massachusetts    3.1 20.8 169.1 231.6 1532.2 2311.3 1140.1 MA E
 Michigan         9.3 38.9 261.9 274.6 1522.7 3159.0 545.5 MI M
 Minnesota        2.7 19.5 85.9 85.8 1134.7 2559.3 343.1 MN M
 Mississippi     14.3 19.6 65.7 189.1 915.6 1239.9 144.4 MS S
 Missouri         9.6 28.3 189.0 233.5 1318.3 2424.2 378.4 MO M
 Montana          5.4 16.7 39.2 156.8 804.9 2773.2 309.2 MT W
 Nebraska         3.9 18.1 64.7 112.7 760.0 2316.1 249.1 NE M
 Nevada          15.8 49.1 323.1 355.0 2453.1 4212.6 559.2 NV W
 New Hampshire    3.2 10.7 23.2 76.0 1041.7 2343.9 293.4 NH E
 New Jersey       5.6 21.0 180.4 185.1 1435.8 2774.5 511.5 NJ E
 New Mexico       8.8 39.1 109.6 343.4 1418.7 3008.6 259.5 NM W
 New York         10.7 29.4 472.6 319.1 1728.0 2782.0 745.8 NY E
 North Carolina  10.6 17.0 61.3 318.3 1154.1 2037.8 192.1 NC S
 North Dakota     0.9 9.0 13.3 43.8 446.1 1843.0 144.7 ND M
 Ohio            7.8 27.3 190.5 181.1 1216.0 2696.8 400.4 OH M
 Oklahoma         8.6 29.2 73.8 205.0 1288.2 2228.1 326.8 OK S
 Oregon          4.9 39.9 124.1 286.9 1636.4 3506.1 388.9 OR W
 Pennsylvania     5.6 19.0 130.3 128.0 877.5 1624.1 333.2 PA E
 Rhode Island     3.6 10.5 86.5 201.0 1489.5 2844.1 791.4 RI E
 South Carolina  11.9 33.0 105.9 485.3 1613.6 2342.4 245.1 SC S
 South Dakota     2.0 13.5 17.9 155.7 570.5 1704.4 147.5 SD M
 Tennessee       10.1 29.7 145.8 203.9 1259.7 1776.5 314.0 TN S
```

```
Texas          13.3 33.8 152.4 208.2 1603.1 2988.7 397.6 TX S
Utah           3.5 20.3 68.8 147.3 1171.6 3004.6 334.5 UT W
Vermont        1.4 15.9 30.8 101.2 1348.2 2201.0 265.2 VT E
Virginia       9.0 23.3 92.1 165.7 986.2 2521.2 226.7 VA S
Washington     4.3 39.6 106.2 224.8 1605.6 3386.9 360.3 WA W
West Virginia  6.0 13.2 42.2 90.9 597.4 1341.7 163.3 WV E
Wisconsin      2.8 12.9 52.2 63.7 846.9 2614.2 220.7 WI M
Wyoming        5.4 21.9 39.7 173.9 811.6 2772.2 282.0 WY W
;



*principle analysis;
proc princomp data=crime cov out=aa outstat=aa_stat;
var  murder rape robbery assault burglary larceny auto;
run;
proc score data=crime score=aa_stat out=FScore;
var   murder rape robbery assault burglary larceny auto;
run;


GOPTIONS RESET=ALL;
proc corr data=aa;        *the principal components are uncorrelated/independent;
 var prin1-prin7;
run;
SYMBOL1 pointlabel=('#state')V=circle C=black I=none;
TITLE1 "Crime Rates per 100,000 Population by State";
TITLE2 "Plot of the First Two Principal Components";;
proc gplot;
 title 'PLOT OF the First Two PRINCIPAL COMPONENTS';
 plot prin2*prin1;
* symbol1 v=1 c=red;
run;

**pcs for crime data by region;
SYMBOL1 pointlabel=('#region')V=star C=black I=none;
SYMBOL2 pointlabel=('#region')V=circle C=red I=none;
SYMBOL3 pointlabel=('#region')V=square C=blueviolet I=none;
SYMBOL4 pointlabel=('#region')V=triangle C=green I=none;
title "95% Control Ellipse";
title2 "First and second principal components";
proc sgplot data=aa  noautolegend;
   scater x=prin1 y=prin2/datalabel=region;
   ellipse x=prin1 y=prin2/alpha=0.05;          /* default is ALPHA=0.05 */
run;
```

```
proc gplot;
title 'Principals for crime data by region';
 plot prin2*prin1=REGION/href=0 vref=0;
run;
quit;

title "95% Prediction Ellipse";
title2 "First and second principal components";
proc sgplot data=aa  noautolegend;
   scater x=prin1 y=prin2/datalabel=state;
   ellipse x=prin1 y=prin2/alpha=0.05;          /* default is ALPHA=0.05 */
run;
*factor analysis;
proc factor data=crime  ;
var murder rape robbery assault burglary larceny auto ;
run;

proc factor data=crime rotate=v reorder n=2 out=ff outstat=ff_1 plot;
 var murder rape robbery assault burglary larceny auto  ;
run;

proc factor data=crime rotate=v reorder method=ml heywood n=2  ;
var murder rape robbery assault burglary larceny auto ;
run;

proc score data=crime score=ff_1 out=FScore;
var  murder rape robbery assault burglary larceny auto;
run;
proc print data=FScore;
run;
proc corr data=ff;
run;


proc gplot;
title 'PLOT OF FACTORS FOR CRIME DATA';
 plot factor2*factor1=state;
run;
 **********LANBLE REGION******;
SYMBOL1 pointlabel=('#region')V=star C=black I=none;
SYMBOL2 pointlabel=('#region')V=circle C=red I=none;
SYMBOL3 pointlabel=('#region')V=square C=blueviolet I=none;
SYMBOL4 pointlabel=('#region')V=triangle C=green I=none;
proc gplot;
title 'Factors for crime data by region';
 plot factor2*factor1=REGION/href=0 vref=0;
```

```
run;

******discriminant analysis*****;
proc stepdisc data=crime ;
 *this will test the equality of the variance covariance matrices;
 class region;
 var  murder rape robbery assault burglary larceny auto ;
run;


proc discrim data=crime method=normal  pool=test slpool=0.05;*wcov short;
 *this will test the equality of the variance covariance matrices;
 class region;
 var  murder rape robbery assault burglary larceny auto ;
run;

proc discrim data=crime pool=yes out=out outd=outd;
 *we will pool in this example if though it is INappropriate;
 class region;
 var  murder rape robbery assault burglary larceny auto ;
* priors  'E'=0.25 'M'=0.25  'S'=0.25 'W'=0.25;
run;

proc discrim data=ff pool=test wcov short ;
 class region;
 var factor2 factor1;
run;


proc discrim data=aa pool=test wcov short ;
 class region;
 var prin1 prin2;
run;

proc discrim data=aa pool=test wcov short ;
 class region;
 var prin1 prin2;
run;
***MANOVA**;
proc glm data=crime;
class region;
model murder rape robbery assault burglary larceny auto= region/ss3;
manova h=_all_/printe;
means region;
run;
```