

MicroCase exercise #2:  
An exploration of self-reported delinquency

INTRODUCTION

In these data analysis exercises, we will use Wave VII data from the National Youth Survey (Elliott, 1996) to examine the distribution and correlates of crime that are discussed in Part IV.

DESCRIPTION OF THE DATASET

In 1976, a team of researchers administered the first wave of the National Youth Survey. This comprehensive survey asked juveniles about their attitudes toward and participation in deviant acts. The survey also queried the youths about their home lives, friends, and work/school activities. In 1987, the seventh wave of the survey was administered and focused on events and behaviors that occurred during 1986, when the participants were in their twenties.

The original National Youth Survey Wave VII data included 1,730 variables for 1,725 individuals. The data for this exercise in this book do not include the 343 cases for which many of the self-reported delinquency data are missing. This leaves a total of 1,382 cases. If you are interested in obtaining the entire dataset, you may do so by contacting ICPSR, the Inter-University Consortium for Political and Social Research. Enough about the dataset, let's do some analyses.

Before we start, we'll need to load up our dataset. We'll do this the same way we did for the first MicroCase exercise (#1), except we need the dataset called "NYSCRIME."

EXPLORING THE CORRELATES OF CRIME: GETTING A HANDLE ON THE DATA

Now, let's take a quick look at some demographic factors to get a handle on the data. To do this, run frequencies on MALE, ETHNIC and AGE. If you need to refresh your memory regarding how to run frequencies, see Blueprint 1 in the first MicroCase exercise. The frequency tables appear below; they look a bit different from the ones generated by MicroCase, but have all the details you need for the exercises.

```
-----  
MALE          y7_2: SEX  
  
                Frequency  Percent  Cum  
                682      49.3     Percent  
Female  
Male           700      50.7     100.0  
                -----  
Total         1382     100.0     100.0  
  
Valid cases    1382      Missing cases    0  
-----
```

ETHNIC		y7_3: ETHNICITY		
		Frequency	Percent	Cum Percent
Anglo		1123	81.3	81.3
Black		190	13.7	95.0
Hispanic		48	3.5	98.5
American Indian		6	.4	98.9
Asian		13	.9	99.9
Other		2	.1	100.0
Total		1382	100.0	100.0
Valid cases	1382	Missing cases		0

AGE		y7_7: AGE (in years)		
		Frequency	Percent	Cum Percent
21		210	15.2	15.2
22		218	15.8	31.0
23		214	15.5	46.5
24		195	14.1	60.6
25		195	14.1	74.7
26		190	13.7	88.4
27		160	11.6	100.0
Total	1382	1382	100.0	100.0
Valid cases	1382	Missing cases		0

From the frequencies, we can see that our sample is nearly equally divided between males and females. We can also see that our data contain responses from a large percentage (81%) of Anglos and smaller numbers of Blacks, Hispanics, American Indians, Asians and Others. In fact, there are so few American Indians, Asians and Others that we will exclude them from any analyses on ethnicity because there are not enough of them to allow for meaningful analyses. Along these lines, I created a new variable WHITE that includes Anglos (coded as '1') and Blacks (coded as '0'), but codes the other ethnicities as missing so they will no longer appear in our analyses.

We can also see that our sample contains responses from individuals aged 21 to 27. This is important information; we now know that we cannot meaningfully test for age effects!

#### EXPLORING THE CORRELATES OF CRIME WITH CROSSTABULATION TABLES

Now, let's get into some replication and exploration. Replication means that we will reproduce the research of other scholars; in our case, we will run some of the analyses performed by the authors whose writings are in Part IV. We will also run some exploratory analyses of our own.

For our first exploratory activity, let's see whether men are more likely than women to commit crimes. Many researchers have found this to be true. We will look at a total of three crimes: (1) whether the respondent has ever stolen something worth more than \$50.00 (THEFT50); (2) whether the respondent has ever committed a violent offense (VIOLENCE); and (3) whether the

respondent used marijuana during the past year (USEDMJ). In this section we will only look at marijuana use, leaving violent offenses and thefts for you to look at for the "further exploration" questions. USED MJ required modification to make it easier to use in our crosstabulations. The raw number of times that the respondent used marijuana during the previous year was recoded into two categories: (0) those who said they did not use marijuana during the previous year, and (1) those who said they used it one or more times.

Before we see whether men are more likely than women to report having committed a crime, run a frequency table for USED MJ to see how it is distributed because we need to make sure there are enough cases in each category to make our analyses worthwhile. If there are very few cases, statistical analyses are not valid. For example, only 7 respondents said they had ever physically hurt or threatened to hurt someone to get them to have sex. Any tables run on such a variable would be invalid due to small cell sizes. The truncated frequency table for USED MJ appears below.

```

-----
USED MJ      used marijuana during in past year

              Frequency  Percent  Cum
              Frequency  Percent  Percent
no marijuana use      873      63.2      63.2
1 + times              509      36.8      100.0
-----
              Total      1382      100.0      100.0

Valid cases      1382      Missing cases      0
-----

```

As we can see from our table, more than one-third said they had used marijuana at least once during the previous year, so both of our categories are large enough for meaningful analyses. Now, to see if any of the correlates of crime discussed in Part IV play a role in marijuana use. To explore this possibility, we'll use our old friend, the crosstabulation table.

Race and sex are relatively straightforward variables in the NYS dataset, but I had to calculate income levels through respondents' hourly wages (and hours worked weekly) for all jobs held during the previous year, and adding income from other sources to that amount. The resulting numbers were recoded into three categories using census bureau data: (1) below the 1986 poverty threshold of \$5,701; (2) above the poverty threshold but below the 1986 mean income of \$16,908 and (3) more than the 1986 mean income. I used census data from 1986 because that's the year for which the job data was collected. Note that we cannot meaningfully test the effects of age because the respondents are too similar in age, as revealed through our frequency tables above.

Using the directions in Blueprint 2 (in the first MicroCase exercise) if necessary, run crosstabulation tables with USED MJ as the dependent (row) variable and MALE, WHITE and INCOME as independent (column) variables. Make sure to click on column percents, then summary statistics to get the Cramer's V and probability. The tables appear below (again, they look different from the MicroCase tables but have all the important information), followed by descriptions for each.

-----  
 USED MJ used marijuana during in past year  
 by MALE y7\_2: SEX

		MALE			
		Female	Male	Row	
		0	1	Total	
USED MJ					
	0	469	404	873	
no marijuana use		68.8	57.7	63.2	
	1	213	296	509	
1 + times		31.2	42.3	36.8	
	Column Total	682	700	1382	
		49.3	50.7	100.0	

Statistic	Value	Approximate Significance
Cramer's V	.11458	.00002

Number of Missing Observations: 0

-----  
 USED MJ used marijuana during in past year  
 by AGE y7\_7: AGE (in years)

		AGE							
		21	22	23	24	25	26	27	Row Total
USED MJ									
no MJ use	0	132	128	128	114	132	136	103	873
		62.9	58.7	59.8	58.5	67.7	71.6	64.4	63.2
1 + times	1	78	90	86	81	63	54	57	509
		37.1	41.3	40.2	41.5	32.3	28.4	35.6	36.8
	Column Total	210	218	214	195	195	190	160	1382
		15.2	15.8	15.5	14.1	14.1	13.7	11.6	100.0

Statistic	Value	Approximate Significance
Cramer's V	.09454	.05458

Number of Missing Observations: 0

-----

USEDMJ used marijuana during in past year  
 by WHITE Respondent race: Black or Anglo

	Count Col Pct	WHITE		Row Total
		Black	Anglo	
		0	1	
USEDMJ				
no marijuana use	0	126 66.3	698 62.2	824 62.8
1 + times	1	64 33.7	425 37.8	489 37.2
	Column Total	190 14.5	1123 85.5	1313 100.0

Statistic	Value	Approximate Significance
Cramer's V	.03028	.27258

Number of Missing Observations: 69

First, let's consider the effects of gender. In 1998, Daniel Mears, Matthew Ploeger, and Mark Warr (one of many research teams who have looked at this topic) report that women are less likely than men to commit crimes. Is this finding supported by the NYS data? To find out, look at the table for MALE. Were men more likely than women to have smoked marijuana during the past year? It seems pretty safe to conclude that women are less likely than men to commit at least this offense. The process we just completed is one part of replication- - when researchers see if the findings reported by other researchers are true for their own data. Replication is a very important part of research because it shows the consistency of findings and allows for sound theories to be developed.

Now, let's look at the effects of WHITE. Robert Sampson and William Julius Wilson noted that blacks are more often enmeshed in the criminal justice system, but also argued that variables describing the communities in which blacks live account for a great deal of the crimes they commit. Does their finding hold true using the NYS data? To find out, look at the table for WHITE. Were Blacks more likely than Whites to have smoked marijuana during the past year? It does not appear that we can easily conclude that race affects the incidence of all crimes because this idea did not hold true for our crime.

POINTS TO PONDER: Can you think of some reasons why race might not be related to marijuana smoking?

What if we consider urban inequality or neighborhood factors like Sampson and Wilson advocate doing? You'll be doing that for the "on your own" questions.

Finally, we will turn our attention to INCOME. In his article, John Hagan argues that class is an important factor to consider. It seems logical that poor individuals would be more likely to commit thefts, but let's see if the NYS data support this idea. Is there a significant relationship between INCOME and marijuana use? It is important to note that the lack of findings for a given variable or variables in one dataset does not invalid any theory. It may indicate, for example, that social class is not best measured through income alone as we did in this example. Indeed, researchers are often frustrated by the lack of suitable variables in data collected by others who did not have their particular theory in mind!

#### FURTHER EXPLORATION OF THE CORRELATES OF CRIME

Are other crimes affected by the correlates we examined above? For the "further exploration" questions, you will run the analyses we ran above for two additional crimes, theft and violent offenses. This is an important part of research; we need to know if our findings generalize to other offenses or whether they are limited to a few select crimes.

#### THE CORRELATES OF CRIME ON YOUR OWN

What else might influence the commission of crime?

Now that we have explored some correlates of crime, you could expand your list of independent variables. Try running the analyses we ran for some other potential correlates of crime, and try adding some independent variables of your own. Perhaps you feel that the level of urban development where the respondents live (RURAL) or their marital status (MARSTAT) has effects on crime commission. Explore the variables and find out for yourself. While you are exploring the correlates of crime, reflect on the readings and how they help us understand crime.

#### References

Elliott, D. (1996). National Youth Survey [United States]: Wave VII, 1987 [Computer file]. ICPSR version. Boulder, CO: Behavioral Research Institute [producer], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1996.

**[PLEASE DO NOT TURN IN THESE PRECEDING SHEETS WITH YOUR ASSIGNMENT- THEY ARE FOR YOU TO KEEP]**

Homework for Microcase #2:  
General questions for 'correlates of crime'

Name: \_\_\_\_\_

Date: \_\_\_\_\_

Directions: Answer the following questions by filling in the blanks or circling the appropriate responses. A couple of answers have been filled in for you to make sure you're on the right track.

Exploring the correlates of crime: Getting a handle on the data:

1. In Wave VII of the NYS dataset, there were 700 males, and they comprise \_\_\_\_% of the sample. There were \_\_\_\_ females; they were \_\_\_\_% of the sample. This division *is / is not* roughly equal.
  
2. In Wave VII of the NYS dataset, \_\_\_\_% of the sample was Anglo; \_\_\_\_% were Black; \_\_\_\_% were Hispanic; \_\_\_\_% were American Indian; \_\_\_\_% were Asian; and \_\_\_\_% were Other. Because this division contained some small categories, only the Anglos and Whites were included in the analysis. The breakdown of the new variable, WHITE, was \_\_\_\_% Anglo and \_\_\_\_% Black.
  
3. In Wave VII of the NYS dataset, the age grouping *did / did not* contain a wide range of ages.
  
4. In Wave VII of the NYS dataset, \_\_\_\_% were in the lowest income category (below the 1986 poverty threshold); \_\_\_\_% were in the middle income category and \_\_\_\_% were in the highest income category (more than the 1986 mean income).

Exploring the correlates of crime with crosstabulation tables:

1. In Wave VII of the NYS dataset, \_\_\_\_\_% reported no marijuana use during the previous year, \_\_\_\_\_% used it 1 or more times.
  
2. In the MALE --> USED MJ crosstabulation, 213 (\_\_\_\_%) of the female respondents said they had smoked marijuana during the previous year, compared to \_\_\_\_\_ (\_\_\_\_%) of the male respondents. The difference between the two percentages is \_\_\_\_\_%, which appears to be *negligible / potentially interesting*. From the table, it appears that *males / females* are less likely to smoke marijuana. The Cramer's V value is \_\_\_\_\_, which means that we can do \_\_\_\_\_% better predicting marijuana use when we know a respondent's gender. This relationship is *weak / moderate / strong*. The approximate significance is \_\_\_\_\_, which means there is a \_\_\_\_\_% chance that the relationship between MALE and USED MJ is due to chance. This relationship *is / is not* statistically significant.

In their article, Daniel Mears and his co-authors reported that *men / women* were more likely to break the law, similar to other studies on the gender/criminality link (by the way, this article is based on Wave III of the NYS; the authors obtained the data, tested their theory of exposure to delinquent peers, and published their work!). The relationship between gender and crime (as measured by marijuana use) found using Wave VII of the NYS data *is similar to / differs from* the findings reported by Daniel Mears and his colleagues.

3. In the WHITE --> USED MJ crosstabulation, \_\_\_\_\_ (\_\_\_\_%) of the Anglo respondents said they had smoked marijuana during the previous year, compared to \_\_\_\_\_ (\_\_\_\_%) of the Black respondents. The difference between the two percentages is \_\_\_\_\_%, which appears to be *negligible / potentially interesting*. The Cramer's V value is \_\_\_\_\_, which means that we can do \_\_\_\_\_% better predicting marijuana use when we know a respondent's race. This relationship is *weak / moderate / strong*. The approximate significance is \_\_\_\_\_, which means there is a \_\_\_\_\_% chance that the relationship between WHITE and USED MJ is due to chance. This relationship *is / is not* statistically significant.

In their article, Robert Sampson and William Julius Wilson noted that Blacks are *more / less* likely to engage in violence, but that variables describing the communities in which they live account for much of that difference. The relationship between race and crime (as measured by marijuana use) found using Wave VII of the NYS data *is similar to / differs from* the findings reported by Sampson and Wilson.

4. In the INCOME --> USED MJ crosstabulation, \_\_\_\_\_ (\_\_\_\_%) of the lower income respondents said they had smoked marijuana during the previous year, compared to \_\_\_\_\_ (\_\_\_\_%) of the middle income group, and \_\_\_\_\_ (\_\_\_\_%) of the highest income respondents. The largest difference between the three percentages is \_\_\_\_\_%, which appears to be *negligible / potentially interesting*. The Cramer's V value is \_\_\_\_\_, which means that we can do \_\_\_\_\_% better predicting USED MJ when we know a respondent's income category. This relationship is *weak / moderate / strong*. The approximate significance is \_\_\_\_\_, which means there is a \_\_\_\_\_% chance that the relationship between

INCOME and USED MJ is due to chance. This relationship *is / is not* statistically significant.

In his article, John Hagan stated that social class *has / has not* been linked to criminality. The relationship between income and crime (as measured by marijuana use) found using Wave VII of the NYS data *is similar to / differs from* the statements of Hagan.

POINTS TO PONDER: Can you come up with a possible explanation for the results we found using the NYS data?

5. Overall, we can say our findings regarding correlates for USED MJ *are / are not* consistent with prior research.

Homework for MicroCase #2: "Further exploration" questions  
(correlates of crime)

Name: \_\_\_\_\_

Date: \_\_\_\_\_

TASK: Run the analyses we ran above for marijuana use for two additional crimes: theft and violent offenses.

NOTE: THEFT50 is unchanged from the original dataset, but VIOLENCE was recoded to be '2' for respondents who said they had committed any of six violent offenses included in the survey.<sup>1</sup> Respondents who said they had never committed any of those offenses were coded as '1'.

Answer the following questions by filling in the blanks or circling the appropriate responses. A couple of answers have been filled in for you to make sure you're on the right track.

Further exploration of the correlates of crime:

1. In Wave VII of the NYS dataset, \_\_\_\_\_% of the sample had ever stolen goods worth more than \$50.00 and \_\_\_\_\_% had ever committed a violent offense.
2. In the MALE --> THEFT50 crosstabulation, 111 (\_\_\_\_%) of the male respondents said they had stolen something worth more than \$50.00, compared to \_\_\_\_\_ (\_\_\_\_%) of the female respondents. This relationship is *weak / moderate / strong*. This relationship *is / is not* statistically significant.

The MALE --> THEFT50 table *does / does not* show that men are more likely than women to report having committed a theft of an item worth more than \$50.00.

---

<sup>1</sup> Those offenses were: "attacked someone with the idea of seriously hurting or killing them, gotten involved in a gang fight, used force or strongarm methods to get money or things from someone, had or tried to have sexual relations against someone's will, physically hurt or threatened to hurt someone to get them to have sex, hurt someone to get sex, or deliberately injured spouse/girlfriend/boyfriend".

3. In the WHITE --> THEFT50 crosstabulation, \_\_\_\_\_ (\_\_\_\_%) of the Anglo respondents said they had stolen something worth more than \$50.00, compared to \_\_\_\_\_ (\_\_\_\_%) of the Black respondents. This relationship is *weak / moderate / strong*. This relationship *is / is not* statistically significant.

The WHITE --> THEFT50 table *does / does not* show that Blacks are more likely than Anglos to report having committed a theft of an item worth more than \$50.00.

4. In the INCOME --> THEFT50 crosstabulation, \_\_\_\_\_ (\_\_\_\_%) of the lower income respondents said they had stolen something worth more than \$50.00, compared to \_\_\_\_\_ (\_\_\_\_%) of the middle income group, and \_\_\_\_\_ (\_\_\_\_%) of the highest income respondents. This relationship is *weak / moderate / strong*. This relationship *is / is not* statistically significant.

The INCOME --> THEFT50 table *does / does not* show that poor individuals are more likely than more wealthy people to report having committed a theft of an item worth more than \$50.00.

POINTS TO PONDER: Can you come up with any explanations for the finding in table INCOME --> THEFT50? Could white-collar crime account for some of the finding?

5. In the MALE --> VIOLENCE crosstabulation, 176 (\_\_\_\_%) of the male respondents said they had committed a violent offense, compared to \_\_\_\_ (\_\_\_\_%) of the female respondents. This relationship is *weak / moderate / strong*. This relationship *is / is not* statistically significant.

The MALE --> VIOLENCE table *does / does not* show that men are more likely than women to report having committed a violent offense.

6. In the WHITE --> VIOLENCE crosstabulation, \_\_\_\_ (\_\_\_\_%) of the Anglo respondents said they had committed a violent offense, compared to \_\_\_\_ (\_\_\_\_%) of the Black respondents. This relationship is *weak / moderate / strong*. This relationship *is / is not* statistically significant.

The WHITE --> VIOLENCE table *does / does not* show that Blacks are more likely than Whites to report having committed a violent offense.

7. In the INCOME --> VIOLENCE crosstabulation, \_\_\_\_ (\_\_\_\_%) of the lower income respondents said they had committed a violent offense, compared to \_\_\_\_ (\_\_\_\_%) of the middle income group, and \_\_\_\_ (\_\_\_\_%) of the highest income respondents. This relationship is *weak / moderate / strong*. This relationship *is / is not* statistically significant.

The INCOME --> VIOLENCE table *does / does not* show that poor individuals are more likely than more wealthy people to report having committed a violent offense.

8. Overall, we can say that our findings for THEFT50 *are / are not* consistent with prior research. Overall, we can say that our findings for VIOLENCE *are / are not* consistent with prior research.

Homework for MicroCase #2: "On your own" questions  
(correlates of crime)

Name: \_\_\_\_\_

Date: \_\_\_\_\_

TASK: Expand your list of independent variables to include other potential correlates of crime (e.g., RURAL or MARSTAT). You may choose any variables you feel make sense.

Directions: Answer the following questions.

NOTE: Please print out and include your tables with these questions so your work can be graded.

The correlates of crime on your own:

1. Which variables did you choose as your independent variables?
2. Why did you select these variables?
3. How did your first independent variable affect commission of crime? Make sure to provide a description that includes the percentages, Cramer's V value, the strength of the relationship, and the significance value.
4. How did your second independent variable affect criminality? Make sure to provide a description that includes the percentages, Cramer's V value, the strength of the relationship, and the significance value.

If you included more than two independent variables, you may summarize the findings on the back of this sheet for future reference.