SECI2143 PROBABILITY AND STATISTICAL DATA ANALYSIS
Project II – Inferential Statistics

USA's crime relationship

Muhammad Rafiy Athalla
Section 05

School of Computing, Faculty of Engineering
Universiti Teknologi Malaysia

INTRODUCTION

Crime has happened during the beginning of our civilizations. An act of sin that was done by a man to fulfill his needs in such inappropriate way, an unlawful act punishable by a state or other authority, and many other description was describe for the word 'crime'. There are many kinds of crime that has happened for example robbery, burglary, murder, rape and etc. The police and many other organizations that was working under the law, has taken many steps to prevent them. The crime did reduce but it didn't just disappear. The question now, is there any connection between those crime, are they related to one another, if a person did one of a crime is it possible for they to act another type of crimes?. So, to answer this question, I came up by making this project. The objective of this project is to see is there any connection between the kind of crime to one another, to prove the number of crime that was written in the data exceeded the number of expectations. And maybe by knowing this, we may focused to some crime to reduced the other and maybe one day there will be no crime that happened in our beautiful world.

METHODOLOGY

To carry out the inferential test, my data(Crime data from 1997-2016 in USA) was taken from https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/topic-pages/tables/table-1. The data from the website contains the information about a number of crime that was happen in the USA that was collected by the FBI. An example of the variable that is in the data is the number of violent crime, the number of burglary, the number of murder and the number for other kind of crime that happened in the past.

HYPOTHESIS TESTING 1-SAMPLE

From all of the crime that ever existed, violent crime maybe the most crime that ever been done by any of the criminals and if I may say, all of the crimes contained a violent crime in it even the slightest of them. In this hypothesis testing 1 sample, I want to test either the mean of the violent crime from my dataset (that was recorded by the FBI from the past few year, exactly from the year 1997 until 2016) will be the same as the amount of the mean that I predicted which is about 1 500 000 amount of violent crimes happen within a year with a significance level of 0.05 or 95% confidence level.

μ is the population amount of violent crimes that happen within a year

H0: μ=1 500 000

H1: $\mu \neq$1 500 00

α=0.05

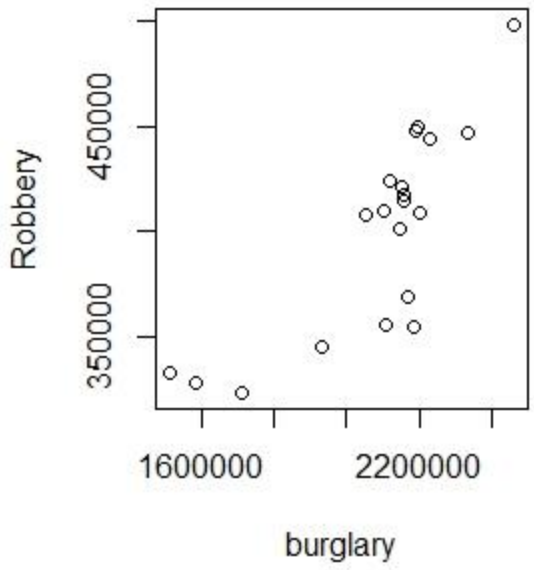
```
> t=(meanhypo1-meanpop)/(sdhypo1/sqrt(n))
> t
[1] -5.167296
>
> t.alpha=qt(alpha/2,n-1)
> t.alpha
[1] -2.093024
```

From the picture above we can see that the t0.025,19(critical) is -2.093024. The H0 will be rejected if the t0*< -t(critical)= -2.093 or if t0*> t(critical)= 2.093024.

Since t0*= -5.167296 < -t(critical)= -2.093024 < t(critical)= 2.093024, we can conclude that it fall within the critical region and we reject H0. There is sufficient evidence to conclude that the mean of violent crimes in USA from 1997 until 2016 is not same as the amount that has been predicted (1 500 00 amount of violent crimes).

CORRELATION

For the correlation, two types of crimes have been chosen by a consideration that these two type of crimes happened to be a similar type of crime. The two type of crimes that I have chosen is burglary and robbery that happen in the USA from 1997 until 2016. Because the type of the data is nominal and not ordinal, so the type of correlation coefficient that will be used in this report is Pearson's product-moment correlation coefficient.



To proof whether the result of the Pearson's product-moment correlation coefficient is reliable we will conduct a test with 0.05 significance level.

H0: There is no linear correlation between burglary and robbery

H1: There is a linear correlation between burglary and robbery

```
> t=corel/sqrt((1-corel^2)/(jum-2))
> t
[1] 6.000347
>
> t.alpha=qt(alpha/2,jum-2)
> t.alpha
[1] -2.100922
```

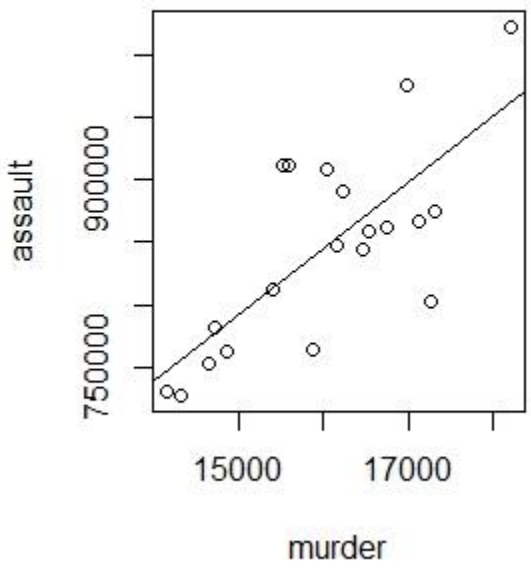
```
> burglary<-(data$Burglary)
> Robbery<-(data$Robbery)
> plot(burglary,Robbery)
> corel=cor(burglary,Robbery)
> corel
[1] 0.8165123
```

From the result of above calculation, since t(critical)= -2.100922 < t(critical)= 2.100922< t0*= 6.000347 and it fall in the critical region, so we reject H0. There s sufficient evidence to proof that there is a linear correlation between burglary and robbery.

Because now we are sure there is a correlation between burglary and robbery, now we c an analyze the scatter plot and the correlation coefficient. By judging from the scatter plot, we ca n know it is a positive correlation because when the amount of the burglary goes up so goes the r obbery. From the result of the correlation coefficient, r =+ 0.8165123, we can conclude that there is a relatively strong positive linear relationship between burglary and robbery. And the amount of the robbery will be increase if the amount of burglary has been increased as well.

REGRESSION

To find a relationship and its effect to another type of crime, this report will show the regression between murder and no negligent manslaughter & aggravated assault. As the aggravated assault as the y and the murder and no negligent manslaughter will be the x. From this, we may see how much the growth of the amount of aggravated assault might happen if a murder and no negligent manslaughter happened, because if it's happened there may some people that be inspire and think that hurting others is fine and they may will not be afraid of the law too later on.




```
> cor(murder,assault)
[1] 0.7383918
> model<-lm(assault~murder)
> model

Call:
lm(formula = assault ~ murder)

Coefficients:
(Intercept)       murder
   -10910.0         53.5
```

From the calculation and the plot above we know the equation for the regression(y= β0 + β1) between murder and no negligent manslaughter and aggravated assault is y= -10910 + 53. 5 x. β0 (estimate value of regression intercept) equal to -10910 indicates the amount of aggravate d assault if there is no murder or no negligent manslaughter. On the other hand the β1(estimate v alue of regression slope) which equal to +53.5 indicates the average amount that grow when ther e happened to a murder or no negligent manslaughter. The coefficient of determination(R^2) for t his relationship is +0.7383918 which mean it is a positive medium linear relationship.

ANNOVA

From the FBI dataset that has been used for this project, three type of crimes have been chosen because a probability that these three type of crimes may have the same population mean. By using a significance level of 0.05(confidence level 95%), an ANOVA test will be conducted for these three types of crime. And these three crimes are larceny theft, motor vehicle theft and property crime.

μ1 will represent population mean for larceny theft

μ2 will represent population mean for motor vehicle theft

μ3 will represent population mean for property crime

α=0.05

H0: μ1= μ2= μ3

H1: At least one of them is different

```
> f=ns/sp
> f
[1] 847.469
>
> num= k-1
> denom=k*(no-1)
> f.alpha=qf(0.95,num,denom)
> f.alpha
[1] 3.158843
```

From the calculation above, we can see that -f(critical)= -3.158843 < f(critical)= 3.158843 < f(test) = 847.469 and it fall n the critical region, so we reject H0. There is a sufficient evidence to conclude that there is neither a same population mean between the mean of these three types of crime. And these three crime have a really different value for its population mean.

CONCLUSION

So , from this project we can conclude that some of these crime may have a connection and some of them may not. Even though the crime look alike, it is not wise to conclude that they have some connection or have a same value of population mean without testing them. Starting from the hypothesis1 sample, we know that violent crime mean between 1997-2016 did not exceed 1 500 000 amount of incident. Next, the burglary and robbery happens to be really related

with a strong relationship. The same as the aggravated assault with murder and no negligent manslaughter, the aggravated assault will grow by consideration with an act of a murder people will see that aggravated assault is not such a bad thing since someone else have committed something even worse. So, we have to try not to increase neither of them, because either increase will affect the other and its decrease will affect otherwise. The last is the difference mean between larceny theft, motor vehicle theft and property crime. Even though they are not too different type of crime but their mean amount of incident is different which proof they did not happen as much as each other.

Appendix

For a better view, suggested to see the file (data.xls) that submitted with the report

---

Year	Population	**Violent_crime**	*Violent_crime_rate*	Murder_and_nonnegligent _manslaughter	*Murder_and_nonnegligent_manslaughter_rate*
1997	267,783,607	1,636,096	611.0	18,208	6.8
1998	270,248,003	1,533,887	567.6	16,974	6.3
1999	272,690,813	1,426,044	523.0	15,522	5.7
2000	281,421,906	1,425,486	506.5	15,586	5.5
2001[5]	285,317,559	1,439,480	504.5	16,037	5.6
2002	287,973,924	1,423,677	494.4	16,229	5.6
2003	290,788,976	1,383,676	475.8	16,528	5.7
2004	293,656,842	1,360,088	463.2	16,148	5.5
2005	296,507,061	1,390,745	469.0	16,740	5.6
2006	299,398,484	1,435,123	479.3	17,309	5.8
2007	301,621,157	1,422,970	471.8	17,128	5.7
2008	304,059,724	1,394,461	458.6	16,465	5.4
2009	307,006,550	1,325,896	431.9	15,399	5.0
2010	309,330,219	1,251,248	404.5	14,722	4.8
2011	311,587,816	1,206,005	387.1	14,661	4.7
2012	313,873,685	1,217,057	387.8	14,856	4.7
2013	316,497,531	1,168,298	369.1	14,319	4.5
2014	318,907,401	1,153,022	361.6	14,164	4.4
2015[6]	320,896,618	1,199,310	373.7	15,883	4.9
2016	323,127,513	1,248,185	386.3	17,250	5.3

Rape(revised_definition)	*Rape(revised_definition)_rate*	Rape_(legacy_definition)	*Rape_(legacy_definition)_rate*	Robbery	*Robbery_rate*	Aggravated_assault	*Aggravated_assault_rate*
		96,153	35.9	498,534	186.2	1,023,201	382.1
		93,144	34.5	447,186	165.5	976,583	361.4
		89,411	32.8	409,371	150.1	911,740	334.3
		90,178	32.0	408,016	145.0	911,706	324.0
		90,863	31.8	423,557	148.5	909,023	318.6
		95,235	33.1	420,806	146.1	891,407	309.5
		93,883	32.3	414,235	142.5	859,030	295.4
		95,089	32.4	401,470	136.7	847,381	288.6
		94,347	31.8	417,438	140.8	862,220	290.8
		94,472	31.6	449,246	150.0	874,096	292.0
		92,160	30.6	447,324	148.3	866,358	287.2
		90,750	29.8	443,563	145.9	843,683	277.5
		89,241	29.1	408,742	133.1	812,514	264.7
		85,593	27.7	369,089	119.3	781,844	252.8
		84,175	27.0	354,746	113.9	752,423	241.5
		85,141	27.1	355,051	113.1	762,009	242.8
113,695	35.9	82,109	25.9	345,093	109.0	726,777	229.6
118,027	37.0	84,864	26.6	322,905	101.3	731,089	229.2
126,134	39.3	91,261	28.4	328,109	102.2	764,057	238.1
130,603	40.4	95,730	29.6	332,198	102.8	803,007	248.5

Property_crime	*Property_crime_rate*	Burglary	*Burglary_rate*	Larceny-theft	*Larceny-theft_rate*	Motor_vehicle_theft	*Motor_vehicle_theft_rate*
11,558,475	4,316.3	2,460,526	918.8	7,743,760	2,891.8	1,354,189	505.7
10,951,827	4,052.5	2,332,735	863.2	7,376,311	2,729.5	1,242,781	459.9
10,208,334	3,743.6	2,100,739	770.4	6,955,520	2,550.7	1,152,075	422.5
10,182,584	3,618.3	2,050,992	728.8	6,971,590	2,477.3	1,160,002	412.2
10,437,189	3,658.1	2,116,531	741.8	7,092,267	2,485.7	1,228,391	430.5
10,455,277	3,630.6	2,151,252	747.0	7,057,379	2,450.7	1,246,646	432.9
10,442,862	3,591.2	2,154,834	741.0	7,026,802	2,416.5	1,261,226	433.7
10,319,386	3,514.1	2,144,446	730.3	6,937,089	2,362.3	1,237,851	421.5
10,174,754	3,431.5	2,155,448	726.9	6,783,447	2,287.8	1,235,859	416.8
10,019,601	3,346.6	2,194,993	733.1	6,626,363	2,213.2	1,198,245	400.2
9,882,212	3,276.4	2,190,198	726.1	6,591,542	2,185.4	1,100,472	364.9
9,774,152	3,214.6	2,228,887	733.0	6,586,206	2,166.1	959,059	315.4
9,337,060	3,041.3	2,203,313	717.7	6,338,095	2,064.5	795,652	259.2
9,112,625	2,945.9	2,168,459	701.0	6,204,601	2,005.8	739,565	239.1
9,052,743	2,905.4	2,185,140	701.3	6,151,095	1,974.1	716,508	230.0
9,001,992	2,868.0	2,109,932	672.2	6,168,874	1,965.4	723,186	230.4
8,651,892	2,733.6	1,932,139	610.5	6,019,465	1,901.9	700,288	221.3
8,209,010	2,574.1	1,713,153	537.2	5,809,054	1,821.5	686,803	215.4
8,024,115	2,500.5	1,587,564	494.7	5,723,488	1,783.6	713,063	222.2
7,919,035	2,450.7	1,515,096	468.9	5,638,455	1,745.0	765,484	236.9