UTM

UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING, FACULTY OF ENGINEERING,
UNIVERSITY TECHNOLOGY MALAYSIA

SECI2143-01 PROBABILITY AND STATISTICAL DATA ANALYSIS

SEM2 2019/2020

PROJECT 2: HOMICIDE CASE REPORT

NAME: NURUL LIYANA BINTI MOHD YUSOF
MATRIC NO: A19EC0144
LECTURER: DR SHARIN HAZLIN RUPIN

INTRODUCTION

The homicide report has reflecting the police classification of homicide deaths as either murder, manslaughter or non-negligent manslaughter with deaths caused by suicide or accidents. Reported statistical data derive from various source, including the FBI's Uniform Crime Reporting (UCR) Program and the FBI's Supplementary Homicide Report from 1976 which compile and made available by Murder Accountability Project. The homicide figure reported in these database is estimated, rather than exact number as the classification are based on police investigation rather than coroner findings or judicial determinations. Despite the limitations, homicide is the least underreported of any serious crime in United States. Homicide data from 1976 to 1999 indicate that nearly 75% of homicide victims are males and 90% to be offenders compare to females. United States also reported to have 4.9 per 100,000 inhabitants for homicide rate and higher murder rates compared to other selected developed countries.

METHODOLOGY

To carry out the statistical data analysis, secondary data is obtained from websites (https://www.kaggle.com/murderaccountabolity/homicide-reports). The target population is citizens of United States of America. 500 out of 638454 data are chosen from 2010 to 2014 (approximately 5 years) of solved cases which are selected randomly from the database. Variables such as gender, relationship, weapon, type of crime and victim race are selected and processed to be analysed. Inferential statistics are carried out by using Hypothesis testing, Correlation, Regression, Chi Square test of independence and Goodness-Fit Chi Square Test.

RESULT AND DISCUSSION

1. *Hypothesis Testing of 2 Sample Test (using 2 Proportions)*

Two sample test are used to support the claim "Males are more likely to be killed than females" which can be expressed as $p_1 > p_2$. If the $p_1 < p_2$, then the statement is false.

The hypothesis can be written as:

$$H_0: p_1 = p_2$$

$$H_1: p_1 > p_2$$

The significant level is $\alpha = 0.05$

	Group1 : Male	Group2 : Female
Murder /Manslaughter	x1 = 345	x2 = 132
Subjects in group	n1 = 365	n2 = 135

Table1: Number of victims in homicide case based on gender

```
        2-sample test for equality of proportions with continuity correction

data:  c(345, 132) out of c(365, 135)
X-squared = 1.6981, df = 1, p-value = 0.9037
alternative hypothesis: greater
95 percent confidence interval:
 -0.06627044  1.00000000
sample estimates:
   prop 1    prop 2
0.9452055 0.9777778
```

Figure1: Calculation for Hypothesis Testiing

The variables used is gender and Homicide case (murder or manslaughter) with total n_1=365, n_2=135 and each variable holds value x_1 (male) = 365 while x_2 (female) = 135. The p-value for test statistic are obtained by using R. The p-value (p=0.9037) is a two-tailed p-value testing the null hypothesis. Since the p-value is more than conventional 0.05, we reject the H_0. There is sufficient evidence to support the claim that males are more likely to be killed than females.

2. Correlation Analysis

A random sample of 500 homicide cases are selected for this test. The type of correlation coefficients we used in this report is Pearson's product-moment correlation coefficient to test the relationship between homicide case and number of cases with weapon involved.

The hypothesis can be written as:

$$H_0: \rho = 0 \text{ (No linear correlation)}$$

$$H_1: \rho \neq 0 \text{ (Linear correlation exist)}$$

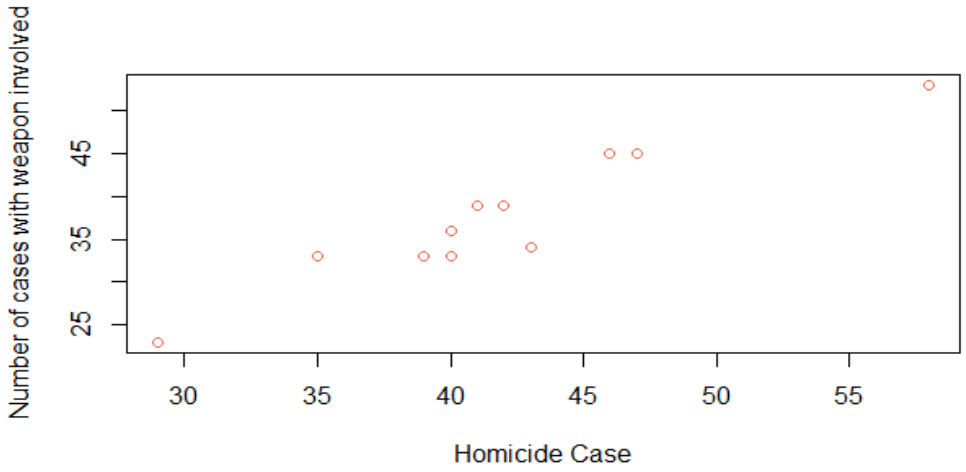
The significant level is $\alpha = 0.05$

```
        Pearson's product-moment correlation

data:  x and y
t = 9.5815, df = 10, p-value = 2.349e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8257252 0.9861048
sample estimates:
      cor
0.9496174
```

Figure2: Calculation for Correlation Analysis

The result above is the significance test of the correlation of x and y. x represents variable homicide case whereas y represents number of cases with weapon involved. The p-value is p = 2.34^{-06} which is less than the significance level $\alpha = 0.05$, thus we reject H_0. Therefore, there is sufficient evidence of linear correlation exist between homicide cases and number of cases with weapon involved with correlation coefficient of +0.9496174 at the 5% level confidence.



Graph1: Scatter plot of homicide cases and number of cases with weapon involved

Based on the graph above, we can conclude that the use of weapon in United States resulting in increment in homicide case since US are able to obtain personal arms. From the scatter plot and the correlation coefficient value, r = +0.9496174 significant at the level of 0.05, we can state that there is strong positive linear relationship between x, homicide cases and y, number of cases with weapon involved.

3. *Linear Regression Model*

A random sample of 500 victims are selected from homicide database. The independent variable (x) in this case is victim race whereas the dependent variable (y) is the white perpetrator. The analysis is to aim the existence of a linear relationship between the variable x and y.

H_0: $\beta = 0$ (No linear relationship)

H_1: $\beta \neq 0$ (Linear relationship exist)

The significant level is $\alpha = 0.05$

```
[1] 0.7554818
> mod<- lm( reg$`White Perpretrator`~reg$`Victim Race`)
> summary (mod)

Call:
lm(formula = reg$`White Perpretrator` ~ reg$`Victim Race`)

Residuals:
      1       2       3       4       5
  5.426 -82.711  -1.288  11.425  67.148

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -10.1392    39.8235  -0.255    0.815
reg$`Victim Race`   0.5714     0.2861   1.997    0.140

Residual standard error: 61.95 on 3 degrees of freedom
Multiple R-squared:  0.5708,    Adjusted R-squared:  0.4277
F-statistic: 3.989 on 1 and 3 DF,  p-value: 0.1397

> attributes(mod)
$names
 [1] "coefficients"  "residuals"     "effects"       "rank"          "fitted.values" "assign"
 [7] "qr"            "df.residual"   "xlevels"       "call"          "terms"         "model"

$class
[1] "lm"

> mod$coef
      (Intercept) reg$`Victim Race`
        -10.139201           0.571392
> coef(mod)
      (Intercept) reg$`Victim Race`
        -10.139201           0.571392
```

Figure3: Calculation for Regression

Graph2: White Perpetrator against Victim Race

From the regression graph, this model has shown a positive linear relationship with an equation of $y = -10.139201 + 0.571392x$. The value of β_0 of -10.139201 indicates that the range of white perpetrators on victim race. The value of slope coefficient, β_1 of +0.571392 just indicates the victim race that were killed by white perpetrator. It shown that white perpetrator is intra-racial which 47.6% of white victim were killed by white perpetrator. Therefore, we know that there is a relationship between the variables x and y based on the calculated value of β. The coefficient of determination, R^2, in this relationship is 0.7554818 which is close to 1 and can be consider as a moderate relationship.

4. Chi-Square Test of Independence

A random sample of 500 victims is selected for this test. In this study, we would like to find out whether there is a relationship between relation of victim and offender and gender of victim.

The hypothesis can be written as:

H_0: No relation between relation of victim and offender and gender

H_1: Relationship between relation of victim and offender and gender exist

The significant level is $\alpha = 0.05$

```
> mon
                            Male Female
Have relationship with victim  141     73
No relationship with victim    224     62
> chisq.test(mon)

        Pearson's Chi-squared test with Yates' continuity correction

data:  mon
X-squared = 8.9809, df = 1, p-value = 0.002728

> alpha <-0.05
> x2.alpha<-qchisq(alpha, df=1,lower.tail=FALSE)
> x2.alpha
[1] 3.841459
>
```

Figure4: Calculation for Goodness-Fit Chi-Square Test

From the result above, the value of chi-square test is x-squared = 8.9809 and critical value = 3.841459. Since the value of x-squared > critical value, we reject H_0. Therefore, we can conclude that Relationship exist between relation of victim and offender and gender. In short, the relationship between relation of victim and offender and gender implies higher homicide case.

5. *Goodness-fit Chi Square Test*

A sample of 500 homicide cases has been selected to observed male victims that were killed with equal proportion for 12 months in 2010 to 2014.

Month	**Jan**	**Feb**	**March**	**Apr**	**May**	**June**	**July**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Observed Cases	26	23	35	26	45	34	24	28	30	39	26	29

Table2: Observed Cases on Male Victims

H_0: $p_1=p_2=p_3=p_4=p_5=p_6=p_7=p_8=p_9=p_{10}=p_{11}=p_{12}$

H_1: At least one of the probabilities is different from others

The significant level is $\alpha = 0.05$

```
> res

        Chi-squared test for given probabilities

data:  x
X-squared = 15.877, df = 11, p-value = 0.1458

> res$observed
 [1] 26 23 35 26 45 34 24 28 30 39 26 29
> res$expected
 [1] 30.41667 30.41667 30.41667 30.41667 30.41667 30.41667 30.41667 30.41667 30.41667 30.41667 30.41667 30.41667
> res$p.value
[1] 0.14577
> alpha <-0.05
> x2.alpha<-qchisq(alpha, df=11,lower.tail=FALSE)
> x2.alpha
[1] 19.67514
>
```

Figure5: Calculation of Goodness-Fit Test

From the result above, the value of goodness-fit test is x-squared = 15.877 with critical value = 19.67514. Since the value of x-squared < critical value, thus, we fail to reject null hypothesis. Therefore, there is sufficient evidence to support the claim that the male victims were killed with equal proportion for 12 months in 2010 to 2014. Hence, the number of male victims are steady with no decrement.

CONCLUSION

In conclusion, United States remarks the male victim as the majority of the victim. We observe that sex difference rates of homicide victimization have not change. The trend of homicide remains the same. In correlation analysis, the graph showed positive relation. The human rights to own personal arms has affected the country by increment of homicide victimization. The use of personal arms by irresponsible citizens compose access to murder and increase the risk of others life. Moreover, the regression showed the positive relation between white perpetrator and victim race which indicates that race and crime are interrelated and most of the cases are between their own race. The pattern of race is a bit different as previously statistic recorded that 13% out of 50% population of blacks commit murder. However, the data show that whites commit more murder than blacks according to database of 500 victims. A final component of homicide within the trends in risk is relationship of victim and offender and gender. It displays that relationship is important factors that allow homicide cases flop especially when the relation is unknown or stranger.