ADVANCE CRIME STATISTICS-LEGAL INFORMATICS AND FORENSIC INSTITUTE

Alex Ogbole

12/13/2019

DEFINITIONS OF VARIABLES:

Criminal damage & arson: Criminal damage is the intentional and malicious damage to the home, other property or vehicles and includes graffiti. Arson is the act of deliberately setting fire to property, including buildings and vehicles.

Public disorder and weapons:Public disorder is violent activity such as rioting or fighting in public place. Weapons offence is any offensive weapon or any article made, adapted, or intended for incapacitating someone.

Shoplifting: the criminal action of stealing goods from a store while pretending to be a customer.

Other theft: relates to theft from a person without the threat or use of force or violence which includes theft by an employee, blackmail and making off without payment.

Drugs: offense refers to the possession, use, sale or furnishing of any drug or intoxicating substance or drug paraphernalia, that is prohibited by law

Research Question: Does Public disorder and weapons affect Shoplifting?

Goal of Research

The goal of the research is to Predict correlation between Public disorder and weapons and shoplifting and which potentially could be used in decision making in the criminal justice system in the United Kingdom

Hypothesis:

Assuming that Public disorder is a potential factor that may be delibrate in causing distraction in a particular area

There may be a correlation between Shoplifting and other dependent variables.

There may possibly be a relationship between variables that could provide a linear regression

There may be a possible relationship between the variables that may provide a Multiple linear regression

Methodology

1.Identify Sample set and Population size

Daily neighbourhood crime data of the United kingdom from jan-dec 2012 is used for this experiment.

2.Collection and cleaning of datasets and categorizing data types to be used which would be numeric.

Datasets are collected from United Kingdom police website and required variables are selected programmatically in r.

1a.THE CORRELATION OF VARIABLES

```r
cor_of_data<-cor(binded_files[c('Criminal.damage.and.arson',
                                'Shoplifting', 'Other.theft',
                                'Drugs', 'Public.disorder.and.weapons')],
                 method = c("pearson"))
round(cor_of_data,2)
```

```
##                             Criminal.damage.and.arson Shoplifting
## Criminal.damage.and.arson                        1.00        0.55
## Shoplifting                                      0.55        1.00
## Other.theft                                      0.49        0.67
## Drugs                                            0.59        0.51
## Public.disorder.and.weapons                      0.65        0.69
##                             Other.theft Drugs Public.disorder.and.weapons
## Criminal.damage.and.arson          0.49  0.59                        0.65
## Shoplifting                        0.67  0.51                        0.69
## Other.theft                        1.00  0.61                        0.67
## Drugs                              0.61  1.00                        0.65
## Public.disorder.and.weapons        0.67  0.65                        1.00
```

It appears that at 0.69 correlation coefficient between public.disorder.and.weapons and shoplifting there is a positive relationship. If the scatterplot doesn't indicate there's at least somewhat of a linear relationship, the correlation doesn't mean much.

1b. Let's compute the linear regression model between public.disorder.and.weapons and shoplifting

General linear formular y = ax + b

Note: "a" intercept, b = coefficients and x = predictor

```
pub.dis.w<- binded_files$Public.disorder.and.weapons
shoplifting<- binded_files$Shoplifting
lmodel<-lm(shoplifting~pub.dis.w)
print(summary(lmodel))
```

```
##
## Call:
## lm(formula = shoplifting ~ pub.dis.w)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -81.817  -2.582  -0.813   1.187 199.969
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.813352   0.035441   22.95   <2e-16 ***
## pub.dis.w   1.768164   0.007446  237.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.872 on 63255 degrees of freedom
## Multiple R-squared:  0.4713, Adjusted R-squared:  0.4713
## F-statistic: 5.639e+04 on 1 and 63255 DF,  p-value: < 2.2e-16
```

a=0.813352 b=1.768164

Derived equation: shoplifting = 0.813352 * Public.disorder.and.weapons + 1.768164

THE LINEAR MODEL EVALUATES THE FOLLOWING

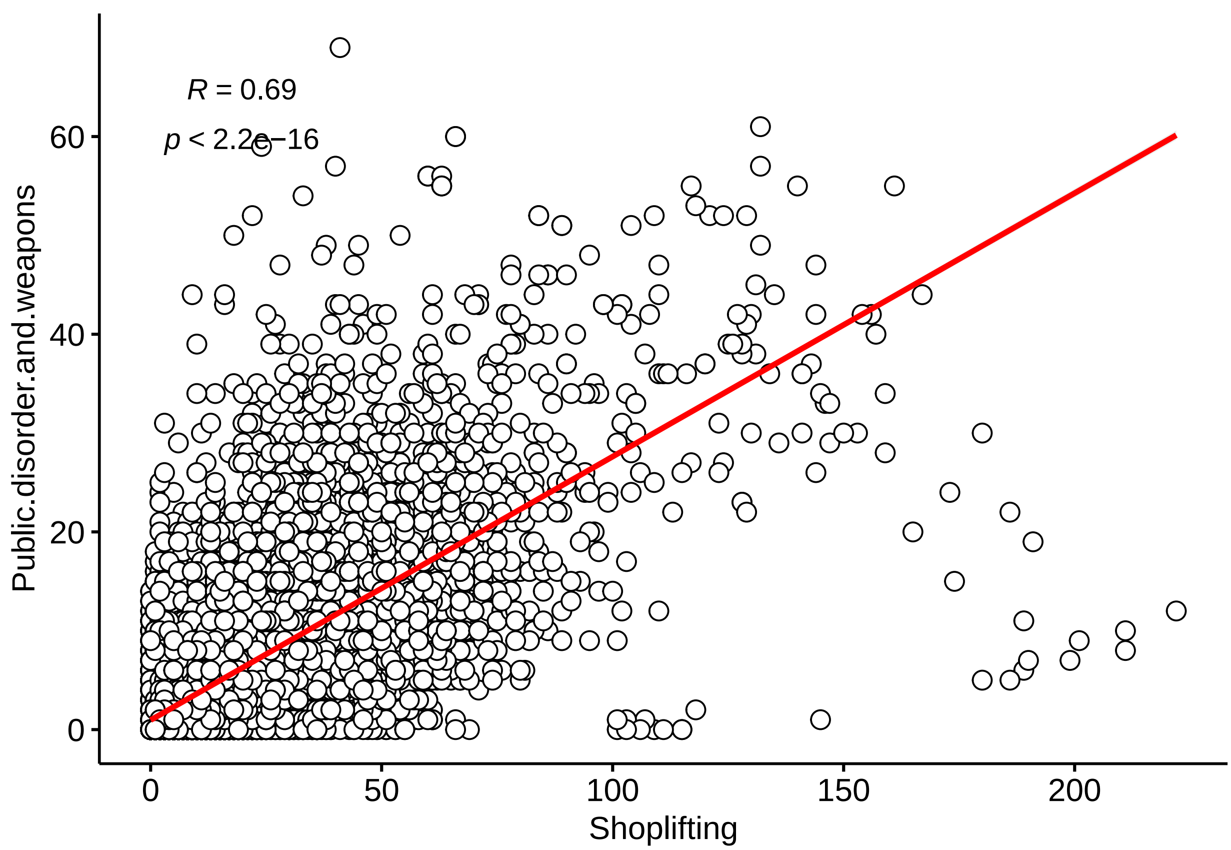
1. t-value = 237.46
2. Degree of freedom = 63255

3. p-value (Significance level of t-value) = 2.2e-16
4. F-Statistic: 5.639e+04 on 1

#A LINEAR GRAPH SHOWING THE RELATIONSHIP

```
df<-binded_files[, c("Public.disorder.and.weapons", "Shoplifting")]

library(ggpubr)
ggscatter(df, x = "Shoplifting", y = "Public.disorder.and.weapons",
   color = "black", shape = 21, size = 3,
   add = "reg.line",
   add.params = list(color = "red", fill = "lightgray"),
   conf.int = TRUE,
   cor.coef = TRUE,
   cor.coeff.args = list(method = "pearson", label.x = 3, label.sep = "\n")
   )
```



From the graph, there isn't a very clear linear representation given the distribution of the variables. Now let's add some more variables to the model. This multiple model might increase or decrease the existing relationship.

#1c. MULTIPLE REGRESSION

General multiple model formular y = a + s1x1 + s2x2 +…snxn Note "a" is the intercept, s1 and s2 are coefficients and x1,x2 predictors.

```r
cri.arson<-binded_files$Criminal.damage.and.arson
other.t<-binded_files$Other.theft
drugs<-binded_files$Drugs
mmodel<-lm(shoplifting~pub.dis.w+other.t+cri.arson+drugs)
print(summary(mmodel))
```

```
##
## Call:
## lm(formula = shoplifting ~ pub.dis.w + other.t + cri.arson +
##     drugs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -88.779  -2.147  -0.381   0.886 145.508
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.378253   0.035576  -10.63   <2e-16 ***
## pub.dis.w    0.927145   0.010868   85.31   <2e-16 ***
## other.t      0.189917   0.001823  104.19   <2e-16 ***
## cri.arson    0.134231   0.003226   41.61   <2e-16 ***
## drugs       -0.075835   0.006534  -11.61   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.136 on 63252 degrees of freedom
## Multiple R-squared:  0.5656, Adjusted R-squared:  0.5656
## F-statistic: 2.059e+04 on 4 and 63252 DF,  p-value: < 2.2e-16
```

Negative t-value in response variable means there is a reversal in the directionality of the model, when three independent variables were introduced thereby causing a negetive intercept on the regression slope.

TESTING THE VARIABLES INDEPENDENTLY ON THE MODEL

```r
mmmodel<-lm(shoplifting~pub.dis.w+cri.arson)
print(summary(mmmodel))
```

```
##
## Call:
## lm(formula = shoplifting ~ pub.dis.w + cri.arson)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -83.719  -2.341  -0.398   0.830 203.538
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.088956   0.038263   2.325   0.0201 *
## pub.dis.w   1.479624   0.009648 153.355   <2e-16 ***
## cri.arson   0.154460   0.003362  45.948   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 7.744 on 63254 degrees of freedom
## Multiple R-squared:  0.4884, Adjusted R-squared:  0.4884
## F-statistic: 3.019e+04 on 2 and 63254 DF,  p-value: < 2.2e-16
```

Criminal damage and arson as an added variable clamps the effect of the model with a t-value of 2.32 and intercept of 0.088 not being significant.

```
mmmmodel<-lm(shoplifting~pub.dis.w+drugs)
print(summary(mmmmodel))
```

```
##
## Call:
## lm(formula = shoplifting ~ pub.dis.w + drugs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -79.323  -2.373  -0.599   1.007 202.079
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.598953   0.035909   16.68   <2e-16 ***
## pub.dis.w   1.577302   0.009753  161.72   <2e-16 ***
## drugs       0.197207   0.006572   30.01   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.817 on 63254 degrees of freedom
## Multiple R-squared:  0.4787, Adjusted R-squared:  0.4787
## F-statistic: 2.904e+04 on 2 and 63254 DF,  p-value: < 2.2e-16
```

Drugs as an added variable clamps the effect of the model with a t-value of 16.68 and intercept of 0.59 being only slightly significant.

```
mmmmmodel<-lm(shoplifting~pub.dis.w+other.t)
print(summary(mmmmmodel))
```

```
##
## Call:
## lm(formula = shoplifting ~ pub.dis.w + other.t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -100.403   -2.069   -0.548    0.864  142.944
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.168081   0.033105   5.077 3.84e-07 ***
## pub.dis.w   1.103735   0.009198 120.000  < 2e-16 ***
## other.t     0.190136   0.001759 108.077  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 7.232 on 63254 degrees of freedom
## Multiple R-squared:  0.5537, Adjusted R-squared:  0.5537
## F-statistic: 3.924e+04 on 2 and 63254 DF,  p-value: < 2.2e-16
```

Other crime as an added variable also does not improve the model with a t-value of 5.077 and intercept of 0.16 .

CONCLUSION

At the end of this experiment, a linear model upheld the hypothesis that Public disorder and weapons affect Shoplifting with a correlation coefficient of 0.69 and t-value of 237.46. This is sufficient enough to create model for duducing shoplifting crime giving the x amount of public disorder.

In the multiple model, using four variables as predictors appeared not to have a significant effect on the model. Using the variables individually also did not improve the accuracy of the model.