

Advance Crime Statistics Final Assessment

Luckson

12/13/2019

1.INTRODUCTION This paper is based on crime statistical analysis looking at Department of Corrections facilities for both man and women for at least 4 years with excel data that 5 variables which are the facilities.

2.RESEARCH QUESTIONS Crime related Question from Incarceration data Having male incarcerated in Central Detention Facility give rise to their female counterparts detention??

2.i.HYPOTHESIS- Central Detention Facility_Males numbers has no significant effect to the number of people in Correctional Treatment Facility_Males

3.RESEARCH METHODOLOGY The study shall use R with R-Markdown for analysing different crime hypotheses to come out with Department of Corrections facilities relational variables given the numbers of people that are checked in on a daily basis.Also to be used is Linear and multiple regression to get correlations,mean and levels of significance on different variables.An hp laptop with a 64bit capacity is to be used.

```
## # A tibble: 5,179 x 5
##   CDF_M CTF_M CDF_J CTF_F CDF_F
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  1817   320    29   320   329
## 2  1810   320    29   320   330
## 3  1809   328    28   187   196
## 4  1810   352    28   186   193
## 5  1831   353    28   188   195
## 6  1829   331    28   186   193
## 7  1826   329    29   185   192
## 8  1832   329    29   185   192
## 9  1839   324    28   187   194
## 10 1833   331    28   190   198
## # ... with 5,169 more rows
```

KEY = Central Detention Facility_Males(CDF_M)

##Correctional Treatment Facility_M(CTF_M) ##Central Detention Facility_J(CDF_J) ##Correc-
tional Treatment Facility_F(CTF_F) ##Central Detention Facility_F(CDF_F)

Summary of the data frame (incarceration)

```
summary (incarceration)
```

```
##           CDF_M           CTF_M           CDF_J           CTF_F
##  Min.      : 977    Min.      :219.0    Min.      : 9.00    Min.      : 83.0
## 1st Qu.:1246    1st Qu.:325.0    1st Qu.:19.00    1st Qu.:103.0
##  Median :1614    Median :354.0    Median :23.00    Median :132.0
##   Mean   :1560    Mean   :397.1    Mean   :24.99    Mean   :140.3
## 3rd Qu.:1861    3rd Qu.:409.0    3rd Qu.:33.00    3rd Qu.:160.0
```

```
## Max. :2166 Max. :747.0 Max. :41.00 Max. :320.0
## CDF_F
## Min. : 89.0
## 1st Qu.:134.0
## Median :167.0
## Mean :178.7
## 3rd Qu.:238.0
## Max. :330.0
```

```
corincarceration <- incarceration [,c("CDF_M","CTF_M","CDF_J","CDF_F","CTF_F")]
corincarceration
```

```
## # A tibble: 5,179 x 5
## CDF_M CTF_M CDF_J CDF_F CTF_F
## <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1817 320 29 329 320
## 2 1810 320 29 330 320
## 3 1809 328 28 196 187
## 4 1810 352 28 193 186
## 5 1831 353 28 195 188
## 6 1829 331 28 193 186
## 7 1826 329 29 192 185
## 8 1832 329 29 192 185
## 9 1839 324 28 194 187
## 10 1833 331 28 198 190
## # ... with 5,169 more rows
```

```
cor(corincarceration)
```

```
## CDF_M CTF_M CDF_J CDF_F CTF_F
## CDF_M 1.0000000 0.25624273 0.53881789 0.1919651 0.6517332
## CTF_M 0.2562427 1.00000000 0.09721482 0.4133950 0.1508112
## CDF_J 0.5388179 0.09721482 1.00000000 0.2530362 0.3673701
## CDF_F 0.1919651 0.41339502 0.25303622 1.0000000 0.2382757
## CTF_F 0.6517332 0.15081120 0.36737011 0.2382757 1.0000000
```

```
## coefficient in a regression indicates that the variable in
## question is linearly related to the other variables with contributions
## to change to the other variable.
## the coefficients in multi-regression mean that variables
## make add value to any hypothesis
```

```
linearMod <- lm(CDF_M ~ CDF_F, data=corincarceration)
print(linearMod)
```

```
##
## Call:
## lm(formula = CDF_M ~ CDF_F, data = corincarceration)
##
## Coefficients:
## (Intercept) CDF_F
## 1347.784 1.185
```

```
# linear regression model
```

```
summary (linearMod)
```

```
##  
## Call:  
## lm(formula = CDF_M ~ CDF_F, data = corincarceration)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -648.09 -288.63    1.39  285.97  686.57   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 1.348e+03  1.578e+01  85.44  <2e-16 ***  
## CDF_F        1.185e+00  8.422e-02  14.07  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 340.3 on 5177 degrees of freedom  
## Multiple R-squared:  0.03685,    Adjusted R-squared:  0.03666   
## F-statistic: 198.1 on 1 and 5177 DF,  p-value: < 2.2e-16
```

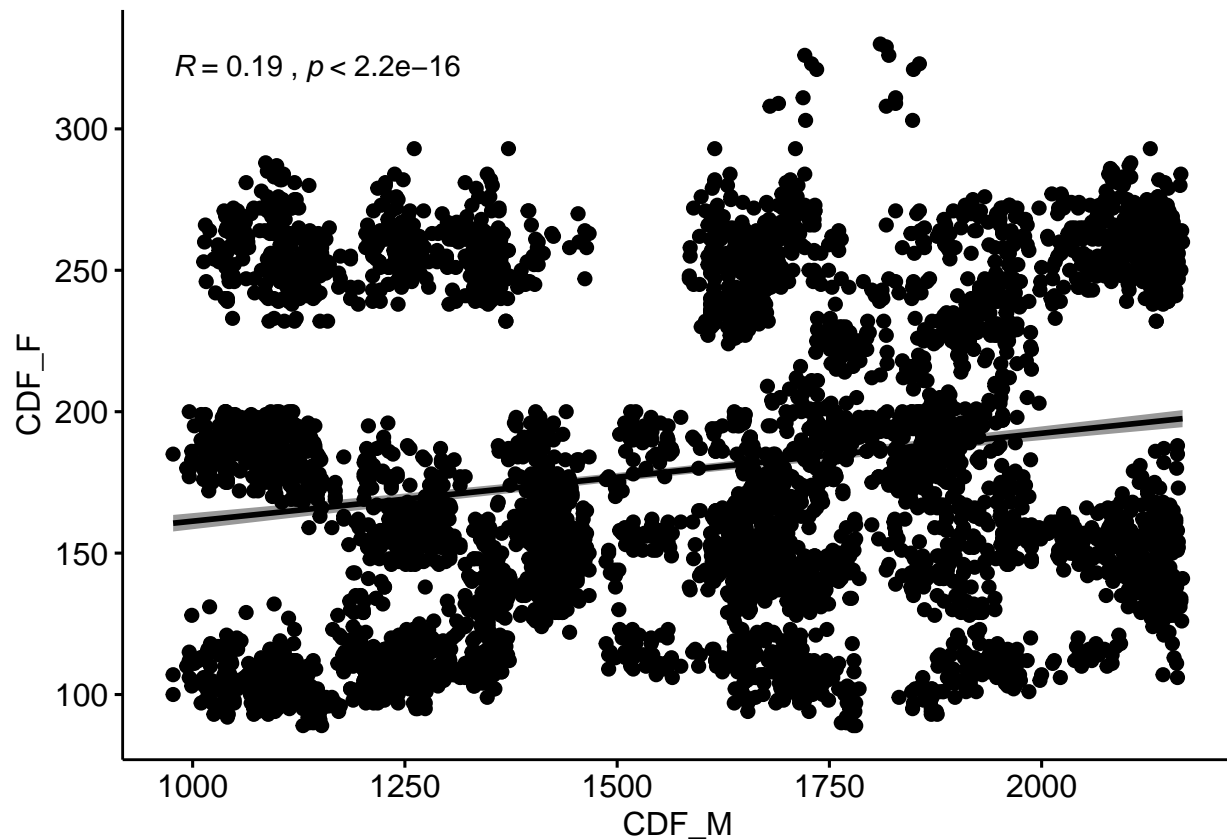
```
##p-value: < 2.2e-16 shows is above the pre-determined  
#significance level of 0.05  
##the CDF_F variable has more significance in this linear  
#model,the more the NUMBER in CDF-M the increase in number of CDF_F
```

```
library("ggpubr") ## using Pearson model
```

```
## Loading required package: ggplot2
```

```
## Loading required package: magrittr
```

```
ggscatter(corincarceration, x = "CDF_M", y = "CDF_F",  
          add = "reg.line", conf.int = TRUE,  
          cor.coef = TRUE, cor.method = "pearson")
```



```
corgraph <- cor.test(corincarceration$CDF_M, corincarceration$CDF_F)
                    method = ("pearson")
corgraph
```

```
##
## Pearson's product-moment correlation
##
## data: corincarceration$CDF_M and corincarceration$CDF_F
## t = 14.074, df = 5177, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1655949 0.2180610
## sample estimates:
##      cor
## 0.1919651
```

```
##Is women incarceration in CDF_F affected by changes in
#other variables like (CDF_M,CTF_F,CDF_F) which is the
#question that I would like to ask ?? ## Multiple Regression
```

```
MultiModel_1 <- lm(CDF_F ~ CDF_M+CTF_F+CDF_F+CDF_J, data=corincarceration)
```

```
## Warning in model.matrix.default(mt, mf, contrasts): the response appeared
## on the right-hand side and was dropped
```

```
## Warning in model.matrix.default(mt, mf, contrasts): problem with term 3 in
## model.matrix: no columns are assigned
```

```
## This 1st model of multi-regression assumes that Y is CDF_F AND X are all other variables
print(MultiModel_1)
```

```
##
## Call:
## lm(formula = CDF_F ~ CDF_M + CTF_F + CDF_F + CDF_J, data = corincarceration)
##
## Coefficients:
## (Intercept)      CDF_M      CTF_F      CDF_J
## 122.887742    -0.007076     0.237580     1.340990
```

```
##(Intercept)122.887742      CDF_M-0.007076      CTF_F 0.237580      CDF_J 1.340990
summary (MultiModel_1)
```

```
##
## Call:
## lm(formula = CDF_F ~ CDF_M + CTF_F + CDF_F + CDF_J, data = corincarceration)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -112.093  -42.184   -6.582   36.466  143.103
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 122.887742   3.460867  35.508  <2e-16 ***
## CDF_M        -0.007076   0.003128  -2.262   0.0237 *
## CTF_F         0.237580   0.021773  10.912  <2e-16 ***
## CDF_J         1.340990   0.102319  13.106  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 53.6 on 5175 degrees of freedom
## Multiple R-squared:  0.08934,    Adjusted R-squared:  0.08881
## F-statistic: 169.2 on 3 and 5175 DF,  p-value: < 2.2e-16
```

```
##This 1st model shows the inability of 3 variables to have a significant effect on Y
```

```
MultiModel_2 <- lm(CDF_F ~ CDF_M+CTF_F, data=corincarceration)
## The 2nd hypothesis will only have 2 variables to find out their levels of significance to Y
print(MultiModel_2)
```

```
##
## Call:
## lm(formula = CDF_F ~ CDF_M + CTF_F, data = corincarceration)
##
## Coefficients:
## (Intercept)      CDF_M      CTF_F
## 128.24409      0.01032      0.24482
```

```
##This result shows the ability of these two variables having a stronger level of significance
##(Intercept)      CDF_M      CTF_F
## 128.24409      0.01032      0.24482
summary (MultiModel_2)
```

```
##
## Call:
## lm(formula = CDF_F ~ CDF_M + CTF_F, data = corincarceration)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -107.11  -46.05  -10.35   43.64  156.73
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.282e+02  3.493e+00  36.716 < 2e-16 ***
## CDF_M       1.032e-02  2.879e-03   3.586 0.000338 ***
## CTF_F       2.448e-01  2.212e-02  11.067 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.48 on 5176 degrees of freedom
## Multiple R-squared:  0.05911,    Adjusted R-squared:  0.05875
## F-statistic: 162.6 on 2 and 5176 DF,  p-value: < 2.2e-16
```

```
##This hypothesis with two variables which answers our 1st question
```

Conclusion

Incarceration data for people who would have committed various criminal acts with both male and females being held at two distinct Department of Corrections Facilities. So the data used showed that women numbers being held up are much less than that of their male counterparts. Central Detention Facility_Females was mainly a constant variable with other variables being predictors coming out with different results or outcomes.