Case Study 2: Describing and Testing the Association Between Trust in Police and Assault Victimisation in Algeria Using Crosstabulations

Nicholas Tratjenberg-Pareja, Tomáš Diviák

2024-02-05

This activity shows how to undertake bivariate and inferential statistical analysis. The idea is to analyse the relationship between the experience of being victimized and perceptions of trust in public institutions in Africa.

There is a long tradition in criminology of research on public attitudes toward criminal justice and legal institutions given its role on compliance and cooperation with law enforcement (Bradford & Jackson, 2010; Sherman, 2002; Tyler, 2006). One of the key determinants of trust in Criminal Justice Institutions is crime experiences. Previous studies have shown that citizens that were victims of crime are more likely to show more negative perceptions of police and less confidence in them (Berthelot et al., 2018; Callahan & Rosenberg, 2011; O’Connor, 2008; Sprott & Doob, 2009). Alternatively, some research has shown that those that trust less in the police are less likely to report violent crimes (Kääriäinen & Sirén, 2011). Yet, most of the research has been conducted in North America and Europe. Exploring the relationship between victimization and trust in police in the global south is worth considering the elevated levels of violence observed, particularly in Latin America and Africa (UNOCD, 2O23) and also by the under-resourced and weak criminal justice institutions in these regions (Bergman, 2018). Some recent research has shown that trust in police is also higher among citizens that have not been victimized in african countries such as South Africa (Olutola & Bello, 2016), and even that this difference in trust in police due to victimization is a characteristic of developing countries and not so much in more industrialized ones (Alda et al., 2017).

library(tidyverse) # for data transformations etc.  
library(haven) # for importing data  
library(ggplot2) # for visualisations  
library(gmodels) # for crosstables  
library(lsr) # Cramer's V

In this context, we are interested to know if there is a relationship between victimization and trust in a criminal justice institution in Algeria. This exercise will allow us to explore creation of new variables collapsing categories, building cross tables, and analysis of association and testing independence.

In order to explore the relationship between victimisation and trust in criminal justice institutions in Algeria, we will use the Algerian part of the Afrobarometer dataset (Hammani et al., 2015). Using the haven package in R (Wickham & Miller, 2020), we will load the dataset and inspect the two variables of interest: having been a victim of assault (named Q11B in the original data) and trust in police (Q52H).

alg <- read\_sav("alg\_r6\_data.sav")  
View(alg)  
  
# Q52H = trust in police  
# 0 = not at all; 1 = a little; 2 = somewhat; 3 = a lot; 9 = NA  
table(alg$Q52H)

##   
## 0 1 2 3 9   
## 156 330 337 323 54

prop.table(table(alg$Q52H))

##   
## 0 1 2 3 9   
## 0.1300000 0.2750000 0.2808333 0.2691667 0.0450000

str(alg$Q52H)

## dbl+lbl [1:1200] 2, 1, 2, 3, 2, 2, 9, 3, 1, 1, 9, 3, 1, 2, 1, 1, 3, 2, 2, ...  
## @ label : chr "Q52h. Trust police"  
## @ format.spss : chr "F2.0"  
## @ display\_width: int 7  
## @ labels : Named num [1:7] -1 0 1 2 3 9 98  
## ..- attr(\*, "names")= chr [1:7] "Missing" "Not at all" "Just a little" "Somewhat" ...

# Q11B = how often have been physically attacked  
# 0 = no; 1 = once; 2 = twice; 3 = three or more; 9 = NA  
table(alg$Q11B)

##   
## 0 1 2 3 9   
## 1045 100 39 12 4

prop.table(table(alg$Q11B))

##   
## 0 1 2 3 9   
## 0.870833333 0.083333333 0.032500000 0.010000000 0.003333333

str(alg$Q11B)

## dbl+lbl [1:1200] 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## @ label : chr "Q11b. Have been physically attacked"  
## @ format.spss : chr "F2.0"  
## @ display\_width: int 7  
## @ labels : Named num [1:7] -1 0 1 2 3 9 98  
## ..- attr(\*, "names")= chr [1:7] "Missing" "No" "Yes, once" "Yes, twice" ...

Both the variables have five different possible values as shown in tables N and M. In terms of experience with having been assaulted, almost 80% of respondents declared having no such experience in the past year, whereas 14% of respondents claimed to have been victims once with the remaining options not even adding up to 1%. The variable measuring trust in police displays quite an equal distribution of respondents declaring at least some level of trust in police with their share of respondents fluctuating between 25% and 30%, while only 13% percent of respondent declared no trust in police at all. Values ‘9’ refer to values that we would like to filter out from the analysis as they essentially denote a missing observation. To do this, we use the tidyverse suite of R packages (Wickham et al., 2019) and we also simultaneously select only the two variables we will be working with in this example using the pipe operator (‘%>%’).

# removing 9s (NAs) from variables  
algS <- alg %>%   
 # select only the columns to be analysed  
 select(Q11B, Q52H) %>%   
 # filter out only valid responses (9 = NA)  
 filter(Q11B != 9, Q52H != 9)  
  
table(algS$Q11B)

##   
## 0 1 2 3   
## 996 98 37 12

table(algS$Q52H)

##   
## 0 1 2 3   
## 156 328 336 323

str(algS$Q11B)

## dbl+lbl [1:1143] 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## @ label : chr "Q11b. Have been physically attacked"  
## @ format.spss : chr "F2.0"  
## @ display\_width: int 7  
## @ labels : Named num [1:7] -1 0 1 2 3 9 98  
## ..- attr(\*, "names")= chr [1:7] "Missing" "No" "Yes, once" "Yes, twice" ...

str(algS$Q52H)

## dbl+lbl [1:1143] 2, 1, 2, 3, 2, 2, 3, 1, 1, 3, 1, 2, 1, 1, 3, 2, 2, 1, 3, ...  
## @ label : chr "Q52h. Trust police"  
## @ format.spss : chr "F2.0"  
## @ display\_width: int 7  
## @ labels : Named num [1:7] -1 0 1 2 3 9 98  
## ..- attr(\*, "names")= chr [1:7] "Missing" "Not at all" "Just a little" "Somewhat" ...

# convert to labelled factors  
algS <- algS %>%   
 mutate(Q11Bf = as\_factor(Q11B),  
 Q52Hf = as\_factor(Q52H))  
  
# univariate description  
table(algS$Q52Hf)

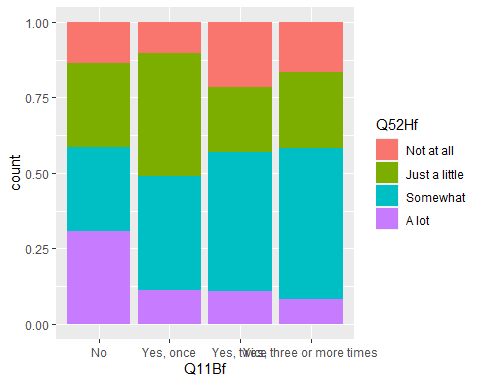
##   
## Missing Not at all   
## 0 156   
## Just a little Somewhat   
## 328 336   
## A lot Don't know/Haven't heard enough   
## 323 0   
## Refused   
## 0

table(algS$Q11Bf)

##   
## Missing No Yes, once   
## 0 996 98   
## Yes, twice Yes, three or more times Don’t know   
## 37 12 0   
## Refused   
## 0

A suitable way to visualise the relationship between two categorical variables is to use stacked or clustered barplot. In figure N, we use a stacked barplot as implemented in the ggplot2 package (Wickham, 2016). This visualisation suggests that the distribution of assault victimisation is quite similar across the values of trust in police with the exception of respondents who trust police ‘a lot’, among whom the share of victimised respondents is substantially smaller.

# stacked bar chart  
ggplot(algS, aes(x = Q11Bf, fill = Q52Hf)) +  
 geom\_bar(position = "fill")

 Visualisation and numeric description of variables can be informative, but it is not sufficient for testing the hypothesis of whether there is an association between these two variables. This is usually done with the use of contingency tables also known as crosstables. The gmodels package in R (Warnes et al., 2019) allows users to create highly informative and visually clear crosstables like those known from other statistical packages such as SPSS or Stata.

# crosstable with residuals, row percent + a full table  
CrossTable(algS$Q52Hf, algS$Q11Bf, format = c("SPSS"), chisq = TRUE,   
 prop.c = FALSE, prop.chisq = FALSE, prop.t = FALSE, asresid = TRUE)

## Warning in chisq.test(t, correct = FALSE, ...): Chi-squared approximation may be  
## incorrect

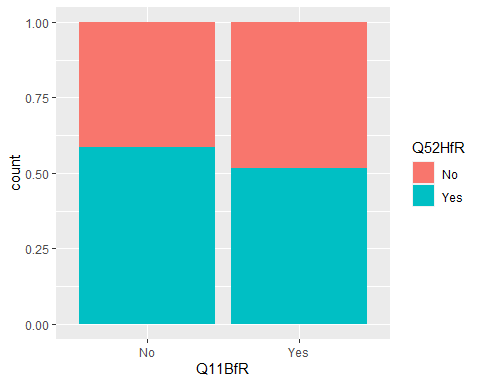
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## | Adj Std Resid |  
## |-------------------------|  
##   
## Total Observations in Table: 1143   
##   
## | algS$Q11Bf   
## algS$Q52Hf | No | Yes, once | Yes, twice | Yes, three or more times | Row Total |   
## --------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|  
## Not at all | 136 | 10 | 8 | 2 | 156 |   
## | 87.179% | 6.410% | 5.128% | 1.282% | 13.648% |   
## | 0.016 | -1.039 | 1.436 | 0.306 | |   
## --------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|  
## Just a little | 277 | 40 | 8 | 3 | 328 |   
## | 84.451% | 12.195% | 2.439% | 0.915% | 28.696% |   
## | -1.722 | 2.774 | -0.967 | -0.285 | |   
## --------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|  
## Somewhat | 276 | 37 | 17 | 6 | 336 |   
## | 82.143% | 11.012% | 5.060% | 1.786% | 29.396% |   
## | -3.256 | 1.900 | 2.246 | 1.575 | |   
## --------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|  
## A lot | 307 | 11 | 4 | 1 | 323 |   
## | 95.046% | 3.406% | 1.238% | 0.310% | 28.259% |   
## | 5.012 | -3.917 | -2.396 | -1.541 | |   
## --------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|  
## Column Total | 996 | 98 | 37 | 12 | 1143 |   
## --------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 35.17103 d.f. = 9 p = 5.558301e-05   
##   
##   
##   
## Minimum expected frequency: 1.637795   
## Cells with Expected Frequency < 5: 4 of 16 (25%)

The crosstabulation (table O) displays the categories of trust in police in rows and assault victimisation in columns. Each cell contains the number of observations, its row percentage, and its adjusted standardised residual respectively. Below the table, we can also find the information related to Pearson’s chi-squared test of independence. While the p-value of this test would suggest that we can reject the null hypothesis of independence between these two variables, the information that follows indicates that the crosstable is not sufficiently filled and therefore the test results are not valid. This can be solved by transforming the variables so that there are less categories in each and thus the resulting table could be better filled. This is done using the following code in the tidyverse package which we immediately follows with a stacked barchart of the transformed variables.

# collapsing the categories using mutate  
algS <- algS %>%  
 mutate(Q11BfR = recode\_factor(Q11Bf,   
 "No" = "No",   
 "Yes, once" = "Yes",  
 "Yes, twice" = "Yes",  
 "Yes, three or more times" = "Yes"))  
  
algS <- algS %>%  
 mutate(Q52HfR = recode\_factor(Q52Hf,   
 "Not at all" = "No",   
 "Just a little" = "No",  
 "Somewhat" = "Yes",  
 "A lot" = "Yes"))

The stacked bar chart (figure O) suggests that there is no discernible association between the transformed trust in police and assault victimisation. This can be corroborated by a new crosstable below. Compared to table O, table P also displays the column percentages and while it indeed displays a sufficiently filled table, the null hypothesis of Pearson’s chi-squared test of independence cannot be rejected at any of the standard significance levels. We can therefore conclude that the two variables are independent of each other. Finally, we also used the lsr package (Navarro, 2015) to calculate one of the most frequently used coefficients of association for categorical variables – Cramér’s V. Using a coefficient of association allows us to complement the statistical test with a measure of effect size. The resulting value of 0.04 further strengthens the evidence that the relationship between trust in police and being a victim of assault is rather weak.

# stacked bar chart  
ggplot(algS, aes(x = Q11BfR, fill = Q52HfR)) +  
 geom\_bar(position = "fill")



# crosstable with residuals, row percent + a full table  
CrossTable(algS$Q52HfR, algS$Q11BfR, format = c("SPSS"), chisq = TRUE,   
 prop.chisq = FALSE, prop.t = FALSE, asresid = TRUE)

##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## | Column Percent |  
## | Adj Std Resid |  
## |-------------------------|  
##   
## Total Observations in Table: 1143   
##   
## | algS$Q11BfR   
## algS$Q52HfR | No | Yes | Row Total |   
## -------------|-----------|-----------|-----------|  
## No | 413 | 71 | 484 |   
## | 85.331% | 14.669% | 42.345% |   
## | 41.466% | 48.299% | |   
## | -1.565 | 1.565 | |   
## -------------|-----------|-----------|-----------|  
## Yes | 583 | 76 | 659 |   
## | 88.467% | 11.533% | 57.655% |   
## | 58.534% | 51.701% | |   
## | 1.565 | -1.565 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 996 | 147 | 1143 |   
## | 87.139% | 12.861% | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 2.450039 d.f. = 1 p = 0.117522   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 2.178133 d.f. = 1 p = 0.1399841   
##   
##   
## Minimum expected frequency: 62.24672

# coefficient of association  
tab <- table(algS$Q52HfR, algS$Q11BfR)  
# necessary to subselect only the relevant columns  
tabR <- tab[1:2, 1:2]   
cramersV(tabR)

## [1] 0.04365351