In [2]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
from scipy.stats import chi2_contingency
```

Chi-Squared - Elliot Linsey

This notebook creates cross tabulated tables of crime counts by year, then calculates the chi-squared statistic, p-values and expected frequencies.

In [3]:
```
df = pd.read_csv(r"D:\EOY Datasets\Full Datasets\Everything_cleaned.csv",index_col=0)
```

```
C:\Users\ellio\anaconda3\lib\site-packages\numpy\lib\arraysetops.py:583: FutureWarning: elementwise comparison failed; returning scalar instead, but in the future will perform elementwise comparison
  mask |= (ar1 == a)
```

In [4]:
```
df2 = pd.read_csv(r"D:\EOY Datasets\Full Datasets\Everything_NonCovid.csv",index_col=0)
```

In [5]:
```
df.head()
```

Out[5]:

	Crime_ID	Month	Reported_by	Crime_type	Outcome_type	Year	Level_of_Lockdown	Region
0	e9a50727f2189e0c50f704e3661bc1a8ae3a39aece4866…	2019-01-01	Avon and Somerset	Violence and sexual offences	No suspect identified	2019	No Lockdown	South West
1	fb06f54e9d633a961109ef74171e4beb40b2fe44e57a10…	2019-01-01	Avon and Somerset	Violence and sexual offences	Unable to prosecute suspect	2019	No Lockdown	South West
2	4655ba031d5c8d00b247577a22472fbb8f9130d98b7d95…	2019-01-01	Avon and Somerset	Violence and sexual offences	No suspect identified	2019	No Lockdown	South West
3	2830f28a4a93138717182d1cc51b6d77cfaf3c27398934…	2019-01-01	Avon and Somerset	Theft offences	No suspect identified	2019	No Lockdown	South West
4	ca01d599d45f6e1d1bc5b05ddc6a1b559870fc831baa8e…	2019-01-01	Avon and Somerset	Theft offences	No suspect identified	2019	No Lockdown	South West

In [6]:
```
import dataframe_image as dfi
```

In [24]:
```
tab = pd.crosstab(df.Crime_type,df.Year)
```

In [31]:
```
tab = tab[[2019,2020]]
tab
```

Out[31]:

Year	2019	2020
Crime_type		
Criminal damage and arson	441936	427073
Drug offences	149840	169180
Public order offences	340925	391677
Theft offences	1094973	965238
Violence and sexual offences	1516608	1624924

First value is the very high chi statistic, then the p-value, then the degrees of freedom, then the expected frequency counts.

In [32]:
```
chi2_contingency(tab)
```

Out[32]:
```
(16686.684072423726,
 0.0,
 4,
 array([[ 432441.90161005,  436567.09838995],
        [ 158752.80400046,  160267.19599954],
        [ 364562.16449234,  368039.83550766],
        [1025215.57608488, 1034995.42391512],
        [1563309.55381225, 1578222.44618775]]))
```

In [33]:
```
tab_expected = pd.DataFrame(np.round(chi2_contingency(tab)[3]),columns=tab.columns,index=tab.index)
tab_expected
```

Out[33]:

Year	2019	2020
Crime_type		
Criminal damage and arson	432442.0	436567.0
Drug offences	158753.0	160267.0
Public order offences	364562.0	368040.0
Theft offences	1025216.0	1034995.0
Violence and sexual offences	1563310.0	1578222.0

In [9]:
```
df2.head()
```

Out[9]:

	Crime_ID	Month	Reported_by	Crime_type	Outcome_type	Region	Year
0	9500972e2e3c87dc23b65885dd61085ee8a39329c03792…	2017-01	Avon and Somerset Constabulary	Theft offences	Other Outcome	South West	2017
1	4c3b7072a3bfcf441b1036e67fd642cf1e42fdccf16af1…	2017-01	Avon and Somerset Constabulary	Criminal damage and arson	No suspect identified	South West	2017
2	8e6ae250f68623b82068cd3c79b64b05a559f6d9c9393c…	2017-01	Avon and Somerset Constabulary	Criminal damage and arson	No suspect identified	South West	2017
3	0f5c81322cd60d96ef8877d94eb37e98cad395d1773efc…	2017-01	Avon and Somerset Constabulary	Violence and sexual offences	No suspect identified	South West	2017
4	b46b395f390b0dc9112686c30d0939f0ded0de42877df7…	2017-01	Avon and Somerset Constabulary	Violence and sexual offences	No suspect identified	South West	2017

In [35]:
```
tab2 = pd.crosstab(df2.Crime_type,df2.Year)
```

In [36]:
```
tab2 = tab2[[2017,2018]]
tab2
```

Out[36]:

Year	2017	2018
Crime_type		
Criminal damage and arson	515876	483533
Drug offences	119272	124341
Public order offences	302082	355207
Theft offences	1349965	1267009
Violence and sexual offences	1260729	1469880

In [11]:
```
# tab2 = tab2/10000
# tab2
```

Still significantly different even without covid.

In [37]:
```
chi2_contingency(tab2)
```

Out[37]:
```
(20915.10952633738,
 0.0,
 4,
 array([[ 489221.72108422,  510187.27891578],
        [ 119251.24862643,  124361.75137357],
        [ 321750.21020396,  335538.78979604],
        [1281037.61754463, 1335936.38245537],
        [1336663.20254077, 1393945.79745923]]))
```

In [38]:
```
tab2_expected = pd.DataFrame(np.round(chi2_contingency(tab2)[3]),columns=tab2.columns,index=tab2.index)
tab2_expected
```

Out[38]:

Year	2017	2018
Crime_type		
Criminal damage and arson	489222.0	510187.0
Drug offences	119251.0	124362.0
Public order offences	321750.0	335539.0
Theft offences	1281038.0	1335936.0
Violence and sexual offences	1336663.0	1393946.0

In [39]:
```
dfi.export(tab,"Counts_covid.png")
dfi.export(tab_expected,"Counts_covid_expected.png")
dfi.export(tab2,"Counts_noncovid.png")
dfi.export(tab2_expected,"Counts_noncovid_expected.png")
```

In [13]:
```
tab3 = pd.crosstab(df2.Outcome_type,df2.Year)
tab3
```

Out[13]:

Year	2017	2018	2019
Outcome_type			
No suspect identified	1865956	1783831	1303229
Other Outcome	347690	348220	282513
Suspect charged	364989	323366	265690
Unable to prosecute suspect	969289	1244553	1240866

In [14]:
```
chi2_contingency(tab3)
```

Out[14]:
```
(126395.93628964067,
 0.0,
 6,
 array([[1699477.56664325, 1772308.54219341, 1481229.89116334],
        [ 335716.24626042,  350103.33921362,  292603.41452596],
        [ 327351.67321651,  341380.30305917,  285313.02372432],
        [1185378.51387982, 1236177.81553379, 1033151.67058639]]))
```