Contingency tables

NHH TECH3

Contingency table

Is a black driver more likely to be searched when they are pulled over by the police compared to a white driver?

Searched by police	Black driver	White driver
FALSE	36 244	239 241
TRUE	1 219	3 108

Searched by police	Black driver	White driver
FALSE	0.130	0.855
TRUE	0.00436	0.0111

Data source: https://openpolicing.stanford.edu/

NHH TECH3

Contingency table

Is a black driver more likely to be searched when they are pulled over by the police compared to a white driver?

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(observed_{ij} - expected_{ij}\right)^2}{expected_{ij}}$$

S: Driver is searched by the police

B: Driver is black

S: Driver is searched by the police

B: Driver is black

H_0: S and B are independent

H_A: S and B are dependent

H_0: S and B are independent

$$P(S \cap B) = P(S) \cdot P(B)$$

Searched by police	Black driver	White driver
FALSE	0.130	0.855
TRUE	0.00436	0.0111

NHH TECH3

H_0: S and B are independent

$$P(S \cap B) = P(S) \cdot P(B)$$

Searched by police	Black driver	White driver
FALSE	0.130	0.855
TRUE	0.00436	0.0111
P(B)	0.13436	0.8661

NHH TECH3

H_0: S and B are independent

$$P(S \cap B) = P(S) \cdot P(B)$$

Searched by police	Black driver	White driver	P(S)
FALSE	0.130	0.855	0.985
TRUE	0.00436	0.0111	0.015
P(B)	0.134	0.866	1.00

H_0: S and B are independent

$$P(S^c \cap B) = P(S^c) \cdot P(B)$$

Searched by police	Black driver	White driver	P(S)
FALSE	$0.985 \cdot 0.134 =$ 0.132		0.985
TRUE			0.015
P(B)	0.134	0.866	1.00

H_0: S and B are independent

$$P(S^c \cap B^c) = P(S^c) \cdot P(B^c)$$

Searched by police	Black driver	White driver	P(S)
FALSE	$0.985 \cdot 0.134 =$ 0.132	$0.985 \cdot 0.866 =$ 0.853	0.985
TRUE			0.015
P(B)	0.134	0.866	1.00

NHH
TECH3

H_0: S and B are independent

$$P(S \cap B) = P(S) \cdot P(B)$$

Searched by police	Black driver	White driver	P(S)
FALSE	$0.985 \cdot 0.134 = 0.132$	$0.985 \cdot 0.866 = 0.853$	0.985
TRUE	$0.015 \cdot 0.134 = 0.00201$		0.015
P(B)	0.134	0.866	1.00

NHH TECH3

H_0: S and B are independent

$$P(S \cap B^c) = P(S) \cdot P(B^c)$$

Searched by police	Black driver	White driver	P(S)
FALSE	$0.985 \cdot 0.134 = 0.132$	$0.985 \cdot 0.866 = 0.853$	0.985
TRUE	$0.015 \cdot 0.134 = 0.00201$	$0.015 \cdot 0.866 = 0.0130$	0.015
P(B)	0.134	0.866	1.00

NHH TECH3

H_0: S and B are independent

Searched by police	Black driver	White driver
FALSE	0.132	0.853
TRUE	0.00201	0.130

NHH
TECH3

H_0: S and B are independent

Searched by police	Black driver	White driver
FALSE	$0.132 \cdot n$	$0.853 \cdot n$
TRUE	$0.00201 \cdot n$	$0.0130 \cdot n$

$$n = 279\,812$$

NHH TECH3

Expected frequencies under the null hypothesis:

Searched by police	Black driver	White driver
FALSE	36 935	238 680
TRUE	562	3 635

H_0: S and B are independent

NHH TECH3

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(observed_{ij} - expected_{ij})^2}{expected_{ij}} \sim \chi^2_{(r-1)\times(c-1)}$$

Observed

Searched by police	Black driver	White driver
FALSE	36 244	239 241
TRUE	1 219	3 108

Expected under H_0

Searched by police	Black driver	White driver
FALSE	36 935	238 680
TRUE	562	3 635

NHH TECH3

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(observed_{ij} - expected_{ij})^2}{expected_{ij}} = 858.7$$

The distribution is

$$\chi^2_{(r-1)\times(c-1)} = \chi^2_{(2-1)\times(2-1)} = \chi^2_1$$

P-value:

```
from scipy import stats
print(1-stats.chi2.cdf(858.7,df=1))
0.0
```

NHH
TECH3

S: Driver is searched by the police

B: Driver is black

~~H_0: S and B are independent~~

H_A: S and B are dependent

NHH

TECH3

Sondre Hølleland
Geir Drage Berentsen