Chi-Square Test

31 May 2023 10:49

χ^2 - test for Independence of Attribute

H_0: The row attribute & column attribute are independent of each other

H_1: The row & col^n attributes dependent on each other

column attribute

row attribute

AirBags	Driver & Passenger	Driver only	None
Type			
Compact	2	9	5
Large	4	7	0
Midsize	7	11	4
Small	0	5	16
Sporty	3	8	3
Van	0	3	6

H_0: AirBags & Type are indep.

H_1: ——"—— are dep.

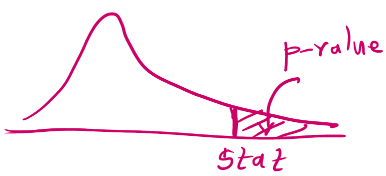
AirBags	Driver & Passenger	Driver only	None	All
Type				
Compact	O_{11} 2	O_{12} 9	O_{13} 5	16 $= R_1$
Large	O_{21} 4	O_{22} 7	O_{23} 0	11 $= R_2$
Midsize	O_{31} 7	O_{32} 11	O_{33} 4	22 $= R_3$
Small	O_{41} 0	O_{42} 5	O_{43} 16	21 $= R_4$
Sporty	O_{51} 3	O_{52} 8	O_{53} 3	14 $= R_5$
Van	O_{61} 0	O_{62} 3	O_{63} 6	9 $= R_6$
All	16	43	34	93 $= n$
	C_1	C_2	C_3	

O_{ij} : Observed Frequency of ith row & j^{th} col^n cell

E_{ij} : Expected Frequency of ith row & j^{th} cell cell

$$E_{11} = \frac{R_1 C_1}{n} = \frac{16 \times 16}{93}, \quad E_{32} = \frac{43 \times 22}{93}, \quad E_{62} = \frac{43 \times 9}{93}$$

$$E_{ij} = \frac{R_i C_j}{n}$$



$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$r = 6, \; c = 3$

Degrees of freedom: df=(r-1)(c-1)

```
In [7]: test_statistic, p_value, df, expected_frequencies = chi2_contingency(ctab)
   ...: print("P-Value =", p_value)
P-Value = 0.000272287749055816
```

< 0.05

$\therefore$ We reject H_0 at 5% l.o.s.

∴ We reject H_0 at 5% l.o.s.

Conclusion: Attributes of AirBag & Type may be dependent.

```
df_bar = pd.melt(ctab.reset_index(),id_vars="Type")
sns.barplot(x="AirBags",
            y="value",
            hue="Type",
            data=df_bar)
plt.show()
```