Astik Jadhav

2018110019

Q2)

	male	female	Total
fiction	250(90)	200(360)	450
non-fiction	50(210)	1000(840)	1050
Total	300	1200	1500

$$e_{ij} = \frac{count(A = a_i) \times count(B = b_j)}{n}$$

$$e_{11} = \frac{300 \times 450}{1500} = 90$$

$$e_{21} = \frac{300 \times 1050}{1500} = 210$$

$$e_{22} = \frac{1200 \times 1050}{1500} = 840$$

$$e_{21} = \frac{1200 \times 450}{1500} = 360$$

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

(O_{ij}: observed, e_{ij}: estimated)

$$\chi^2 = \frac{(250-90)^2}{90} + \frac{(50-210)^2}{210} + \frac{(200-360)^2}{360} + \frac{(1000-840)^2}{840}$$

$$= 284.44 + 121.90 + 71.11 + 30.48 = 507.93$$

Astik Jadhav
20181100 19

DOF = $(2-1) \times (2-1) = 1$

For DOF X^2 value needed to reject the hypothesis at 0.001 significance is 10.828

since computed value is much higher than this we can safely reject the hypothesis that gender and preferred reading are independant and conclude that they are correlated.

Q1)

	Ontime	Late	Very Late	Cancelled
Prob.	14/20 = 0.7	2/20 = 0.1	3/20 = 0.15	1/20 = 0.05
Weekday	9/14 = 0.64	1/2 = 0.5	3/3 = 1	0/1 = 0
Winter	2/14 = 0.14	2/2 = 1	2/3 = 0.67	0/1 = 0
High	4/14 = 0.29	1/2 = 0.5	1/3 = 0.33	1/1 = 1
None	5/14 = 0.36	1/2 = 0.5	1/3 = 0.33	0/1 = 0

Case 1 Ontime = 6.548×10^{-3}
Late = 0.0125
Very Late = 0.0109
Cancelled. = 0

Late is the highest hence the accurate class is Late.