Assignment - 1

Siddhant Even
2019120024

Attributes taken :- Day, season, fog, rain

Categories -

ON Time	Late	very late	Cancelled
14/20	2/20	3/20	1/20

$$= P(y|x_1 \dots x_n) = \frac{P(x_1|y)[P(x_2|y) \dots P(x_n|y)](P(y))}{P(x_1)P(x_2)P(x_3) \dots P(x_n)}$$

	Attribute	ON Time	LATE	very Late	Cancelled
			class		
Day	weekday	9/14 = 0.64	1/2 = 0.5	3/3 = 1	0/1 = 0
	saturday	2/14 = 0.14	1/2 = 0.5	0/3 = 0	1/1 = 1
	sunday	1/14 = 0.07	0/2 = 0	0/3 = 0	0/1 = 0
	holiday	2/14 = 0.14	0/2 = 0	0/3 = 0	0/1 = 0
Season	spring	4/14 = 0.29	0/2 = 0	0/3 = 0	0/1 = 0
	summer	6/14 = 0.43	0/2 = 0	0/3 = 0	0/1 = 0
	Autumn	2/14 = 0.14	0/2 = 0	1/3 = 0.33	0/1 = 0
	winter	2/14 = 0.14	2/2 = 1	2/3 = 0.67	0/1 = 0
Fog	None	5/14 = 0.36	0/2 = 0	0/3 = 0	0/1 = 0
	High	4/14 = 0.29	1/2 = 0.5	1/3 = 0.33	1/1 = 1
	Normal	5/14 = 0.36	1/2 = 0.5	2/3 = 0.67	0/1 = 0
Rain	None	5/14 = 0.36	1/2 = 0.5	1/3 = 0.33	0/1 = 0
	slight	8/14 = 0.57	0/2 = 0	0/3 = 0	0/1 = 0
	heavy	1/14 = 0.07	1/2 = 0.5	2/3 = 0.67	1/1 = 1
	.	.	.	.	.
		14/20 = 0.7	2/20 = 0.1	3/20 = 0.15	1/20 = 0.05

$$x^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

2. H_0 = Reading preferred & gender are not correlated in the group.

H_1 = Both are correlated.

Computing the x^2,

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

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

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

For 2×2 = Degree of freedom = $(2-1)(2-1)$
$= 1$

For 1 degree of freedom, x^2 value needed to reject the hypothesis at 0.01 significance level is 10.828 (from table)

We reject the null hypothesis that gendere and preferred reading are independent (C: the calculated value is above it)

∴ We conclude that 2 attributes are correlated for the given group.