UE19CS203 – STATISTICS FOR DATA SCIENCE
Unit-4 - Hypothesis and Inference

QUESTION BANK – SOLVED

Chi-squared Test :

Exercises for section 6.10: [Text Book Exercise 6.10– Pg. No. [466 – 468]]

1. For the given table of observed values,
   a. Construct the corresponding table of expected values.
   b. If appropriate, perform the chi-square test for the null hypothesis that the row and column outcomes are independent. If not appropriate, explain why.

Observed Values	1	2	3
A	15	10	12
B	3	11	11
C	9	14	12

[Text Book Exercise – Section 6.10 – Q. No.8 – Pg. No. 467]

Solution:

The hypotheses of a chi-square test for independence are

H_0: The variables are independent.
H_1: The variables are dependent.

Determine the observed frequencies O.

The expected frequencies E are the product of the column and row total, divided by the sample size of n.

Chi-square test statistic is given by

$$x^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

a)

O	E	O-E	(O-E)²/E
25	28	-3	0.3214
4	4	0	0
11	8	3	1.125
3	7	-4	2.2857
3	1	2	4
4	2	2	2
42	35	7	1.4
3	5	-2	0.8
5	10	-5	2.5

b) All expected counts are not greater than 5, then it is not appropriate to perform the chi-square test for independence.

2. Because of printer failure, none of the observed values in the following table were printed, but some of the marginal totals were. Is it possible to construct the corresponding table of expected values from the information given? If so, construct it. If not, describe the additional information you would need.

| | Observed Values | | | |
	1	2	3	Total
A	—	—	—	25
B	—	—	—	—
C	—	—	—	40
D	—	—	—	75
Total	50	20	—	150

[Text Book Exercise – Section 6.10 – Q. No.10 – Pg. No. 467]

Solution:

The row total of row B is missing. We can determine this row total by decreasing the table total by the other row totals.

Row total B= 150-25-40-75=10

The column total of column 3 is missing. We can determine this column total by decreasing the table total by the other column totals.
Column total 3 =150-50-20=80

The expected value is the product of the row total and column total, divided by the table total of 150.

A1: $\dfrac{25 \times 50}{150} \approx 8.3333$

A2: $\dfrac{25 \times 20}{150} \approx 3.3333$

A3: $\dfrac{25 \times 80}{150} \approx 13.3333$

B1: $\dfrac{10 \times 50}{150} \approx 3.3333$

B2: $\dfrac{10 \times 20}{150} \approx 1.3333$

B3: $\dfrac{10 \times 80}{150} \approx 5.3333$

C1: $\dfrac{40 \times 50}{150} \approx 13.3333$

C2: $\dfrac{40 \times 20}{150} \approx 5.3333$

C3: $\dfrac{40 \times 80}{150} \approx 21.3333$

D1: $\dfrac{75 \times 50}{150} \approx 25$

D2: $\dfrac{75 \times 20}{150} \approx 10$

D3: $\dfrac{75 \times 80}{150} \approx 40$