# Chi-Square Assignment.

## PROBLEM I:-

**#1.** A poker-dealing machine is supposed to deal cards at random, asif from an infinite deck. In a test, you counted 1600 cards, and observed the following:-

    Spades    404
    Hearts    420
    Diamonds  400
    Clubs     376

Could it be that the suits are equally likely? or are these discrepencies too much to be random?

In ideal case, 1600 card will be distributed as,—

|          | Observed | Expected/Actual |
|----------|----------|-----------------|
| Spade    | 404      | 400             |
| hearts   | 420      | 400             |
| Diamonds | 400      | 400             |
| Clubs    | 376      | 400             |

Here, degrees of freedom, $df = 3$

$$x^2 = \sum \frac{(E - O)^2}{E}$$
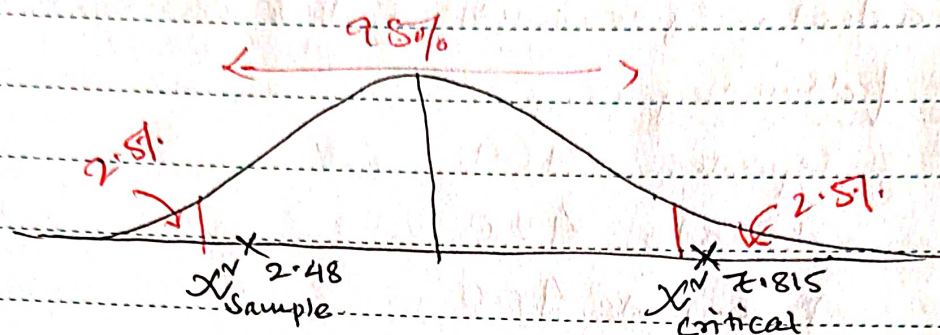
$$= \frac{(404-400)^2 + (420-400)^2 + (400-400)^2 + (376-400)^2}{400}$$

$$= \frac{(4)^2 + (20)^2 + 0 + (24)^2}{400}$$

$$= \frac{992}{400} = \boxed{2.48}$$

From, Chi-Square Critical Table,
$\check{X}$ at 5% Significance for 3 degrees of freedom
is ⇒ 7.815



Here, $\check{X}_{sample} < \check{X}_{critical}$

Hence, accept the NULL Hypothesis,
Which means, there is no major variation
or variation is not Significant.

※ In the titanic Dataset, do a crosstab for embarked
and survival rate. Using chi-square test, determine
whether both of them are dependent or independent
From, panda cross tab, you will get below
Observed frequencies —

| Survived | 0 | 1 | All |
| --- | --- | --- | --- |
| Embarked | | | |
| C | 75 | 93 | 168 |
| Q | 47 | 30 | 77 |
| S | 427 | 217 | 644 |
| All | 549 | 340 | 889 |

So, Now, what is expected frequencies,—

for, 75 $= \dfrac{168 \times 549}{889} = 103.74$

47 $= \dfrac{77 \times 549}{889} = 47.55$

427 $= \dfrac{644 \times 549}{889} = 397.70$

93 $= \dfrac{168 \times 340}{889} = 64.25$

30 $= \dfrac{77 \times 340}{889} = 29.44$

217 $= \dfrac{644 \times 340}{889} = 246.29$

$$x^2 = \sum \dfrac{(E-O)^2}{E}$$

$$= \dfrac{(103.74-75)^2 + (47.55-47)^2 + (397.70-427)^2}{\phantom{889}}$$
$$\dfrac{(64.25-93)^2 + (29.44-30)^2 + (246.29-217)^2}{889}$$

$$= \dfrac{(28.74)^2 + (0.55)^2 + (-29.3)^2 + (-28.75)^2 + (-0.56)^2 + (29.29)^2}{889}$$

$(6-1)(4-1)$

$$= \dfrac{3369.56}{889} = 3.79$$

Here, df = 2,

∅ $x^2_{critical} = 5.99$
(from $x^2$ critical Table)

$x^2_{sample} = 3.79$ & $x^2_{critical} = 5.99$

Hence, it's INDEPENDENT?

**XX.** Same as before, but this time jokers are included, and you counted (1662) cards, with these results:-

Spades 404
Hearts 420
Diamond 400
Clubs 356
Jokers 82

a) How many jokers would you expect out of (1662) random cards? How many of each suit?

b) Is it possible that the cards are really random? Or are the discrepancies too large?

| | Observed (O) | Expected (E) | (E-O) | $(E-O)^2$ | $(E-O)^2/E$ |
|---|---|---|---|---|---|
| Spades | 404 | 400 | -4 | 16 | 0.04 |
| Hearts | 420 | 400 | -20 | 400 | 1 |
| Diamond | 400 | 400 | 0 | 0 | 0 |
| Clubs | 356 | 400 | 44 | 1936 | 4.84 |
| Jokers | 82 | 62 | -20 | 40 | 6.45 |

a) Expected ⇒ 62 Jokers                    Total 12.33
&              ⇒ 400 each suit

b)

$$X^2 = \sum \frac{(E-O)^2}{E} = 12.33$$

$$= \frac{(400-404)^2 + (400-420)^2 + (400-400)^2 + (400-356)^2 + (62-82)^2}{}$$

$X^2$ critical value at $\alpha = 0.05$, df = 4
                    is ⇒ 9.488

Since, $X^2_{sample} > X^2_{cr}$, Hence Reject Null Hypothesis. Cards are not random, there are too many Jokers and not enough clubs.

2.5%                    2.5%



$X^2_{cr} = 9.488$
$X^2_{sample} = 12.33$
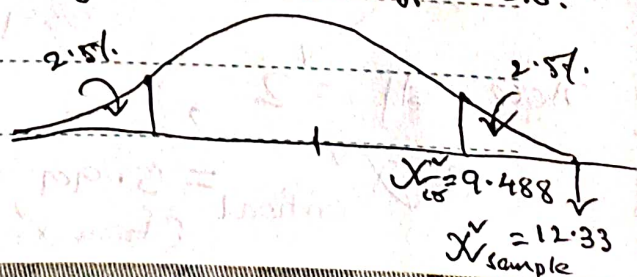
**3.** A genetic engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio ④ stripes only : ③ spots only : ⑨ both stripes and spots. When the cross was performed and she counted the individuals she found ㊿ with stripes only, ㉑ with spots only and ㊺ with both. According to the chi-square test, did she get the predicted outcome?

| ratio | Observed | Expected | E-0 | $(E-0)^2$ | $(E-0)^2/E$ |
|-------|----------|----------|-----|-----------|-------------|
| 4 stripes | 50 | 44 | -6 | 36 | 0.82 |
| 3 spots | 41 | 33 | -8 | 64 | 1.94 |
| 9 stripes/spots | 85 | 99 | 14 | 196 | 1.98 |
| Total 16 | 176 | 176 | 0 | | 4.74 |

Expected # of stripes = 4/16 × 176 = 44
Expected # of spots = 3/16 × 176 = 33
Expected # of stripes/spots = 9/16 × 176 = 99

df = 3-1 = 2

$X^2$ critical value, for $\alpha = 0.05$, df 2
        is → 5.991

Since, $X^2$ sample value $< X^2$ critical value
Hence, accept the null hypothesis.
    She got her predicted outcome.

**4.** In the garden pea, yellow cotyledon color is dominant to green and inflated pod shape is dominant to the constricted form. Considering both of these traits jointly in self-fertilized dihybrid the progeny appeared in the following numbers:—

193 green inflated, 184 yellow constricted, 556 yellow inflated, 61 green constricted

Do these genes assort independently? Support your answer using Chi-square analysis.

Note:— Genes assort independently if they follow the 9:3:3:1 rule (or the 16 square Punnett Square) resulting from a dihybrid cross.