EXPERIMENT NO. 11

A. CHI-SQUARE ANALYSIS

A chi square ($\chi2$) statistic is used to investigate whether distributions of categorical variables differ from one another.
The Chi Square statistic compares the tallies or counts of categorical responses between two (or more) independent groups. (note: Chi square tests can only be used on actual numbers and not on percentages, proportions, means, etc.). $\chi2$ is used to test whether the number of individuals in different categories fit a **null hypothesis** (an expectation of some sort). The value of $\chi2$ can be checked in the table.

Null hypothesis

-The null hypothesis is a term that statisticians often use to indicate the statistical hypothesis tested.
-The purpose of most statistical tests is to determine if the obtained results provide a reason to reject the hypothesis and that they are merely a product of chance factors.

Significance level

In hypothesis testing, the significance level is the criterion used for rejecting the null hypothesis. The significance level is used in hypothesis testing as follows: First, the difference between the results of the experiment and the null hypothesis is determined. Then, assuming that the null hypothesis is true, the probability of a difference that is large or larger is computed. Finally, this probability is compared to the significance level. If the probability is less than or equal to the significance level, then the null hypothesis is rejected and the outcome is said to be statistically significant. Traditionally, experimenters have used either the .05 level (sometimes called the 5% level) or the .01 level (1% level), although the choice of levels is largely subjective. The lower the significance level, the more the data must diverge from the null hypothesis to be significant. Therefore, the .01 level is more conservative than the .05 level. The Greek letter alpha is sometimes used to indicate the significance level.

Degree of Freedom

That measure of **variability** which merely expresses the number of options available within a **variable** or space. In a **system** with N states the degree of freedom is N. In statistics of the N cells of a table of probabilities only N-1 can be arbitrarily filled, the last being determined by the requirement that probabilities must add to 1, **probability**, hence the degree or freedom is N-1.

Chi Square Goodness of Fit (One Sample Test)

This test allows us to compare a collection of categorical data with some theoretical expected distribution. This test is often used in genetics to compare the results of a cross with the theoretical distribution based on genetic theory. Suppose you performed a simple monohybrid cross between two individuals that were heterozygous for the trait of interest.

Aa x Aa

If

Oi (i=1, 2, 3....n) is a set of observed (experimental) frequencies

Ei (i=1, 2, 3,....n) is the corresponding set of expected frequencies, then chi-square is defined by

$$\chi 2 = \sum_{i=1}^{n} \frac{(Oi-Ei)^2}{Ei}$$

Question:- In pea-breeding experiments, Mendel got the following ratios:-

Round and yellow = 315
Wrinkled and yellow= 101
Round and green = 108
Wrinkled and green = 32
Total = 556
Verify 9:3:3:1
($\chi 2$ for 3 df at 5% level = 7.815)

Answer:-

Expected frequencies

$\frac{9 X 556}{16}$, $\frac{3 X 556}{16}$, $\frac{3 X 556}{16}$, $\frac{1 X 556}{16}$

i.e., 313, 104, 104, 35.

Ratios	9	3	3	1
Oi	315	101	108	32
Ei	313	104	104	104

Therefore,

$$\chi 2 = \frac{(315-313)^2}{313} + \frac{(101-104)^2}{104} + \frac{(108-104)^2}{104} + \frac{(32-35)^2}{35}$$

$= 0.013 + 0.09 + 0.15 + 0.26 = 0.513$

Degree of freedom = no. of observations-1

$= 4 - 1 = 3$

($\chi 2$ for 3 df at 5% level = 7.815) [from table]

Since, the calculated value of $\chi 2$ is less than the table value, the hypothesis may be accepted. Hence, there is correspondence between theory and experiment.

2 x 2 Contingency Table

There are several types of chi square tests depending on the way the data was collected and the hypothesis being tested. We'll begin with the simplest case: a 2 x 2 contingency table. If we set the 2 x 2 table to the general notation shown below in Table 1, using the letters a, b, c, and d to denote the contents of the cells, then we would have the following table:

General notation for a 2 x 2 contingency table.

Variable	Data type 1	Data type 2	Totals
Category 1	a	b	a + b
Category 2	c	d	c + d
Total	a + c	b + d	a + b + c + d = N

For a 2 x 2 contingency table the Chi Square statistic is calculated by the formula:

- $$\chi 2 = \frac{(ad-bc)2(a+b+c+d)}{(a+b)(c+d)(b+d)(a+c)}$$

Note: notice that the four components of the denominator are the four totals from the table columns and rows.

Suppose you conducted a drug trial on a group of animals and you hypothesized that the animals receiving the drug would survive better than those that did not receive the drug. You conduct the study and collect the following data:

	Dead	Alive	Total
Treated	36	14	50
Not treated	30	25	55
Total	66	39	105

Applying the formula above we get:

$\chi 2 = 105[(36)(25) - (14)(30)]^2 / (50)(55)(39)(66) = 3.418$

Before we can proceed we need to know how many degrees of freedom we have. When a comparison is made between one sample and another, a simple rule is that the degrees of freedom equal (number of columns minus one) x (number of rows minus one) not counting the totals for rows or columns. For our data this gives (2-1) x (2-1) = 1.

We now have our chi square statistic ($\chi 2 = 3.418$), our predetermined alpha level of significance (0.05), and our degrees of freedom (df = 1). Entering the Chi square distribution table with 1 degree of freedom and reading along the row we find our value of $\chi 2$ (3.418) lies between 2.706 and 3.841. The corresponding probability is $0.10<P<0.05$. This is below the conventionally accepted significance level of 0.05 or 5%, so the null hypothesis that the two distributions are the same is verified. In other words, when the computed $\chi 2$ statistic exceeds the critical value in the table for a 0.05 probability level, then we can reject the null hypothesis of equal distributions. Since our $\chi 2$ statistic (3.418) did not exceed the critical value for 0.05 probability level (3.841) we can accept the null hypothesis that the survival of the animals is independent of drug treatment (i.e. the drug had no effect on survival).

Chi Square distribution table.

HOW TO READ A TABLE?

$X^2=3.418$

Probability level (α)

Df of our example ⇨

Df	0.5	0.1	0.05	0.02	0.01	0.001
1.	0.455	2.706	3.841	5.412	6.635	10.827
2.	1.386	4.605	5.991	7.824	9.210	13.815
3.	2.366	6.251	7.815	9.837	11.345	16.268
4.	3.357	7.779	9.488	11.668	13.277	18.465
5.	4.351	9.236	11.070	13.388	15.086	20.517

EXERCISE.....

1. The theory predicts the proportion of beans in four groups A, B, C and D should be 9:3:3:1 in an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287 and 118. Does the experiment result support the theory?
   ($\chi 2$ for 3 df at 5% level = 7.815).