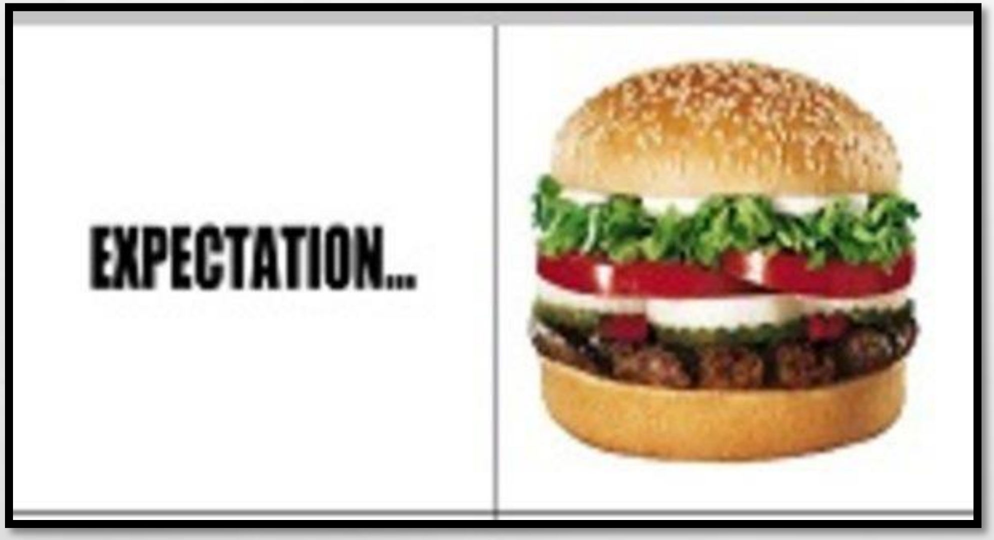
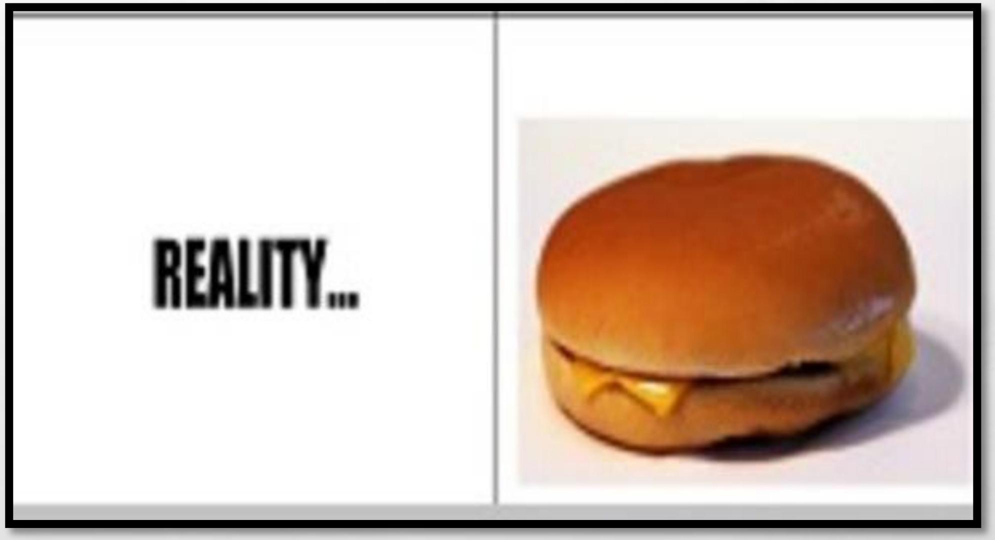

Expectation
VS
Reality


EXPECTATION...


REALITY...


NEW YEARS
RESOLUTION
expectation
reality


EXPECTATION

REALITY


You don't go on Facebook for a week:
Expectations:
facebook
25
35
95
Search
Reality:
facebook
1
Search


Expectation

Reality


REALITY
VS.
EXPECTATION

tests the significant difference between expected value and observed value

EXAMINING DIFFERENCE: CHI-SQUARED

χ^2

INTRODUCTION

- The chi-square test was first used by

KARL PEARSON in 1900.

Karl Pearson

INTRODUCTION

"X" in greek represents 'chi' , square of this is **<u>'chi square' or ' chi-squared test'</u>** .

What is Chi Square?

- Chi square test is the test which tests the significant difference between expected value and observed value which can be used for categorical data.

- *used to test whether a significant difference exists between the* ***OBSERVED NUMBER OF SAMPLES*** *and the* ***EXPECTED NUMBER OF RESPONSES.***

What is Chi Square?

Chi-squared is used to examine differences ***between what you actually find in your study and what you expected to find.***

The chi-square test uses frequency data to generate a statistical result

WHEN TO USE CHI-SQUARED?

Look at the list of questions below. If the answer is **YES** to each question, a chi-squared test is appropriate:

- ✓ ***Are you trying to see if there is a difference between what you have found and what would be found in a random pattern?***
- ✓ ***Is the data gathered organised into a set of categories?***
- ✓ ***In each category, is the data displayed as frequencies (not percentages)?***
- ✓ ***Does the total amount of data collected (observed data) add up to more than 20?***
- ✓ ***Does the expected data for each category exceed four?***

The chi-square is considered a unique test due to its 3 functions:

- ✓ TEST OF GOODNESS-OF-FIT
- ✓ TEST OF HOMOGENEITY
- ✓ TEST OF INDEPENDENCE

THE CHI-SQUARE TEST FOR GOODNESS-OF-FIT

GOODNESS OF FIT refers to how close the observed data are to those predicted from a hypothesis

Note:

- It evaluates to what extent the data and the hypothesis have a good fit

CHI-SQUARE GOODNESS of fit test is applied when you have ONE CATEGORICAL VARIABLE from a single population. It is used to determine whether sample data are consistent with a hypothesized distribution.

Chi-Square Test for Goodness of Fit

1. State **null hypothesized proportions** for each category. Alternative is that at least one of the proportions is different than specified in the null.
2. **Calculate the expected counts** for each cell .
3. **Calculate the χ^2 statistic**
4. **Compute the p-value**
5. Interpret the p-value in context.

WHEN DO WE USE THE TEST OF GOODNESS-OF-FIT?

- **When ratio is employed**

HOW DO WE USED THE TEST OF GOODNESS-OF-FIT?

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

- **Where**

  χ2-= the Chi-Square test

  O = observed frequencies / data in each category

  E = expected frequencies / data in each category based on experimenter's hypothesis

  ∑ = Sum of the calculations for each category

EXAMPLE 1

- A certain machine is supposed to mix peanuts, hazelnuts, cashews and pecans in the ratio of 4:3:2:1. A can containing 500 of these mixed nut was found to have 270 peanuts, 110 hazelnuts, 80 cashew and 40 pecans. At .05 level of significance, test the hypothesis that the machine is mixing the nuts at the ratio of 4:3:2:1.

SOLVING BY THE STEPWISE METHOD

I. **PROBLEM:** IS THE MACHINE MIXING THE NUTS AT THE RATIO OF 4:3:2:1?

The null hypothesis in a chi-square goodness-of-fit test should state a claim about the distribution of a single categorical variable in the population of interest.

HO: THE MACHINE IS MIXING THE NUTS AT THE RATIO OF 4:3:2:1.

The alternative hypothesis in a chi-square goodness-of-fit test is that the categorical variable does *not* have the specified distribution.

RATIO OF 4:3:2:1.

the null hypothesis (H_0) specifies the proportion of observations at each level of the categorical variable. The alternative hypothesis (H_a) is that *at least* one of the specified proportions is not true.

= 4-1

= 3

$\chi^2_{.05}$ = 7.815 *tabular value*

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Nuts			
Peanuts			
Hazelnut			
Cashew			
Pecans			
Total			

HOW TO COMPUTE FOR THE EXPECTED VALUE:

1. Add all the ratio
2. Divide the total ratio to total observed and multiply the ratio of each item

10 / 500 = 50

For expected:

50 x 4 = 200

50 x 3 = 150

50 x 2 = 100

50 x 1 = 50

SOLUTION

Nuts	Ratio	Observed	Expected
Peanuts	:4	270	200
Hazelnut	:3	110	150
Cashew	:2	80	100
Pecans	:1	40	50
Total	**:10**	**500**	**500**

$\chi 2=\sum(O-E)2/E$

$$\chi 2=\frac{(270-200)^2}{200}+\frac{(110-150)^2}{150}+\frac{(80-100)^2}{100}+\frac{(40-50)^2}{50}$$

$\chi 2 = 24.5 + 10.66 + 4 + 2$

$\chi 2 = 41.167$

DECISION RULE

- ***If the chi-square computed value is greater than the chi-square tabular value, disconfirm the Ho***

CONCLUSION

- The chi-square computed value of 41.167 is greater than the chi-square tabular value of 7.815 at .05 level of significance with 3 degrees 9 of freedom, so the research hypothesis is confirmed which means that the machine is not mixing the nuts in the ratio of 4:3:2:1. It implies that the machine is not in good order because it does not anymore mix the nuts as expected.

EXAMPLE 2

- The penotypic ratio is 85 of the A type and 15 of the a-type (homozygous recessive). In a monohybrid cross between two heterozygotes, however, it was predicted a 3:1 ratio of phenotypes. In other words, it was expected to get 75 A type and 25 a-type. Are or results different?

SOLVING BY THE STEPWISE METHOD

I. **PROBLEM:** Is the observed values of monohybrid cross are the same at the ration of 3:1?

II. **HYPOTHESES:**

HO: Is the observed values of monohybrid cross is the same as the theoretical distribution of a 3:1 ratio

H1: Is the observed values of monohybrid cross is NOT the same at the ration of 3:1?

III. LEVEL OF SIGNIFICANCE:

$\alpha = .05$

df =2-1

= 2-1

= 1

$\chi^2_{.05} = 3.481$ *tabular value*

CALCULATE THE CHI SQUARE STATISTIC X2 BY COMPLETING THE FOLLOWING STEPS:

1. For each observed number in the table subtract the corresponding expected number (O — E).
2. Square the difference [(O —E)2].
3. Divide the squares obtained for each cell in the table by the expected number for that cell [(O - E)2 / E].
4. sum all the values for (O - E)2 / E. This is the chi square statistic.

	Observed	Expected	(O — E)	(O — E)2	(O — E)2/ E
A-type	85	75	10	100	1.33
a-type	15	25	10	100	4.0
Total	100	100			5.33

probability level (alpha)

Df	0.5	0.10	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

- The computed x^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 x^2 statistic (5.33) exceeded the critical value for 0.05 probability level (3.841) we can reject the null hypothesis that the observed values of our cross are the same as the theoretical distribution of a 3:1 ratio.


Do not underestimate
yourself by comparing
yourself with others.
It's our differences
that make us unique
and define just
how special we are.

The chi-square test for goodness-of-fit (cont.)

the null hypothesis (H_0) specifies the proportion of observations at each level of the categorical variable. The alternative hypothesis (H_a) is that *at least* one of the specified proportions is not true.

- **The null hypothesis specifies the proportion of the population that should be in each category.**

- **The proportions from the null hypothesis are used to compute expected frequencies that describe how the sample would appear if it were in perfect agreement with the null hypothesis.**