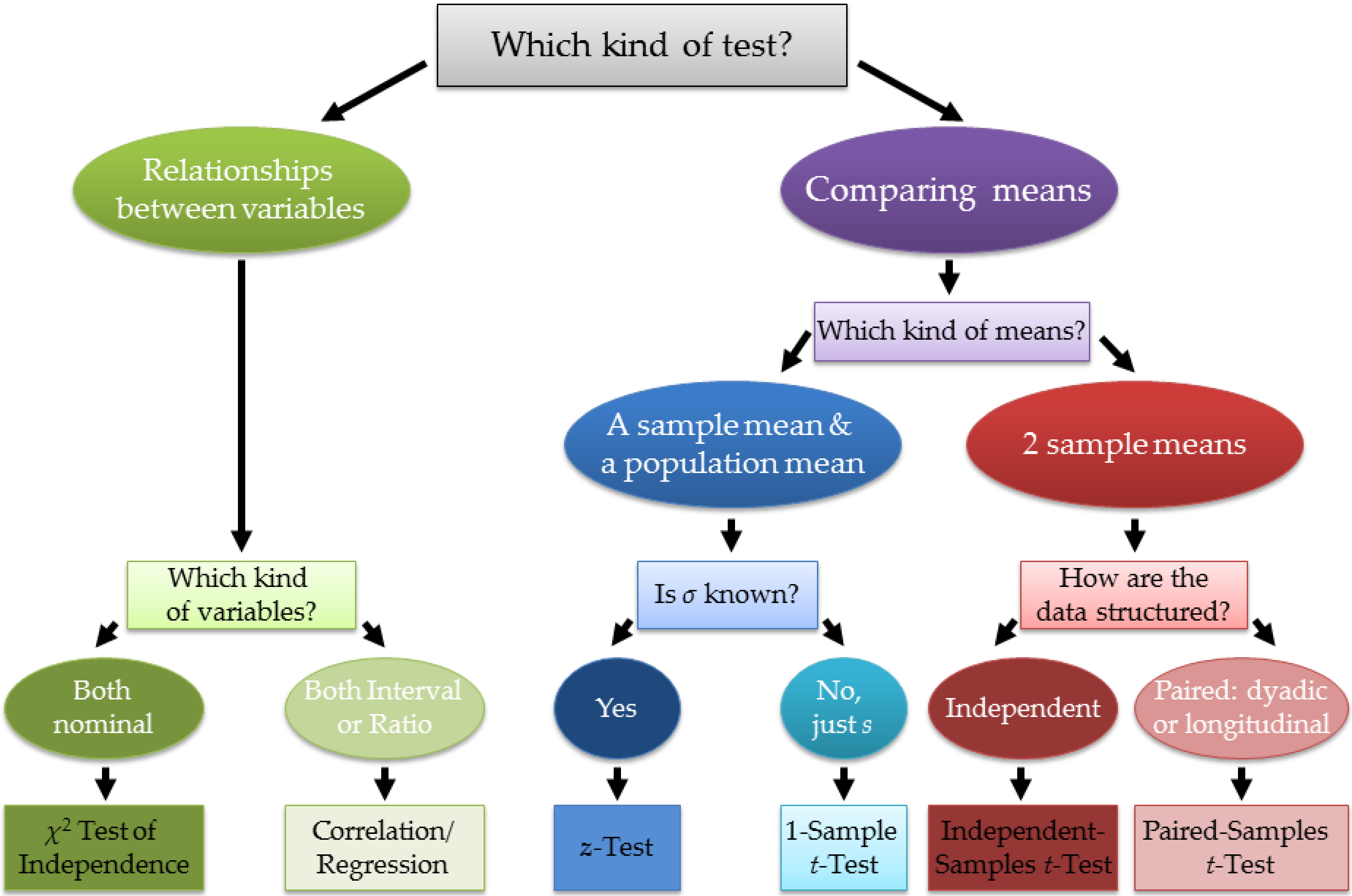
Chi Square (X2) Distribution & Chi Square Test

Prof Dr Najlaa Fawzi

Decision Tree



The testing hypotheses about means, using *t*- and *z*-tests.
These tests share three common features:

--Their hypotheses refer to a population parameter: the population mean.

For this reason, such tests are called parametric tests.

• Their hypotheses concern interval or ratio scale data, such as weight, blood pressure, IQ, per improvement.capita income, measures of clinical

---They make certain assumptions about the distribution of the data of interest in the population—principally,
that the population data are normally distributed.

There are other statistical techniques that do not share these features:

They do not test hypotheses concerning parameters, so they are known as nonparametric tests.

- **They do not assume that the population is normally distributed, so they are also called distribution -free tests.**
- **They are used to test nominal or ordinal scale data.**

Introduction to Nonparametric Stats

--- Parametric and nonparametric are two broad classifications of statistical procedures.

--- Parametric tests are based on assumptions about the distribution of the underlying population from which the sample was taken. The most common parametric assumption is that data are approximately normally distributed.

--- Nonparametric tests do not rely on assumptions about the shape or parameters of the underlying population distribution.

--- If the data deviate strongly from the assumptions of a parametric procedure, using the parametric procedure could lead to incorrect conclusions.

----should be aware of the assumptions associated with a parametric procedure and should learn methods to evaluate the validity of those assumptions.

---If determine that the assumptions of the parametric procedure are not valid, use an analogous nonparametric procedure instead.

---The parametric assumption of normality is particularly worrisome for small sample sizes (n < 30). Nonparametric tests are often a good option for these data.
----It can be difficult to decide whether to use a parametric or nonparametric procedure in some cases.

--- Nonparametric procedures generally have less power for the same sample size than the corresponding parametric procedure if the data truly are normal.

--- Interpretation of nonparametric procedures can also be more difficult than for parametric procedures.

Used when the assumptions for a parametric test have not been met:

- **Data not on an interval or ratio scale**
- **Observations not drawn from a normally distributed population**
- **Variance in groups being compared is not homogeneous**
- **Chi-Square test is the most commonly used when nominal level data is collected**

Analysis Type	Example	Parametric Procedure	Nonparametric Procedure
Compare means between two distinct/independent groups	Is the mean systolic blood pressure (at baseline) for patients assigned to placebo different from the mean for patients assigned to the treatment group?	Two-sample t-test	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Was there a significant change in systolic blood pressure between baseline and the six-month follow-up measurement in the treatment group?	Paired t-test	Wilcoxon signed-rank test

Compare means between three or more distinct/independent groups	If our experiment had three groups (e.g., placebo, new drug #1, new drug #2), we might want to know whether the mean systolic blood pressure at baseline differed among the three groups?	Analysis of variance (ANOVA)	Kruskal-Wallis test
Estimate the degree of association between two quantitative variables	Is systolic blood pressure associated with the patient's age?	Pearson coefficient of correlation	Spearman's rank correlation

Why don’t we always use nonparametric tests?

They have two main drawbacks.

--- The first is that they generally are less statistically powerful than the analogous parametric procedure when the data truly are approximately normal.

“Less powerful” means that there is a smaller probability that the procedure will tell us that two variables are associated with each other when they in fact truly are associated.

If you are planning a study and trying to determine how many patients to include, a nonparametric test will require a slightly larger sample size to have the same power as the corresponding parametric test.

The second drawback associated with nonparametric tests is that their results are often less easy to interpret than the results of parametric tests.

Many nonparametric tests use rankings of the values in the data rather than using the actual data.

Chi Square is used when both variables are measured on a nominal scale.

It can be applied to interval or ratio data that have been categorized into a small number of groups.

It assumes that the observations are randomly sampled from the population.

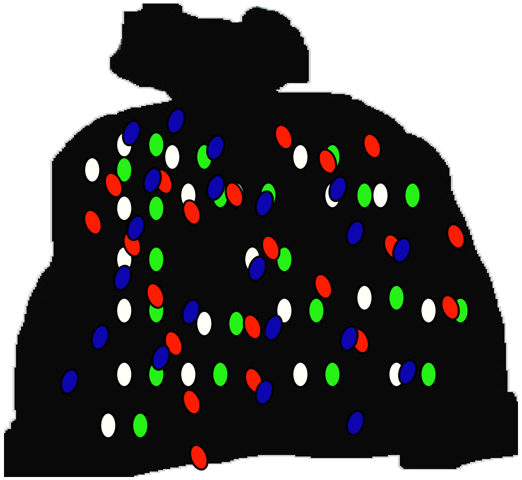
All observations are independent (an individual can appear only once in a table and there are no overlapping categories).

It does not make any assumptions about the shape of the distribution nor about the homogeneity of variances.

Different Scales, Different Measures of Association

Scale of Both Variables	Measures of Association
Nominal Scale	Pearson Chi-Square: χ^2
Ordinal Scale	Spearman's rho
Interval or Ratio Scale	Pearson r

Pearson Chi-Square:



- **Frequencies**

  No mean and SD

statistics → χ^2

No assumption of normality

  Non-parametric test

The X2 test tells the presence or absence of an association between two events or characters but <u>does not measures the strength of association</u> .

The statistical finding or relationship , <u>does not indicate the cause and</u> <u>effect</u> .

The X2 values do not tell why smoking and cancer are associated , it only tells the probability (p) of occurrence of association by chance .

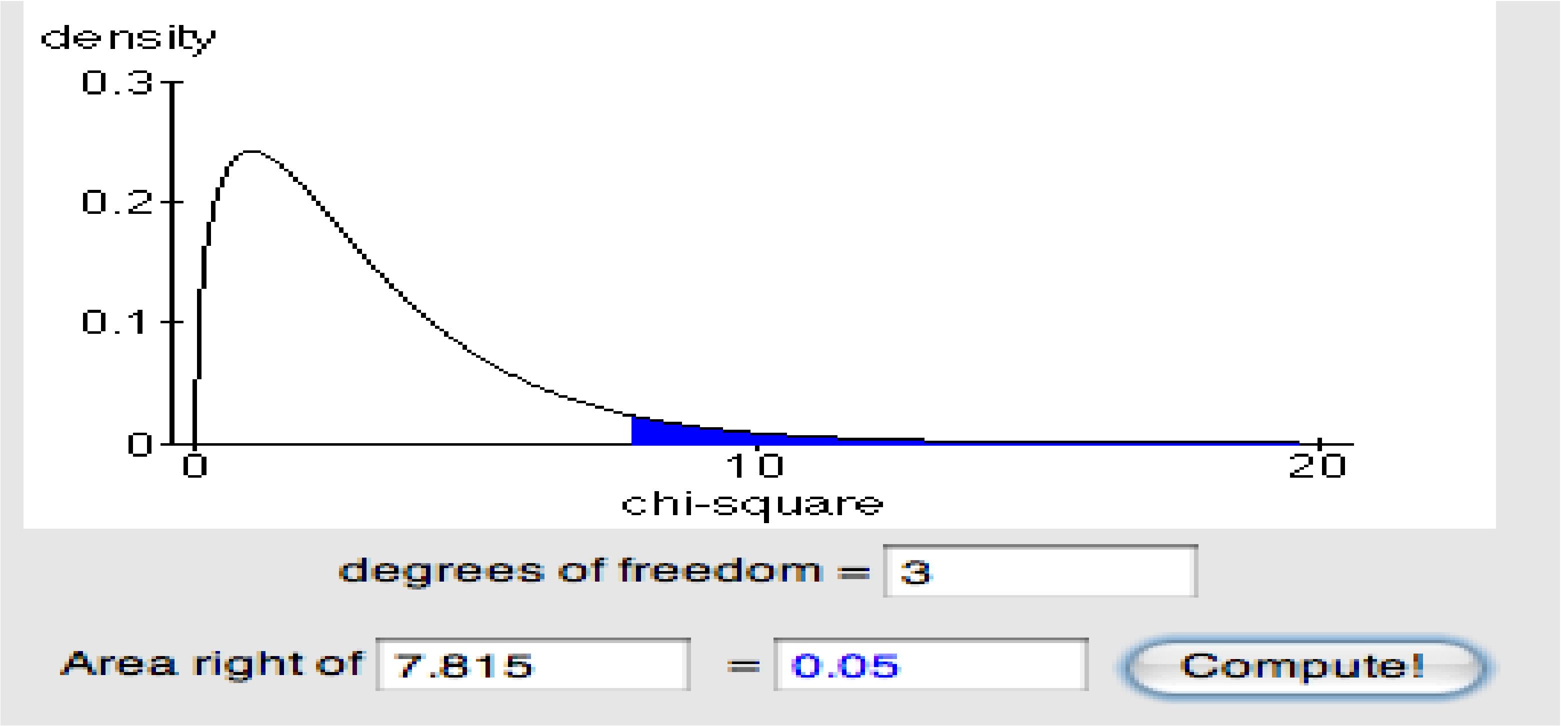
Chi Square (x^2) Distribution & Chi Square Test

The most widely used X2 test is test of independence.

Properties:

1. It is a non-parametric test (deals with frequencies).
2. It is one of the most widely used tests in statistical application.
3. Derived from normal distribution.

4. X^2 assumes values between zero and + ∞, i.e. no negative values, and has one tailed curve.


density
0.3
0.2
0.1
0
0
10
20
chi-square
degrees of freedom = 3
Area right of 7.815 = 0.05
Compute!

5. X2 relates to the frequencies of occurrence individuals or events in categories of one or more Of variables.

6. X2 is used to test agreements between the observed frequencies with certain characteristics and expected frequencies under certain hypothesis.

Categorical data may be displayed in contingency tables

The chi-square statistic compares the observed count in each table cell to the count which would be expected under the assumption of no association between the row and column classifications

The chi-square statistic may be used to test the hypothesis of no association between two or more groups, populations, or Criteria.

Observed counts are compared to expected counts.

	Smoking (Y_1)	No (Y_2)	Total
M.I (X_1)	a	b	a + b
Not (X_2)	c	d	c + d
totals	a + c	b + d	a + b + c + d

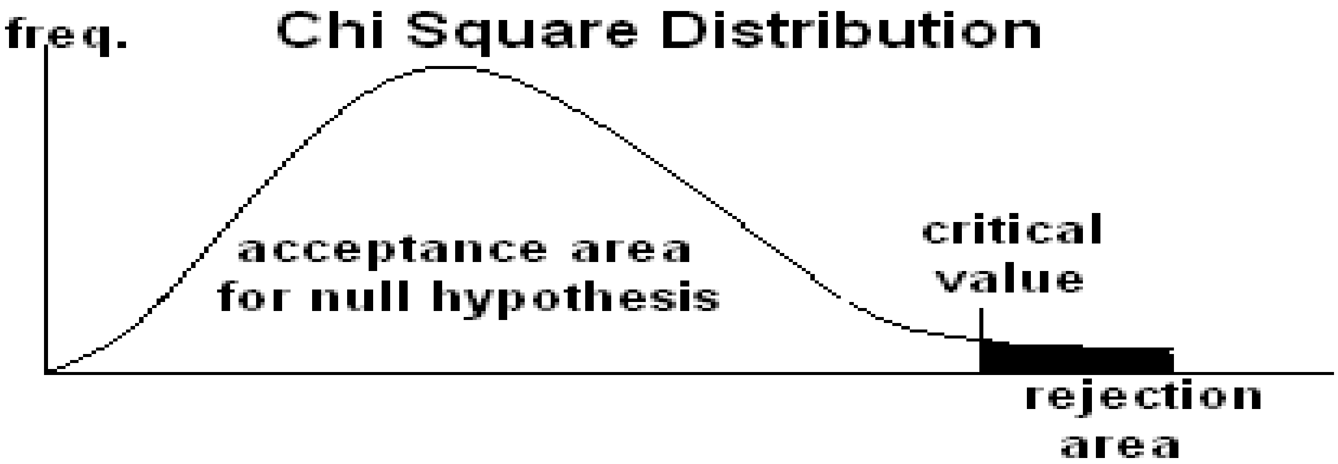
GENERALLY: two criteria classification, when applied to the same set of entities, are independent (no association). In other words; if a sample of n size drawn from a population; the frequency of occurrence of entities are cross classified on the basis of the two variables of interest (X &Y).

	Smoking (Y_1)	No (Y_2)	Total
M.I (X_1)	a	b	a + b
Not (X_2)	c	d	c + d
totals	a + c	b + d	a + b + c + d

The corresponding cells are formed by the intersection of rows and columns & constructed table is a contingency table as the adjacent cells are interrelated.

Hypothesis and conclusion are stated on in terms of association or lack of association of the two variables.

(H0: no association & HA: there is an association).



Critical value = Tabulated X^2 = d. f X^2 (1- α) (from the X^2 table)

X^2 distribution curve is a single tail curve so α is not divided by 2.

Tabulated X^2 = 3.841 for 2X2 table with α = 0.05 (df 1 X^2 0.95)

d. f = (r-1)(c-1) , where r = no. of rows & c = no. of columns, so d. f always equals to 1 in 2X2 table.

Observed frequencies

Expected frequency

Calculated $X^2 = \sum [(O - E)^2 / E]$, where for each cell:

Expected frequency

O = observed frequency in the table
E = expected frequency

Calculation of expected frequencies is based on the probability theory.

E = (row marginal total x column marginal total) / grand total

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

χ^2 = the test statistic $\sum$ = the sum of

O = Observed frequencies E = Expected frequencies

Types of Chi-Square Tests

1- Tests of goodness – of-fit

2-Tests of independence

The X^2 goodness of fit test.

– **Used when we have distributions of frequencies across two or more categories on *one* variable.**

– **Test determines how well a hypothesized distribution fits an obtained distribution.**

We use it to refer to a comparison of a sample distribution to some theoretical distribution that it assumed describes the population from which the sample came.

Could test whether *two* proportions are the same using a two-proportion z test…. but we have 3 groups.

Chi-square goodness-of-fit tests against given proportions (theoretical models) …. but we want to know if choices have *changed*.

So… we'll use a chi-square test of homogeneity. Homogeneity means that things are the same so we have a built-in null hypothesis – the distribution does not change from group to group. This test looks for differences too large from what we might expect from random sample-to-sample variation.

NOTES:

A. Find the association = chi-squared test.

B. No association in this example is to expect these findings to match the hypothetical or theoretical distribution (observed – expected = 0).

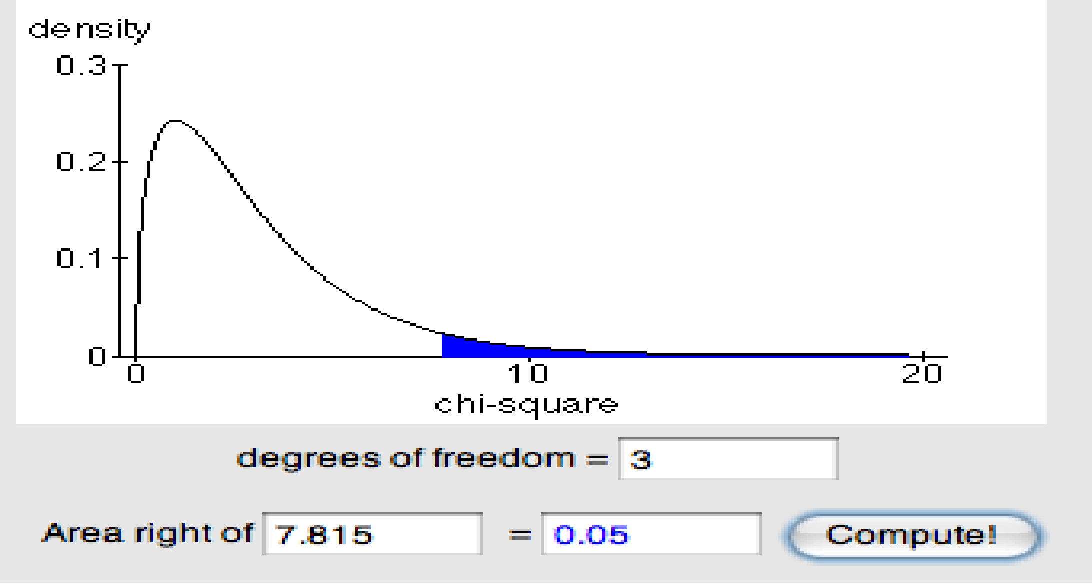
C. The expected values (theoretical or hypothetical) can be either:

- **Given: ratio, percent, proportion, incidence, prevalence or certain laws or rules.**
- **Not given & here we assume NO difference i.e. equality 50:50 if two groups, 25,25,25,25 if 4 groups, etc.**

D. To apply chi-squared test for goodness of fit, we need 3 criteria:

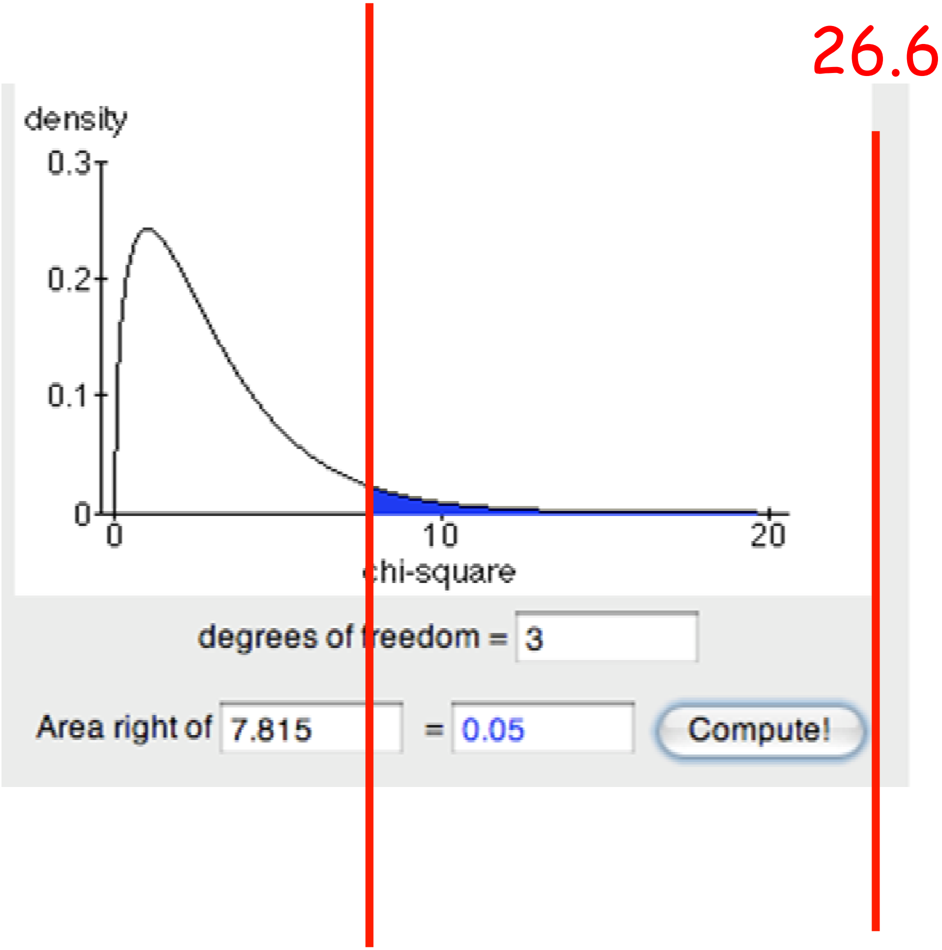
- **Qualitative variables (frequency, proportion, percent…).**
- **One group of data (which is divided into subgroups).**
- **The expected distribution (Theoretical, hypothetical, etc.).**

The df for this test are k-1, where k is the number of cells.



Critical value = 7.81

χ = 26.66
p< 0.001



- **Chi-Square test for goodness of fit is like one sample t-test**
- **You can test your sample against any possible expected values**

TABLE 2.1

Chi Square Values and Probability

Degrees of Freedom	*P* = 0.99	0.95	0.80	0.50	0.20	0.05	0.01
1	0.000157	0.00393	0.0642	0.455	1.642	**3.841**	**6.635**
2	0.020	0.103	0.446	1.386	3.219	**5.991**	**9.210**
3	0.115	0.352	1.005	2.366	4.642	**7.815**	**11.345**
4	0.297	0.711	1.649	3.357	5.989	**9.488**	**13.277**
5	0.554	1.145	2.343	4.351	7.289	**11.070**	**15.086**
6	0.872	1.635	3.070	5.348	8.558	**12.592**	**16.812**
7	1.239	2.167	3.822	6.346	9.803	**14.067**	**18.475**
8	1.646	2.733	4.594	7.344	11.030	**15.507**	**20.090**
9	2.088	3.325	5.380	8.343	12.242	**16.919**	**21.666**
10	2.558	3.940	6.179	9.342	13.442	**18.307**	**23.209**
15	5.229	7.261	10.307	14.339	19.311	**24.996**	**30.578**
20	8.260	10.851	14.578	19.337	25.038	**31.410**	**37.566**
25	11.524	14.611	18.940	24.337	30.675	**37.652**	**44.314**
30	14.953	18.493	23.364	29.336	36.250	**43.773**	**50.892**

From Fisher, R. A., and Yates, F. (1943) *Statistical Tables for Biological, Agricultural, and Medical Research.* Oliver and Boyd, London.

Example: In health district reported the numbers of vaccine – preventable influenza cases are : December 62, January 84, February 17, March 16 , April 21 . Total reported cases 200.

We are interested in knowing whether the numbers of flu cases in the district are equally distributed among the five flu season months.

Data: **31%, 42%, 8.5% , 8%,10.5%**

Assumptions: we assume that the reported cases of flu selected randomly from a population.

Hypotheses: **Ho** : flu cases of are uniformly distributed over the flu season months.

HA: flu cases are not uniformly distributed ....

Level of significance:

$\alpha = 0.05 \rightarrow$ 5% chance factor effect

95% influencing factor effect

d. f = k-1= 5-1=4,   where k = no. of subgroups

Critical point = tabulated X^2 df 4 α 0.05 = 9.49

Testing for significance:

Expected frequency =

total freq x probability individual falls in the category

200/5 = 40 cases per month

Calculated $X^2 = \sum [(O - E)^2 / E] =$

$(62-40)^2/40 + (84-40)^2/40 + (17-40)^2/40 + (16-40)^2/40 + (21-40)^2/40$

$X^2 = 9.025$

Statistical decision: Cal value , is smaller than the tab value ,

We not reject Ho. $p > 0.05$

Conclusion: **we conclude that the occurrence of flu cases follow a uniform distribution.**

The greater the differences between the observed and expected observations , the larger the value of X2 and the less likely it is that the difference is due to chance.

Tests of Independence

The most frequent use of the chi-square test is to test the Ho that two criteria of classification , when applied to the same set of entities , are independent.

- When we have two or more sets of categorical data (IV,DV both categorical)

We have 2x2 and {ax b} or we called[C x r] depending on the number of rows and columns.

EXAMPLE: **two randomly selected samples: 50 child with leukemia 20 males and 30 females, and another 50 healthy children 24 males and 26 females.**

Is the occurrence of leukemia is affected by sex?

	M	F	Total
Leukemia	**30** (27)	**20** (23)	**50**
Healthy	**24** (27)	**26** (23)	**50**
Totals	**54**	**46**	**100**

Data: the two randomly selected samples, 1st of 50 leukemic children consisting of 30 M & 20 F, and 2nd sample of 50 healthy children consisting of 24 M & 26 F.

Assumption: the two samples represent 2 independent groups are taken randomly from 2 independent populations.

Hypotheses:

HO: no significant difference in M & F frequencies with and without leukemia.

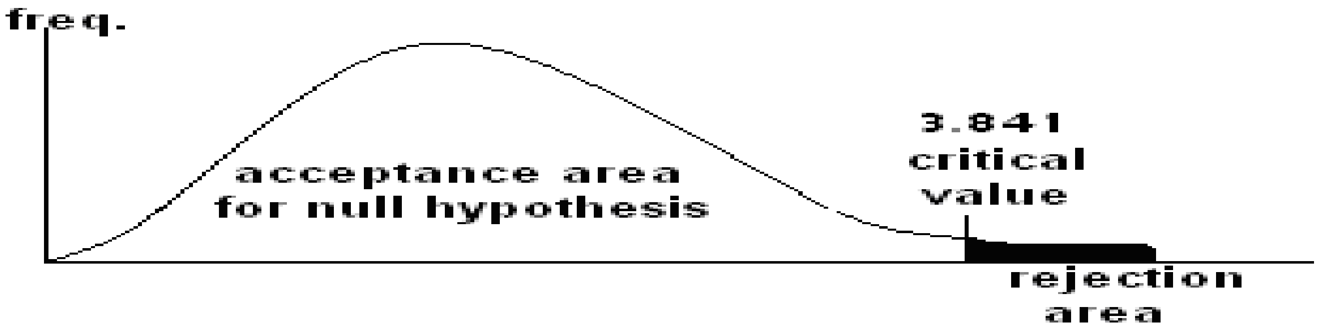
Or there is no association between leukemia and gender type.

HA: there is significant…

Level of significance: α= 0.05 → 5% chance factor effect
95% influencing factor effect

d. f = (C-1)(r -1) = (2-1)(2-1) = 1x 1 = 1

Critical point = tabulated X2 = = 3.841



Expected value for each cell = multiplication of marginal totals/ grand total

So for cell a we call it O_1= 30 , E_1 = 50x54/100= 27

Testing for significance:

Calculated $X^2 = \sum [(O - E)^2 / E]$

= (30-27)2/27 + (20-23)2/23 + (24-27)2/27 + (26-23)2/23 = 1.448

As calculated X2 < tabulated X2→ p >0.05, so we not reject Ho

Conclusion: **there is no association between leukemia & gender type.**

Short cut chi-square formula for 2x2 table

$X2 = n(ad - b\ c)^2 / [(a+c)\ (b+d)\ (a+b)\ (c+d)$

X2- test is <u>avoided</u> if:

<u>Small Expected Frequencies:</u> For contingency tables with more than 1 degree of freedom, a minimum expectation of 1 is allowable if no more than 20% of the cells have expected frequencies of less than 5.

To meet this rule, adjacent rows and / or adjacent columns may be combined . If X2 is based on less than 30 df ,expected frequencies as small as 2 can be tolerated .

<u>Expected frequency</u> of any cell is <u>less than 1</u> whatever the sample size was.

Examples on cells modification if they have small frequencies or zero making chi-squared test not applicable:

Disease	*M*	*F*		*Disease*	*M*	*F*
Leukemia	20 (15.3)	30 (34.6)	→→	Leukemia	20	30
Healthy	0 (15.3)	50 (34.6)		Non-leukemia	26	74
Diabetic	26 (15.3)	24 (34.6)				

Smoking	*M*	*F*		*Smoking state*	*M*	*F*
cigarette	3	5	→→	smoker	3 (5)	8 (6)
pipe	0	3		Non-smoker	7 (5)	4 (6)
Non smokers	7	4				

Interpret x2 test with caution if sample total or total of values in all the cells, is less than 50

Fisher's exact test

Is an alternative for X^2 test for 2x2 tables if:

1. **Grand total was less than 20.**

2. **Grand total was 20 < N < 40 and expected frequency for any cell was less than 5.**

	Guessed correctly	Guessed incorrectly	TOTAL
Letterman	8	1	9
Shaffer	4	5	9
TOTAL	12	6	18

Player:	Result		
	Correct	Incorrect	All
David	**8**	**1**	**9**
	6.00	3.00	9.00
Paul	**4**	**5**	**9**
	6.00	3.00	9.00
Total	**12**	**6**	**18**
	12.00	6.00	18.00

Chi-Square = 4.000, DF = 1, P-Value = 0.046

2 cells with expected counts less than 5.0

The resulting p-value, 0.046, from the test indicates there is a statistically significant difference (at the α = 0.05 level) in the success rates between Letterman and Shaffer.

Some practitioners will experience a problem when an expected value is less than five

Sometimes it's appropriate to group certain categories to avoid the problem, but this is clearly not possible when there are only two categories.

There are two cells in which the expected counts are less than five.

Fisher's exact test considers all the possible cell combinations that would still result in the marginal frequencies as highlighted (namely 9, 9 and 12, 6).

The test is exact because it uses the exact hyper geometric distribution rather than the approximatechi-square distribution to compute the p-value.

The resulting p-value using Fisher's exact test is 0.1312. Therefore, you would fail to reject the null hypothesis of equal proportions at the α = 0.05 level.

This contradicts the X2 results from the test and indicates the test provided a poor approximation to the exact results.

The computations involved in Fisher's exact test may be extremely time consuming to calculate by hand.

Implications

It's appropriate to use Fisher's exact test, in particular when dealing with small counts.

The X2 test is basically an approximation of the results from the exact test, so erroneous results could potentially be obtained from the few observations.

TABLE	ASSOCIATED PROBABILITY
9 0 3 6	9! *9!* 12!*6!/ 18!*9!*0!*3!*6! =0.004524*
7 2 5 4	9!*9!*12!*6!/ 18!*7!*2!*5!*4!= 0.244343
6 3 6 3	9!*9!*12!*6! / 18!* 6!* 3!*6!*3! = 0.38009
5 4 7 2	9!*9!*12!*6! / 18!* 5!* 4!*7!*2! =0.24434
4 5 8 1	9!*9!*1!*2!*6! / 18!*4!*5!*8!*1!= 0.06108*
3 6 9 0	9!*9!*1!*2!*6! / 18!*3!*6!*9!*0! = 0.00452*

This particular p-value is 0.13122.

Yates's Correction: used if one of observed frequency is small , lower than 5 in 2x2 contingency table.

It involves subtracting 0.5 from the difference between O and E frequencies in the numerator of X2 before squaring .

$$X2 = \sum [(O - E - 0.5)^2 / E]$$

It has the effect of making the value for X2 smaller . Smaller value means that the Ho will not be rejected as often as it is with large in corrected X2 .

It is more conservative , the risk of a type I error is smaller , while the risk of type II error is increase.

McNemar test : **it is used for dichomatous data , it is similar to paired t test.**

The difference being that paired t test is used for variable which is normally distributed in population such as mean systolic B p ,serum sodium concentration where as McNemar test is used when the data are dichomatous form (yes & no- either the person is hypertensive or non hypertensive , either he is diseased or not diseased).

Severe cold at age 12	Severe cold at age 14 Yes	No	Total
Yes	212	144	356
No	256	707	963
Total	468	851	1319

What do these mean?

Statistics for Table of Sex by Eyes

Statistic	DF	Value	Prob
Chi-Square	2	0.7504	0.6872
Likelihood Ratio Chi-Square	2	0.7583	0.6845
Mantel-Haenszel Chi-Square	1	0.0015	0.9689
Phi Coefficient		0.1370	
Contingency Coefficient		0.1357	
Cramer's V		0.1370	

Likelihood Ratio Chi Square

$$G^2 = 2 \sum i \sum j \; n_{ij} \ln(n_{ij}/e_{ij})$$

- G^2 follows the Chi-Square distribution with (R-1)(C-1) degrees of freedom
- Builds on the likelihood of the data under the null hypothesis relative to the maximum likelihood
- Result is usually slightly larger than the Pearson Chi-Square
- Suggested that the Likelihood Ratio Chi-square statistic could be more effective than the Pearson Chi-square statistic when observed or expected frequencies in cells are less than 5 (Ozdemir and Eyduran 2005)

Continuity-Adjusted Chi-Square Test

$$Q_c = \sum i \sum j \frac{(\max(0,|nij - eij| - 0.5))^2}{eij}$$

- This is the Yates Correction
- Use when there is only 1 degree of freedom in a 2x2 contigency table
- Most useful for small sample sizes
- As the sample size increases, becomes more similar to Pearson Chi-square
- Follows chi-square distribution with (R-1)(C-1) degrees of freedom

Mantel- Haenszel Chi-Square Test

$$Q_{MH} = (n-1)r^2$$

r^2 is the Pearson correlation coefficient (which also measures the linear association between row and column)

Tests alternative hypothesis that there is a linear association between the row and column variable

Follows a Chi-square distribution with 1 degree of freedom

Phi Coefficient

Pearson Chi-Square provides information about the existence of relationship between 2 nominal variables, but not about the magnitude of the relationship

Phi coefficient is the measure of the strength of the association. Appropriate for measuring degree of association between two binary variables.

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

Cramer's V

When the table is larger than 2 by 2, a different index must be used to measure the strength of the relationship between the variables. One such index is Cramer's V.

If Cramer's V is large, it means that there is a tendency for particular categories of the first variable to be associated with particular categories of the second variable.

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

No.of cases

Smallest no.of rows or columnes