- <u>Contingency Tables:</u>
- In many situations, enumeration data collected simultaneously for two variables, and it is desired to test the hypothesis that the frequencies of occurrence categories of one variable are independent of frequencies in the second variable.
- <u>*The 2X2 Contingency table*</u>: It is commonly encountered in biological research. For such tables, the computing formula:

Second criteria of classification	First criteria of Classification 1	2	Total
1	a	b	a+b
2	c	d	c+d
Total	a+c	b+d	n

$$\chi^2 = \frac{n(ad - bc)^2}{(a+c)(b+d)(a+b)(c+d)} \quad \text{-------------- (1)}$$

$$\chi^2 = \frac{n(|ad - bc| - 0.5n)^2}{(a+c)(b+d)(a+b)(c+d)} \quad \text{-------------- (2) Yate's Correction}$$

- <u>Note</u>: If $|ad - bc| \leq n/2$, Yate's correction will increase χ^2, rather than decrease it.
- Equation (1) should be used.

- Example: A group of 350 adults who participated in a health survey were asked whether or not they were on diet. The responses by sex are as follows.

	Male	Female	Total
On diet	14	25	39
Not on diet	159	152	311
Total	173	177	350

- Do these data suggest that being on diet is independent on sex? Let $\alpha=0.05$.
  - H_0: Sex and diet are independent. H_1: Sex and diet are dependent.

$$|14*152-25*159|>350/2$$

- $$c^2=\frac{350\left(|14*152-25*159|-0.5*350\right)^2}{173*177*39*311}=2.634$$
- *df* = (r-1) (c-1) = (2-1) (2-1) = 1

From precentiles of the chi - sequare distribution table,

$$c^2_{(1-a),v}=c^2_{(0.95),1}=3.841$$

- *Statistical decision*: Accept H0, since $\chi^2 < 3.841$
- *Clinical decision*: Conclude that, on the bases of these data, The sex and diet are independent. (P>0.05).

- <u>Test of Homogeneity:</u>
- The test of homogeneity is concerned with the question. Are the samples drawn from *<u>populations</u>* are homogeneous with respect to some criterion of classification? *Either row or column total may be under control of the investigator, that is the investigator may specify independent samples be drawn from each of several populations.*
- <u>Example:</u> In air pollution study, a random sample of 200 household was selected from each of two communities. A respondent in each household was asked whether or not anyone in the household was bothered by air pollution. Can the researches conclude that the two communities differ with respect to variable of interest? Let α =0.01.

	Any Member of Household Bothered by Air Pollution		
Community	Yes	No	Total
I	43	157	200
II	81	119	200
Total	124	276	400

H_0: the two populations are homogenous.

H_1: the two populations are not homogenous.

$$|43*119 - 81*157| > 400/2$$

$$c^2 = \frac{400\left(|43*119 - 81*157| - 0.5*400\right)^2}{124*276*200*200} = 16$$

From precentiles of the chi - sequare distribution table,

$$c^2_{(1-a),v} = c^2_{(0.99),1} = 6.635$$

- *Statistical decision*: reject H_0, since $c^2 > 6.635$
- *Clinical decision*: Conclude that, on the bases of these data, The two populations are not homogenous with respect to household bothered by air pollution. (P<0.01).

Note:

1- Yarnold (1970), suggested that the minimum cell expectation should be greater than 5s/r c, where s is the number of cell expectation less than 5.

2- Fisher mentioned that, some times we have data that can be summarized in a 2X2 contingency table, but these are derived from very small samples. For example, n less than 20 or if n is between 20 and 40 and one of expected frequencies is less than 5 the Chi-square test is avoided.

Example: retrospective assessment of smoking frequency. The table displays the daily average number of cigarettes for lung cancer patients and control patients.

Daily No. cigarettes

	None	< 5	5-14	15-24	25-49	50+	Total
Cancer	7	55	489	475	293	38	1357
	(34)*	(92)	(529.5)	(453)	(223.5)	25	1357
Control	61	129	570	431	154	12	1357
	(34)	(92)	(529.5)	(453)	(223.5)	25	1357
Total	68	184	1059	906	447	50	2714

we want to test whether the smoking frequency is the same for each of the populations sampled. **Let α = 0.05.**

H0: smoking frequency same in both groups.

HA: smoking frequency not the same.

- **(34)= (1357*68)/ 2714**

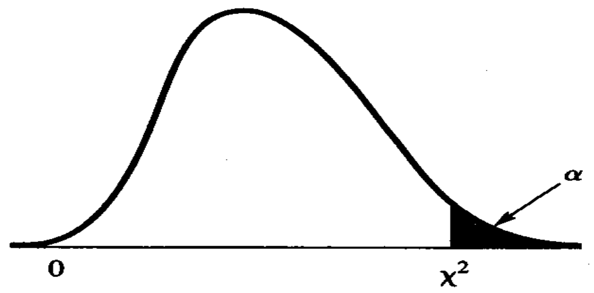
$$\chi^2 = \frac{(7-34)^2}{34} + \frac{(55-92)^2}{92} + \dots + \frac{(12-25)^2}{25} = 137.7$$

Looking in $\chi^2_{(95),5}$ table, we find that = 11.07, H_o is rejected, since $\chi^2 > 11.07$

§*Clinical decision*: Conclude that, on the bases of these data, The two populations are not homogenous with respect to two groups. (P<0.05).

Chi-Square Table

Critical values of $\chi\chi^2$ at 1 df: 3.841 at p = 0.05, and 6.635 at p = 0.01



df	α .05	α .01	df	α .05	α .01
1	3.841	6.635	16	26.296	32.000
2	5.991	9.210	17	27.587	33.409
3	7.815	11.345	18	28.869	34.805
4	9.488	13.277	19	30.144	36.191
5	11.070	15.086	20	31.410	37.566
6	12.592	16.812	21	32.671	38.932
7	14.067	18.475	22	33.924	40.289
8	15.507	20.090	23	35.172	41.638
9	16.919	21.666	24	36.415	42.980
10	18.307	23.209	25	37.652	44.314
11	19.675	24.725	26	38.885	45.642
12	21.026	26.217	27	40.113	46.963
13	22.362	27.688	28	41.337	48.278
14	23.685	29.141	29	42.557	49.588
15	24.996	30.578	30	43.773	50.892

Source: R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research,* 4th ed., Longman Group Ltd., London (previously published by Oliver & Boyd, Edinburgh), Table IV, by permission of the authors and the publisher.

The Chi-Square Distribution

	Probability	
Degrees of Freedom	.05	.01
1	3.84	6.64
2	5.99	9.21
3	7.82	11.34
4	9.45	13.28
5	11.07	15.09
6	12.59	16.82
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21
11	19.68	24.73
12	21.03	26.22
13	22.36	37.69
14	23.69	29.14
15	25.00	30.58
16	26.30	32.00
17	27.59	33.41
18	28.87	34.81
19	30.14	36.19
20	31.41	37.57
21	32.67	38.93
22	33.92	40.29
23	35.17	41.64
24	36.42	42.98
25	37.65	44.31