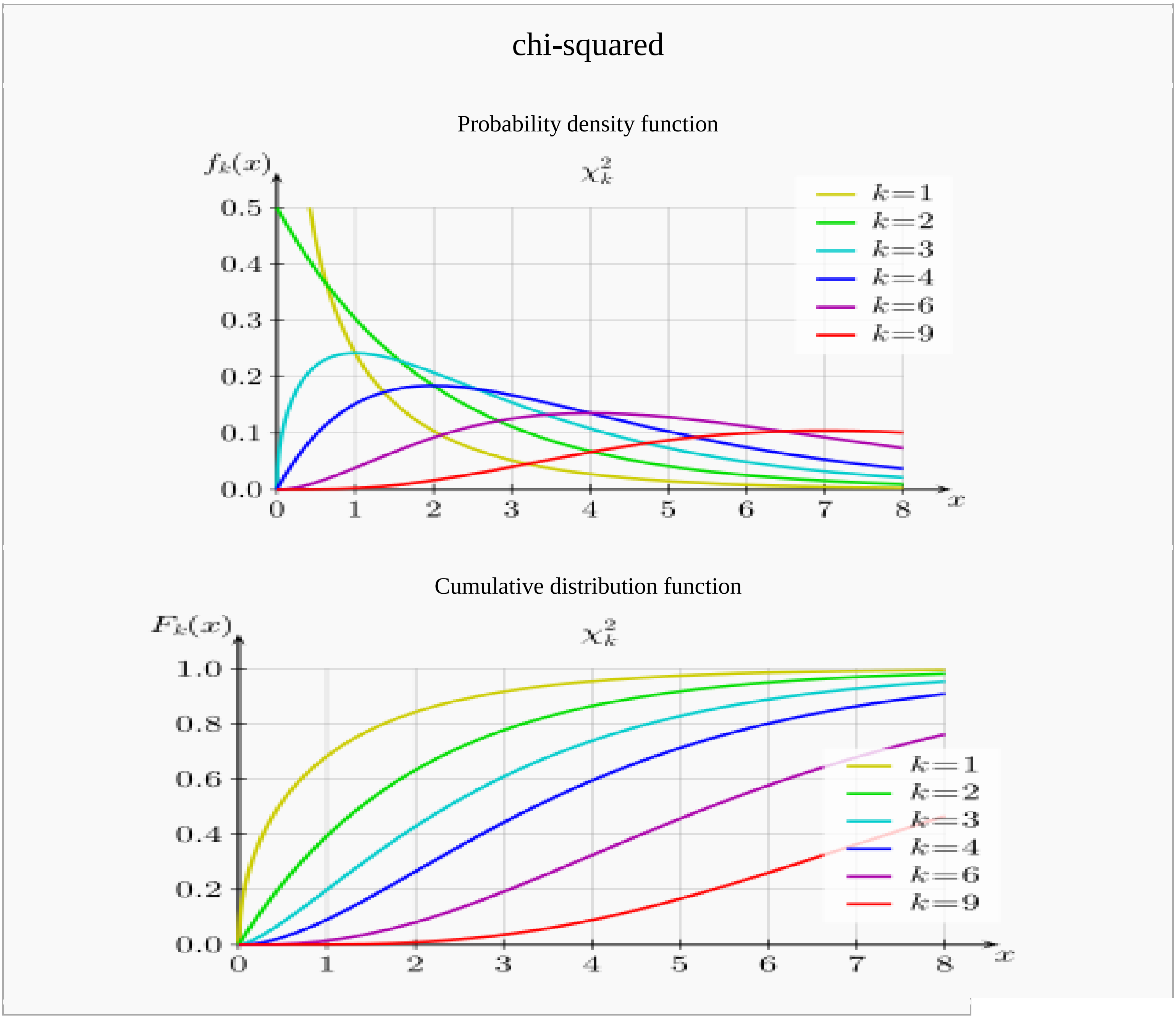
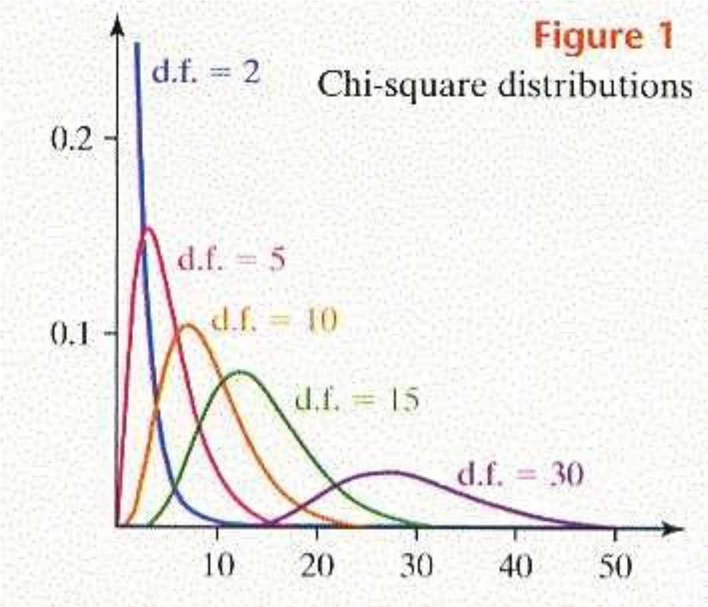
Lecture 11:

Chi-squared distribution

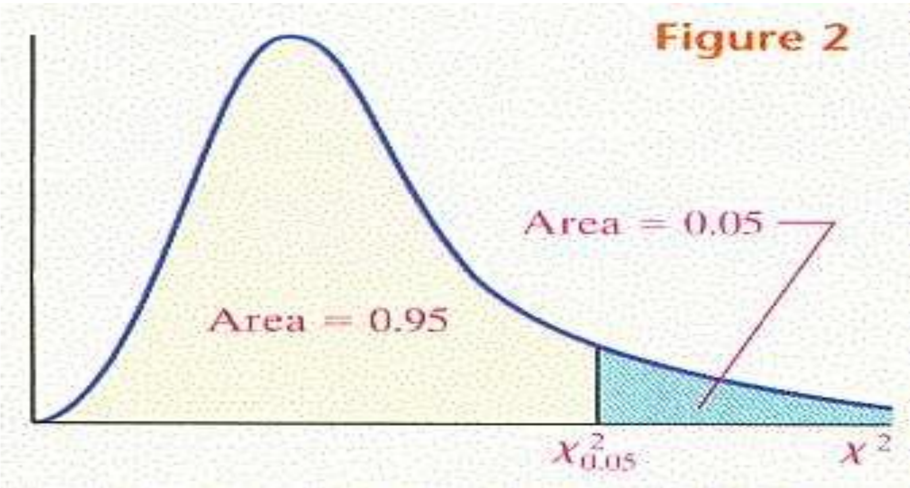


Characteristics of the Chi-Square Distribution:

1. It is not symmetric.
2. The values of X^2 are non-negative
3. The chi-square distribution is to the horizontal axis on the right-hand-side.
4. The shape of the chi-square distribution depends upon the degrees of freedom, just like Student’s t-distribution and Fisher’s F-distribution.
5. As the number of degrees of freedom increases, the chi-square distribution becomes more
6. Total area under the curve is equal to 1.0



Finding Critical Values of the Chi-Square Distribution:



Find the critical value of chi-square for a one-tail (right-tail) test with = 0.05 and df=15.

Figure 3

Degrees of Freedom	Area to the Right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.365	7.215	8.231	9.288	10.265	25.200	28.601	31.595	24.265	36.456

Chi-Square: A statistical test that exams the whole distributions, and the relationship between two distributions. In doing this, the data is not summarized into a single measure such as the mean, standard deviation or proportion. The whole distribution of the variable is examined, and inferences concerning the nature of the distribution are obtained

A goodness-of-fit test is a procedure used to determine whether a frequency distribution follows a claimed distribution. It is a test of the agreement or conformity between the observed frequencies (O) and the expected frequencies (E) for several classes or categories

Chi- square ($\times^2$**); a** statistical test used for category data that based on comparison of frequencies observed & expected in various categories.

- A statistical test of significance for 2 qualitative variables as if there is association, effect.

- Use 2 x 2 table , calculate d.f by (raw-1) × (Colum-1)
- For 2 variables with more 2 categories use K × K table

Uses and Applications

- Used when you have frequency distribution of qualitative type of variables
- In 2X2 table, it is used to test whether there is an association between the row and the column variables; ie whether the distribution of individuals among the categories of one variable is independent of their distribution among the categories of the other

Example: **Influenza &vaccination trial**

	Influenza	No influenza	Total
Vaccine	**20**	**220**	**240**
Placebo	**80**	**140**	**220**
Total	**100**	**360**	**460**

The questionis:

Is the difference (in the percentages of influenza) due to vaccination or occurred by chance? Vaccinated had influenza= 20 × 100 / 100 = 20% while non vaccinated had influenza= 80 × 100 / 100 = 80%

Steps of test:

- Assume $\times^2$ distributions. **-** Use 2 x 2 table

- H_O: there is no association between 2 variables, H_A: there is association between 2 variables

- Level of significance (alpha) =0.05

- Calculate K---- df = (raw– 1) × (Colum– 1)

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

- Conclusion : Compare $\times^2$ calculated with $\times^2$ tabulated : If $\times^2$ calculated is > $\times^2$ tabulated reject the H_o & accept H_A If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is association between variables, If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is no association between variables .

	disease +ve	disease -ve	
Test +ve	A	B	A+B
Test -ve	C	D	C+D
	A+C	B+D	A+B+C+D= N

(X²) = $\sum \frac{[O-E]^2}{E}$

Where O = observed number

E = Expected number

Ea= $\frac{(A+C) \times (A+B)}{N}$, Or **Ea = total raw a x total column a / total**

Eb = $\frac{(B+A) \times (B+D)}{N}$, Or **Eb = total raw b x total column b / total**

EC= $\frac{(C+A) \times (C+D)}{N}$, Or **Ec = total raw c x total column c / total**

ED = $\frac{(D+C) \times (D+B)}{N}$, Or **Ed = total raw d x total column d / total**

<u>Example;</u> To assess the possible association between 100% oxygen therapy & development of retinal fibroplasia of 135 premature infants in intensive care units that the result.
(alpha) =0.05

<u>Steps of test</u> : - Assume $\times^2$ distributions. **-** Use 2 x 2 table

- H_o: there is no association between 2 variables, H_A: there is association between 2 variables

- Level of significance (alpha) =0.05

- Calculate K----df = (raw– 1) × (Colum– 1)

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

- Conclusion : Compare$\times^2$ calculated with $\times^2$ tabulated : If $\times^2$ calculated is >$\times^2$ tabulated reject the H_0 & accept H_A If $\times^2$ calculated is < $\times^2$ tabulated , accept H_0 &There is association between variables, If $\times^2$ calculated is <$\times^2$ tabulated , accept H_0 If $\times^2$ calculated is < $\times^2$ tabulated , accept H_0 &There is no association between variables.

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

oxygen therapy	Retinal fibroplasia **+ve**	 -ve	**Total**
+ve	A 36	B 31	**A+B 67**
-ve	C 22	D 46	**C+D 68**
Total	**A+C 58**	**B+D 77**	**N = 135**

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

Ea = $\frac{(a+b) \times (a+c)}{N}$ or **total raw a x total column a / total** =$\frac{(67) X (58)}{135}$=28. 785

Eb = $\frac{(b+a) \times (b+d)}{N}$ or **total raw b x total column b / total** =$\frac{(67) X (77)}{135}$= 38.215

Ec= $\frac{(c+a) \times (c+d)}{N}$ or total raw c x total column c / total =$\frac{(58) \times (68)}{135}$= 29.215

Ed = $\frac{(d+c) \times (d+b)}{N}$ or total raw d x total column d / total =$\frac{(68) \times (77)}{135}$= 38.785

CELL	Observed	Expected	(O-E)	([O-E] ²	([O-E] ²)/E
A	36	28. 785	7.215	52.215	1.808
B	31	38.215	- 7.215	52.215	1.362
C	22	29.215	- 7.215	52.215	1.781
D	46	38.785	7.215	52.215	1.342
				TOTAL	**6.293**

$(\times^2)$ = 6.293 d f = (2-1) x (2-1) =1

(X^2) Cqi -Square Distribution Table

D F			Probability (P Value)		
	0.50	**0.10**	**0.05**	**0.01**	**.001**
1	0.455	2.7o6	3.841	6.63	10.83
2	1.386	4.605	5.991	9.21	13.82
3	2.366	6.251	7.815	11.34	16.27
4	3.357	7.779	9.448	13.28	18.47
5	4.351	9.236	11.o7o	15.o9	20.51

From **Chi -Square Distribution Table (**3.841), calculated ($\times^2$)= 6.293> tabulated

($\times^2$)= 3.841 so $p < 0.05$& there is association between developments of retinol fibroplasia in premature infants & receiving 100% oxygen with non received.

Example; The following table shows mothers on contraceptive pills & their infants developed jaundice? The question if their relation or association between jaundice &pills & what is the **confidence interval** that the proportion of using pills was 57%.

C. P.P	Jaundice +ve-ve		**Total**
Pills +ve	A 33	B 26	**57**
Pills –ve	C 14	D 45	**59**
Total	**47**	**69**	**116**

Steps of test: - Assume $\times^2$ distributions. **-** Use 2 x 2 table

- H_o: there is no association between 2 variables, H_A: there is association between 2 variables

- Level of significance (alpha) =0.05

- Calculate K---- df = (raw– 1) × (Colum– 1)

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

- Conclusion : Compare$\times^2$ calculated with $\times^2$ tabulated : If $\times^2$ calculated is >$\times^2$ tabulated reject the H_o & accept H_A If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is association between variables, If $\times^2$ calculated is <$\times^2$ tabulated , accept H_o If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is no association between variables .

Ea= $\frac{(A+C) \times (A+B)}{N}$ (total raw a x total column a / total) = $\frac{(47) \times (57)}{116}$ = 23. 09

Eb= $\frac{(B+A) \times (B+D)}{N}$(total raw a x total column a / total) = $\frac{(57) \times (67)}{116}$ = 33.91

Ec= $\frac{(C+A) \times (C+D)}{N}$(total raw a x total column a / total) = $\frac{(47) \times (59)}{116}$ = 23.91

Ed = $\frac{(D+C) \times (D+B)}{N}$(total raw a x total column a / total) = $\frac{(59) \times (69)}{116}$ = 35.09

Cell	Observed	Expected	(O-E)	([O-E] ²	([O-E] ²)/E
A	33	23. 09	9.91	98.21	4.25
B	24	33.91	-9.91	98.21	2.90
C	14	23.91	-9.91	98.21	4.11
D	45	35.09	9.91	98.21	2.73
				TOTAL	**13.99**

$(X^2) = 13.99$ d f = 1,   (alpha) = 0.05

χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

	Significance level		
Degrees of freedom	5%	1%	0.1%
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	11.070	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588

Calculated (X^2) 13.99 > tabulated (X^2) 3.841so reject Ho

There is real association between using c.c.p & develop of jaundice in infants

X^2 :<u>Practical</u>

Q: A sample of 150 carriers of a certain antigen and a sample of 500 non\ carriers the following blood group distributions .

Blood group	**Carriers**	**Non Carriers**	**Total**
O	**72 a**	**230 b**	**302**
A	**54 c**	**192 d**	**246**
B	**16 e**	**63 f**	**79**
AB	**8 g**	**15 h**	**23**
Total	**150**	**500**	**650**

Can one conclude from these data that the two populations from which the samples were drawn differ with respect to blood group distribution? $\alpha = 0.05$

Steps of test : - Assume $\times^2$ distributions. - Use 2 x 2 table

- H_o: there is no association between 2 variables, H_A: there is association between 2 variables

- Level of significance (alpha) =0.05

- Calculate K----df = (raw– 1) × (Colum– 1)

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

- Conclusion : Compare$\times^2$ calculated with $\times^2$ tabulated : If $\times^2$ calculated is >$\times^2$ tabulated reject the H_o & accept H_A If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is association between variables, If $\times^2$ calculated is <$\times^2$ tabulated , accept H_o If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is no association between variables.

The answer: dF = (column -1) x (raw -1) = (4 -1) x (2 -1) = 3

$$\sum \frac{[O-E]^2}{E}$$

X^2 = Ea = total raw a x total column a / total = 302 x 150 / 650 = 69.69

Eb = total raw b x total column b / total = 302 x 500 / 650 = 232.31

Ec = total raw c x total column c / total = 246 x 150 / 650 = 56.77

Ed = total raw d x total column d / total = 246 x 500 / 650 = 189.23

Ee = total raw e x total column e / total = 79 x 150 / 650 = 18.23

Ef = total raw f x total column f / total = 79 x 500 / 650 = 60.77

Eg = total raw g x total column g / total = 23x150/650= 5.31

Eh = total raw h x total column h / total = 23x500/650= 17.69

$X^2 = \sum \frac{[O-E]^2}{E}$

a = $(72- 69.69)^2$ / 69.69= 0.076

b =$(230-232.31)^2$/ 232.31=0.023

c = $(54-56.77)^2$ / 56.77 = 0.14

d = $(192-189.23)^2$ /189.23= 0.041

e = $(16-18.23)^2$ / 18.23= 0.27

f = $(63-60.77)^2$ / 60.77= 0.08

g = $(8-5.31)^2$ / 5.31= 1.36

h = $(15-17.69)^2$/ 17.69 =0.41

X^2= 2.4
Tabulated X^2 = 7.815
So calculated X^2 (2.4) <Tabulated X^2 (7.815)
So accept Ho that there is no association between antigen &blood groups

Q: A sample of 500 college students participated in a study designed to evaluate the level college students, knowledge of a certain group of common disease. The following table shows the students classify by major field of study and level of knowledge of the group of diseases:

Knowledge of Diseases			
Major	**Good**	**Poor**	**Total**
Premedical	**31 a**	**91 b**	**122**
Other	**19 c**	**359 d**	**378**
Total	**50**	**450**	**500**

The answer steps:

<u>Steps of test</u> : - Assume $\times^2$ distributions. **-** Use 2 x 2 table

- H_o: there is no association between 2 variables, H_A: there is association between 2 variables

- Level of significance (alpha) =0.05

- Calculate K----df = (raw– 1) × (Colum– 1)

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

- Conclusion : Compare$\times^2$ calculated with $\times^2$ tabulated : If $\times^2$ calculated is >$\times^2$ tabulated reject the H_o & accept H_A If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is association between variables, If $\times^2$ calculated is <$\times^2$ tabulated , accept H_o If $\times^2$ calculated is < $\times^2$ tabulated , accept H_o &There is no association between variables.

The answer:

DF = (column -1) x (raw -1) =(2 -1) x (2 -1) = 1

$$X^2 = \sum \frac{[O-E]^2}{E}$$

Ea = total raw a x total column a / total = 122 x 50 / 500 = 12.2

Eb = total raw b x total column b / total = 122 x 450 / 500 = 109.8

Ec = total raw c x total column c / total = 378 x 50 / 500 = 37.8

Ed = total raw d x total column d / total = 378x 450 / 500 = 340.2

$$X^2 = \sum \frac{[O-E]^2}{E}$$

$a = (31- 12.2)^2 / 12.2 = 28.97$
$b = (91-109.8)^2/ 109.8 =3.22$
$c = (19 -37.8)^2 / 37.8= 9.35$
$d = (359-340.2)^2 /340.2= 1.04$

X^2= 42.58
Tabulated X^2 = 3.841
So calculated X^2 (42/58) >Tabulated X^2= (3.841)
So accept HA that there is association between 2 groups

(X^2) CHI-SQUARE Distribution Table

D F			PROBABLITY (P Value)		
	o.50	0.10	0.05	0.01	.001
1	0.455	2.7o6	3.841	6.63	10.83
2	1.386	4.605	5.991	9.21	13.82
3	2.366	6.251	7.815	11.34	16.27
4	3.357	7.779	9.448	13.28	18.47
5	4.351	9.236	11.o7o	15.o9	20.51