(Lect. 38)
10.12 TEST FOR INDEPENDENCE

The chi-squared test procedure can also be used to test the hypothesis of independence of two variables of classification.

Procedure:

Step 1: Select a fixed significance level α

Step 2: State the Null Hypothesis H_0 and alternative hypothesis H_1 that is we have to test the Null Hypothesis H_0 against the alternative hypothesis H_1

Step 3: Determine

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

where χ^2 is a value of a random variable whose sampling distribution is approximated very closely by the chi-squared distribution with $v = (r-1)(c-1)$ degrees of freedom.

r represents number of row and
c represents number of column.
The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the i^{th} cell.
$k = rc$ represents number of cell.
e_i is obtained by the following formula:

$$e_i = \frac{(Row\ Total) \times (Column\ Total)}{Grand\ Total} = \frac{(RT) \times (CT)}{GT}$$

Step 4: Determine critical value ${\chi_\alpha}^2$ using the following equation
$P(\chi^2 > {\chi_\alpha}^2) = \alpha$ with $(r-1)(c-1)$ degrees of freedom (Use χ^2 distribution table)

Step 5: Determine the critical region and fail to reject region based on α, using χ^2-distribution table with $(r-1)(c-1)$ degrees of freedom.
Here Critical region is $\chi^2 > {\chi_\alpha}^2$
Fail to reject null hypothesis H_0 region is $\chi^2 \leq {\chi_\alpha}^2$

Question No.87

A random sample of 90 adults is classified according to gender and the number of hours of television watched during a week:

	Gender	
	Male	Female
Over 25 hours	15	29
Under 25 hours	27	19

(0.1)

Use a 0.01 level of significance and test the hypothesis that the time spent watching television is independent of whether the viewer is male or female.

Solution:
Step 1: significance level α=0.01
Step 2: we have to test the Null Hypothesis H_0:the time spent watching television is independent of whether the viewer is male or female against the alternative hypothesis
H_1:the time spent watching television is not independent of whether the viewer is male or female.
Step 3:

Observed and expected frequencies

	Male	Female	Total
Over 25 hours	15(20.5)	29(23.5)	$r_1 = 44$
Under 25 hours	27(21.5)	19(24.5)	$r_2 = 46$
Total	$c_1 = 42$	$c_2 = 48$	90

(0.2)

Here o_1= 15, o_2=29, o_3=27 and o_4=19
r_1= first row total=15+29=44
r_2= second row total=27+29=46
c_1=first column total=15+27=42
c_2=second column total=29+19=48
Grand total=15+29+27+19=90=GT
e_1=expected frequency of $(1,1)$ cell=

$$\frac{r_1c_1}{GT} = \frac{44 \times 42}{90} = 20.5$$

e_2=expected frequency of $(1,2)$ cell=

$$\frac{r_1c_2}{GT} = \frac{44 \times 48}{90} = 23.5$$

e_3=expected frequency of $(2,1)$ cell=

$$\frac{r_2c_1}{GT} = \frac{46 \times 42}{90} = 21.5$$

e_4=expected frequency of $(2,2)$ cell=

$$\frac{r_2c_2}{GT} = \frac{46 \times 48}{90} = 24.5$$

Now

$$\chi^2 = \sum_{i=1}^{4} \frac{(o_i - e_i)^2}{e_i} = \frac{(15-20.5)^2}{20.5} + \frac{(29-23.5)^2}{23.5} + \frac{(27-21.5)^2}{21.5} + \frac{(19-24.5)^2}{24.5} = 5.47$$

Step 4: Now we have to determine the critical value ${\chi_\alpha}^2$ using the following equation $P(\chi^2 > {\chi_\alpha}^2) = \alpha = 0.01$ with $(r-1)(c-1) = (2-1)(2-1) = 1$ degrees of freedom (Use χ^2 distribution table)
$\Rightarrow {\chi_\alpha}^2 = 6.635$
Step 5: As here $\chi^2 = 5.47 < {\chi_\alpha}^2 = 6.635$,
we have Fail to reject null hypothesis H_0 .
So our conclusion is the time spent watching television is independent of whether the viewer is male or female.

Test for Homogeneity

Homogeneous means the same in structure or composition. This test gets its name from the null hypothesis, where we claim that the distribution of the responses are the same (homogeneous) across groups.
NOTE: PROCEDURE IS SAME AS TEST FOR INDEPENDENCE

Question No.93

To determine current attitudes about prayer in public schools, a survey was conducted in four Virginia counties. The following table gives the attitudes of 200 parents from Craig County, 150 parents from Giles County, 100 parents from Franklin County, and 100 parents from Montgomery County:

	County			
Attitude	Craig	Giles	Franklin	Mont.
Favor	65	66	40	34
Oppose	42	30	33	42
No opinion	93	54	27	24

(0.3)

Test for homogeneity of attitudes among the four counties concerning prayer in the public schools using significance level α=0.01.

Solution:

Step 1: significance level α=0.01
Step 2: we have to test the Null Hypothesis H_0 : The attitudes among the four countries are homogeneous against the alternative H_1 : The attitudes among the four countries are not homogeneous.

Observed and expected frequencies					
	County				
Attitude	Craig	Giles	Franklin	Montgomery	Total
Favor	65(74.5)	66(55.9)	40(37.3)	34(37.3)	$r_1 = 205$
Oppose	42(53.5)	30(40.1)	33(26.7)	42(26.7)	$r_2 = 147$
No Opinion	93(72.0)	54(54.0)	27(36.0)	24(36.0)	$r_3 = 198$
Total	$c_1 = 200$	$c_2 = 150$	$c_3 = 100$	$c_4 = 100$	$GT = 550$

(0.4)

Here o_1= 65, o_2=66, o_3=40 and o_4=34 o_5= 42, o_6=30, o_7=33 and o_8=42 o_9= 65, o_{10}=54, o_{11}=27 and o_{12}=24
r_1= first row total=65+66+40+34=205
r_2= second row total=42+30+33+42=147
r_3= third row total=93+54+27+24=198
c_1=first column total=65+42+93=200

c_2=second column total=66+30+54=150
c_3=third column total=40+33+27=100
c_4= fourth column total=34+42+24=100
Grand total=205+147+198=GT
e_1=expected frequency of $(1,1)$ cell=

$$\frac{r_1c_1}{GT}=\frac{205\times 200}{550}=74.5$$

e_2=expected frequency of $(1,2)$ cell=

$$\frac{r_1c_2}{GT}=\frac{205\times 150}{550}=55.9$$

e_3=expected frequency of $(1,3)$ cell=

$$\frac{r_1c_3}{GT}=\frac{205\times 100}{550}=37.3$$

e_4=expected frequency of $(1,4)$ cell=

$$\frac{r_1c_4}{GT}=\frac{205\times 100}{550}=37.3$$

e_5=expected frequency of $(2,1)$ cell=

$$\frac{r_2c_1}{GT}=\frac{147\times 200}{550}=53.5$$

e_6=expected frequency of $(2,2)$ cell=

$$\frac{r_2c_2}{GT}=\frac{147\times 150}{550}=40.1$$

e_7=expected frequency of $(2,3)$ cell=

$$\frac{r_2c_3}{GT}=\frac{147\times 100}{550}=26.7$$

e_8=expected frequency of $(2,4)$ cell=

$$\frac{r_2c_4}{GT}=\frac{147\times 100}{550}=26.7$$

e_9=expected frequency of $(3,1)$ cell=

$$\frac{r_3c_1}{GT}=\frac{198\times 200}{550}=72$$

e_{10}=expected frequency of $(3,2)$ cell=

$$\frac{r_3 c_2}{GT} = \frac{198 \times 150}{550} = 54$$

e_{11}=expected frequency of $(3,3)$ cell=

$$\frac{r_3 c_3}{GT} = \frac{198 \times 100}{550} = 36$$

e_{12}=expected frequency of $(3,4)$ cell=

$$\frac{r_3 c_4}{GT} = \frac{198 \times 100}{550} = 36$$

$$\chi^2 = \frac{(65-74.5)^2}{74.5} + \frac{(66-55.9)^2}{55.9} + \frac{(40-37.3)^2}{37.3} + \frac{(34-37.3)^2}{37.3} + \frac{(42-53.5)^2}{53.5} + \frac{(30-40.1)^2}{40.1}$$
$$+ \frac{(33-26.7)^2}{26.7} + \frac{(42-26.7)^2}{26.7} + \frac{(93-72)^2}{72} + \frac{(54-54)^2}{54} + \frac{(27-36)^2}{36} + \frac{(24-36.0)^2}{36.0} = 31.17$$

Step 4: Now we have to determine the critical value ${\chi_\alpha}^2$ using the following equation

$P(\chi^2 > {\chi_\alpha}^2) = \alpha = 0.01$ with $(r-1)(c-1) = (3-1)(4-1) = 6$ degrees of freedom (Use χ^2 distribution table)

$\Rightarrow {\chi_\alpha}^2 = 16.812$

Step 5 As here $\chi^2 = 31.17 > {\chi_\alpha}^2 = 16.812$, it satisfy critical region

so we have to reject null hypothesis H_0 .

So our conclusion is the attitudes among the four countries are not homogeneous.