(Lect. 37)
10.11 Goodness-of-Fit Test

Here we will discuss a test, which is based on how good a fit we have between the frequency of occurrence of observations in an observed sample and the expected frequencies obtained from the hypothesized distribution.
A goodness-of-fit test between observed and expected frequencies is based on the quantity.
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 $v = k - 1$ degrees of freedom.
The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the i^{th} cell.
k represents number of cell.
Step 4:
Determine critical value ${\chi_\alpha}^2$ using the following equation
$P(\chi^2 > {\chi_\alpha}^2) = \alpha$ with $k - 1$ 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 $k - 1$ degrees of freedom.
Critical region is $\chi^2 > {\chi_\alpha}^2$
Fail to reject region is $\chi^2 \leq {\chi_\alpha}^2$

Question No 80
The grades in a statistics course for a particular semester were as follows:

Grade	*A*	*B*	*C*	*D*	*F*
f	14	18	32	20	16

(0.1)

Test the hypothesis, at the 0.05 level of significance, that the distribution of grades is uniform.

Solution:
Step 1: Given $\alpha = 0.05$
Step 2:
we have to test the Null Hypothesis
H_0:Distribution of grades is uniform against the alternative hypothesis
H_1: Distribution of grades is not uniform

Step 3:

Here k=5= number of cells
o_i: represents the observed frequencies, $i = 1, 2, 3, 4, 5$
o_1=14, o_2=18, o_3=32, o_4=20, o_5=16
e_i: represent expected frequencies, $i = 1, 2, 3, 4, 5$

$$e_i = \frac{14 + 18 + 32 + 20 + 16}{5} = 20$$

$$\chi^2 = \sum_{i=1}^{5} \frac{(o_i - e_i)^2}{e_i} = \frac{(14-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(32-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(16-20)^2}{20} = 10$$

Step 4:

Determine critical value ${\chi_\alpha}^2$ using the following equation
$P(\chi^2 > {\chi_\alpha}^2) = 0.05$ with $k - 1 = 5 - 1 = 4$ degrees of freedom (Use χ^2 distribution table)
Now ${\chi_\alpha}^2 = {\chi_{0.05}}^2 = 9.488$

Step 5:
As here $\chi^2 = 10 > {\chi_\alpha}^2 = 9.488$.
So it satisfy critical region (reject region)($\chi^2 > {\chi_\alpha}^2$)
So we reject the null hypothesis and not to fail alternative hypothesis.
Our conclusion is the distribution of grades is not uniform.
Question No 83

A coin is thrown until a head occurs and the number X of tosses recorded. After repeating the experiment 256 times, we obtained the following results:

x	1	2	3	4	5	6	7	8
f	136	60	34	12	9	1	3	1

(0.2)

Test the hypothesis, at the 0.05 level of significance, that the observed distribution of X may be fitted by the geometric distribution $g(x; 1/2)$, $x = 1, 2, 3, \ldots$.
Solution:
Step 1:
Given $\alpha = 0.05$
Step 2:
We have to test the Null Hypothesis
H_0: Observed distribution of X is fitted by the geometric distribution
$g(x; 1/2)$, $x = 1, 2, 3, \ldots$ against the alternative hypothesis
H_1: Observed distribution of X is not fitted by the geometric distribution
$g(x; 1/2)$, $x = 1, 2, 3, \ldots$.

Also we represent $H_0 : f(x) = g(x; 1/2)$, $x = 1, 2, 3, \ldots$
and $H_1 : f(x) \neq g(x; 1/2)$, $x = 1, 2, 3, \ldots$.

Step 3:
We know the geometric distribution $g(x;p) = pq^{x-1}$
p=Probability of getting head $=\frac{1}{2}$
q=Probability of getting tail $=\frac{1}{2}$

$$g(x;1/2) = pq^{x-1} = \frac{1}{2}(\frac{1}{2})^{x-1} = \frac{1}{2^x}, x = 1,2,3,....$$

$$g(x;1/2) = \frac{1}{2^x}, x = 1,2,3,...$$

$$g(1;1/2) = \frac{1}{2^1} = \frac{1}{2}$$

$$g(2;1/2) = \frac{1}{2^2} = \frac{1}{4}$$

$$g(3;1/2) = \frac{1}{2^3} = \frac{1}{8}$$

$$g(4;1/2) = \frac{1}{2^4} = \frac{1}{16}$$

$$g(5;1/2) = \frac{1}{2^5} = \frac{1}{32}$$

$$g(6;1/2) = \frac{1}{2^6} = \frac{1}{64}$$

$$g(7;1/2) = \frac{1}{2^7} = \frac{1}{128}$$

$$g(8;1/2) = \frac{1}{2^8} = \frac{1}{248}$$

Here k=8= number of cells
o_i: the observed frequencies,$i = 1,2,3,4,5,6,7,8$
o_1=136, o_2=60, o_3=34, o_4=12, o_5=9, o_6=1, o_7=3, o_8=1
Total number of observed frequencies=$136 + 60 + 34 + 12 + 9 + 1 + 3 + 1 = 256$

Now we have to calculate all the expected frequencies, e_i, i=1,2,3,4,5,6,7,8

$$e_1 = 256(g(1; 1/2)) = 256(\frac{1}{2}) = 128$$

$$e_2 = 256(g(2; 1/2)) = 256(\frac{1}{4}) = 64$$

$$e_3 = 256(g(3; 1/2)) = 256(\frac{1}{8}) = 32$$

$$e_4 = 256(g(4; 1/2)) = 256(\frac{1}{16}) = 16$$

$$e_5 = 256(g(5; 1/2)) = 256(\frac{1}{32}) = 8$$

$$e_6 = 256(g(6; 1/2)) = 256(\frac{1}{64}) = 4$$

$$e_7 = 256(g(7; 1/2)) = 256(\frac{1}{128}) = 2$$

$$e_8 = 256(g(8; 1/2)) = 256(\frac{1}{248}) = 1$$

Now

$$\chi^2 = \sum_{i=1}^{8} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(136-128)^2}{128} + \frac{(60-64)^2}{64} + \frac{(34-32)^2}{32} + \frac{(12-16)^2}{16} + \frac{(9-8)^2}{8} + \frac{(1-4)^2}{4} + \frac{(3-2)^2}{2} + \frac{(1-1)^2}{1} = 3.125$$

Step 4: Determine critical value ${\chi_\alpha}^2$ using the following equation
$P(\chi^2 > {\chi_\alpha}^2) = 0.05$ with $k-1 = 8-1 = 7$ degrees of freedom (Use χ^2 distribution table) Now
${\chi_\alpha}^2 = {\chi_{0.05}}^2 = 14.067$
Step 5:
As here $\chi^2 = 3.125 < {\chi_\alpha}^2 = 14.067$.
So it satisfy fail to reject region ($\chi^2 \leq {\chi_\alpha}^2$)
So we Fail to reject the null hypothesis.
Our conclusion is $f(x) = g(x; 1/2)$.