IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

χ^2 Goodness of Fit Test

Dr. A. Ramesh

DEPARTMENT OF MANAGEMENT STUDIES

Agenda

- Python demo for testing GOF for Poisson distribution
- Understanding goodness of fit test for:
  - Uniform
  - Normal
- Python demo for testing GOF for uniform and normal distribution

Goodness of fit for Uniform Distribution

- Milk Sales Data

Month	Litres
January	1,610
February	1,585
March	1,649
April	1,590
May	1,540
June	1,397
July	1,410
August	1,350
September	1,495
October	1,564
November	1,602
December	1,655
	18,447

Hypotheses and Decision Rules

H_o: The monthly milk figures for milk sales are uniformly distributed

H_a: The monthly milk figures for milk sales are not uniformly distributed

$$\alpha = .01$$

$$\begin{aligned} df &= k - 1 - p \\ &= 12 - 1 - 0 \\ &= 11 \end{aligned}$$

$$\chi^2_{.01,11} = 24.725$$

If $\chi^2_{\text{Cal}} > 24.725$, reject H_o.

If $\chi^2_{\text{Cal}} \le 24.725$, do not reject H_o.

Python code

```
In [1]: from scipy.stats import chi2

In [2]: import pandas as pd
        import numpy as np

In [3]: chi2.ppf(0.99,11)

Out[3]: 24.724970311318277
```

Calculations

Month	f_o	f_e	$(f_o - f_e)^2/f_e$
January	1,610	1,537.25	3.44
February	1,585	1,537.25	1.48
March	1,649	1,537.25	8.12
April	1,590	1,537.25	1.81
May	1,540	1,537.25	0.00
June	1,397	1,537.25	12.80
July	1,410	1,537.25	10.53
August	1,350	1,537.25	22.81
September	1,495	1,537.25	1.16
October	1,564	1,537.25	0.47
November	1,602	1,537.25	2.73
December	1,655	1,537.25	9.02
	18,447	18,447.00	74.38

$$f_e = \frac{18447}{12}$$

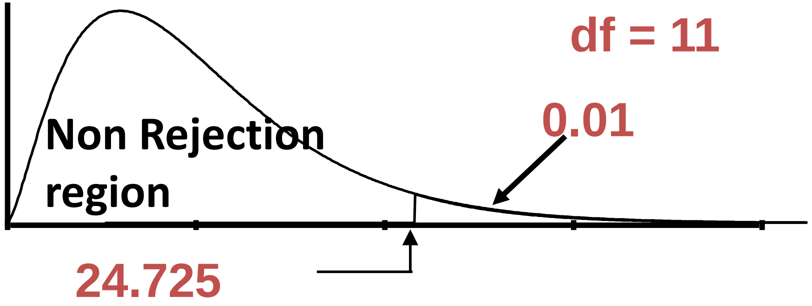
$$= 1537.25$$

$$\chi^2_{Cal} = 74.37$$

Python code

```
In [6]: x =[1610,1585,1649,1590,1540,1397,1410,1350,1495,1564,1602,1655]

In [7]: np.mean(x)

Out[7]: 1537.25

In [8]: exp_f=[1537.25,1537.25,1537.25,1537.25,1537.25,1537.25,1537.25,1537.25,1537.25,1537.25,1537.25,1537.25]

In [9]: from scipy.stats import chisquare
        chisquare(x,exp_f)

Out[9]: Power_divergenceResult(statistic=74.37583346885673, pvalue=1.78545252783034e-11)
```

Conclusion



$$\chi^2_{Cal} = 74.37 > 24.725, \text{ reject } H_o.$$

Goodness of Fit Test: Normal Distribution

1. Set up the null and alternative hypotheses.
2. Select a random sample and
    a. Compute the mean and standard deviation.
    b. Define intervals of values so that the expected frequency is at least 5 for each interval.
    c. For each interval record the observed frequencies
3. Compute the expected frequency, e_i , for each interval.

Goodness of Fit Test: Normal Distribution

4. Compute the value of the test statistic.

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

$k-1-p$

5. Reject H_0 if $\boxed{\chi^2 \geq \chi^2_\alpha}$

(where α is the significance level and there are k - 3 degrees of freedom)

Normal Distribution Goodness of Fit Test

- Example: IQL Computers

IQL Computers manufactures and sells a general purpose microcomputer. As part of a study to evaluate sales personnel, management wants to determine, at $\alpha = 0.05$ significance level, if the annual sales volume (number of units sold by a salesperson) follows a normal probability distribution.

Normal Distribution Goodness of Fit Test

A simple random sample of 30 of the salespeople was taken and their numbers of units sold are below.

33	43	44	45	52	52	56	58	63	64
64	65	66	68	70	72	73	73	74	75
83	84	85	86	91	92	94	98	102	105

(mean = 71, standard deviation = 18.23)

Python code

```
In [12]: A =[33, 43, 44, 45, 52, 52, 56, 58, 63, 64, 64, 65, 66, 68, 70, 72, 73, 73, 74, 75, 83, 84, 85, 86, 91, 92, 94, 98, 102, 105]

In [13]: mean = np.mean(A)
         mean

Out[13]: 71.0

In [14]: std = np.std(A)
         std

Out[14]: 18.226354544998845
```

Normal Distribution Goodness of Fit Test

- Hypotheses

H_0: The population of number of units sold has a normal distribution with mean 71 and standard deviation 18.23

H_a: The population of number of units sold does not have a normal distribution with mean 71 and standard deviation 18.23

Normal Distribution Goodness of Fit Test

- Interval Definition

To satisfy the requirement of an expected frequency of at least 5 in each interval we will divide the normal distribution into 30/5 = 6 equal probability intervals.

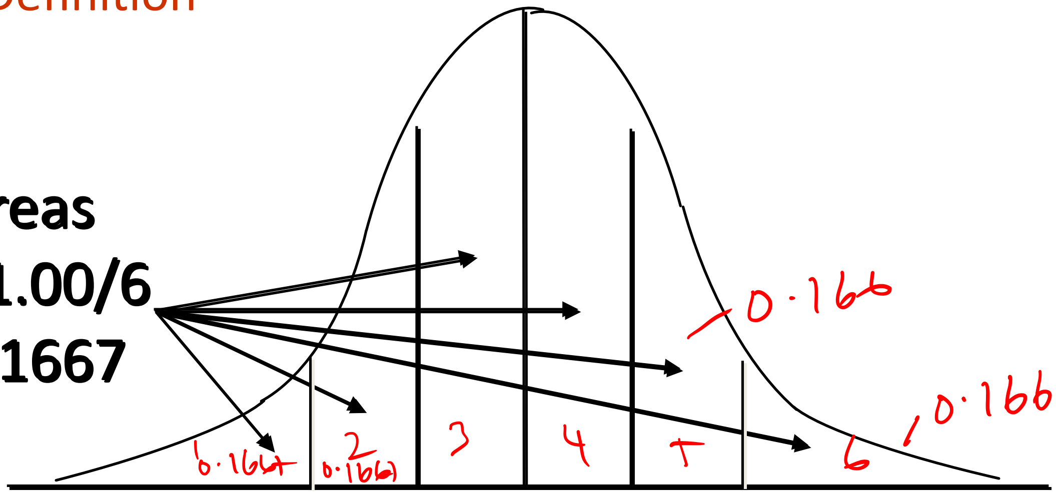
Normal Distribution Goodness of Fit Test

- Interval Definition

Areas

= 1.00/6

= .1667



$\bar{x} - 2\sigma$

53.367   **71**   **88.63 = 71 + .97(18.24)**

71 - .43(18.23) = 63.149   **78.85**

Python code

```
In [15]: x = 1/6 #for 6 equal probability intervals.

In [16]: for j in range(1,6):
             Prob_intervals = [scipy.stats.norm.ppf(j*x, mean, std)]
             print ( Prob_intervals)

[53.36743154175236]
[63.14941153083116]
[71.0]
[78.85058846916884]
[88.63256845824763]
```

Normal Distribution Goodness of Fit Test

- Observed and Expected Frequencies

i	f_i	e_i	$f_i - e_i$
Less than 53.02	6	5	1
53.02 to 63.03	3	5	-2
63.03 to 71.00	6	5	1
71.00 to 78.97	5	5	0
78.97 to 88.98	4	5	-1
More than 88.98	6	5	1
Total	30	30	

$\frac{30}{6} = 5$

Python code

```
In [17]: Expected_Freq = [5,5,5,5,5,5] #will divide the normal distribution into 6 intervals at frequency 5 in each

In [18]: Obs_f = [6,3,6,5,4,6]

In [19]: scipy.stats.chisquare(Obs_f, Expected_Freq)

Out[19]: Power_divergenceResult(statistic=1.5999999999999999, pvalue=0.9012493445012737)
```

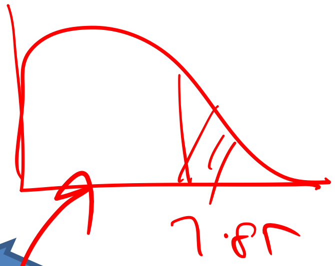
$\alpha = 5\%$

Normal Distribution Goodness of Fit Test

- Rejection Rule

With $\alpha = .05$ and $k - p - 1 = 6 - 2 - 1 = 3$ d.f.

(where k = number of categories and p = number of population parameters estimated), $\chi^2_{.05} = 7.815$




```
In [5]: chi2.ppf(0.95,3)
Out[5]: 7.8147279032511765
```

Reject H_0 if p-value $\leq .05$ or $\chi^2 \geq 7.815$.

- Test Statistic

$$\chi^2 = \frac{(1)^2}{5} + \frac{(-2)^2}{5} + \frac{(1)^2}{5} + \frac{(0)^2}{5} + \frac{(-1)^2}{5} + \frac{(1)^2}{5} = \boxed{1.600}$$

Thank you