Subject	Statistics
Year	1st Year B.Sc
Paper no	05
Paper Name	Probability Distributions-1
Topic no	26
Topic name	Fitting Standard Univariate Continuous Distributions such as Normal and Exponential Distributions
SME	Suguna Nayak
ID	Syeda Shafia Kausar

E-Learning Module on Fitting Standard Univariate Continuous Distributions such as Normal and Exponential Distributions

Learning Objectives

By the end of this session, you will be able to:

- Explain the fitting of exponential distribution when the parameter is known or unknown
- Explain the fitting of normal distribution when parameters are known and unknown

Fitting Normal Distribution

In order to fit normal distribution to the given data, we first calculate mean μ and standard deviation σ from the given data. Then, the normal curve fitted to the given data is given by,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; -\infty < x < \infty$$

To calculate the expected normal frequencies, we first find the standard normal variates corresponding to the lower limits of each of the class intervals, i.e. we compute $z=(x-\mu)/\sigma$, where x is the upper limit of the class interval.

Then, the areas under the normal curve to the left of the ordinate z, say $\Phi(z)=P(Z\leq z)$ are computed from the tables. Finally, the areas for the successive class intervals are obtained by subtraction $\Phi(z)- \Phi(z-1)$ and on multiplying these areas by N, the total frequency, we get the expected normal frequencies.

Fitting Exponential Distribution

In order to fit an exponential distribution to the given data, we first estimate the parameter of the distribution by its maximum m.l.e $\hat{\theta} = \frac{1}{\bar{x}}$

Then, the exponential curve fitted to the given data is given by, $f(x)=\theta e^{-x\theta}$, $x>0$.

To calculate expected frequencies, we first write the upper limits of each class interval. Then, the probability below the upper limit is calculated using the formula $P(X<x)=1-e^{-x\theta}$.

Exercise - 1

Fit an exponential distribution with parameter θ for the following data.

CI	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	130	60	34	12	9	3	2	0

Solution:

Given X~Exp(θ). Hence, pdf is given by, $f(x)=\theta e^{-x\theta}$, $-\infty<x<\infty$

Since parameter is not known, we estimate using its m.l.e $\theta=\frac{1}{\bar{x}}$.

CI	Frequency (f)	Mid-point (x)	fx
0-10	130	5	650
10-20	60	15	900
20-30	34	25	850
30-40	12	35	420
40-50	9	45	405
50-60	3	55	165
60-70	2	65	130
70-80	0	75	0
Total	**250**	**-**	**3520**

The mean of the distribution is given by,

$$\overline{X} = \frac{\sum fx}{N} = \frac{3520}{250} = 14.08$$

$$\therefore \theta = \frac{1}{\overline{x}} = \frac{1}{14.08} = 0.071$$

Hence, equation fitted to given data is

$f(x) = \theta e^{-x\theta} = 0.071 e^{-0.071x}$

CI	Upper limit	$F(x)=1-e^{-0.071x}$	$\Delta F(x)$	Expected Frequencies
0-10	10	0.508	0.508	127
10-20	20	0.758	0.250	62
20-30	30	0.881	0.123	31
30-40	40	0.942	0.060	15
40-50	50	0.971	0.030	7
50-60	60	0.986	0.015	4
60-70	70	0.993	0.007	2
70-80	80	0.997	0.004	1
80 and above	∞	1.000	0.003	1
Total				**250**

Exercise - 2

Fit an exponential distribution with mean 17 for the following data and obtain expected frequencies.

CI	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	22	13	5	5	3	2

Solution:

Here we need to fit an exponential distribution with mean 17 i.e. parameter 1/17. Hence, the equation of the exponential curve fitted to the given data is $f(x) = \frac{1}{17} e^{-x/17}; x > 0$

CI	Upper limit(x)	$F(x)=1-e^{-x/17}$	$\Delta F(x)$	Expected Frequency
0-10	10	0.445	0.445	22
10-20	20	0.692	0.247	12
20-30	30	0.829	0.137	7
30-40	40	0.905	0.076	4
40-50	50	0.947	0.042	2
50-60	60	0.971	0.023	1
60 and above	∞	1	0.029	2
Total				50

Exercise - 3

Obtain equation of a normal curve that may be fitted to the following data.

CI	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
Frequency	3	21	150	335	326	135	26	4

Also obtain the expected normal frequencies.

Solution:

For the given data, let us obtain mean and standard deviation, which are taken as mean and standard deviation of the normal distribution, which are given by,

$$\mu = \frac{\Sigma fx}{N} \ \& \ \sigma = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

CI	Frequency (f)	x	fx	fx^2
60-65	3	62.5	187.5	11718.75
65-70	21	67.5	1417.5	95681.25
70-75	150	72.5	10875	788437.5
75-80	335	77.5	25962.5	2012094
80-85	326	82.5	26895	2218838
85-90	135	87.5	11812.5	1033594
90-95	26	92.5	2405	222462.5
95-100	4	97.5	390	38025
Total	**1000**	**-**	**79945**	**6420850**

$$\mu = \frac{\Sigma fx}{N} = \frac{79945}{1000} = 79.945$$

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2} = \sqrt{\frac{6420850}{1000} - \left(\frac{79945}{1000}\right)^2} = 5.545$$

If the normal curve is fitted to the given data, then the equation is given as,

$$f(x) = \frac{1}{5.545\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-79.945}{5.545}\right)^2}$$

Now, consider the following table, where we calculate expected frequencies.

CI	Upper limit	Z=(x-79.945)/5.545	Φ(z)	ΔΦ(z)	Expected frequency
below 60	60	-3.60	0.0002	0.0002	0
60-65	65	-2.70	0.0030	0.0028	3
65-70	70	-1.79	0.0341	0.0311	31
70-75	75	-0.89	0.1819	0.1478	148
75-80	80	0.01	0.5040	0.3221	322
80-85	85	0.91	0.8233	0.3193	319
85-90	90	1.81	0.9674	0.1441	144
90-95	95	2.72	0.9971	0.0297	30
95-100	100	3.62	0.9999	0.0028	3
100 and above	∞	∞	1.0000	0.0001	0
Total					**1000**

Exercise - 4

Fit a Standard normal distribution for the following data.

CI	-5-(-3)	-3-(-1)	-1-1	1-3	3-5
Frequency	5	9	19	10	7

Solution:

Here we need to fit standard normal distribution. The standard normal curve fitted to the given data is, $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

CI	Upper limit (z)	Φ(z)	ΔΦ(z)	Expected Frequency
below -5	-5	0.000	0.000	0
-5-(-3)	-3	0.001	0.001	0
-3-(-1)	-1	0.159	0.158	8
-1-1	1	0.813	0.654	33
1-3	3	0.999	0.186	9
3-5	5	1.000	0.001	0
5 & above	∞	1.000	0.000	0
Total				50