Fitting Distributions

Fitting of a distribution to the data given. While fitting the distribution, theoretical frequency or expected frequency is determined.

1. The following data are the number of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. Fit a binomial distribution to these data.

x :	0	1	2	3	4	5	6	7	8	9	10
f:	6	20	28	12	8	6	0	0	0	0	0

Here n = 10 and $N = \sum f_i = 80$

$$Mean = \frac{\sum f_i x_i}{\sum f_i} = \frac{20+56+36+32+30}{80} = \frac{174}{80} = 2.175$$

Mean of the binomial distribution = np

np = 10p=2.175

therefore p=0.2175 , q=1- p = 0.7825

Hence the binomial distribution to be fitted is

$N(q+p)^n = 80(0.7825 + 0.2175)^{10}$

$$80.10_{C_0}(0.7825)^{10} + 80.10_{C_1}(0.7825)^9(0.2175)^1 + 80.10_{C_2}(0.7825)^8(0.2175)^2 +$$

$$80.10_{C_3}(0.7825)^7(0.2175)^3 + 80.10_{C_4}(0.7825)^6(0.2175)^4 +$$

$$80.10_{C_5}(0.7825)^5(0.2175)^5 +$$

$$80.10_{C_6}(0.7825)^4(0.2175)^6 + 80.10_{C_7}(0.7825)^3(0.2175)^7 +$$

$$80.10_{C_8}(0.7825)^2(0.2175)^8 +$$

$$80.10_{C_9}(0.7825)^1(0.2175)^9 + 80.10_{C_{10}}(0.2175)^{10}$$

$= 6.895 + 19.13 + 23.94 + ... + 0.0007 + 0.00002$

Therefore the successive terms in the expansion give the expected or theoretical frequencies which are

x:	0	1	2	3	4	5	6	7	8	9	10
f:	6.9	19.1	24	17.8	8.6	2.9	0.7	0.1	0	0	0

2. The mistakes committed by a typist follow a Poisson distribution as given below :

x : Number of mistakes per page	: 0	1	2	3	4	5
f: Number of pages	: 142	156	69	27	5	1

$$\overline{x} = \frac{\Sigma f x}{\Sigma f} = \frac{156 + 138 + 81 + 20 + 5}{400} = 1$$

$\lambda = 1$ $\quad\quad\therefore\ e^{-\lambda} = e^{-1} = 0.3679$

Expected frequencies :

$$N\ (x = 0) = 400 \times 0.3679 = 147.16$$

$$N\ (x = 1) = 400 \times \lambda = 147.16 = 147.16$$

$$N\ (x = 2) = \frac{147.16\ \times\ \lambda}{2} = 73.58$$

$$N\ (x = 3) = \frac{73.58 \times \lambda}{3} = 24.53$$

$$N\ (x = 4) = \frac{24.53 \times \lambda}{4} = 6.13$$

$$N\ (x = 5) = \frac{6.13 \times \lambda}{5} = 1.23$$

Converting to whole numbers, the expected frequencies are given as follows:

x:	0	1	2	3	4	5
f :	147	147	74	25	6	1

3. Obtain the equation of the normal probability curve that can be fitted to the following data and the theoretical frequencies:

Variable	4	6	8	10	12	14	16	18	20	22	24
Frequency	1	7	15	22	35	43	38	20	13	5	1

Variable x	4	6	8	10	12	14	16	18	20	22	24	Total
Frequency f	1	7	15	22	35	43	38	20	13	5	1	200
X=(x – 4)/2	-5	-4	-3	-2	-1	0	1	2	3	4	5	
fX	-5	-28	-45	-44	-35	0	38	40	39	20	5	-15
fX^2	25	112	135	88	35	0	38	80	117	80	25	735

$$Mean = \mu = 14 + 2\left(\frac{-15}{200}\right) = 13.85$$

$$Variance = \sigma^2 = 2^2\left[\frac{735}{200} - \left(\frac{-15}{200}\right)^2\right] = 4 \times 3.6694$$

$\therefore$ The standard deviation = σ = 3.83

The equation of the normal probability curve is

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{3.83\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-13.85}{3.83}\right)^2}$$

x	$t = \left(\frac{x-13.85}{3.83}\right)$	$\phi = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$	$Y = \frac{200}{3.83}\phi$	*Theoretical frequency* $= Y \times$ *class width* $= 2Y$
4	-2.5710	0.01466	0.7656	2
6	-2.0488	0.04892	2.5530	5
8	-1.5266	0.12445	6.4975	13
10	-1.0044	0.24094	12.5800	25
12	-0.4822	0.35513	18.5430	37
14	0.0400	0.39860	20.8200	42
16	0.5622	0.34058	17.7860	36
18	1.0844	0.22364	11.6800	23
20	1.6066	0.10978	5.7204	12
22	2.1288	0.04141	2.1620	4
24	2.6510	1.01187	0.6198	1

$$Y = = \frac{N}{3.83\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-13.85}{3.83}\right)^2} = \frac{N}{\sigma}\phi(t) = \frac{200}{3.83}\phi$$

Y is the ordinate corresponding to the value *x*.
$\therefore$ The frequency corresponding to x is the total frequency over the interval (x – 1, x+1) whose length is 2. This is equal to 2Y.

The values of x increases by 2 from 4 to 24.

$\therefore$ The values of t increases by $\frac{2}{3.83} = 0.5222$ from -2.5710 to 2.6510.

The frequency corresponding to x = 20 with maximum fractional is rounded to 12 in order to get the total frequency, 200