Unit 15: Comparison of observed frequency distribution with fitted frequency distribution

Meaning of the terms ‘theoretical distribution’, ‘model’, ‘observed frequencies’ and ‘fitted or expected frequencies’ should be given as the first insight. Illustration is better done by using a simple example, e.g. rolling a fair die several times.

Idea of fitting a theoretical distribution to a set of observed data should be taught in an intuitive way. The procedures of fitting should be given in details, especially the way of calculating the fitted (expected) frequencies.

Considerations should be limited to binomial and Poisson distribution for discrete cases and normal distribution for continuous cases. In each case, students should be able to distinguish whether the parameters of the distribution are given or not. If the parameters are not given, students should know how to estimate the values of the parameters and then use them in further calculations. Examples for each case involving different distributions should be given. These examples should also include comparison of two or more fitted distributions to a specific set of observed data.

Descriptive treatment for goodness-of-fit is expected. The comparison of observed frequencies and fitted frequencies should be done by inspection or by graphs only. How the differences between the observed frequencies and the expected frequencies determine the fitness of the model should be explained to the students. The reason behind should be clarified but formal statistical test is not required.

Students are expected to know how to make prediction and draw conclusion from the fitted distribution.

<u>Example 1</u>

Eggs are packed in boxes. Each box contains 10 eggs. Table 1 shows the frequency distribution of rotten eggs in each of the 120 randomly selected boxes.

(i) It is suggested that the number of the rotten eggs could be modelled by a binomial distribution with the probability that an eggs is rotten being 0.15. Fill in the missing expected frequencies, in the third column of Table 1.

(ii) A buyer claims that the number of rotten eggs could be approximated by a Poisson distribution with mean λ. He has calculated some expected frequencies as shown in Table 1.
   (a) Find λ, correct to 1 decimal place.
   (b) Fill in the missing expected frequencies, in the fourth column of Table 1.

(iii) The buyer compares the two distributions in (i) and (ii) and adopts the one which fits the observed data better. He classifies a box of eggs as *good* if there are less than 2 rotten eggs in the box. He buys 5 boxes of eggs.
   (a) Find the probability that at least 4 of the these boxes are *good.*
   (b) Suppose that at least 4 of these 5 boxes are *good* and the buyer buys 5 more boxes. Find the probability that exactly 9 of these 10 boxes are *good.*

Table 1 Observed and expected frequencies of rotten eggs in each of 120 randomly selected boxes.

Number of Rotten Eggs	Observed Frequency	Expected Frequency *	
		Binomial	Poisson
0	35	23.6	32.7
1	42	41.7	
2	28		27.6
3	11	15.6	
4	3	4.8	
5	1	1.0	
6	0		
7 or more	0	0	

* correct to 1 decimal place

<u>Example 2</u>

Two machines X and Y fill boxes with detergent powder. Machine X is set to fill each box with μ_0 kg of the powder. The net weight varies according to a normal distribution, with a standard deviation of 0.15 kg . Machine Y is set to fill each box with μ_1 kg of the powder. The net weight also varies according to a normal distribution, but with a standard deviation of 0.1 kg.

(i) It is desired that no more than 2.5% of the boxes filled by each machine should weigh less than 2.854 kg net. Determine the minimum values of μ_0 and μ_1.

(ii) The minimum values of μ_0 and μ_1 determined in (i) are used for production. A shipment of the product was filled by one and the same machine, but it is not known whether the machine was X or Y. In order to determine the source of the shipment, a random sample of 50 boxes of detergent powder was taken and their weights in kg were recorded as follows:

Table 2

Weight in kg	Observed Frequency	Expected Frequency	
		Machine X	Machine Y
2.7 – 2.8	1	0.4	
2.8 – 2.9	4	2.0	
2.9 – 3.0	10		
3.0 – 3.1	17		19.2
3.1 – 3.2	13		12.1
3.2 – 3.3	4		3.0
3.3 – 3.4	1		0.3

(a) Fill in the expected frequencies in Table 2.

(b) Compare the observed and the two expected frequency distributions by drawing histograms. Hence determine the source of the shipment.