Chapter-4

Analysis of Data (Computations)

§ 4.1 Introduction:

After the data collected have been scrutinized, the next task completed is the analysis of the scrutinized data. The analysis of data mostly, in this study, consists of numerical computations covering computations of estimated values / projected values, computations of the values of test statistics in testing of the acceptance of the estimated values / projected values that are obtained and interpretations of the results / findings. The interpretations of the results / findings have been presented in *Chapter*-5 while the results / findings obtained in the study have been presented in *Chapter*-6. Here, the steps in the numerical computations have been outlined. In this study, the associated data are of two categories viz,

(i) Data on population

(ii) Data on Temperature and Rainfall

respectively. Accordingly, the steps in the numerical computations have been outlined separately for these two categories of data.

§ 4.2 Computations and Results in Population Projection:

Computations in population projection have been done in the following broad steps:

Step-4.2.1

The data, collected in the study, relate to

(i) total population (i.e. total number of persons)

and (ii) number of persons age-sex wise

of India and of its states since the year 1901 to the year 2001. For the interest of maintaining consistency in data, the same have been retained for the years 1951, 1961, 1971, 1981, 1991 & 2001. Also due to the same reason, the age groups viz.

0 – 10, 10 – 20, 20 – 30, 40 – 50, 50 – 60 etc.

have been chosen in the classification of the number of persons into age and sex.

Step-4.2.2

In this step formula given by equation (2.7.6) has been applied to obtain stable value of the relative frequency corresponding to each of the states of India. The stable values, obtained, have been retained as the values of the corresponding probabilities $P(E_i)$. The values of $P(E_i)$ so obtained have been shown in **Table-6.1**.

Step-4.2.3

Using the values, obtained in Step-4.2.2, total population of each the states of India corresponding to each of the years

1961, 1971, 1981, 1991 & 2001

has been estimated by the formula given by (2.7.4)

Step-4.2.4

In order to assess the acceptability of the estimates, the significance of the difference between the observed figures and the corresponding estimated figures has been tested for each year separately using the test statistic given by (2.12.1) as discussed in § 2.12 viz.

$$\chi^2 = [\sum_{i=1}^{n} \{(O_i - E_i)^2\}/E_i],$$

$$(\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i)$$

(where O_i, (i = 1, 2, 3, ..., n) is a set of observed (experimental) frequencies and E_i, (i = 1, 2, 3, ..., n) is the set of the corresponding expected (theoretical or hypothetical) frequencies)

follows chi-square (χ^2) distribution with (n-1) *d.f.* (***Ref.*** 91, 92 & 127).

It has been found that the said difference for each year is insignificant if the total population is considered in thousand or in higher units. Therefore, the estimated figures have been retained in the unit of thousand. The estimated total populations (in thousand) of the states, obtained, have been shown in **Table-6.2**.

Step-4.2.5

It has been required to know for how far future projections on the total populations mentioned above can be computed so that the computed projected values are acceptable in the sense of statistical philosophy. For this purpose, total population of each *the states of India corresponding to each of the years*

1961, 1971, 1981, 1991 & 2001

has been estimated by the formula (2.7.4) treating the value of the relative frequency corresponding to the year 1961 as the value of $P(E_i)$ and chi square test of **goodness of fit** have been applied. The same has been done treating the value of the relative frequency corresponding to each of the remaining years

1971, 1981 & 1991

separately. It has been found that projected values with the sense of statistical acceptance can be obtained for next thirty years from the period up to which data are available Therefore, it has been decided to compute the projected values for the years 2011, 2021 & 2031.

<u>**Step-4.2.6**</u>

In order to obtain projection on total population of India, logistic curve formulated by *Pearl* and *Reed* described by equation (2.11.1) viz.

$$N(R:t) = \frac{L}{1 + \exp\{r.(\beta - t)\}}$$

(L, r and β are the parameters of the curve to be estimated on the basis of observed data) where

(iv) $N(R:t)$ is the total population of the region R under study at time t,

(v) L is the upper limit of $N(R:t)$,

(vi) β is the value of t for which $N(R:t)$ is $L/2$

and (iv) r is the value of

$$\frac{1}{N(R:t)} \quad \frac{d}{dt} N(R:t)$$

when $N(R:t) \to L$.

has been considered and attempt has been made to estimate the parameters L, r and β. For this purpose, the two constants A and B where

$$A = \{1 - \exp(r)\}$$

$$\& \; B = \exp(r)$$

have been estimated by the equations (2.10.3) and (2.10.4) as discussed in § 2.10 viz.

$$B = [\{\sum_{t=1}^{n-1} (Y_t - \bar{Y})^2\} / \sum_{t=1}^{n-1} (X_t - \bar{X})^2\}]^{1/2}$$

$$\& \; A = \bar{Y} - B.\bar{X}$$

where $Y_t = 1 / N(R:t)$

$$\& \ X_t = 1 / N(R : t-1) .$$

and then attempt has been made to estimate the parameters L and r by the equation

$$A = \{1 - \exp(r)\}$$

$$\& \ B = \exp(r).$$

Next, attempt has been made to estimate the parameter β by the equation (2.11.5) viz.

$$\beta = (1/nr). \sum_{t=1}^{n-1} \log(z_t - 1) + (n-1)/2$$

where $\quad z_t = L/N_t .$

However, it has been found that one of the three parameters viz. β cannot be estimated on the basis of the total populations of India observed.

Step-4.2.7

In this step, the exponential curve given by equation (2.11.6) viz.

$$N(R : t) = \mu.\exp(-\lambda.t),$$

$$\mu \geq 0 , \ \lambda \geq 0$$

(where μ and λ are the parameters which are to be determined on the basis of the observed data)

has been considered and attempt has been made to estimate the parameters μ and. For this purpose, the two constants v and λ where

$$v = \log \mu$$

have been estimated by the equations (2.11.8) and (2.11.9) as discussed in § 2.11 viz.

$$\sum_{t=1}^{n} y_t = n v - \lambda \sum_{t=1}^{n} t$$

$$\& \ \sum_{t=1}^{n} t y_t = v \sum_{t=1}^{n} t - \lambda \sum_{t=1}^{n} t^2$$

where $\quad y_t = \log N(R : t)$

$$\& \ y_t = \log N_t$$

Here, (t , N_t) , $(t = 1, 2 , \ldots\ldots, n)$ are the observed data on $\{t , N(R : t)\}$ i.e. the total population of India.

Next, an estimate of the parameter μ has been computed by equation (2.11 10) viz.

$$v = \log \mu$$

Then, the projected values on the total population of India have been computed for the years

2011, 2021 & 2031.

Step-4.2.8

In this step, a set of underestimates and another set of overestimates on the total population of India for the years

2011, 2021 & 2031

have been computed by the formulae given in the techniques described in § 2.9. Then, the projected interval on the total population of India for each of these years have been determined by selecting the maximum of the underestimates as the lower bound and the minimum of the overestimates as the upper bound of the corresponding interval. The projected intervals on total populations (in thousand) of India, obtained, have been shown in **Table-6.3**.

Step-4.2.9

The projected point values computed, by the exponential law, in Step-4.2.7 have been compared with the corresponding projected intervals computed in Step-4.2.8. It has been found that many of the projected figures computed by the exponential law fall outside the corresponding projected intervals. Therefore, the projected figures obtained by the exponential law have been rejected.

Step-4.2.10

In this step, to determine projections on total population of India for the years

2011, 2021 & 2031

the formulation given in § 2.8 has been considered. In order to determine n, the number of subintervals in an interval, the values of d and a given by (2.8.15) and (2.8.16) viz.

$$d = \Delta^2 N(t_{m-2})/n^2$$

$$\text{and} \quad a = \frac{2.n.\Delta N(t_{m-1}) - (n-1)\Delta^2 N(t_{m-2})}{2.n^2}$$

respectively have been computed for the arbitrary values

2, 3, 4, 5, 6, 7, ...

of n. The fractional parts of the values of a and d, obtained after computation, have been found to be minimum for

$$m = 56.$$

Therefore, the number '56' has been taken as the required estimated value of n. Using this estimated value of n and the corresponding values of a and d, projected total population of India for the years

2011, 2021 & 2031

have been computed by the formula (2.7.17) described in § 2.7 viz.

$$N(t_{m+k}) = N(t_m) + [2(a + 2md) + (n - 1)d].(n/2)$$

Step-4.2.11

The projected point values computed, by the method described in § 2.8, in Step-4.2.10 have been compared with the corresponding projected intervals computed in Step-4.2.8. It has been found that the projected point values computed by this method fall within the corresponding projected intervals. Therefore, the projected figures obtained by this method have been considered for acceptance.

Step-4.2.12

In this step, projected total population of each of the states in India for the years

2011, 2021 & 2031

have been computed by the formula for $N(E: t)$ given by equation (2.7.4) using the values of $P(E: t)$ obtained in Step-4.2.2.

The projected total populations (in thousand) of India and of its states, obtained in Step-4.2.10 and Step-12, have been shown in **Table-6.4**.

Step-4.2.13

In order to determine point values on the projected total population of India, the method described in § 2.10 has been applied. For this purpose, the formula described by equation (2.10.4) has been applied upon the underestimates and the overestimates taking together for each of the years

2011, 2021 & 2031

obtained in Step-4.2.8 in order to obtain the first approximation of each of the point values on the projected total population of India. The said formula has been applied repeatedly upon the approximations, obtained, in order to obtain the stabilized value of the approximations. Finally, the stabilized values obtained have been retained as the point values on the projected total population of India for each of the years 2011, 2021 & 2031respectively.

Step-4.2.14

The projected point values computed, by the method described in § 2.10, in Step-4.2.13 have been compared with the corresponding projected intervals computed in Step-4.2.8. It has been found that the projected point values computed by this method fall within the corresponding projected intervals. Therefore, the projected figures obtained by this method have been considered for acceptance.

Step-4.2.15

In this step, projected total population of the states in India for the years

2011, 2021 & 2031

have been computed by the formula for $N(E: t)$ given by (2.7.4) using the projected total populations of India obtained in Step-4.2.14 and the values of $P(E: t)$ obtained in Step-4.2.2.

The projected total populations (in thousand) of India and of its states, obtained in Step-4.2.14 and Step-4.2.15, have been shown in **Table-6.5**.

Step-4.2.16

In this step, the values of $P(S_i : g_i)$ where

$$P(S_i : g_i) = P(\text{a person belongs to } S_i \text{ sex and } g_i \text{ age group})$$

have been computed for each state by the same method as in Step-4.2.2 (treating the state as the whole region). Then projections on number of persons with respect to age and sex have been computed for the years

2011, 2021 & 2031.

applying the same method as in Step-4.2.13. The projected populations (in thousand), obtained in this step, have been shown in **Table-6.6**.

Step-4.2.17

Projected intervals on total populations (in thousand) have also been computed for the states of India by applying the method described in § 2.9. These projected intervals have been shown in **Table-6.3**.

§ 4.3 Computations and Results in Forecasting on Temperature and Rainfall:

Computations in projection on temperature and rainfall have been done in the following broad steps:

Step-4.3.1

Collected data on

(4) Mean Maximum Temperature,

(5) Highest Maximum Temperature,

(6) Mean Minimum Temperature,

(4) Lowest Minimum Temperature,

(7) Number of Rainy Days in the Month

and (6) Heaviest 24 Hours Rainfall in the Month

for each of the stations have been arranged in two-way table viz. month wise and year wise. On the other hand, the collected data on monthly total rainfall have been converted to yearly total rainfall and arranged in one-way table (year wise).

Step-4.3.2

In this step, the two parameters viz. μ and σ have been evaluated for each of the characteristics viz. Mean Maximum Temperature, Highest Maximum Temperature, Mean Minimum Temperature, Lowest Minimum Temperature, Number of Rainy Days (Monthly) & Heaviest 24 Hours Rainfall (Monthly) corresponding to each of the given twelve months and at each of the stations under study. The parameters μ and σ for a specified characteristic corresponding to a specified month and at a specified station are nothing but the population mean and population variance respectively of the specified characteristic corresponding to the specified month and at the specified station. The collected data on each of the characteristics corresponding to each of the given months and at each of the stations under study can be treated to be the corresponding whole-population data if the picture for the period from 1969 to 2002 is to be studied. Thus, the mean and the variance of the corresponding data will be the value of corresponding μ and σ respectively. Therefore, μ and σ have been evaluated for each of the characteristics viz. Mean Maximum Temperature, Highest Maximum Temperature, Mean Minimum Temperature, Lowest Minimum Temperature, Number of Rainy Days (Monthly) & Heaviest 24 Hours Rainfall (Monthly) corresponding to each of the given twelve months and at each of the stations under study computing the mean and the variance of the corresponding data.

Step-4.3.3

In this step, 2.58σ limits and 3σ limits of each of Mean Maximum Temperature, Highest Maximum Temperature, Mean Minimum Temperature, Lowest Minimum Temperature, Number of Rainy Days (Monthly) and Heaviest 24 Hours Rainfall (Monthly) corresponding to each of the given twelve months and at each of the stations under study have been computed by the formula

(i) $(\mu - 2.58\,\sigma\ ,\ \mu + 2.58\,\sigma)$

& (ii) $(\mu - 3\,\sigma\ ,\ \mu + 3\,\sigma)$

respectively.

Step-4.3.4

The intervals computed in Step-4.3.3 have been observed. Only a few intervals have been found where hardly one of the corresponding observations falls outside the limits.

Step-4.3.5

The intervals mentioned in Step-4.3.4 have been computed, by the same method as computed in Step-4.3.3, afresh for those cases that contain outliers excluding the corresponding outliers. The other intervals have been retained as it is.

Step-4.3.6

The intervals obtained in Step-4.3.5 have been accepted as the 99% confidence and the 99.73% confidence intervals (shown in **Table-6.7**) for the Characteristics, mentioned in Step-4.3.3, under study.

Step-4.3.7

In this step, analysis of variance has been carried out for each of Mean Maximum Temperature, Highest Maximum Temperature, Mean Minimum Temperature, Lowest Minimum Temperature, Number of Rainy Days (Monthly) and Heaviest 24 Hours Rainfall (Monthly) and for each station. The findings of Analysis of variance (applying F test) have been shown in **Table-6.9** while the findings of Analysis of variance (applying χ^2 test) have been shown in **Table-6.10**.

Step-4.3.8

In this step, ***t* test**, described in § 2.14.1, has been carried out for each pair of years where the difference between the years has been found to be significant by analysis of variance in the case of each of Mean Maximum Temperature, Highest Maximum Temperature, Mean Minimum Temperature, Lowest Minimum Temperature, Number of

Rainy Days (Monthly) and Heaviest 24 Hours Rainfall (Monthly) at each station. The difference between the years in case of only one pair of years, in each of the cases, has been found to be significant on applying *t* test. Thus, the difference between the effects of the years, in each of the cases, has been found to be insignificant overall.

Step-4.3.9

In this step, the pairs of the years where the difference between the years have been found to be significant on applying *t* test in Step-4.3.8 have been compared with the outliers described in Step-4.3.4 & Step-4.3.5. It has been found that the pairs of the years where the difference between the years have been found to be significant on applying *t* test are nothing but the outliers obtained in Step-4.3.4. These findings make the 99% confidence and the 99.73% confidence intervals, obtained in Step-4.3.6, more confident in terms of their acceptance.

Step-4.3.10

In this step, 2.58σ limits and 3σ limits of Yearly Total Rainfall have been computed for each station by applying the same method as in Step-4.3.3. Then these limits have been observed and modified in a similar manner as in Step-4.3.5 to obtain 99% confidence intervals and 99.73% confidence intervals (shown in **Table-6.8**) for yearly total rainfall at the stations under study.