## DISCOVERY AND DUPLICATION OF ANTIBIOTICS : A CORRELATION

In another note[1] we have reported the fitting of a modified Poisson Distribution to the frequency of duplication of discovery of antibiotics. The present note reports on a linear relation between the number of discoveries and the number of duplications of discovery of antibiotics from micro-organisms over the thirty-year period 1937-66.

The source of data on the report of discovery of antibiotics is a recent compilation by G. S. Rao and others.[2] Table I gives year-wise data on the number of reported discoveries of antibiotics in year $t$ ($X_t$), the number of new discoveries of antibiotics in year $t$ ($Y_t$), and the number of duplications in year $t$ ($Z_t$). In this note the discovery of an antibiotic is considered a duplication if it is identified with another antibiotic and so reported in published documents, irrespective of whether the compounds were produced by one and the same or different species of micro-organisms.

The data given in Table I indicate that $X_t$, $Y_t$, and $Z_t$ generally increase with $t$. A regression analysis was done to examine the relationship between the three variables. Three scatter diagrams were plotted for ($X_t$, $Y_t$), ($Z_t$, $Y_t$), and ($X_t$, $Z_t$) respectively. The diagrams indicated the existence of a linear relation between the respective pairs of variables. To examine this, the following linear relations were assumed :

$$Y_t = \alpha_0 + \beta_0 X_t$$
$$Y_t = \alpha_1 + \beta_1 Z_t$$
$$Z_t = \alpha_2 + \beta_2 Y_t.$$

Least square estimates of the constants $\alpha_0$, $\alpha_1$, $\alpha_2$, $\beta_0$, $\beta_1$, $\beta_2$ were obtained using the data given in Table I.

**Table I**

*Year-wise data on reports of discovery, new discovery and duplication*

| Serial number | Year | $X_t$ | $Y_t$ | $Z_t$ |
|---|---|---|---|---|
| 1 | 1937 | 2 | 1 | 1 |
| 2 | 38 | 3 | 2 | 1 |
| 3 | 39 | 8 | 8 | 0 |
| 4 | 40 | 5 | 2 | 3 |
| 5 | 41 | 3 | 3 | 0 |
| 6 | 42 | 15 | 7 | 8 |
| 7 | 43 | 10 | 6 | 4 |
| 8 | 44 | 20 | 13 | 7 |
| 9 | 45 | 25 | 17 | 8 |
| 10 | 46 | 27 | 25 | 2 |
| 11 | 47 | 54 | 35 | 19 |
| 12 | 48 | 44 | 36 | 8 |
| 13 | 49 | 51 | 35 | 16 |
| 14 | 50 | 42 | 30 | 12 |
| 15 | 51 | 53 | 37 | 16 |
| 16 | 52 | 58 | 46 | 12 |
| 17 | 53 | 68 | 54 | 14 |
| 18 | 54 | 84 | 62 | 22 |
| 19 | 55 | 82 | 60 | 22 |
| 20 | 56 | 72 | 48 | 24 |
| 21 | 57 | 119 | 91 | 28 |
| 22 | 58 | 79 | 57 | 22 |
| 23 | 59 | 95 | 70 | 25 |
| 24 | 60 | 89 | 69 | 20 |
| 25 | 61 | 62 | 45 | 17 |
| 26 | 62 | 93 | 67 | 28 |
| 27 | 63 | 107 | 77 | 30 |
| 28 | 64 | 100 | 80 | 20 |
| 29 | 65 | 83 | 60 | 23 |
| 30 | 66 | 77 | 60 | 17 |
| $n = 30$; | | $\Sigma X_t = 1{,}630$; | $\Sigma Y_t = 1{,}208$; | $\Sigma Z_t = 427$ |

The estimated equations were as follows:

$$Y_t = -0 \cdot 8075 + 0 \cdot 7529\, X_t \quad (1)$$

$$Y_t = 2 \cdot 084 + 2 \cdot 6711\, Z_t \quad (2)$$

$$Z_t = 0 \cdot 0814 + 0 \cdot 2671\, Y_t . \quad (3)$$

Using (1), (2) and (3) the residuals were computed in each case. The following values were obtained:

$$\Sigma (Y_t - \hat{Y}_t) = 0 \cdot 008$$

$$\Sigma (Y_t - \hat{Y}_t) = 0 \cdot 008$$

$$\Sigma (Z_t - \hat{Z}_t) = 0 \cdot 002.$$

This confirmed the fit of a straight line for the linear relation between the pairs of variables concerned

Analysis of variance was used to test the Null Hypothesis

$$H_{0i} = \beta_1 = 0 \quad \text{for} \quad 1, 2, 3, \quad (4)$$

The F values obtained in the ANOVA for (1), (2) and (3) were as follows:

| For equation | F | |
|---|---|---|
| | Calculated | Expected (1%)[3] |
| (1) | 2,539·26 | 7·64 |
| (2) | 144·96 | 7·64 |
| (3) | 273·75 | 7·64 |

Thus, in each of the cases the Null Hypothesis (4) was rejected, indicating a strong relationship between the pairs of variables. The relationship, within the observed range of variation, can be used for prediction of the discovery and duplication of antibiotics.

Documentation Res. and Training Centre,
Indian Statistical Institute,
112 Cross Road 11,
Bangalore-3, India, *July* 6, 1970.

A. Neelameghan.
V. V. Buche.
B. S. S. Gupta.

1. Neelameghan, A., Buche, V. V. and Gupta, B. S. S., "Distribution pattern of the duplication of discoveries," *Nature*, 1970, **227**, 1060.
2. Rao, G. S., Bannur, B. B. and Purandare, G. M., "Index to antibiotics producing micro-organisms," *Hindustan Antibiotics Bulletin*, August, 1967, **10** (1), 1.
3. Rao, C. R., Mitra, S. K. and Matthai, A., *Formulæ and Tables for Statistical Work*, Calcutta, Statistical Publishing Society, 1966.