# matters arising

## Do molecular biologists come of age in Aries?

WINDSOR has concluded that, "More molecular biologists were born under the sign of Aries than any other sign. More taxonomists were born under the sign of Cancer than any other sign and relatively few were born under Scorpio." Lest acceptance of these findings induces prospective parents to time the birth of their offspring to maximise their chances of becoming molecular or taxonomic biologists, a modest statistical analysis of Windsor's data seems in order.

Table 1 contains the frequencies, arranged by Sun sign and biological bent, of the birthdates of 812 biologists in the sample observed by Windsor[1]. The relative frequencies of births for the general population of the United States for 1934 by Sun sign were estimated by adding the products of the monthly relative frequencies by the respective fraction of each of the two months contributing to a given Sun sign interval as given by Parker and Parker[2].

In the $\chi^2$ tests for goodness of fit[3] to the data in Table 1, the critical value is $\chi^2_{0.05,11} = 19.675$ if $P \leq 0.05$ is the significance level to reject the null hypothesis, $H_0$: that each sample has the same frequencies as the general population. Since $0.75 \leq P(\chi^2 \geq 7.282) \leq 0.90$ for the taxonomists, $0.10 \leq P(\chi^2 \geq 16.368) \leq 0.25$ for the molecular biologists, and $0.25 \leq P(\chi^2 \geq 11.580) \leq 0.50$ for the combined sample of 812 biologists, the null hypotheses cannot be rejected. That the samples of taxonomists and molecular biologists could have come from the same population is confirmed by the heterogeneity $\chi^2$ test: $0.25 \leq P(\chi^2 \geq 12.070) \leq 0.50$.

As might be expected, the same conclusion that the biological discipline of the 812 scientists is independent of the Sun sign of birth is borne out by a 2×12 contingency table[3,4] which involves no assumptions about the birth frequencies per Sun sign for the general population. In this case, $0.25 \leq P(\chi^2_{11} \geq 11.495) \leq 0.50$.

Even if the $\chi^2$ tests had permitted the rejection of the null hypotheses, the validity of Windsor's conclusions would have been restricted to biologists in *American men and women of science*[5] whose birthdays were listed. Other factors such as age and sex proportions among taxonomists, molecular biologists, and the general population would have complicated analysis and interpretation even further.

**Table 1** Frequencies ($f$) of birthdates by Sun sign from a sample of 812 biologists[1] and expected frequencies ($F$) if the proportions were the same as in the general population (see text)

| Sun sign | General population | Taxonomists | | Molecular biologists | | Cumulative | |
|---|---|---|---|---|---|---|---|
| | % | $f$ | $F$ | $f$ | $F$ | $f$ | $F$ |
| Aquarius | 9.05 | 26 | 30.98 | 35 | 42.57 | 61 | 73.55 |
| Pisces | 8.18 | 31 | 28.00 | 36 | 38.48 | 67 | 66.48 |
| Aries | 8.11 | 28 | 27.76 | 58 | 38.15 | 86 | 65.91 |
| Taurus | 8.46 | 30 | 28.96 | 32 | 39.80 | 62 | 68.76 |
| Gemini | 8.08 | 31 | 27.66 | 39 | 38.01 | 70 | 65.67 |
| Cancer | 8.84 | 38 | 30.26 | 41 | 41.59 | 79 | 71.85 |
| Leo | 8.67 | 32 | 29.68 | 32 | 40.79 | 64 | 70.46 |
| Virgo | 8.98 | 31 | 30.74 | 42 | 42.24 | 73 | 72.98 |
| Libra | 8.17 | 25 | 27.97 | 41 | 38.43 | 66 | 66.40 |
| Scorpio | 7.98 | 18 | 27.32 | 41 | 37.54 | 59 | 64.86 |
| Sagittarius | 7.75 | 27 | 26.53 | 40 | 36.46 | 67 | 62.99 |
| Capricorn | 7.64 | 25 | 26.15 | 33 | 35.94 | 58 | 62.09 |
| Total | 99.91 | 342 | 342.00 | 470 | 470.00 | 812 | 812.00 |
| $\chi^2$ | | 7.282 | | 16.368 | | 11.580 | |

While I hope that this modest analysis will be instructive and entertaining to students of biostatistics, I regret that astrologers may be disappointed, although they usually stress[2] that birth place and planets' positions are the determining factors of the birth chart or 'horoscope' used to suggest the likelihood of certain events.

ANTONIO E. COLAS

*University of Wisconsin*
*Centre for Health Sciences,*
*Madison, Wisconsin 73706*

1 Windsor, D. A., *Nature*, **248**, 788 (1974).
2 Parker, D., and Parker, J., *The complete astrologer's sun-signs guide*, 5–8 (Crown Publishers, New York, 1973).
3 Zar, J. H., *Biostatistical analysis*, 41–69 (Prentice-Hall, Englewood Cliffs, 1974).
4 Snedecor, G. W., and Cochran, W. G., *Statistical methods*, sixth ed., fifth printing, 228–257 (Iowa State University Press, Ames, 1972).
5 *American men and women of science: the physical and biological sciences*, twelfth ed. (edit. by Jacques Cattell Press), **1-6**, (Bowker, New York, 1971–73).

DR WINDSOR REPLIES—Dr Colás certainly provides an appropriate criticism of my Sun sign findings.

The statistical method I used, before publication, was a non-parametric analysis of a manifold population as described by Tate and Clelland[1]. Considering the 12 Sun signs as a manifold population, the hypothesis tested was that the distribution of birthdays was uniform, $H_0 = \text{Ar} = \text{T} = \text{G} \ldots = \text{P} = 1/12$. $\chi^2 = [(f^0 - f^e)^2]/f^e$, where $f^0$ is the frequency of characters in the sample and $f^e$ is the frequency expected under the null hypothesis, that is, the average.

$$\chi^2 = [(\text{Ar} - f^e)^2]/f^e + [(T - f^e)^2]/f^e + [(\text{G} - f^e)^2]/f^e \ldots + [(\text{P} - f^e)^2]/f^e.$$

$\chi^2$ for molecular biologists = 13.818; $\chi^2$ for taxonomists = 8.666. At 11 degrees of freedom the probabilities of the hypotheses being true are 0.025 and 0.600, respectively. Since the hypothesis is rejected in both cases, therefore, the distributions were not random.

I compared molecular biologists with themselves and I compared taxonomists with themselves. I did not compare them with each other or with the general population. Molecular biologists and taxonomists are simply different kinds of biologists and must be considered independently or else the distinction is lost. Comparing them with the general population is unnecessary because it is obvious that, for the most part, they did indeed come from the same general population. The problem is not to identify similarities and establish origins, but to separate components and identify their characteristics. I used *American men and women of science* as my data base because it is a readily accessible, reproducible sample. Its validity can only be determined by gathering additional data. To cite one instance of possible invalidity, many practising molecular biologists are probably not listed as such, but appear classified as biochemists or physiologists, since the scientists describe their own disciplines.

My data were, admittedly, meagre, but it was hoped that their publication would inspire, or even goad, scientists with more resources to pursue this line of research by acquiring and analysing larger data bases. The basic concept of extraterrestrial influences on human activities is so important that it certainly deserves much more investigation than is now being performed.

*PO Box 604,*
*Norwich, New York 13815.*

1 Tate, M. W., and Clelland, R. C., *Non-parametric and shortcut statistics*, 51–53 (Interstate, Danville, Illinois, 1957).