What should we know about the famous *Iris* data?

Marcin Kozak and Barbara Łotocka

The *Iris* dataset is famous among statisticians and pattern recognition scientists. It has been used to illustrate a variety of techniques (established and novel), for instance, in multivariate statistics[1,2], pattern recognition[3,4] and even visualization[5,6]. The set was popularized by Fisher[7] as a powerful example for discriminant analysis, but the data themselves originated from research by Anderson[8,9], who studied morphologic variations in *Iris* flowers of three species, *Iris setosa*, *I. versicolor* and *I. virginica*. Not only has the dataset been employed often (*Web of Science* lists around 4000 citations to Fisher's article), it is also now offered as a standard example dataset in various software and other sources, and so has an important contribution not only to science, but also to teaching statistics. Perhaps it was due to this popularity that some mistakes were incorporated into the data, as discussed by Bezdek *et al*.[10].

The dataset consists of samples of 50 flowers for each species, the information describing each flower being the length and width of sepal and petal, all four measured in centimetres. The scatterplot matrix of these traits is presented in Figure 1.

The problem is, however, that many (not all, as we will discuss later) botanists would disagree with a statement that *Iris* flowers have sepals! This can be confusing for statisticians who would prefer such famous datasets, used all over the world, to be correct from every single point of view – and now it occurs this does not have to be the case with the *Iris* data. Here we show that there are two sides of the coin, and that the current knowledge does not support either of the two claims: that *Iris* flowers do or do not have sepals.

What are the traits?

Both sepals and petals are modified leaves; a collective name for both of them is perianthium or perianth[11]. Usually green, sepals protect reproductive parts developing within the flower bud. Petals are usually brightly coloured, which signals to the pollinating animals that the pollen (and nectar) is ready to be collected. However, such perianth is a general feature of dicot species but not of all flowering plants. In monocots, the perianth is usually composed of identical members which are called tepals. Since the terms may be a bit confusing for a non-botanist, the morphology of dicot and monocot flowers is shown and explained in Figure 2.

Where is the problem?

The problem is that although some botanists say that *Iris* flowers do have sepals[12], others do not agree[13]. In numerous angiosperm species, the morphological distinction of perianthium and perigonium, or sepaloid and petaloid perigon is not clear[11]. Intermediary forms of both perianthium and perigonium evolved in many species and classification of such forms may be made differently by different authors. Hopefully, the molecular mechanisms underlying the formation of sepals and petals or tepals will provide less arbitrary classification criteria. It is known that formation of sepals is governed by A-class floral identity genes. In the generative meristem, the sepals are formed in those places where the A-class genes are expressed. For the formation of petals, both A-class and B-class genes have to be expressed. In the few monocots having petaloid tepals that have been studied (*Iris* was not among them), the B-class domain is extended to encompass both whorls of perigon[14]. Therefore, in molecular terms such monocots have only petals in their perigon. In the other monocots the molecular background of sepaloid perigon formation is not clear.

Perhaps in time, after more molecular data have been collected, the distinction between perianth with sepals + petals and perigon with tepals will be neglected.

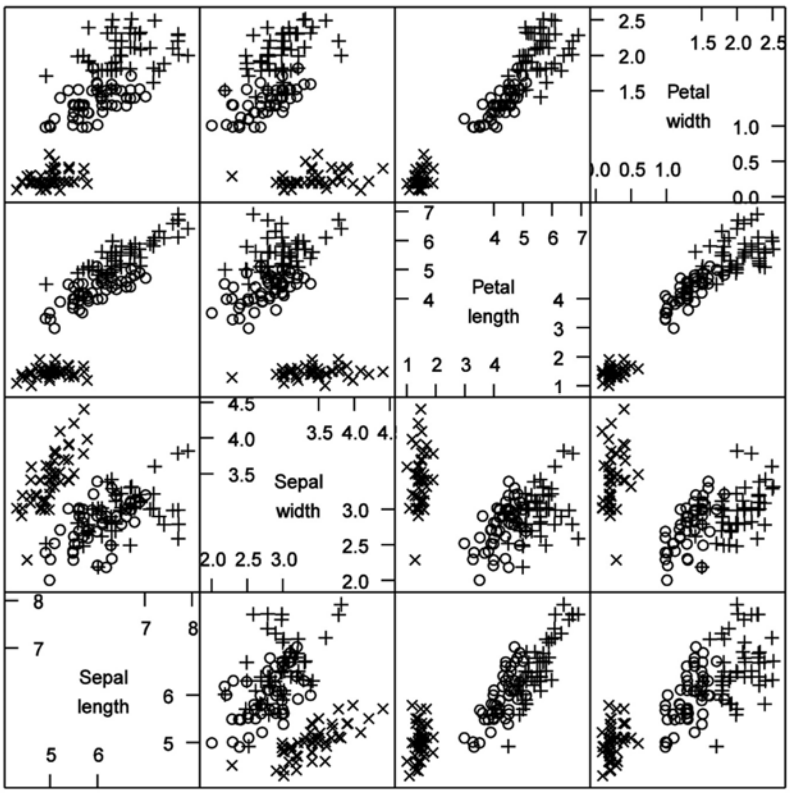

Figure 1. Scatterplot matrix of the *Iris* dataset. The traits (all in cm) were slightly jittered to remove overlap. ○, *Iris versicolor*; ×, *Iris setosa*; +, *Iris virginica.*

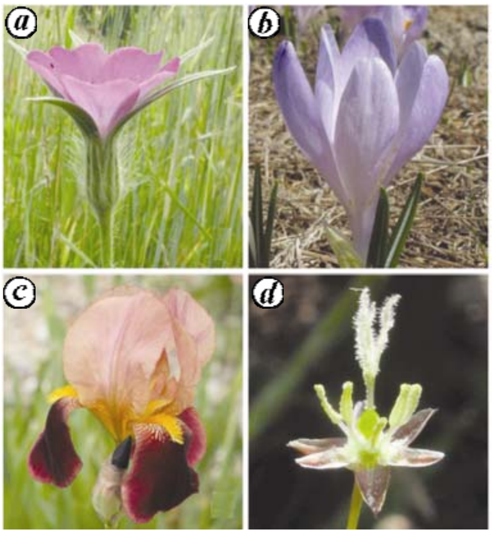

Figure 2. Flower morphology in dicots and monocots. ***a***, *Agrostemma githago* – An insect-pollinated flower, dicot family Caryophyllaceae. The perianth is two-whorled and clearly differentiated into an outer calyx composed of five free sepals (green and leaf-like, hairy) and an inner corolla composed of five petals (pink). ***b***, *Crocus scepusiensis* – An insect-pollinated flower, monocot family Iridaceae. Each flower has six tepals, all of them identical. The perianth is two-whorled, but it is not differentiated into calyx and corolla. Such a perianth is called a perigon and its members are termed tepals. In ornamental monocots, tepals are coloured and thus petal-like. The *Crocus* has a petaloid perigon composed of identical tepals. ***c***, *Iris germanica*, group *barbata elatior* – An insect-pollinated flower, monocot family Iridaceae. The petaloid perigon is composed of two whorls of three tepals each, but the inner and outer tepals differ in shape and colour. Therefore, some botanists name them sepals and petals. ***d***, *Luzula* sp. flower, monocot family Juncaceae. In this wind-pollinated species, the perigon is composed of two whorls of identical sepal-like tepals. Therefore this perigon is sepaloid.

Conclusion

Our aim is not to claim that *Iris* flowers have or do not have sepals, although we agree that such a discussion is needed. Quite likely only molecular techniques will enable one to say with certainty whether or not *Iris* flowers do have sepals.

It appears that the dataset itself is not interesting for botanists, and only statistics and some other related disciplines are familiar with it and show any interest in it. Thus we want to direct attention of statisticians, who usually are so far from botany, to this discussion, still contemporary. From a historical point of view, it is interesting that such a famous dataset, so often used and so well known, is problematic in such a simple aspect. As it seems, this aspect is not so simple, and the history of this dataset is still alive.

1. Krzanowski, W. J. and Radley, D., *Biometrics*, 1989, **45**, 1163–1173.
2. Zhang, J. and Fang, W., *Commun. Stat.: Theor. Methods*, 2003, **32**, 435–457.
3. Hathaway, R. J. and Bezdek, J. C., *Pattern Recognition Lett.*, 2003, **24**, 1563–1569.
4. Priebie, C. E., *Am. Stat.*, 2011, **65**, 167–169.
5. Blanchard, F., Herbin, M. and Lucas, L., *Inf. Vis.*, 2005, **4**, 257–265.
6. Kozak, M., *IEEE Comput. Graphics Appl.*, 2011, **30**, 3–7.
7. Fisher, R. A., *Ann. Eugenics*, 1936, **7**, 179–188.
8. Anderson, E., *Bull. Am. Iris Soc.*, 1935, **59**, 2–5.
9. Anderson, E., *Ann. Mo. Bot. Gard.*, 1936, **23**, 457–509.
10. Bezdek, J. C., Keller, J. M., Krishnapuram, R., Kuncheva, L. I. and Pal, N. R., *IEEE Trans. Fuzzy Syst.*, 1999, **7**, 368–369.
11. Cronk, Q. C. B., *The Molecular Organography of Plants*, Oxford University Press, 2009.
12. Uno, G. E., *Am. J. Bot.*, 1982, **69**, 818–823.
13. Stevens, P. F., Angiosperm phylogeny website, version 12, Monocotyledons, July 2012; available at http://www.mobot.org/MOBOT/research/APweb/
14. Kanno, A., Saeki, H., Kameya, T., Saedler, H. and Theissen, G., *Plant Mol. Biol.*, 2003, **52**, 831–841.

Marcin Kozak is in the Department of Quantitative Methods in Economics, University of Information Technology and Management in Rzeszow, Sucharskiego 2, 35-225 Rzeszów, Poland; Barbara Łotocka is in the Warsaw University of Life Sciences – SGGW, Faculty of Agriculture and Biology, Department of Botany, 159 Nowoursynowska Str, 02-776, Warsaw, Poland.*
**e-mail: nyggus@gmail.com*

Initiating exploration of the gut microbiome in Indians

Sudarshan A. Shetty, Nachiket P. Marathe and Yogesh S. Shouche

Human gut harbours approximately 10^{14} bacterial cells which outnumber the total number of human body cells[1]. The gut microbiota plays a vital role in development, nutritional, health and disease status of its host. Assemblage of these microbes is known to be predisposed by genetic make-up of an individual, age, dietary habit, ethnicity and environmental factors. India is distinguished from the rest of the world with vast diversity in the genetic composition of its human population, coupled with the enormous diversity in dietary habits, cultural affiliations with various religions and geographic scattering. However, there are hardly any studies on gut microbial communities in the Indian population. Five of the recent publications have addressed this issue. The results of these indicate differences in gut microflora of Indians from the previous studies in other populations[2–6]. Although these are preliminary studies, the observations are significant in terms of the gut microbiota in Indian diaspora. These studies have highlighted the necessity to further explore the gut microbiota of Indian individuals.

Colonization of gut by microbes takes place immediately after birth and is