**The iris flower data set**

This data set was collected by Edgar Anderson to quantify the morphologic variation of Iris flowers of three related species. Two of the species were collected from the same pasture, on the same day, and measured at the same time by the same person with the same apparatus.

The iris data set consists of 50 samples each, of three iris species (Iris setosa, Iris virginica and Iris versicolor). From every sample four features were measured: the length and width of both the sepals and petals, in centimetres.

British statistician and biologist Ronald Fisher used this dataset to develop a linear discriminant model to distinguish the species from each other. Ronald Fisher is one of the most celebrated statisticians of all time. He is described as “a genius who almost single-handedly created the foundations for modern statistical science”.

The data set has some key features that make it very useful in all sorts of data analytics, from data visualisation to machine learning. It is small, but not trivial. It consists of real data of good quality. The goal of the dataset, to discriminate between three species of Iris, based on measurements is simple, yet challenging. Also, the fact someone as famous as Ronald Fisher used the dataset makes it more interesting to a lot of statisticians. No wonder it is one of the most used data sets in the world of statistics.

**Sources:**

<https://en.wikipedia.org/wiki/Iris_flower_data_set>

 Edgar Anderson (1936). ["The species problem in Iris"](http://biostor.org/reference/11559). [Annals of the Missouri Botanical Garden](https://en.wikipedia.org/wiki/Annals_of_the_Missouri_Botanical_Garden). **23** (3): 457–509. [doi](https://en.wikipedia.org/wiki/Digital_object_identifier):[10.2307/2394164](https://doi.org/10.2307%2F2394164). [JSTOR](https://en.wikipedia.org/wiki/JSTOR) [2394164](https://www.jstor.org/stable/2394164).

 Edgar Anderson (1935). "The irises of the Gaspé Peninsula". Bulletin of the American Iris Society. **59**: 2–5.

<https://stats.stackexchange.com/questions/74776/what-aspects-of-the-iris-data-set-make-it-so-successful-as-an-example-teaching>

[Hald, Anders](https://en.wikipedia.org/wiki/Anders_Hald" \o "Anders Hald) (1998). A History of Mathematical Statistics. New York: Wiley. [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [978-0-471-17912-2](https://en.wikipedia.org/wiki/Special:BookSources/978-0-471-17912-2).