TAMS42 (Probability and Statistics) Vinjett 1
—the concept of distribution

- This vinjett is suggested to be done with the help of suitable softwares (such as **Matlab** and **R**).

Suppose that we examine a PHEV (a Plug-in-Hybrid Electrical-Vehicle): Toyota Corolla Hybrid 5-Dörras, see the below link for the vehicle descriptions

`https://www.toyota.se/nya-bilar/corolla/index.json`

According to the descriptions, the fuel consumption is claimed to be: 4.9 l/100 km. Of course, such fuel consumption depends on, among other things, driving style and driving conditions. That is to say, for every 100 km that such a vehicle is driven, the fuel consumption may not be exactly 4.9 l. Such a 4.9 l represents the theoretical average of the distribution of fuel consumption of such a vehicle. Actually one can check and read the real-time fuel consumption anytime during driving from the inside car monitor.

In order to understand the distribution of fuel consumption of such a vehicle, suppose that Simon recently purchased such a vehicle with plate number "ABC123", and he has read the fuel consumption from the car monitor 200 times, and the data are:

```
fuel=[4.1158 5.3870 3.8805 6.1539 4.3330 4.8955 4.2330 3.8311 4.5341 3.8649 ...
      4.2773 4.4627 4.3155 4.4042 4.8113 5.4049 4.4573 4.5578 4.5212 4.3175 ...
      5.3217 4.8924 5.0314 4.2952 6.8844 5.8291 5.6043 4.7810 5.2716 4.8309 ...
      5.4883 4.0849 3.5138 6.0620 3.0882 6.0662 4.7859 4.5314 5.8869 4.1475 ...
      5.6805 5.8745 4.9562 4.0015 5.1551 5.4857 4.1390 3.0272 4.7226 5.0671 ...
      4.6904 5.6923 5.2422 3.0994 3.8872 4.9788 6.0091 3.7994 5.0093 4.7214 ...
      5.5096 5.4003 5.7402 5.2557 3.9886 3.2347 4.3960 5.6182 6.4031 6.3457 ...
      5.4458 4.9968 3.9515 5.0297 4.7583 7.3530 3.6018 4.1621 2.9481 4.2048 ...
      6.0216 3.8829 7.0471 4.5047 3.6660 3.9128 3.6561 3.3744 4.2035 5.5276 ...
      5.3947 4.6908 2.6555 4.9297 5.1155 5.9086 5.6006 2.7920 4.0381 5.8566 ...
      2.3784 4.4947 4.4573 3.8883 4.2309 3.6248 3.7941 3.4644 4.8369 4.9606 ...
      4.2308 5.1101 4.3319 6.6625 6.6106 4.4874 4.3422 4.4491 4.7733 4.1961 ...
      5.9661 3.6897 3.8159 5.1562 4.5296 5.1378 4.7852 3.8350 5.2119 4.7599 ...
      5.4793 4.1202 5.5627 6.2677 4.1122 6.4106 5.1785 6.1298 4.5123 6.9122 ...
      5.3271 5.2200 3.5264 5.8719 6.1488 4.1759 4.0956 5.0037 6.3840 5.5607 ...
      3.9509 5.9595 5.0886 4.3558 6.5410 5.0789 5.0778 4.2815 6.2453 4.3989 ...
      3.0371 4.3107 3.2845 4.5748 6.6271 4.6203 5.4804 6.3891 4.9233 4.6326 ...
      4.5572 4.6974 7.2891 4.0919 2.7671 4.9166 5.9177 4.4199 3.8854 5.5051 ...
      5.4789 2.9622 5.3474 4.1286 6.7186 5.5851 4.9462 4.1733 4.7450 3.3730 ...
      4.4194 5.0278 6.2798 4.3090 5.8309 4.1447 5.7124 6.1805 4.4324 5.1870]
```

To understand the distribution of fuel consumption, what do we do with these data?

(a) **First, check the histogram of the data.** What can you see from the histogram? Are these 200 fuel consumptions around 4.9 l? Roughly speaking, such a histogram is an approximation of the real distribution of fuel consumption of the car "ABC123". One can think of the real distribution as the limit of these histograms as the data size n tends to infinity.

(b) **Second, look at some important characteristics of the data.** Discuss what kind of characteristics of data are usually taken into account? These characteristics can help us understand the real distribution of fuel consumption of the car "ABC123".