Analysis of the Cereal Dataset

Information on the distribution of various groups of nutrients among 77 commonly available breakfast cereals was obtained from the *statistical graphics exposition* in 1993. The number of sugars/serving, was recorded.

We estimated the mean, the median and the standard deviation of the sugars/serving. The Gini standard deviation was also estimated because the simple standard deviation is sensitive to outliers. In addition, the standard errors and 95% confidence intervals (CI) of the sample mean can be easily computed but others are not available. However, the SE and 95% CI of **ALL** the statistics can be easily estimated by using bootstrapping (1000 replicates).

Table 1 present the results. The bootstrap estimates are based on the simulation, so sometimes the bootstrap SE will be larger, and sometimes it will be smaller. However, The 2 set (analytical & bootstrap) of statistics for the mean must be **very close**. And the bootstrap precision is similar to the analytic precision. These 2 set of statistics for mean have similar shape of sampling distributions. They are all normal distribution (bell curve) and the bootstrap distribution may be slightly left (in my sampling) or right skewed. But their shapes should be very close.

Table 1. Statistics about sugars/serving and comparison of measures of precision computed using bootstrapping. (NA indicates statistic was not available)

	Estimate	SE	LCL	UCL	Boot.SE	Boots.LCL	Boot.UCL
mean	7.03	0.5	6.03	8.03	0.50	6.01	7.93
median	7.00	NA	NA	NA	0.92	5.00	8.50
sd	4.38	NA	NA	NA	0.23	3.85	4.78
gsd	4.37	NA	NA	NA	0.26	3.79	4.79