Drill 3

By minimizing D_1 and D_2 with respect to μ

$$D_1 = \sum_{i=1}^{n} |\mu - X_i| \quad \text{and} \quad D_2 = \sum_{i=1}^{n} (\mu - X_i)^2,$$

we have the sample median

$$\hat{\mu}_1 = \underset{i=1,2,\ldots,n}{\text{median}}\, X_i$$

and the sample mean

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The HL estimator (Hodges and Lehmann, 1963) is also widely used for a location estimation. It is defined as the median of all pairwise averages given by

$$\hat{\mu}_3 = \underset{i<j}{\text{median}} \left(\frac{X_i + X_j}{2} \right).$$

We assume that $X_i \sim N(\mu, \sigma^2)$. Find the variance of the three estimators above and compare them (if theoretical derivation is not available, R program can be used). Which one is the best in a sense of smaller variance? One can also refer to Park et al. (2021) and Park and Wang (2020).

References

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics*, 34:598–611.

Park, C., Kim, H., and Wang, M. (2021). Investigation of finite-sample properties of robust location and scale estimators. *Communication in Statistics – Simulation and Computation*, To appear. doi:10.1080/03610918.2019.1699114.

Park, C. and Wang, M. (2020). `rQCC`: Robust quality control chart. `https://CRAN.R-project.org/package=rQCC`. R package version 1.20.7 (published on July 5, 2020).