

# 1 Introduction: Assignment

## Instructions

Submit your answers to the seven first exercises through Aula Global. Remember to round longer decimals to the second place after the comma. The remainder of the assignment concern self-studies.

## Descriptive statistics

Imagine you measured the (average) voice pitch of 20 Korean speakers (in Hz)<sup>1</sup>. The data looks as follows:

208, 158, 153.3, 259.7, 102.9, 86.1, 231.9, 285.1, 181.2, 266, 252.5, 185.5, 110.7, 236.8, 207.5

Codified as a vector, called `pitch`, in R:

```
pitch <- c(208, 158, 153.3, 259.7, 102.9,
           86.1, 231.9, 285.1, 181.2, 266,
           252.5, 185.5, 110.7, 236.8, 207.5)
```

Report the following summaries of this sample

1. Mean
2. Median
3. Variance
4. Standard deviation

Answer the following questions:

5. Does the mean give a good summary of this sample? Why (not)?
6. What is suggested by the mean and median leading to different summaries?
7. Complete the following sentence with the first word that comes to your mind: *It's raining. I'm going to take the ...*

## Self-study

1. Work through Chapter 1, Chapter 2 (until the end of subsection 2.2) and Chapter 3 of [Introduction to Data Analysis](#) (Franke 2021);
2. Make sure you have a working R environment: either locally –on your computer– or by having made sure you can execute code in a “colab” document
3. Look at the guidelines for the analysis report and read the illustrative report. Think of a general question about language that interests you and write it down.

## References

- Franke, Michael. 2021. *An Introduction to Data Analysis*.
- Winter, Bodo, and Sven Grawunder. 2012. “The Phonetic Profile of Korean Formal and Informal Speech Registers.” *Journal of Phonetics* 40 (6): 808–15. <https://doi.org/10.1016/j.wocn.2012.08.006>.

---

<sup>1</sup>This data was randomly sampled from the pre-processed and shortened version of the study of Winter and Grawunder (2012), as provided in Franke (2021). The data is accessible [here](#).