MECH481A6: Engineering Data Analysis in R

Chapter 10 Homework: Measurement

Student Name

30 October, 2023

Load packages

Chapter 10 Homework

This homework will give you practice at working with a measurement dataset: `airlift_mass_repeatability.csv`. This data set represents repeated measures of "blank" air sampling filters.

A couple notes to consider when reporting answers in response to questions. The microbalance used to make these measurements reads out to the nearest microgram (μg), which is 0.000001 g or 0.001 mg. Thus, be careful when reporting descriptive statistics so as not to overstate your **precision**. Use the `round()` function to avoid reporting more than 0.1 μg of precision (or 0.0001 mg). Here is some example code that uses the `across()` function from `dplyr::` to round numeric output to just four digits (appropriate for mg units in this exercise):

```
dplyr::mutate(across(.cols = where(is.numeric), .fns = round, 3))
```

Question 1

Import the `airlift_mass_repeatability.csv` file into a data frame called `blanks` and perform the following data wrangling in a single pipe:

- retain only the first 3 columns of data;
- rename the columns with the names `date`, `id`, and `mass_mg`;
- convert the `date` column vector into a date class object using `lubridate::`
- convert the `id` variable to a class `factor` (this can be accomplished using `base::as.factor()` or `purrr::as_factor()`)
- create a new column vector named `mass_mg` by rescaling the `mass_g` data (i.e., convert g to mg by multiplying `mass_g` by 1000)

Question 2:

2a. Are there any NAs present in the data frame?
2b. How many unique filter IDs are present in this data frame?
2c. How many samples are present for each filter ID? Hint: look up the `dplyr::count()` function.
2d. Over how long of a period were these blank measurements made? Hint: this can be done in base R with a `max() - min()` or with `lubridate::interval() %>% as.duration()`.

Question 3

Group the `blanks` data frame by `id` and calculate mean, median, and standard deviations for each filter id. Hint: use `group_by() %>% summarise()` to do this efficiently.

Question 4

Calculate the limit of detection (LOD) for this measurement method. Note: you will need to calculate standard deviations for each filter `id` (as done in question 3) and then estimate LOD from $LOD = 3 \cdot \sigma_b$ where σ_b is calculated for each filter `id`.