**More Exercises: Weight loss**

Consider a dataset of sequential measurements of a person’s weight while on a diet (the ’weightloss’ dataset, see attached).

1. Read the dataset (‘weightloss.csv’), and convert the ’Date’ variable to the Date class...
2. Add a new variable to the dataset, with the subjects’s weight in kilograms (kg) ( kg = 2.204 pounds).
3. Produce a line plot that shows weight (in kg) versus time.
4. The problem with the plot you just produced is that all measurements are connected by a line, although we would like to have line breaks for the days where the weight was not measured. To do this, construct a dataframe based on the weightloss dataset that has daily values. Hints:

• Make an entirely new dataframe, with a Date variable, ranging from the first to last days in the weightloss dataset, with a step of one day

• Using merge, paste the Weight data onto this new dataframe. Check for missing values. Use the new dataframe to make the plot.

1. Based on the new dataframe you just produced, graph the daily change in weight versus time. Also add a dashed horizontal line at y=0

Writing functions and using lists

1. You learned that you can take subset of a string using the substr function. First, using that function to extract the first character of a bit of text. Then, write a function called firstTwoChars that extracts the first two characters of any bit of text.
2. Recall the functions head and tail. Write a function called middle that shows a few rows around (approx.) the ’middle’ of the dataset. *Hint: use* nrow, print*, and possibly* floor*.*
3. First read the following list:

veclist <- list(x=1:5, y=2:6, z=3:7)

1. Using sapply, check that all elements of the list are vectors of the same length. Also calculate the sum of each element.
2. Add an element to the list called ’norms’ that is a vector of 10 numbers drawn from the standard normal distribution
3. Using the pupae data (see below), use a t-test to find if PupalWeight varies with temperature treatment, separate for the two CO2 treatments (so, do two t-tests). Use split and lapply.
4. Read the data coweta.csv (see below). Then, split the data by species, to produce a list called coweeta\_sp. Keep only those species that have at least 10 observations.(Hint: first count the number of observations per species, save that as a vector, second which are at least 10, and use that to subscript the list.) If you don’t know how to do this last step, skip it and continue to the next item.
5. Using the split Coweeta data, perform a linear regression of log10(biomass) on log10(height), separately by species.
6. Run this code to get two vectors:

x <- rnorm(100)

y <- x + rnorm(100)

Run a linear regression y = f(x), save the resulting object. Look at the structure of this object, and note the names of the elements. Extract the residuals and make a histogram.

1. From question 10 write a function that takes an lm object as an argument, and plots a histogram of the residuals.

**File** : ‘pupae.csv’

This dataset is from an experiment where larvae were left to feed on *Eucalyptus* leaves, in a glasshouse that was controlled at two different levels of temperature and CO2 concentration. After the larvae pupated (that is, turned into pupae), their body weight was measured, as well as the cumulative ‘frass’(larvae excrement) over the entire time it took to pupate.

**Variables:**

• T\_treatment - Temperature treatments (’ambient’ and ’elevated’)

• CO2\_treatment - CO2 treatment (280 or 400 ppm).

• Gender - The gender of the pupae : 0 (male), 1 (female)

• PupalWeight - Weight of the pupae (g)

• Frass - Frass produced (g)

**File:** ‘coweeta.csv’

Tree measurements in the Coweeta LTER.

**Variables:**

• species One of 10 tree species

• site Site abbreviation

• elev Elevation (m asl)

• age Tree age (yr)

• DBH Diameter at breast height (cm)

• height Tree height (m)

• folmass Foliage mass (kg)

**lm**

Fitting Linear Models

**lm** is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance (although [aov](https://www.rdocumentation.org/link/aov?package=stats&version=3.6.2) may provide a more convenient interface for these).

Keywords

[regression](https://www.rdocumentation.org/search/keywords/regression)