Hypothesis testing in R.
Comparing means and fitting distributions

There is a restriction to the length of the report: 4 pages

200 mice received a treatment "Nutritional Supplement" for 6 months. We want to know whether the treatment has an impact on the weight of the mice.

Another 200 rats received the same treatment, but results seem to differ.

To answer to this question, the weight of the mice has been measured before and after the treatment. This gives us 200 sets of values before treatment and 200 sets of values after treatment from measuring twice the weight of the **same mice**, and another 200 sets of values before treatment and 200 sets of values after treatment from measuring twice the weight of the **same rats.**

For this test you are required to create two datasets for each set of mice and rats. The data will be created from artificial data.

To obtain full marks you will need to write all relevant underlying theory, explain the hypothesis test (critical values, test statistics and p-values) in detail and all underlying assumptions. Just reporting the test output and interpretation will not be enough to get full marks.

Create the following:

Task 1: Data Generation

a. The weights of **mice** as "before" and "after" the treatment, coming from a normal distribution with mean = 20, and variance = 2.
   For the "after" treatment add to the mean 1 unit and to the variance 0.5 units, that is mean = 21 and variance = 2.5.
b. The weights of **rats** as "before" and "after" the treatment, coming from a Weibull distribution with shape = 10, and scale = 20.
   For the "after" treatment remove from the shape 1 unit and add to the scale 1 unit, that is shape = 9 and scale = 21.
c. Using the function 'qplot' with 'geom = density' (from your lecture notes) compare for each of your data sets mice(before, after) and rats(before, after)
d. Perform the same operation using 'geom = boxplot'.

Task 2: Appropriateness for Hypothesis t-testing

a. For your **mice** data set (combined "before" + "after") examine whether the data passes normality qualitatively (QQ plot) and quantitatively (Shapiro-Wilk test).
b. For your **rats** data set (combined "before" + "after") examine whether the data passes normality qualitatively (QQ plot) and quantitatively (Shapiro-Wilk test).
c. Explain the output of your analysis for each of the dataset and discuss the appropriate test to test your hypothesis.

Task 3: Hypothesis testing

a. For the normal data set (**mice**) examine perform a paired t-test and explain your findings. You need to extract and comment on all output of the t-test:
    a. T-test statistic
    b. Degrees of freedom
    c. P-value
    d. Confidence Interval
    e. Sample estimates

b. For the **rats** dataset perform a non-parametric -test and comment on your finding.

Task 4: Fitting distributions

a. Lastly, for the **rats'** datasets use the function 'fitdist' (from the 'fitdistrplus' R package) and examine the best-fit distribution (even if we know the TRUE distribution). Fit a Weibull, a lognormal and a Gamma distribution and discuss your findings making use of the package comparison tool for a Density, CDF, QQ, and PP.