## Exercise 6 : Hypothesis testing of numerical data using T- tests

For each task, clearly state the null, the alternative, the p value, and the acceptance or rejection of the null, and also the final conclusion of the test in simple words. Also **describe your code and results.**

**TASK 1**

a = c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)

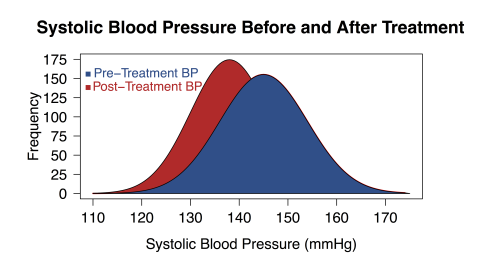
b = c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)

t.test(a,b, var.equal=TRUE, paired=FALSE)

**TASK 2**

Let’s say that we work at a largel health clinic and we’re testing a new drug, Procardia, that’s meant to reduce hypertension. We find 100 individuals with a high systolic blood pressure ( mean=145 or mean=145mmHg, SD=9 or SD=9mmHg), we give them Procardia for a month, and then measure their blood pressure again. We find that the mean systolic blood pressure has decreased to 139mmHg with a standard deviation 8mmHg. The datasets for both of these treatments are available on canvas in the assignment 6 section in a file called ProcardiaStudy.csv

We can visualize this difference with a kernel density plot as:



Here, we would conduct a t-test using:

preTreat <- ProcardiaStudy$preTreat  
postTreat <- ProcardiaStudy$postTreat

hist(preTreat,probability=TRUE)

hist(postTreat,probability=TRUE)

plot(density(preTreat), col="red",

main="Systolic Blood Pressure Before and After Treatment",  
 xlab = "Systolic Blood Pressure (mmHg)",  
 ylab = "Frequency",  
 xlim = c(110, 160))

lines(density(postTreat), col="blue")

t.test(preTreat, postTreat, paired = TRUE)

Which means the new drug may/may not lead to change in blood pressure.

**TASK 3**

An outbreak of Salmonella-related illness was attributed to ice cream produced at a certain factory. Scientists measured the level of Salmonella in 9 randomly sampled batches of ice cream. Usually anything above 0.3 MPN/g is considered unsafe. The levels (in MPN/g) were: 0.593 0.142 0.329 0.691 0.231 0.793 0.519 0.392 0.418 Is there evidence that the mean level of Salmonella in the ice cream is greater than 0.3 MPN/g?

Let be the mean level of Salmonella in all batches of ice cream. Here the hypothesis of interest can be expressed as:

H0: = 0.3

Ha: > 0.3

Hence, we will need to include the options alternative="greater", mu=0.3. Below is the relevant R-code:

x = c(0.593, 0.142, 0.329, 0.691, 0.231, 0.793, 0.519, 0.392, 0.418)

t.test(x, alternative="greater", mu=0.3)

**TASK 4**

Subjects were given a drug (Treatment group) and an additional 6 subjects a placebo (Control group). Their reaction time to a stimulus was measured (in ms). We want to perform a two-sample t-test for comparing the means of the treatment and control groups. Let T be the mean of the population taking medicine and C the mean of the untreated population. Here the hypothesis of interest can be expressed as:

H0: no difference in mean of the two groups

Ha: mean of group C is lower than mean of group T

Here we will need to include the data for the treatment group and the data for the control group. We will also need to include the options alternative="less", mu=0. Finally, we need to decide whether or not the standard deviations are the same in both groups. Below is the relevant R-code when assuming equal standard deviation:

Control = c(91, 87, 99, 77, 88, 91)

Treat = c(101, 110, 103, 93, 99, 104)

t.test(Control,Treat,alternative="less", var.equal=TRUE)

Below is the relevant R-code when not assuming equal standard deviation:

t.test(Control,Treat,alternative="less")

Here the first t-test and the second t-test give roughly the same results (p-value = 0.00313 and 0.00339, respectively)

**TASK 5**

You go to the Galapagos Islands. You measure size of beaks for 30 finches(in inches). Make a CSV file that records this data. Then you measure size (in inches) of beaks for 30 finches after a drought. Make another CSV file that records this data too with slightly lower numbers.

-Import both CSV files into R.

-Find the mean and standard deviations of the datasets.

-Plot histograms of the two datasets.

-By eyeballing figure out whether they are normally distributed or not.

-Find out whether there is any statistical difference between mean beak sizes before and after drought. In other words, do a t-test to compare the means of the two CSV files that are imported into R.