In this document, we'll focus on using `R` to run hypothesis tests.

You will struggle to complete all these questions in the workshop if you try to do them all on your own. Instead, find some way to split the questions between the members of your group. By the end of Week 4, make sure each member of your group has access to the solutions to every question attempted by the group. How you go about doing this within your group is up to you. Remember that the deadline for your first groupwork-based individual assignment is **noon Monday 30th October**, and that the assignment requires you reflect on how you contributed to this week and last week's groupwork within the workshops.

Part 1: One Sample Tests

In this section, we will focus on hypothesis tests performed on a single sample (this will include paired hypothesis tests, which as you saw in Video 4.6 can be reduced to a single hypothesis test).

These questions are all similar to the one I went through at the beginning of the workshop, so you might find it useful to take a look at the code I made use of when answering them.

1. Every morning and evening, our dog Quiz is taken for a walk. A dog of his breed should get two 45 minute walks a day. Over six days, I measure how far Quiz walks while out on his two walks, leading to the following values in minutes

   $$45.6, 41.8, 46.4, 41.7, 44.0, 44.6, 46.8, 42.9, 45.1, 45.7, 45.1, 41.6.$$

   Note that these times are in decimal values, not in minutes and seconds.

   (a) Use a Q-Q plot and the Shapiro-Wilk test to check whether this data can be considered as normally distributed.

   (b) Making the assumption that this data can indeed be considered normally distributed, find the test statistic for a hypothesis test with $H_0 : \mu = 45$.

   (c) I want the average walk length to be 45 minutes - longer than that and Quiz gets too tired, shorter than that and Quiz has too much energy at the end of the day. Describe an appropriate alternative hypothesis for this situation.

(d) Find the critical value for the hypothesis test with the alternative hypothesis you identified in part (c), and use it to decide whether H_0 can be rejected here.

2. The `faithful` data, already available in `R`, contains two variables. Both relate to Old Faithful, a geyser in Yellowstone National Park, Wyoming, USA. The geyser erupts every 45 to 120 minutes, throwing thousands of gallons of water into the air, and is a very popular tourist attraction.

   The variable `eruptions` gives the length in minutes of 272 minutes. The variable `waiting` gives the length in minutes of 272 gaps between eruptions.

   (a) Calculate the test statistic for a hypothesis test, to test the null hypothesis that the average duration of an eruption is 3.5 minutes.

   (b) For alternative hypothesis $H_1 : \mu < 3.5$, find the p-value for this test statistic, using the t-distribution.

   (c) Hence state whether H_0 can be rejected at the 5% level.

   (d) State which form of tailed test $H_1 : \mu < 3.5$ represents. Give a reason why we might want to apply this form of tailed test in the context of Old Faithful eruption times.

   (e) Redo questions b) and c) using a standard normal distribution instead of a t-distribution. What do you notice about the changes in values?

3. Geologists have theorised that waiting times for Old Faithful follow a simple pattern - a short waiting time follows a long waiting time, which follows another short waiting time, and so on.

   We can investigate this using the `faithful` data once again. We can divide the waiting times into two sets, representing odd eruptions and even eruptions within the data. The fact the `faithful` data is in chronological order means we can pair this data, with each odd eruption time paired with the even eruption time that followed it.

   (a) Create two vectors `faithfulodd` and `faithfuleven`, which contain all odd numbered waiting times and even numbered waiting times, respectively. (**Hint:** try using the function `seq` to generate all odd numbers between 1 and 271, and all even numbers between 2 and 272.)

   (b) We can check the geologists' theory by running a hypothesis test, with the null hypothesis being that there exists no difference between the average odd-numbered waiting period, and the average even-numbered waiting period. Give a brief explanation of how this can be done using `faithfulodd` and `faithfuleven`, including a description of the relevant null hypothesis.

(c) Run a t-test at the 1% significance level, with a two-tailed alternative hypothesis, in order to test the null hypothesis you gave in (a). Is that null hypothesis rejected?

Part 2: Two Sample Tests

In this section, we will focus on hypothesis tests performed on a two samples, which are not paired.

These questions are all similar to the one I went through at the beginning of the second half of the workshop, so you might find it useful to take a look at the code I made use of when answering them. In fact, Question 1 follows directly from that code.

1. Try re-running the polar bear example I demonstrated multiple times, using different distributions for at least one of the 2013 and 2023 data. Try to use distributions which have a different variance to each other. Roughly how different do the population variances have to be before the Levene's test starts rejecting the null hypothesis that they are equal? In such cases, how different are the results from the pooled t-test (which would be inappropriate in such an instance) and the unpooled t-test (which is the appropriate choice in such an instance)?

   (**Hint**: You can find the mean and variance for a discrete distribution `prob` relating to possible values `cubnum` with the following code:

   ```
   meanvalue<-cubnum%*%prob
   ```

   ```
   varvalue<-(cubnum**2)%*%prob-meanvalue**2
   ```

   Why not see if you can figure out why this code works, as an exercise?)

2. This question uses the `Guyer` data set, which is available after loading the `car` package into your library. The Guyer data comes from an experiment in cooperation. 30 teams of 4 people played the "Prisoner's Dilemma". The details of this game are very interesting, and I recommend looking them up. For this question, though, what matters is that each player can choose to either cooperate, or not to cooperate. Cooperation gets you a better outcome if other people co-operate as well, but a worse outcome if other people do not cooperate.

   In this experiment, each player played 30 times, leading to 120 choices per group as to whether or not to cooperate. Some groups made their choices anonymously, others had to make their decisions publicly.

(a) Test at the 10% significance level the null hypothesis that men are equally likely to cooperate as women, against the alternative hypothesis that they are not.

(b) Test at the 10% significance level the null hypothesis that people are equally likely to cooperate if their decisions are anonymous as if they are made publicly, against the alternative hypothesis that people are less likely to cooperate if their decisions are anonymous as if they are made publicly.

In both cases note that if a Levene's test does not allow you to assume sample variances are equal, the small number of values here means a normal test needs to be passed for each set of sample values before we apply an unpooled t-test.