One and Two-sample t-tests

The R function t.test() can be used to perform both one and two sample t-tests on vectors of data.

The function contains a variety of options and can be called as follows:

> t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95)

Here x is a numeric vector of data values and y is an optional numeric vector of data values. If y is excluded, the function performs a one-sample t-test on the data contained in x, if it is included it performs a two-sample t-tests using both x and y.

The option mu provides a number indicating the true value of the mean (or difference in means if you are performing a two sample test) under the null hypothesis. The option alternative is a character string specifying the alternative hypothesis, and must be one of the following: "two.sided" (which is the default), "greater" or "less" depending on whether the alternative hypothesis is that the mean is different than, greater than or less than mu, respectively. For example the following call:

> t.test(x, alternative = "less", mu = 10)

performs a one sample t-test on the data contained in x where the null hypothesis is that $\mu=10$ and the alternative is that $\mu<10$.

The option paired indicates whether or not you want a paired t-test (TRUE = yes and FALSE = no). If you leave this option out it defaults to FALSE.

The option var.equal is a logical variable indicating whether or not to assume the two variances as being equal when performing a two-sample t-test. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch (or Satterthwaite) approximation to the degrees of freedom is used. If you leave this option out it defaults to FALSE.

Finally, the option conf.level determines the confidence level of the reported confidence interval for μ in the one-sample case and $\mu_1-\mu_2$ in the two-sample case.

A. One-sample t-tests

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

H_0: $\mu = 0.3$
H_a: $\mu > 0.3$

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)

    One Sample t-test

data:  x
t = 2.2051, df = 8, p-value = 0.02927
alternative hypothesis: true mean is greater than 0.3
```

From the output we see that the p-value = 0.029. Hence, there is moderately strong evidence that the mean Salmonella level in the ice cream is above 0.3 MPN/g.

B. Two-sample t-tests

Ex. 6 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 μ_1 be the mean of the population taking medicine and μ_2 the mean of the untreated population. Here the hypothesis of interest can be expressed as:

H_0: $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 - \mu_2 < 0$

Here we will need to include the data for the treatment group in x and the data for the control group in y. 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)

        Two Sample t-test

data:  Control and Treat
t = -3.4456, df = 10, p-value = 0.003136
alternative hypothesis: true difference in means is less than 0
```

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

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

         Welch Two Sample t-test

data:  Control and Treat
t = -3.4456, df = 9.48, p-value = 0.003391
alternative hypothesis: true difference in means is less than 0
```

Here the pooled t-test and the Welsh t-test give roughly the same results (p-value = 0.00313 and 0.00339, respectively).

C. Paired t-tests

There are many experimental settings where each subject in the study is in both the treatment and control group. For example, in a matched pairs design, subjects are matched in pairs and different treatments are given to each subject in the pair. The outcomes are thereafter compared pair-wise. Alternatively, one can measure each subject twice, before and after a treatment. In either of these situations we can't use two-sample t-tests since the independence assumption is not valid. Instead we need to use a paired t-test. This can be done using the option paired =TRUE.

Ex. A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with either regular or premium gas, decided by a coin toss, and the mileage for that tank was recorded. The mileage was recorded again for the same cars using the other kind of gasoline. We use a paired t-test to determine whether cars get significantly better mileage with premium gas.

Below is the relevant R-code:

```
> reg = c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
> prem = c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
> t.test(prem,reg,alternative="greater", paired=TRUE)

        Paired t-test

data:  prem and reg
t = 4.4721, df = 9, p-value = 0.000775
alternative hypothesis: true difference in means is greater than 0
```

The results show that the t-statistic is equal to 4.47 and the p-value is 0.00075. Since the *p*-value is very low, we reject the null hypothesis. There is strong evidence of a mean increase in gas mileage between regular and premium gasoline.