Big Data Analytics

What is one-sample t-test?

one-sample **t-test** is used to compare the ***mean*** of one sample to a known standard (or theoretical/hypothetical) mean (μ).

Theoretical mean comes from:

- a previous experiment. For example, compare whether the mean weight of mice differs from 200 mg, a value determined in a previous study.
- or from an experiment where you have control and treatment conditions. If you express your data as "percent of control", you can test whether the average value of treatment condition differs significantly from 100.

Note that, one-sample t-test can be used only, when the data are normally distributed.

Research questions and statistical hypotheses

1.whether the mean (m) of the sample *is* *equal* to the theoretical mean (μ)?

2.whether the mean (m) of the sample *is* *less than* the theoretical mean (μ)?

3.whether the mean (m) of the sample *is* *greater than* the theoretical mean (μ)?

In statistics, we can define the corresponding ***null*** *hypothesis* (H_0) as follow:

1. $H_0: m = \mu$
2. $H_0: m \leq \mu$
3. $H_0: m \geq \mu$

The corresponding alternative hypotheses (Ha) are as follow:

1. $H_a: m \neq \mu$ (different)
2. $H_a: m > \mu$ (greater)
3. $H_a: m < \mu$ (less)

Note that:

- Hypotheses 1) are called **two-tailed tests**
- Hypotheses 2) and 3) are called **one-tailed tests**

Formula of one-sample t-test

The t-statistic can be calculated as follow:

$$t = \frac{m - \mu}{s/\sqrt{n}}$$

where,

- **m** is the sample **mean**
- **n** is the sample **size**
- **s** is the sample **standard deviation** with $n-1$ **degrees of freedom**
- μ is the **theoretical value**

We can compute the p-value corresponding to the absolute value of the **t-test statistics** (|t|) for the **degrees of freedom** (df): $df=n-1$.

R function to compute one-sample t-test

- `t.test(x, mu = 0, alternative = "two.sided")`

•**x**: a numeric vector containing your data values
•**mu**: the theoretical mean. Default is 0 but you can change it.
•**alternative**: the alternative hypothesis. Allowed value is one of “two.sided” (default), “greater” or “less”.

```r
# One-sample t-test
res <- t.test(my_data$weight, mu = 25)
# Printing the results
res
```

```
    One Sample t-test
data:  my_data$weight
t = -9.0783, df = 9, p-value = 7.953e-06
alternative hypothesis: true mean is not equal to 25
95 percent confidence interval:
 17.8172 20.6828
sample estimates:
mean of x
    19.25
```

In the result above :

- **t** is the **t-test statistic** value (t = -9.078),
- **df** is the degrees of freedom (df= 9),
- **p-value** is the significance level of the **t-test** (p-value = 7.95310^{-6}).
- **conf.int** is the **confidence interval** of the mean at 95% (conf.int = [17.8172, 20.6828]);
- **sample estimates** is he mean value of the sample (mean = 19.25).

•if you want to test whether the mean weight of mice is less than 25g (one-tailed test), type this:

```
t.test(my_data$weight, mu = 25, alternative = "less")
```

•Or, if you want to test whether the mean weight of mice is greater than 25g (one-tailed test), type this:

```
t.test(my_data$weight, mu = 25, alternative = "greater")
```

Access to the values returned by t.test() function

- *# printing the p-value*
- res$p.value
- [1] 7.953383e-06
- *# printing the mean*
- res$estimate
- mean of x
- 19.25
- *# printing the confidence interval*
- res$conf.int
- [1] 17.8172 20.6828
- attr(,"conf.level")
- [1] 0.95