Summary of the t-test chapter

1 One Sample t-test

Consider a sample $x_1, x_2, \dots, x_n$, which is continuous valued with unknown variance. The sample is either normally distributed (this can be checked on a Q-Q plot or with a goodness of fit test for instance), or the sample size n is large ($n \geq 30$ as a rule of thumb). The null hypothesis H_0 is that the true expectation (mean) μ of the random distribution from which the x_i values are drawn is equal to a specified value μ_0, i.e.,

$$H_0 : \mu = \mu_0.$$

The test statistic for the one-sample t-test is given by:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}},$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, and n is the sample size. Under H_0, the test statistic t follows a Student's t-distribution with $n - 1$ degrees of freedom.

Link with the z-test
When the population variance is known, the t-test can be approximated by a z-test, which uses the standard normal distribution.

⌨ **Python.** `scipy.stats.ttest_1samp(sample_data, mu_0)`

⌨ **Python.** `stats.probplot(sample_data, dist="norm", plot=plt)`

Selecting the Alternative Hypothesis
In hypothesis testing, the alternative hypothesis H_1 (or H_A) represents the statement we want to test against the null hypothesis H_0. Depending on the research question, the alternative hypothesis can take one of the following forms:

- **Two-sided (non-directional) alternative:** $H_1 : \mu \neq \mu_0$
  This is used when we are interested in detecting any significant difference from the specified value μ_0, regardless of direction. For a two-sided test, the critical region is split between both tails of the t-distribution. The rejection region is determined by the quantiles $t_{1-\alpha/2,n-1}$ and $t_{\alpha/2,n-1}$, where α is the significance level of the test. Specifically, we reject H_0 if the test statistic t falls outside the interval $[-t_{1-\alpha/2,n-1}, t_{1-\alpha/2,n-1}]$.

- **Right-sided (upper-tailed) alternative:** $H_1 : \mu > \mu_0$
  This is used when we want to test if the true mean is greater than μ_0. For a right-sided test, we reject H_0 if the test statistic t is greater than the critical value $t_{1-\alpha,n-1}$, which corresponds to the upper tail of the t-distribution.
- **Left-sided (lower-tailed) alternative:** $H_1 : \mu < \mu_0$
  This is used when we are interested in testing if the true mean is less than μ_0. For a left-sided test, we reject H_0 if the test statistic t is less than the critical value $t_{\alpha,n-1}$, which corresponds to the lower tail of the t-distribution.

The choice of alternative hypothesis affects both the critical values used and how the test is performed. In a two-sided test, since we are concerned with deviations in both directions, the significance level α is split between the two tails of the distribution, resulting in a more conservative test compared to a one-sided test where all of α is allocated to one tail.

Bonus: Type II Error in One Sample t-test
Type II error occurs when we fail to reject the null hypothesis H_0 when it is false. If the true mean is μ_1 (under the alternative hypothesis), the Type II error probability β can be derived using the fact that the distribution of the test statistic under μ_1 is also a Student's t-distribution with $n-1$ degrees of freedom. The new test statistic under the alternative hypothesis is:

$$t' = \frac{\bar{x} - \mu_1}{\frac{s}{\sqrt{n}}}.$$

By calculating the probability that t' falls within the acceptance region of the original t-distribution, we can determine β.

2 Two Samples t-test

Consider two independent samples, $x_{11}, x_{21}, \ldots, x_{n_a1}$ and $x_{12}, x_{22}, \ldots, x_{n_b2}$, which are continuous valued with unknown and unequal variances. The samples are either normally distributed or n_a, n_b are large.
The null hypothesis H_0 is that the true expectations (means) μ_1 and μ_2 of the random distributions from which the x_{i1} and x_{i2} values are drawn are equal, i.e.,

$$H_0 : \mu_1 = \mu_2.$$

The test statistic for the Welch's t-test is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_a} + \frac{s_2^2}{n_b}}},$$

where $\bar{x}_1$ and $\bar{x}_2$ are the sample means, s_1^2 and s_2^2 are the sample variances, and n_a and n_b are the sample sizes.
Under H_0, the test statistic follows a Student's t-distribution with Welch-Satterthwaite degrees of freedom:

$$\mathrm{df} \approx \frac{\left(\frac{s_1^2}{n_a} + \frac{s_2^2}{n_b}\right)^2}{\frac{\left(\frac{s_1^2}{n_a}\right)^2}{n_a-1} + \frac{\left(\frac{s_2^2}{n_b}\right)^2}{n_b-1}}.$$

Python. `scipy.stats.ttest_ind(sample1, sample2, equal_var=False)`

Equal Variance Case
If the variances of the two samples are assumed to be equal, the test statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}},$$

where s_p^2 is the pooled sample variance given by:

$$s_p^2 = \frac{(n_a - 1)s_1^2 + (n_b - 1)s_2^2}{n_a + n_b - 2}.$$

Under H_0, the test statistic follows a Student's t-distribution with $n_a + n_b - 2$ degrees of freedom.

Python. `scipy.stats.ttest_ind(sample1, sample2, equal_var=True)`

Remark
The assumption of equal variances should be reasonable based on the nature of the samples or should be verified. One common method to verify this assumption is using a Fisher test (F-test), which compares the ratio of the two sample variances. The test statistic for the F-test is:

$$F = \frac{s_1^2}{s_2^2},$$

where s_1^2 and s_2^2 are the sample variances of the two groups.
The F-test follows an F-distribution with $n_a - 1$ and $n_b - 1$ degrees of freedom. If the F-test is significant, it suggests that the variances are not equal, and hence the Welch's t-test should be preferred.

Drawback of the F-test
Performing an F-test adds an additional layer of testing, which increases the complexity of the analysis. Furthermore, the F-test itself is sensitive to deviations from normality, which might lead to incorrect conclusions about the equality of variances.

Bonus: Type II Error in Two Samples t-test
For the two-sample case, the Type II error probability β can be derived when the true means are μ_1 and μ_2 under the alternative hypothesis. The test statistic under this scenario will still follow a Student's t-distribution with the same degrees of freedom, and β can be calculated by finding the probability that the test statistic falls within the acceptance region of the original t-distribution.

3 Bonus: What if we have more than two groups?

Option 1: Multiple Pairwise Tests with Bonferroni Correction

The Bonferroni correction adjusts the significance level α by dividing it by the number of comparisons m. For k groups, $m = \frac{k(k-1)}{2}$ pairwise comparisons are made, and the adjusted significance level is α/m.

Option 2: Analysis of Variance (ANOVA)

ANOVA is used to determine if there are any statistically significant differences between the means of k groups.

- **Between-group variability (MSB):**

$$\text{MSB} = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{k-1}$$

- **Within-group variability (MSW):**

$$\text{MSW} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{N-k}$$

- **Test Statistic:**

$$F = \frac{\text{MSB}}{\text{MSW}}$$

- **Sampling Distribution:** Under H_0, F follows an F-distribution with $(k-1, N-k)$ degrees of freedom.