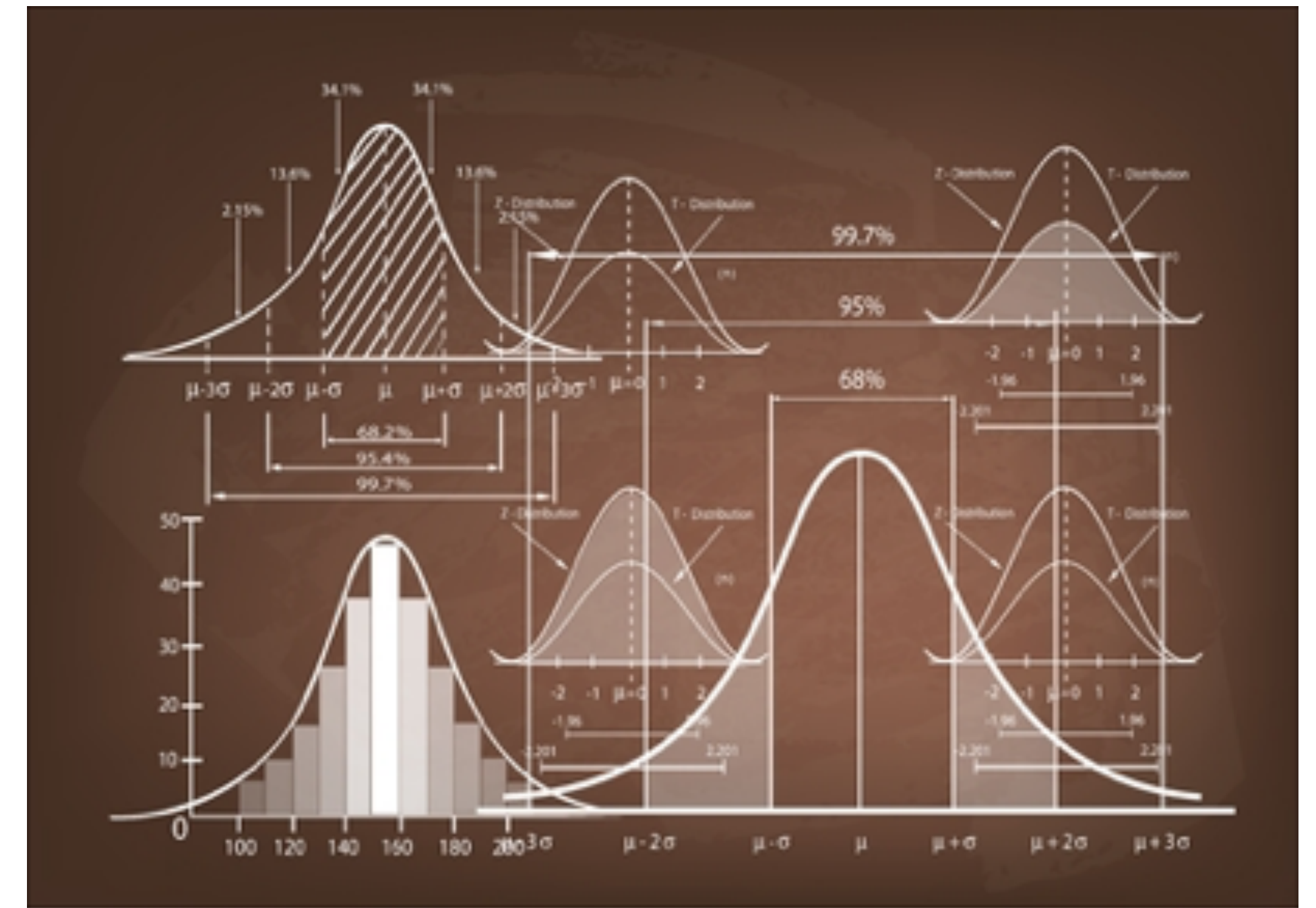


Business Statistics

MLS 3 - Common statistical tests





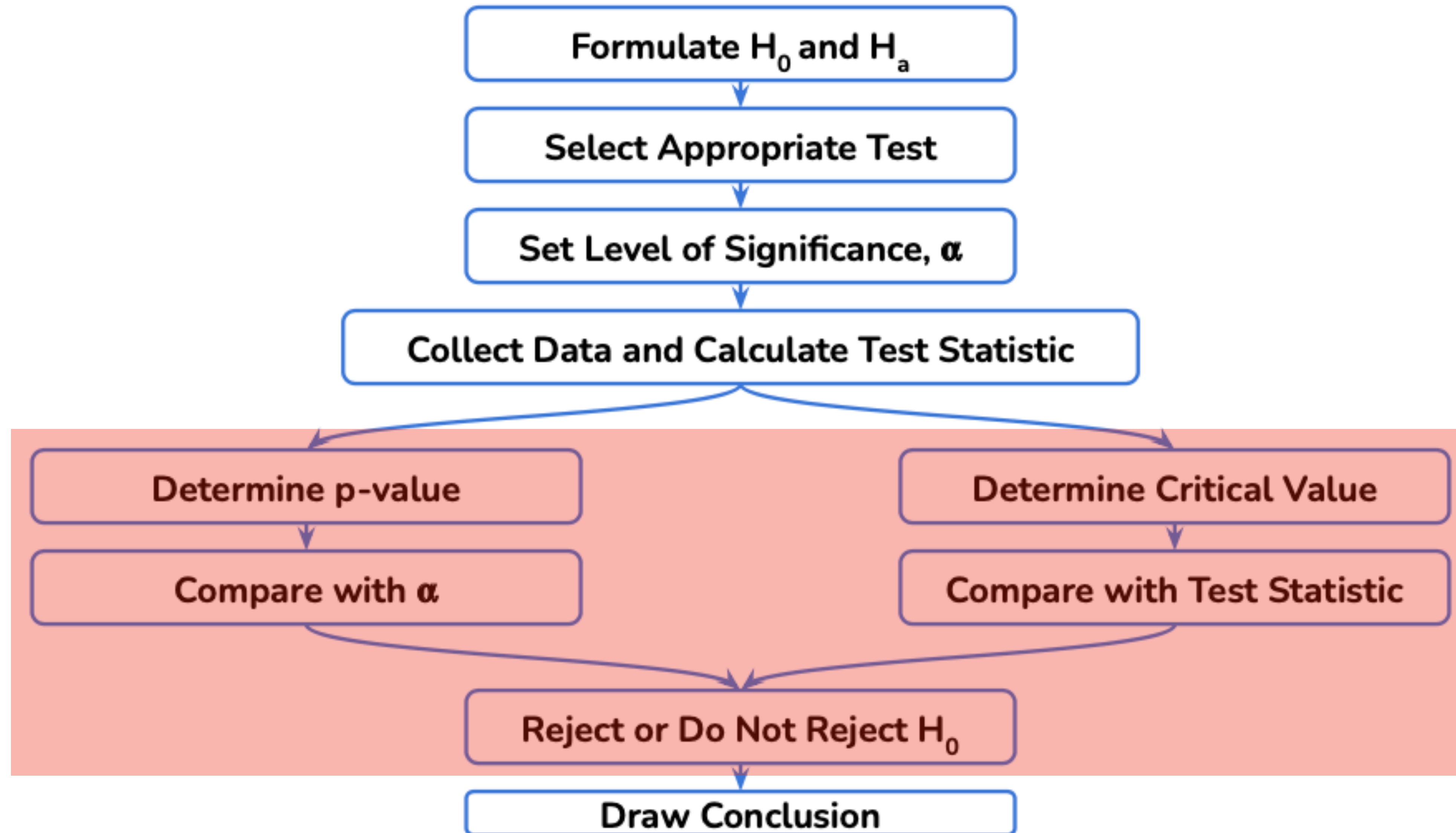
Session objectives

Mentored Learning Session 3: Common statistical tests

Learning Objectives

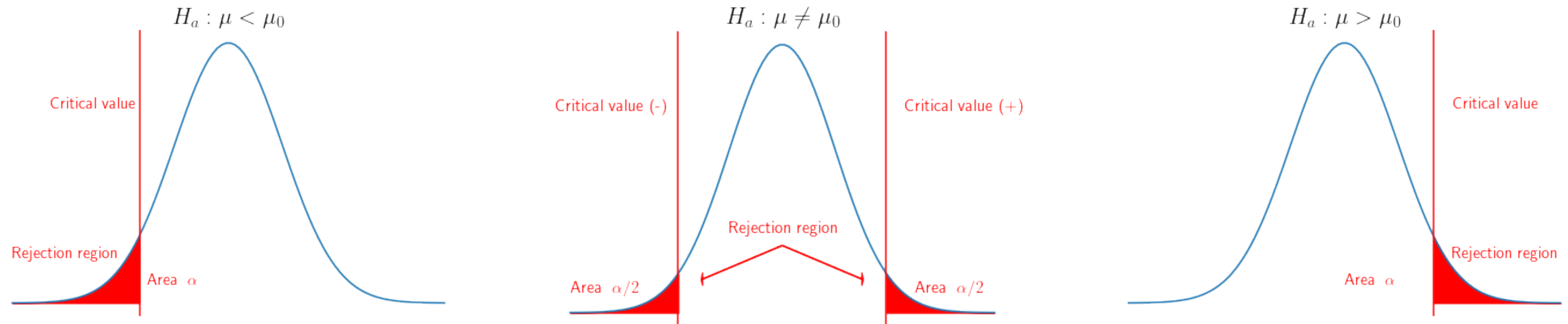
- Review and better understand the basics of hypothesis testing
- Learn about some of the standard tests used for hypothesis testing
- **Hands-on case Studies**
 - **Titan insurance**
 - **Mobile Internet**
 - **Diest Case study**

Hypothesis testing steps



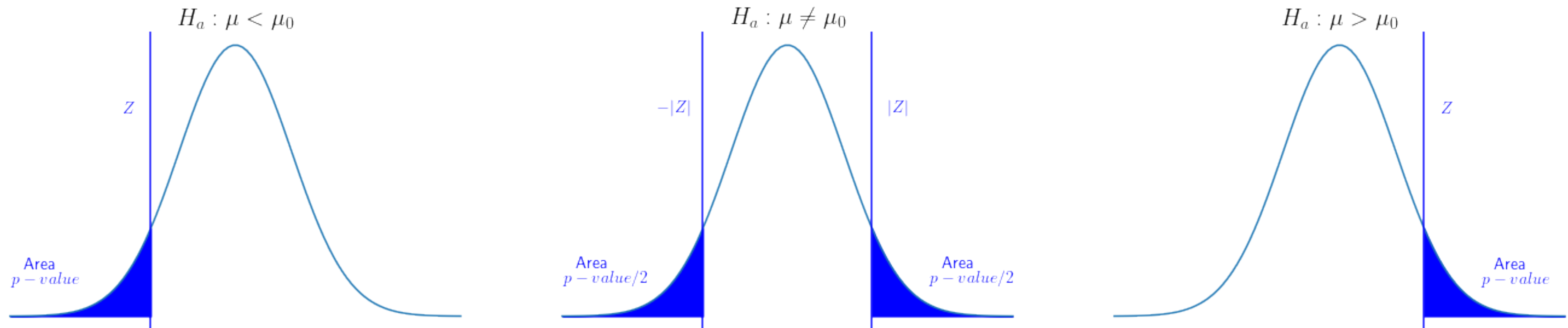
Hypothesis testing: rejection region approach

- In the **rejection region approach**, we define a region whose total area is equal to the significance level α (in red below)
- The location of the rejection region depends on the alternative hypothesis H_a
- **Decision of the test:** reject the null hypothesis H_0 when the test statistic lies in the rejection region



Hypothesis testing: p-value approach

- In the **p-value approach**, we calculate the likelihood (*p-value*) of the test statistic Z given the assumption of the null hypothesis H_0
- Low *p-values* are obtained for *extreme* test statistics with respect to H_0
- The area used to compute the *p-value* depends on the alternative hypothesis H_a (in blue below)
- **Decision of the test:** reject the H_0 when *p-value* < α



Hypothesis testing: An example

- Null hypothesis H_0

$$\mu \geq 0$$

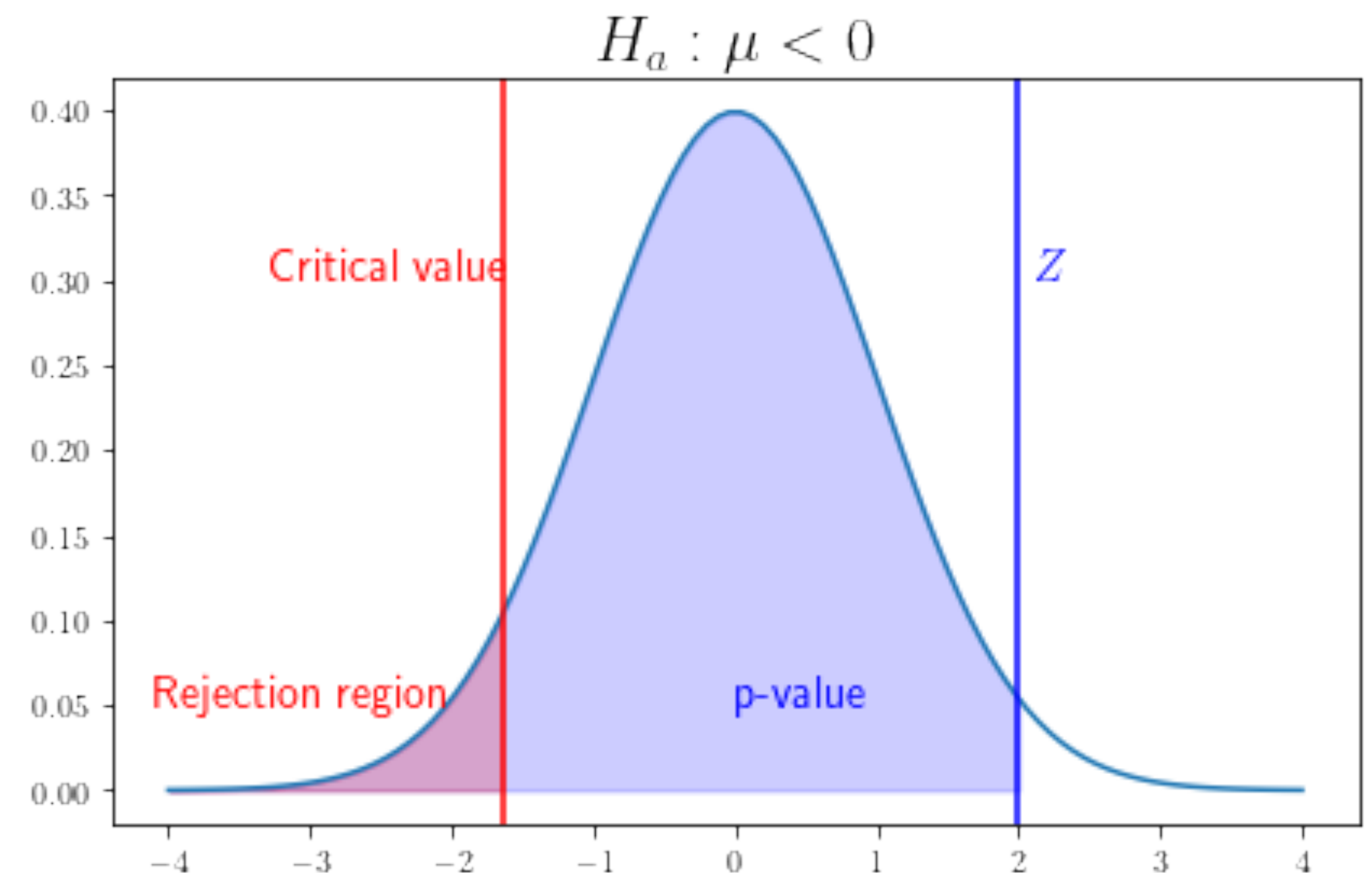
- Alternative hypothesis H_a

$$\mu < 0$$

- Observed test statistic: $Z = 2$

Decision:

- Rejection region: the test statistic is not in the rejection region, hence we do not reject H_0
- P-value: the blue area (p-value) is larger than the red area (level of significance), hence we do not reject H_0



Standard hypothesis tests

Tests for the means

Test	Usage	Python
1-sample z-test	Compare the sample mean to the population mean when std is known or $n > 30$	<code>statsmodels.stats.weightstats.ztest</code>
1-sample t-test	Compare the sample mean to the population mean when std is unknown and $n < 30$	<code>scipy.stats.ttest_1samp</code>
2-sample ind. z-test	Compare the sample means for 2 independent samples when their std are known	<code>statsmodels.stats.weightstats.ztest</code>
2-sample ind. t-test	Compare the sample means for 2 independent samples when their std are unknown	<code>scipy.stats.ttest_ind</code>
Paired t-test	Compare the sample means for 2 dependent samples when their std are unknown	<code>scipy.stats.ttest_rel</code>
Anova test	Compare the sample means for 2 or more independent normally distributed populations with equal variances	<code>scipy.stats.f_oneway</code>
Tuckey'sHSD	Pairwise sample means comparison test for 2 or more independent normally distributed populations	<code>statsmodels.stats.multicomp.pairwise_tuckeyhsd</code>

Standard hypothesis tests

Test for normality

Test	Usage	Python
1-sample z-test	Check the normality of the sample	scipy.stats.shapiro

Tests for variances

Test	Usage	Python
Leven’s test	Asses the equality of variances of a variable for 2 or more groups	scipy.stats.levene
Chi-square test	Compare the sample variance to the known normally distributed population variance	Using the SciPy.stats.chi2 distribution with the calculated test statistics
F-test	Compare the sample variances for 2 independent and normally distributed populations	Using the SciPy.stats.F distribution with the calculated test statistics

Tests for proportions and frequencies

Test	Usage	Python
1-sample z-test	Compare the sample proportion to the population proportion in the normality approximation	statsmodels.stats.proportion.proportions_ztest
2-sample z-test	Compare the sample proportions from two populations in the normality approximation	statsmodels.stats.proportion.proportions_ztest
chi-Square test of independence	Check whether the categorical variables from a population are independent	scipy.stats.chi2_contingency

A flowchart to help you choose

