

# Distribution, significance & statistical power

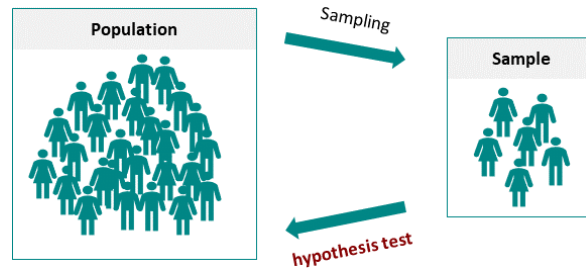
By Cristina Domínguez

## Understanding Hypothesis Testing

**Statistical hypothesis testing** is used to test an assumption regarding a population parameter. It is a form of statistical inference that uses data from a sample to draw a conclusion about a population.

In order to carry out a hypothesis testing, we need to follow a few steps:

- Formulate the two hypotheses, null and alternative ( $H_0$  and  $H_a$ ).
- Formulate a statistical analysis plan to evaluate the data.
- Analyse the sample data.
- Reject the null hypothesis, or state that the analysis failed to reject the null hypothesis.



The aim of hypothesis testing is to disprove the null hypothesis by computing a range (with a 95% confidence level) that would contain the experimental results, assuming that the null hypothesis is true.

A **test statistic** describes how closely the distribution of your data matches the expected/predicted distribution if the null hypothesis is true.

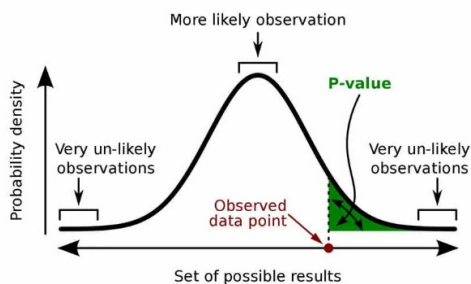
Different **statistical tests** predict different types of distributions, so it's important to choose the right statistical test for your hypothesis. For example, we would conduct an ANOVA test when comparing the variation between two or more groups, which uses a F-value.

Test statistic	Null and alternative hypotheses	Statistical tests that use it
<b>t value</b>	<b>Null:</b> The means of two groups are equal <b>Alternative:</b> The means of two groups are not equal	<ul style="list-style-type: none"> <li>• T test</li> <li>• Regression tests</li> </ul>
<b>z value</b>	<b>Null:</b> The means of two groups are equal <b>Alternative:</b> The means of two groups are not equal	<ul style="list-style-type: none"> <li>• Z test</li> </ul>
<b>F value</b>	<b>Null:</b> The variation among two or more groups is greater than or equal to the variation between the groups <b>Alternative:</b> The variation among two or more groups is smaller than the variation between the groups	<ul style="list-style-type: none"> <li>• ANOVA</li> <li>• ANCOVA</li> <li>• MANOVA</li> </ul>
<b><math>\chi^2</math>-value</b>	<b>Null:</b> Two samples are independent <b>Alternative:</b> Two samples are not independent (i.e., they are correlated)	<ul style="list-style-type: none"> <li>• Chi-squared test</li> <li>• Non-parametric correlation tests</li> </ul>

## Difference between Power and Confidence Level

### p-Value

The **p-value** is a number, calculated from a statistical test, that describes how likely you are to obtain an observed value (or more extreme) if the null hypothesis ( $H_0$ ) was true. P-value is used in hypothesis testing to help decide whether to reject the null hypothesis. The smaller the p-value, the more likely you are to reject the null hypothesis and the more statistically relevant your results will be.



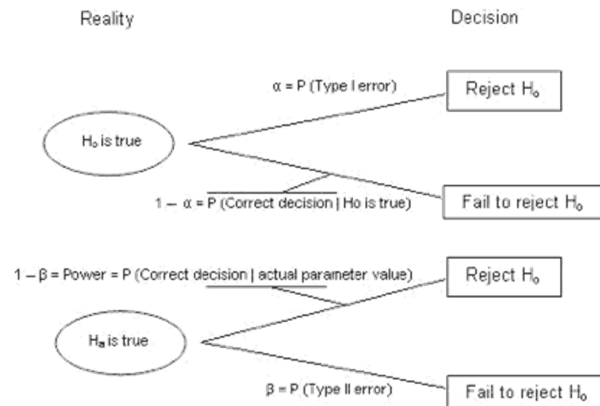
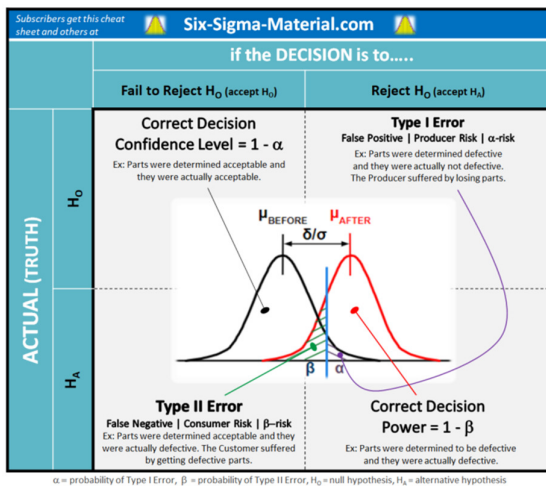
Values of p	Inference
$p > 0.10$	No evidence against the null hypothesis.
$0.05 < p < 0.10$	Weak evidence against the null hypothesis
$0.01 < p < 0.05$	Moderate evidence against the null hypothesis
$0.05 < p < 0.001$	Good evidence against null hypothesis.
$0.001 < p < 0.01$	Strong evidence against the null hypothesis
$p < 0.001$	Very strong evidence against the null hypothesis

## Significance ( $\alpha$ ) & Power ( $1-\beta$ )

**Hypothesis testing** helps us automate decision making about whether the experimental data supports our predictions, with null and alternative hypothesis. As a rule, what we want to demonstrate is placed as the alternative hypothesis (eg. There is a significant difference between the groups). Risks are unavoidable when making statistical decisions. The probability of making a **Type I error** is the significance level, or alpha ( $\alpha$ ), while the probability of making a **Type II error** is beta ( $\beta$ ).

**Significance level ( $\alpha$ )** is the probability that we reject the null hypothesis when it is true, also known as “false positive”. The **confidence level ( $1 - \alpha$ )** is the probability of obtaining the same results if the tests were repeated over and over again (eg. Probability to fail to reject a true null hypothesis).

**Beta ( $\beta$ )** is the probability of rejecting the null hypothesis while it is false, also known as “false negative”. The **power** of a hypothesis test is  $1 - \beta$ . If the power is close to 1, the hypothesis test is highly capable of detecting and rejecting a false null hypothesis. Beta is commonly set at 0.2. Power may be altered by significance level ( $\alpha$ ), sample size, variance and effect size. A test with a 95% confidence level means that there is a 5% chance of getting a type 1 error.



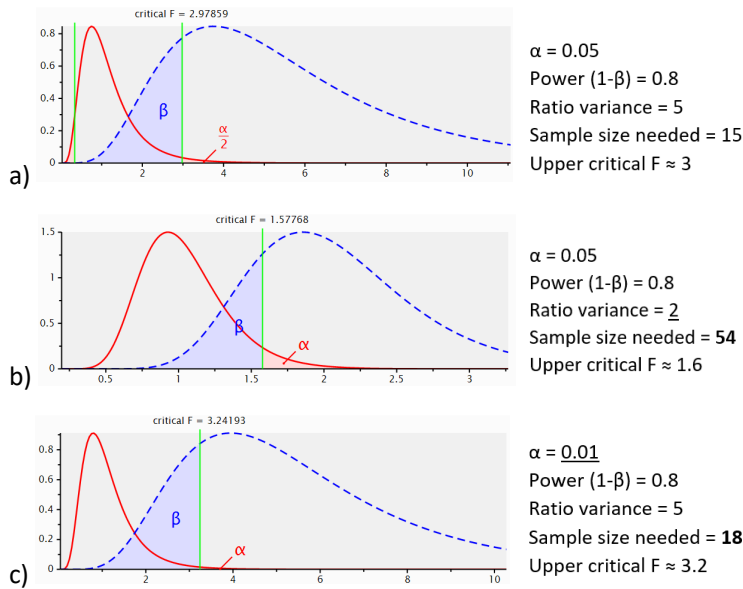
## F-test / F distribution

When we run a hypothesis test, often we see a **bell curve** (also called **standard normal distribution** or Gaussian distribution). However, the type of test that we use will determine the distribution of the data.

For example, the **F-distribution** is a ratio of variances, but it can also be used to determine whether several group means are equal. This probability distribution is nonsymmetric since it is skewed to the right. When  $F=1$ , the group means are equal. The **critical F** is the value of the probability  $\alpha$  of wrongly rejecting a true null hypothesis. If the obtained F value is equal or larger than the critical F value, then the result is significant at that level of probability (i.e.,  $F_{observed} > F_{critical}$  is equivalent to  $pval < \alpha$ ).

Some parameters (such as  $\alpha$  and  $\beta$ ) must be set before conducting the experiment, while others, such as the p-value, are obtained afterwards. In the curves below, obtained through the software G\*Power, we defined a significance level of 5% and a statistical power of 80%.

By varying these parameters, we can study how the shape of the curve, the sample size and the critical F value change. Decreasing the ratio variance increases the sample size needed and decreases the critical F value (see curve b). Also, the smaller the significance, the greater the sample size needed (see curve c).



**F-test in one-way ANOVA**

Lee Ds, Kim YK and Jung YW. Simvastatin, sildenafil and their combination in monocrotaline-induced pulmonary arterial hypertension. *Korean Circulation Journal*, 2010; 40: 658- 664.

The study was designed to investigate single and **combination therapy** of **simvastatin** and **sildenafil** in monocrotaline (MCT)-induced **PAH** (pulmonary arterial hypertension). One of the parameters studied to evaluate the efficacy of the treatments was the systolic pressure in the right ventricle.

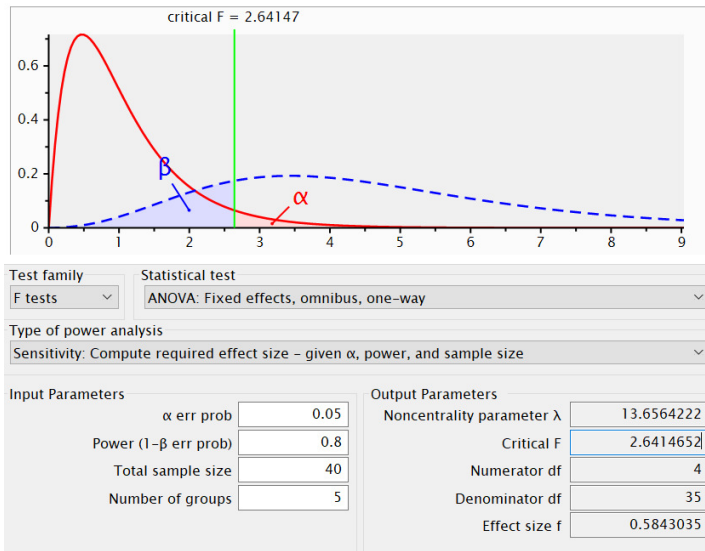
For this study, **ANOVA** (F-test in one-way ANOVA) is used to assess and compare several treatments. The null hypothesis is that none of the treatments have a significant therapeutic effect.

1-way ANOVA:  $X_{ij} = \mu + \alpha_j + \varepsilon_{ij}$

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_K$$

$$H_1 : \exists i, j \mid \alpha_i \neq \alpha_j$$

RVP increased in MCT-induced PAH group compared to the control. RVP suppressed in both single and combined treatment groups. However, combination therapy was more effective than single therapy with simvastatin and sildenafil.



RVP values (mmHg) ( $p < .05$ )	
Control	22.2 ± 0.7
MCT-induced	51.5 ± 1.5
MCT + Sim	32.4 ± 1.2
MCT + Sil	31.4 ± 1.0
MCT + Sim + Sil	33.4 ± 0.7

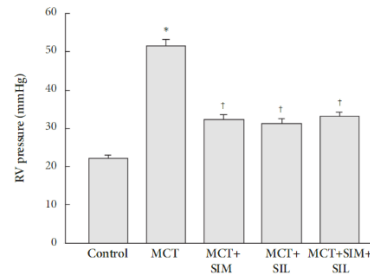


Fig. 1. Right systolic ventricle pressure. Simvastatin (SIM), sildenafil (SIL) and combination therapy (SIM+SIL) prevents the development of pulmonary arterial hypertension in monocrotaline (MCT) treated rats. However, combination therapy does not show additive effects over monotherapy. Values are expressed as mean±SD. \* $p < 0.05$  vs. control, † $p < 0.05$  vs. MCT. RV: right ventricle.

ANOVA can be used to compute the **effect size** and the **sample size**. In this study,  $\alpha$ , power and sample size (5 groups,  $n=40$ ) were given, so we calculated the effect size. There are different ways to measure the effect size:

Effect size measure	Small effect size	Medium effect size	Large effect size	Very large effect size
Odds ratio	1.5	2.5	4	10
Cohen's <i>d</i> (or one of its variants)	0.20	0.50	0.80	1.30
<i>r</i>	0.10	0.30	0.50	0.70
Cohen's <i>f</i>	0.10	0.25	0.40	—
Eta-squared	0.01	0.06	0.14	—

<sup>a</sup>Cohen, 1992, 1988; Rosenthal, 1996.

“Effect size *f*”, also called Cohen’s *f*, is a ratio between the sum of squares (sum of the squared differences between the independent observations and the group mean). ANOVA interprets *f* as being a small size effect ( $\approx 0.10$ ), medium size effect ( $\approx 0.25$ ) and large size effect ( $\approx 0.40$ ). Our calculated Cohen’s *f* value can be interpreted as being a large size effect ( $0.58 > 0.40$ ).

Previously, we had rejected the null hypothesis and affirmed that the treatments were different. The effect size *f* helps us quantify the size of the difference between treatments.

### Website guide

Introduction to Statistical Hypothesis Testing: <https://data-flair.training/blogs/hypothesis-testing-in-r/>

Alpha and beta risks: <https://www.six-sigma-material.com/Alpha-and-Beta-Risks.html>

F-Test: <https://www.six-sigma-material.com/F-Test.html>

T-test, ANOVA and Chi Squared: <https://www.youtube.com/watch?v=iJeEYFnS2v4>

Article used :

[https://www.researchgate.net/publication/49785806\\_Simvastatin\\_Sildenafil\\_and\\_Their\\_Combination\\_in\\_Monocrotaline\\_Induced\\_Pulmonary\\_Arterial\\_Hypertension](https://www.researchgate.net/publication/49785806_Simvastatin_Sildenafil_and_Their_Combination_in_Monocrotaline_Induced_Pulmonary_Arterial_Hypertension)