

Interpreting Significant Results

Prerequisites

[Introduction to Hypothesis Testing](#), [Statistical Significance](#), [Type I and II Errors](#), [One and Two-Tailed Tests](#)

When a [probability value](#) is below the α level, the effect is *statistically significant* and the [null hypothesis](#) is rejected. However, not all statistically significant effects should be treated the same way. For example, you should have less confidence that the null hypothesis is false if $p = 0.049$ than $p = 0.003$. Thus, rejecting the null hypothesis is not an all-or-none proposition.

If the null hypothesis is rejected, then the alternative to the null hypothesis (called the [alternative hypothesis](#)) is accepted. Consider the *one-tailed test* in the [James Bond](#) case study: Mr. Bond was given 16 trials on which he judged whether a martini had been shaken or stirred and the question is whether he is better than chance on this task. The null hypothesis for this one-tailed test is that $\pi \leq 0.5$ where π is the probability of being correct on any given trial. If this null hypothesis is rejected, then the alternative hypothesis that $\pi > 0.5$ is accepted. If π is greater than 0.50 then Mr. Bond is better than chance on this task.

Now consider the two-tailed test used in the [Physicians' Reactions](#) case study. The null hypothesis is:

$$\mu_{\text{obese}} = \mu_{\text{average}}.$$

If this null hypothesis is rejected, then there are two alternatives:

$$\begin{aligned}\mu_{\text{obese}} &< \mu_{\text{average}} \\ \mu_{\text{obese}} &> \mu_{\text{average}}.\end{aligned}$$

Naturally, the direction of the sample means determines which alternative is adopted. If the sample mean for the obese patients is significantly lower than the sample mean for the average-weight patients, then one should conclude that the population mean for the obese patients is lower than the sample mean for the average-weight patients.

There are many situations in which it is very unlikely two conditions will have exactly the same population means. For example, it is practically impossible that aspirin and acetaminophen provide exactly the same degree of pain relief. Therefore, even before an experiment comparing their effectiveness

is conducted, the researcher knows that the null hypothesis of exactly no difference is false. However, the researcher does not know which drug offers more relief. If a test of the difference is significant, then the direction of the difference is established. This point is also made in the section on the relationship between [confidence intervals and significance tests](#).

Optional

Some textbooks have incorrectly stated that rejecting the null hypothesis that two population means are equal does not justify a conclusion about which population mean is larger. Instead, they say that all one can conclude is that the population means differ. The validity of concluding the direction of the effect is clear if you note that a two-tailed test at the 0.05 level is equivalent to two separate one-tailed tests each at the 0.025 level. The two null hypotheses are then

$$\begin{aligned}\mu_{\text{obese}} &\geq \mu_{\text{average}} \\ \mu_{\text{obese}} &\leq \mu_{\text{average}}.\end{aligned}$$

If the former of these is rejected, then the conclusion is that the population mean for obese patients is lower than that for average-weight patients. If the latter is rejected, then the conclusion is that the population mean for obese patients is higher than that for average-weight patients.