

Interpreting Non-Significant Results

Prerequisites

[Introduction to Hypothesis Testing](#), [Significance Testing](#), [Type I and II Errors](#)

When a significance test results in a high probability value, it means that the data provide little or no evidence that the null hypothesis is false. However, the high probability value is not evidence that the null hypothesis is true. The problem is that it is impossible to distinguish a null effect from a very small effect. For example, in the [James Bond Case Study](#), suppose Mr. Bond is, in fact, just barely better than chance at judging whether a Martini was shaken or stirred. Assume he has a 0.51 probability of being correct on a given trial ($\pi = 0.51$). Let's say Experimenter Jones (who did not know $\pi = 0.51$) tested Mr. Bond and found he was correct 49 times out of 100 tries. How would the significance test come out? The experimenter's significance test would be based on the assumption that Mr. Bond has a 0.50 probability of being correct on each trial ($\pi = 0.50$). Given this assumption, the probability of his being correct 49 or more times out of 100 is 0.62. This means that the probability value is 0.62, a value very far higher than the conventional significance level of 0.05. This result, therefore, does not give even a hint that the null hypothesis is false. However, we know (but Experimenter Jones does not) that $\pi = 0.51$ and not 0.50 and therefore that the null hypothesis is false. So, if Experimenter Jones had concluded that the null hypothesis were true based on the statistical analysis, he or she would have been mistaken. Concluding that the null hypothesis is true is called *accepting the null hypothesis*. To do so is a serious error.

[Binomial Calculator](#)

[Further argument for not accepting the null hypothesis](#)

**Do not accept the null hypothesis
when you do not reject it.**

So how should the non-significant result be interpreted? The experimenter should report that there is no credible evidence Mr. Bond can tell whether a martini was shaken or stirred, but that there is no proof that he cannot. It is generally impossible to prove a negative. What if I claimed to have been Socrates in an earlier life? Since I have no evidence for this claim, I would have great difficulty convincing anyone that it is true. However, no one would be able

to prove definitively that I was not.

Often a non-significant finding increases one's confidence that the null hypothesis is false. Consider the following hypothetical example. A researcher develops a treatment for anxiety that he or she believes is better than the traditional treatment. A study is conducted to test the relative effectiveness of the two treatments: 20 subjects are randomly divided into two groups of 10. One group receives the new treatment and the other receives the traditional treatment. The mean anxiety level is lower for those receiving the new treatment than for those receiving the traditional treatment. However, the difference is not significant. The statistical analysis shows that a difference as large or larger than the one obtained in the experiment would occur 11% of the time even if there were no true difference between the treatments. In other words, the probability value is 0.11. A naive researcher would interpret this finding as evidence that the new treatment is no more effective than the traditional treatment. However, the sophisticated researcher, although disappointed that the effect was not significant, would be encouraged that the new treatment led to less anxiety than the traditional treatment. The data support the thesis that the new treatment is better than the traditional one even though the effect is not statistically significant. This researcher should have more confidence that the new treatment is better than he or she had before the experiment was conducted. However, the support is weak and the data are inconclusive. What should the researcher do? A reasonable course of action would be to do the experiment again. Let's say the researcher repeated the experiment and again found the new treatment was better than the traditional treatment. However, once again the effect was not significant and this time the probability value was 0.07. The naive researcher would think that two out of two experiments failed to find significance and therefore the new treatment is unlikely to be better than the traditional treatment. The sophisticated researcher would note that two out of two times the new treatment was better than the original treatment. Moreover, two experiments each providing weak support that the new treatment is better, when taken together, can provide strong support. Using a method for combining probabilities, it can be determined that combining the probability values of 0.11 and 0.07 results in a probability value of 0.045. Therefore, these two non-significant findings taken together result in a significant finding.

Although there is never a statistical basis for concluding that an effect is exactly zero, a statistical analysis can demonstrate that an effect is most likely

small. This is done by computing a [confidence interval](#). If all effect sizes in the interval are small, then it can be concluded that the effect is small. For example, suppose an experiment tested the effectiveness of a treatment for insomnia. Assume that the mean time to fall asleep was 2 minutes shorter for those receiving the treatment than for those in the control group and that this difference was not significant. If the 95% confidence interval ranged from -4 to 8 minutes, then the researcher would be justified in concluding that the benefit is eight minutes or less. However, the researcher would not be justified in concluding the null hypothesis is true, or even that it was supported.