

Comparing Alpha and the False Alarm Rate

Part 1. Alpha is not the False Alarm Rate

To get the most out of this activity, you should be somewhat familiar with these concepts related to statistical hypothesis testing: null hypothesis, alternative hypothesis, alpha, t -test, statistical significance, p -value, Type I error, statistical power.

Suppose a researcher did a study to measure the effectiveness of a new treatment for anxiety. The researcher compared two similar groups of 40 people with anxiety. One group received the new treatment, and one group received a placebo treatment as a control. After gathering all of the data, the researcher did a standard t -test. The null hypothesis was that the new treatment has no effect. The alpha level for the test was .05. The researcher summarized the result like this: "Compared to the control group, the treatment group had significantly lower anxiety scores, $t(78) = 2.34, p < .05$. This result suggests that the new treatment is effective."

There was a statistically significant difference, but how likely is it that the new treatment actually has no effect? In other words, **What is the probability the researcher's conclusion is a false alarm?** You might think the answer is 5%, because the alpha level was 5%. But, alpha is not the false alarm rate.

► Alpha is not the false alarm rate

Here are definitions for these two terms:

alphathe probability a **TRUE** null hypotheses will be **REJECTED**

false alarm ratethe probability a **REJECTED** null hypotheses is **TRUE**

It is important to understand why alpha is not the false alarm rate, because it improves your ability to make sense of the statistical information that is such an important component of scientific research.

► Finding the false alarm rate

Alpha is selected by the researcher before doing a statistical test. Researchers often use an alpha of .05. However, you can't just select a false alarm rate. It depends on alpha, plus two other factors.

The app for this activity will help you find the false alarm rate. To use the app, imagine that 1000 researchers each conducted an experiment to test new anxiety treatments. All of the researchers also conducted a t -test. For this example, set the menus to these three values: prior = .20, power = .40, alpha = .05.

Prior	<input type="text" value=".20"/>	Power	<input type="text" value=".40"/>	Alpha	<input type="text" value=".05"/>
-------	----------------------------------	-------	----------------------------------	-------	----------------------------------

Prior is an abbreviation for "the prior probability that there is an effect." This is an overall estimate of the proportion of tested anxiety treatments that actually work. Some of the treatments will work, others will not. Because we can estimate this probability *before* we see the results of any experiment, it is called the prior probability. For this example, assume 20% of the treatments that get tested actually work. In the bar graph, the gray bars represent the number of experiments in which the treatment has an effect (20% of 1000 = 200), and the 800 experiments where there is no effect. Of course, for any one experiment, we don't know if the treatment works or not—that's why we do the experiment. Setting the prior to 20% means that overall, we expect 20% of the treatments will have an effect.

Power means statistical power, the proportion of false null hypotheses that are rejected. Power is set at 40% in this example. If a treatment has an effect, there is a 40% chance that it will produce a statistically significant difference. In this example, the treatment has an effect in 200 experiments, so 80 of them (40% of 200) produce a statistically significant difference. These are called *correct rejections* because the null hypothesis, which is false, has been correctly rejected. The blue bar shows the 80 correct rejections.

Alpha is the proportion of true null hypotheses that are rejected (the Type I error rate). Alpha is set at 5% in this example. There are 800 experiments in which the null hypothesis is true (the treatment doesn't work) and 40 of them (5% of 800) produce a statistically significant difference. The orange bar shows these 40 Type I errors.

As the pie chart shows, there are 120 statistically significant results, 80 correct rejections and 40 Type I errors. So one-third (40/120) of the statistically significant differences are Type I errors. Because these are results that might lead a researcher to conclude that an ineffective treatment works, they are called false alarms (or false positives). In this example, the false alarm rate is 33%, much higher than the alpha level of 5%.

► **Questions:** Try to answer these questions from memory, then refer to this handout if necessary.

1. Alpha is the probability a _____ null hypothesis will be rejected.
2. The false alarm rate is the probability a _____ null hypothesis _____.
3. When you see " $p < .05$ ", let the p remind you of the other two values needed to estimate the false alarm rate, _____ and _____.
4. Suppose you read a study that found a "statistically significant" difference ($p < .05$).
 - (a) Does this mean there is a 5% chance the result is a false alarm? Explain.
 - (b) An alpha level of .05 means there is a 5% chance that _____.
5. In the anxiety treatment example, the prior of .20 means that 20% of the treatments _____.
6. In the anxiety treatment example, the false alarm rate was 33%. What does this mean?

List any aspects of this activity that you would like the instructor to clarify.

Comparing Alpha and the False Alarm Rate

Part 2. Changing the False Alarm Rate

In Part 1 of this activity you learned why alpha is not the false alarm rate. Now you will investigate how the false alarm rate changes when you change the prior, power, and alpha.

Prior	<input type="text" value=".20"/>	Power	<input type="text" value=".40"/>	Alpha	<input type="text" value=".05"/>
-------	----------------------------------	-------	----------------------------------	-------	----------------------------------

► Change the prior

1. Set the app menus to: prior = .20, power = .40, alpha = .05. Now, gradually increase the prior from .20 to .90.
 - a. Power stays the same, so why does the number of correct rejections rise (the blue bar)?
 - b. Alpha stays the same, so why does the number of false alarms drop (the orange bar)?
 - c. Is there a direct or an inverse relationship between the prior and the false alarm rate?

► Change power

2. Set the menus to: prior = .50, power = .10, alpha = .05. Now, gradually increase power from .10 to .40.
 - a. Why is the blue bar changing, but not the orange bar?
 - b. Using the anxiety treatment experiments from Part 1, explain what a correct rejection means.
 - c. First predict, then verify with the app, the number of correct rejections when you change power to 50%, and then 70%. Were your predictions correct? If not, why?
 - d. Is there a direct or an inverse relationship between power and the false alarm rate?

► Change alpha

3. Set the menus to: prior = .20, power = .60, alpha = .05.
 - a. Explain what a 25% false alarm rate means? What does a 5% alpha mean?
 - b. Predict the number of correct rejections, the number of false alarms, and the false alarm rate, when you increase alpha to 10%. Were you right? If not, why?
 - c. Predict the number of correct rejections, the number of false alarms, and the false alarm rate, when you set alpha to 1%. Were you right? If not, why?
 - d. Is there a direct or an inverse relationship between alpha and the false alarm rate?

► Example study 1

Suppose you read a study that tested the effect of a new drug. The manufacturer claims the new drug can improve memory in elderly adults. Using an alpha level of .05, a researcher found a statistically significant improvement in memory in elderly adults who took the new drug (compared to a placebo control group).

4. This result would be a false alarm if _____.
5. To estimate the probability the outcome is a false alarm, what three values do you need to specify?

Because the researcher tested a large number of people, the estimated power of the statistical test was very high at .90. Also, it turns out that the drug is a homeopathic medication, and homeopathic medicine has never been shown to work, because it contains little or no active ingredients. This information leads you to estimate the prior probability at .01.

6. Based on this information (and before using the app), predict how likely is it that the research finding is a false alarm. Then use the app to verify your answer.
7. Is the alpha level of 5% similar to the estimated false alarm rate?

► Example study 2

Suppose a student conducted a study to see if the persistence of a behavior is affected by the pattern in which the behavior is rewarded. During 5 days of training, 10 rats were rewarded every time they pressed a switch. Another 10 rats were rewarded every tenth time they pressed the switch. Then the student stopped all rewards, and counted how many times each rat pressed the switch over the next 2 hours. The student found a statistically significant difference ($p < .05$) between the two groups. On average, the rats that were rewarded every tenth time were more persistent—they were still pressing the switch long after the rewards had stopped. The student concluded that partial reinforcement increases persistence. Is this result a false alarm?

The "partial reinforcement" effect is well-known, and has been found in many experiments. For this reason, you estimate the prior probability to be high, 70%. Also assume the study had a lot of statistical power, 70%. Finally, the student used an alpha level of 5% for the statistical test.

8. Predict how likely is it that the research finding is a false alarm. Then use the app to verify your answer.
9. Is the alpha level of 5% similar to the estimated false alarm rate?

► Review Questions

10. Alpha is the probability that a _____ null hypothesis will be _____.
11. The false alarm rate is the probability that a _____ null hypothesis is _____.
12. List three different changes that will decrease the false alarm rate.

► Final Reminders

Alpha is not the probability that a statistically significant result is a false alarm. The false alarm rate is often much higher than alpha. When you see something like " $p < .05$ ", let the p remind you of the **prior** and **power**, the other two values needed to estimate the probability the finding is a false alarm.