

What Is an Alpha-Level?

Donald J. Wheeler

One thing burned into the brains of those who survive a statistics class is that you have to specify an alpha-level before you do anything statistical. And when it comes to statistical inference they are correct. But just what does the alpha-level represent? What does it mean in practice? Read on to find out.

THE DECISION PROBLEM

With experimental data, the essence of statistical analysis is the decision regarding the relationship between a factor and some response variable. For example, if changes in the level of Factor A do not result in changes in the level of the Response Variable Y , then Factor A does not have an effect upon Variable Y . (This is usually called the Null State.) If changes in Factor A do result in changes in the level of Variable Y , then Factor A does have an effect upon the Response Variable. (Call this the Alternate State.) The experimenter must decide which state of nature exists. Thus we have the decision matrix shown in Figure 1. On the left side we have the two possible decisions we can make and across the top we have the two states of nature. The resulting two-by-two grid contains two correct decisions and two incorrect decisions.

DECISION MADE	POSSIBLE STATES OF NATURE	
	Null State Factor A has no effect upon Response Y	Alternate State Factor A affects Response Y
1. Factor A affects Y	False Alarm	Correct Decision
2. Factor A does not affect Y	Correct Decision	Missed Signal

Figure 1: Consequences of Different Decisions

A study will be said to be *conservative* when the analysis is performed in such a way that there should be very few false alarms. Here it is more important to be right than to find all of the signals, so we want to be confident that we are likely to be correct when we say Factor A affects the Response Y .

Given a finite amount of data, the only way to minimize false alarms will be to minimize the number of times that Decision One is made. By avoiding the first decision, we will automatically minimize the opportunities for a false alarm. A conservative analysis is appropriate for experimental data when confirmation from subsequent experiments is not readily available.

A study will be said to be *exploratory* when the analysis is performed in such a way that there

should be few missed signals. Here it is more important to find all of the signals than to avoid false alarms, so we want to be confident that we are likely to be correct when we say that Factor A does not affect Response Y.

Given a finite amount of data, we obtain an exploratory analysis by traditionally using decision rules which encourage the first decision: "Factor A affects Y." By making Decision One more often, Decision Two will be made less often, and there will be fewer opportunities for a missed signal.

In order to characterize the four outcomes for the decision matrix in Figure 1 more completely, let θ represent the *a priori* probability that the Null State is true. This means that the *a priori* probability for the Alternate State will be $[1 - \theta]$.

$$\text{Probability that Null State is true} = \theta \quad [1]$$

$$\text{Probability that Alternate State is true} = 1 - \theta$$

Furthermore, under the condition that the Null State is true, let the probability of Decision One be denoted by α . This will result in a conditional probability for Decision Two of $[1 - \alpha]$.

$$\text{Probability of making Decision One given that Null State is true} = \alpha \quad [2]$$

$$\text{Probability of making Decision Two given that Null State is true} = 1 - \alpha$$

Finally, under the condition that the Alternate State is true, let the probability of Decision Two be denoted by β . This will result in a conditional probability for Decision One of $[1 - \beta]$.

$$\text{Probability of making Decision One given that Alternate State is true} = 1 - \beta \quad [3]$$

$$\text{Probability of making Decision Two given that Alternate State is true} = \beta$$

Using the three values of α , β , and θ we can characterize the probabilities of each of the four outcomes in the decision matrix as shown in Figure 2.

DECISION MADE	POSSIBLE STATES OF NATURE	
	Null State Factor A has no effect on Y $Pr\{null\} = \theta$	Alternate State Factor A affects Response Y $Pr\{alt\} = (1-\theta)$
1. Factor A affects Y	False Alarm $Pr\{False\ Alarm\} = \alpha\theta$	Correct Decision $Pr\{Dec.1\ is\ right\} = (1-\beta)(1-\theta)$
2. Factor A does not affect Y	Correct Decision $Pr\{Dec.2\ is\ right\} = (1-\alpha)\theta$	Missed Signal $Pr\{Missed\ Signal\} = \beta(1-\theta)$

Figure 2: The Probabilities of Different Outcomes Prior to Making a Decision

Before we make a decision we have four possible outcomes. Two of these are correct and two are incorrect. Thus, prior to making a decision, the probability of making an incorrect decision is:

$$Pr\{Incorrect\ Decision\} = Pr\{False\ Alarm\} + Pr\{Missed\ Signal\} = \alpha\theta + \beta(1-\theta) \quad [4]$$

Other considerations require that β must be less than $[1 - \alpha]$. Because of this, for any fixed amount of data, any increase in α will result in a decrease in β , and any decrease in α will result in an increase in β . This means that you cannot simultaneously reduce both terms in equation [4] except by collecting more data. However, equation [4] shows the rationale behind exploratory and conservative analyses.

If we believe the Null State is correct then θ will be large, and to avoid a false alarm we will want a small value for α .

If we believe the Alternate State is correct then $[1 - \theta]$ will be large and to avoid a missed signal we will want a small value for β . Since β must be less than $1 - \alpha$ we can always make β smaller by choosing a larger value for α .

If we must perform our analysis in such a way that we can convince a skeptic, then we will adopt the skeptic's position, assume that the Null State is true, and perform a conservative analysis using a small value for α .

If we have the luxury of easily confirming our decision when we make Decision One, so the problem of an occasional false alarm is not so severe, we might take a more exploratory approach and use a larger alpha-level in order to reduce the risk of missing a signal.

WHAT DOES THE ALPHA-LEVEL DO?

As noted in the last two statements, the alpha-level specifies our attitude to the analysis. We use the value of alpha to characterize our *approach* to the decision problem. *The value of alpha does not, by itself, characterize any aspect of the final results of our decision making process.*

Consider, for example, what happens if we make Decision One. If we make Decision One then we hope that the Alternate State is true, and according to the calculus of probabilities the probability of having made a correct decision is:

$$\text{Prob}\{\text{Alternate State is true given Decision One}\} = \frac{(1-\beta)(1-\theta)}{\alpha\theta + (1-\beta)(1-\theta)} \quad [5]$$

while the probability that we have made an incorrect decision is:

$$\text{Prob}\{\text{Null State is true given Decision One}\} = \frac{\alpha\theta}{\alpha\theta + (1-\beta)(1-\theta)} \quad [6]$$

On the other hand, when we make Decision Two we hope that the Null State is true, and the probability that we have made a correct decision becomes:

$$\text{Prob}\{\text{Null State is true given Decision Two}\} = \frac{(1-\alpha)\theta}{(1-\alpha)\theta + \beta(1-\theta)} \quad [7]$$

while the probability that we have made an incorrect decision is:

$$\text{Prob}\{\text{Alternate State is true given Decision Two}\} = \frac{\beta(1-\theta)}{(1-\alpha)\theta + \beta(1-\theta)} \quad [8]$$

The point of writing out these four formulas is to make it clear that none of these four probabilities is uniquely determined by the value of α . Therefore, when we call α the "significance level for the test" we are saying one thing while meaning something else. Whenever we specify the alpha-level for a decision we are not defining the probability that we will make the right or wrong decision, but

we are rather characterizing our attitude toward the analysis by specifying our willingness to risk a false alarm under the condition that the Null State is true.

In traditional inference techniques we have to specify a value for alpha in order to use a probability model to obtain critical values that we can use to make the decision. In finding these critical values we usually assume that:

- (1) the measurements do not display any discreteness,
- (2) the measurements come from a process that does not change over time, and
- (3) the measurements are normally and independently distributed.

Under these assumptions it is possible to obtain decision rules that will yield specific theoretical values for α and β . Unfortunately, actual data are always discrete, they are never normally distributed, and processes do change over time. However, given the necessity of filtering out at least some of the noise, and lacking any other guidelines on how to obtain a decision rule for separating potential signals from background noise, we use decision rules based upon the probability models. The probability models can be thought of as providing a theoretical *approximation* which can be used as a guide for practice. The decision rules obtained in this manner are probably not the optimum decision rules for any given real-world situation, but they at least have the property that they are logical and appropriate in the theoretical world, and they are reasonably efficient in practice.

Thus, in describing the decision rules based upon probability models, we have to specify a value for alpha. *In practice, the alpha-level for a procedure simply tells the reader how the experimenter defined the decision rule for interpreting the data.* Conservative analyses will use a small alpha-level and exploratory analyses will use a large alpha-level. Therefore, while the alpha-level does not actually give the probability of a false alarm, or define the confidence level for a set of results, it does rank the sensitivity of the analysis. For a given data set, larger alpha-levels will yield a more sensitive analysis while smaller alpha-levels will yield a less sensitive analysis. Conservative analyses will generally use a 1% alpha-level. Traditional analyses will generally use a 5% alpha-level, and exploratory analyses will generally use a 10% alpha-level.

In the industrial context, where today's experimental results can often be verified by tomorrow's production, most experimental studies should use the exploratory approach. With the exploratory approach you are less likely to miss any effects that might be of interest, and any false alarms will be quickly identified as such by the accumulation of subsequent data.

In those situations where subsequent experiments are difficult to arrange it is common to build the replication into the experiment by using multiple experimental units in a single study. This model is frequently used in agricultural and biomedical experiments. Here a traditional or conservative approach to analysis is appropriate.

HOW DOES THE ALPHA-LEVEL DIFFER FROM THE P-VALUE?

Say you have an F-test or a t-test. When performing this test manually you would compare the observed test statistic (F-observed) with a critical value (F-critical). The critical value would depend upon the alpha-level for your test. If the observed test statistic (F-observed) is more extreme than the critical value (F-critical) you would say that you have found a detectable effect at whatever alpha-level you were using. Here you would make Decision One. Otherwise you

would have failed to detect an effect and would make Decision Two.

Most software packages make the comparison above by reporting a p-value for the test. This p-value is the probability of exceedance for the test statistic. This means that the p-value is a transformed version of the observed test statistic (F-observed). Instead of comparing the observed test statistic with a critical value, the p-value becomes the test statistic and it is directly compared to the alpha-level chosen by the investigator. When the p-value is smaller than the alpha-level you make Decision One, and when the p-value is larger than the alpha-level you make Decision Two.

So, once again, the alpha-level defines the investigator's attitude to the analysis. It defines the user's cut-off between the potential signals and the probable noise. A p-value that is smaller than the chosen alpha-level will lead to Decision One, and a p-value that is larger than the chosen alpha-level will lead to Decision Two.

Neither the p-value nor the alpha-level define the significance (importance) of the result, they merely allow you to separate the potential signals from the probable noise. They allow you to identify the detectable signals. With a lot of data you may detect trivial signals, and with small amounts of data you may fail to detect important signals. This is why I do not use the language of "statistical significance." It is completely misleading and a continual source of confusion. To say that something is "statistically significant at a 5% alpha-level" simply means that you have *detected a potential signal* by using critical values that will give false alarm 5% of the time when the Null State is true. The actual probability that you have made the correct decision is given by equation [5] above. While a small value for alpha will tend to make equation [5] large, we cannot always compute the value for equation [5] in practice. So stop talking about "significant results" and start talking about "detectable signals." The resulting clarity will be appreciated by your clients.

