α Errors and β Errors

Masaru Okada

October 26, 2025

Contents

Statistical Significance

For example, in a large corporation, information that 'a factor causing a one-yen difference in annual sales has been found' cannot be said to be very meaningful.

> However, even if this difference of only one yen seems meaningless, if it is unlikely to have occurred by chance data variability, it is said to be statistically 'significant'.
>
> Typically, it is quite normal for differences in representative values, such as means or proportions, to arise between two divided groups, simply because they are different groups. But if that difference is large, such as 'two standard deviations ($\pm$ 2SD) or more', it may be said to be statistically significant.

Statistical Power

In reality, the means of groups that must be compared are rarely separated by as much as two standard deviations.

> If they were that far apart, one would likely notice the difference even without statistical analysis.

Therefore, the challenge is how to detect a statistically significant difference—one that is smaller than two standard deviations yet still practically meaningful—from a minimal amount of data. In other words, how can **statistical power** be increased?

> Statistical power is 'the probability of being able to correctly state that a significant difference exists, given that the hypothesis of some difference existing is true'.

The Error of the 'Alarmist'

Simply increasing statistical power is not the whole story.

- There is a simple, yet harmful, method to maximize power—that is, 'to be able to find a significant difference 100% of the time whenever any hypothesis of a difference is true'.
- This method involves 'continuously and irresponsibly asserting everything one thinks of, without any basis in data'.
- If the hypothesis happens to be correct, one would have successfully found a meaningful difference 100% of the time.
- The world (even in companies, on television, or in parliament) is full of people who plausibly assert baseless ideas; they are, in essence, creatures who only maximize statistical power.

The Error of the 'Alarmist'

The world (even in companies, on television, or in parliament) is full of people who plausibly assert baseless ideas; they are, in essence, creatures who only maximize statistical power.

> Mark Twain's famous quote: 'Even a broken clock is right twice a day'.
>
> An economic analyst who constantly predicts 'a recession is coming soon' will inevitably have said 'a recession is coming soon' in the year right before a recession actually arrives.

α Errors and β Errors

Maximizing statistical power is harmful.

The reason this approach is harmful is not only because of the error of 'missing a correct hypothesis'.

It fails to consider the opposite error: 'accepting a false hypothesis as true', in other words, claiming a difference exists when there is none.

In statistics, this error of 'claiming a difference exists when there is none' is called an α **error**.

On the other hand, the error of 'missing a difference that truly exists' is called a β **error**.

α Errors and β Errors

Most textbooks introduce these using a Japanese mnemonic that links the initial letters: the α error is called the '**a**larmist's error', and the β error is called the '**b**lockhead's error'.

> Based on this expression, the people mentioned earlier who spout baseless hypotheses can be described as being too 'alarmist' in order to reduce their risk of being a 'blockhead' (a β error) to zero.

Conversely, there is also a way to reduce the α error to zero, but this is harmful as well.

> This is the approach of only ever saying, 'Regardless of who asserts what, or on what basis, we cannot know for sure, so let's just continue to discuss it cautiously'.
>
> In essence, such people assert no hypotheses, nor do they act in belief of any hypothesis.
>
> Therefore, while their risk of being an 'alarmist' (an α error) is zero, they will continue to be 'blockheads' and dismiss any truth, even when it is presented right before their eyes.

The Error of the 'Blockhead'

In many of our decisions, failing to make the best judgment right here and now results in losses accumulating moment by moment.

If a doctor merely continues to observe a patient cautiously, most patients will eventually die.

Significance Level

The 'alarmist's' error and the 'blockhead's' error are in a trade-off relationship.

When dealing with events involving variability, it is impossible to reduce both errors to zero simultaneously.

Statistics provides a formulation for how to make realistically correct judgments between these 'alarmist' and 'blockhead' errors.

In statistics, one first decides the permissible limit for the α error. This permissible limit is called the '**significance level**'.

Within that significance level, one seeks to minimize the β error, or conversely, maximize the statistical power.

Statistical Testing

Simply increasing the data used for analysis increases statistical power. However, to avoid missing the truth (like a 'blockhead') even with limited data, different methods are used depending on the hypothesis.

In statistics, the methodology used to judge whether a hypothesis can be considered correct is generally called **statistical testing**.

At a given significance level, the testing method with the highest statistical power is called the **most powerful test**.

Example of Testing in Marketing

Suppose that as a result of an A/B test, a new design increased the conversion rate from 0.10% to 0.11%.

> This is a mere 0.1% difference, but if it is a truly meaningful difference, it might increase the service's sales 1.1-fold.
>
> Conversely, if it is 'just a coincidence', one might fall into a vicious cycle of continuing useless design changes.

A statistical test is what is used to determine whether this slight 0.1% difference should be considered a 'significant difference' or just a statistically meaningless difference due to chance.

References I

Statistics is the most powerful discipline (Practical Edition) - Kei Nishiuci