Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

5.2 Type I and Type II errors

In any hypothesis test there are two types of **inferential decision error** which could be committed.

Clearly, we would like to reduce the probabilities of these errors as much as possible. These two types of error are called a Type I error and a Type II error.

- **Type I error**: rejecting H_0 when it is true. This can be thought of as a 'false positive'. Denote the probability of this type of error by α.
- **Type II error**: failing to reject H_0 when it is false. This can be thought of as a 'false negative'. Denote the probability of this type of error by β.

Both errors are undesirable and, depending on the context of the hypothesis test, it could be argued that either one is worse than the other.

However, on balance, a Type I error is usually considered to be more problematic.[1]

The possible **decision space** can be presented as:

		Decision made	
		H_0 not rejected	H_0 rejected
True state of nature	H_0 true	Correct decision	Type I error
	H_1 true	Type II error	Correct decision

For example, if H_0 was being 'not guilty' and H_1 was being 'guilty', a Type I error would be finding an innocent person guilty (bad for him/her), while a Type II error would be finding a guilty person innocent (bad for the victim/society, but admittedly good for him/her!).

[1] Thinking back to trials by jury, conventional wisdom is that it is better to let 100 guilty people walk free than to convict a single innocent person. While you are welcome to disagree, this view is consistent with Type I errors being more problematic.

The complement of a Type II error, that is $1 - \beta$, is called the **power** of the test – the probability that the test will reject a false null hypothesis. Hence power measures the ability of the test to reject a false H_0, and so we seek the most powerful test for any testing situation.

Unlike α, we do not control test power.However, we can increase it by increasing the sample size, n (a larger sample size will inevitably improve the accuracy of our statistical inference).

These concepts can be summarised as **conditional probabilities**.

	Decision	
	H_0 not rejected	H_0 rejected
H_0 true	$1 - \alpha$	$P(\text{Type I error}) = \alpha$
H_1 true	$P(\text{Type II error}) = \beta$	Power $= 1 - \beta$

We have:

$$P(\mathrm{H}_0 \text{ not rejected} \,|\, \mathrm{H}_0 \text{ is true}) = 1 - \alpha$$

$$P(\mathrm{H}_0 \text{ rejected} \,|\, \mathrm{H}_0 \text{ is true}) = \alpha$$

$$P(\mathrm{H}_0 \text{ not rejected} \,|\, \mathrm{H}_1 \text{ is true}) = \beta$$

$$P(\mathrm{H}_0 \text{ rejected} \,|\, \mathrm{H}_1 \text{ is true}) = 1 - \beta.$$

Other things equal, if you decrease α you increase β and vice-versa. Hence there is a trade-off.

Significance level

Since we control for the probability of a Type I error, α, what value should this be?

Well, in general we test at the $100\alpha\%$ significance level, for $\alpha \in [0, 1]$. The default choice is $\alpha = 0.05$, i.e. we test at the 5% significance level. Of course, this value of α is subjective, and a different significance level may be chosen. The severity of a Type I error in the context of a specific hypothesis test might for example justify a more conservative or liberal choice for α.

In fact, noting our look at confidence intervals in Section 4.6, we could view the **significance level as the complement of the confidence level**.[2] For example:

- a 90% confidence level equates to a 10% significance level
- a 95% confidence level equates to a 5% significance level
- a 99% confidence level equates to a 1% significance level.

[2]Strictly speaking, this would apply to so-called 'two-tailed' hypothesis tests.