Open topic with navigation

P Values

The P value, or calculated probability, is the probability of finding the observed, or more extreme, results when the **null hypothesis (H_0)** of a study question is true – the definition of 'extreme' depends on how the hypothesis is being tested. P is also described in terms of rejecting **H_0** when it is actually true, however, it is not a direct probability of this state.

The null hypothesis is usually an hypothesis of "no difference" e.g. no difference between blood pressures in group A and group B. Define a null hypothesis for each study question clearly before the start of your study.

The only situation in which you should use a **one sided** P value is when a large change in an unexpected direction would have absolutely no relevance to your study. This situation is unusual; if you are in any doubt then use a **two sided** P value.

The term **significance level (alpha)** is used to refer to a pre-chosen probability and the term "P value" is used to indicate a probability that you calculate after a given study.

The **alternative hypothesis (H_1)** is the opposite of the null hypothesis; in plain language terms this is usually the hypothesis you set out to investigate. For example, question is "is there a significant (not due to chance) difference in blood pressures between groups A and B if we give group A the test drug and group B a sugar pill?" and alternative hypothesis is " there is a difference in blood pressures between groups A and B if we give group A the test drug and group B a sugar pill".

If your P value is less than the chosen significance level then you reject the null hypothesis i.e. accept that your sample gives reasonable evidence to support the alternative hypothesis. It does NOT imply a "meaningful" or "important" difference; that is for you to decide when considering the real-world relevance of your result.

The choice of significance level at which you reject H_0 is arbitrary. Conventionally the 5% (less than 1 in 20 chance of being wrong), 1% and 0.1% ($P < 0.05$, 0.01 and 0.001) levels have been used. These numbers can give a false sense of security.

In the ideal world, we would be able to define a "perfectly" random sample, the most appropriate test and one definitive conclusion. We simply cannot. What we can do is try to optimise all stages of our research to minimise sources of uncertainty. When presenting P values some groups find it helpful to use the asterisk rating system as well as quoting the P value:

P < 0.05 *

P < 0.01 **

P < 0.001

Most authors refer to **statistically significant** as $P < 0.05$ and **statistically highly significant** as $P < 0.001$ (less than one in a thousand chance of being wrong).

The asterisk system avoids the woolly term "significant". Please note, however, that many statisticians do not like the asterisk rating system when it is used without showing P values. As a rule of thumb, if you can quote an exact P value then do. You might also want to refer to a quoted exact P value as an asterisk in text narrative or tables of contrasts elsewhere in a report.

At this point, a word about error. **Type I error** is the false rejection of the null hypothesis and **type II error** is the false acceptance of the null hypothesis. As an aid memoir: think that our cynical society rejects before it accepts.

The significance level (alpha) is the probability of type I error. The power of a test is one minus the probability of type II error (beta). Power should be maximised when selecting statistical methods. If you want to estimate sample sizes then you must understand all of the terms mentioned here.

The following table shows the relationship between power and error in hypothesis testing:

		DECISION
TRUTH	Accept H_0:	Reject H_0:
H_0 is true:	**correct decision P** *1-alpha*	**type I error P** *alpha (significance)*

H_0 is false:	**type II error P** *beta*	**correct decision P** *1-beta (power)*

H_0 = null hypothesis
P = probability

If you are interested in further details of probability and sampling theory at this point then please refer to one of the general texts listed in the reference section.

You must understand **confidence intervals** if you intend to quote P values in reports and papers. Statistical referees of scientific journals expect authors to quote confidence intervals with greater prominence than P values.

Notes about **Type I error**:

- is the incorrect rejection of the null hypothesis
- maximum probability is set in advance as alpha
- is not affected by sample size as it is set in advance
- increases with the number of tests or end points (i.e. do 20 rejections of H_0 and 1 is likely to be wrongly significant for alpha = 0.05)

Notes about **Type II error**:

- is the incorrect acceptance of the null hypothesis
- probability is beta
- beta depends upon sample size and alpha
- can't be estimated except as a function of the true population effect
- beta gets smaller as the sample size gets larger
- beta gets smaller as the number of tests or end points increases