What is the P-value?

Bengt Ringnér

May 23, 2007

When testing some hypothesis, for instance that the sex of a newly-born baby is a girl or a boy has equal probabilities, the result is often reported as a P-value, which is the probability of getting a result as least extreme as the one observed. For instance, among 1024 babies one would expect 512 boys with a standard deviation of $\sqrt{1024 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 16$. According to the normal approximation, there is a probability of 5% of obtaining a number $512 - 2 \cdot 16 = 480$ or lower, or a number $512 + 2 \cdot 16 = 544$ or higher. If this happens, one rejects the hypothesis with error probability 5%. Similarly, the probability of getting an observation outside three times the standard deviation is 0.2%.

- Does the above calculation give the probability that the hypothesis is true?

Let us look at an example: Suppose that one has made 25 independent repetitions of an experiment which can either succeed or not, and that one has observed 18 successes. In other words, 18 is an observation of a random variable X, which has a binomial distribution $\text{Bin}(25, p)$ where p is the probability of success. The P-value when testing $p = .5$ against $p > .5$ is

$$P = \mathbb{P}(X \geq 18) = 0.0216$$

from a table over the binomial distribution. Now consider some situations:

1. A person claims that he can feel the difference between tea made from bags and tea made in a pot. A statistician who doubts this makes an experiment in which the person at 25 occasions gets one cup of each without knowing which one is which.

2. Another person claims that he has telepatic talents and can tell whether a coin tossed in another room falls heads or tails.

3. A third situation involves the question whether sunspots can affect activity on Earth.

In all three cases the P-values are the same, but their interpretation is, loosely speaking, the probability that data is in error. This probability has to be combined with information from other sources as to how probable H_0 is. In some cases this is even less probable than data being in error. Therefore one proceeds in steps:

1. First one makes objective calculations and reports them.

2. Then one makes more or less subjective judgments.

An alternative approach is the Bayesian one.

- What is the probability of the hypothesis being tested?

This question can only be answered if one, prior to making observations, can quantify the plausibility of the hypothesis in terms of probability. Let us start with a simplified situation: We want to test

$$H_0 : p = 0.5$$

against

$$H_1 : p = 0.7.$$

Suppose that these hypotheses have prior probabilities π and $1 - \pi$, respectively. Then Bayes' theorem gives

$$\mathbb{P}(H_0|X = 18) = \frac{\mathbb{P}(H_0)\mathbb{P}(X = 18|H_0)}{\mathbb{P}(X = 18)} = \frac{\pi\mathbb{P}(X = 18|H_0)}{\pi\mathbb{P}(X = 18|H_0) + (1 - \pi)\mathbb{P}(X = 18|H_1)}.$$

Later in the course we shall study more realistic situations, namely

$$H_0 : p \leq 0.5$$

against

$$H_1 : p > 0.5,$$

in which the P-value is approximately the probability of H_0, and

$$H_0 : p = 0.5$$

against

$$H_1 : p \neq 0.5,$$

where things are quite different.