

Lsn36

Clark

Admin

Assume we are testing $H_0 : \theta = \theta_0$ vs $H_a : \theta < \theta_0$ and our test statistic is $W = T(X)$ some function of our data. Under H_0 we typically choose our test statistic because we know the distribution of it, generically, $W \sim F(.)$. After we collect our data we calculate the realization of our test statistic, say w_0 . Note that W is a random variable, for example we might use:

as our test statistic. We then conduct the experiment and perhaps we observe:

Formally the *P-Value* is:

Note the similarities with α . The biggest difference is when we were discussing Type-1 errors earlier we either fixed α and found our rejection region, or fixed our rejection region and found α . In either case, **our conversations occurred prior to data being collected**. When we compute a P-value we are conducting calculations after our experiment has occurred.

Oftentimes the interpretation that of a P-value being the probability under a specified statistical model that a statistical summary of the data (e.g. the sample mean difference between two compared groups) would be equal to or more extreme than its observed value

Let's assume we wanted to determine if more than 50 percent of the corps thought that our SHARP training last Friday was effective. So 10 Cadets were asked whether the SHARP training we conducted last Friday was effective and 7 said that it was. What is the P-value associated with this test? What does it mean?

For large samples we can again use the Z statistic.

Two sets of elementary schoolchildren were taught to read by using different methods, 50 by each method. At the conclusion of the instructional period, a reading test yielded the results $\bar{y}_1 = 74$, $\bar{y}_2 = 71$, $s_1 = 9$, and $s_2 = 10$. What is the P-value associated with this test?

Most statisticians don't like P-values. They get used and abused. In fact the *American Statistical Association*,

in 2016, came out with six principles about P-values.

1. P-values can indicate how incompatible the data are with a specified statistical model
2. P-values do NOT measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone
3. Scientific conclusions and business or policy decisions should NOT be based only on whether a p-value passes a specific threshold
4. Proper inference requires full reporting and transparency
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

More recently, leaders within Academia and the ASA published further guidance on how we move to a world beyond $p < .05$. In sum, they state, don't say statistically significant. Don't say it, and don't use it.

But rather, **accept uncertainty** and use statistical principles to quantify the uncertainty; **be thoughtful**, clearly state your research objectives, recognize and be truthful about doing exploratory studies vs pre-planned rigorous studies; **be open** and transparent in how you did your analysis and how your data was collected, share your data, share your code, get over your ego, report P-values as continuous descriptive statistics as you would \bar{X} ; **be modest** a statistical model is an abstraction of reality, it is not correct and, of necessity, they are extremely simplified relative to the complexity of actual study conduct, remember that statistical tools have limitations.

$P < 0.05$ as a measure for a good study vs a poor study is slowly going away but it will take awhile. A p-value, much like anything we study in this course, has its uses but is just one of hundreds of concepts in statistical inference. It is not, nor should it be, the desired end state. It is ONLY a tool that tells us how incompatible our data are with the null model under consideration.