

Technical Support Document

Taking the Perplexity out of P-Values

In Minitab, the conclusion for many statistical tests is reported in terms of the p-value, rather than in terms of rejecting or failing to reject the null hypothesis. Minitab technical support frequently receives calls from users who aren't aware that they can use the p-value to decide whether or not to reject the null hypothesis.

In a basic t-test, for example, the null hypothesis, H_0 , and the alternative hypotheses, H_A , are usually $H_0: \mu = \mu_0$ and $H_A: \mu \neq \mu_0$, where μ (mu) is the population mean for the group under study and where μ_0 is the hypothesized population mean.

Let's say you want to determine if a new gasoline additive has an effect on gas mileage. If the known gas mileage for this particular class of cars is 25 miles per gallon (m.p.g.), the hypotheses for this study are $H_0: \mu = 25$ and $H_A: \mu \neq 25$.

You test 35 cars and find that the m.p.g.'s range from 14.4 to 28.8. After putting the data in the column MPG, you run Minitab's t-test (the menu command Stat > Basic Statistics > 1-Sample t, or the session command TTEST) and get these results:

Output from T-Test

```
Test of mu = 25.000 vs mu not = 25.000
Variable      N      Mean    StDev   SE Mean      T      P-Value
MPG           35     23.754   3.156   0.533     -2.34    0.026
```

Now What?

Interpreting the P-Value

The results show that the mean of the 35-car sample is 23.754. But the mean m.p.g. of all cars of this type (μ) may still be 25—you need to know if there is enough sample evidence to reject H_0 . There are two usual ways to determine that, and both involve the significance level, α (alpha). α is the probability of rejecting H_0 when H_0 is true. In this case, that's the probability of concluding that the population mean is NOT 25 m.p.g., when in fact it is.

The first way to determine if you have enough sample evidence to reject H_0 is to compute the value of the test statistic and compare it to a value from the appropriate statistical table at a pre-specified value of α .

The second method is to compute the p-value, often called the attained significance level of the test, and compare it to α . The p-value is an indication of how well the sample evidence supports the rejection of H_0 . In general, the smaller the p-value, the heavier the weight of the sample evidence for rejecting H_0 . More specifically, the p-value is the smallest value of α that results in the rejection of H_0 . For any value of $\alpha < p$ -value, fail to reject H_0 , and for any value of $\alpha \geq p$ -value, reject H_0 .

Technical Support Document

Taking the Perplexity out of P-Values

The p-value is also the proportion of the test statistics from samples of the same size (taken from the same population) that would yield a more extreme value of the test statistic under the assumption that H_0 is true. In our t-test example, the test statistic is a function of the mean, and the p-value is .026. This means that 2.6% of the means from samples of size 35, drawn from a population where $\mu = 25$, will produce a mean which indicates that μ is not equal to 25. Ask yourself which is most likely: that you were lucky enough to draw one of the few samples that indicate that μ is not equal to 25; or that μ really is not equal to 25?

The p-value is traditionally compared to α values of less than .05 or .01, depending on the field of study (your mileage may vary!). Check journal articles in your field for acceptable values.

In our example, let's assume a .05 value of α . The p-value of .026 indicates that the mean m.p.g. of all cars of this type (and not just the mean of the 35 cars in the study) is probably not equal to 25. A more statistically correct way to state this is "at a significance level of .05, the mean m.p.g. appears to be significantly different from 25."

Using P-Values in Other Tests

The null hypothesis varies from test to test. Interpreting the p-value depends on knowing what the null and alternative hypotheses are. The Minitab output for regression, for example, includes a hypothesis test for the estimates of β_0, β_1, \dots . In the example below, the variable DIAMETER is very statistically significant, because the p-value is essentially 0.

Output from Regression

The regression equation is
 VOLUME = - 36.9 + 5.07 DIAMETER

Predictor	Coef	Stdev	t-ratio	p
Constant	-36.943	3.365	-10.98	0.000
DIAMETER	5.0659	0.2474	20.48	0.000

s = 4.252 R-sq = 93.5% R-sq(adj) = 93.3%

In a test for normality, such as the Anderson-Darling test or the Kolmogorov-Smirnov test, the null hypothesis is that the data have a normal distribution. In a Chi-squared test of independence, the null hypothesis is that the two factors in the study—for example, age and voting preference—are independent of one another.

If the test results in a small enough p-value, then there is enough sample evidence to reject the null hypothesis in favor of the alternative hypothesis ("the data have a non-normal distribution" or "the factors are not independent").

Using p-values, then, is simple if you know two key facts: the values of α that are acceptable in your field, and the null and alternative hypotheses for the tests you are using.