

11.4: One- and Two-Tailed Tests

Learning Objectives

- Define **Type I** and **Type II** errors
- Interpret significant and non-significant differences
- Explain why the null hypothesis should not be accepted when the effect is not significant

In the James Bond case study, Mr. Bond was given 16 trials on which he judged whether a martini had been shaken or stirred. He was correct on 13 of the trials. From the binomial distribution, we know that the probability of being correct 13 or more times out of 16 if one is only guessing is 0.0106. Figure 11.4.1 shows a graph of the binomial distribution. The red bars show the values greater than or equal to 13. As you can see in the figure, the probabilities are calculated for the upper tail of the distribution. A probability calculated in only one tail of the distribution is called a "one-tailed probability."

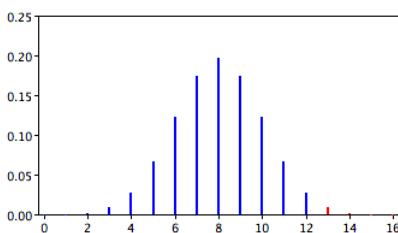


Figure 11.4.1: The binomial distribution. The upper (right-hand) tail is red

Binomial Calculator

A slightly different question can be asked of the data: "What is the probability of getting a result as extreme or more extreme than the one observed?" Since the chance expectation is 8/16, a result of 3/16 is equally as extreme as 13/16. Thus, to calculate this probability, we would consider both tails of the distribution. Since the binomial distribution is symmetric when $\pi = 0.5$, this probability is exactly double the probability of 0.0106 computed previously. Therefore, $p = 0.0212$. A probability calculated in both tails of a distribution is called a "two-tailed probability" (see Figure 11.4.2).

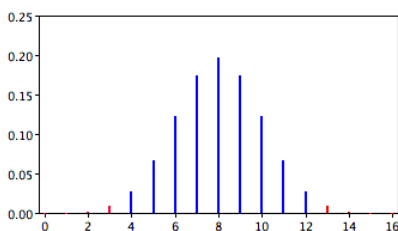


Figure 11.4.2: The binomial distribution. Both tails are red

Should the one-tailed or the two-tailed probability be used to assess Mr. Bond's performance? That depends on the way the question is posed. If we are asking whether Mr. Bond can tell the difference between shaken or stirred martinis, then we would conclude he could if he performed either much better than chance or much worse than chance. If he performed much worse than chance, we would conclude that he can tell the difference, but he does not know which is which. Therefore, since we are going to reject the null hypothesis if Mr. Bond does either very well or very poorly, we will use a two-tailed probability.

On the other hand, if our question is whether Mr. Bond is better than chance at determining whether a martini is shaken or stirred, we would use a one-tailed probability. What would the one-tailed probability be if Mr. Bond were correct on only 3 of the 16 trials? Since the one-tailed probability is the probability of the right-hand tail, it would be the probability of getting 3 or more correct out of 16. This is a very high probability and the null hypothesis would not be rejected.

The null hypothesis for the two-tailed test is $\pi = 0.5$. By contrast, the null hypothesis for the one-tailed test is $\pi \leq 0.5$. Accordingly, we reject the two-tailed hypothesis if the sample proportion deviates greatly from 0.5 in either direction. The one-tailed hypothesis is rejected only if the sample proportion is much greater than 0.5. The alternative hypothesis in the two-tailed test is $\pi \neq 0.5$. In the one-tailed test it is $\pi > 0.5$.

You should always decide whether you are going to use a one-tailed or a two-tailed probability before looking at the data. Statistical tests that compute one-tailed probabilities are called one-tailed tests; those that compute two-tailed probabilities are

called two-tailed tests. Two-tailed tests are much more common than one-tailed tests in scientific research because an outcome signifying that something other than chance is operating is usually worth noting. One-tailed tests are appropriate when it is not important to distinguish between no effect and an effect in the unexpected direction. For example, consider an experiment designed to test the efficacy of a treatment for the common cold. The researcher would only be interested in whether the treatment was better than a placebo control. It would not be worth distinguishing between the case in which the treatment was worse than a placebo and the case in which it was the same because in both cases the drug would be worthless.

Some have argued that a one-tailed test is justified whenever the researcher predicts the direction of an effect. The problem with this argument is that if the effect comes out strongly in the non-predicted direction, the researcher is not justified in concluding that the effect is not zero. Since this is unrealistic, one-tailed tests are usually viewed skeptically if justified on this basis alone.

This page titled [11.4: One- and Two-Tailed Tests](#) is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by [David Lane](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.