**Probability** is the measure of the likeliness that an event will occur.Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).The higher the probability of an event, the more certain we are that the event will occur. A simple example is the toss of a fair coin. Since the two outcomes are equally probable, the probability of "heads" equals the probability of "tails", so the probability is 1/2 (or 50%) chance of either "heads" or "tails"

**Frequency Distributions**

One important set of statistical tests allows us to test for deviations of observed frequencies from expected frequencies. To introduce these tests, we will start with a simple, non-biological example. We want to determine if a coin is fair. In other words, are the odds of flipping the coin heads-up the same as tails-up. We collect data by flipping the coin 200 times. The coin landed heads-up 108 times and tails-up 92 times. At first glance, we might suspect that the coin is biased because heads resulted more often than than tails. However, we have a more quantitative way to analyze our results, a chi-squared test.

To perform a chi-square test (or any other statistical test), we first must establish our null hypothesis. In this example, our null hypothesis is that the coin should be equally likely to land head-up or tails-up every time. The null hypothesis allows us to state expected frequencies. For 200 tosses, we would expect 100 heads and 100 tails.

The next step is to prepare a table as follows.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Heads | Tails | Total |
| Observed | 108 | 92 | 200 |
| Expected | 100 | 100 | 200 |
| Total | 208 | 192 | 400 |

The Observed values are those we gather ourselves. The expected values are the frequencies expected, based on our null hypothesis. We total the rows and columns as indicated. It's a good idea to make sure that the row totals equal the column totals (both total to 400 in this example).

Using probability theory, statisticians have devised a way to determine **if a frequency distribution differs from the expected distribution**. To use this chi-square test, we first have to calculate chi-squared.

Chi-squared = … (observed-expected)2/(expected)

We have two classes to consider in this example, heads and tails.

Chi-squared = (100-108)2/100 + (100-92)2/100 = (-8)2/100 + (8)2/100 = 0.64 + 0.64 = **1.28**

Now we have to consult a table of critical values of the chi-squared distribution. Here is a portion of such a table.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **df/prob.** | **0.99** | **0.95** | **0.90** | **0.80** | **0.70** | **0.50** | **0.30** | **0.20** | **0.10** | **0.05** |
| **1** | 0.00013 | 0.0039 | 0.016 | 0.64 | 0.15 | 0.46 | 1.07 | 1.64 | 2.71 | 3.84 |
| **2** | 0.02 | 0.10 | 0.21 | 0.45 | 0.71 | 1.39 | 2.41 | 3.22 | 4.60 | 5.99 |
| **3** | 0.12 | 0.35 | 0.58 | 1.00 | 1.42 | 2.37 | 3.66 | 4.64 | 6.25 | 7.82 |
| **4** | 0.3 | 0.71 | 1.06 | 1.65 | 2.20 | 3.36 | 4.88 | 5.99 | 7.78 | 9.49 |
| **5** | 0.55 | 1.14 | 1.61 | 2.34 | 3.00 | 4.35 | 6.06 | 7.29 | 9.24 | 11.07 |

The left-most column list the degrees of freedom (df). We determine the degrees of freedom by subtracting one from the number of classes (df = n - 1). In this example, we have two classes (heads and tails), so our degrees of freedom is 1.

Our chi-squared value is 1.28. Move across the row for 1 df until we find critical numbers that bound our value. In this case, 1.07 (corresponding to a probability of 0.30) and 1.64 (corresponding to a probability of 0.20). We can interpolate our value of 1.24 to estimate a probability of 0.27. This value means that there is a 73% chance that our coin is biased. In other words, the probability of getting 108 heads out of 200 coin tosses with a fair coin is 27%. In biological applications, a probability ¾ 5% is usually adopted as the standard. Because the chi-squared value we obtained in the coin example is greater than 0.05 (0.27 to be precise), we accept the null hypothesis as true and conclude that our coin is fair.