The choice of statistical test is often a matter of evaluating the relative suitability of a number of possible options. We wish to explain why we use the binomial distribution to evaluate the null hypothesis that male and female effigies are equally likely to have cheek decorations, instead of alternatives such as the Chi-square Test and Fisher’s Exact Probability. First, the binomial distribution is a particularly elegant method for evaluating the hypothesis (i.e., it is appropriate). Second, the Chi-square Test and the Fisher Exact Probability are less sensitive in this case. Consider the 2 x 2 matrix corresponding to our binomial analysis:

|  |  |  |
| --- | --- | --- |
|  | Sex | |
| Cheek Decoration | Female | Male |
| Present | 42 | 43 |
| Absent | 3 | 10 |

The probabilities of a Chi-square Test and Fisher’s Exact Probability of this matrix are .07 and .05, which are not significant given α = .05. However, neither test is actually comparing the probability of association as specified in our hypothesis. The abundances being compared correlate with the probability, but these tests are less sensitive than the binomial distribution for such associations. To illustrate this, consider flipping an unbiased coin and a coin biased towards heads at a probability of .7, 20 times each. The expected 2x2 matrix is:

|  |  |  |
| --- | --- | --- |
|  | Coin | |
| Result | Unbiased | Biased |
| Heads | 10 | 14 |
| Tails | 10 | 6 |

A Chi-square Test and Fisher’s Exact Probability produce probabilities of .20 and .11, both of which are greater than α, thereby producing Type II statistical errors in which meaningful differences are not detected. However, using the binomial distribution to evaluate whether 14 heads/6 tails is expected given p(head) = .5 produces a probability of .02, and a correct decision to reject the null hypothesis. A larger sample size would eventually allow both the Chi-square Test and the Fisher’s Exact Probability to correctly identify the different probabilities of a head, but the Binomial Distribution is a more sensitive method of doing so.

The Chi-square Goodness of Fit is also less useful than the binomial distribution, because it requires expected values of 5 or greater. This assumption is rarely met in our data. For example, alternating zigzag lines on the cheek occur with 8 males but no females (Table 1). The Chi-Square Goodness of Fit test produces expected values of 4, which violates the test’s assumptions:

|  |  |  |
| --- | --- | --- |
| Effigy Sex | Observed | Expected |
| Female | 0 | 4 |
| Male | 8 | 4 |

However, the Binomial Distribution has no such assumptions, causing it to be a more robust statistical approach.