Cornell Statistical Consulting Unit

The Hardy-Weinberg Principle in Population Genetics

Jing Yang

1 Introduction

The Hardy-Weinberg principle is an important concept in population genetics. It states that allele and genotype frequencies in an ideal population will remain constant from one generation to the next without any evolutionary factors such as non-random mating, natural selection, mutations, gene flow, etc. An allele is one of two or more possible forms of a gene at the same place on a chromosome; a genotype is the genetic makeup of an individual and can refer to an organism's entire genetic makeup or the alleles at a particular locus.

This state of equilibrium is also called Hardy-Weinberg Equilibrium (HWE). If genotype frequencies differ from what we would expect under HWE, we assume that one or more of these evolutionary factors might occur, which can be interesting and useful in many different areas of research.

2 Allele and genotype frequences under HWE

There are two important equations which describe the HWE: one that relates to allele frequencies and the other that relates to genotype frequencies. Consider the simplest case, a population of diploid individuals, where each individual consists of two alleles at each gene locus, which has two possible types denoted A and B. Since there are only two possible alleles, the frequency of one plus the frequency of the other must equal 1. Therefore, we have

$$\text{freq}(A) + \text{freq}(B) = p + q = 1,$$

where p and q denote the frequencies of alleles A and B, respectively.

Ideally, the expected genotype frequencies in the next generation are $\text{freq}(AA) = p^2$, $\text{freq}(BB) = q^2$ and $\text{freq}(AB) = 2pq$. These genotype frequencies are also called the Hardy-Weinberg proportions, and these also should sum to 1

$$p^2 + 2pq + q^2 = 1.$$

2.1 Testing for HWE

A statistical test for HWE compares a population's observed genetic structure with the genetic structure we would expect if the population were in HWE. For instance, tests of HWE have been widely used for detecting possible genotyping errors in the quality control process, detecting potential genetic variants under natural selection, and helping to identify candidate genes associated with Mendelian and complex human diseases.

2.1.1 Example

This example illustrates how to determine whether a population is in HWE.

Table 2.1 contains counts of the number of individuals with each of three genotypes.

Table 2.1: Example genotype frequencies

Genotype	Count (n)
AA	725
AB	123
BB	7
Total	855

First, we calculate the frequencies of the two alleles A and B:

$$\text{freq}(A) = p = \frac{2n_{AA} + n_{AB}}{2n_{\text{total}}} = \frac{2 \times 725 + 123}{2 \times 855} \approx 0.92,$$

$$\text{freq}(B) = q = \frac{n_{AB} + 2n_{BB}}{2n_{\text{total}}} = \frac{123 + 2 \times 7}{2 \times 855} \approx 0.08$$

Then the expected genotype frequencies under HWE are

$$\text{freq}(AA) = p^2 = 0.92^2,$$

$$\text{freq}(AB) = 2pq = 2 \times 0.92 \times 0.08,$$

and

$$\text{freq}(BB) = q^2 = 0.08^2.$$

For a total of 855 individuals in the sample, we expect that approximately 724 individuals ($p^2 \times n_{\text{total}} = 0.92^2 \times 855 = 723.67$) have the AA genotype, 126 ($2pq \times n_{\text{total}} = 2 \times 0.92 \times 0.08 \times 855 = 125.86$) individuals have the AB genotype, and 5 individuals ($q^2 \times n_{\text{total}} = 0.08^2 \times 855 = 5.47$) have the BB genotype.

A HWE test is typically conducted to check how these expected numbers compare to the observed numbers in Table 2.1. Generally, we use a Chi-square test to compare them for a relatively large sample size. The Chi-square test statistic can be calculated as

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{(725 - 723.67)^2}{723.67} + \frac{(123 - 125.86)^2}{125.86} + \frac{(7 - 5.47)^2}{5.47} \approx 0.49$$

where O_i denotes the observed counts and E_i denotes the expected counts. Comparing the test statistic of 0.49 to a Chi-square distribution with one degree of freedom, the P-value is about 0.104 and we conclude that the expected and observed counts are not significantly different from one another for this sample. In other words, we cannot reject the null hypothesis that the allele and genotype frequencies at this given locus are in HWE. If the sample size is not large enough, exact tests for HW proportions are available (Engels 2009).

For a more general case of n distinct alleles in m-ploid individuals, the expected genotype frequencies under HWE can be similarly derived from individual terms in a multinomial

expansion of the expression $(p_1 + p_2 + \cdots + p_n)^m$ and a HWE test can still be applied accordingly.

Reference: Genetics (2009) 183: 1431–1441, Exact Tests for Hardy–Weinberg Proportions. W.R. Engels.

Created October 2013. Last updated April 2022.