StatNews #87

The Hardy-Weinberg Principle in Population Genetics

October 2013

The Hardy-Weinberg principle is an important concept in population genetics. It states that, allele[1] and genotype[2] 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. 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.

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 to 100%, or 1. Therefore, we have

$$freq(A) + freq(B) = p + q = 1 \qquad (1)$$

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

Ideally, the expected possible genotype frequencies in the next generation are freq(AA)= p^2, freq(BB)= q^2, and freq(AB)= $2pq$. These genotype frequencies are also called the Hardy-Weinberg proportions, and we have

$$p^2 + 2pq + q^2 = (p + q)^2 = 1 \qquad (2)$$

---

[1] Allele: an allele is one of two or more forms of a gene or a genetic locus.

[2] Genotype: the genetic makeup of an individual. Genotype can refer to an organism's entire genetic makeup or the alleles at a particular locus.

A HWE test enables us to compare a population's observed genetic structure with the genetic structure that 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, for testing for potential genetic variants under natural selection, and for helping to identify candidate genes associated with Mendelian and complex human diseases.

We illustrate how to determine whether or not a population is in HWE using the example below. The table contains counts of the number of individuals with each of three genotypes.

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

First, we calculate the frequencies of the two alleles A and B by simply counting,

$$freq(A) = p = \frac{n_{AA}*2 + n_{AB}*1 + n_{BB}*0}{n_{total}*2} = \frac{(725*2 + 123*1 + 7*0)}{855*2} \approx 0.92,$$

$$freq(B) = q = \frac{n_{AA}*0 + n_{AB}*1 + n_{BB}*2}{n_{total}*2} = \frac{(725*0 + 123*1 + 7*2)}{855*2} \approx 0.08,$$

Then the expected genotype frequencies under HWE (Equation 2) are

$$freq(AA) = p^2 = (0.92)^2,$$

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

$$freq(BB) = q^2 = (0.08)^2.$$

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

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

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

where O_i denotes the observed counts and E_i denotes the expected counts. The p-value is around 0.104 for a test statistic of 0.49 according a Chi-square distribution with one degree freedom, so we conclude that the observed and expected 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 equation $(p_1 + p_2 + \cdots + p_n)^m$ and a HWE test can still be applied accordingly.

As always if you would like assistance with this topic or any other statistical consulting question, feel free to contact statistical consultants at CSCU.
Author: Jing Yang

(This newsletter was distributed by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at cscu@cornell.edu. Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at http://www.cscu.cornell.edu).

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