Advance Algorithm in Bioinformatics – CSC 8540 – Fall 2022

September 6 2022

Pardis Sadatian Moghaddam

Panther ID: 002722641

HW2:

(a) p_A=0.7, p_a =0.3, p_AA = 0.6, p_aa=0.01, p_Aa=0.39
Find probability that the population of size 100 is in HWE.

p_A=0.7

p_a=0.3

p_AA=0.6

p_aa=0.01

p_Aa=0.39

Hardy-Weinberg Equation:

p^2 + 2*p*q + q^2 = 1

p + q = 1

p_AA = 0.6 => numbers of AA genotype is 60

p_Aa = 0.39 => numbers of Aa genotype is 39

p_aa = 0.01 => numbers of aa genotype is 1

Numbers of AA genotype = 60 => total of 120 in population.

Numbers of Aa genotype = 39

Numbers of A alleles = 120 + 39 = 159

Numbers of aa genotype =1 => total of 2 in population.

Numbers of Aa genotype = 39

Numbers of a alleles = 2 + 39 = 41

The total numbers = 120 + 2 + 39*2 = 200

We have **dominant allele** and **recessive allele**

dominant allele = p_A = 159/200 = 0.795 ≈ 0.7

recessive allele = p_a = 41/200 = 0.205 ≈ 0.3

Quick Check => p+q = 1 => 0.795 + 0.205 = 1

p = P_A = 0.795 => p_AA = 0.632025

p = p_A ≈ 0.7 => p_AA = 0.49

q = p_a = 0.205 => p_aa = 0.042025

q = p_a ≈ 0.3 => p_aa = 0.09

2*p*q = P_Aa = 2*0.632025 * 0.042025 = 0.05312170125

P_aA = 0.42

At first we consider it p_AA = 0.6, p_aa = 0.01 and p_Aa = 0.39. The genotype should be p_AA = 0.49, p_aa = 0.9 and p_Aa = 0.49 if the population size 100 is In Hardy-Weinberg.

The X test (HW Chi Squared)

Population
P^2*Total = 0.49*100=49
2*p*q*Total = 0.42*100= 42

q^2*Total = 0.09*100=9

X^2 = ∑(observed - expected)^2 – (expected)

X^2 = (60-49)^2/49 + (1-9)^2/9 + (39-42)^2/42 = 2.469 + 7.111 + 0.214 = **9.794**

Degree of freedom = n-1 = 2-1 = 1

X^2 = 9.794

Pr (x^2> 9.794) = 0.001751 which is less than 0.05 so it is not accepted.

b) Given genotype population: AA=100, Aa=50, aa=50. Based on allele frequencies and assuming HWE, What is the expected genotype frequencies?

Perform X^2 test – do you accept hypothesis that the population is in HWE?

Numbers of AA genotype = 199 => total of 200 in population.

Numbers of Aa genotype = 50

Numbers of A alleles = 200 + 50 = 250

Numbers of aa genotype =50 => total of 100 in population.

Numbers of Aa genotype = 50

Numbers of a alleles = 2 + 39 = 150

The total numbers = 200 + 100 + 50*2 = 400

We have **dominant allele** and **recessive allele**

dominant allele = p_A = 250/400 = 0.625

recessive allele = p_a = 150/400 = 0.375

The genotype frequencies are:

AA **(0.625*0.625 = 0.390625)**	Aa **(0.625*0.375 = 0.234375)**	aa **(0.375*0.375 = 0.140625)**

The genotype should be p_AA = 0.390625, p_aa = 0.140625 and p_Aa = 0.234375 In Hardy-Weinberg.

The X test (HW Chi Squared)

P = 0.625

q = 0.375

Population

P^2*Total = 0.390625*200=78.125
2*p*q*Total = 2*0.625*0.375*200= 93.75
q^2*Total = 0.140625*200=28.125

X^2 = ∑(observed - expected)^2 – (expected)

X^2 = (100-78.125)^2/78.125 + (50-93.75)^2/93.75 + (50-28.125)^2/28.125 = 6.125 + 20.416 + 17.013 = **43.554**

Degree of freedom = n-1 = 2-1 = 1

X^2 = 43.554

Pr (x^2> 43.554) = 0.00001 which is less than 0.05 so it is not accepted.