

Week 4 Homework Solutions

Ex. 14, 22, 23, 24 Allman Chapter 6.4 Ex. 1,3,5

Chapter 6.2: Probability Distributions in Genetics

6.2.14. a. $\mathcal{P}(\text{age at death} = 0) = 1/2$;

$$\mathcal{P}(\text{age at death} = 1) = (1/2)(3/4) = 3/8;$$

$$\mathcal{P}(\text{age at death} = 2) = (1/2)(1/4)(3/4) = 3/32;$$

$$\mathcal{P}(\text{age at death} = 3) = (1/2)(1/4)(1/4)(1) = 1/32.$$

These probabilities add to 1 since the events are disjoint and exhaust all possibilities.

b. $0(1/2) + 1(3/8) + 2(3/32) + 3(1/32) = .65625$

6.2.22. a. Progeny of the specified cross should display the dominant phenotype with probability $3/4$. The number of progeny in 1000 that display the dominant phenotype is a random variable with binomial distribution, so its expected value is $(1000)(3/4)=750$. We thus expect 750 dominant and 250 recessive phenotypes.

If 700 dominant and 300 recessive phenotypes are observed, then

$$\chi^2 = \frac{(700 - 750)^2}{750} + \frac{(300 - 250)^2}{250} = 13.333.$$

With $\alpha = .05$ and 1 degree of freedom, the critical value is $\chi_{crit}^2 = 3.841$. Since this is less than our computed value, we find the data is not in accord with the Mendelian model at the .05 significance level.

b. If instead 725 dominant phenotypes are observed in 1000 progeny, $\chi^2 = 3.333$. This is less than the critical value in part (a), so the data is in accord with the Mendelian model at the .05 significance level.

c. If N dominant phenotypes are observed in 1000 progeny, then

$$\chi^2 = \frac{(N - 750)^2}{750} + \frac{(1000 - N - 250)^2}{250} = \frac{4(N - 750)^2}{750}.$$

Thus the data is in accord with the Mendelian model at the .05 significance level when

$$\frac{4(N - 750)^2}{750} < 3.841$$

$$(N - 750)^2 < 720.1875$$

$$|N - 750| < 26.8363$$

$$723.1637 < N < 776.8363$$

Since N must be an integer, this means $724 \leq N \leq 776$.

- 6.2.23. In the table, moving across a row means α decreases (with the degrees of freedom not changing). Since a smaller α means a χ^2 value computed from data is *less likely* to exceed the critical value, the critical value must get larger. Moving down a column means the degrees of freedom increases (with α held fixed). Since a larger degrees of freedom means more terms are added to compute χ^2 , we expect χ^2 values to typically be larger. Thus for a fixed significance level, the critical value must be larger in order that computed χ^2 values exceed it the same percentage of times.
- 6.2.24. With 556 individuals included in the table, and independent assortment of the genes implying expected proportions of 9/16, 3/16, 3/16, and 1/16 of the phenotypes in the order listed, the expected frequencies are 312.75, 104.25, 104.25, and 34.75. This gives $\chi^2 = .47$. This is well below the critical value if α is taken to be .01, .05, or even .10. Thus the data is judged to be in accord with independent assortment at any of these significance levels.
- 6.4.1. a. Let p be the frequency of ct in the population. Then $p^2 = 9/450$, so $p \approx .1414$.
 b. The percentage of the population heterozygous for the gene is $2p(1-p) \approx .2428$.
- 6.4.3. a. $2p(1-p) = .4$ implies $p^2 - p + .2 = 0$, so $p = (1 \pm \sqrt{1 - .8})/2 = (1 \pm \sqrt{.2})/2 \approx .2764$ or $.7236$, with q being the other value.
 For $2p(1-p) = H$, the values of p and q are $(1 \pm \sqrt{1 - 2H})/2$.
 b. $H = 2p(1-p)$ is maximized when $p = 1/2$, $q = 1/2$. This can be seen either by graphing the parabola, or by using calculus.
- 6.4.5. a. Let p , q , and r denote the frequencies of the alleles I^A , I^B , and I^O . Assuming random mating in the population,

$$p^2 + 2pr = .32, \quad q^2 + 2qr = .15, \quad 2pq = .04, \quad r^2 = .49.$$

Solving these gives $r = .7$, $p = .2$, and $q = .1$. Note that even though there are 4 equations in only 3 unknowns here, these values makes all equations hold.

b. The equations to be solved are

$$p^2 + 2pr = .40, \quad q^2 + 2qr = .11, \quad 2pq = .05, \quad r^2 = .44.$$

From the last we find $r = .6633$. Then the first gives $p = .2532$, and the second gives $q = .0783$. With these values $2pq = .0396$, so the third equation is *not* satisfied. (Also, $p + q + r \neq 1$.) Thus the system has no exact solution.

It could be that the population is not in a Hardy-Weinberg equilibrium, or that the data is flawed. Given the relative ease of collecting bloodtype data, and the doubtfulness of the random mating assumption applying to the U.S. population, the first is more likely.