p8119_hw1_jsg2145

Jared Garfinkel

10/4/2020

```
tibble("Chapter" = c(2,3,4), "Problems" = c("4, 5, 6, 8", "1, 2, 4, 10", "1, 2, 3, 7, 8, 11, 12"))
```

```
## # A tibble: 3 x 2
##   Chapter Problems
##     <dbl> <chr>
## 1       2 4, 5, 6, 8
## 2       3 1, 2, 4, 10
## 3       4 1, 2, 3, 7, 8, 11, 12
```

Chapter 2

Problem 4

How many genotypes are possible with a 3-allele marker? With K alleles?

There are 3 alleles in a genotype, so they can take the form 1, 2, and 3. Each offspring can inherit one allele from each parent.

By hand, the possible genotypes are (1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3). There are 6 possible genotypes.

For K alleles, the number of possible genotypes is the sum of n from n = 1 to K.

Number of genotypes =

$$\Sigma_{n=1}^{K} n$$

Problem 5

Given:

Pr(Y = 1 | any variant) = 1 Pr(no variant) ~ 1 Pr(early onset AD) = 0.001 Pr(any variant | AD) = 0.07

Find:

Pr(disease | no variant) = Pr(AD | no variant)

Bayes Rule:

P(A|B) = $\frac{P(B|A)P(A)}{P(B)}$

$$\frac{Pr(no\ variant \mid AD)\ *\ Pr(AD)}{Pr(no\ variant)}$$

Pr(no variant | AD) = 1 - Pr(any variant | AD) = 1 - 0.07 = 0.93

0.93 x 0.001 / 1 = 0.00093

Problem 6

a) The probability of a random person being affected is the prevalence = 0.000001. If the person is affected, the sibling of

that person is more likely to have it, but the probability of the sibling being affected is still quite low.

b) Given:

Pr(disease | 2 variants) = 1/2 Pr(phenocopy) = 0

The probability of a person having the disease who is selected at random remains the same. However, the chances of the sibling having

the disease is even lower than before although still greater than the background probability of disease, since the genotype associated

with the disease only produces the disease phenotype in 1/2 of those with 2 variants.

Problem 8

Penetrance functions allow for differences in mean and variance as a function of genotype.

$f(y|\mu_G, \sigma^2_G)$

Chapter 3

Problem 1

p =

$$\frac{2n_{AA} + n_{Aa}}{2n}$$

$\text{var}(\hat{p})$ =

$$\frac{pq}{2n}$$

Problem 2

Given:

Pr(sickle cell anemia | dd) = 1

Pr(sickle cell anemia) = 0.01 = Pr(dd) = q^2

Pr(d) = ?

Using:

$q = \sqrt{q^2} = \sqrt{0.01} = 0.1$

Problem 4

In a population with minor allele frequency, p, and population N, a rare recessive disease would be expected in $N * p^2$ individuals.

If there is an inbreeding coefficient of F, the number affected by the disease would be calculated from the equation:

$$F * p \ + \ (1 \ - \ F) * p^2$$

Given:

p = 0.0001

N = 10^6

F = 0.2

Assuming p is the probability of the recessive allele, the expected number of indiviuals with this disease will be $0.2 * 0.0001 + (1 - 0.2) * 0.0001^2$

$= 20.008 = 20.008 \sim 21$

Problem 10

Prove:

var(X) = 2pq + var(p_k)

By definition:

p =

$$\sum_k s_k p_k$$

Using:

Pr(X=0) = p(aa) =

$$\sum_k s_k q_k^2$$

= 1 - 2p + E[p_k^2] - p^2 + p^2 = q^2 + var(p_k)

Pr(X=1) = P(Aa) =

$$2 \sum_k s_k p_k q_k$$

= 2p - 2E[p_k^2] + 2p^2 - 2p^2 = 2pq - 2var(p_k)

Pr(X=2) = P(AA) =

$$\sum_k s_k p_k^2$$

= E[p_k^2] - p^2 + p^2 = p^2 + var(p_k)

Therefore:

E[X] = 0 * (q^2 + var(p_k)) + 1 * (2pq - 2var(p_k)) + 2 * (p^2 + var(p_k)) = 2p^2 + 2pq = 2p

Var(X) = E[X^2] - $E[X]^2$ = 0 * (q^2 + var(p_k)) + 1 * (2pq - 2var(p_k)) + 4 * (p^2 + var(p_k)) - $E[X]^2$

Var(X) = 2pq + 2 var(p_k)

Chapter 4

Problem 1

$$\lambda_s = \frac{P(Y_1 \ = \ 1, Y_2 \ = \ 1)}{P(Y_1 \ = \ 1)^2}$$

$P(Y_1 = 1) = (1-p)^4$

$P(dd, dd) = \frac{\frac{1}{16}4p^2(1-p)^2 + \frac{1}{4}4p(1-p)^3 + (1-p)^4}{(1-p)^4}$

$P(dd, dd) = \frac{1}{4}p^2 - p + 1 = (\frac{1}{2}p - 1)^2$

Problem 2

Part a

Prove:

$Cov(Y_1, Y_2) = P(Y_1 = Y_2 = 1) - K^2$

Definitions:

$K = P(Y_1 = 1) = P(Y_2 = 1)$

$Cov(X, Y) = E[XY] - E[X]E[Y]$

$P(Y_1 = Y_2 = 1) - K^2 = P(Y_1 = 1, \ Y_2 = 1) - P(Y_1 = 1)P(Y_2 = 1)$

$E[Y_i] = 1 * P(Y_{i1} = 1) + 0 * P(Y_{i0} = 0) = P(Y_{i1} = 1)$

Therefore,

$Cov(Y_1, Y_2) = P(Y_1 = Y_2 = 1) - K^2$

Part b

Express λ_R as a function of covariance and K

$\lambda_R = P(Y_2 = 1 | Y_1 = 1)/K$

Definitions:

$P(A|B) = P(B|A)P(A)/P(B)$

Then,

$P(Y_2 = 1 | Y_1 = 1) = P(Y_2 = 1, Y_1 = 1)/P(Y_1 = 1)$

$P(Y_2 = 1, Y_1 = 1) = Cov(Y_1, Y_2) + K^2$

So,

$\lambda_R = \frac{Cov(Y_1, \ Y_2) + K^2}{K^2}$

Problem 3

We might expect dizygotic twins to have the same recurrence risk as siblings, because they are genetically unique. However, they share the same conditions in the womb and this may contribute to the observed relative risk in dizygotic twins bigger than that for siblings.

Problem 7

Given:

$h^2 = V_A/Var(Y) = 2\rho$

$X_1 = X_2$

$M = 1$

Definition:

$\rho_{XY} = \sigma_{XY}/\sigma_X\sigma_Y$

$Cov(Y_O, Y_P) = V_A/2$

Further, for identical twins,

$\sigma_X = \sigma_Y$

So,

$\rho_{XY} = Cov(Y_P, Y_O)/Var(Y_P) = \frac{1}{2}V_A/Var(Y_P)$

Therefore, the heritability, $h^2 = \frac{1}{2}2\rho = \rho$

Problem 8

Find the heritability of a trait in identical twins with a correlation of 0.8.

Since $h^2 = \rho$ in identical twins, the heritability of the trait is 0.8.

Problem 11

a) At least one parent had the dominant allele.

b) Autosomal dominant traits can skip a generation if the parents are both heterozygous.

c) The children of two parents with autosomal recessive traits will have the recessive trait.

d) An autosomal recessive trait cannot skip a generation unless one of the parents has the dominant trait.

e) The children will inherit at least one recessive allele from that parent.

Problem 12

Pedigree A: dominant, not recessive because 7 and 8 would have all recessive offspring.

Pedigree B: recessive, not dominant, because 3 and 4 would have homozygotic offspring since a dominant allele will cause disease in all carriers, and neither parent is a carrier in 3 and 4.

Pedigree C: recessive, not dominant, because carriers would have the disease and thus 3 and 4 would be homozygotic.

Pedigree D: dominant or recessive

Pedigree E: recessive, not dominant, because 13 and 14 would have no dominant allele to pass to 20.

Pedigree F: neither, because 3 and 4 would have all recessive alleles in the recessive model, while 6 and 7 would have all recessive alleles in the dominant model.