

POPULATION GENETICS PROBLEM #2

All of you are probably familiar with the story of sickle cell anemia. It's associated with a single amino acid substitution in the β chain of hemoglobin.¹ Individuals heterozygous for the “normal” and the “sickle” allele suffer from a mild anemia that makes them more resistant to malaria than those homozygous for the “normal” allele. Individuals homozygous for the “sickle” allele suffer from a severe, debilitating anemia. The “sickle” allele is common only in human populations where malaria is also very common. What you may not know is that there are other genetic polymorphisms in humans that provide a similar protective effect, for example, a polymorphism at the locus coding for glucose-6-phosphate dehydrogenase.

In a recent study of populations in Mali, Guindo et al. [?] compared the frequency of uncomplicated malaria with the frequency of severe malaria in individuals with different genotypes at the G6PD locus.² The G6PD locus is on the X chromosome, so men carry only one copy of the allele. The $A-$ allele had been shown in previous studies to confer some resistance to severe malaria in men who were hemizygous. In this study, women were also included. The data are summarized in the table below.

Condition	Men		Women		
	A	$A-$	AA	$AA-$	$A - A-$
Severe	217	15	170	27	3
Uncomplicated	1356	201	1375	221	44

For purposes of this problem we will assume (1) that 90% of those diagnosed with severe malaria do not survive and (2) that anyone who is going to get severe malaria contracts it before they have a chance to reproduce. Given those assumptions and the data above, answer the following questions.

1. What are the fitnesses of the two male genotypes and the three female genotypes?

¹This is the simplified version of the story. There are at least three alleles segregating at the β -globin locus in humans. Ask me about it if you want to hear more details.

²Uncomplicated malaria is not life-threatening. Severe malaria is associated with very high loads of *Plasmodium falciparum* (> 500,000 parasites per μ l) or with cerebral malaria, severe anemia, respiratory distress, or prostration.

Fitness in this context is probability of survival, so

$$\begin{aligned}
 w.m(A) &= 0.1 * 217 / (217 + 1356) + 1356 / (217 + 1356) = 0.876 \\
 w.m(A-) &= 0.1 * 15 / (15 + 201) + 201 / (15 + 201) = 0.938 \\
 w.f(AA) &= 0.1 * 170 / (170 + 1375) + 1375 / (170 + 1375) = 0.901 \\
 w.f(AA-) &= 0.1 * 27 / (27 + 221) + 221 / (27 + 221) = 0.902 \\
 w.f(A - A-) &= 0.1 * 3 / (3 + 44) + 44 / (3 + 44) = 0.943
 \end{aligned}$$

Corresponding Bayesian estimates (from WinBUGS, see code below) are

$$\begin{aligned}
 w.m(A) &= 0.876(0.860, 0.891) \\
 w.m(A-) &= 0.934(0.900, 0.961) \\
 w.f(AA) &= 0.901(0.886, 0.914) \\
 w.f(AA-) &= 0.900(0.861, 0.932) \\
 w.f(A - A-) &= 0.926(0.844, 0.979)
 \end{aligned}$$

2. Is there evidence for resistance to severe malaria in heterozygous females?

This calls for a χ^2 contingency test (or Fisher's exact test) comparing the frequency of severe malaria in women with genotypes AA and AA-. The data (observed) looks like this:

	AA	AA-	Total
Severe	170	27	197
Uncomplicated	1375	221	1596
Total	1545	248	1793

The expected values are calculated from the marginal frequencies.

	AA	AA-	Total
Severe	$1793 * (1545 / 1793) * (197 / 1793) = 169.75$	$1793 * (248 / 1793) * (197 / 1793) = 27.25$	$1793 * (197 / 1793) = 197$
Uncomplicated	$1793 * (1545 / 1793) * (1596 / 1793) = 1375.25$	$1793 * (248 / 1793) * (1596 / 1793) = 220.75$	$1793 * (1596 / 1793) = 1596$
Total	$1793 * (1545 / 1793) = 1545$	$1793 * (248 / 1793) = 248$	1793

So

$$\begin{aligned}
 \chi_1^2 &= \frac{(170 - 169.75)^2}{169.75} + \frac{(27 - 27.25)^2}{27.25} + \frac{(1375 - 1375.25)^2}{1375.25} + \frac{(221 - 220.75)^2}{220.75} \\
 &= 0.00299
 \end{aligned}$$

This value is much smaller than the critical value necessary for rejection of the null hypothesis that the frequency of severe malaria is the same in the two genotypes.

If you know R, the statistical package, calculating this statistic is a *lot* easier.

```
> x <- matrix(c(170,27,1375,221), byrow=TRUE, ncol=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: x
X-squared = 0.003, df = 1, p-value = 0.956
```

Alternatively, you can look at $w.f(AA)$ and $w.f(AA-)$ from the Bayesian analysis and notice that the two point estimates are virtually identical (0.901 *versus* 0.900) and that the 95% credible intervals are broadly overlapping, indicating that we have no evidence that $w.f(AA)$ and $w.f(AA-)$ are statistically distinguishable.

3. If this were the only form of selection acting on this locus in the populations from which these samples were taken, what would the ultimate genetic composition of the population be?

There is directional selection in favor of the A- allele in both males and females. If this were the only selection operating on this locus, we'd expect the A- allele to be fixed eventually.

4. **Bonus question:** What allele frequencies would you predict in males and females of the next generation? Pretend that generations are non-overlapping and that everything else about Hardy-Weinberg applies, except for the selection you've just estimated.³

The frequency of A- in males and females after selection are

$$\begin{aligned}
 p.m.after &= \frac{x.m(A-) * w.m(A-)}{x.m(A-) * w.m(A-) + x.m(A) * w.m(A)} \\
 &= \frac{226 * 0.935}{226 * 0.934 + 1573 * 0.876} \\
 &= 0.133 \\
 p.f.after &= \frac{x.f(A-A-) * w.f(A-A-) + x.f(AA-) * w.f(AA-)/2}{x.f(A-A-) * w.f(A-A-) + x.f(AA-) * w.f(AA-) + x.f(AA) * w.f(AA)} \\
 &= \frac{47 * 0.926 + 248 * 0.900/2}{47 * 0.926 + 248 * 0.900 + 1545 * 0.901} \\
 &= 0.0935
 \end{aligned}$$

³Remember that since G6PD is sex-linked, male offspring will get their one allele from their mother, while female offspring will get one allele from their mother and one from their father.

In the next generation, males receive their chromosomes from their moms, so

$$p.m' = 0.0935 \quad .$$

Females receive one chromosome from their mom and one from their dad, so

$$\begin{aligned} p.f' &= \frac{p.f.after + p.m.after}{2} \\ &= \frac{0.0935 + 0.133}{2} \\ &= 0.113 \quad . \end{aligned}$$

The estimates from WinBUGS are very close to these values:

$$\begin{aligned} p.m' &= 0.0942(0.0836, 0.1053) \\ p.f' &= 0.111(0.102, 0.121) \quad . \end{aligned}$$

WinBUGS code

```
model {
  # likelihoods for disease
  for (i in 1:2) {
    sick.m[i] ~ dbin(x.m.sick[i], n.m[i])
  }
  for (i in 1:3) {
    sick.f[i] ~ dbin(x.f.sick[i], n.f[i])
  }

  # priors for disease frequencies
  for (i in 1:2) {
    x.m.sick[i] ~ dunif(0,1)
  }
  for (i in 1:3) {
    x.f.sick[i] ~ dunif(0,1)
  }

  # fitnesses
```

```

for (i in 1:2) {
  w.m[i] <- w*x.m.sick[i] + (1.0-x.m.sick[i])
}
for (i in 1:3) {
  w.f[i] <- w*x.f.sick[i] + (1.0-x.f.sick[i])
}

# likelihoods for genotype frequencies
n.m[1] ~ dbin(x.m, n.males)
n.f[1:3] ~ dmulti(x.f[], n.females)
n.males <- sum(n.m[])
n.females <- sum(n.f[])

# priors for genotype frequencies
x.m ~ dunif(0,1)
for (i in 1:3) {
  a[i] ~ dexp(1)
  x.f[i] <- a[i]/sum(a[])
}

# allele frequencies before selection
p.m.before <- x.m
p.f.before <- x.f[1] + x.f[2]/2

# allele frequencies after selection
p.m.after <- x.m*w.m[1]/(x.m*w.m[1] + (1-x.m)*w.m[2])
p.f.after <- (x.f[1]*w.f[1] + x.f[2]*w.f[2])/2
              /(x.f[1]*w.f[1] + x.f[2]*w.f[2] + x.f[3]*w.f[3])

# allele frequencies in next generation
p.m.prime <- p.f.after
p.f.prime <- (p.m.after + p.f.after)/2
}

list(sick.m=c(15,217), n.m=c(216,1573), sick.f=c(3,27,170),
      n.f=c(47,248,1545), w=0.1)

```