HANDOUT I.4. Comparing Hypothesis: genes and longevity

Yashin et al.[1] compared frequencies of various genetic markers among centenarians to frequencies in younger individuals. At the SOD1 locus, which codes for a sodium oxide dismutase, allele D21S1 was found to occur with frequency 0.025 among younger individuals. In a random sample of 197 centenarians, 19 were found to carry the D21S1 allele. Is this allele associated with greater longevity? If not then the difference in frequency between younger individuals and the sampled centenarians should be readily explained by sampling variation alone. If the allele is associated with greater longevity, then the allele should much more frequent among the centenarians. We will use the likelihood ratio test to examine this hypotheses.

Under the null hypothesis (H_0: no effect on longevity), D21S1 should have the same frequency among younger individuals (p) as centenarians (p^*). That is, under H_0 $p^* = p = 0.025$ and the probability of the centenarian data ($n_1 = 19$ carriers and $n_2 = 197–19 = 178$ non-carriers) follows a binomial distribution, giving the likelihood

(1) $$L\left(p^* = p = 0.025 \middle| n_1 = 19, n_2 = 178\right) = \binom{197}{19}(0.025)^{19}(0.975)^{178}$$

Now suppose D21S1 carriers are w times more likely to live to age 100 than non-carriers. This means the ratio of carriers to non-carriers among centenarians compared to the same ratio among younger individuals is

(2) $$\frac{p^*}{1-p^*} = w\frac{p}{1-p}.$$

Solving (2) for p^* shows that the frequency of D21S1 among centenarians would be $p^* = pw/\left(pw+1-p\right)$. Notice that $p^* = p$ when $w = 1$, so H_0 corresponds to setting $w = 1$.

In general, the likelihood of the centenarian data is

(3) $$L\left(p^* = \frac{pw}{pw+1-p}\middle| n_1, n_2\right) = \binom{n_1+n_2}{n_1}\left(\frac{pw}{pw+1-p}\right)^{n_1}\left(\frac{1-p}{pw+1-p}\right)^{n_2}$$

From (3), it can be shown[2] that the maximum likelihood estimate of w is

(4) $$\hat{w} = \frac{n_1}{n_2}\frac{1-p}{p} = \frac{19}{178}\frac{0.975}{0.025} = 4.2$$

That is, carriers of D21S1 are more than 4 times as likely as non-carriers to become centenarians. But is the apparent advantage statistically meaningful? From (1) and (3) using $p = 0.025$, $n_1 = 19$, and $n_2 = 178$, the likelihood ratio statistic is

(5) $$G = 2\ln\left[\frac{\binom{197}{19}(0.097)^{19}(0.903)^{178}}{\binom{197}{19}(0.025)^{19}(0.975)^{178}}\right] = 24.2$$

Since the test statistic G exceeds the critical value $\chi^2_{1,.0.001} = 10.83$, the data strongly suggest that carriers of DS121 live longer than non-carriers ($P < 0.001$).

[1] A.I. Yashin et al. 2000. Genes and longevity: lessons from studies of centenarians. Journal of Gerontology 55A:B319-B328.

[2] From (3), the log-likelihood has the form

$$l(w) = n_1 \ln(pw) - (n_1 + n_2)\ln(pw + 1 - p) + [\text{terms with no } w]$$

so

$$\frac{dl(w)}{dw} = \frac{n_1}{w} - \frac{(n_1 + n_2)p}{pw + 1 - p}.$$

Setting $dl(w)/dw$ = 0 and solving for w gives the maximum-likelihood estimate (4).