Question 9 B242 2003 - Model answer

It does seem to me that students of "evolution" -- meaning changes in gene frequencies -- ought to be able to estimate gene frequencies from a set of genotypic data, and estimate the change between two generations if given the data from those two generations. This is the purpose of this exercise.

			A1A1	A1A2	A2A2	Total	
HARDY-WEINBERG CALCULATIONS 2003:							
			A1A1	A1A2	A2A2	Total	
	Selection model	(in i)					
	O		1231	76	1	1308	
(a)	p(C)=	**0.0298**	[= (1231 + 76/2)/1308]				**(2½ marks)**
	p(c)=	**0.9702**					**(2½ marks)**
	Sum(check)	1.0000					
(b)	Exp Genot freqs		**0.9413**	**0.0579**	**0.0009**		**(2 marks)**
	E		1231.1628	75.6743	1.1628	1308.0000	(sum check)
(c)	**X^2**	**(2 marks)**	**0.0000**	**0.0014**	**0.0228**	**0.0242**	total χ^2
(d)		**(3 marks)**				**P approx 0.9 (>0.05)**	

Common mistakes: (1) Not being able to calculate expected numbers. Expected numbers are not typically whole numbers: if I toss a coin three times, I expect 1½ heads, right? If I said it's 1 head or it's 2 heads, I would be wrong!   (2) Thinking that a low value of χ^2 means highly significant, or vice-versa. Remember: if χ^2 is big, |O-E| is big, so O is *FAR* from E! (3) Not being able to express what the P value means. (4) I did not take marks off for "we therefore accept the null hypothesis", but I don't like this way of expressing it! All we know is that *the data are not sufficient to reject the null hypothesis*. This doesn't mean we have to "accept" it, since more data might show that it was clearly rejected. You either reject the null hypothesis or you don't. You can't really "accept" something on this basis.

e) Chi-square contingency table analysis **(4 marks)**

	Obs				
	no. alleles	A1	A2	sum	
	1955	623	7	630	
	1956	2538	78	2616	
	Total	3161	85	3246	
	Expected				
	no. alleles				
	1955	613.5028	16.49723		
	1956	2547.497	68.50277		
				3246	
		chi-square			
		0.14702	5.467423		
		0.035406	1.316696		
f)	**(1 mark)**			6.966546	1 df

Expected values are here calculated as follows:

For the A1 allele, the expected number in 1955 is 3161 x 630/3246. And so on.

Common mistakes: 1) the alleles add to 315 in 1955 and 1308 in 1956, leading to 311½ A1 alleles in 1955, etc. NO! The populations are diploid, so there are two alleles per individual! 2) Not knowing how to do a 2x2 contingency table analysis. 3) Doing the analysis on allele frequencies, rather than numbers.

g) Yes, because P<0.01! **(1 mark)**

h) **(3 marks)**

	p	q
	p	q
1955	0.988889	0.011111
1956	0.970183	0.029817
$\Delta p=$	**− 0.01871**	

i) Drift, or selection. However, the change is quite significant (see above), as is the estimated selection (see below) and the population sizes (at least judging by the large sample sizes, which are presumably much smaller than the actual population) are very large, so it is unlikely to be drift. More likely selection. Note: the sample sizes do not tell you anything about the total population, necessarily, so using "population size" changes as evidence for "relaxed selection" is bogus. **(2 marks)**

j) **(2 marks)**

	$\Delta p=$	$-sq/(1-s)$	
Therefore	$s=$	$\Delta p/(\Delta p-q)$	**= 0.62735**

(a large selection pressure)

Common mistakes. (1) Not being able to do simple algebra rearrangement. One person even iterated to get the answer! (but got it right so I gave it full marks) (2) Dropping a minus sign leading to selection values greater than one. (3) Forgetting that Δp is negative, and ALSO dropping a minus sign, to get the correct answer overall.