We have a Genome Wide Association Study (GWAS) with 50 people. Each person has 2 alleles at LPA variant chr6:160985526 (because everyone has 2 copies of every chromosome).

1. What is the frequency of the G nucleotide in this population? (We call this the allele frequency)

A A A A A A A A A A A A A A A A A A A A A A A A A
A G A A A A A A A A A A A A A A A A A A A A A A A
A A G A A A A A A A A A A A A A A G A A A G A A
A A A A A A A A G A A A A A A A A A G A A G A A

2. The red letters belong to people who have coronary artery disease (CAD) and the blue letters belong to healthy controls. Assume every person with CAD has the genotype A/G, what is the prevalence of CAD in this population? (Hint: prevalence=Number of people with disease/total number of people)

3. Complete the following 2x2 table with counts for each category.

	A	G
Case		
Control		

Bonus: Perform a Chi-Square test on the table. In every cell, compare the observed counts to the expected using the following expected table and the observed table in #3. If $\chi^2 > 3.84$ we can reject the null hypothesis that DNA nucleotide and CAD are independent, assuming $\alpha = 0.05$. In other words, we could say the letter G is associated with increased risk of CAD.

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

	A	G
Case	13	1
Control	80	6