Math Minute 3.1 How Can You Tell if Base Compositions Are Different?

To say that the ratios of the four bases in the O-islands and K-islands are different from the backbone ratio, you must do some statistical analysis. You really want to be able to say they are *significantly* different. The standard tool for testing whether two sets of ratios, or frequencies, are significantly different is the **chi-square test** of homogeneity.

To illustrate how the chi-square test is performed, suppose that you had a 4 kb sequence containing 1,000 bases of each type, and a 3 kb sequence containing 600 A's, 800 C's, 700 G's, and 900 T's (Table MM3.1). For each of the eight cells in the interior of the table, compute the expected frequency of that cell as the cell's row total times the cell's column total divided by the grand total. For example, the expected frequency of A in Sequence 1 is $1{,}600 \times 4{,}000/7{,}000 \approx 914.29$. Subtract the expected frequency from the actual frequency ($1{,}000 - 914.29 \approx 85.71$), square the result ($85.71^2 \approx 7{,}346.2$), and divide this number by the expected frequency ($7{,}346.2/914.29 \approx 8.03$). Repeat this process for each of the eight cells, and sum the results. If this sum is larger than a cut-off value from the chi-square distribution, the frequencies are said to be significantly different.

This is the math behind the scenes of the simple statement that the ratios of the four bases in most O-islands and K-islands are different from the ratios in the backbone. The investigators performed the chi-square test for 108 O-islands longer than 1 kb, and found 101 of them to have a base composition significantly different from that of the backbone.

Table MM3.1 Base frequencies.

Base	Sequence 1	Sequence 2	Total
A	1,000	600	1,600
C	1,000	800	1,800
G	1,000	700	1,700
T	1,000	900	1,900
Total	4,000	3,000	7,000