Group 6:
Tiffany Cook
Drashti Mehta
Joshua Mikombo

BINF 6201 - Lab 7

Question 1: Calculate the transition probabilities for the CpG data set.

	A	C	G	T
A	0.1875	0.25	0.4375	0.125
C	0.1428	0..3714	0.2857	0.2
G	0.1714	0.3429	0.3714	0.114
T	0.0714	0.4286	0.4286	0.0714

Question 2: Calculate the transition probabilities for the non-CpG data set.

	A	C	G	T
A	0.3448	0.1724	0.2414	0.2414
C	0.35	0.25	0.05	0.35
G	0.3	0.2	0.25	0..25
T	0.1935	0.1935	0.2258	0.3871

Question 3: Get the log odds ratio for a short test sequence: **CGCG**

- **Log odds ratio: *log*((0.2857/0.05)*(0.3429/0.2)*(0.2857/0.05)) = 1.748**

Question 4: Get the log odds ratio for a 100 bp sequence

- CG occurrences: 8; GC occurrences: 12

- **Log odds ratio: *log*(8(0.2857/0.05)*(12(0.3429/0.2))) = 2.973**

Question 5: Find the CpG island region in an unknown sequence

- Question 5A. How many total windows are you going to need to test?
  - **We'll need 19 total windows**
- Question 5B. Does the first 100 bp window fit the CpG model or the non-CpG model?
  - **Because the GC ratio is less than 0.6 (0.36) the first 100bp window fits the non-CpG model**

Question 5C. Calculate the log odds ratio for all of the windows. How many have an odds ratio >0?

- CG total occurrences: 37; GC total occurrences: 67
  - **Window 1 (4.58)**
  - **Window 2 (3.38)**
  - **Window 3 (0)**
  - **Window 4 (0)**
  - **Window 5 (2.97)**
  - **Window 6 (3.668)**
  - **Window 7 (4.36)**
  - **Window 8 (3.89)**
  - **Window 9 (0)**
  - **Window 10 (5.97)**
  - **Window 11 (7.18)**
  - **Window 12 (7.31)**
  - **Window 13 (7.11)**
  - **Window 14 (6.53)**
  - **Window 15 (6.53)**
  - **Window 16 (6.28)**
  - **Window 17 (4.92)**
  - **Window 18 (5.28)**
  - **Window 19 (5.68)**
- **16 windows have an odds ratio of >0**

Which window has the highest odds ratio value?

  - **Window 12 (7.31)**

Question 5D. Roughly **where** in the unknown sequence is the location of the CpG island? (i.e., what are the start and end position(s) of the windows that show evidence for the CpG model)?

- **Window 11: start & end positions → 501-600**
- **Window 12: start & end positions → 551-650**
- **Window 13: start & end positions → 601-700**
- **Window 14: start & end positions → 651-750**
- **Window 15: start & end positions → 701-800**
- **Window 16: start & end positions → 751-850**