Collin Yen

804034634

Project 1

**Problem 1**

**Part A:**

Case = P(d) \* N = 0.1 \* 1000 = **100**

**Part B:**

**Pseudo Code:**

dataIn[] = Generate 1000x200 matrix where 0 = no minor allele and 1 = minor allele

controlMinorAllele[] = binomial distribution with p = 0.25 and n = 1 and size = 900

caseMinorAllele[] = binomial distribution with p = 0.95 and n = 1 and size = 100

hasMinor[] = controlMinorAllele + caseMinorAllele

Assign first index of dataIn to be hasMinor since s1 is the causal SNP

**Part C:**

**Pseudo Code:**

Using the data generated from above:

For each SNP

Calculate the significance of that SNP

Return the max SNP and compare it to the threshold given by alpha

**Significance =** 20.034537475

**Threshold =** 3.480756

Significance > Threshold so we reject the null hypothesis

**Part D:**

**Number of pairs:** 39

**Tag Selection:** [9, 15, 70, 84, 123, 170, 181, 187, 0, 1, 7, 10, 27, 41, 42, 54, 55, 75, 77, 78, 86, 88, 94, 98, 103, 112, 120, 121, 135, 182, 197, 2, 3, 4, 5, 6, 8, 11, 12, 13, 14, 16, 18, 19, 20, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 72, 73, 74, 76, 79, 80, 81, 82, 85, 90, 91, 92, 93, 95, 96, 99, 101, 102, 104, 105, 106, 107, 109, 110, 111, 113, 114, 115, 117, 122, 125, 126, 128, 129, 131, 132, 136, 137, 138, 139, 140, 141, 142, 143, 147, 148, 149, 150, 151, 152, 154, 155, 156, 158, 159, 160, 162, 164, 165, 168, 169, 171, 172, 174, 175, 176, 177, 179, 180, 183, 184, 185, 186, 190, 191, 192, 193, 195, 196, 198]

**Pseudo Code:**

Determine the indices where the abs cor is over 0.1

For each correlation pair (x, y) in SNPs

Add the associated SNP to the dictionary map

Determine number of correlated SNPs for each SNP

While there are unmatched SNPs::

Find the SNP with the maximum correlated SNPs

Add the SNP to the list of tags

Set the number of Correlated SNPs for each of the associated SNP for the found tag to 0

Return the list of found tags

**Part E:**

**Power:** 0.007263

**Pseudo Code:**

Calculate the SNP of s0

For each SNP

Find the correlation between tag s0 and the current tag

Compute the NCP of the current tag

Calculate the power for this SNP

Add it to the total power

Return the average of the SNP powers