# N0 and N1 calculation

### N0 and N1 Calculations

Let $X = (X_1, X_2, ..., X_m$ be the informatiopn on the number of copies of m cSNV, e.g. $X = (0, 0, 1)$ or $X = (0, 1, 1)$ for m = 3.

m = number of cSNV's in the model.

$X_j \in 0, 1, 2$ for j = 1,2, . . . , m

Logistic model to measure the effect of number of copies of cSNV on the probability of having disease:

$logit(P(D|X)) = \beta_0 + \beta_1 \sum_{i=1}^{m} X_i$

$P(D)$ is the probability of disease in the population of 3100 individuals.

$P(D) = \int P(D|X)P(X)dX = E[P(D|X)] = \frac{1}{N} \sum_{j=1}^{N} P(D|X_j)$

where $P(D|X_j)$ is the probability of disease for the j-th individual. Let us assume each person carries zero or one cSNV, then:

if $\sum_{i=1}^{m} X_i = 0$ we have, $logitP(D|X) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$

if $\sum_{i=1}^{m} X_i = 1$ we have, $logitP(D|X) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$

and we can write P(D) in the following form:

$P(D) = pd = \frac{1}{N} \sum_{j=1}^{N} P(D|X_j) = \frac{1}{N}(N_0 \times a + N_1 \times b)$

where in this formula:

$a = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
$b = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$

As $N_0 + N_1 = 3100$ we can calculate $N_1$ as follows:

$N_1 = \frac{3100 \times (pd-a)}{b-a}$

A function to calculate $N_1$ for different values of pd:

```
N1_Calc = function(pd){

  beta0 = -10
  beta1 = 16

  a = exp(beta0)           /   (  1 + exp(beta0)          )
  b = exp(beta0+beta1)     /   (  1 + exp(beta0+beta1)    )

  res = ( 3100 * (pd - a) ) / (b - a)
  return(res)

}
```

The $N_1$ value if the probability of disease in the population is 0.05:

```r
round( N1_Calc(0.05) )
```

```
## [1] 155
```