

# Population Genetics Basics

*Kajal Claypool*

```
mat = matrix(c(270,436,957,1398,771,1170), nrow = 2)
colnames(mat) = c('CC', 'CT', 'TT')
rownames(mat) = c('Bipolar', 'Controls')
mat
```

```
##           CC   CT   TT
## Bipolar  270  957  771
## Controls 436 1398 1170
```

## (1) Allele Frequency

```
# Populations
pop_bip = sum(mat[1,])*2
pop_ctrl = sum(mat[2,])*2
```

```
# Bipolar - C frequency
c_bip = 270*2 + 957
c_bip/pop_bip
```

```
## [1] 0.3746246
```

```
# Bipolar - T frequency
t_bip = 957 + 771*2
t_bip/pop_bip
```

```
## [1] 0.6253754
```

```
# Controls - C frequency
c_ctrl = 436*2 + 1398
c_ctrl/pop_ctrl
```

```
## [1] 0.3778296
```

```
# Controls - T frequency
t_ctrl = 1398 + 1170*2
t_ctrl/pop_ctrl
```

```
## [1] 0.6221704
```

```
# ==> T is the major allele
```

## Hardy Weinberg

Hardy-Weinberg

$$p^2 + 2pq + q^2 = 1$$

$$p = \frac{(2X + Y)}{(2X + 2Y + 2Z)}$$

$$q = 1 - p$$

Observed:

- $TT = X$
- $CT = Y$
- $CC = Z$
- $N = X + Y + Z$

Expected:

- $E(TT) = N \cdot p^2$
- $E(CT) = N \cdot 2pq$
- $E(CC) = N \cdot q^2$

```
# Controls
X = mat[2,1]
Y = mat[2,2]
Z = mat[2,3]
N = X + Y + Z

p = (2*X + Y)/(2*X + 2*Y + 2*Z) ## same as # Controls - C frequency above
q = 1-p

EX = N * p^2
EY = N * 2*p*q
EZ = N * q^2

matrix(c(mat[2,], EX, EY, EZ), byrow = T, nrow = 2)

##          [,1]      [,2]      [,3]
## [1,] 436.0000 1398.000 1170.000
## [2,] 428.8366 1412.327 1162.837

hw_chi2 = (((X-EX)^2)/EX) + (((Y-EY)^2)/EY) + (((Z-EZ)^2)/EZ)
pchisq(hw_chi2, df=1, lower.tail = F)

## [1] 0.5782178

# Deviation from HW is unlikely (p = 0.58)

## Odds Ratio
mat2 = matrix(c(c_bip, c_ctrl, t_bip, t_ctrl), nrow = 2)
colnames(mat2) = c('Minor', 'Major')
rownames(mat2) = c('Bipolar', 'Controls')
mat2

##           Minor Major
## Bipolar   1497  2499
## Controls  2270  3738

or = (mat2[1,1]*mat2[2,2])/(mat2[1,2]*mat2[2,1])
or

## [1] 0.9864362
```

Association test

```
chisq.test(mat2)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  mat2
## X-squared = 0.091794, df = 1, p-value = 0.7619
# p-value = 0.76 ==> no statistically significant associations
```