GENE 210: Genomics and Personalized Medicine

How to do a GWAS

By: Alicia Martin, adapted from notes by Konrad Karczewski

For one trait, bitter taste ability, the results from last year are as follows:

```
            X4988235
bitter       AA AG GG NULL
  bitter_no   2  3  7    1
  bitter_yes  4  9 11    1
==================================
            X7495174
bitter       AA AG GG NULL
  bitter_no   9  2  1    1
  bitter_yes 13  8  3    1
==================================
            X713598
bitter       CC CG GG NULL
  bitter_no   7  4  1    1
  bitter_yes  1 13 10    1
==================================
            X17822931
bitter       CC CT NULL TT
  bitter_no   6  3    1  3
  bitter_yes 13  7    1  4
==================================
            X4481887
bitter       AA AG GG NULL
  bitter_no   0  6  6    1
  bitter_yes  6 10  8    1
==================================
```

Now, let's determine the association between bitter taste and the first SNP (rs4988235) using a χ^2 test:

```
          X4988235
bitter       AA AG GG NULL
  bitter_no   2  3  7    1
  bitter_yes  4  9 11    1
```

1. Get **observed** allele counts
   - There are 2 As for AA homozygotes and 1 A for AG heterozygotes. Similarly, there is 1 G for AG heterozygotes and 2 Gs for GG homozygotes.

	Bitter taster	**Non-taster**	**TOTAL**
A	2*4+1*9 = 17	2*2+1*3 = 7	24
G	2*11+1*9 = 31	2*7+1*3 = 17	48
TOTAL	48	24	72

2. Get observed frequencies
   - Normalize by the total

	Bitter taster	**Non-taster**	**TOTAL**
A	17/72 = 23.6%	7/72 = 9.7%	33.3%
G	31/72 = 43.1%	17/72 = 23.6%	66.7%
TOTAL	66.7%	33.3%	100%

3. Get expected frequencies
   - Treat events as independent and calculate the contingency table.
   - I.e. P(Bitter taster ∩ A) = P(Bitter taster)*P(A)

	Bitter taster	**Non-taster**	**TOTAL**
A	.333*.667=22.2%	.333*.333=11.1%	33.3%
G	.667*.667=44.5%	.333*.667=22.2%	66.7%
TOTAL	66.7%	33.3%	100%

4. Get **expected** counts
   - Multiply the frequency by the total allele counts

	Bitter taster	**Non-taster**	**TOTAL**
A	.222*72=15.98	.111*72=8.00	23.98
G	.445*72=32.04	.222*72=15.98	48.02
TOTAL	48.02	23.98	72

5. Use the χ^2 equation. Then lookup χ^2 value to get p-value using 1 degree of freedom (can plug this number into an online calculator, here for example: http://www.danielsoper.com/statcalc3/calc.aspx?id=11).
   - $\chi^2 = \sum(O - E)^2/E$
   - O = observed, E = expected

$\chi^2 = (17 - 15.98)^2/15.98 + (7 - 8)^2/8 + (31 - 32.04)^2/32.04 + (17 - 15.98)^2/15.98$
$= 0.07 + .13 + .03 + .07$
$= \mathbf{0.3}$
$P(\chi^2=0.3) = \mathbf{0.58}$

Interpretation

These SNPs are all strong affects and were likely discovered in GWAS with 100s of thousands of markers (let's say 500,000). If we were only looking at one SNP, this would not be significant at α=0.05. Right now, we're looking at 5 SNPs, so we would require a p-value of 0.05/5 = 0.01. However, these were non-randomly ascertained from many SNPs. If we want a 5% false discovery rate (FDR) across 500,000 SNPs, we need a p-value of 0.05/500,000 = $1 * 10^{-7}$. Therefore, rs4988235 is not significantly associated with bitter taste.

Class GWAS on your own

Now that you know how to calculate the significance of association between a phenotype (bitter taste) and a genotype (rs4988235), use the combined class data from this year and last year to determine whether any SNPs are significantly associated with bitter taste.

Fisher's exact test

We typically use a χ^2 test as it is highly reliable and easy to calculate for large sample sizes. However, the χ^2 test is only an approximation of the significance of the results, since the sampling of these data are not exactly equal to a χ^2 distribution. Therefore, we would ideally use a Fisher's exact test, especially when sample sizes are small. Fisher's exact test is too complicated to calculate by hand for this exercise, but we could use a program such as R or MATLAB to do the test.

Examples in R

- `chisq.test(matrix(c(top-left, top-right, bottom-left, bottom-right), nrow=2), correct=FALSE), i.e.`
- `chisq.test(matrix(c(17,7,31,17), nrow=2), correct=FALSE)`
    - Normally we would not set correct=FALSE, but on paper we weren't applying a continuity correction, so this will give us the results we calculated by hand.
- `fisher.test(matrix(c(17,7,31,17), nrow=2))`

Output

```
     Pearson's Chi-squared test

data:  matrix(c(17, 7, 31, 17), nrow = 2)
X-squared = 0.2812, df = 1, p-value = 0.5959
These are in line with our on paper calculations! Woohoo!
```

Odds ratios

Once we find a SNP that is significantly associated with the trait of interest, we can compute the odds ratio for that SNP. This is found using a simple application of Bayes' rule, where the probability of having a trait given a genotype

$$P(Trait \mid A)$$

is the probability of having the trait and the genotype, divided by the probability of having A.

$$P(Trait \mid A) = \frac{P(Trait \ \& \ A)}{P(A)}$$

Next, we can convert this probability to an odds by dividing it by 1---itself (Think "Vegas": 75% chance means a 3:1 odds of it happening). So:

$$Odds(A) = \frac{P(Trait \mid A)}{1 - P(Trait \mid A)}$$

Then, we find the odds ratio by dividing the odds of both genotypes:

$$Odds\ Ratio = \frac{Odds(A)}{Odds(B)}$$

P(Taster I A) = 17/(17+7) = .708
P(Taster I G) = 31/(31+7) = .816
Odds(A) = 2.42
Odds(G) = 4.43
Odds ratio = Odds(A)/Odds(G) = .55

This does not exactly mean that people with the A allele are .55 times less likely to have the trait. We'll discuss the exact application of these onto a personal genome (including different risk models such as likelihood ratios), but for now, just consider that the A allele is not significantly associated with bitter taste.