HWE TESTS

Hardy-Weinberg Law

For a random mating population, expect that genotype frequencies are products of allele frequencies.

For a locus with two alleles, A, a:

$$\begin{aligned} P_{AA} &= (p_A)^2 \\ P_{Aa} &= 2p_A p_a \\ P_{aa} &= (p_a)^2 \end{aligned}$$

This is equivalent to $f = 0$.

Inference about HWE

If $\widehat{f}$ is the MLE of the within-population inbreeding coefficient f, it has a normal distribution for large sample sizes n. It can be transformed into a standard normal variable z by

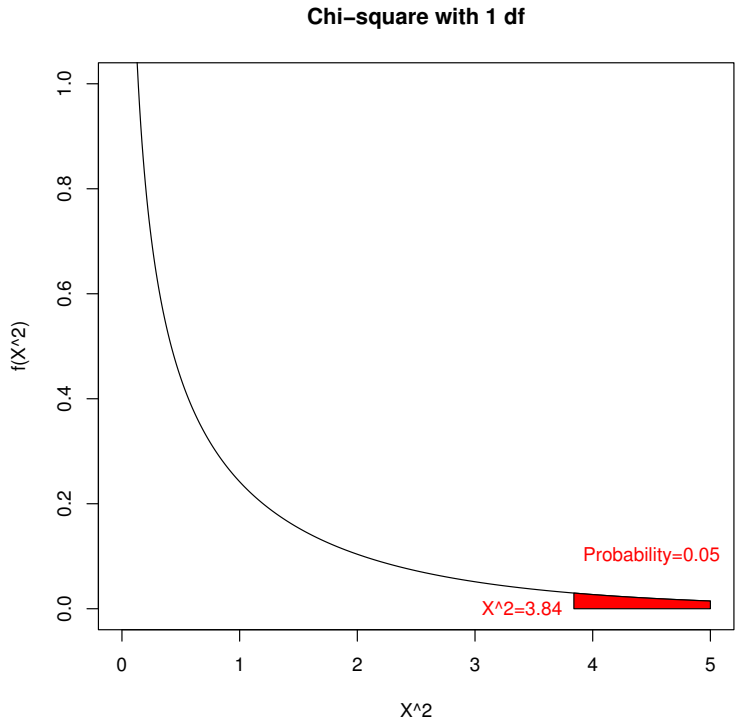
$$z = \frac{\widehat{f} - f}{\sqrt{\mathrm{Var}(\widehat{f})}}$$

If the true value f is zero, then $\mathrm{Var}(\widehat{f}) = 1/n$, and $X^2 = z^2$ has a chi-square distribution with one degree of freedom:

$$X^2 = \left(\frac{\widehat{f} - 0}{\sqrt{1/n}}\right)^2 = n\widehat{f}^2 \sim \chi 2_{(1)}$$

The HWE hypothesis is rejected at the 5% significance level if $X^2 > 3.84$.

Significance level of HWE test



The area under the chi-square curve to the right of $X^2 = 3.84$ is the probability of rejecting HWE when HWE is true. This is the significance level of the test.

Example

Does a sample of 6 AA, 3 Aa, 1 aa support Hardy-Weinberg?

First need to estimate allele frequencies:

$$\tilde{p}_A = \tilde{P}_{AA} + \frac{1}{2}\tilde{P}_{Aa} = 0.75$$

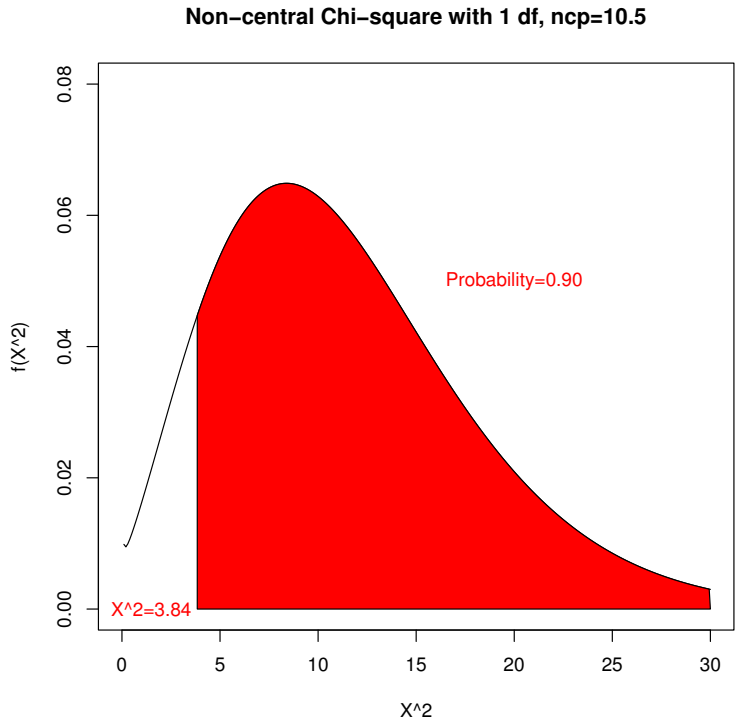
$$\tilde{p}_a = \tilde{P}_{aa} + \frac{1}{2}\tilde{P}_{Aa} = 0.25$$

$$\hat{f} = 1 - \frac{\tilde{P}_{Aa}}{2\tilde{p}_A\tilde{p}_a} = 0.2$$

$$X^2 = n\hat{f}^2 = 0.4 \text{ ns}$$

Sample size determination

When the Hardy-Weinberg hypothesis is not true, the test statistic $n\widehat{f}^2$ has a non-central chi-square distribution with one degree of freedom (df) and non-centrality parameter $\lambda = nf^2$. To reach 90% power with a 5% significance level, for example, it is necessary that $\lambda \geq 10.51$.



Significance Levels and p-values

The *significance level* α of a test is the probability of a false rejection. It is specified by the user, and along with the null hypothesis, it determines the rejection region. The specified, or "nominal" value may not be achieved for an actual test.

Once the test has been conducted on a data set, the probability of the observed test statistic, *or a more extreme value*, if the null hypothesis is true is the p-*value*. The chi-square and normal tests shown above give approximate p-values because they use a continuous distribution for discrete data.

An alternative class of tests, "exact tests," use a discrete distribution for discrete data and provide accurate p-values. It may be difficult to construct an exact test with a particular nominal significance level.

HWE Exact Test

If the counts of genotypes AA, Aa, aa are n_{AA}, n_{Aa}, n_{aa} in a sample of n individuals, and if the sample allele counts are $n_A = 2n_{AA} + n_{Aa}$ and $n_a = 2n_{aa} + n_{Aa}$, then the probability of the genotypic data *conditional on the allele counts* if there is HWE is

$$\Pr(n_{AA}, n_{Aa}, n_{aa} | n_A, n_a) = \frac{n!}{n_{AA}! n_{Aa}! n_{aa}!} \frac{2^{n_{Aa}} n_A! n_a!}{(2n)!}$$

HWE is rejected if this probability is amongst the smallest probabilities for all possible sets of genotype counts for those allele counts.

The p-value for the dataset is this probability plus probabilities for other possible sets of genotype counts that are smaller than this probability.

Exact HWE Test Example

As another example, the sample with $n_{AA} = 6, n_{Aa} = 3, n_{aa} = 1$ has allele counts $n_A = 15, n_a = 5$. There are two other sets of genotype counts possible and the probabilities of each set for a HWE population are:

n_{AA}	n_{Aa}	n_{aa}	n_A	n_a	$\Pr(n_{AA}, n_{Aa}, n_{aa} \mid n_A, n_a)$
7	1	2	15	5	$\frac{10!}{7!1!2!}\frac{2^1 15!5!}{20!} = \frac{15}{323} = 0.047$
6	3	1	15	5	$\frac{10!}{6!3!1!}\frac{2^3 15!5!}{20!} = \frac{140}{323} = 0.433$
5	5	0	15	5	$\frac{10!}{5!5!0!}\frac{2^5 15!5!}{20!} = \frac{168}{323} = 0.520$

The p-value is $0.433 + 0.047 = 0.480$. Compare this to the chi-square p-value for $X^2 = 0.40$:

```
> pchisq(0.4,1)
[1] 0.4729107
```

Exact HWE Test Example

For a sample of size $n = 100$ with minor allele frequency of 0.07, there are 8 sets of possible genotype counts:

			Exact		Chi-square	
n_{AA}	n_{Aa}	n_{aa}	Prob.	p value	X^2	p value
93	0	7	0.0000	0.0000*	100.00	0.0000*
92	2	6	0.0000	0.0000*	71.64	0.0000*
91	4	5	0.0000	0.0000*	47.99	0.0000*
90	6	4	0.0002	0.0002*	29.07	0.0000*
89	8	3	**0.0051**	**0.0053***	14.87	0.0001*
88	10	2	0.0602	0.0655	**5.38**	**0.0204***
87	12	1	0.3209	0.3864	0.61	0.4348
86	14	0	0.6136	1.0000	0.57	0.4503

So, for a nominal 5% significance level, the actual significance level is 0.0053 for an exact test that rejects when $n_{Aa} \leq 8$ and is 0.0204 for an exact test that rejects when $n_{AB} \leq 10$.

Modified Exact HWE Test

Traditionally, the p-value is the probability of the data plus the probabilities of all the less-probable datasets. The probabilities are all calculated assuming HWE is true and are conditional on the observed allele frequencies. More recently Graffelman and Moreno showed that the test has a significance value closer to the nominal value if the p-value is half the probability of the data plus the probabilities of all datasets that are less probably under the null hypothesis.

Usual vs Mid p values

				p value	
AA	Aa	aa	$\Pr(n_{Aa}\|n, n_A)$	Usual	Mid
5	5	0	0.520	1.000	0.740
6	3	1	0.433	0.480	0.287
7	1	2	0.047	0.047	0.023

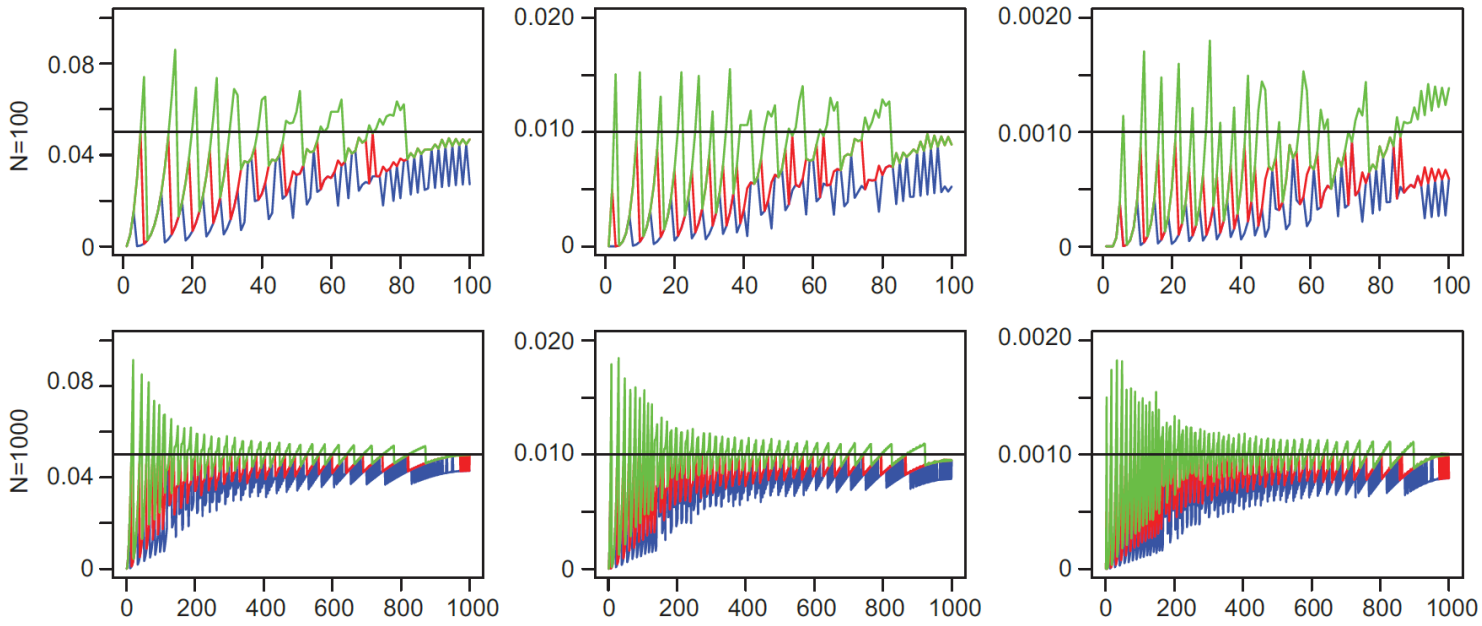
Modified Exact HWE Test Example

For a sample of size $n = 100$ with minor allele frequency of 0.07, there are 8 sets of possible genotype counts:

			Exact		Chi-square	
n_{AA}	n_{Aa}	n_{aa}	Prob.	Mid p value	X^2	p value
93	0	7	0.0000	0.0000*	100.00	0.0000*
92	2	6	0.0000	0.0000*	71.64	0.0000*
91	4	5	0.0000	0.0000*	47.99	0.0000*
90	6	4	0.0002	0.0002*	29.07	0.0000*
89	8	3	0.0051	0.0028*	14.87	0.0001*
88	10	2	**0.0602**	**0.0353***	**5.38**	**0.0204***
87	12	1	0.3209	0.2262	0.61	0.4348
86	14	0	0.6136	0.6832	0.57	0.4503

So, for a nominal 5% significance level, the actual significance level is 0.0353 for an exact test that rejects when $n_{Aa} \leq 10$ and is 0.0204 for a chi-square test that also rejects when $n_{AB} \leq 10$.

Graffelman and Moreno, 2013



Type I error rate against minor allele count for sample sizes 100 and 1000 and significance levels (0.05, 0.01, and 0.001) for exact tests with standard two-sided (red), doubled one-sided (blue) and mid p-values (green).

Power of Exact Test

Calculating the power of an HWE test is easy for the chi-square test statistic as it follows from the non-central chi-square distribution.

It is more complicated for the exact test, and the power depends on the quantity $\psi = P_{Aa}/(\sqrt{P_{AA}P_{aa}})$, involving the genotype probabilities in the population. This quantity depends on both the inbreeding coefficient f and the allele probabilities p_A, p_a in the population.

Power of Exact Test

Once the rejection region has been determined, the power of the test (the probability of rejecting) can be found by adding these probabilities for all sets of genotype counts in the region. HWE corresponds to $\psi = 2$. What is the power to detect HWE when $\psi = 1 (f > 0)$, the sample size is $n = 10$ and the sample allele frequencies are $\tilde{p}_A = 0.75, \tilde{p}_a = 0.25$?

			$\Pr(n_{Aa} \mid n_A, n)$	
n_{AA}	n_{Aa}	n_{aa}	$\psi = 2$	$\psi = 1$
7	1	2	0.047	0.374
6	3	1	0.433	0.364
5	5	0	0.520	0.262

The $\psi = 2$ column shows that the rejection region is $n_{Aa} = 1$, and significance level is 4.7%.

The $\psi = 1$ column shows that the power (the probability $n_{Aa} = 1$ when $\psi = 1$) is 37.4%.

Multiple Testing

When multiple tests are performed, each at significance level α, a proportion α of the tests are expected to cause rejection even if all the hypotheses are true.

Bonferroni correction makes the overall (experimentwise) significance level equal to α by adjusting the level for each individual test to α'. If α is the probability that at least one of the L tests causes rejection, it is also 1 minus the probability that none of the tests causes rejection:

$$\begin{aligned} \alpha &= 1 - (1 - \alpha')^L \\ &\approx L\alpha' \end{aligned}$$

provided the L tests are independent.

If $L = 10^6$, the "genome-wide significance level" is 5×10^{-8} in order for $\alpha = 0.05$.

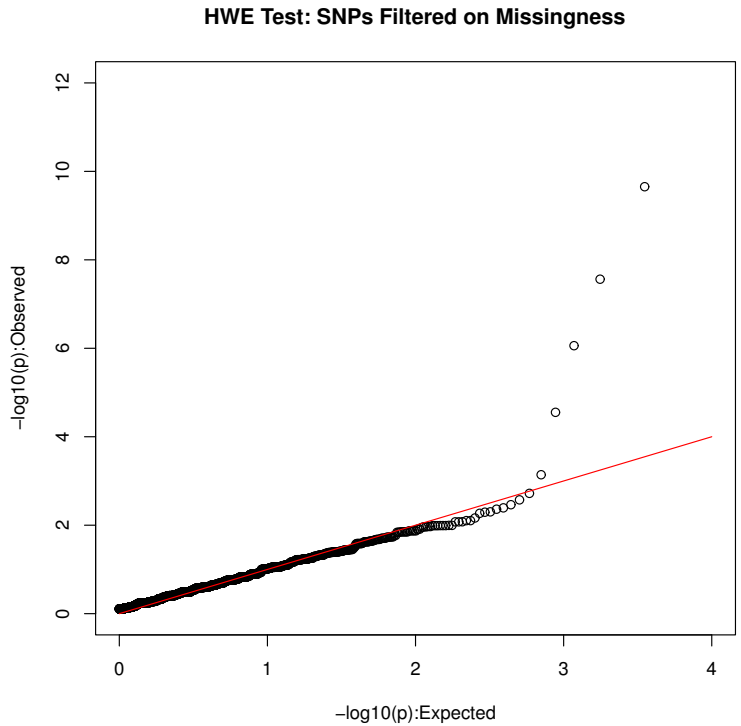
QQ-Plots

An alternative approach to considering multiple-testing issues is to use QQ-plots. If all the hypotheses being tested are true then the resulting p-values are uniformly distributed between 0 and 1.

For a set of n tests, the n p-values are expected to be evenly spread p values between 0 and 1 e.g. $1/2n, 3/2n, \ldots, (2n-1)/2n$. The observed p-values can be plotted against these expected values: the smallest against $1/2n$ and the largest against $(2n-1)/2n$. It is more convenient to transform to $-\log_{10}(p)$ to accentuate the extremely small p values. The point at which the observed values start departing from the expected values is an indication of "significant" values in a way that takes into account the number of tests.

A useful diagnostic for QQ-plots is the "genomic control" quantity λ. This is the ratio of the median of the observed p-values to the median of the expected values. If the expected p-values have a uniform distribution on [0,1], under the null hypothesis of HWE, the median is 0.5. The λ ratio should be 1.

QQ-Plots



There are 7446 SNPs and Bonferroni would reject if $-\log_{10}(p) \geq 5.2$. All five outliers had zero counts for the minor allele homozygote and at least 32 heterozygotes in a sample of size 50.

Linkage Disequilibrium

For alleles A and B are two loci, the usual measure of linkage disequilibrium is

$$D_{AB} = P_{AB} - p_A p_B$$

Whether or not this is zero does not provide a direct statement about linkage between the two loci.

Estimation of LD

With random sampling of gametes, gamete counts have a multinomial distribution and

$$\hat{D}_{AB} = \tilde{P}_{AB} - \tilde{p}_A \tilde{p}_B$$

$$X^2_{AB} = \frac{\hat{D}^2_{AB}}{\text{Var}(\hat{D}_{AB})} \sim \chi^2_{(1)}$$

When $D_{AB} = 0$, $\text{Var}(\hat{D}_{AB}) = p_A(1-p_A)p_B(1-p_B)/n$ and the test statistic is calculated as

$$X^2_{AB} = \frac{n\hat{D}^2_{AB}}{\tilde{p}_A(1-\tilde{p}_A)\tilde{p}_B(1-\tilde{p}_B)}$$

This can be written as $X^2_{AB} = nr^2_{AB}$, by analogy to the test statistic $X^2 = n\hat{f}^2$ for Hardy-Weinberg equilibrium.

Composite Disequilibrium

When genotypes are scored, it is often not possible to distinguish between the two double heterozygotes AB/ab and Ab/aB, so that gametic frequencies cannot be inferred.

Under the assumption of random mating, in which genotypic frequencies are assumed to be the products of gametic frequencies, it is possible to estimate gametic frequencies with the EM algorithm. To avoid making the random-mating assumption, however, it is possible to work with a set of composite disequilibrium coefficients.

Composite Disequilibrium

Although the separate digenic frequencies p_{AB} (one gamete) and $p_{A,B}$ (two gametes) cannot be observed, their sum can be since

$$\begin{aligned} p_{AB} &= P_{AB}^{AB} + \frac{1}{2}P_{Ab}^{AB} + \frac{1}{2}P_{aB}^{AB} + \frac{1}{2}P_{ab}^{AB} \\ p_{A,B} &= P_{AB}^{AB} + \frac{1}{2}P_{Ab}^{AB} + \frac{1}{2}P_{aB}^{AB} + \frac{1}{2}P_{aB}^{Ab} \\ p_{AB} + p_{A,B} &= 2P_{AB}^{AB} + P_{Ab}^{AB} + P_{aB}^{AB} + \frac{P_{ab}^{AB} + P_{aB}^{Ab}}{2} \end{aligned}$$

Digenic disequilibrium is measured with a composite measure Δ_{AB} defined as

$$\begin{aligned} \Delta_{AB} &= p_{AB} + p_{A,B} - 2p_A p_B \\ &= D_{AB} + D_{A,B} \end{aligned}$$

which is the sum of the gametic ($D_{AB} = p_{AB} - p_A p_B$) and nongametic ($D_{A,B} = p_{A,B} - p_A p_B$) coefficients.

Composite Disequilibrium

If the counts of the nine genotypic classes are

	BB	Bb	bb
AA	n_1	n_2	n_3
Aa	n_4	n_5	n_6
aa	n_7	n_8	n_9

the count for pairs of alleles in an individual being A and B, whether received from the same or different parents, is

$$n_{AB} = 2n_1 + n_2 + n_4 + \frac{1}{2}n_5$$

and the MLE for Δ is

$$\hat{\Delta}_{AB} = \frac{1}{n}n_{AB} - 2\tilde{p}_A\tilde{p}_B$$

Composite LD and Allele Dosage

The allele dosage for a SNP is the number of copies of the (say) the reference allele carried by an individual. If A is the reference allele for SNP **A**, then genotypes AA, Aa, aa have dosages X_A of 2,1,0.

The covariance of allele dosages X_A, X_B for loci **A, B** is

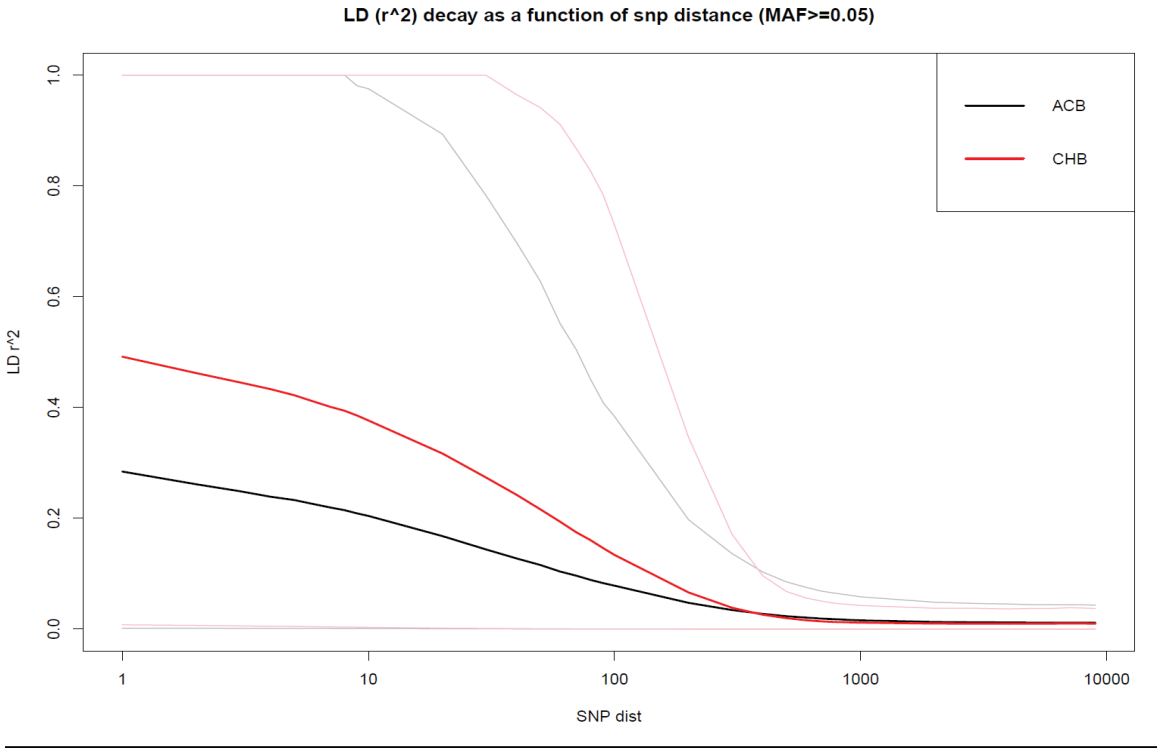
$$\text{Cov}(X_A, X_B) \;=\; 2\Delta_{AB}$$

By analogy to the tests for within-population inbreeding and for gametic linkage disequilibrium, a test statistic for composite LD is

$$X^2_{AB_c} \;=\; nr^2_{AB_c}$$

where r_{AB_c} is the sample correlation coefficient for allele dosages at the two loci over the n individuals in a sample.

1000 Genomes Data



Allele dosage squared correlations for pairs of SNPs on chromosomes 21 and 22 of the 1000 Genomes ACB and populations. Heavy lines: means. Light lines: 5th and 95th percentiles.