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.

Aside: Exact HWE Test

The preferred test for HWE is an exact one. The test rests on the assumption that individuals are sampled randomly from a population so that genotype counts have a multinomial distribution:

$$\Pr(n_{AA}, n_{Aa}, n_{aa}) = \frac{n!}{n_{AA}!n_{Aa}!n_{aa}!}(P_{AA})^{n_{AA}}(P_{Aa})^{n_{Aa}}(P_{aa})^{n_{aa}}$$

This equation is always true, and when there is HWE ($P_{AA} = p_A^2$ etc.) there is the additional result that the allele counts have a binomial distribution:

$$\Pr(n_A, n_a) = \frac{(2n)!}{n_A!n_a!}(p_A)^{n_A}(p_a)^{n_a}$$

Aside: Exact HWE Test

Putting these together gives the conditional probability

$$\begin{aligned}
\Pr(n_{AA}, n_{Aa}, n_{aa} | n_A, n_a) &= \frac{\Pr(n_{AA}, n_{Aa}, n_{aa} \text{ and } n_A, n_a)}{\Pr(n_A, n_a)} \\
&= \frac{\frac{n!}{n_{AA}! n_{Aa}! n_{aa}!}(p_A^2)^{n_{AA}}(2p_A p_a)^{n_{Aa}}(p_a^2)^{n_{aa}}}{\frac{(2n)!}{n_A! n_a!}(p_A)^{n_A}(p_a)^{n_a}} \\
&= \frac{n!}{n_{AA}! n_{Aa}! n_{aa}!}\frac{2^{n_{Aa}} n_A! n_a!}{(2n)!}
\end{aligned}$$

Reject the Hardy-Weinberg hypothesis if this quantity, the probability of the genotypic array conditional on the allelic array, is considered too small to allow that outcome if HWE holds. Is the probability for the data among the smallest of its possible values?

Exact HWE Test Example

For convenience, write the probability of the genotypic array, conditional on the allelic array and HWE, as $\Pr(n_{Aa}|n, n_A)$. Reject the HWE hypothesis for a data set if this value is among the smallest probabilities.

As an example, consider $(n_{AA} = 1, n_{Aa} = 0, n_{aa} = 49)$. The allele counts are $(n_A = 2, n_a = 98)$ and there are only two possible genotype arrays:

AA	Aa	aa	$\Pr(n_{Aa}\|n, n_A)$
1	0	49	$\frac{50!}{1!0!49!}\frac{2^0 2!98!}{100!} = \frac{1}{99}$
0	2	48	$\frac{50!}{0!2!48!}\frac{2^2 2!98!}{100!} = \frac{98}{99}$

The p-value is 1/99=0.01 and HWE is rejected at the 5% level.

Exact HWE Test Example

In this example, $\widehat{f} = 0$ and the chi-square test statistic is $X^2 = 50$. The resulting p-value is 1.54×10^{-12}, substantially different from the exact value of 0.01.

```
> 1-pchisq(50,1,0)
[1] 1.537437e-12
```

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

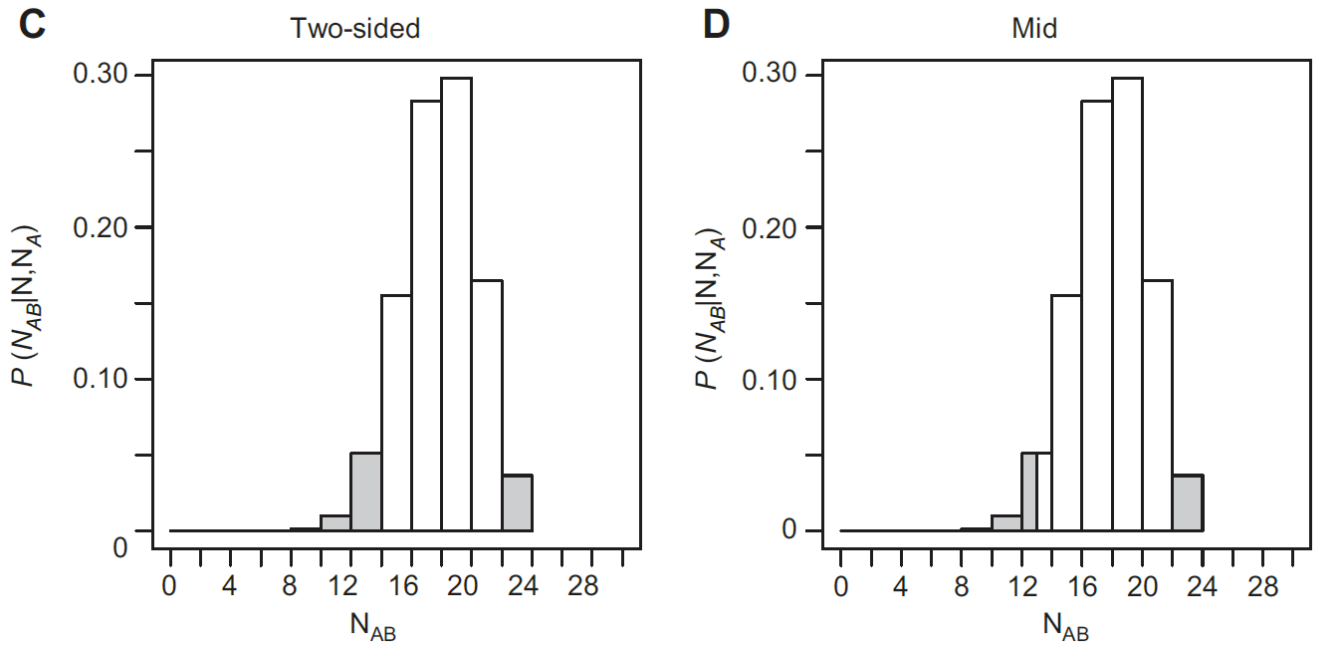
So, for a nominal 5% significance level, the actual significance level is 0.0053 for an exact test that rejects when $n_{Aa} \leq 8$ and is 0.0204 for an exact test that rejects when $n_{AB} \leq 10$.

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. For the $(n_{AA} = 1, n_{Aa} = 0, n_{aa} = 49)$ example then, the p-value is 1/198.

Graffelman J, Moreno V. 2013. Statistical Applications in Genetics and Molecular Biology 12:433-448

Graffelman and Moreno, 2013



Computation of the p-value in an exact test for HWP, for a sample of 50 individuals with a minor allele count of 23, for which 13 heterozygotes were observed. (C) Standard two-sided p-value, (D) Mid p-value based on half the probability of the observed sample.

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

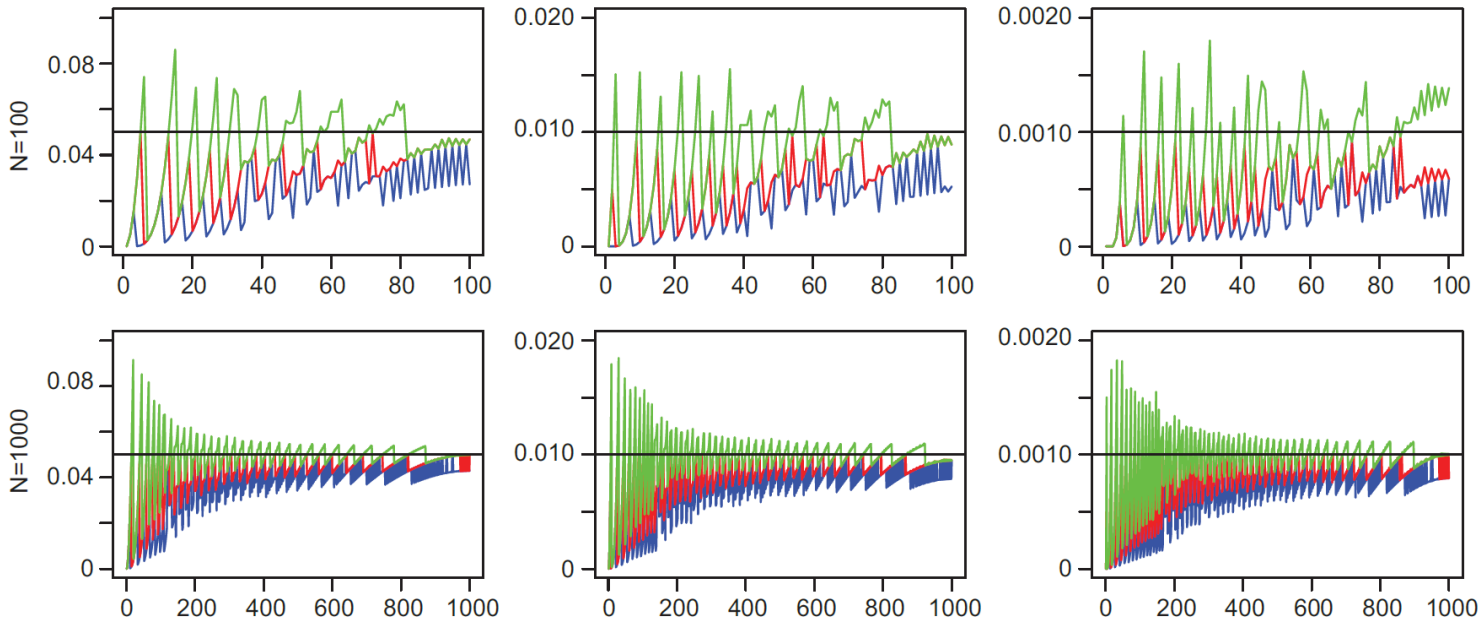
So, for a nominal 5% significance level, the actual significance level is 0.0353 for an exact test that rejects when $n_{Aa} \leq 10$ and is 0.0204 for a chi-square test that also rejects when $n_{AB} \leq 10$.

Effect of Minor Allele Frequency

Even though the nominal significance level for a HWE test may be set at 0.05, for example, the actual significance level can be quite different. (e.g. 0.0353 vs 0.05 on the previous slide.)

The difference between nominal and actual values depends on the sample size and the minor allele frequency, as shown on the next slide.

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.

Aside: Power of exact test

If there is not HWE:

$$\begin{aligned}
\Pr(n_{Aa}|n_A, n_a) &= \frac{n!}{n_{AA}!n_{Aa}!n_{aa}!}(P_{AA})^{n_{AA}}(P_{Aa})^{n_{Aa}}(P_{aa})^{n_{aa}} \\
&= \frac{n!}{n_{AA}!n_{Aa}!n_{AA}!}(P_{AA})^{\frac{n_A - n_{Aa}}{2}}(P_{Aa})^{n_{Aa}}(P_{aa})^{\frac{n_a - n_{Aa}}{2}} \\
&= \frac{n!}{n_{AA}!n_{Aa}!n_{aa}!}\sqrt{P_{AA}^{n_A}}\sqrt{P_{aa}^{n_a}}\left(\frac{P_{Aa}}{\sqrt{P_{AA}P_{aa}}}\right)^{n_{Aa}} \\
&= \frac{C\psi^{n_{Aa}}}{n_{AA}!n_{Aa}!n_{aa}!}
\end{aligned}$$

where $\psi = P_{Aa}/(\sqrt{P_{AA}P_{aa}})$ measures the departure from HWE. The constant C makes the probabilities sum to one over all possible n_{Aa} values: $C = 1/[\sum_{n_{Aa}} \psi^{n_{Aa}}/(n_{AA}!n_{Aa}!n_{aa}!)]$.

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%.

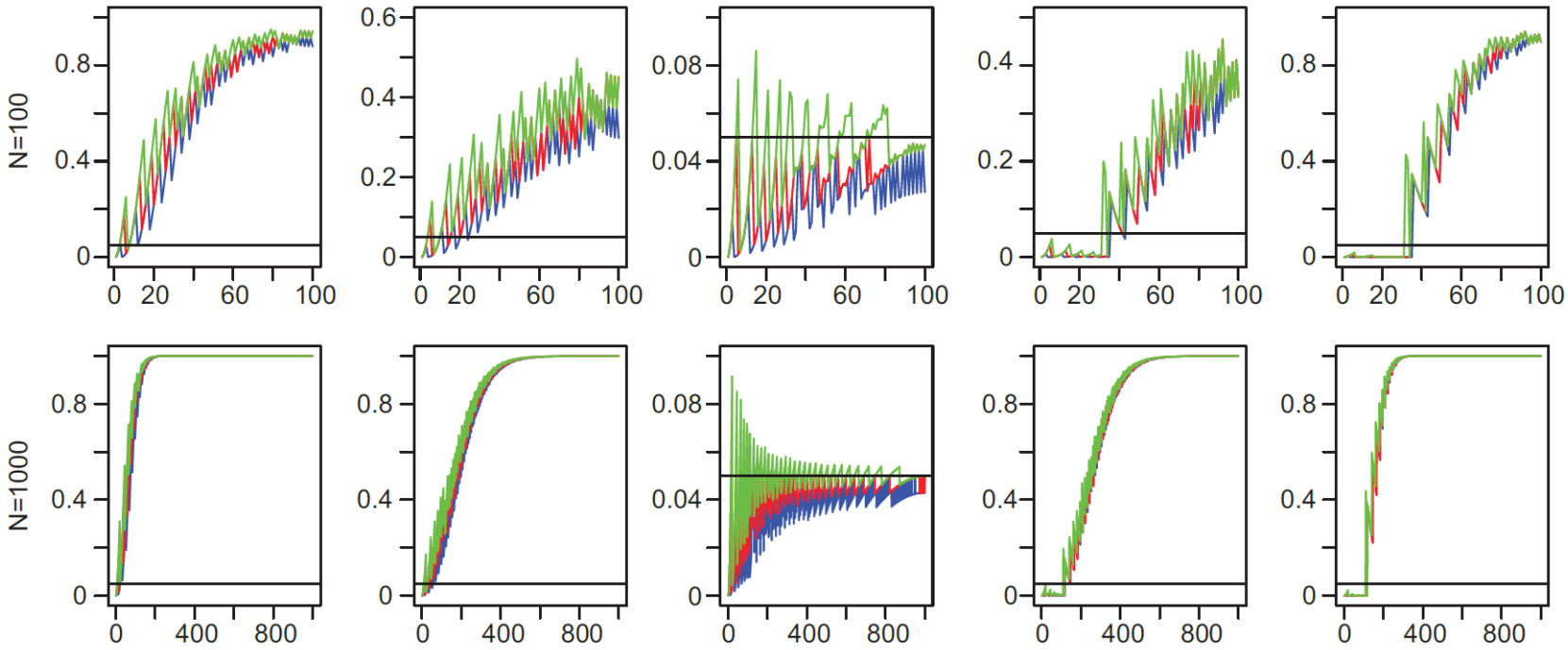
Power Examples

For given values of n, n_a, the rejection region is determined from null hypothesis and the power is determined from the multinomial distribution.

		$\Pr(n_{Aa} \mid n_a = 16, n = 100)$						
	ψ	.250	.500	1.000	2.000	4.000	8.000	16.000
n_{Aa}	f	.631	.398	.157	.000	$-.062$	$-.081$	$-.085$
0		.0042	.0000	.0000	.0000	.0000	.0000	.0000
2		.0956	.0026	.0000	.0000	.0000	.0000	.0000
4		.3172	.0349	.0003	.0000	.0000	.0000	.0000
6		.3568	.1569	.0056	.0000	.0000	.0000	.0000
8		.1772	.3116	.0441	.0008	.0000	.0000	.0000
10		.0433	.3047	.1725	.0123	.0003	.0000	.0000
12		.0054	.1506	.3411	.0974	.0098	.0007	.0000
14		.0003	.0356	.3223	.3681	.1485	.0422	.0109
16		.0000	.0032	.1142	.5214	.8414	.9571	.9890
	Power*	.9943	.8107	.2225	.0131	.0003	.0000	.0000

* $\Pr(n_{Aa} \leq 10)$.

Graffelman and Moreno, 2013



Power ofmHWE exact tests against minor allele count for sample sizes 100 and 1000 and disequilibria 1,2,4,8,16. Standard two-sided (red), double one-sided (blue) and mid p-values (green).

Permutation Test

For large sample sizes and many alleles per locus, there are too many genotypic arrays for a complete enumeration and a determination of which are the least probable 5% arrays.

A large number of the possible arrays is generated by permuting the alleles among genotypes, and calculating the proportion of these permuted genotypic arrays that have a smaller conditional probability than the original data. If this proportion is small, the Hardy-Weinberg hypothesis is rejected.

This procedure is not needed for SNPs with only 2 alleles. The number of possible arrays is always less than about half the sample size.

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 $\alpha = 0.05$.

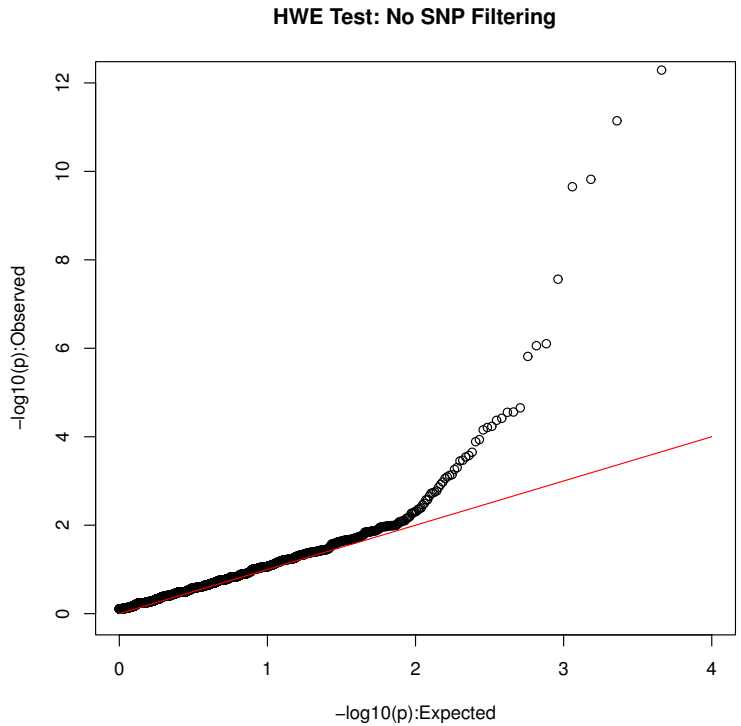
QQ-Plots

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.

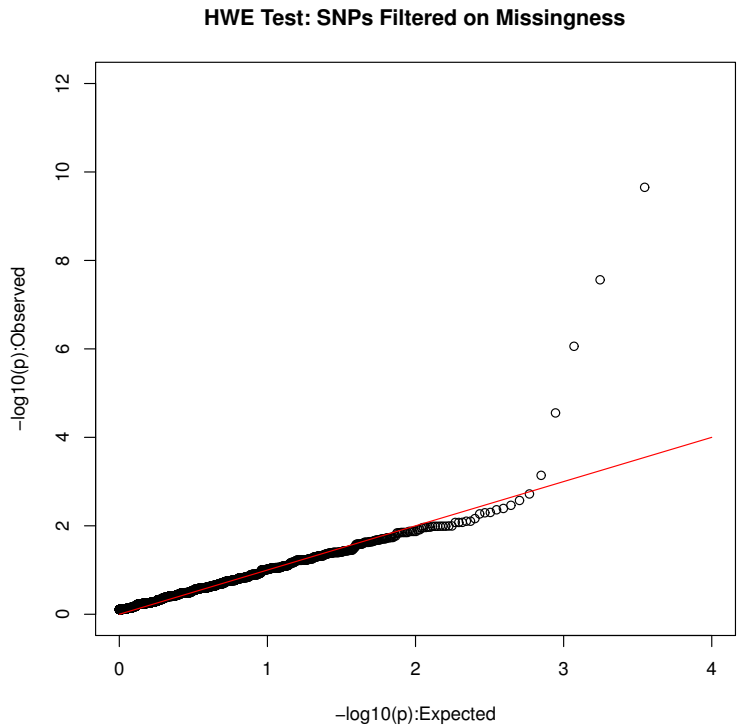
A useful diagnostic for QQ-plots is the "genomic control" quantity λ. This is the ratio of the median of the observed p-values to the median of the expected values. If the expected p-values have a uniform distribution on [0,1], under the null hypothesis of HWE, the median is 0.5. The λ ratio should be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

QQ-Plots



The results for 9208 SNPs on human chromosome 1 for 50 AMD controls ($\lambda = 0.86$). Bonferroni would suggest rejecting HWE when $p \leq 0.05/9208 = 5.4 \times 10^{-6}$ or $-\log_{10}(p) \geq 5.3$.

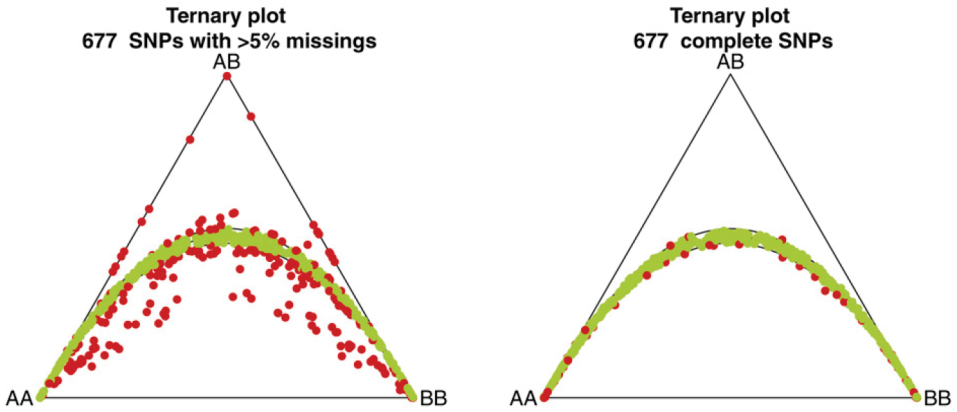
QQ-Plots



The same set of results as on the previous slide except now that any SNP with any missing data was excluded ($\lambda = 1.035$, closer to 1 than for all the SNPs). Now 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.

Imputing Missing Data

Instead of discarding an individual for any SNP when there is no genotype call, it may be preferable to use neighboring SNPs to impute the missing values. Graffelman applied this procedure to a study on pre-term birth:

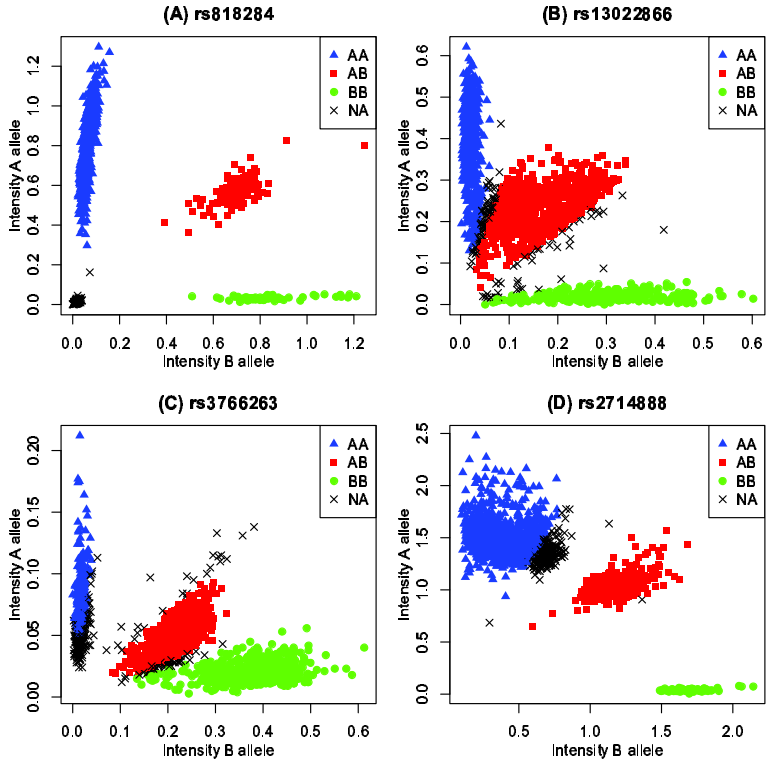


Significant markers are red and nonsignificant markers are green ($\alpha = 0.05$).

Ternary plot: distance of point to side of triangle is frequency of genotype shown on opposite vertex.

Graffelman J, et al. 2015, G3 (Genes, Genomes, Genetics) 5:2365-2373.

Imputing Missing Data



	p-value		
SNP	Discard	Impute	Comment
rs818284	0.000	0.000	Null alleles
rs13022866	0.046	0.571	Het deficiency
rs3766263	0.020	0.539	Het excess
rs2714888	0.192	0.007	Hom deficiency