Categorical data analysis: classical approach.

The multiple test results in GWATCH are created from the original data in form of four P-values and the quantative association statistics (QAS) for codominant, dominant and recessive alternatives, and for allelic approach.

1: The original data

Before starting cflculations of multiple tests the input clynical data should be checked. The phenotype should be a factor of d-levels. Number of the phenotype levels are calculated on the preliminary stage. Denote the common variant is "+" and the minor variant is "−". The genotype levels for the codominant, dominant and recessive alternatives and for the allelic approach are given in the following table.

Test \ Variant	"++"	"+−"	"−−"	Missed
Codominant	0	1	2	3
Dominant	0	1	1	3
Recessive	0	0	1	3
Allelic	$(0,0)$	$(1,0)$	$(1,1)$	$(3,3)$

Under the allelic approach each observaton produces two with the same phenotype and genotype belonging to $\{0,1\}$, so the population size is doubled.

The GWATCH forms $d \times 3$ contingency table $\{n_{ij}\}$ for the codominant alternative and $d \times 2$ contigency table $\{n_{ij}\}$ for other approaches, where n_{ij} are the corresponding counts. Before starting multiple tests GWATCH classify the phenotype: *binary* or *multilevel*. GWATCH removes missed data for each SNP-test separately. For each genetic marker GWATCH calculates P-value of the association test and QAS.

2: The binary phenotype

Currently, GWATCH uses the χ^2 test for *codominant* alternative, and the combined two sided χ^2 and Fisher's exact test by default in other cases. Let $n_{\cdot j} = \sum_i n_{ij}$, $n_{i\cdot} = \sum_j n_{ij}$ and $n = n_\bullet = \sum_{ij} n_{ij}$. For the combined test: the χ^2 test is used under $n_{i\cdot}n_{\cdot j} \geq 5$ for all $i, j \in \{1,2\}$ and Fisher's exact test is used otherwise. The QAS is the Fisher's z-transformation of Pearson's correlation coefficient $z' = \log((1+\rho)/(1-\rho))/2$, where $\rho = \frac{n_{12}+2n_{13}-n_{2\cdot}t/n}{\sqrt{n_{2\cdot}(1-n_{2\cdot}/n)(n_{\cdot 2}+4n_{\cdot 3}-t^2/n)}}$ and $t = n_{\cdot 2} + 2n_{\cdot 3}$, for *codominant* test, and the log odds ratio $lor = \log(n_{11}n_{22}/(n_{21}n_{12}))$ for other tests.

3: The multilevel phenotype

Currently, GWATCH uses the χ^2 test under multilevel phenotype, even in the case of 2×2 table after removing missed data. The P-value is calculated from the obtained contigency table $d' \times l'$ ($d', l' > 1$; $d' \leq d$; $l' \leq 3$ under the codominant alternative and $l' \leq 2$ otherwise) The QAS is the Fisher's z-transformation of Pearson's correlation coefficient $z' = \log((1+\rho)/(1-\rho))/2$, where ρ is the Pearson's correlation coefficient:

$$\rho = \begin{cases} \frac{\sum_i (n_{i2}+2n_{i3})(i-1) - st/n}{\sqrt{\sum_i n_{i\cdot}(i-1)^2 - s^2/n)(n_{\cdot 2}+4n_{\cdot 3}-t^2/n)}}, & \text{under codominant alternative,} \\ \frac{\sum_i n_{i2}(i-1) - st/n}{\sqrt{\left(\sum_i n_{i\cdot}(i-1)^2 - s^2/n\right)n_{\cdot 2}(1-n_{\cdot 2}/n)}}, & \text{otherwise,} \end{cases}$$

where $s = \sum_i n_{i\cdot}(i-1)$ and $t = n_{\cdot 2} + 2n_{\cdot 3}$.